

Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder

R. C. Kessler^{1*}, H. M. van Loo², K. J. Wardenaar², R. M. Bossarte³, L. A. Brenner⁴, D. D. Ebert^{1,5},
P. de Jonge², A. A. Nierenberg⁶, A. J. Rosellini¹, N. A. Sampson¹, R. A. Schoevers², M. A. Wilcox⁷ and
A. M. Zaslavsky¹

¹ Department of Health Care Policy, Harvard Medical School, Boston, MA, USA

² Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

³ Department of Veterans Affairs, Office of Public Health, Washington, DC, USA

⁴ VISN 19 Mental Illness Research Education and Clinical Center, University of Colorado, Anschutz Medical Campus, Anschutz, CO, USA

⁵ Department of Psychology, Clinical Psychology and Psychotherapy, Friedrich-Alexander University Nuremberg-Erlangen, Erlangen, Germany

⁶ Department of Psychiatry and Depression Clinical and Research Program, Harvard Medical School and Massachusetts General Hospital, Boston, MA, USA

⁷ Department of Epidemiology, Janssen Research and Development, Titusville, NJ, USA

Backgrounds. Clinicians need guidance to address the heterogeneity of treatment responses of patients with major depressive disorder (MDD). While prediction schemes based on symptom clustering and biomarkers have so far not yielded results of sufficient strength to inform clinical decision-making, prediction schemes based on big data predictive analytic models might be more practically useful.

Method. We review evidence suggesting that prediction equations based on symptoms and other easily-assessed clinical features found in previous research to predict MDD treatment outcomes might provide a foundation for developing predictive analytic clinical decision support models that could help clinicians select optimal (personalised) MDD treatments. These methods could also be useful in targeting patient subsamples for more expensive biomarker assessments.

Results. Approximately two dozen baseline variables obtained from medical records or patient reports have been found repeatedly in MDD treatment trials to predict *overall* treatment outcomes (i.e., intervention *v.* control) or *differential* treatment outcomes (i.e., intervention A *v.* intervention B). Similar evidence has been found in observational studies of MDD persistence-severity. However, no treatment studies have yet attempted to develop treatment outcome equations using the full set of these predictors. Promising preliminary empirical results coupled with recent developments in statistical methodology suggest that models could be developed to provide useful clinical decision support in personalised treatment selection. These tools could also provide a strong foundation to increase statistical power in focused studies of biomarkers and MDD heterogeneity of treatment response in subsequent controlled trials.

Conclusions. Coordinated efforts are needed to develop a protocol for systematically collecting information about established predictors of heterogeneity of MDD treatment response in large observational treatment studies, applying and refining these models in subsequent pragmatic trials, carrying out pooled secondary analyses to extract the maximum amount of information from these coordinated studies, and using this information to focus future discovery efforts in the segment of the patient population in which continued uncertainty about treatment response exists.

Received 25 November 2015; Accepted 6 January 2016; First published online 26 January 2016

Key words: Depression, epidemiology, evidence-based psychiatry, research design and methods, treatment allocation.

Introduction

Patients with major depressive disorder (MDD) vary substantially in treatment response and illness course. This heterogeneity of treatment effects (HTE)

complicates clinical decision-making. Clinicians have consistently identified the absence of dealing with this variation as a critical gap in personalising MDD treatment (Altshuler *et al.* 2001; Perlis, 2007; Hetrick *et al.* 2011; Kuiper *et al.* 2013). Researchers have tried to address this gap by searching for depression subtypes defined by presumed causes (e.g., postnatal depression) (Cooper & Murray, 1995; Cooper *et al.* 2007), clinical presentations (e.g., atypical or melancholic depression (Fink *et al.* 2007; Uher *et al.* 2011))

* Address for correspondence: Dr R. C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA.
(Email: Kessler@hcp.med.harvard.edu)

or empirically derived symptom profiles (e.g., based on cluster analysis (Andreasen & Grove, 1982), factor analysis (Romera *et al.* 2008) and latent class analysis (Lamers *et al.* 2012)) in hopes of predicting differential treatment response, but results have been disappointing (Baumeister & Gordon, 2012; van Loo *et al.* 2012). More recent efforts have searched for genetic, neuroendocrine, electrophysiological and brain imaging biomarkers of treatment response (Pizzagalli, 2011; Souslova *et al.* 2013; Breitenstein *et al.* 2014; Perlis, 2014), but have failed so far to yield results of sufficient strength to inform clinical decision-making (Simon & Perlis, 2010). Guidelines for MDD treatment selection consequently continue to be based on simple clinical observations about overall MDD severity (National Institute for Health and Clinical Excellence (NICE), 2009; American Psychiatric Association, 2010).

Another promising approach for studying predictors of differential treatment response has received much less attention: to use supervised machine learning methods to develop multivariate prediction equations of treatment outcomes based on symptoms and other easily assessed clinical features that have been found in previous research to predict MDD treatment outcomes (Strobl *et al.* 2009; Zhang & Singer, 2010; van der Laan & Rose, 2011; James *et al.* 2013). Although such methods have been used in this way in other areas of medicine (Chang *et al.* 2012; Chao *et al.* 2012), applications to MDD have so far been based on samples too small to realise the potential of the methods (Andreescu *et al.* 2008a; Rabinoff *et al.* 2011; Riedel *et al.* 2011; Nelson *et al.* 2012; Jain *et al.* 2013). Yet, promising preliminary results exist in clinical (Moos & Cronkite, 1999; Perlis, 2013) and community epidemiological (Angst *et al.* 2011; van Loo *et al.* 2014) studies designed to predict MDD persistence-severity. In addition, innovative statistical methods for building such models exist but have not yet been applied to MDD (Kent *et al.* 2010; van der Laan & Gruber, 2010; Diaz Munoz & van der Laan, 2011; Willke *et al.* 2012; Burke *et al.* 2014; Neugebauer *et al.* 2014). We review these developments in the current report.

Self-report predictors of heterogeneity of MDD treatment effects

We reviewed the literature on self-reported predictors (i.e., assessed by survey or questionnaire; non-biomarker) of MDD treatment response beginning with a PubMed search using the search string: depress* AND predict* AND ('treatment outcome' OR 'treatment response' OR 'course') AND ('self-report' OR 'survey' OR 'questionnaire'). Abstracts were then reviewed and articles read in full if the abstract indicated that: (i) participants underwent treatment for

depression (randomised controlled trials, uncontrolled treatment trials, observational studies in which participants were in a treatment during the follow-up period); and (ii) associations were examined between baseline self-report constructs and MDD treatment outcomes. We also accessed and read any studies that were cited in these papers to have examined baseline self-reported predictors of MDD treatment outcomes. All examined associations were recorded on a spreadsheet. If both bivariate and multivariate models were estimated, we recorded the results from the multivariate models.

Replicated significant associations were found between roughly two dozen baseline self-reported constructs and subsequent MDD treatment outcomes (Table 1). It is noteworthy, though, that the typical study reviewed considered only a handful of these modifiers and no single study included all modifiers. Analyses considering only a small number of modifiers are unlikely to provide reliable clinical guidance due to the existence of many HTE predictors, while more complex subgroup analyses are precluded by the small size of MDD treatment trials (Simon & Perlis, 2010; Cuijpers *et al.* 2012).

Prior attempts to develop models of heterogeneity of MDD treatment effects

While, as noted above, no prior study of MDD HTE has included all the predictors in Table 1, encouraging preliminary results nonetheless exist. The first effort along these lines was that of Perlis (Perlis, 2013), who carried out a secondary analysis of the STAR*D dataset, where MDD treatment response was predicted with an area under the receiver operating characteristic curve (AUC) of 0.71 using a simple logistic regression equation containing a small number of easily accessible patient self-report measures (socio-demographics, depressive symptoms, comorbidity and prior MDD history). An AUC of 0.71 is similar to the levels of prediction accuracy found in a number of widely used risk prediction models in other areas of medicine (Anothaisintawee *et al.* 2012; Siontis *et al.* 2012; Echouffo-Tcheugui & Kengne, 2013). However, this analysis focused on *overall* treatment response rather than *differential* response across multiple treatments.

In comparison, Kraemer (2013) developed an approach to estimate MDD HTE in a treatment trial comparing the relative effectiveness of exactly two treatment types. The three-step approach began by estimating a conventional modifier model for each potential modifier one at a time (i.e., including predictor variables for treatment type (a dummy variable), the modifier and an interaction term between treatment and the modifier). The second step then consisted of

Table 1. Baseline constructs associated with poor overall depression treatment response and/or differential treatment responses in two or more studies

Baseline construct ^a	References demonstrating direct effect on (poor) overall course/treatment response	References demonstrating interaction of baseline construct with treatment type (Treatment A v. Treatment B differential outcome) ^b
Socio-demographics		
Older age	Dew et al. (1997), Fournier et al. (2009), Thase et al. (1997), Vuorilehto et al. (2009)	Cuijpers et al. (2012) (PSY + MED > MED)
Less than full-time employment	Jarrett et al. (2013), Rush et al. (2008), Trivedi et al. (2006)	Fournier et al. (2009) (CT > SSRI)
Low income	Cohen et al. (2009), Cohen et al. (2006), Trivedi et al. (2006)	
Low education	Agid & Lerer (2003), Marquett et al. (2013), Perlis (2013), Szadoczky et al. (2004)	
Non-white race	Perlis (2013), Rush et al. (2008), Trivedi et al. (2006)	
Unmarried	Frank et al. (2011), Rush et al. (2008), Trivedi et al. (2005)	Barber & Muenz (1996) (IPT > CT)
Married/cohabitating		Fournier et al. (2009) (CT > SSRI); Barber & Muenz (1996) (CT > IPT)
Depression history, triggers and symptoms		
High baseline severity	Agid & Lerer (2003), Barber & Muenz (1996), Bernecker et al. (2014), Hollon et al. (2014), Jarrett et al. (2013), Perlis (2013), Rush et al. (2008), Sotsky et al. (1991), Thase et al. (1997), Thase et al. (1994), Vuorilehto et al. (2009)	Luty et al. (2007) (CBT > IPT); Dimidjian et al. (2006) (BA > CT); Elkin et al. (1989) (IPT > CBT); Hollon et al. (2014) (CT + MED > MED); Thase et al. (1997) (IPT + MED > IPT); Thase et al. (2007) (SNRI > SSRI); Kennedy et al. (2006) (Second generation SSRI > Conventional SSRI)
High psychomotor symptoms/low energy	Frank et al. (2011), Perlis (2013)	Frank et al. (2011) (SSRI > IPT)
High anhedonia/melancholic features	Rush et al. (2008), Vrieze et al. (2014)	
Atypical depression		McGrath et al. (1996) (MAOI > TCA); Quitkin et al. (1990) (MAOI > TCA); Quitkin et al. (1991) (MAOI > TCA)
Chronic/double depression/longer episodes	Fournier et al. (2009), Lowe et al. (2005), Sung et al. (2012), Trivedi et al. (2005)	Cuijpers et al. (2012) (PSY + MED > MED, PSY + MED > PSY, MED > PSY)
Endogenous depression	Dew et al. (1997), Ionescu et al. (1994)	
Current/past suicidality	Frank et al. (2011), Rush et al. (2008)	
Greater number prior/recurrent episodes	Carter et al. (2011), Jarrett et al. (2013), Perlis (2013)	
Earlier age of onset	Andreescu et al. (2008b), Dew et al. (1997)	
Comorbid disorders/symptoms		
Personality disorder/symptoms	Hoencamp et al. (1994), Joyce et al. (2007)	Cuijpers et al. (2012) (PSY + MEDS > MEDS); Carter et al. (2011) (CBT > IPT); Joyce et al. (2007) (CBT > IPT); Fournier et al. (2008) (SSRI > CT); Bellino et al. (2008) (SSRI + IPT > SSRI); Kool et al. (2005) (MED + PA > MED); Barber & Muenz (1996) (CT > IPT)
Psychotic disorder/symptoms	Frank et al. (2011), Perlis (2013)	

Continued

Table 1. Continued

Baseline construct ^a	References demonstrating direct effect on (poor) overall course/treatment response	References demonstrating interaction of baseline construct with treatment type (Treatment A v. Treatment B differential outcome) ^b
Anxiety disorder/symptoms	Andreescu <i>et al.</i> (2007), Andreescu <i>et al.</i> (2008b), Barber & Muenz (1996), Cohen <i>et al.</i> (2009), Dew <i>et al.</i> (1997), Frank <i>et al.</i> (2011), Perlis (2013), Rush <i>et al.</i> (2008), Smits <i>et al.</i> (2012), Trivedi <i>et al.</i> (2006)	Papakostas <i>et al.</i> (2008) (SSRI > AAD)
Substance use/abuse	Trivedi <i>et al.</i> (2006)	Rush <i>et al.</i> (2008) (SSRI > SNRI)
Sleep problems	Andreescu <i>et al.</i> (2008b), Dew <i>et al.</i> (1997), Troxel <i>et al.</i> (2012)	
High somatic anxiety	Feske <i>et al.</i> (1998)	Frank <i>et al.</i> (2011) (IPT > SSRI)
Stress and adversity		
Childhood maltreatment/trauma	Carter <i>et al.</i> (2011), Johnstone <i>et al.</i> (2009), Nanni <i>et al.</i> (2012)	Nemeroff <i>et al.</i> (2003) (CBT > SSRI)
High stress	Dew <i>et al.</i> (1997)	Fournier <i>et al.</i> (2009) (CT > SSRI)
Poor physical functioning	Lowe <i>et al.</i> (2005), Trivedi <i>et al.</i> (2006)	
Poor social support/relationships	Denton <i>et al.</i> (2010), Dew <i>et al.</i> (1997), Marquett <i>et al.</i> (2013)	
Personality traits		
Non-secure attachment style	Constantino <i>et al.</i> (2013)	McBride <i>et al.</i> (2006) (CBT > IPT)
High negative affect/neuroticism	Joyce <i>et al.</i> (2007), Szadoczky <i>et al.</i> (2004), Vrieze <i>et al.</i> (2014)	Bagby <i>et al.</i> (2008) (SSRI > CBT)
Low openness to experience	Bagby <i>et al.</i> (2008), Marquett <i>et al.</i> (2013)	
Other		
High impairment	Feske <i>et al.</i> (1998), Frank <i>et al.</i> (2011), Jarrett <i>et al.</i> (2013), Lowe <i>et al.</i> (2005), Sotsky <i>et al.</i> (1991), Trivedi <i>et al.</i> (2006)	
High cognitive dysfunction/low intelligence	Fournier <i>et al.</i> (2009)	Cuijpers <i>et al.</i> (2012) (PSY + MED > MED)

IPT, interpersonal psychotherapy; CT, cognitive therapy; CBT, cognitive-behavioural therapy; BA, behavioural activation; PA, psychoanalysis; PSY, psychotherapy (non-specific); SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin–norepinephrine reuptake inhibitors; TCA, tricyclic antidepressant; AAD, atypical antidepressant (e.g., bupropion); MAOI, monoamine oxidase inhibitor; MED, pharmacotherapy (non-specific); +, combined treatment.

^aPredictor and outcome measures varied by study, and only constructs with statistically significant ($p < 0.05$) associations with depression treatment outcome (overall or differential response) in two or more studies are presented here.

^bDifferential treatment response depending on the baseline construct is shown in parentheses. Treatment type is operationalised based on broad classes of psychotherapy (e.g., any PSY, IPT, CT, CBT, BA, PA) and pharmacotherapy (e.g., any MED, SSRI, SNRI, TCA, AAD, MAOI). The treatment associated with the better response (among patients with the baseline construct) is listed before the >. In other words, $X > Y$ means that treatment X is favoured relative to treatment Y if the (row) construct is present at baseline.

estimating a multivariate model to create regression weights for all modifiers judged to be important in the first step. These modifiers were then combined in a third step into a single composite HTE measure for each patient by summing the products $b_m \times M_{im}$ where b_m is the slope of the treatment outcome on modifier M and M_{im} is the score of respondent i on

modifier M . 'Importance' of individual modifiers in the first step was defined by the standardised correlation between modifier scores and differences in treatment effect across the two treatment types, where the latter association was estimated in a person-pair dataset for all $n_1 \times n_2$ pairs of patients in either Treatment A (n_1) or Treatment B (n_2).

In an illustration of this approach applied to a small treatment trial in which patients were randomised to receive either interpersonal therapy (IPT) or a selective serotonin reuptake inhibitor (SSRI) and 32 potential treatment effect modifiers were assessed at baseline, Kraemer showed that even though the overall effect size was approximately 0 (i.e., patients responded equivalently in the aggregate to the two treatments), the effect size was 0.50 favouring IPT over SSRI in the segment of the sample in which a composite HTE score (made up of 8 of the 32 original baseline measures) favoured IPT (representing 44% of the patients in this particular trial) and 0.48 favouring SSRI over IPT in the remainder of the sample. (See also Wallace *et al.* 2013 for a more substantive presentation of the same results.)

While this illustration makes it clear that baseline information could be of great value in guiding clinical decision-making about MDD treatment selection, it is also important to point out that the Kraemer approach to defining individual-level HTE is limited in that it provides no practical way to estimate an optimal clinical decision support model for choosing among the wider range of treatments available for MDD (e.g., IPT, cognitive therapy (CT), behavioural activation, cognitive-behavioural therapy, or some other type of psychotherapy; SSRI, serotonin-norepinephrine reuptake inhibitor (SNRI) or some other type of pharmacotherapy; any combination of a particular psychotherapy with a particular pharmacotherapy). Nor does the Kraemer approach allow for the estimation of stable models that make use of the large number of potential modifiers, some of which might be highly inter-correlated, in ways that consider the possible existence of complex non-linear and/or non-additive multivariate associations (e.g., three-way interactions) with response to particular types of treatment.

DeRubeis *et al.* (2014) proposed an approach to MDD HTE estimation very similar to the Kraemer approach in that it began by estimating a conventional modifier model for each potential modifier one at a time. However, it differed from the Kraemer approach in that subsequent steps of model-building that combined important modifiers (where 'important' was defined initially as significant at the 0.20 level when modifiers were considered one at a time, at the 0.10 level when included in subsequent within-domain multivariate models, and at the 0.05 level when included in final cross-domain models) were carried out at the person level rather than, as in the Kraemer approach, at the person-pair level. This person-level analysis allowed DeRubeis to generate a predicted treatment outcome score for each patient based on the final model separately for the actual type of treatment received as well as based on the counter-factual

assumption that the patient had received another type of treatment. Individual-level comparison of these two predicted scores then allowed DeRubeis to determine the preferred treatment for each patient.

In an application of this approach to a small treatment trial in which patients were randomised to receive either antidepressant medication or cognitive behaviour therapy (CBT), interactions of treatment type with 38 potential baseline modifiers (as detailed in Fournier *et al.* 2009) were estimated initially one at a time and then in sequential multivariate models to arrive at a final model that included nine significant (0.05 level in third-step models) predictors either having interactions with type of treatment (five predictors) or associated with treatment outcome equivalently for both types of treatment (four predictors). Roughly 60% of patients had predicted outcome scores based on the model that differed between the two types of treatment by an amount considered clinically significant (three points on the Hamilton Rating Scale for Depression). In the aggregate, patients in this 60% of the sample who were randomised to the type of treatment to which they were predicted by the model to have better response had a treatment effect size 0.58 greater than that of patients who were randomised to the other treatment.

This research team subsequently applied the same method to another small treatment trial that randomised patients with MDD either to IPT or CT (Huibers *et al.* 2015). A total of 43 potential baseline modifiers were available and the final model included 13 of them (eight with significant interactions and five others that were associated with treatment outcomes equivalently across the two treatments). In the aggregate, patients who were randomised to the type of treatment to which they were predicted by the model to have better response had a treatment effect size 0.51 greater than that of patients who were randomised to the other treatment.

As with the Kraemer study, the two studies by DeRubeis and colleagues illustrate the potential value of using baseline information to help clinicians select personalised MDD treatments. As with the Kraemer study, though, it is quite likely that the method used by DeRubeis and colleagues would lead to model overfitting; that is, to a situation in which application of the models in independent patient datasets would lead to prediction accuracy being lower, perhaps substantially so, than in the sample in which the models were built. As we discuss later, machine learning methods are designed to address this problem of overfitting. The *ad hoc* stepwise model-building procedures used by Kraemer and DeRubeis are far inferior to methods designed explicitly to maximise prediction accuracy in independent samples.

It is noteworthy in this regard that the last step of the DeRubeis approach used the leave-one-out (LOO) method to impute individual-level predicted outcome scores. DeRubeis and colleagues asserted that this method addressed the problem of overfitting, but this assertion is incorrect. Overfitting almost certainly occurred at the level of variable selection, where step-wise analysis was used to create models with interactions involving 5 of 38 (the CBT *v.* SSRI trial) and 8 of 43 (the IPT *v.* CT trial) initially examined predictors in very small treatment samples ($n=154$ in the CBT *v.* SSRI trial; $n=134$ in the IPT *v.* CT trial). This means that any attempt to use the coefficients in these models to predict differential treatment response in a new sample of patients would almost certainly yield less positive effects than those suggested by the results of studies. The use of the LOO method to estimate the likely strength of the model in an independent sample is an invalid approach when LOO is applied *after* selection of the final model predictors. At the level of variable selection, furthermore, use of LOO is widely recognised to be suboptimal compared with other types of cross-validation due to the fact that it has high variance (Hastie *et al.* 2009).

Is there a better way?

Best practices recommendations for HTE analysis call for a different approach. In the simple case of a single treatment *v.* control evaluation, these recommendations call for a three-step approach: (i) estimate the joint effects of baseline predictors in multivariate prediction equations applied either to an independent sample of people with the disorder (Kent *et al.* 2010) or, if the clinical trial sample is large enough, to the control group of the trial in which the predictors are being studied (Burke *et al.* 2014); (ii) apply the predicted probabilities of treatment outcomes from these equations to both intervention and control patients; and (iii) plot treatment outcomes separately in the intervention and control groups to examine differences in absolute risk reduction (ARR) as a function of these predicted probabilities. Patients with high predicted probabilities of recovery will have low ARR because they will recover even without treatment. Patients with low probabilities of recovery might also have low ARR due to available treatments being ineffective in these difficult cases. Depending on the proportions of patients at these tails of the distribution, the trial might be negative overall even though ARR is significant among patients with intermediate predicted probabilities of recovery. This approach to multivariate HTE analysis has proven useful in guiding personalised treatment planning in other areas of medicine (Hayward *et al.* 2006; Dorresteijn *et al.* 2011) even

though the prediction equations have largely focused exclusively on *overall* treatment response rather than *differential* treatment response.

An expansion of this approach to HTE involving multiple types of treatment would either require estimation: (i) of a separate model for each type of treatment *v.* controls; (ii) of a separate within-treatment model for each type of treatment; or (iii) of a pooled model across active comparator treatments that allowed for interactions of dummy variables for treatment type with baseline variables. In the ideal case, these models would be estimated using modern machine learning methods rather than the *ad hoc* methods used by Kraemer and DeRubeis in order to reduce the problem of overfitting and maximise out-of-sample performance when applied to independent patient samples (Ritchie, 2005; Upstill-Goddard *et al.* 2013).

The results of such models could be applied in subsequent patient samples by comparing estimated treatment outcomes for each patient separately for each treatment option to arrive at an estimate of the optimal treatment for each patient. In order to do this, though, the logic would require a large enough trial to obtain stable coefficient estimates within treatment-specific subsamples and the application of the coefficients from that trial to subsequent trials. This is infeasible in the case of MDD treatment trials, though, because MDD treatment trials are too small to support such an analysis. Another problem is that MDD treatment trials do not use a stable set of baseline measures of the sort outlined in Table 1. The problem of small sample size is largely responsible for the fact that Kraemer, in the approach described above, carried out the analysis using patient-pair data, as the approach needed to conserve degrees of freedom in a sample that consisted of only $n=291$ patients randomised between two conditions and the analysis examined the modifying effects of 32 baseline predictors. And the problem of inconsistency in baseline measure is largely responsible for the fact that no efforts have been made to pool results across a large number of MDD treatment trials to estimate complex MDD HTE models.

This problem of small sample size has been addressed in other areas of medicine either by developing interactive HTE models based on very large trials (e.g., the *Use of Statins in Prevention* trial, which randomised 17 802 initially healthy men and women to statins or placebo for 10 years to evaluate the effect of early statin use in preventing cardiac events (Dorresteijn *et al.* 2011) or by using previously developed external risk scores based on prediction equations developed either in large observational samples or in pooled samples that combine data across the multiple observational studies and/or clinical trials (Perel *et al.* 2006; Prieto-Merino & Pocock, 2012).

The latter would be the more practical approach for MDD HTE, possibly beginning with large observational samples of patients beginning MDD treatment, administering self-report surveys of constructs found previously to predict MDD HTE, and following these patients through treatment to assess treatment outcomes. These data would then be analysed using statistical methods recently developed to estimate comparative treatment effectiveness in observational studies (van der Laan & Gruber, 2010; Neugebauer *et al.* 2014). Inspection of between-patient differences in predicted outcomes pooled across all treatments could be used to study individual differences in *overall* treatment outcomes, while inspection of between-patient differences in treatments associated with highest predicted probabilities of recovery could be used to study individual differences in *differential* treatment outcomes. Internal cross-validation could be used to evaluate in-sample performance. If clinically meaningful individual differences were documented in this way, the same approach could be used in subsequent MDD clinical trials to support HTE analyses.

The above approach would need to begin with large thoughtfully constructed (Madigan *et al.* 2014) *observational* samples because the sample sizes of even the largest MDD clinical trials would be much too small to provide stable estimates of predicted HTE (Madigan *et al.* 2014). Although HTE estimates are biased in observational studies if treatment assignment is informatively non-random, statistical methods exist to adjust for this bias (Picciotto *et al.* 2014). This is true even for non-random variation in dynamic treatment assignment (e.g., due to side effects or lack of early treatment response) (Suarez *et al.* 2008; Liu *et al.* 2014) and for unmeasured determinants (Lin & Chen, 2014; Tchetgen Tchetgen, 2014). Consistent with these observations, a recent Cochrane review concluded that treatment effect size estimates based on well-analysed observational studies are very similar to those based on randomised controlled trials (Anglemyer *et al.* 2014).

Preliminary results

Although we are aware of no existing efforts to develop a multivariate model of MDD treatment response in a clinical trial sample along the lines suggested above, a potentially useful model can be found in a series of studies designed to examine multivariate predictors of long-term depression persistence-severity in secondary analyses of the 1990–1992 National Comorbidity Survey (NCS; Kessler *et al.* 1994), 2001–2003 NCS follow-up survey (NCS-2; Kessler *et al.* 2003), 2001–2003 NCS Replication (NCS-R; Kessler *et al.* (2004)), and WHO World

Mental Health (WMH) surveys (Demyttenaere *et al.* 2004). We briefly review the results of these studies in this section of the paper and then discuss prospects for extending the methods used to examine of MDD HTE.

The NCS and NCS-R were nationally representative community epidemiological surveys of common mental disorders in the USA. The NCS-2 was a follow-up survey of NCS 10–12 years after baseline. The WMH surveys were national or regional surveys based on NCS-R in 15 other countries. Initial exploratory analyses based on unsupervised clustering found patterns suggesting that significant associations existed between retrospective reports about incident episode symptoms and subsequent illness course in the NCS-R data. These results were sufficiently promising that subsequent supervised machine learning analyses were carried out to maximise the prediction of MDD persistence-severity from retrospectively reported information on incident episode symptoms in the much larger WMH series, where there were 8261 respondents with lifetime DSM-IV MDD (van Loo *et al.* 2014; Wardenaar *et al.* 2014).

Two machine learning algorithms (ensemble recursive partitioning, penalised regression) were used to examine associations of the outcomes with predictors that consisted of retrospectively reported parental history of depression, temporally primary comorbid disorders, and characteristics of incident MDD episodes. The outcomes were two measures of retrospectively reported subsequent MDD persistence (number of years with episodes and with episodes lasting most days throughout the year) and two measures of subsequent MDD severity (hospitalisation; work disability). K-means cluster analysis of the four predicted values found three risk strata that parsimoniously characterised multivariate associations. The high-risk cluster (32.4% of cases) accounted for 56.6–72.9% of high persistence-severity, with area under the receiver operating characteristic curve (AUC) of 0.63–0.70.

As these WMH results were retrospective, a validation study was subsequently undertaken in the NCS/NCS-2 panel. Predicted outcome scores were generated from information collected in the baseline survey scored using model coefficients estimated in the WMH analysis. Associations of these predicted values with outcomes over the intervening 10–12 years were then examined using reports obtained in the NCS-2 follow-up survey. These prospective associations were comparable to the retrospective associations found in WMH (Kessler *et al.* 2016). Importantly, meaningful discrimination was found both at the upper and lower ends of the predicted outcome distributions. For example, the respondents classified at baseline as being in the top quintile of risk accounted

for 55.8% of all suicide attempts over the subsequent 10–12 years, while the respondents in the lowest baseline risk quintile accounted for only 1.5% of subsequent suicide attempts.

It is instructive to compare these NCS/NCS-2 results to those of other prospective studies that used baseline measures to predict MDD persistence-severity over 10+ years in samples of initially depressed patients (Moos & Cronkite, 1999; Mueller *et al.* 1999; Klein *et al.* 2008; Cronkite *et al.* 2013) or community residents (Mattisson *et al.* 2007; Bradvik *et al.* 2008; Eaton *et al.* 2008; Fichter *et al.* 2010; Angst *et al.* 2011). Although these studies were all quite small ($n = 87\text{--}424$) and none reported AUC, AUC could be computed *post hoc* from two of them. The first study was a 50-year follow-up of the 191 respondents in the Lundby community study with baseline MDD, 20 of whom subsequently died by suicide (Bradvik *et al.* 2008). A composite measure of baseline depression severity predicted subsequent suicide with $AUC = 0.69$ compared with $AUC = 0.70$ for the most comparable NCS-2 outcome (attempted suicide). The second study followed 313 depressed outpatients 1, 4 and 10 years after baseline and defined *chronic* depression as either (i) meeting full criteria for MDD at any 2 follow-ups or (ii) meeting full criteria at the 10-year follow-up and partial criteria at both earlier assessments (Moos & Cronkite, 1999). Twenty baseline predictors (depressive symptoms, self-concept, social function and coping) predicted chronicity with $AUC = 0.70$ compared to $AUC = 0.66$ for the most comparable NCS-2 outcome (high persistence of episodes). In making these comparisons, it is important to remember that the AUCs in these other studies were not validated in independent samples.

It is also noteworthy that the predictors in the WMH and NCS/NCS-2 studies as well as in the above studies were much less comprehensive than those in Table 1. This means that the estimates of prediction strength in these studies are likely to be lower bounds. A preliminary expansion of the NCS/NCS-2 analysis to go beyond the incident episode predictors considered so far and include all the predictors in the baseline survey listed in Table 1. While still only a subset of all the predictors in Table 1, AUC increased to more than 0.80 for each NCS-2 outcomes when the predictors were expanded in this way.

Implications for developing models of heterogeneity of MDD treatment effects

Given the above results, one potentially useful next step in studying MDD HTE would be to develop a self-report questionnaire based on the predictors in Table 1, administer that questionnaire to large observational samples of patients at the beginning of MDD

treatment, monitor treatment types and responses, and analyse these data to generate predicted MDD treatment outcome scores that could be used as the basis of HTE analyses in subsequent clinical trials. If many different researchers carrying out prospective observational studies and controlled MDD treatment trials used a consistent questionnaire of this type, results could be pooled to predict HTE. There is precedent for this kind of pooling of observational and controlled studies to study consistency of estimated treatment effects and the roles of observational study confounding, compositional differences and variation in treatments in accounting for between-study discrepancies (Prentice *et al.* 2006; Toh & Manson, 2013).

Another possible extension would be to carry out subsequent pragmatic trials (Lurie & Morgan, 2013) in the same treatment systems where prior observational studies were carried out by randomising participating clinicians either to receive or not to receive actuarial information about optimal treatments based on HTE models for individual patients based on administration of questionnaires prior to initiation of treatment. The treatment outcomes of the patients included in this randomisation could then be tracked to evaluate the effects of making this personalised clinical decision support tool available to clinicians. These predictions could also be used to determine which patients should be targeted for randomisation to interventions involving expensive biomarkers (Uher *et al.* 2010; Williams *et al.* 2011; Dunlop *et al.* 2012; Kennedy *et al.* 2012; Wallace *et al.* 2013) that would only be needed if the actuarial model based on self-report questionnaire data yielded equivocal results (Van Staa *et al.* 2012). It would also be valuable in this context to evaluate the *incremental* value of promising biomarkers in improving prediction beyond the level achieved less expensively using only self-report data (Li & Lu, 2010; Steyerberg *et al.* 2014).

Conclusions

Significant associations exist between numerous self-report measures and subsequent MDD outcomes. These associations have been documented both in the controlled treatment trials, where the outcomes were measures of treatment response, and in observational studies, where the outcomes were more general measures of MDD persistence-severity. Although no large-scale prospective study has been carried out to evaluate the joint effects of all these predictors at once, the preliminary results reviewed above make a good case that the resulting multivariate equations might be of clinical value in predicting both absolute and differential treatment response. The use of recent advances in machine learning methods to detect

interactions could be valuable in refining these equations, while the use of recent advances in statistical methods to make causal inferences from observational data could help reduce bias in estimating HTE due to non-random treatment assignment and informative loss to follow-up. These equations could then be used both to generate individual-level predicted outcome scores to support the investigation of MDD HTE in subsequent controlled treatment trials and provide useful decision support for clinicians attempting to optimise the treatment of their depressed patients.

Acknowledgements

None.

Financial Support

The NCS data collection was supported by the National Institute of Mental Health (NIMH; R01MH46376). The NCS-2 data collection was supported by the National Institute on Drug Abuse (NIDA; R01DA012058). Dr de Jonge was supported by a VICI grant (no. 91812607) from the Netherlands organisation for Scientific research (NWO-ZonMW). The NCS surveys were carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative, which was supported by the National Institute of Mental Health (R01MH070884), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, the US Public Health Service (R13MH066849, R01MH069864 and R01DA016558), the Fogarty International Center (FIRCA R03TW006481), the Pan American Health Organization, Eli Lilly and Company, Ortho-McNeil Pharmaceutical, Inc., GlaxoSmithKline, and Bristol-Myers Squibb. Dr de Jonge was supported by a VICI grant (no. 91812607) from the Netherlands organization for Scientific research (NWO-ZonMW). Preparation of this report was supported by Janssen Pharmaceuticals.

Conflict of Interest

In the past 3 years, Dr Kessler has been a consultant for Johnson & Johnson Wellness and Prevention, Shire Pharmaceuticals, and the Lake Nona Institute. Dr Nierenberg has been a consultant for Abbott Laboratories, American Psychiatric Association, Appliance Computing Inc. (Mindsite), Basliea, Brain Cells, Inc., Brandeis University, Bristol Myers Squibb, Clintara, Corcept, Dey Pharmaceuticals, Dainippon Sumitomo (now Sunovion), Eli Lilly and Company, EpiQ, L.P./Mylan Inc., Forest, Genaissance,

Genentech, GlaxoSmithKline, Hoffman LaRoche, Infomedic, Lundbeck, Janssen Pharmaceutica, Jazz Pharmaceuticals, Medavante, Merck, Methylation Sciences, Naurex, Novartis, PamLabs, Pfizer, PGx Health, Ridge Diagnostics Shire, Schering-Plough, Somerset, Sunovion, Takeda Pharmaceuticals, Targacept, and Teva; consulted through the MGH Clinical Trials Network and Institute (CTNI) for Astra Zeneca, Brain Cells, Inc, Dianippon Sumitomo/Sepracor, Johnson and Johnson, Labopharm, Merck, Methylation Science, Novartis, PGx Health, Shire, Schering-Plough, Targacept and Takeda/Lundbeck Pharmaceuticals; had grant/research support from the American Foundation for Suicide Prevention, AHRQ, Brain and Behavior Research Foundation, Bristol-Myers Squibb, Cederroth, Cephalon, Cyberonics, Elan, Eli Lilly, Forest, GlaxoSmithKline, Janssen Pharmaceutica, Lichtwer Pharma, Marriott Foundation, Mylan, NIMH, PamLabs, PCORI, Pfizer Pharmaceuticals, Shire, Stanley Foundation, Takeda, and Wyeth-Ayerst; received honoraria from Belvoir Publishing, University of Texas Southwestern Dallas, Brandeis University, Bristol-Myers Squibb, Hillside Hospital, American Drug Utilization Review, American Society for Clinical Psychopharmacology, Baystate Medical Center, Columbia University, CRICO, Dartmouth Medical School, Health New England, Harold Grinspoon Charitable Foundation, IMEDEX, International Society for Bipolar Disorder, Israel Society for Biological Psychiatry, Johns Hopkins University, MJ Consulting, New York State, Medscape, MBL Publishing, MGH Psychiatry Academy, National Association of Continuing Education, Physicians Postgraduate Press, SUNY Buffalo, University of Wisconsin, University of Pisa, University of Michigan, University of Miami, University of Wisconsin at Madison, APSARD, ISBD, SciMed, Slack Publishing and Wolters Kluwer Publishing; owns stock in Appliance Computing, Inc. (MindSite); Brain Cells, Inc., Medavante; and owns the following copyrights: Clinical Positive Affect Scale and the MGH Structured Clinical Interview for the Montgomery Asberg Depression Scale exclusively licensed to the MGH Clinical Trials Network and Institute (CTNI). Dr Wilcox is an employee of Janssen Pharmaceuticals. The remaining authors report nothing to disclose. The remaining authors report nothing to disclose.

The funders/sponsors were not involved in any aspect of the design and conduct of the study other than for the participation of Dr Wilcox as a scientific collaborator.

Ethical Standard

The authors assert that all procedures contributing to this work comply with the ethical standards of the

relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Disclaimer

The views, opinions and/or findings contained in this article are those of the author(s) and should not be construed as an official Department of Veterans Affairs position, policy, or decision unless so designated by other documentation, nor should they be construed to represent the views of any of the sponsoring organizations, agencies, or US Government.

Additional Information

A complete list of NCS and NCS-2 publications can be found at <http://www.hcp.med.harvard.edu/ncs>. A complete list of WMH publications can be found at <http://www.hcp.med.harvard.edu/wmh>

References

- Agid O, Lerer B** (2003). Algorithm-based treatment of major depression in an outpatient clinic: clinical correlates of response to a specific serotonin reuptake inhibitor and to triiodothyronine augmentation. *The International Journal of Neuropsychopharmacology/Official Scientific Journal of the Collegium Internationale Neuropsychopharmacologicum* **6**, 41–49.
- Altshuler LL, Cohen LS, Moline ML, Kahn DA, Carpenter D, Docherty JP, Ross RW** (2001). Treatment of depression in women: a summary of the expert consensus guidelines. *Journal of Psychiatric Practice* **7**, 185–208.
- American Psychiatric Association** (2010). *Practice Guideline for the Treatment of Patients with Major Depressive Disorder*. American Psychiatric Association: Arlington, VA.
- Andreasen NC, Grove WM** (1982). The classification of depression: traditional *versus* mathematical approaches. *The American Journal of Psychiatry* **139**, 45–52.
- Andrescu C, Lenze EJ, Dew MA, Begley AE, Mulsant BH, Dombrovski AY, Pollock BG, Stack J, Miller MD, Reynolds CF** (2007). Effect of comorbid anxiety on treatment response and relapse risk in late-life depression: controlled study. *The British Journal of Psychiatry: The Journal of Mental Science* **190**, 344–349.
- Andrescu C, Chang CC, Mulsant BH, Ganguli M** (2008a). Twelve-year depressive symptom trajectories and their predictors in a community sample of older adults. *International Psychogeriatrics/IPA* **20**, 221–236.
- Andrescu C, Mulsant BH, Houck PR, Whyte EM, Mazumdar S, Dombrovski AY, Pollock BG, Reynolds CF III** (2008b). Empirically derived decision trees for the treatment of late-life depression. *The American Journal of Psychiatry* **165**, 855–862.
- Anglemyer A, Horvath HT, Bero L** (2014). Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *The Cochrane Database of Systematic Reviews* **4**, MR000034.
- Angst J, Gamma A, Rossler W, Ajdacic V, Klein DN** (2011). Childhood adversity and chronicity of mood disorders. *European Archives of Psychiatry and Clinical Neuroscience* **261**, 21–27.
- Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A** (2012). Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Research and Treatment* **133**, 1–10.
- Bagby RM, Quilty LC, Segal ZV, McBride CC, Kennedy SH, Costa PT** (2008). Personality and differential treatment response in major depression: a randomized controlled trial comparing cognitive-behavioural therapy and pharmacotherapy. *Canadian Journal of Psychiatry/Revue Canadienne de Psychiatrie* **53**, 361–370.
- Barber JP, Muenz LR** (1996). The role of avoidance and obsessiveness in matching patients to cognitive and interpersonal psychotherapy: empirical findings from the treatment for depression collaborative research program. *Journal of Consulting and Clinical Psychology* **64**, 951–958.
- Baumeister H, Gordon P** (2012). Meta-review of depressive subtyping models. *Journal of Affective Disorders* **139**, 126–140.
- Bellino S, Paradiso E, Bogetto F** (2008). Efficacy and tolerability of pharmacotherapies for borderline personality disorder. *CNS Drugs* **22**, 671–692.
- Bernecker SL, Constantino MJ, Pazzaglia AM, Ravitz P, McBride C** (2014). Patient interpersonal and cognitive changes and their relation to outcome in interpersonal psychotherapy for depression. *Journal of Clinical Psychology* **70**, 518–527.
- Bradvik L, Mattisson C, Bogren M, Nettelbladt P** (2008). Long-term suicide risk of depression in the Lundby cohort 1947–1997 – severity and gender. *Acta Psychiatrica Scandinavica* **117**, 185–191.
- Breitenstein B, Scheuer S, Holsboer F** (2014). Are there meaningful biomarkers of treatment response for depression? *Drug Discovery Today* **19**, 539–561.
- Burke JF, Hayward RA, Nelson JP, Kent DM** (2014). Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circulation: Cardiovascular Quality and Outcomes* **7**, 163–169.
- Carter JD, Luty SE, McKenzie JM, Mulder RT, Frampton CM, Joyce PR** (2011). Patient predictors of response to cognitive behaviour therapy and interpersonal psychotherapy in a randomised clinical trial for depression. *Journal of Affective Disorders* **128**, 252–261.
- Chang YJ, Chen LJ, Chung KP, Lai MS** (2012). Risk groups defined by Recursive Partitioning Analysis of patients with colorectal adenocarcinoma treated with colorectal resection. *BMC Medical Research Methodology* **12**, 2.
- Chao ST, Koefman SA, Woody N, Angelov L, Soeder SL, Reddy CA, Rybicki LA, Djemil T, Suh JH** (2012). Recursive partitioning analysis index is predictive for overall survival in patients undergoing spine stereotactic

- body radiation therapy for spinal metastases. *International Journal of Radiation Oncology, Biology, Physics* **82**, 1738–1743.
- Cohen A, Houck PR, Szanto K, Dew MA, Gilman SE, Reynolds CF** (2006). Social inequalities in response to antidepressant treatment in older adults. *Archives of General Psychiatry* **63**, 50–56.
- Cohen A, Gilman SE, Houck PR, Szanto K, Reynolds CF** (2009). Socioeconomic status and anxiety as predictors of antidepressant treatment response and suicidal ideation in older adults. *Social Psychiatry and Psychiatric Epidemiology* **44**, 272–277.
- Constantino MJ, Adams ML, Pazzaglia AM, Bernecker SL, Ravitz P, McBride C** (2013). Baseline patient characteristics as predictors of remission in interpersonal psychotherapy for depression. *Psychotherapy Research: Journal of the Society for Psychotherapy Research* **23**, 190–200.
- Cooper C, Jones L, Dunn E, Forty L, Haque S, Oyebode F, Craddock N, Jones I** (2007). Clinical presentation of postnatal and non-postnatal depressive episodes. *Psychological Medicine* **37**, 1273–1280.
- Cooper PJ, Murray L** (1995). Course and recurrence of postnatal depression. Evidence for the specificity of the diagnostic concept. *The British Journal of Psychiatry: The Journal of Mental Science* **166**, 191–195.
- Cronkite RC, Woodhead EL, Finlay A, Timko C, Unger Hu K, Moos RH** (2013). Life stressors and resources and the 23-year course of depression. *Journal of Affective Disorders* **150**, 370–377.
- Cuijpers P, Reynolds CF III, Donker T, Li J, Andersson G, Beekman A** (2012). Personalized treatment of adult depression: medication, psychotherapy, or both? A systematic review. *Depression and Anxiety* **29**, 855–864.
- Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, Angermeyer MC, Bernert S, de Girolamo G, Morosini P, Polidori G, Kikkawa T, Kawakami N, Ono Y, Takeshima T, Uda H, Karam EG, Fayyad JA, Karam AN, Mneimneh ZN, Medina-Mora ME, Borges G, Lara C, de Graaf R, Ormel J, Gureje O, Shen Y, Huang Y, Zhang M, Alonso J, Haro JM, Vilagut G, Bromet EJ, Gluzman S, Webb C, Kessler RC, Merikangas KR, Anthony JC, Von Korff MR, Wang PS, Brugha TS, Aguilar-Gaxiola S, Lee S, Heeringa S, Pennell BE, Zaslavsky AM, Ustun TB, Chatterji S** (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA* **291**, 2581–2590.
- Denton WH, Carmody TJ, Rush AJ, Thase ME, Trivedi MH, Arnow BA, Klein DN, Keller MB** (2010). Dyadic discord at baseline is associated with lack of remission in the acute treatment of chronic depression. *Psychological Medicine* **40**, 415–424.
- DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L** (2014). The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS ONE* **9**, e83875.
- Dew MA, Reynolds CF III, Houck PR, Hall M, Buysse DJ, Frank E, Kupfer DJ** (1997). Temporal profiles of the course of depression during treatment. Predictors of pathways toward recovery in the elderly. *Archives of General Psychiatry* **54**, 1016–1024.
- Diaz Munoz I, van der Laan MJ** (2011). Super learner based conditional density estimation with application to marginal structural models. *The International Journal of Biostatistics* **7**, Article 38.
- Dimidjian S, Hollon SD, Dobson KS, Schmalzing 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.
- Dorresteijn JA, Visseren FL, Ridker PM, Wassink AM, Paynter NP, Steyerberg EW, van der Graaf Y, Cook NR** (2011). Estimating treatment effects for individual patients based on the results of randomised clinical trials. *BMJ* **343**, d5888.
- Dunlop BW, Binder EB, Cubells JF, Goodman MM, Kelley ME, Kinkead B, Kutner M, Nemeroff CB, Newport DJ, Owens MJ, Pace TW, Ritchie JC, Rivera VA, Westen D, Craighead WE, Mayberg HS** (2012). Predictors of remission in depression to individual and combined treatments (PRE-DICT): study protocol for a randomized controlled trial. *Trials* **13**, 106.
- Eaton WW, Shao H, Nestadt G, Lee HB, Bienvenu OJ, Zandi P** (2008). Population-based study of first onset and chronicity in major depressive disorder. *Archives of General Psychiatry* **65**, 513–520.
- Echouffo-Tcheugui JB, Kengne AP** (2013). Comparative performance of diabetes-specific and general population-based cardiovascular risk assessment models in people with diabetes mellitus. *Diabetes & Metabolism* **39**, 389–396.
- Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB** (1989). National institute of mental health treatment of depression collaborative research program. General effectiveness of treatments. *Archives of General Psychiatry* **46**, 971–982; discussion 983.
- Feske U, Frank E, Kupfer DJ, Shear MK, Weaver E** (1998). Anxiety as a predictor of response to interpersonal psychotherapy for recurrent major depression: an exploratory investigation. *Depression and Anxiety* **8**, 135–141.
- Fichter MM, Quadflieg N, Fischer UC, Kohlboeck G** (2010). Twenty-five-year course and outcome in anxiety and depression in the Upper Bavarian Longitudinal Community Study. *Acta Psychiatrica Scandinavica* **122**, 75–85.
- Fink M, Rush AJ, Knapp R, Rasmussen K, Mueller M, Rummans TA, O'Connor K, Husain M, Biggs M, Bailine S, Kellner CH** (2007). DSM melancholic features are unreliable predictors of ECT response: a CORE publication. *The Journal of ECT* **23**, 139–146.
- Fournier JC, DeRubeis RJ, Shelton RC, Gallop R, Amsterdam JD, Hollon SD** (2008). Antidepressant medications *v.* cognitive therapy in people with depression

- with or without personality disorder. *The British Journal of Psychiatry: The Journal of Mental Science* **192**, 124–129.
- Fournier JC, DeRubeis RJ, Shelton RC, Hollon SD, Amsterdam JD, Gallop R** (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting and Clinical Psychology* **77**, 775–787.
- Frank E, Cassano GB, Rucci P, Thompson WK, Kraemer HC, Fagiolini A, Maggi L, Kupfer DJ, Shear MK, Houck PR, Calugi S, Grochocinski VJ, Scocco P, Buttenfield J, Forgione RN** (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine* **41**, 151–162.
- Hastie T, Tibshirani R, Friedman J** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. Springer: New York.
- Hayward RA, Kent DM, Vijan S, Hofer TP** (2006). Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Medical Research Methodology* **6**, 18.
- Hetrick SE, Simmons M, Thompson A, Parker AG** (2011). What are specialist mental health clinician attitudes to guideline recommendations for the treatment of depression in young people? *The Australian and New Zealand Journal of Psychiatry* **45**, 993–1001.
- Hoencamp E, Haffmans PM, Duivenvoorden H, Knegtering H, Dijken WA** (1994). Predictors of (non-) response in depressed outpatients treated with a three-phase sequential medication strategy. *Journal of Affective Disorders* **31**, 235–246.
- Hollon SD, DeRubeis RJ, Fawcett J, Amsterdam JD, Shelton RC, Zajecka J, Young PR, Gallop R** (2014). Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. *JAMA Psychiatry* **71**, 1157–1164.
- Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P, DeRubeis RJ** (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PLoS ONE* **10**, e0140771.
- Ionescu R, Popescu C, Jipescu I** (1994). Predictors of outcome in depression. *Romanian Journal of Neurology and Psychiatry/Revue Roumaine de Neurologie et Psychiatrie* **32**, 153–173.
- Jain FA, Hunter AM, Brooks JO III, Leuchter AF** (2013). Predictive socioeconomic and clinical profiles of antidepressant response and remission. *Depression and Anxiety* **30**, 624–630.
- James G, Witten D, Hastie T, Tibshirani R** (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer: New York.
- Jarrett RB, Minhajuddin A, Kangas JL, Friedman ES, Callan JA, Thase ME** (2013). Acute phase cognitive therapy for recurrent major depressive disorder: who drops out and how much do patient skills influence response? *Behaviour Research and Therapy* **51**, 221–230.
- Johnstone JM, Luty SE, Carter JD, Mulder RT, Frampton CM, Joyce PR** (2009). Childhood neglect and abuse as predictors of antidepressant response in adult depression. *Depression and Anxiety* **26**, 711–717.
- Joyce PR, McKenzie JM, Carter JD, Rae AM, Luty SE, Frampton CM, Mulder RT** (2007). Temperament, character and personality disorders as predictors of response to interpersonal psychotherapy and cognitive-behavioural therapy for depression. *The British Journal of Psychiatry: The Journal of Mental Science* **190**, 503–508.
- Kennedy SH, Andersen HF, Lam RW** (2006). Efficacy of escitalopram in the treatment of major depressive disorder compared with conventional selective serotonin reuptake inhibitors and venlafaxine XR: a meta-analysis. *Journal of Psychiatry & Neuroscience* **31**, 122–131.
- Kennedy SH, Downar J, Evans KR, Feilotter H, Lam RW, MacQueen GM, Milev R, Parikh SV, Rotzinger S, Soares C** (2012). The Canadian Biomarker Integration Network in Depression (CAN-BIND): advances in response prediction. *Current Pharmaceutical Design* **18**, 5976–5989.
- Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA** (2010). Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* **11**, 85.
- Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS** (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry* **51**, 8–19.
- Kessler RC, Merikangas KR, Berglund P, Eaton WW, Koretz DS, Walters EE** (2003). Mild disorders should not be eliminated from the DSM-V. *Archives of General Psychiatry* **60**, 1117–1122.
- Kessler RC, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell BE, Walters EE, Zaslavsky A, Zheng H** (2004). The US National Comorbidity Survey Replication (NCS-R): design and field procedures. *International Journal of Methods in Psychiatric Research* **13**, 69–92.
- Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, Ebert DD, Hwang I, Li J, de Jonge P, Nierenberg AA, Petukhova MV, Rosellini AJ, Sampson NA, Schoevers RA, Wilcox MA, Zaslavsky AM** (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*.
- Klein DN, Shankman SA, Rose S** (2008). Dysthymic disorder and double depression: prediction of 10-year course trajectories and outcomes. *Journal of Psychiatric Research* **42**, 408–415.
- Kool S, Schoevers R, de Maat S, Van R, Molenaar P, Vink A, Dekker J** (2005). Efficacy of pharmacotherapy in depressed patients with and without personality disorders: a systematic review and meta-analysis. *Journal of Affective Disorders* **88**, 269–278.
- Kraemer HC** (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Statistics in Medicine* **32**, 1964–1973.

- Kuiper S, McLean L, Fritz K, Lampe L, Malhi GS (2013). Getting depression clinical practice guidelines right: time for change? *Acta Psychiatrica Scandinavica. Supplementum* 24–30.
- Lamers F, Burstein M, He JP, Avenevoli S, Angst J, Merikangas KR (2012). Structure of major depressive disorder in adolescents and adults in the US general population. *The British Journal of Psychiatry: The Journal of Mental Science* 201, 143–150.
- Li C, Lu Y (2010). Evaluating the improvement in diagnostic utility from adding new predictors. *Biometrical Journal. Biometrische Zeitschrift* 52, 417–435.
- Lin HW, Chen YH (2014). Adjustment for missing confounders in studies based on observational databases: 2-stage calibration combining propensity scores from primary and validation data. *American Journal of Epidemiology* 180, 308–317.
- Liu Y, Nie Z, Zhou J, Farnum M, Narayan VA, Wittenberg G, Ye J (2014). Sparse generalized functional linear model for predicting remission status of depression patients. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 364–375.
- Lowe B, Schenkel I, Bair MJ, Gobel C (2005). Efficacy, predictors of therapy response, and safety of sertraline in routine clinical practice: prospective, open-label, non-interventional postmarketing surveillance study in 1878 patients. *Journal of Affective Disorders* 87, 271–279.
- Lurie JD, Morgan TS (2013). Pros and cons of pragmatic clinical trials. *Journal of Comparative Effectiveness Research* 2, 53–58.
- Luty SE, Carter JD, McKenzie JM, Rae AM, Frampton CM, Mulder RT, Joyce PR (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *The British Journal of Psychiatry: The Journal of Mental Science* 190, 496–502.
- Madigan CD, Jolly K, Lewis AL, Aveyard P, Daley AJ (2014). A randomised controlled trial of the effectiveness of self-weighing as a weight loss intervention. *The International Journal of Behavioral Nutrition and Physical Activity* 11, 125.
- Marquett RM, Thompson LW, Reiser RP, Holland JM, O'Hara RM, Kesler SR, Stepanenko A, Bilbrey A, Rengifo J, Majoros A, Thompson DG (2013). Psychosocial predictors of treatment response to cognitive-behavior therapy for late-life depression: an exploratory study. *Aging & Mental Health* 17, 830–838.
- Mattisson C, Bogren M, Horstmann V, Munk-Jorgensen P, Nettelbladt P (2007). The long-term course of depressive disorders in the Lundby Study. *Psychological Medicine* 37, 883–891.
- McBride C, Atkinson L, Quilty LC, Bagby RM (2006). Attachment as moderator of treatment outcome in major depression: a randomized control trial of interpersonal psychotherapy versus cognitive behavior therapy. *Journal of Consulting and Clinical Psychology* 74, 1041–1054.
- McGrath PJ, Nunes EV, Stewart JW, Goldman D, Agosti V, Ocepek-Welikson K, Quitkin FM (1996). Imipramine treatment of alcoholics with primary depression: a placebo-controlled clinical trial. *Archives of General Psychiatry* 53, 232–240.
- Moos RH, Cronkite RC (1999). Symptom-based predictors of a 10-year chronic course of treated depression. *The Journal of Nervous and Mental Disease* 187, 360–368.
- Mueller TI, Leon AC, Keller MB, Solomon DA, Endicott J, Coryell W, Warshaw M, Maser JD (1999). Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *The American Journal of Psychiatry* 156, 1000–1006.
- Nanni V, Uher R, Danese A (2012). Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. *The American Journal of Psychiatry* 169, 141–151.
- National Institute for Health and Clinical Excellence (NICE) (2009). *Depression: Treatment and Management of Depression in Adult*. National Institute for Health and Clinical Excellence: London.
- Nelson JC, Zhang Q, Deberdt W, Marangell LB, Karamustafalioglu O, Lipkovich IA (2012). Predictors of remission with placebo using an integrated study database from patients with major depressive disorder. *Current Medical Research and Opinion* 28, 325–334.
- Nemeroff CB, Heim CM, Thase ME, Klein DN, Rush AJ, Schatzberg AF, Ninan PT, McCullough JP Jr., Weiss PM, Dunner DL, Rothbaum BO, Kornstein S, Keitner G, Keller MB (2003). Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. *Proceedings of the National Academy of Sciences of the United States of America* 100, 14293–14296.
- Neugebauer R, Schmittiel JA, van der Laan MJ (2014). Targeted learning in real-world comparative effectiveness research with time-varying interventions. *Statistics in Medicine* 33, 2480–2520.
- Papakostas GI, Stahl SM, Krishen A, Seifert CA, Tucker VL, Goodale EP, Fava M (2008). Efficacy of bupropion and the selective serotonin reuptake inhibitors in the treatment of major depressive disorder with high levels of anxiety (anxious depression): a pooled analysis of 10 studies. *The Journal of Clinical Psychiatry* 69, 1287–1292.
- Perel P, Edwards P, Wentz R, Roberts I (2006). Systematic review of prognostic models in traumatic brain injury. *BMC Medical Informatics and Decision Making* 6, 38.
- Perlis RH (2007). Use of treatment guidelines in clinical decision making in bipolar disorder: a pilot survey of clinicians. *Current Medical Research and Opinion* 23, 467–475.
- Perlis RH (2013). A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biological Psychiatry* 74, 7–14.
- Perlis RH (2014). Pharmacogenomic testing and personalized treatment of depression. *Clinical Chemistry* 60, 53–59.
- Piccioletto S, Eisen EA, Chevrier J (2014). 0351 G-estimation: why does it work and what does it offer? *Occupational and Environmental Medicine* 71 (Suppl. 1), A120–A121.
- Pizzagalli DA (2011). Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 36, 183–206.
- Prentice RL, Langer RD, Stefanick ML, Howard BV, Pettinger M, Anderson GL, Barad D, Curb JD, Kotchen J,

- Kuller L, Limacher M, Wactawski-Wende J** (2006). Combined analysis of Women's Health Initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. *American Journal of Epidemiology* **163**, 589–599.
- Prieto-Merino D, Pocock SJ** (2012). The science of risk models. *European Journal of Preventive Cardiology* **19**, 7–13.
- Quitkin FM, McGrath PJ, Stewart JW, Harrison W, Tricamo E, Wager SG, Ocepek-Welikson K, Nunes E, Rabkin JG, Klein DF** (1990). Atypical depression, panic attacks, and response to imipramine and phenelzine. A replication. *Archives of General Psychiatry* **47**, 935–941.
- Quitkin FM, Harrison W, Stewart JW, McGrath PJ, Tricamo E, Ocepek-Welikson K, Rabkin JG, Wager SG, Nunes E, Klein DF** (1991). Response to phenelzine and imipramine in placebo nonresponders with atypical depression. A new application of the crossover design. *Archives of General Psychiatry* **48**, 319–323.
- Rabinoff M, Kitchen CM, Cook IA, Leuchter AF** (2011). Evaluation of quantitative EEG by classification and regression trees to characterize responders to antidepressant and placebo treatment. *The Open Medical Informatics Journal* **5**, 1–8.
- Riedel M, Moller HJ, Obermeier M, Adli M, Bauer M, Kronmuller K, Brieger P, Laux G, Bender W, Heuser I, Zeiler J, Gaebel W, Schennach-Wolff R, Henkel V, Seemuller F** (2011). Clinical predictors of response and remission in inpatients with depressive syndromes. *Journal of Affective Disorders* **133**, 137–149.
- Ritchie MD** (2005). Bioinformatics approaches for detecting gene-gene and gene-environment interactions in studies of human disease. *Neurosurgical Focus* **19**, E2.
- Romera I, Delgado-Cohen H, Perez T, Caballero L, Gilaberte I** (2008). Factor analysis of the Zung self-rating depression scale in a large sample of patients with major depressive disorder in primary care. *BMC Psychiatry* **8**, 4.
- Rush AJ, Wisniewski SR, Warden D, Luther JF, Davis LL, Fava M, Nierenberg AA, Trivedi MH** (2008). Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. *Archives of General Psychiatry* **65**, 870–880.
- Simon GE, Perlis RH** (2010). Personalized medicine for depression: can we match patients with treatments? *The American Journal of Psychiatry* **167**, 1445–1455.
- Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP** (2012). Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ* **344**, e3318.
- Smits JA, Minhajuddin A, Thase ME, Jarrett RB** (2012). Outcomes of acute phase cognitive therapy in outpatients with anxious versus nonanxious depression. *Psychotherapy and Psychosomatics* **81**, 153–160.
- Sotsky SM, Glass DR, Shea MT, Pilkonis PA, Collins JF, Elkin I, Watkins JT, Imber SD, Leber WR, Moyer J, et al.** (1991). Patient predictors of response to psychotherapy and pharmacotherapy: findings in the NIMH Treatment of Depression Collaborative Research Program. *The American Journal of Psychiatry* **148**, 997–1008.
- Souslova T, Marple TC, Spiekerman AM, Mohammad AA** (2013). Personalized medicine in Alzheimer's disease and depression. *Contemporary Clinical Trials* **36**, 616–623.
- Steyerberg EW, Vedder MM, Leening MJ, Postmus D, D'Agostino RB Sr., Van Calster B, Pencina MJ** (2014). Graphical assessment of incremental value of novel markers in prediction models: from statistical to decision analytical perspectives. *Biometrical Journal. Biometrische Zeitschrift*.
- Strobl C, Malley J, Tutz G** (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* **14**, 323–348.
- Suarez D, Haro JM, Novick D, Ochoa S** (2008). Marginal structural models might overcome confounding when analyzing multiple treatment effects in observational studies. *Journal of Clinical Epidemiology* **61**, 525–530.
- Sung SC, Haley CL, Wisniewski SR, Fava M, Nierenberg AA, Warden D, Morris DW, Kurian BT, Trivedi MH, Rush AJ** (2012). The impact of chronic depression on acute and long-term outcomes in a randomized trial comparing selective serotonin reuptake inhibitor monotherapy versus each of 2 different antidepressant medication combinations. *The Journal of Clinical Psychiatry* **73**, 967–976.
- Szadoczky E, Rozsa S, Zambori J, Furedi J** (2004). Predictors for 2-year outcome of major depressive episode. *Journal of Affective Disorders* **83**, 49–57.
- Tchetgen Tchetgen E** (2014). The control outcome calibration approach for causal inference with unobserved confounding. *American Journal of Epidemiology* **179**, 633–640.
- Thase ME, Reynolds CF III, Frank E, Simons AD, McGeary J, Fasiczka AL, Garamoni GG, Jennings JR, Kupfer DJ** (1994). Do depressed men and women respond similarly to cognitive behavior therapy? *The American Journal of Psychiatry* **151**, 500–505.
- Thase ME, Greenhouse JB, Frank E, Reynolds CF III, Pilkonis PA, Hurley K, Grochocinski V, Kupfer DJ** (1997). Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Archives of General Psychiatry* **54**, 1009–1015.
- Thase ME, Pritchett YL, Ossanna MJ, Swindle RW, Xu J, Detke MJ** (2007). Efficacy of duloxetine and selective serotonin reuptake inhibitors: comparisons as assessed by remission rates in patients with major depressive disorder. *Journal of Clinical Psychopharmacology* **27**, 672–676.
- Toh S, Manson JE** (2013). An analytic framework for aligning observational and randomized trial data: application to postmenopausal hormone therapy and coronary heart disease. *Statistics in Biosciences* **5**.
- Trivedi MH, Morris DW, Pan JY, Grannemann BD, John Rush A** (2005). What moderator characteristics are associated with better prognosis for depression? *Neuropsychiatric Disease and Treatment* **1**, 51–57.
- 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. *The American Journal of Psychiatry* **163**, 28–40.
- Troxel WM, Kupfer DJ, Reynolds CF III, Frank E, Thase ME, Miewald JM, Buysse DJ** (2012). Insomnia and objectively measured sleep disturbances predict treatment outcome in depressed patients treated with psychotherapy or psychotherapy-pharmacotherapy combinations. *The Journal of Clinical Psychiatry* **73**, 478–485.
- Uher R, Perroud N, Ng MY, Hauser J, Henigsberg N, Maier W, Mors O, Placentino A, Rietschel M, Souery D, Zagar T, Czerski PM, Jerman B, Larsen ER, Schulze TG, Zobel A, Cohen-Woods S, Pirlo K, Butler AW, Muglia P, Barnes MR, Lathrop M, Farmer A, Breen G, Aitchison KJ, Craig I, Lewis CM, McGuffin P** (2010). Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. *The American Journal of Psychiatry* **167**, 555–564.
- 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.
- Upstill-Goddard R, Eccles D, Fliege J, Collins A** (2013). Machine learning approaches for the discovery of gene-gene interactions in disease data. *Briefings in Bioinformatics* **14**, 251–260.
- van der Laan MJ, Gruber S** (2010). Collaborative double robust targeted maximum likelihood estimation. *The International Journal of Biostatistics* **6**, Article 17.
- van der Laan MJ, Rose S** (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer: New York.
- van Loo HM, de Jonge P, Romeijn JW, Kessler RC, Schoevers RA** (2012). Data-driven subtypes of major depressive disorder: a systematic review. *BMC Medicine* **10**, 156.
- van Loo HM, Cai T, Gruber MJ, Li J, de Jonge P, Petukhova M, Rose S, Sampson NA, Schoevers RA, Wardenaar KJ, Wilcox MA, Al-Hamzawi AO, Andrade LH, Bromet EJ, Bunting B, Fayyad J, Florescu SE, Gureje O, Hu C, Huang Y, Levinson D, Medina-Mora ME, Nakane Y, Posada-Villa J, Scott KM, Xavier M, Zarkov Z, Kessler RC** (2014). Major depressive disorder subtypes to predict long-term course. *Depression and Anxiety* **31**, 765–777.
- Van Staa TP, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, Delaney B, Smeeth L** (2012). Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ* **344**.
- Vrieze E, Demyttenaere K, Bruffaerts R, Hermans D, Pizzagalli DA, Sienaert P, Hompes T, de Boer P, Schmidt M, Claes S** (2014). Dimensions in major depressive disorder and their relevance for treatment outcome. *Journal of Affective Disorders* **155**, 35–41.
- Vuoriolehto MS, Melartin TK, Isometsa ET** (2009). Course and outcome of depressive disorders in primary care: a prospective 18-month study. *Psychological Medicine* **39**, 1697–1707.
- Wallace ML, Frank E, Kraemer HC** (2013). A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry* **70**, 1241–1247.
- Wardenaar KJ, van Loo HM, Cai T, Fava M, Gruber MJ, Li J, de Jonge P, Nierenberg AA, Petukhova MV, Rose S, Sampson NA, Schoevers RA, Wilcox MA, Alonso J, Bromet EJ, Bunting B, Florescu SE, Fukao A, Gureje O, Hu C, Huang YQ, Karam AN, Levinson D, Medina Mora ME, Posada-Villa J, Scott KM, Taib NI, Viana MC, Xavier M, Zarkov Z, Kessler RC** (2014). The effects of co-morbidity in defining major depression subtypes associated with long-term course and severity. *Psychological Medicine* **44**, 3289–3302.
- Williams LM, Rush AJ, Koslow SH, Wisniewski SR, Cooper NJ, Nemeroff CB, Schatzberg AF, Gordon E** (2011). International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: rationale and protocol. *Trials* **12**, 4.
- Willke RJ, Zheng Z, Subedi P, Althin R, Mullins CD** (2012). From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer. *BMC Medical Research Methodology* **12**, 185.
- Zhang H, Singer BH** (2010). *Recursive Partitioning and Applications*. Springer: New York.