Published in final edited form as:

Pharmacogenomics. 2012 January ; 13(2): 233–240. doi:10.2217/pgs.11.161.

Biomarkers predicting treatment outcome in depression: what is clinically significant?

Rudolf Uher^{*,1,2}, Katherine E Tansey¹, Karim Malki¹, and Roy H Perlis³

¹Institute of Psychiatry, King's College London, 16 De Crespigny Park, SE5 8AF, London, UK

²Brain Repair Centre, Dalhousie University, Halifax, NS, Canada Laboratory of Psychiatric

³Pharmacogenomics, Department of Psychiatry, Massachusetts General Hospital & Harvard Medical School, Boston, MA, USA

Abstract

Aim—To extend to biomarker studies the consensus clinical significance criterion of a three-point difference in Hamilton Rating Scale for Depression.

Materials & methods—We simulated datasets modeled on large clinical trials.

Results—In a typical clinical trial comparing active treatment and placebo, a difference of three Hamilton Rating Scale for Depression (HRSD) points at the end of treatment corresponds to 6.3% of variance in outcome explained. To achieve a similar explanatory power, genotypes with minor allele frequencies of 5, 10, 20, 30 and 50% need to attain a per allele difference of 4.7, 3.6, 2.8, 2.4 and 2.2 HRSD points, respectively. A normally distributed continuous biomarker will need an effect size of 1.5 HRSD points per standard deviation. A number needed to assess of three suggests that with this effect size, a biomarker will significantly improve the prediction of outcome in one out of every three patients assessed.

Conclusion—This report provides guidance on assessing clinical significance of biomarkers predictive of outcome in depression treatment.

Keywords

antidepressant medication; biomarkers; clinical significance; major depressive disorder; number needed to assess

Depression is one of the most common and disabling diseases worldwide. Although a number of pharmacological and psychosocial treatment options are available, no single treatment is universally effective and most individuals with depression undergo multiple treatments before achieving remission. The substantial individual variation of treatment outcomes has motivated investigation of biomarkers with the hope that these can help select a treatment that a given individual is more likely to benefit from. However, with a large enough sample, even weak predictive effects can be found to be statistically significant. While the utility of any particular test depends on the context in which it is applied, there is

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

a need for a benchmark that could determine if a predictive biomarker is likely to be meaningful in clinical settings [1].

The distinction between statistical and clinical significance has been made in the context of controlled clinical trials assessing the difference of outcome between active treatment and control condition [2–5]. It has been proposed that for a treatment to be considered clinically significant, it has to improve outcome by a degree that is noticeable to the individual or close others and that makes a meaningful difference to the individual's ability to function and participate in society [2]. NICE in the UK had defined clinically significant effect size in comparison of two treatments for depression as a difference of three points or more on the 17-item Hamilton Rating Scale for Depression (HRSD) at the end of treatment [5,6]. This definition has been widely adopted as a benchmark for comparisons of antidepressant treatment against placebo [3,7,8]. The aim of this article is to extend the use of this clinical significance criterion to the study of biomarkers and other predictors of outcome.

This extension is not entirely straightforward. Unlike the typically equal groups receiving active treatment and comparator in a clinical trial, biomarkers are often unequally distributed in the population. Therefore, in addition to the effect size of prediction per unit of biomarker, we need to take into account its distribution. This is especially relevant to the decision whether to carry out a test to measure a biomarker. For example, consider a genotype that is associated with an improvement of three HRSD points less than alternative genotypes. If this genotype is carried by 50% of patients, it may be worthwhile genotyping it to improve prediction of outcome. Conversely, if the same genotype is only carried by 1% of patients, genotyping will be noninformative for the vast majority of patients and it will not be worth testing 100 patients to inform the prediction of outcome in a single individual unless it indicates a high probability of an extreme and dangerous outcome. In the present study, we establish thresholds for clinical significance for treatment outcome prediction by biomarkers that are either genetic (genotype is typically a three-level categorical variable) or of the type of protein levels (continuous variable).

Materials & methods

Standard

We took as the starting point the consensus recommendation by NICE that a difference of three points or more on the HRSD from control condition be the criterion for a clinically significant effect of a treatment for major depressive disorder [5,6]. Although this criterion was proposed as a 'rule of thumb', it was subsequently adopted by influential meta-analyses and obtained a status of a recognized benchmark for discussions of the efficacy of antidepressant treatments [3,7,8]. It also corresponds to approximately half the difference between bands on the Clinical Global Impression Scale, meaning that a difference greater than this threshold means that subject becomes more likely to transit into a different band of clinical impression, (e.g., from moderate to mild depression) [9]. Therefore, we make the assumption that a biomarker that predicts a difference of at least three HRSD points, or its equivalent on other measures [10,11], at the end of treatment for a substantial proportion of individuals is likely to be clinically significant.

Simulated datasets

We modeled simulated datasets on recent largescale trials, the STAR*D [12] and the GENDEP [13], conducted in the USA and Europe, respectively. The characteristics of recruited patients and outcomes of 12 weeks of treatment with antidepressants were similar in these two large studies. Accordingly, we simulated samples with a mean baseline HRSD of 23.0 (standard deviation [SD] = 5.0) censored at the minimum value of 14, corresponding

to the usual eligibility threshold for studies of antidepressants, and a mean exit HRSD score of 11.5 (SD = 7.0) censored at the natural minimum value of 0, with a typical correlation of 0.4 (Pearson product-moment correlation) between baseline and end point values, corresponding to the mean of STAR*D and GENDEP values. We simulated a range of potential predictor variables including three-level genotypes with varying minor allele frequency distributed according to Hardy-Weinberg equilibrium and normally distributed continuous variables (these may represent a level of protein in the blood, an electrophysiological measure, a measure of cerebral blood flow, but also a continuous composite measure derived from multiple genetic or nongenetic markers). The predictors were unrelated to baseline values of depression severity, but the relationship between the predictors and outcome was systematically varied over simulation sets. For comparison, we have also simulated an active treatment to control comparison in a typical randomized controlled trial, where half the subjects are randomly allocated to receive the active treatment and the other half receive control condition (e.g., placebo). Measures of effect size are independent of sample size. However, to obtain stable estimates, we simulated relatively large clinical trials with 1000 subjects. Each data point is based on 10,000 independently simulated datasets and analyzed with a routine statistical approach. For continuous outcomes, linear regression was used with end-of-treatment HRSD score as the dependent variable and the biomarker (genotype or continuous) as the predictor of interest. Logistic regression was used for categorical outcomes.

Measures of effect size

In comparisons of active treatment to control condition, clinical significance is defined by the 'd family' group of effect size measures, including measures of between group difference, such as Cohen's d, odds ratio or risk ratio, which range from minus infinity to plus infinity and can be readily translated one to another [14]. By contrast, biomarkers aim to explain variation within a single group of patients who receive the same treatment and the effect size of such prediction is typically expressed with a measure from the 'r family', such as correlation coefficients and other measures of association, which range from minus one to plus one [14]. Translation between effect size measures from the d and r families requires a set of assumptions that vary from setting to setting. The aim of this report is to provide a criterion for clinical significance of genetic and other biomarkers that is relevant to the setting of a typical depression treatment study and corresponds to the criterion of clinical significance established for comparisons of active treatment and control conditions in clinical trials.

We started with an absolute difference expressed as a number of points on the HRSD by which equally sized groups receiving active treatment and comparator in a randomized controlled trial differed at the end of treatment. Assuming no between group differences at baseline, a standardized measure of effect size from the d family, such as Cohen's d, is calculated as the absolute difference between active treatment and comparator divided by the pooled SD at the end of treatment (e.g., for a group difference of three HRSD points and pooled SD of seven: $d = 3/7 = 0.43$) [14].

To compare the overall affect of the active treatment in the randomized controlled trial with the predictive power of variously distributed biomarkers in the population of patients with depression, we next calculated the proportion of variance explained (also known as coefficient of determination) as r^2 in a linear regression of outcome on treatment or other predictor variable. Proportion of variance explained allows comparison of population-based effect sizes across predictors with varying distribution. We use proportion of variance explained as the common metric to compare the predictive power of active treatment in a randomized controlled trial with the effect of genetic or continuous biomarkers with various degrees of relationship to outcome.

Finally, there is a group of effect size measures designed to facilitate communication with clinicians, including the number needed to treat (more accurately termed number needed to assess [NNA] for predictive measures) and area under the receiver operating characteristic curve (AuROC), which are mutually convertible [1,4,14]. These effect size measures were originally developed for the relationship between two binary variables (e.g., treatment and remission) and have been extended to the relationship between one binary and one continuous variable [4]. However, there is no direct method of calculating these measures for the relationship between two variables that are both continuous or have more than two levels. To approximate these measures, we also derived a categorical outcome of remission based on the standard definition of an HRSD score of seven or less at the end of treatment [15]. We calculated the AuROC for the relationship between predictors and this dichotomized outcome and we used McFadden's pseudo r^2 measure from a logistic regression to approximate the proportion of variance explained. We used previously published equations to convert AuROC to number needed to treat/NNA, number needed to treat = $1/(2 \times \text{AUC} - 1)$ [16].

Results

Treatment comparison in a randomized controlled trial

In a typical clinical trial comparing equally large groups of patients treated with active treatment and placebo, a difference of three HRSD points at the end of treatment corresponds to 6.3% of the variance in outcome explained by treatment allocation (Figure 1). We use this as a benchmark to be translated into predictions by genetic and continuous biomarkers.

Prediction of treatment outcome by genotypes

For genetic predictors, the proportion of variance explained depends on the minor allele frequency and the difference in outcome per allele (Figure 2). To explain 6.3% of the variance in outcome, genotypes with minor allele frequencies of 50, 30, 20, 10 and 5% need to attain a per-allele difference of 2.2, 2.4, 2.8, 3.6 and 4.7 HRSD points, respectively, under an additive genetic model (Figure 2). For very low minor allele frequencies, the required absolute effect sizes increase dramatically. For genotypes with minor allele frequencies of 4, 3 and 2%, effects of 5.2, 6.1 and 7.3 HRSD points per allele are needed to explain 6.3% of the variance in outcome. A genotype with a minor allele frequency of 1% does not attain this explanatory power even with effects exceeding ten HRSD points per allele. With a difference held constant at three HRSD points per allele, a genotype with a minor allele frequency of 13% explains a proportion of variance in outcome corresponding to a clinically significant difference between placebo and active treatment.

Prediction of treatment outcome by continuous biomarkers

For a normally distributed continuous predictor, the proportion of variance explained is exponentially related to the absolute effect size expressed in HRSD point difference per one SD of the predictor (Figure 3). To explain 6.3% of the variance in outcome, an effect size of 1.52 HRSD points per one SD of a continuously distributed predictor is needed (Figure 3).

Continuous & categorical outcomes

The categorical outcome of remission (end point HRSD of seven or less) was constructed in the same set of simulated datasets with continuous biomarkers as predictors. There was a perfect linear relationship between the proportion of variance explained in the categorical outcome and proportion of variance explained in the continuous outcome (Figure 4). The effect size of the categorical outcome prediction, expressed as proportion of variance

explained estimated as pseudo r^2 , was 50.6% (95% CI: 50.3–51.0) of that for the linear outcome (Figure 4).

Proportion of variance explained & NNA

The NNA (calculated from AuROC) was in a negative exponential relationship to the proportion of variance in remission explained in a logistic model estimated as pseudo r^2 (Figure 5). Proportion of variance explained of 6.3% corresponded to a NNA of three.

Discussion

Using simulations incorporating parameters from real-world studies, we have translated the consensus criterion for clinical significance from comparative clinical trials to predictive biomarker research using genetic or continuously distributed biomarkers. In practice, the clinical applicability of predictive biomarkers will depend not only on effect size, but also on the biomarker availability and the cost, burden, delays associated with testing and the specificity of prediction to one treatment over another [1,4,17,18]. However, the present results provide several approximate rules that allow the translation of clinical significance between different types of predictors.

At the level of the whole patient population, the predictive power of genotypes depends on genotype frequency and absolute difference in outcome per each allele. A genotype with a minor allele frequency of 50% reaches the clinically significant prediction level if it is associated with a 2.2 HRSD point difference in outcome per allele. A genotype with a minor allele frequency of 10% needs a difference of 3.6 HRSD points per allele to reach the clinical significance criterion equivalent to that established in clinical trials. A normally distributed continuous biomarker that predicts a 1.5 HRSD point difference in outcome per SD has an explanatory power corresponding to the clinically significant difference of three HRSD points between equally sized groups. The graphs provided in this article (and an online calculator [101]) can serve to translate smaller or larger effect sizes for genetic or continuous predictors and compare their predictive power with a placebo–drug difference in randomized controlled trials.

If the continuous outcome measured on a depression rating scale is dichotomized to a categorical outcome of remission defined by a final score below a clinical cutoff, the predictive power of biomarkers is approximately halved. This is in line with previous studies demonstrating that dichotomization of continuous variables leads to a substantial loss of information and of statistical power [19–22]. Dichotomous outcomes allow an approximate translation between the proportion of variance explained (estimated as pseudo r^2 in logistic models) and the clinically meaningful effect size measure of number needed to assess. The proportion of variance explained corresponding to the previously established clinical significance criterion (6.3%) corresponds approximately to a NNA of three. This means that for every three patients assessed for a biomarker, one significantly more accurate prediction of outcome can be made.

The literature on the pharmacogenetics of antidepressants suggests that a single genetic marker is unlikely to achieve clinically significant prediction [23–25]. Therefore, polygenic scores summarizing information from multiple markers may replace single genotypes as predictors of outcome [26,27]. The normally distributed continuous predictors in the current study are applicable to such polygenic scores. Similarly, the continuous biomarker results can be applied to any linear combination of multiple biomarkers and clinical variables in a predictive score [1,28].

The relationship between the distribution of a biomarker in the population and its clinical usefulness may change dramatically in the future. If genetic biomarkers become routinely available and do not require additional testing, the absolute effect size of the prediction in a given individual becomes more important than the population-based explanatory power that is primarily considered in this article. However, the proportion of variance explained will remain a useful measure as it is applicable to multivariate models that may combine a large number of genotypes or other biomarkers to achieve a clinically meaningful prediction [1,18].

The applicability of the present results is limited to studies of similar character to those that served as a basis for the simulation. However, since the two studies that were considered were real-world pragmatic trials with relatively broad inclusion criteria [12,13], the conclusions should relatively well generalize to the population of patients treated in routine primary and secondary care settings. Our conclusions about the benchmark for clinical significance are based on the assumption that similar effect size is relevant for predictive biomarkers as for drug–placebo differences. Since the proposed clinical significance is based on a difference that is noticeable for patients, people close to the patient and clinicians [2,6], this assumptions appears reasonable. The simulations of genetic biomarkers in the present study have been limited to an additive genetic model, which assumes that heterozygotes are intermediate between the two homozygous groups. We chose an additive model, since this is the most commonly applied genetic model in practice and most recessive or dominant effects can be seen with an additive test. Extension of the present results to additive and dominant models depends on minor allele frequency and would be difficult to estimate for genetic markers with very low minor allele frequency owing to the very low number of homozygotes. In the present study, we have not separately considered the role of additional clinical variables that may contribute to prediction of outcome, such as number of previous episodes, duration of present episode, age, subtypes of depression and symptom dimensions [1,29–31]. Similar estimates for biomarker effects that are conditional on any such variables will have to be calculated with respect to the distribution of each such additional variable. When interpreting the results, it is important to keep in mind that the outcome of antidepressant treatment is measured with a certain error. Therefore, 100% of variance could never be explained even by a perfect predictor and the relatively low percentage of variance explained may represent a relatively larger proportion of what is explainable. For example, with a typical reliability of measurement of 0.80, 36% of variance in outcome is owing to measurement error and a clinically significant predictor would explain 10% rather than 6.3% of the theoretical outcome that is measured with perfect accuracy. However, since it is unlikely that depression severity will ever be measured without error, we keep the results in the raw metrics that of necessity is attenuated by measurement error. A final limitation is that the conclusions regarding the proportion of variance explained in categorical outcomes depend on the estimation of pseudo r^2 from logistic models, which is inexact, and should therefore be treated only as approximation.

Conclusion

In conclusion, we propose a set of standards for clinical significance of biomarkers predictive of depression treatment outcome. The clinical applicability of biomarkers will depend on a cost–benefit analysis where the effect size of prediction is weighted against the cost and risk of testing.

Future perspective

We hope that the current study will highlight the need for robust prediction with a sufficiently large effect size for biomarker research to significantly inform clinical practice.

We expect that this will lead to two important developments in biomarker research. First, since it appears unlikely that a single common biomarker will achieve the effect size needed to significantly inform clinical decision-making, the research will move towards combining multiple biomarkers into predictive score or algorithms that may achieve clinically significant prediction [1]. The estimates for continuous biomarkers can be used to assess the clinical significance of such multivariate prediction. Second, the applicability of any prediction increases if an alternative treatment is available for those with a predicted poor treatment outcome. It is therefore likely that biomarker research will move to comparative studies of different treatments. The comparison of psychological and pharmacological approaches may be especially promising [28].

Executive summary

Background

- The benefits of any biomarker-based prediction for an individual patient depend on effect size rather than on statistical significance.
- Three points difference on the Hamilton Rating Scale for Depression (HRSD) has been proposed as a consensus criterion for clinical significance in drug comparisons. A similar criterion is needed to assess the clinical significance of biomarkers.

Materials & methods

- To translate the clinical significance criterion to biomarker research, we have simulated datasets modeled on large pragmatic clinical trials.
- We used proportion of variance explained as a common metric to translate explanatory power between variously distributed predictors, including genotypes and continuous biomarkers.

Results

- Clinically significant drug–placebo difference corresponds to 6.3% of variance in outcome explained.
- Genotypes with minor allele frequencies of 5, 10, 20, 30 and 50% need to attain a per-allele difference of 4.7, 3.6, 2.8, 2.4 and 2.2 HRSD points, respectively, to reach clinically significant prediction.
- A normally distributed continuous biomarker (or a predictive score composed from multiple biomarkers) will need an effect size of 1.5 HRSD points per standard deviation to reach clinical significance.
- A number needed to asses of three suggests that, with this effect size, a biomarker will significantly improve the prediction of outcome in one out of every three patients assessed.

Discussion

- The clinical applicability of a predictive biomarker also depends on other factors including test availability, cost and treatment delays associated with testing.

Future perspective

- It is likely that multiple biomarkers combined in predictive scores or algorithms will be needed to achieve clinically meaningful prediction.

- Biomarker research should move towards comparative studies of different treatments (e.g., comparison of psychological and pharmacological treatment).

Acknowledgments

The authors thank M Gray, who has provided technical support with creating an online calculator that accompanies this article.

The research leading to these results has received support from the Innovative Medicine Initiative Joint Undertaking (IMI-JU) under grant agreement No. 115008, of which, resources are composed of EU and the European Federation of Pharmaceutical Industries and Associations (EFPIA) in-kind contribution and financial contribution from the EU's Seventh Framework Program (FP7/2007-2013). RH Perlis is supported by the National Institute of Mental Health MH086026. R Uher consults for the WHO. RH Perlis has received consulting fees from Proteus Biomedical, Concordant Rater Systems and RIDventures and research funding from National Institute of Mental Health. None of these organizations have direct interest in the content of this publication.

References

Papers of special note have been highlighted as:

- of interest
- ■ of considerable interest

1. Perlis RH. Translating biomarkers to clinical practice. *Mol. Psychiatry*. 2011; 16(11):1076–1087. [PubMed: 21709685] ■ ■ Reviews the application of biomarkers and the conditions that need to be fulfilled for biomarkers to inform clinical practice. It suggests the combination of multiple clinical variables and biomarkers into predictive scores or algorithms.
2. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* 1991; 59(1):12–19. [PubMed: 2002127] ■ Delineates the concept of clinical significance as separate from statistical significance.
3. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 2008; 5(2):e45. [PubMed: 18303940]
4. Kraemer HC. Reporting the size of effects in research studies to facilitate assessment of practical or clinical significance. *Psychoneuroendocrinology*. 1992; 17(6):527–536. [PubMed: 1287676] ■ Reviews the importance of effect size measures for clinical applications. It also describes conversions between various effect-size measures.
5. The Treatment and Management of Depression in Adults. NICE, Department of Health; London, UK: 2009. National Institute for Clinical Excellence.
6. National Institute for Clinical Excellence. Depression: Management of Depression in Primary and Secondary Care. 2004NICE, Department of HealthLondon, UK ■ Evidence-based guideline that introduced the consensus criterion for clinical significance in depression treatment outcome.
7. Fountoulakis KN, Moller HJ. Efficacy of antidepressants: a re-analysis and reinterpretation of the Kirsch data. *Int. J. Neuropsychopharmacol.* 2011; 14(3):405–412. [PubMed: 20800012]
8. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010; 303(1):47–53. [PubMed: 20051569]
9. Muller MJ, Himmerich H, Kienzle B, Szegedi A. Differentiating moderate and severe depression using the Montgomery–Asberg depression rating scale (MADRS). *J. Affect. Disord.* 2003; 77(3): 255–260. [PubMed: 14612225]
10. Carmody TJ, Rush AJ, Bernstein I, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur. Neuropsychopharmacol.* 2006; 16(8):601–611. [PubMed: 16769204]
11. Uher R, Farmer A, Maier W, et al. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol. Med.* 2008; 38(2):289–300. [PubMed: 17922940]

12. Rush AJ, Fava M, Wisniewski SR, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin. Trials*. 2004; 25(1):119–142. [PubMed: 15061154]
13. Uher R, Maier W, Hauser J, et al. Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. *Br. J. Psychiatry*. 2009; 194(3):252–259. [PubMed: 19252156]
14. Ellis, PD. *The Essential Guide to Effect Size*. Ellis, PD., editor. Cambridge, UK: Cambridge University Press; 2010.
15. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Arch. Gen. Psychiatry*. 1991; 48(9):851–855. [PubMed: 1929776]
16. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol. Psychiatry*. 2006; 59(11):990–996. [PubMed: 16368078]
17. Perlis RH, Patrick A, Smoller JW, Wang PS. When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost–effectiveness analysis based on data from the STAR*D study. *Neuropsychopharmacology*. 2009; 34(10):2227–2236. [PubMed: 19494805]
18. Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry*. 2010; 167(12):1445–1455. [PubMed: 20843873]
19. Kunz M. On responder analyses when a continuous variable is dichotomized and measurement error is present. *Biom. J.* 2011; 53(1):137–155. [PubMed: 21259314]
20. Deyi BA, Kosinski AS, Snapinn SM. Power considerations when a continuous outcome variable is dichotomized. *J. Biopharm. Stat.* 1998; 8(2):337–352. [PubMed: 9598427]
21. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006; 332(7549): 1080. [PubMed: 16675816]
22. Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can. J. Psychiatry*. 2002; 47(3):262–266. [PubMed: 11987478]
23. Ising M, Lucae S, Binder EB, et al. A genome wide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch. Gen. Psychiatry*. 2009; 66(9): 966–975. [PubMed: 19736353] ■ The first genome-wide pharmacogenetic study in depression. Suggests that a polygenic score is a more replicable predictor of outcome than any single genetic polymorphism.
24. Garriock HA, Kraft JB, Shyn SI, et al. A genome wide association study of citalopram response in major depressive disorder. *Biol. Psychiatry*. 2010; 67(2):133–138. [PubMed: 19846067] ■ The largest genome-wide pharmacogenetic study in depression to date. It reports a genome-wide analysis of predictors to a single treatment – the antidepressant citalopram.
25. Uher R, Perroud N, Ng MY, et al. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. *Am. J. Psychiatry*. 2010; 167(5):555–564. [PubMed: 20360315] ■ The largest comparative genome-wide pharmacogenetic study in depression. It reports genetic predictors of response to two alternative antidepressants: the serotonin-reuptake inhibitor, escitalopram, and the tricyclic antidepressant, nortriptyline. It has reported the only genome-wide significant predictor of outcome published to date.
26. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460(7256):748–752. [PubMed: 19571811]
27. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* 2009; 18(18):3525–3531. [PubMed: 19553258]
28. Uher R. Genes, environment, and individual differences in responding to treatment for depression. *Harv. Rev. Psychiatry*. 2011; 19(3):109–124. [PubMed: 21631158] ■ Comprehensive review of environmental, clinical and genetic predictors of treatment outcome. It concludes that different types of predictors may need to be combined in predictive scores or algorithms to significantly inform personalized treatment.
29. Joyce PR, McKenzie JM, Carter JD, et al. Temperament, character and personality disorders as predictors of response to interpersonal psychotherapy and cognitive-behavioural therapy for depression. *Br. J. Psychiatry*. 2007; 190:503–508. [PubMed: 17541110]

30. Uher R, Perlis RH, Henigsberg N, et al. Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychol. Med.* 2011; 20:1–14.
31. Uher R, Dernovsek MZ, Mors O, et al. Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. *J. Affect. Disord.* 2011; 132(1–2):112–120. [PubMed: 21411156]

Website

101. Online calculator. www.depressiontools.org

Figure 1. Variance explained in a drug-placebo comparison

The figure shows variance explained by HRSD point drug-placebo difference in a sample with 50% of individuals allocated to active drug and 50% to placebo. The dashed lines mark the point where the difference in outcome scores between active drug and comparator (placebo) reaches the consensus level for clinically significant difference of three HRSD points; the proportion of variance explained at this point is 0.0633. HRSD: Hamilton Rating Scale for Depression.

Figure 2. Variance explained by genotypes

Relationship between absolute change in outcome per each allele (in HRSD score points) and proportion of outcome variance explained by additive genetic effects for genotypes with various MAFs. The horizontal line marks the threshold for clinically significant prediction (6.33% variance explained).

HRSD: Hamilton Rating Scale for Depression; MAF: Minor allele frequency.

Figure 3. Variance explained by a continuous biomarker

The x-axis plots the outcome difference (HRSD) per one standard deviation of the biomarker. The dashed horizontal and vertical lines mark the point where the proportion of variance in outcome explained equals that explained by an active drug allocation in a clinical trial with a clinically significant difference of three HRSD points (i.e., 6.3%). The effect size (in HRSD points per biomarker standard deviation) for which this explanatory power is attained is 1.52.

HRSD: Hamilton Rating Scale for Depression.

Figure 4. Continuous and categorical outcomes

Proportion variance explained in linear model (r^2) predicting a continuous outcome (Hamilton Rating Scale for Depression score) and logistic model (pseudo r^2) predicting a dichotomized version of the same outcome (remission). For comparison, the equality relationship is shown as a dashed line with a slope of one.

Figure 5. Number needed to assess

The figure shows the NNA and proportion variance explained in a logistic regression model (pseudo r^2) in the prediction of a dichotomized version of the same outcome (remission) by a continuous predictor. The horizontal dashed line marks the criterion for clinical significance (proportion of variance explained of 6.3%). The vertical dashed line and 2.98 mark the NNA corresponding to the clinical significance criterion. NNA: Number needed to assess.