Pathways to Neuroprediction: Opportunities and challenges to prediction of treatment response in depression

Scott A. Langenecker, Ph.D.[1], **Natania A. Crane, Ph.D.**[1], **Lisanne M. Jenkins, Ph.D.**[1], **K. Luan Phan, M.D.**[1], and **Heide Klumpp, Ph.D.**[1]

[1]The University of Illinois at Chicago

Abstract

Purpose of review—We set out to review the current state of science in neuroprediction, using biological measures of brain function, with task based fMRI to prospectively predict response to a variety of treatments.

Recent findings—Task-based fMRI neuroprediction studies are balanced between whole brain and ROI specific analyses. The predominant tasks are emotion processing, with ROIs based upon amygdala and subgenual anterior cingulate gyrus, both within the salience and emotion network. A rapidly emerging new area of neuroprediction is of disease course and illness recurrence. Concerns include use of open-label and single arm studies, lack of consideration of placebo effects, unbalanced adjustments for multiple comparisons (over focus on type I error), small sample sizes, unreported effect sizes, overreliance on ROI studies.

Summary—There is a need to adjust neuroprediction study reporting so that greater coherence can facilitate meta analyses, and increased funding for more multiarm studies in neuroprediction.

Keywords

Depression; prediction; fMRI; networks; cognitive behavioral therapy; psychopharmacology

Introduction

A. The State of Neuroscience and Clinical Care in MDD

Major Depressive Disorder (MDD) is a brain disease, or heterogeneous set of brain diseases. It can lead to temporary or permanent disruptions in emotions, problem solving, attention, motivation, and sleep. MDD also has a high prevalence, now with lifetime estimates at nearly 20% (1). Indeed, the majority who experience a depressive episode will re-experience the illness within 2 years, with some estimates at 80% recurrence (2–4). Moreover, it is becoming increasingly clear that environmental and "endogenous" risk factors for MDD

become evident in childhood, even if the majority of individuals who have these risk factors do not express illness.

It is surprising, then, that so little knowledge exists about how to prevent or treat MDD. The research expenditure for MDD has paled in comparison to other diseases that are rarer and less costly – MDD is estimated to result in a 500 million dollar loss of productivity and earnings per year in the US alone. Current treatments for MDD have limited success, and no clinical predictors of treatment response exist that could be used on the individual patient level. Unfortunately, longitudinal studies that might better illustrate the risk and expression factors for MDD are expensive and difficult to maintain using standard extramural funding cycles, so only a limited number of studies have followed samples longitudinally.

One particular challenge for treatment identification and prediction is that MDD is a highly heterogeneous condition that can present with any number and severity of symptoms for treatment identification and prediction. For example, one of the best-studied mechanisms for mood disorders is the Hypothalamic Pituitary Adrenal (HPA) access. Yet even this well-established model has led to a highly heterogeneous set of reports, and recent failure of multiple clinical trials for HPA axis modifiers (5–13). To this end, one could consider MDD as a multidimensional condition, as the neural underpinnings of any of the symptoms are (a) shared with many other conditions, and (b) supported by unique and integrated neural circuits (14). The Research Domain Criteria (RDoC) is one framework by which the field has begun to deconstruct the substrates that are potential risks for MDD, or that might be adversely affected by MDD.

A related challenge is that treatment for MDD has almost uniformly consisted of repurposed treatments for alternative conditions, or accidental findings. Of these treatments, there is not a clear understanding of the mechanisms by which they work. For example, SSRI action happens over the course of hours, whereas the treatment effect is not evident for weeks (15). More recently, alternative treatments like repetitive TMS, magnetic seizure therapy, and ketamine have enabled us to imagine and test different pathways and mechanisms for MDD (16, 17). Furthermore, large clinical trials like I-SPOT have solidified some pathways on treatment response (18, 19), and EMBARC holds promise for new insights (20). It is perhaps a time for some critical reflection and discussion.

B. The Promise of Neuroprediction in MDD

Given the many challenges toward understanding factors involved in risk for and expression of MDD, neuroscientists have designed experiments as bottom-up tests of response in MDD. The idea was to understand what neural mechanisms and pathways mediate or predict (neuroprediction) response for some individuals, so that specific treatments could be targeted for certain individuals. First, there was a hope that neuroprediction might illuminate the mechanisms by which standard treatments effect change in MDD, and for whom. Second, there was also hope that different treatments might have different predictive capacity based upon regions and networks (21). Indeed, there is mounting evidence that brain activity is better than standard clinical measures at predicting treatment outcome (19, 22, 23). Unfortunately, neither of those hypotheses has led to any clear breakthroughs as of

yet – what makes for significant prediction at the group level may not be specific enough to transfer to the individual patient (24).

In addition, in many studies, samples that are eligible for MDD imaging trials tend to be younger, more highly educated, less severely ill, absent many other medical and clinical comorbidities, and with lower body mass index. Each of these factors alone, and in combination, results in a far greater likelihood of a better treatment response (25). In contrast, studies of treatment resistant depression (TRD) are plagued by variable definitions of MDD treatment resistance, and more heterogeneous, chronic representations of the illness (26–30). How might we integrate these disparate results?

C. Predicting Treatment Response in MDD

By our reckoning, there are well over 60 studies that have attempted to use neuroimaging measures to predict treatment response (See Table 1 for a subset of just task-based fMRI studies). The majority of these studies have been open-label, unblinded medication trials, many of which are discussed in two separate meta analyses (31, 32) and three recent review articles (33–35). These studies include a near term prediction of reduction or resolution of depressive symptoms, typically over 4–16 weeks.

Prediction studies fall into four broad categories of RDoC domains – negative valence processing (e.g., emotion reactivity, attentional bias), positive valence processing (e.g., reward), cognitive control and working memory (e.g., inhibitory control), and resting. In addition, the methodologies for measurement of brain function are multifaceted, including EEG, MRI, fMRI (also including ASL, rs-fMRI, ALFF, ICA techniques), PET, SPECT, and fNRIS. We focus in this review on task-based fMRI, the predominant strategy, especially since other reviews have included PET, ASL, EEG, and rs-fMRI. One of the main differences across studies is whether whole brain vs region of interest analyses were used. Moreover, different studies have used different thresholds for significance, which, when combined with variable and often small sample sizes (see below) results in an investigator/team effect outside the effects of measurement. Finally, though, the median number of subjects is about 20 (see Table 1), which limits the nature of accuracy in regression models (24, 36). Or for example, in t test comparisons of responders vs non-responders, for example, if a majority of those enrolled will be treatment "responders" then the clinically more meaningful group (non-responders) are underweighted within the model. Overall, these considerations and variations render integration and interpretation challenging.

D. Existing Studies and Networks in Treatment Prediction in MDD

Some early studies, reviews, and meta-analyses have honed in on key neural circuits involved in neuroprediction of treatment response (31–34), and we will only briefly retouch upon these here, offering a network-based framework for interpretation. The utility of a "key region" (KR) predictive model is balanced by the reality of the level of precision in the data (smoothness), the nature of network-based functioning, and realizations that there is more scanner noise in a given region, particularly for fMRI. For example, the ability to replicate an exact KR is not the same as replicating a performance or self-report predictor in the standard sense of replication. Differences in scanners, software, preprocessing pipelines

(including choices made regarding realignment, slice-timing, normalization, standard templates, smoothing kernel), and quality control procedures, substantially adds to variability in 3D coordinate system replication, beyond the assessment of the effect size. An alternative strategy might be to employ some additional smoothing function when a meta analytic tool, such as GingerAle, is employed. As a result, here we organize the regions by virtue of recent network parcellations. For simplicity, we focus on the three-network model of Menon and colleagues (see Table 1), while acknowledging that other researchers choose to parcellate networks and subnetworks differently (37).

Salience and Emotion Network—Within the concept of a three-network model, there is one network that prioritizes processing and reacting to information that has a high salience level, including self-relevant and often emotionally-laden stimuli. As a result, it can be difficult to dissociate what could be considered emotional and not salient, and even more challenging to segregate out those experiences that might be self-relevant but somehow not salient. One can imagine, then, that these categorizations do not lend themselves to a transparent set of questions or answers within neuroimaging research. As such, we combine these two highly overlapping concepts into one broad network. The key nodes involved in this broad network are amygdala, subgenual cingulate, dorsal cingulate, ventral striatum, and anterior insula (possibly more ventral). The rostral and subgenual anterior cingulate have been implicated in prediction studies across modality and task (24, 25, 31, 34, 38, 39). The amygdala has been a much trickier region to study (more heavily targeted, but inconsistent results) in prediction of response (19, 40–42). The insula has also been observed in a few task paradigms (32, 43, 44). Within studies of reward, ROI approaches have been very common, with hypoactivation in ventral striatum and subgenual cingulate as predictors of poor treatment response (45–47).

Cognitive Control Network—The cognitive control network (CCN), is thought to be a large subnetwork within the "task-positive" network. It is thought to prioritize processing of information in relation to planning, organization, sequencing, stopping and starting, and processing of mental operations. Key nodes are the dorso-lateral prefrontal cortex (DLPFC), the inferior parietal lobule, and dorsal anterior cingulate. Of the few studies that have employed cognitive control or working memory studies, the importance of the right DLPFC in treatment prediction has been reported in several studies, primarily cognitive control and working memory tasks (18, 24, 38, 48). Notably, many negative valence studies have also reported DLPFC activation as a predictor of treatment response (39, 49–52). It is possible and even likely that emotion regulation engages cognitive control regions to aid in managing the emotional response, even if it is somewhat unclear about the level of volitional control a particular patient or control participant might have over such regions.

Default Mode Network (DMN)—The DMN is a distributed neural network encompassing a large amount of medial cortex, proximal to both anterior and posterior aspects of the medial prefrontal cortex. It also includes nodes within medial and lateral temporal and parietal cortex. It is thought to represent a host of functions including memory, self-referential thought, theory of mind (37). Because DMN is a task-negative network, it has been relatively understudied in task domains, yet it does routinely activate in prediction of

treatment response for affective paradigms (34, 50, 53–55). The medial prefrontal cortex is a frequently reported predictor in fMRI task-based studies, and includes/extends into rostral cingulate, mostly for emotion perception/processing studies, primarily in emotion perception, processing and regulation tasks (50, 51). The most ventral aspect of the posterior cingulate, extending into posterior hippocampus, is a reported predictor for both cognitive and affective paradigms (50, 55, 56). The anterior hippocampus, has also been reported as a predictor, primarily in emotion processing studies (49).

Other Regions—Surprisingly, visual cortex and cerebellum appear also in such task-based fMRI studies, irrespective of mechanism (24, 31, 34, 38, 49). Not surprisingly, these results are often underspecified and under-discussed. Potentially due to the lack of theory regarding the potential contributions of these regions to MDD, such findings nevertheless cause us to pause in making assumptions about network or KR specificity in neuroprediction, and require further study.

Recent studies have even looked at comparative and integrative prediction of different neuroimaging approaches (24, 57). The goal is to obtain a treatment prediction accuracy of >95% (binary question of whether this particular patient will achieve remission) so that such predictors could be used in treatment prescriptive studies. More importantly, using combined/comparative treatment studies could identify highly accurate moderators and mediators of treatment response, so that a prescriptive clinical imaging design could be planned. The results of such studies could inform newer guided clinical trials where the key outcome is time to remission. If we could reduce the median time to wellness by weeks or even months, a considerable degree of the "burden" of depression could be reversed.

One recent study by our group combined behavioral, task-fMRI, and task-fMRI with independent component analysis in an integrative predictive model that achieved 89% accuracy in prediction of treatment response, including steps with cross validation (24). Medial and lateral prefrontal cortex synchronization of activation during errors was a positive predictor of degree of treatment responsiveness, and accuracy of prediction significantly increased when combined with poorer behavioral inhibitory control and increased activation in several prefrontal regions. In addition, one of the I-SPOT reports suggested that hypo-reactivity in emotional stimuli within the amygdala was successful in predicting treatment response with 75% accuracy (19).

E. Predicting Risk and Disease Course

Another potential use of neuroimaging studies in MDD is to predict disease course or recurrence of illness. Of the studies that have been conducted, initial results are interesting and potentially promising. However, there are relatively few studies of this type. The distinction between risk and disease course studies, is that they study individuals over a longer period of time (e.g., 6 months up through decades of follow-up), and the goal is to predict a distant event. The likelihood that such studies will yield a positive predictor is quite modest. In fact, this weakened predictive capacity is compounded by attrition, and further limited by the tendency for negative studies to go unpublished (58), and the difficulty in publishing replication studies (see below for more details).

A prior review identified biological markers of vulnerability in at-risk youth (59). Despite the many challenges of publishing longitudinal data, investigator variability in task/physiological/measurement probes that were used, and the marked cost of longitudinal studies, this review noted that there were a significant number of biological predictors that were reported by more than one group. For example, EEG measured alpha band power, P300, and frontal asymmetry all demonstrated some degree of hereditary convergence. Studies that use fMRI measures to predict outcome are few and far between. Recently, the longitudinal assessment of manic symptoms (LAMS) multisite group suggested that self-report and neuroimaging markers could account for 28% of variance in future manic symptoms (60). This study followed 78 at risk youth for an average of 15 months and using cingulum connectivity and connectivity from the ventral striatum to the parietal cortex as predictors. Another study used cross-hemispheric connectivity from subgenual anterior cingulate seeds within a psychophysiological interaction analysis during a self-blame task. Surprisingly, though, the resilient group (no MDD recurrence) was different from the healthy control group, whereas the group with recurrent MDD did not noticeably differ from the HC group (61). Recurrence was predicted with 75% accuracy in this sample. A final example of the utility of fMRI in the prediction of treatment response comes from a study of anterior cingulate volume, which predicted 52% of variance in future depression scores, along with other relevant clinical data (62). While these studies are encouraging, more are needed. As an example of what might be conducted in future studies, a recent paper used discrete-time Markov Chain with finite states (based upon 1 year of monthly self report questionnaires) to define latent symptom classes in 209 adults with bipolar disorder (63). These repeated measures type analyses with MDD combined with biological measures in patients with MDD could be very helpful for predictive future states and course of illness.

F. Open Label Studies, Placebo Response and the Specificity of Clinical Prediction

The majority of predictive studies in MDD with fMRI have been open label studies. Those who have longstanding interest in clinical trials have questioned the internal validity of open label, one-arm predictive studies, because there is no comparison treatment, the treatment is not blinded, and there is no placebo control. We agree that single arm studies have challenges for specificity of prediction – but if replicated these studies still may offer some prescriptive value. We highlight the importance of the control group to evaluate effects of time and maturation independent of the treatment condition (64).

Placebo-controlled designs have several challenges and merits, as there are opportunities in these designs to distinguish treatment specific and more generic effects of help-seeking and return to wellness. The role of placebo responding is an important consideration in treatment prediction modeling. Many studies suggest that placebo responding can be nearly as good as the effects of an active treatment (20, 65). These studies have led to concerns about the biological specificity and clarity of diagnoses and treatments, including with MDD. They have also led to broader concerns with specificity of treatments for MDD.

More recently, our group has focused on whether placebo responding is in fact distinct in any way from response to a standard psychiatric medication (66). Notably, the individuals who are most responsive to a suggestive placebo effect are also the ones who show the

greatest responsiveness to a psychoactive treatment had greater *u*-opioid release during placebo in the nucleus accumbens. The EMBARC study should also be able to address this question in some detail.

Psychometric considerations beyond placebo include natural resolution of illness, regression to the mean, and the close links between hopefulness, behavioral activation, and placebo responding (65). Continued innovation is needed to better understand placebo response, and perhaps how placebo might be marshaled to facilitate, enhance or extend the effects of psychoactive medications and psychotherapies. In summary, although we agree that single arm studies clearly have challenges for specificity of prediction, if replicated these studies still may offer some prescriptive value.

G. Power, Clinical Significance, Effect Sizes, and Adjustments for Multiple Comparisons

There is continued misunderstanding within the imaging field (although it is not relegated solely to imaging studies) about the role of statistical adjustment of accepted type I error rate, the relationships of statistical threshold adjustment to chances for replication, and whether such one-off studies can actually diminish the type I error without negatively compromising a scientific line of inquiry. Statisticians often counsel on the careful selection of a p value to balance out the nature of a false positive, type I error, vs a type II, negative error (36, 67). In addition, the concept of meaningfulness of a significant effect – does it help us to understand illness, treatment with a reasonable degree of precision and effect size, is often lost in the discussion (64, 68). We hope to illustrate that the concerns about type I error are valid, but that they have mislead reviewers and the field into a p value war that can only sacrifice type II error, clinical significance - and will very likely reduce the capability of time tested strategies like replication and meta-analysis. Figure 1, Panel A is an actual illustration of the relationship between sample size and statistical significance using GPower. We set alpha at .005, as our experience suggests that this threshold has a balance between statistical stringency and clinical significance. To achieve significance with an alpha of .005 and power of .80, an effect size of 1.25 (very large) is needed with equal samples of 20. This means that many comparison studies are underpowered for large and medium effect sizes, they would have a higher likelihood for non-significant in this scenario (type II error). This is particularly troubling, as the vast majority of medical treatments have small to moderate effect sizes (Figure 1, Panel B). So, would we counsel throwing away the baby with the bathwater?

This illustration enables us to see what types of effect sizes would be significant with a given sample size. This is compounded by the reality that large effect sizes may be no more or less likely to replicate than very large effect sizes. This creates an unhealthy tension between whole brain analyses and ROI analyses. There may be a temptation to only report ROI analyses to avoid undergoing adjustments for multiple comparisons. Including ROI only analyses - limits the ability to conduct meaningful meta-analyses across many studies. This challenge is compounded, because few groups are capable and motivated toward carrying out treatment studies with biological markers.

A brief comment on adjustments for multiple comparisons. We contend that the unbalanced concern about p value adjustments, with multiple comparisons in mind, has created a

mindset in authors, reviewers and editors that is not conducive to evaluating the relative merits of false positives vs true negatives. There is already a tremendous bias against publication of negative results, often referred to as the file drawer effect. Well-funded labs are left to resort to publishing in paid journals, if they choose to publish the findings at all. Less-well funded labs resort to planting summary results in chapters and reviews, with sparse and under-evaluated methods. No matter the outcome, the lack of published negative studies substantially limits the benefits of meta analytic and qualitative review techniques.

Finally, an unadjusted p value for a new treatment (exploratory) should be viewed differently than an unadjusted value for a known treatment. There is currently a focus on rapid fail clinical trials at the NIMH. A strict evaluation of merit based upon adjusted p value may result in type II error – a promising treatment may be relegated to the dust heap. An evaluation based upon effect size with confidence estimates around the effect size, combined with rapid extension into a replication sample, can help balance type I vs type II error. Moreover, before the rapid extension to multisite trials, it is wise to require a semi-independent replication at a separate site. An extension of R61 to R33, could be followed by a second R33 (or even concurrently run, perhaps a new R mechanism) in an independent lab with input from the PI and team from the R61/R33.

H. Neuroprediction in the RDoC Era: Current Directions and recommendations

The emergence of the RDoC era has placed the question of treatment effect sizes and specificity squarely in the cross-hairs. There are many non-specific effects of intervention; effects of therapeutic alliance, intervention time, effort toward change, regression to the mean, placebo effects, natural resolution of illness, etc. Each of these can contribute to significant "improvement" that is not related to the specific mechanisms of treatment (e.g., domain). RDoC highlights this tension because it shows how many diseases may have common and overlapping domains of illness – therefore they may also have common pathways to wellness (69–73). Anxiety and depression may share similar negative valence domain disruptions, whereas only depression might have positive valence domain dysfunction.

To date, the study of clinical predictors has tended to overly rely on a categorical-polythetic diagnostic nomenclature (e.g., DSM-IV) constricting tests to one disorder, often testing the therapeutic response in terms of rigid measures of symptom change – these are inevitably tied to categorical diagnostic systems. Given the heterogeneity of major depression and dimensional nature of symptomatology, neuropredictors of treatment response may elucidate distinct and shared pathways that interact with particular interventions. Therefore, testing the discriminant and construct validities of several RDoC domains and dimensions (e.g., reward, threat responding, loss, affect regulation) linked to circuits in experimental designs that examine response to interventions with different mechanisms of action (e.g., pharmacotherapy, psychotherapy, neuromodulation) can lead to new insights.

The convergence of anxiety and depression symptoms and effected domains suggests that there may be parallel predictors in treatment response. The RDoC initiative has encouraged us to frame our understanding of treatment mechanisms and predictors to have the broadest impact on the care of patients with major depression and other internalizing

psychopathologies (IPs (74)). More specifically, the RDoC framework is grounded on three postulates of high relevance to neuroprediction: 1) IPs as mental illnesses are disorders of brain circuits (e.g., amygdala-frontal circuitry); 2) tools of clinical neuroscience (e.g., functional neuroimaging, electrophysiology, etc.) can be used to test and advance biosignatures that will guide treatment; 3) brain-based predictors can be enhanced by a multimodal approach that incorporates different units of information that are likely to moderate or mediate neural predictors (75). This framework has the potential to catalyze research that will address knowledge gaps that have hindered progress in incorporating biological predictors into clinical practice.

Additionally, accumulating data from the literature and our teams suggest brain regions implicated in the brain pathophysiology of a disorder or even in treatment-mediated change may not be the same regions that *predict* treatment response (35, 76, 77). For example, those factors that contribute to risk of illness are thought of as endophenotypes. Those that mediate that treatment response are considered treatment targets. Those factors or biological markers that predict treatment response could be endophenotypes. They could also be treatment targets. They could, however be independent treatment predictors, and be unrelated to endophenotypes or treatment targets. RDoC studies with a focus on specific domains of dysfunction (e.g., treatment targets and/or endophenotypes), may in fact best highlight (or even expand) treatment predictors across multiple illnesses and domains. Thus, continued focus on mechanisms for endophenotypic risk is likely a different path than the advancement of biomarkers towards precision medicine (treatment targets or predictors).

I. Conclusions and recommendations

We have attempted to cover a few important issues in neuroprediction studies in MDD. To our view, there are two few studies of the neurobiological predictors of and mechanisms involved in treatment response, even of accepted clinical treatments. Neuroprediction studies offer several windows into disease, illness expression, processes of recovery, and maintenance of wellness. We recommend that concerted effort be focused towards collections of patients with internalizing disorders, often referred to as repositories. These repositories of eligible and interested patients can then be tested with different RDoC paradigms, with sufficient sample sizes, using different treatment strategies, with the idea that accumulated knowledge will improve the matching of treatments to patients for optimal outcomes. In the meantime, a great deal will be learned about how our treatments work, and for whom.

We close with some additional recommendations for uniformity in reporting for neuroprediction studies (Table 2). These can be considered as additional and complementary to already existing reporting guidelines (e.g., COBIDAS (78)), with a specific focus on data that will assist in evaluating clinical specificity, meaningfulness, and can contribute to meta analyses. We highlight again, that such reporting guidelines do not and cannot protect against a failure to replicate. They can only guide better implementation of replication studies, increased rigor. Moreover, we add that replication studies should carefully consider challenges of overfitting, p-hacking, and spatial alignment challenges. A poorly executed replication study (by sample size, design, inclusion and exclusion criteria, treatment fidelity)

has the potential for great harm. As the number and types of therapies for internalizing disorders has expanded, including many different potential mechanisms, we harbor optimism that we will move the needle forward, toward better and more precise treatment matching.

Acknowledgments

Reference List

1. Blazer DG, Kessler RC, McGonagle KA, Swartz MS. The prevalence and distribution of major depression in a national community sample: The National Comorbidity Survey. American Journal of Psychiatry. 1994; 151(7):979–86. [PubMed: 8010383]

2. Klerman GL, Weissman MM. The course, morbidity, and costs of depression. Archives of General Psychiatry. 1992; 49(10):831–4. [PubMed: 1417437]

3. Cassano GB, Akiskal HS, Savino M, Soriani A, Musetti L, Perugi G. Single episode of major depressive disorder. First episode of recurrent mood disorder or distinct subtype of late-onset depression? Eur Arch Psychiatry Clin Neurosci. 1993; 242(6):373–80. [PubMed: 8323988]

4. Greden JF. Diagnosing and treating depression earlier and preventing recurrences: still neglected after all these years 5. CurrPsychiatry Rep. 2004; 6(6):401–2.

5. Carrol B, Curtis G, Mendels J. Cerebrospinal fluid and plasma free cortisol concentrations in depression. Psychological Medicine. 1976; 6:235–44. [PubMed: 1005564]

6. Halbreich U, Asnis GM, Shindledecker R, Zumoff B, Nathan RS. Cortisol secretion in endogenous depression. I. basal plasma levels. Archives of General Psychiatry. 1985; 42

7. Akil H, Haskett RF, Young EA, Grunhaus L, Kotun J, Weinberg V, et al. Multiple HPA profiles in endogenous depression: Effect of age and sex on cortisol and beta-endorphin. Biol Psychiatry. 1993; 33:73–85. [PubMed: 8382535]

8. Axelson DA, Doraiswamy PM, McDonald WM, Boyko OB, Tupler LA, Patterson LJ. Hypercortisolemia and hippocampal changes in depression. Psychiatry Res. 1993; 47:163–73. [PubMed: 8341769]

9. Posener JA, DeBattista C, Williams GH, Kraemer HC, Kalehzan BM, Schatzberg AF. 24-hour monitoring of cortisol and corticotropin secretion in psychotic and non-psychotic major depression. Archives of General Psychiatry. 2000; 57:755–60. [PubMed: 10920463]

10. Belanoff JK, Kalehzan M, Sund B, Ficek SKF, Schatzberg AF. Cortisol activity and cognitive changes in psychotic major depression. American Journal of Psychiatry. 2001; 158:1612–6. [PubMed: 11578992]

11. Bremner JD, Vythilingam M, Vermetten E, Anderson G, Newcomer JW, Charney DS. Effects of glucocorticoids on declarative memory function in major depression. BiolPsychiatry. 2004; 55(8): 811–5.

12. Young EA, Abelson JL, Cameron OG. Effect of comorbid anxiety disorders on the hypothalamic-pituitary-adrenal axis response to a social stressor in major depression. BiolPsychiatry. 2004; 56(2):113–20.

13. Zorrilla EP, Koob GF. Progress in corticotropin-releasing factor-1 antagonist development. Drug Discovery Today. 2010; 15:371–83. [PubMed: 20206287]

14. Langenecker SA, Jacobs RH, Passarotti AM. Current Neural and Behavioral Dimensional Constructs across Mood Disorders. Current Behavioral Neuroscience Reports. 2014; 1:114–53.

15. Stahl SM. Mechanism of action of serotonin selective reuptake inhibitors. Serotonin receptors and pathways mediate therapeutic effects and side effects. Journal of Affective Disorders. 1998; 51:215–35. [PubMed: 10333979]

16. Leuchter A, Bissett J, Manberg P, Carpenter L, Massaro JM, George M. The relationship between the individual alpha frequency (IAF) and response to synchronized transcranial magnetic

stimulation (sTMS) for treatment of major depressive disorder (MDD). Brain stimulation. 2017; 10:492.

17. Maeng S, Zarate CA Jr, Du J, Schloesser RJ, McCammon J, Chen G, et al. Cellular mechanisms underlying the antidepressant effects of ketamine: role of alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors. BiolPsychiatry. 2008; 63(4):349–52.

18**. Gyurak A, Patenaude B, Korgaonkar MS, Grieve SM, Williams LM, Etkin A. Frontoparietal Activation During Response Inhibition Predicts Remission to Antidepressants in Patients with Major Depression. Biological Psychiatry. 2015; 79:274–81. *I-SPOT study highlights the role of a multi-arm treatment prediction study, and the idea that neuroprediction might be used to guide differential treatment decisions.* [PubMed: 25891220]

19**. Williams LM, Korgaonkar MS, Song YC, Paton R, Eagles S, Goldstein-Piekarski A, et al. Amygdala Reactivity to Emotional Faces in the Prediction of General and Medication-Specific Responses to Antidepressant Treatment in the Randomized iSPOT-D Trial. Neuropsychopharmacology. 2015; 40:2398–408. *I-SPOT study highlights the role of a multi-arm treatment prediction study, and also of testing prediction accuracy so that meaingfulness is not defined soley by p values.* [PubMed: 25824424]

20. Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, et al. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. J Psychiatr Res. 2016; 78:11–23. [PubMed: 27038550]

21**. Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, et al. Functional Connectivity of the Subcallosal Cingulate Cortex And Differential Outcomes to Treatment With Cognitive-Behavioral Therapy or Antidepressant Medication for Major Depressive Disorder. American Journal of Psychiatry. 2017; 174(6):533–45. *The study highlights multi-arm study of standardly available treatments, a easily implemented prediction model (if replicated), and testing prediction accuracy so that meaingfulness is not defined soley by p values.* [PubMed: 28335622]

22. Ball TM, Goldstein-Piekarski AN, Gatt JM, Williams LM. Quantifying person-level brain network functioning to facilitate clinical translation. Transl Psychiatry. 2017; 7(10):e1248. [PubMed: 29039851]

23. Hahn T, Kircher T, Straube B, Wittchen HU, Konrad C, Strohle A, et al. Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. JAMA Psychiatry. 2015; 72(1):68–74. [PubMed: 25409415]

24**. Crane NA, Jenkins LM, Bhaumik R, Dion C, Gowins JR, Mickey BJ, et al. Multidimensional prediction of treatment response to antidepressants with cognitive control and functional MRI. Brain. 2017; 140:472–86. *Study highlights strategy of using mutimodal and multidemensional techniques to enhance prediction, and testing prediction accuracy so that meaingfulness is not defined soley by p values.* [PubMed: 28122876]

25. Young KD, Drevets WC, Bodurka J, Preskorn SS. Amygdala activity during autobiographical memory recall as a biomarker for residual symptoms in patients remitted from depression. Psychiatry Research: Neuroimaging. 2016; 248:159–61. [PubMed: 26809279]

26. Kornstein S, Schneider R. Clinical features of treatment-resistant depression. Journal of Clinical Psychiatry. 2001; 62(S16):S18–S25.

27. Shah PJ, Glabus MF, Goodwin GM, Ebmeier KP. Chronic, treatment-resistant depression and right fronto-striatal atrophy. BrJPsychiatry. 2002; 180:434–40.

28. Fagiolini A, Kupfer DJ. Is treatment-resistant depression a unique subtype of depression? Biological Psychiatry. 2003; 53(8):640–8. [PubMed: 12706950]

29. Duhameau B, Ferré J, Jannin P, Gauvrit J, Vérin M, Millet B, et al. Chronic and treatment-resistant depression: A study using arterial spin labeling perfusion MRI at 3 Tesla. Psychiatry Research: Neuroimaging. 2010; 182:6.

30. de Kwaasteniet BP, Rive MM, Ruhe HG, Schene AH, Veltman DJ, Fellinger L, et al. Decreased Resting-State Connectivity between Neurocognitive Networks in Treatment Resistant Depression. Front Psychiatry. 2015; 6:28. [PubMed: 25784881]

31. Pizzagalli DA. Frontocingulate Dysfunction in Depression: Toward Biomarkers of Treatment Response. Neuropsychopharmacology Reviews. 2010; 36:183–206. [PubMed: 20861828]

32. Fu CH, Steiner H, Costafreda SG. Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. Neurobiology of Disease. 2013; 52:75–83. [PubMed: 22659303]

33**. Phillips ML. Identifying Predictors, Moderators, and Mediators of Antidepressant Response in Major Depressive Disorder: Neuroimaging Approaches. American Journal of Psychiatry. 2016; 172:124–38. *A model driven review of neuroprediction and treatment-induced change, and integrates the idea of mediation and moderation in treatment change in treatments for depression.*

34. Dichter GS, Gibbs D, Smoski MJ. A systematic review of relations between resting-state functional-MRI and treatment response in major depressive disorder. Journal of Affective Disorders. 2015; 172:8–17. [PubMed: 25451389]

35. Wessa M, Lois G. Brain Functional Effects of Psychopharmacological Treatment in Major Depression: a Focus on Neural Circuitry of Affective Processing. Current neuropharmacology. 2015; 13(4):466–79. [PubMed: 26412066]

36. Stevens, J. Applied multivariate statistics for the social sciences. Mahwah, NJ: Lawrence Ehrlbaum Associates; 1996. p. 659

37. Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. Journal of Neuroscience. 2007; 27:2349–56. [PubMed: 17329432]

38. Langenecker SA, Kennedy SE, Guidotti LM, Briceno EM, Own LS, Hooven T, et al. Frontal and limbic activation during inhibitory control predicts treatment response in major depressive disorder. Biol Psychiatry. 2007; 62(11):1272–80. [PubMed: 17585888]

39. Lopez-Sola M, Pujol J, Hernandez-Ribas R, Harrison BJ, Contreras-Rodriguez O, Soriano-Mas C, et al. Effects of Duloxetine Treatment on Brain Response to Painful Stimulation in Major Depressive Disorder. Neuropsychopharmacology. 2010; 2010:2305–17.

40. Victor TA, Furey ML, Fromm SJ, Öhman A, Drevets WC. Relationship between amygdala responses to masked faces and mood state and treatment in major depressive disorder. Archives of General Psychiatry. 2010; 67(11):1128–38. [PubMed: 21041614]

41. Doerig N, Krieger T, Altenstein D, Schlumpf Y, Spinelli S, Spati J, et al. Amygdala response to self-critical stimuli and symptom improvement in psychotherapy for depression. British Journal of Psychiatry. 2016; 208(2):175–81. [PubMed: 26494872]

42. Siegle GJ, Thompson WK, Collier A, Berman SR, Feldmiller J, Thase ME, et al. Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. Arch Gen Psychiatry. 2012; 69(9):913–24. [PubMed: 22945620]

43. Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath C, Choi KS, et al. Functional Connectivity of the Subcallosal Cingulate Cortex And Differential Outcomes to Treatment With Cognitive-Behavioral Therapy or Antidepressant Medication for Major Depressive Disorder. American Journal of Psychiatry. 2017; 174:533–45. [PubMed: 28335622]

44. Davidson RJ, Irwin W, Anderle MJ, Kalin NH. The neural substrates of affective processing in depressed patients treated with venlafaxine. American Journal of Psychiatry. 2003; 160(1):64–75. [PubMed: 12505803]

45. Forbes EE, Olino TM, Ryan ND, Birmaher B, Axelson DA, Moyles DL, et al. Reward-related brain function as a predictor of treatment response in adolescents with major depressive disorder. Cognitive Affective & Behavioral Neuroscience. 2010; 10:107–18.

46. Straub J, Plener PL, Sproeber N, Sprenger L, Koelch MG, Groen G, et al. Neural correlates of successful psychotherapy of depression in adolescents. Journal of Affective Disorders. 2015; 183:239–46. [PubMed: 26025370]

47. Carl H, Walsh E, Eisenlohr-Moul T, Minkel J, Crowther A, Moore T, et al. Sustained anterior cingulate cortex activation during reward processing predicts response to psychotherapy in major depressive disorder. Journal of Affective Disorders. 2016; 203:204–12. [PubMed: 27295377]

48. Walsh ND, Williams SCR, Brammer MJ, Bullmore ET, Kim J, Suckling J, et al. A Longitudinal Functional Magnetic Resonance Imaging Study of Verbal Working Memory in Depression After Antidepressant Therapy. Biological Psychiatry. 2007; 62:1236–43. [PubMed: 17601497]

49. Costafreda SG, Khanna A, Mourao-Miranda J, Fu CH. Neural correlates of sad faces predict clinical remission to cognitive behavioural therapy in depression. Neuroreport. 2009; 20:637–41. [PubMed: 19339907]

50. Miller JM, Schneck N, Siegle GJ, Chen Y, Ogden RT, Kikuchi T, et al. fMRI response to negative words and SSRI treatment outcome in major depressive disorder: a preliminary study. Psychiatry Research. 2013; 214:296–305. [PubMed: 24446548]

51. Ritchey M, Dolcos F, Eddington KM, Strauman TJ. Neural correlates of emotional processing in depression: Changes with cognitive behavioral therapy and predictors of treatment response. Journal of Psychiatric Research. 2011; 45:577–87. [PubMed: 20934190]

52. Fu CH, Mourao-Miranda J, Costafreda S, Khanna A, Marquand A, Williams S, et al. Pattern Classification of Sad Facial Processing: Toward the Development of Neurobiological Markers in Depression. Biological Psychiatry. 2008; 63:7.

53. Dichter GS, Felder JN, Smoski MJ. The effects of Brief Behavioral Activation Therapy for Depression on cognitive control in affective contexts: An fMRI investigation. J Affect Disord. 2010; 126(1–2):236–44. [PubMed: 20421135]

54. Delaveau P, Jabourian M, Lemogne C, Allaili N, Choucha W, Girault N, et al. Antidepressant short-term and long-term brain effects during self-referential processing in major depression. Psychiatry Res. 2016; 247:17–24.

55. Rizvi SJ, Salomons TV, Konarski JZ, Downar J, Giacobbe P, Mcintyre RS, et al. Neural response to emotional stimuli associated with successful antidepressant treatment and behavioral activation. Journal of Affective Disorders. 151:573–581.

56. Keedwell PA, Drapier D, Surguladze S, Giampietro V, Brammer M, Phillips ML. Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. Journal of Affective Disorders. 2010; 120:120–5. [PubMed: 19539998]

57. Gong Q, Wu Q, Scarpazza C, Lui S, Jia Z, Marquand A, et al. Prognostic prediction of therapeutic response in depression using high-field MR imaging. Neuroimage. 2011; 55:1497–503. [PubMed: 21134472]

58. Nissen SB, Magidson T, Gross K, Bergstro CT. Research: Publication bias and the canonization of false facts. eLife. 2016; 5:e21451. [PubMed: 27995896]

59. Jacobs RH, Orr JL, Gowins JR, Forbes EE, Langenecker SA. Biomarkers of Intergenerational Risk for Depression: A Review of Mechanisms in Longitudinal High-Risk (LHR) Studies. Journal of Affective Disorders. 2015; 175:494–506. [PubMed: 25687188]

60. Bertocci MA, Bebko G, Versace A, Fournier JC, Iyengar S, Olino T, et al. Predicting clinical outcome from reward circuitry function and white matter structure in behaviorally and emotionally dysregulated youth. Mol Psychiatry. 2016; 21(9):1194–201. [PubMed: 26903272]

61. Lythe KE, Moll J, Gethin JA, Workman CI, Green S, Ralph MAL, et al. Self-blame–Selective Hyperconnectivity Between Anterior Temporal and Subgenual Cortices and Prediction of Recurrent Depressive Episodes. JAMA psychiatry. 2015; 72(11):1119–26. [PubMed: 26445229]

62. Serra-Blasco M, de Diego-Adelino J, Vives-Gilabert Y, Trujols J, Puigdemont D, Carceller-Sindreu M, et al. Naturalistic Course of Major Depressive Disorder predicted by Clinical and Structural Neuroimaging Data: A Five Year Follow-Up. Depression and Anxiety. 2016; 33:1055–64.

63**. Cochran AL, McInnis MG, Forger DB. Data-driven classification of bipolar I disorder from longitudinal course of mood. Translational Psychiatry. 2016; 6:e912. *Highlights the importance on longitudinal studies in the understanding of disease subtypes, which may be relevant for neuroprediction of treatment response and disease course.* [PubMed: 27727242]

64. Feingold A. Effect Sizes for Growth-Modeling Analysis for Controlled Clinical Trials in the Same Metric as for Classical Analysis. Psychological Methods. 2009; 14:43–53. [PubMed: 19271847]

65. Brunoni AR, Lopes M, Kaptchuk TJ, Fregni F. Placebo response of non-pharmacological and pharmacological trials in major depression: a systematic review and meta-analysis. PLoS One. 2009; 4(3):e4824. [PubMed: 19293925]

66. Peciña M, Bohnert AS, Sikora M, Avery ET, Langenecker SA, Mickey BJ, et al. Placebo-Activated Neural Systems are Linked to Antidepressant Responses: Neurochemistry of Placebo Effects in Major Depression. JAMA psychiatry. 2015; 72(11):1087. [PubMed: 26421634]

67. Cohen J. A power primer. Psychological Bulletin. 1992; 112:155–9. [PubMed: 19565683]

68. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting & Clinical Psychology. 1991; 59:12–9. [PubMed: 2002127]

69. Jenkins LM, Barba A, Campbell MM, Lamar M, Shankman SA, Leow A, et al. Shared white matter alterations across emotional disorders: A voxel-based meta-analysis of fractional anisotropy. NeuroImage: Clinical. 2016; 12:1022–34. [PubMed: 27995068]

70. Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA. Large-Scale Network Dysfunction in Major Depressive Disorder: A Meta-analysis of Resting-State Functional Connectivity. JAMA Psychiatry. 2015

71. Satterthwaite TD, Kable JW, Vandekar L, Katchmar N, Bassett DS, Baldassano CF, et al. Common and Dissociable Dysfunction of the Reward System in Bipolar and Unipolar Depression Reward Dysfunction in Depression. Neuropsychopharmacology. 2015

72. McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A. Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. Am J Psychiatry. 2017; 174(7):676–85. [PubMed: 28320224]

73. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry. 2015; 72(4):305–15. [PubMed: 25651064]

74. Stange JP, MacNamara A, Barnas O, Kennedy AE, Hajcak G, Phan KL, et al. Neural markers of attention to aversive pictures predict response to cognitive behavioral therapy in anxiety and depression. Biol Psychol. 2017; 123:269–77. [PubMed: 27784617]

75. Nathan PJ, Phan KL, Harmer CJ, Mehta MA, Bullmore ET. Increasing pharmacological knowledge about human neurological and psychiatric disorders through functional neuroimaging and its application in drug discovery. Current opinion in pharmacology. 2014; 14:54–61. [PubMed: 24565013]

76**. Klumpp H, Fitzgerald JM, Kinney KL, Kennedy AE, Shankman SA, Langenecker SA, et al. Predicting cognitive behavioral therapy response in social anxiety disorder with anterior cingulate cortex and amygdala during emotion regulation. NeuroImage Clinical. 2017; 15:25–34. *Illustrates how diagnoses – that have overlapping symptoms and similar treatments may also have similar neuroprediction models.* [PubMed: 28462086]

77. Falconer E, Allen A, Felmingham KL, Williams LM, Bryant RA. Inhibitory neural activity predicts response to cognitive-behavioral therapy for posttraumatic stress disorder. J Clin Psychiatry. 2013; 74(9):895–901. [PubMed: 24107763]

78**. Nichols TE, Das S, Eickhoff SB, Evans AC, Glatard T, Hanke M, et al. Best Practices in Data Analysis and Sharing in Neuroimaging using MRI. bioRxiv. 2016:054262. General report of best practices in neuroimaging studies.
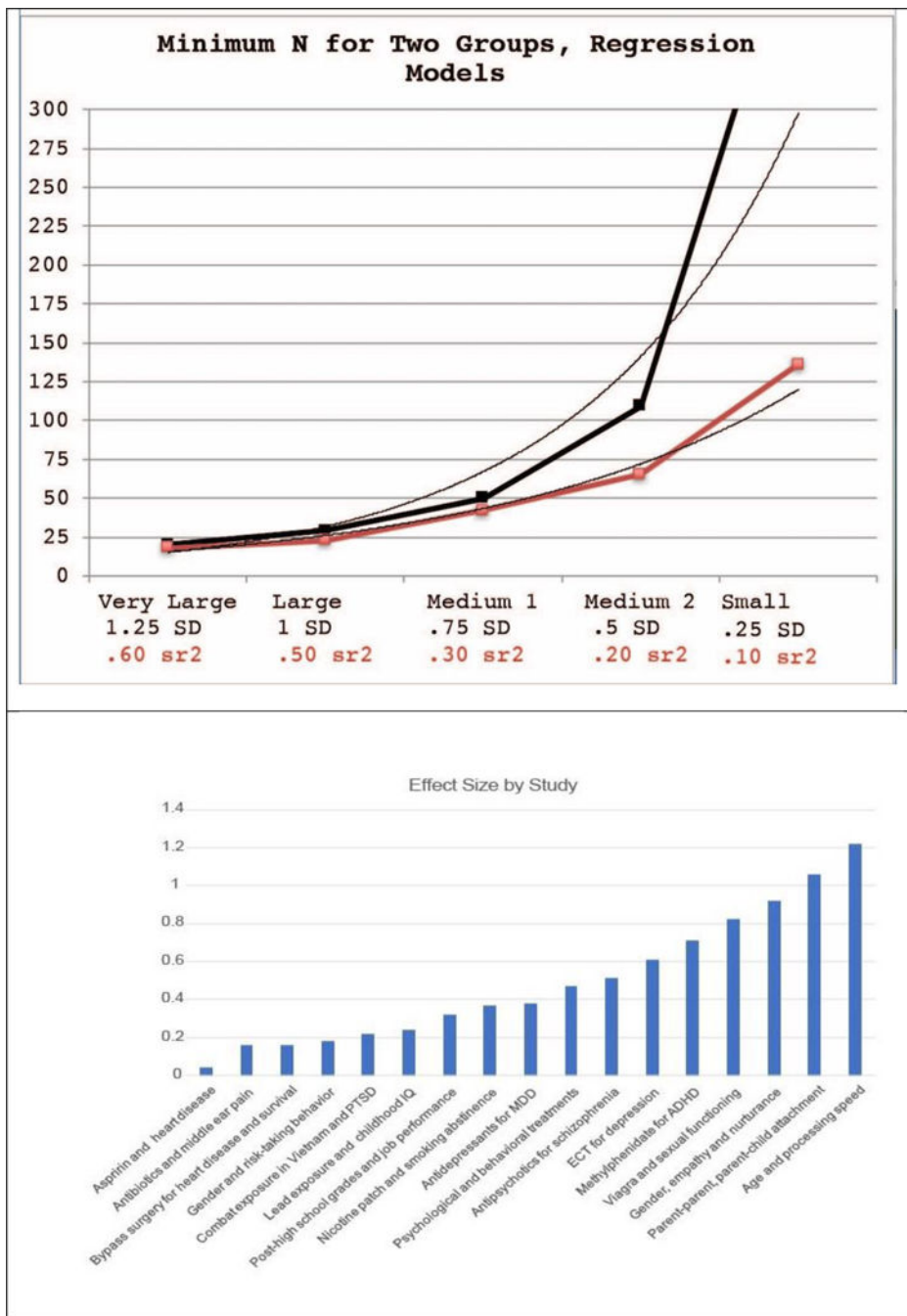
Figure 1.
A. Illustrates the observed N needed to obtain power of .80 to reject the null hypothesis, in a given voxel, based upon alpha < .005. This is independent of adjustment for cluster size. A exponential fit line is include to illustrate the relationship between sample size and power.
B. Effect sizes for comparison with 1B. Most psychotherapies are moderate effect sizes, suggesting that a similar brain effect size would have adequate power with Ns of between 50 and 109. The assumption is that the brain marker would have the same effect size as the

treatment. Brain effect sizes may be larger or smaller, as it is doubtful that they are parametrically linked. Effect sizes from Meyer et al., 2001 (69) and Leuck et al., 2013 (70).

Table 1

Neuroprediction Studies in Major Depressive Disorder with Task-Based fMRI

Study authors	N entered/N used	Task/Probe	Treatment	model (t test/regression)	whole brain or key ROIs	key regions	networks implicated	threshold
Emotion Perception/Regulation								
Chen et al., 2007	17	Sad faces	fluoxetine	regression, t test	whole brain, sgACC	anterior mid-cingulate, pregenual ACC, sgACC, Caudate	SEN, CCN	p < .005 %
Costafreda et al., 2009	16	Sad faces (Intensity-based)	CBT	SVM pattern classification of response vs non-response	whole brain	ACC, mid-cingulate, SFG, MFG, paracentral, PCC, precuneus, SPL, IPL, parahippocam pal, cerebellum	All (diffuse)	binomial distribution
Davidson et al., 2003	14/12	Affective Pictures	venlafaxine	correlation (with extracted betas)	Insular cortex and ACC (from significant group by time interaction)	L ACC	DMN	p < .05 (extracted)
Delaveau et al., 2016	30/16	IAPS pictures	agomelatine or placebo	t test	dmPFC, vmPFC, PCC, ventral ACC, rostral ACC, dorsal ACC, IFG, MFG, SFG	dmPFC, precuneus	DMN	p < .05
Dichter et al., 2010	12	CPT with Faces	Behavioral Activation	correlation (with extracted betas)	those regions with group by time interactions	paracingulate gyrus	DMN	p < .05
Doerig et al., 2016	23/21	personal adjectives	CBT	t test	amygdala	R amygdala R rolandic	SEN	p < .05
Frodl et al., 2011	27/24	Emotion Face Matching	11 mirtazipine, 13 venlafaxine	regression	whole brain	operculum, L fusiform gyrus	Visual	p < .05 *%
Fu et al., 2008a	19	Implicit sad facial recognition	fluoxetine	t test, pattern classification	amygdala, ACC, Visual, whole brain	–	–	p < .05
Fu et al., 2008b	16	Implicit sad facial recognition	CBT	correlation mixed effects	whole brain	R IFG/insula, L putamen/ globu s pallidus	CCN, SEN	p < .001
Fu et al., 2015	32/24	affect faces, emotional stroop (excluding rs-fMRI	duloxetine	model, repeated measures	whole brain	– for task	–	p< .05 *
Furey et al., 2015	16/14	Faces, Houses Stimuli	scopolamine	regression	whole brain, sgACC and occipital	sgACC, middle occipital	SEN, Visual	p < .005, k > 25 %
Godlewsk a et al., 2016	35	Fearful and Happy faces	escitalopram	t test	whole brain	– for task	–	p < .05 %
Keedwell et al., 2010	12	Affective Faces- Sad Faces	"variety of antidepressant treatments"	regression	whole brain	sgACC and visual cortex	SEN, Visual	voxelwise p = 0.05, clusterwise p = 0. 005
Lisiecka et al., 2011	27/23	Emotion Face Matching	13 venlafaxine, 10 mirtazipine	effective connectivity during task	whole brain	L precentral gyrus, R OFC, L cerebellum, thalamus	SEN, Somatomotor	*

Study authors	N entered/N used	Task/Probe	Treatment	model (t test/regression)	whole brain or key ROIs	key regions	networks implicated	threshold
Lopez-Sola et al., 2010	15/13	Pain Perception/Tolerance	duloxetine	t test	whole brain	sgACC and vmPFC, R DLFPC L	SEN, CCN, DMN	$p < .05$, $k > 200$ %
Miller et al., 2013	17/15	Personal relevance rating	escitalopram	correlation	related to task effects, plus BA 25	paracingulate, globus pallidus, post-central gyrus vmPFC, L	DMN	$p < .05$ %
Ritchey et al., 2011	22/15	IAPS pictures	CBT	correlation	Group difference	ATL/vlPFC, R DLPFC	CCN, DMN	$p < .05$
Rizvi et al., 2013	21/15	IAPS pictures	olanzapine/fluoxetine combined treatment	regression, ANOVA	whole brain	Premotor cortex, PCC, insula L dmPFC, R SFG, L PCC, R Calcarinus, R lingual gyrus, L	Sensorimot or, DMN, SEN	$p < .02$
Samson et al., 2011	21	Faces, Houses Stimuli	9 mirtazipine, 12 venlafaxine	regression	whole brain whole	cuneus, L SPL, R STG	DMN, Visual	*, #
Siegle et al., 2006	14	Personal relevance rating	CBT	regression	brain, amygdala, sgACC	sgACC, amygdala	SEN	$p < .0005$, $p < .05$
Siegle et al., 2012	20/17, 41/32	Personal relevance rating	Cognitive therapy	regression	sgACC	sgACC	SEN	$p < .001$
Spies et al.,	23	Emotion Face Matching	escitalopram	regression	whole brain		SEN	$p < .001$ %
Szczepanik et al., 2016	14	Faces, Houses, selective attention	scopolamine	regression	amygdala	amygdala	SEN	$p < .007$
Toki et al., 2014	14	Encoding of Positive words	pharmacotherapy	regression	hippocampus	L hippocampus	DMN	$p < .005$, $k = 16$
Victor et al., 2013	20/10	masked Sad and Happy faces	sertraline	regression	perigenual ACC	perigenual ACC	DMN	$p < .05$
Williams et al., 2015	80	Facial Stimuli (Gur et al., 2002)	27 escitalopram, 27 sertraline, 26 venlafaxine	GLM responder vs non-responder	amygdala	amygdala	SEN	*
Yoshimura et al., 2014	23	self-referential processing	CBT	correlation	whole brain	vACC, mPFC	DMN	$G \times Time$, $p < .01$
Reward								
Carl et al., 2016	38/33	Monetary Incentive Delay	Brief Behavioral Activation	ANOVA (with extracted betas)	striatum, mPFC, OFC, ACC	ACC	SEN	$p < .05$
Forbes et al., 2010	13/6	Card Guessing	7 fluoxetine, 6 CBT plus fluoxetine	regression	striatum, mPFC	striatum	SEN	$p < .05$, $k > 10$ ^
Straub et al., 2015	22	Monetary Incentive	CBT group	t-test (with extracted betas)	sgACC, hippocampus, amygdala, NAcc	sgACC	SEN	$p < .05$
Working Memory/Emotional Working Memory								
Furey et al., 2013	15	Emotional Working Memory	scopolamine	correlation (with extracted betas)	inferior, middle, and superior occipital cortical regions	middle occipital cortex	Visual	$p < .05$, $k > 25$ %

Author Manuscript

Study authors	N entered/N used	Task/Probe	Treatment	model (t test/regression)	whole brain or key ROIs	key regions	networks implicated	threshold
Marquand et al., 2008	16	Variable-load N-Back	fluoxetine	correlation, linear load response modeling	whole brain	–	–	p < .05 [I]
Roy et al., 2010	22/20	IAPS pictures w Memory	citalopram	Regression (with extracted betas)	whole brain	L SFG, L vmPFC, R rostral ACC, ACC, L IFG, MFG, SFG,	CCN, SEN, DMN	p < .005, k > 10
Walsh et al., 2007	20	N-back	fluoxetine	regression	whole brain, ACC	precentral, mid-cingulate, STG, MTG, occipital, cerebellum	CCN, Visual, sensorimotor	p < .005
Executive Functioning								
Crane et al., 2017	49/36	Parametric Go/No-Go	14 duloxetine, 22 s-citalopram	regression (within ICA masks)	ICA masks	dorsal ACC, rostral ACC, MFG	CCN, DMN	p < .01, k > 23 %
Gyurak et al., 2016	80	Go/No-Go, selective oddball, working memory	27 escitalopram, 27 sertraline, 26 venlafaxine	t-test (with extracted betas)	multiple CCN regions	R DLPFC, R IPL perigenual	CCN	p < .001, SVC
Hernandez-Ribas et al. 2013	21	Verbal Expression	10 TMS, 11 Sham TMS	correlations	whole brain	ACC, left medial OFC, R MFG, L ventral putamen IFG, MFG, Insula,	SEN, CCN, DMN	p < .01
Langeneck er et al., 2007	20/15	Parametric Go/No-Go	s-citalopram	regression	whole brain	Amygdala, Cerebellum, ACC	CCN	p < .0001, k=20
Thompson et al., 2015	60/44	WCST set-shifting	CBT	regression	multiple frontal	L IFG, MFG, R SFG	CCN	*extent threshold: k=50
Wagner et al., 2010	20	Stroop	12 citalopram, 8 reboxetine	regression	interference task effect	–	–	p =.001, k=25

*
FWE p < .05,

#
actual p value not reported,

^
FDR corrected, % cluster corrected threshold,

I
Using permutation testing, – no significant findings sg = subgenual, ACC = anterior cingulate cortex, SEN = salience and emotion network, CCN = cognitive control network, SFG = superior frontal gyrus, MFG = middle frontal gyrus, IFG = inferior frontal gyrus, PCC = posterior cingulate cortex, DMN = default mode network, dm = dorsomedial, vm = ventro medial, CBT = cognitive behavioral network, ATL = anterior temporal lobe, vl = ventro lateral, SPL = superior parietal lobule, STG = superior temporal gyrus, NAcc = nucleus accumbens, MTG = middle temporal gyrus,

Table 2

Reporting Recommendations for Neuroprediction Studies

ROC curves	Use of ROC curves enables a quick easy understanding of the clinical relevance of a prediction variable or model.
Effect sizes	Reporting effect sizes can help translate to clinical meaningfulness (cohen's d, or Z score, suggest including confidence intervals).
Standardization of treatment response	Use of reliable change index or multiple measures (e.g., pre-to-post percent reduction on both self-report and clinician-report) standardizes reporting and interpretation
Validation of biomarkers	Training-testing splits or cross-validation for internal validation; testing predictor performance on data sets collected from different laboratories for external validation
Omnibus, multimodal prediction strategies	Newer prediction tools can adjust for the imbalance between number of predictors and number of subjects, adjust for the impact of non-normal distributions with outliers.
Reporting results at $p < .005$ with modest cluster size correction (e.g., $k > 50$)	Supplemental Tables with whole brain, partially adjusted results, guided by clinical effect sizes, not multiple corrections. These will enable users to integrate against a uniform standard for meta-analysis
Reporting total number of analyses conducted and family wise error	Report the number of *a priori* analyses and the adjusted Familywise Error rates. This allows to the reader to better integrate the likelihood of false positive errors.
Using comparison treatments that are expected to have a different mechanism of action	Even if the interaction between treatment arm and predictors is not significant, it can provide meaningful pilot information for future studies of differential mechanism and effectiveness trials.