Objective: Reducing unsuccessful treatment trials could improve depression treatment. Quantitative EEG (QEEG) may predict treatment response and is being commercially marketed for this purpose. The authors sought to quantify the reliability of QEEG for response prediction in depressive illness and to identify methodological limitations of the available evidence.

Method: The authors conducted a meta-analysis of diagnostic accuracy for QEEG in depressive illness, based on articles published between January 2000 and November 2017. The review included all articles that used QEEG to predict response during a major depressive episode, regardless of patient population, treatment, or QEEG marker. The primary meta-analytic outcome was the accuracy for predicting response to depression treatment, expressed as sensitivity, specificity, and the logarithm of the diagnostic odds ratio. Raters also judged each article on indicators of good research practice.

Results: In 76 articles reporting 81 biomarkers, the meta-analytic estimates showed a sensitivity of 0.72 (95% CI=0.67–0.76) and a specificity of 0.68 (95% CI=0.63–0.73). The logarithm of the diagnostic odds ratio was 1.89 (95% CI=1.56–2.21), and the area under the receiver operator curve was 0.76 (95% CI=0.71–0.80). No specific QEEG biomarker or specific treatment showed greater predictive power than the all-studies estimate in a meta-regression. Funnel plot analysis suggested substantial publication bias. Most studies did not use ideal practices.

Conclusions: QEEG does not appear to be clinically reliable for predicting depression treatment response, as the literature is limited by underreporting of negative results, a lack of out-of-sample validation, and insufficient direct replication of previous findings. Until these limitations are remedied, QEEG is not recommended for guiding selection of psychiatric treatment.

Major depressive illness remains a leading worldwide contributor to disability despite the growing availability of medications and psychotherapies (1). The persistent morbidity is partly due to the difficulty of treatment selection. An adequate “dose” of cognitive-behavioral therapy for depression is 10–12 weeks (2). An antidepressant or augmentation medication trial requires at least 4 weeks at an adequate dosage (2). Patients may spend months to years searching through options before responding to treatment (3). Knowing sooner whether a treatment will be effective could increase the speed and possibly the rate of overall treatment response. The high potential value of treatment prediction biomarkers has spurred extensive research. Unfortunately, it has also encouraged commercial ventures that market predictive tests to both patients and physicians, often without the support of evidence of clinical efficacy (4). Inappropriate use of invalid “predictive” tests could easily increase health care costs without benefiting patients (5).

Predictive biomarker research emphasizes pretreatment and “treatment-emergent” biomarkers. Treatment-emergent markers are physiologic changes that precede and predict the response to effective treatment. They may represent physiologic processes mediating the clinical response, whereas pretreatment markers may represent moderating factors. If we could confidently predict a treatment’s efficacy or non-efficacy 1–2 weeks into a treatment trial, we could move much more quickly through clinical decision trees. For novel therapies such as brain stimulation, treatment-emergent markers could also guide “closed-loop” treatment, where an aspect of the stimulation is titrated in direct response to the physiologic marker (6, 7).

Electroencephalography (EEG) is a promising source of psychiatric biomarkers. Unlike serum chemistry or genetic variation, EEG directly measures brain activity. EEG is potentially more cost-effective than neuroimaging techniques.
such as functional MRI (fMRI) and nuclear medicine computed tomography (PET/SPECT), which have also been proposed as biomarkers (8–10). EEG recordings can be more feasibly implemented in a wide variety of clinical settings, and it has essentially no safety concerns, whereas PET involves radiation and MRI cannot be used in the presence of metal foreign bodies.

Psychiatric biomarker studies have emphasized quantitative EEG, or QEEG (see the text box). Baseline and treatment-emergent biomarkers, as qualitatively reviewed in recent years (11–13), include simple measures such as loudness dependence of auditory evoked potentials (LDAEP) (14–22), oscillatory power in the theta and alpha ranges (see the text box) (14, 23–39), and the distribution of those low-frequency oscillations over the scalp (35, 37, 40–45). With the increasing power of modern computers, biomarkers involving multiple mathematical transformations of the EEG signal became available. These include a metric called cordance (23, 26, 46–57) and a proprietary formulation termed the Antidepressant Treatment Response (ATR) index (57–61). Each is based on both serendipitous observations and physiologic hypotheses of depressive illness (11, 12). LDAEP is believed to measure serotoninergic function, oscillations are linked to top-down executive functions (62, 63), and cordance may reflect cerebral perfusion changes related to fMRI signals. ATR and related multivariate markers (64–66) merge these lines of thought to increase predictive power. Recent studies (including the Canadian Biomarker Integration Network in Depression [CAN-BIND], the International Study to Predict Optimized Treatment–Depression [iSPOT-D], and the Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care study [EMBARC]) have sought to create large multicenter data sets that may allow more robust biomarker identification (40, 67–72).

Despite the rich literature, the value of QEEG as a treatment response predictor in depressive illness remains unclear. This is in part because there has been no recent meta-analysis aimed at the general psychiatrist or primary care practitioner. The last formal American Psychiatric Association position statement on EEG was issued in 1991 (73), at which time personal computers had a fraction of the computing power of today’s computers. A 1997 American Academy of Neurology report (74) focused on QEEG in epilepsy and traumatic brain injury. The most recent report, from the American Neuropsychiatric Association, was similarly cognition oriented (75). All of these reports are over a decade old. More recent reviews have delved into the neurobiology of QEEG but have not quantitatively assessed its predictive power (11–13). The closest was a 2011 meta-analysis that combined imaging and EEG to assess the role of the rostral cingulate cortex in major depression (76).

To fill this gap in clinical guidance, we performed a meta-analysis of QEEG as a predictor of treatment response in depression. We cast a broad net, considering all articles on adults with any type of major depressive episode, receiving any intervention, and with any study design or outcome scale. This approach broadly evaluated QEEG’s utility without being constrained to specific theories of depression or specific markers. We complemented that coarse-grained approach with a meta-regression investigating specific biomarkers to ensure that inconsistent results across the entire QEEG field would not mask a single effective marker.

**METHOD**

Our review focused on two primary questions: What is the overall evidence base for QEEG techniques in predicting response or nonresponse in the treatment of depressive episodes? Given recent concerns about reliability in neuroimaging (10, 77), how well did published studies implement practices that support reproducibility and reliability?
We searched PubMed for articles related to EEG, major depression, and response prediction (see the online supplement). We considered articles published in any indexed year. From these, we kept all that reported prediction of treatment response, to any treatment, in any type of depressive illness, using any EEG metric. Our prospective hypothesis was that EEG cannot reliably predict treatment response. We chose broad inclusion criteria to maximize the chance of a signal detection that falsified our hypothesis. That is, we sought to determine whether there is sufficient evidence to recommend the routine use of any QEEG approach to inform psychiatric treatment. This is an important clinical question, given the commercial availability and promotion of psychiatric QEEG. We did not include studies that attempted to directly select patients’ medication based on an EEG evaluation, an approach sometimes termed “referenced EEG” (78). Referenced EEG is not a diagnostic test, and as such does not permit the same form of meta-analysis.

The meta-analysis of diagnostic markers depends on 2×2 tables summarizing correct and incorrect responder and non-responder predictions (79). Two trained raters extracted these from each article, with discrepancies resolved by discussion and final arbitration by the first author. Where necessary, table values were imputed from other data provided in the article (see the online supplement). For articles that examined more than one marker or treatment (19, 29, 52, 57, 60, 67, 80), we considered them as separate studies. We reasoned that treatments with different mechanisms of action (e.g., repetitive transcranial magnetic stimulation [rTMS] versus medication) may have different effects on reported biomarkers, even if studied by a single investigator. For studies that reported more than one method of analyzing the same biomarker (23, 34, 57), we used the predictor with the highest positive predictive value. This further increased the sensitivity and the chance of a positive meta-analytic result. Articles that did not report sufficient information to reconstruct a 2×2 table (14, 15, 17, 21, 25, 28, 32, 33, 42, 43, 81–92) were included in descriptive and study quality reporting but not in the main meta-analysis.

For quality reporting, we focused on whether the study used analytic methods that increase the reliability of conclusions. Chief among these is independent sample verification or cross-validation—reporting the algorithm’s predictive performance on a sample of patients separate from those originally used to develop it. Cross-validation has repeatedly been highlighted as essential in the development of a valid biomarker (10, 11, 61, 74, 93). Our two other markers of study quality were total sample size and correction for multiple hypothesis testing. Small sample sizes falsely inflate effect sizes (93), and correction for multiple testing is a foundation of good statistical practice.

We conducted univariate and bivariate meta-analyses using R’s mada package for analysis and metafor for visualizations (94–96). The univariate analysis summarized each study as the natural logarithm of its diagnostic odds ratio, using a random-effects estimator (79). Bivariate analysis used sensitivity and specificity following the approach of Reitsma et al. (97). From the bivariate analysis, we derived the area under the summary receiver operator curve and computed an area-under-the-curve confidence interval by 500 iterations of bootstrap resampling with replacement. For the univariate analysis, we report $I^2$ as a measure of study heterogeneity. As secondary analyses, we separated studies by biomarker type (LDAEP, power features, ATR, cortance, and multivariate) and by treatment type (medication, rTMS, or other). These were then entered as predictor variables in bivariate meta-regressions. Finally, to assess the influence of publication bias, we plotted log(diagnostic odds ratio) against its precision, expressed as both the standard error (funnel plot) and the effective sample size (98). We tested funnel plot asymmetry with the arc sine method described in Rücker et al. (99), as implemented in the meta package (100). This test has been suggested to be robust in the presence of heterogeneity and is the recommended choice of a recent working group (101). All of the above were preplanned analyses. Our analysis and reporting conform with the PRISMA guidelines (102); the checklist is included in the online supplement. The supplement also reports an alternative approach using standardized mean differences.

RESULTS

Descriptive Study Characteristics

Our initial search produced 995 articles, to which we added 28 articles from other sources (see Figure S1 in the online supplement). Ninety of these appeared to discuss response prediction, and 76 articles, covering 81 biomarkers, were eligible for descriptive analysis. Of these, 53 articles, discussing 57 biomarkers, included sufficient information for meta-analysis. The majority of articles that did not include sufficient 2×2 table information still reported a statistically significant result (22/24, 91.7%).

Studies varied in the degree of treatment resistance, included and excluded diagnoses, details of EEG recording, and analytic and statistical approach (see Table S1 in the online supplement). Seventy percent (57/81) were studies of response to medication, with most of the remaining (17%, 14/81) predicting response to rTMS. Citalopram/escitalopram and venlafaxine were the most commonly studied medications, representing 23% (13/57) and 19% (11/57) of medication studies, respectively. Most reported markers were from resting-state EEG (70%, 57/81) and did not source-localize the EEG data (79%, 64/81). The most heavily represented biomarkers were low-frequency EEG power (31%, 25/81) and cordance (19%, 15/81).

No study was a preplanned independent-sample replication of a previous investigation with identical medication regimens and outcome measures. A few markers, however, were studied repeatedly with similar designs. Three LDAEP studies attempted to predict response to citalopram (19, 21, 103). They had inconsistent results that appeared to be dependent on source-localization technique. Obrich et al. (67) used a vigilance marker that had previously been validated (using different recording/analysis methods) in smaller data sets. Cook et al. used cordance to predict the response to
FIGURE 1. Results of a Meta-Analysis of Quantitative EEG (QEEG) Biomarkers in Depression Treatment: Sensitivity

The figure shows results, presented as forest plots, for sensitivity for prediction of clinical antidepressant response based on QEEG biomarkers. Meta-analytic estimates show modest predictive power for clinical response. (For more information about the studies listed, see Tables S1 and S2 in the online supplement.) Markers are indicated as follows: TMS=transcranial magnetic stimulation; Vfx=venlafaxine; ATR=Antidepressant Treatment Response index; Theta=theta power; Rbx=reboxetine; ACC=anterior cingulate cortex; OFC=orbitofrontal cortex; Bup=bupropion; Esc=Escitalopram; Clo=clomipramine; Map=maprotiline.
FIGURE 2. Results of a Meta-Analysis of Quantitative EEG (QEEG) Biomarkers in Depression Treatment: Specificity

The figure shows results, presented as forest plots, for specificity for prediction of clinical antidepressant response based on QEEG biomarkers. Meta-analytic estimates show modest predictive power for clinical response. (For more information about the studies listed, see Tables S1 and S2 in the online supplement.) Markers are indicated as follows: TMS=transcranial magnetic stimulation; Vfx=venlafaxine; ATR=Antidepressant Treatment Response index; Theta=theta power; Rbx=reboxetine; ACC=anterior cingulate cortex; OFC=orbitofrontal cortex; Bup=bupropion; Esc=Escitalopram; Clo=clomipramine; Map=maprotiline.
FIGURE 3. Results of a Meta-Analysis of Quantitative EEG (QEEG) Biomarkers in Depression Treatment: Log(Diagnostic Odds Ratio)\(^a\)

<table>
<thead>
<tr>
<th>Log(Odds Ratio)</th>
<th>[95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.33 [1.18, 3.47]</td>
<td></td>
</tr>
<tr>
<td>3.91 [0.97, 6.86]</td>
<td></td>
</tr>
<tr>
<td>3.42 [0.42, 6.43]</td>
<td></td>
</tr>
<tr>
<td>3.75 [2.15, 5.35]</td>
<td></td>
</tr>
<tr>
<td>2.59 [0.74, 4.43]</td>
<td></td>
</tr>
<tr>
<td>2.43 [0.59, 4.28]</td>
<td></td>
</tr>
<tr>
<td>3.84 [0.73, 6.95]</td>
<td></td>
</tr>
<tr>
<td>3.42 [0.42, 6.43]</td>
<td></td>
</tr>
<tr>
<td>3.25 [−0.66, 5.46]</td>
<td></td>
</tr>
<tr>
<td>1.75 [0.22, 3.29]</td>
<td></td>
</tr>
<tr>
<td>3.50 [−1.04, 7.63]</td>
<td></td>
</tr>
<tr>
<td>0.07 [−1.32, 1.47]</td>
<td></td>
</tr>
<tr>
<td>0.00 [−0.84, 0.84]</td>
<td></td>
</tr>
<tr>
<td>2.19 [1.22, 3.15]</td>
<td></td>
</tr>
<tr>
<td>2.49 [0.32, 4.65]</td>
<td></td>
</tr>
<tr>
<td>1.29 [0.52, 2.06]</td>
<td></td>
</tr>
<tr>
<td>3.84 [0.63, 7.04]</td>
<td></td>
</tr>
<tr>
<td>2.47 [0.46, 4.48]</td>
<td></td>
</tr>
<tr>
<td>2.80 [0.54, 5.06]</td>
<td></td>
</tr>
<tr>
<td>3.04 [0.81, 5.28]</td>
<td></td>
</tr>
<tr>
<td>3.98 [0.94, 7.02]</td>
<td></td>
</tr>
<tr>
<td>2.43 [0.41, 4.45]</td>
<td></td>
</tr>
<tr>
<td>4.14 [−0.02, 8.30]</td>
<td></td>
</tr>
<tr>
<td>1.89 [−0.49, 4.26]</td>
<td></td>
</tr>
<tr>
<td>2.40 [0.52, 2.48]</td>
<td></td>
</tr>
<tr>
<td>1.57 [0.01, 3.13]</td>
<td></td>
</tr>
<tr>
<td>3.85 [0.81, 6.90]</td>
<td></td>
</tr>
<tr>
<td>1.64 [0.66, 2.62]</td>
<td></td>
</tr>
<tr>
<td>0.75 [−0.12, 1.63]</td>
<td></td>
</tr>
<tr>
<td>3.22 [1.00, 5.44]</td>
<td></td>
</tr>
<tr>
<td>3.74 [1.29, 6.18]</td>
<td></td>
</tr>
<tr>
<td>1.18 [−0.15, 2.52]</td>
<td></td>
</tr>
<tr>
<td>1.31 [−0.04, 2.66]</td>
<td></td>
</tr>
<tr>
<td>1.29 [−0.00, 2.58]</td>
<td></td>
</tr>
<tr>
<td>0.99 [0.04, 1.94]</td>
<td></td>
</tr>
<tr>
<td>2.54 [1.27, 3.81]</td>
<td></td>
</tr>
<tr>
<td>1.37 [0.62, 2.53]</td>
<td></td>
</tr>
<tr>
<td>3.51 [1.52, 5.49]</td>
<td></td>
</tr>
<tr>
<td>0.58 [0.15, 1.00]</td>
<td></td>
</tr>
<tr>
<td>0.01 [−0.62, 0.64]</td>
<td></td>
</tr>
<tr>
<td>0.79 [0.22, 1.35]</td>
<td></td>
</tr>
<tr>
<td>5.11 [0.97, 9.25]</td>
<td></td>
</tr>
<tr>
<td>5.05 [1.92, 8.18]</td>
<td></td>
</tr>
<tr>
<td>2.16 [1.05, 3.27]</td>
<td></td>
</tr>
<tr>
<td>0.50 [−1.10, 2.29]</td>
<td></td>
</tr>
<tr>
<td>1.48 [1.13, 1.82]</td>
<td></td>
</tr>
<tr>
<td>1.89 [1.53, 2.26]</td>
<td></td>
</tr>
<tr>
<td>2.20 [−0.42, 4.81]</td>
<td></td>
</tr>
<tr>
<td>3.66 [0.47, 7.26]</td>
<td></td>
</tr>
<tr>
<td>1.84 [0.46, 3.22]</td>
<td></td>
</tr>
<tr>
<td>4.01 [−0.19, 8.20]</td>
<td></td>
</tr>
<tr>
<td>0.00 [−1.54, 1.54]</td>
<td></td>
</tr>
<tr>
<td>0.42 [−0.68, 1.51]</td>
<td></td>
</tr>
<tr>
<td>1.37 [0.26, 2.49]</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The figure shows results, presented as forest plots, for the log(diagnostic odds ratio) for prediction of clinical antidepressant response based on QEEG biomarkers. Meta-analytic estimates show modest predictive power for clinical response. (For more information about the studies listed, see Tables S1 and S2 in the online supplement.) Markers are indicated as follows: TMS=transcranial magnetic stimulation; Vfx=venlafaxine; ATR=Antidepressant Treatment Response index; Theta=theta power; Rbx=reboxetine; ACC=anterior cingulate cortex; OFC=orbitofrontal cortex; Bup=bupropion; Esc=Escitalopram; Clo=clomipramine; Map=maprotiline.

Am J Psychiatry 176:1, January 2019 ajp.psychiatryonline.org 49
EEG BIOMARKERS FOR TREATMENT RESPONSE PREDICTION IN MAJOR DEPRESSIVE ILLNESS

FIGURE 4. Results of a Meta-Analysis of Quantitative EEG (QEEG) Biomarkers in Depression Treatment: Summary Receiver Operator Curve

The figure shows results for the summary receiver operator curve for sensitivity of prediction of clinical antidepressant response based on QEEG biomarkers. The modest predictive power for clinical response shown in Figures 1–3 is also visible in the summary receiver operator curve, where the area under the curve is estimated at 0.76. rTMS = repetitive transcranial magnetic stimulation.

varying medication protocols, but a series of their studies (26, 54, 55) found better-than-chance prediction using the same equipment, outcome measures, and decision rule (cordance decrease at 1 week of treatment). Bares et al. used different patient populations (bipolar depression and major depressive disorder), treatments, and response definitions but also repeatedly reported successful response prediction with a 1-week cordance decrease (47, 50, 51). A pair of studies with a relatively large sample size found that the ATR predicted response to different medications (59, 60). These studies were based on earlier reports of cordance and power biomarkers by the same researchers using different medication regimens (104). The larger ATR studies reported a slight modification of a previously unpublished version of ATR (version 4.1 in the study reports, compared with version 4.0 in the trial protocol and previous poster presentations). Widge et al. (61) reported that the same version of ATR did not predict response to rTMS. Reports from the iSPOT-D study were hypothesis driven and meant to test biomarkers that were hypothesis driven and meant to test biomarkers that had previously been reported, although they used a different medication protocol (25, 40, 69). Finally, theta power source-localized to the anterior cingulate cortex was reported by multiple laboratories as a predictor of response to different monoaminergic medications (14, 29, 31). A recent report from the EMBARC study (105) (which did not report information necessary for meta-analysis) also found cingulate theta to predict antidepressant response, although theta changes did not differ between patients receiving sertraline and those receiving placebo.

Study Quality
Study sizes were generally small, with a median N of 25. The distribution was trimodal (see Figure S2 in the online supplement), with peaks at approximately N=20, N=85, and N=660. The latter reflects reports from the recently concluded iSPOT-D study (40, 67, 69).

Most studies did not meet the quality metrics. Forty studies reported testing only a single EEG feature or finding no significant results and thus did not require correction for multiple comparisons. Of the 36 studies that tested multiple features, 67% (24/36) did not report use of a statistical correction. Of 71 markers reported to have significant predictive validity, only six (8%) were studied with cross-validation or another out-of-sample verification. Three of these were from the same first author (106–108). One article reported using cross-validation but did not include cross-validated algorithm performance in its main text or abstract (60).

Overall Efficacy
For all biomarkers taken together, the meta-analysis suggested predictive power above chance (Figures 1–4). The meta-analytic estimate of sensitivity was 0.72 (95% CI=0.67–0.76), specificity was 0.68 (95% CI=0.63–0.73), and log(diagnostic odds ratio) was 1.89 (95% CI=1.56–2.21). These correspond to an area under the curve of 0.76 (95% CI=0.71–0.80). The univariate analysis did not suggest study heterogeneity as a driver of results (I²=0%; Q=55.9, p=0.48). This implies that, in general, QEEG may have predictive power for treatment response in depressive illness. No biomarker or treatment type showed significantly greater predictive power than another. In bivariate meta-regressions (see Tables S3 and S4 in the online supplement), the Akaike information criterion increased from its omnibus value of −115.7 to −104.1 for a model split by biomarker type and to −107.7 for a model split by treatment type. Increases in Akaike information criterion imply that model terms have no true explanatory power (109). This is further supported by most meta-regression model coefficients failing to reach significance. We considered the possibility that these results reflect older studies identifying incorrect candidates, with newer studies homing in on true effects. A bivariate meta-regression of diagnostic accuracy against publication year showed no effect (p>0.27, Z-test on regression coefficients).

Funnel-plot analysis suggested that QEEG’s apparent predictive power is driven by small studies with strong positive results. The plot was specifically depleted in studies with smaller effect sizes that may not have reached prespecified significance thresholds (Figure 5A), and there was a tight correlation between effect size and the reciprocal of effective sample size (Figure 5B). The arcsine test for funnel plot asymmetry rejected the null hypothesis (t=6.33, p=4.64×10−10).
DISCUSSION

QEEG is commercially promoted to psychiatrists and our patients as a “brain map” for customizing patients’ depression treatment. Our findings indicate that QEEG, as studied and published to date, is not well supported as a predictive biomarker for treatment response in depression. Use of commercial or research-grade QEEG methods in routine clinical practice would not be a wise use of health care dollars. This conclusion is likely not surprising to experts in QEEG, who are familiar with the limitations of this literature. It is important, however, for practicing psychiatrists to understand the limitations, given the availability of QEEG as a diagnostic test. At present, marketed approaches do not represent evidence-based care. This mirrors other biomarker fields, such as pharmacogenomics and neuroimaging, for which recent reviews (4, 110) suggest that industry claims substantially exceed the evidence base. Like those markers, QEEG may become clinically useful, but only with further and more rigorous study.

We showed that the QEEG literature generally describes tests with reasonable predictive power for antidepressant response (sensitivity, 0.72; specificity, 0.68). This apparent utility, however, may be an artifact of study design and selective publication. We observed a strong funnel plot asymmetry, indicating that many negative or weak studies are not in the published literature. Of those that were published, many have small sample sizes. Small samples inflate effect sizes, which may give a false impression of efficacy (111). This is doubly true given the wide range of options available to EEG data analysts, which can lead to inadvertent multiple hypothesis testing (93). We also identified a common methodological deficit in the lack of cross-validation, which could overestimate predictive capabilities. Taken together, the findings suggest that community standards in this area of psychiatric research do not yet enforce robust and rigorous practices, despite recent calls for improvement (11, 77, 93). Our results indicate that QEEG is not ready for widespread use. Cordance and cingulate theta power are closest to proof of concept, with studies reporting successful treatment prediction across different medication classes and study designs (14, 29, 31, 47–49, 51, 105). ATR has been successful across medication classes, but only when tested by its original developers (58, 59). A direct and identical replication of at least some of those findings is still necessary. These design and reporting limitations suggest that QEEG has not yet been studied or validated to a level that would make it reliable for regular clinical use.

We designed this meta-analysis for maximum sensitivity, because we sought to demonstrate QEEG’s lack of maturity as a biomarker. This makes our omnibus meta-analytic results overly optimistic and obscures three further limitations of QEEG as a response predictor. First, we accepted each individual study’s definition of the relevant marker without enforcing consistent definitions within or between studies. For example, alpha EEG has been defined differently for different sets of sensors within the same patient (34) and at different measurement time points (60). Enforcing consistent definitions would attenuate the predictive signal, because it reduces “researcher degrees of freedom” (77). On the other hand, an important limitation of our meta-analysis is that it could not identify a narrow biomarker. If QEEG can predict response to a single specific treatment or response in a biologically well-defined subpopulation, that finding would be obscured by our omnibus treatment. Marker-specific meta-analysis (as in reference 76) would be necessary to answer that question.
Second, we did not consider studies as negative if they found significant change in the “wrong” direction. For instance, theta cordance decline during the first week of treatment is believed to predict medication response (26, 47, 48, 51, 52, 55). Two studies reported instead that a cordance increase predicted treatment response (46, 53). LDAEP studies have reported responders to have both higher (17, 19, 20) and lower (15) loudness dependence compared with nonresponders. This could be explained by differences in collection technique, or in the biological basis of the interventions (e.g., the inconsistent study used noradrenergic medication, whereas LDAEP is thought to assess serotonergic tone). It could also be explained by true effect sizes of zero, and modeling these discrepancies differently would reduce our estimates of QEEG’s efficacy.

Third, and arguably most important, depression itself is heterogeneous (6, 112). Defining and subtyping it is one of the major challenges of modern psychiatry, and there have been many proposals for possible endophenotypes (6, 9, 12, 113, 114). When we consider that each primary study effectively lumped together many different neurobiological entities, the rationale for QEEG-based prediction is less clear. As an example, a recent attempt to validate an obsessive-compulsive disorder biomarker, using the originating group’s own software, showed a significant signal in the opposite direction from the original study (115). Furthermore, studies often predict antidepressant response for patients receiving medications with diverse mechanisms of action. Considering that patients who do not respond to one medication class (e.g., serotonergic) often respond to another (e.g., noradrenergic or multireceptor), it does not make sense for any single EEG measure to predict response to multiple drug types. Similarly, although the goal of many recent studies is to explicitly select medication on the basis of a single EEG recording (40, 70, 72, 78), this may not be possible given the many ways in which neurotransmitter biology could affect the EEG. Reliable electrophysiologic biomarkers may require “purification” of patient samples to those with identifiable circuit or objective behavioral deficits (6, 116) or use of medications with simple receptor profiles. It may also be helpful to shift from resting-state markers to activity recorded during standardized tasks (6) as a way of increasing the signal from a target cortical region. Task-related EEG activity has good test-retest reliability, potentially improving its utility as a biomarker (71).

We stress that our meta-analysis means that QEEG as currently known is not ready for routine clinical use. It does not mean QEEG research should be stopped or slowed. Many popular QEEG markers have meaningful biological rationales. LDAEP is strongly linked to serotonergic function in animals and humans (117). Cordance was originally derived from hemodynamic measures (11, 54). Neither cordance nor ATR changed substantially in placebo responders, even though both changed in medication responders (26, 59, 117). The theta and alpha oscillations emphasized in modern QEEG markers are strongly linked to cognition and executive function (62, 118). Our results do not imply that QEEG findings are not real; they call into question the robustness and reliability of links between symptom checklists and specific aspects of resting-state brain activity. If future studies can be conducted with an emphasis on rigorous methods and reporting, and with specific attempts to replicate prior results, QEEG still has much potential.

**AUTHOR AND ARTICLE INFORMATION**

From the Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis; the Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston; the Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Mass.; the Department of Psychiatry, Stanford University School of Medicine, Palo Alto, Calif.; Butler Hospital and Warren Alpert Medical School of Brown University, Providence, R.I.; the Department of Psychiatry, University of Wisconsin School of Medicine and Public Health, Madison; the Department of Psychiatry, University of Miami Miller School of Medicine, Miami.

Address correspondence to Dr. Widge (awidge@umn.edu).

Preparation of this work was supported in part by grants from the Brain and Behavior Research Foundation, the Harvard Brain Science Initiative, and NIH (MH109722, NS100548) to Dr. Widge. The authors further thank Farifeth F. Duffy, Ph.D., and Diana Clarke, Ph.D., of the American Psychiatric Association, for critical administrative and technical assistance throughout preparation.

This article is derived from work done on behalf of the American Psychiatric Association (APA) and remains the property of APA. It has been altered only in response to the requirements of peer review. Copyright © 2017 American Psychiatric Association. Published with permission.

Drs. Widge and Deckersbach have pending patent applications related to the use of electrographic markers to characterize patients and select neuromodulation therapies. Dr. Widge has received device donations and consulting income from Medtronic. Dr. Rodriguez has served as a consultant for Allergan, BlackThorn Therapeutics, and Rugen Therapeutics. Dr. Carpenter has served as a consultant for Magstigm and has received research clinical trial support from Cervel, Janssen, NeoSync, and Neurotensics. Dr. Kalin has received research support from NIMH; he has served as a consultant for CME Outfitters, the Pritzker Neuropsychiatric Disorders Research Consortium, the Skyland Trail Advisory Board, and TC MSO (parent company of Actify Neurotherapies); and he receives remuneration from Elsevier as co-editor of the journal Psychoneuro-endoocrinology. Dr. Nemeroff has received grants or research support from NIH and the Stanley Medical Research Institute; he has served as a consultant for Bracket (Clintara), Dainippon Pharma, Fortress Biotech, Intra-Cellular Therapies, Janssen Research and Development, Magstigm, Prismic Pharmaceuticals, Sumitomo Navitor Pharmaceuticals, Sunovion, Taisho Pharmaceutical, Takeda, TC MSO, and Xhale; he has served on advisory boards for the American Foundation for Suicide Prevention (AFSP), the Anxiety Disorders Association of America (ADAA), Bracket (Clintara), the Brain and Behavior Research Foundation, the Laureate Institute for Brain Research, Skyland Trail, and Xhale and on directorial boards for ADAA, AFSP, and Gratitude America; he is a stockholder in AbbVie, Antares, BI Gen Holdings, Celgene, Corcept Therapeutics, OPKO Health, Seattle Genetics, and Xhale; he receives income or has equity of $10,000 or more from American Psychiatric Publishing, Bracket (Clintara), CME Outfitters, Intra-Cellular Therapies, Magstigm, Takeda, and Xhale; and he holds patents on a method and devices for transdermal delivery of lithium (patent 6,375,990B1) and a method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (patent 7,148,027B2). The other authors report no financial relationships with commercial interests.
REFERENCES

1. Roehrig C: Mental disorders top the list of the most costly conditions in the United States: $201 billion. Health Aff (Millwood) 2016; 35:1130–1135


5. Cassel CK, Guest JA: Choosing wisely: helping physicians and patients make smart decisions about their care. JAMA 2012; 307: 1801–1802


80. Ulrich G, Haug H-J, Fähndrich E: Acute vs chronic EEG effects in


93. Blackford JU: Leveraging statistical methods to improve validity and reproducibility of research findings. JAMA Psychiatry 2017; 74:119–120


