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

Greg J. Siegle, PhD; Wesley K. Thompson, PhD; Amanda Collier, BS; Susan R. Berman, MEd; Joshua Feldmiller, BA; Michael E. Thase, MD; Edward S. Friedman, MD

Context: Among depressed individuals not receiving medication in controlled trials, 40% to 60% respond to cognitive therapy (CT). Multiple previous studies suggest that activity in the subgenual anterior cingulate cortex (sgACC; Brodmann area 25) predicts outcome in CT for depression, but these results have not been prospectively replicated.

Objective: To examine whether sgACC activity is a reliable and robust prognostic outcome marker of CT for depression and whether sgACC activity changes in treatment.

Design: Two inception cohorts underwent assessment with functional magnetic resonance imaging using different scanners on a task sensitive to sustained emotional information processing before and after 16 to 20 sessions of CT, along with a sample of control participants who underwent testing at comparable intervals.

Setting: A hospital outpatient clinic.

Patients: Forty-nine unmedicated depressed adults and 35 healthy controls.

Main Outcome Measures: Pretreatment sgACC activity in an a priori region in response to negative words was correlated with residual severity and used to classify response and remission.

Results: As expected, in both samples, participants with the lowest pretreatment sustained sgACC reactivity in response to negative words displayed the most improvement after CT ($R^2=0.29$, $>75$% correct classification of response, $>70$% correct classification of remission). Other a priori regions explained additional variance. Response/remission in cohort 2 was predicted based on thresholds from cohort 1. Subgenual anterior cingulate activity remained low for patients in remission after treatment.

Conclusions: Neuroimaging provides a quick, valid, and clinically applicable way of assessing neural systems associated with treatment response/remission. Subgenual anterior cingulate activity in particular may reflect processes that interfere with treatment (eg, emotion generation) in addition to its putative regulatory role; alternatively, its absence may facilitate treatment response.

Arch Gen Psychiatry. 2012;69(9):913-924
activity to self-report measures are notoriously unreliable roimaging studies because relationships of voxelwise ac-

given a single study with 14 participants, the need for replications before translation to the clinic, particularly in common arguments regarding the importance of 

declines of sgACC activity (goal 5) depend on the somewhat ambiguous function of this region. On the theory that the sgACC inhibits lim-

Activity biomarker, we examined whether sgACC prediction of outcome could be replicated in an efficacy sample (ie, best possible conditions for observing the effect) and an effectiveness sample (ie, more real-world, less ideal conditions). Thus, we examined whether sgACC prediction of depression outcome in CT holds in an efficacy cohort using effectively the same recruitment, design (including task), therapists, sample size, trial selection (eg, measuring reactivity to negative words), data preparation, and analytic techniques, as in Siegle et al.4 In addition to common arguments regarding the importance of replication before translation to the clinic, particularly given a single study with 14 participants, the need for strict replication is particularly great for voxelwise neuroimaging studies because relationships of voxelwise activity to self-report measures are notoriously unreliable (goal 1).13,16 We also examined whether sgACC prediction of outcome holds in an effectiveness cohort with a more heterogeneous, clinically representative sample and community therapists with variable supervision (goal 2), is revealed as one of the measured neural indices that best predict treatment outcome (goal 3), can be formulated in a way that can be easily interpreted by clinicians and used across scanners (goal 4), and can inform, if not confirm, causal inferences regarding the possible role of sgACC change in clinical change by considering changes in treatment with patients with the predictive marker (goal 5).

For prediction (goals 1 and 2), our primary hypothesis was that sgACC activity would be prognostic for outcome in CT in multiple samples robustly enough to garner clinical consideration. Our secondary hypotheses were that other relevant theoretically motivated regions and their functional relationships would add variance to prediction of outcome but would not completely moderate the role of the sgACC (goal 3), that the same predictive associations would be apparent in normalized data (goal 4), and that sgACC activity would normalize after treatment (goal 5).

For the other relevant regions (goal 3), we examined a brain network associated with emotional reactivity and implicated in depression and treatment outcome,27 including the left amygdala, as in previous studies.4,18 and in regions associated with regulatory control that have decreased functioning in depression, including the dor-
solateral prefrontal cortex (DLPFC),8,19,20 the rostral anterior cingulate,21,22 and their functional corticolimbic/corticocortical relationships.24,25 We also performed a voxelwise analysis. Directional hypotheses for change in sgACC activity (goal 5) depend on the somewhat ambiguous function of this region. On the theory that the sgACC inhibits limbic regions,26,28,29 we have suggested that, because CT teaches skills for emotion regulation, individuals who most need CT (ie, those with the lowest pretreatment levels of sgACC activity) may respond best to CT.3 Thus, patients with remitting depression after CT would be expected to show decreased pretreatment but increased posttreatment sgACC activity. Alternately, the sgACC may also support emotion generation/monitoring. In support of this theory, (1) neurofeedback-induced decreases in sgACC activity yield increased sadness,29 (2) trait-related increased sgACC activity is associated with higher sadness30 and increased depressive severity,31 and (3) sgACC inhibition via deep brain stimulation facilitates recovery.32 If this role is primary, low sgACC activity may be needed for voluntary regulation to occur. In this case, remitters would be expected to show decreased pretreatment and posttreatment sgACC activity.

To evaluate whether sgACC activity normalizes in treatment, a sample of healthy never-depressed control participants were recruited to establish a normative baseline. The controls also afforded us the ability to examine pretreatment activity with respect to normative function (ie, as clinically interpretable z scores [goal 4]) and to assess the test-retest reliability outside the context of depression (ie, showing that healthy individuals do not change strongly over time) and whether the measure is reliable enough to make inferences regarding pretreatment-posttreatment measurements (as we have recommended).33 Data from controls were not used in primary prediction analyses.

METHODS

PARTICIPANTS AND THERAPISTS

As shown in Figure 1 (a CONSORT diagram) and Table 1 (listing demographic data), participants from 2 clinical trials underwent testing in 2 cohorts on different scanners, separated by approximately 6 months (full elaboration can be found in eMaterial 1 on the authors’ website [http://www.pitt.edu /~gsiegel/Siegle-PMR-Prediction-Archives12-AuthorMaterial .pdf]). After attrition and data cleaning (fully described in Figure 1), cohort 1 (the efficacy cohort, from M.E.T., princi-
pal investigator; ClinicalTrials.gov identifier: NCT00183664) included 17 patients with recurrent major depressive disorder (diagnoses via the Structured Clinical Interview for DSM-IV Axis I Disorders–Patient Edition [SCID]34) treated by the same 3 therapists (including E.S.F.) from the study by Siegle et al.a on which this study was based; these therapists continued to receive weekly supervision with taped review from the same master clinician (Sandar Kornblith, PhD). Cohort 1 also included 15 healthy controls (no current or historical Axis I disorder via SCID interview). Therapeutic adherence was deemed adequate with a random selection of 10 tapes (ratings of ≥2 tapes per therapist by other therapists or outside raters) receiving high marks on the Cognitive Therapy Scale35 (scores of 40 represent high levels of adherence; mean [SD], 52.7 [8.6]; range, 40-63).

Cohort 2 (the effectiveness cohort) included 32 patients who were more clinically heterogeneous than cohort 1 because they included 23 patients with recurrent major depressive disorder and 9 in their first episode. Participants were drawn from the same trial (G.J.S, principal investigator; ClinicalTrials.gov identifier: NCT00787501 [CT election, given options of treatment with CT or a selective serotonin reuptake inhibitor]). Cohort 2 also included 20 healthy controls (no current or historical Axis I disorder via SCID interview). Patients were treated by 6 community clinicians (with doctor of philosophy, doctor of medicine, master of education, or licensed clinical social worker degrees [including E.S.F.]) who ranged in CT experience from a doctorate founding member of the Academy of Cognitive Therapy to a social worker who took her second CT case as
part of this study. Therapists received group supervision monthly without taped review and as-requested supervision by Dr Kornblith (weekly for 2 therapists). Therapeutic adherence was more variable, with 12 tapes of 4 therapists receiving Cognitive Therapy Scale mean (SD) scores of 43.3 (13.1) and ranging from 23 to 61, with 6 of 12 tapes falling below the cutoff of 40 for adequate adherence, as administered by other study therapists or outside raters.

Participants described no health problems, eye problems, or psychoactive drug abuse in the past 6 months and no history of psychosis or manic or hypomanic episodes. None of the controls or depressed participants had used antidepressants within 2 weeks of testing (6 weeks for fluoxetine hydrochloride) owing to medication naivety or supervised withdrawal from unsuccessful medication therapies. Participants reported no excessive use of alcohol in the 2 weeks before testing and scored in the normal range on a cognitive screen (Verbal IQ equivalent: >85).

**PROCEDURE**

After complete study description, University of Pittsburgh institutional review board–approved written informed consent was obtained followed by a SCID interview, vision test, and unrelated physiological assessment. Patients underwent assessment on a different day with a battery of functional magnetic resonance imaging (fMRI) tasks administered in counterbalanced order. One task is reported herein; the others are described in eMaterial III-A. Participants rated their sad, anxious, and happy affects from 1 (not at all) to 5 (very) before and after the task. The Beck Depression Inventory, second edition (BDI), was administered after fMRI to assess depressive severity (rationale is found in eMaterial II-A). Depressed participants then received CT consisting of 2 sessions/wk for the first 4 weeks followed by 8 weekly sessions for early treatment responders (Hamilton Rating Scale for Depression [HRSD] reduction, <40% at session 9, of 16 total sessions), or 2 sessions/wk for the first 8 weeks followed by 4 weekly sessions for those without early treatment response (20 total sessions). Cognitive therapy followed the guidelines by Beck et al (detailed in eMaterial I), with weekly HRSD ratings. Within 2 weeks of completion (week 16 for controls), all participants completed the same fMRI protocol and BDI again.

**CLINICAL DATA PREPARATION**

In cohort 1, 1 participant's pretreatment and 1 participant's post-treatment BDI scores were missing. Fortunately the BDI-I was administered clinically in the protocol, and thus these values were reconstructed via regression (described in eMaterial II-B). In all other cases, final BDI scores were assessed at the participants' second IMRI session (after completing CT or approximately 12 weeks from their first IMRI session for those who did not complete CT). Final HRSD scores were imputed on the basis of the trajectory of weekly responses as described in the eMaterial II-D. Response was defined as a 50% reduction in the initial BDI or HRSD score (as opposed to the early response criterion of 40% reduction used to determine the number of sessions), and remission was defined as a final BDI score less than 10 (rationale in eMaterial II-C) or a final HRSD score less than

<table>
<thead>
<tr>
<th>CONSORT Diagram</th>
<th>CT indicates cognitive therapy; fMRI, functional magnetic resonance imaging; HRSD, Hamilton Rating Scale for Depression; and SSRI, selective serotonin reuptake inhibitor.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enrollment</strong></td>
<td>278 Considered for CT via initial evaluation during multistudy triage enrollment</td>
</tr>
<tr>
<td>219 Excluded after initial evaluation</td>
<td></td>
</tr>
<tr>
<td>88 Elected to participate in another study</td>
<td></td>
</tr>
<tr>
<td>77 No show</td>
<td></td>
</tr>
<tr>
<td>32 Other</td>
<td></td>
</tr>
<tr>
<td>13 Axis I rule-outs</td>
<td></td>
</tr>
<tr>
<td>9 Currently in treatment and wish to remain</td>
<td></td>
</tr>
<tr>
<td><strong>Allocation</strong></td>
<td>59 Consented</td>
</tr>
<tr>
<td>47 Depressed cohort 1 CT</td>
<td></td>
</tr>
<tr>
<td>28 Did not participate</td>
<td></td>
</tr>
<tr>
<td>27 Elected not to participate because of disinterest or inconvenience</td>
<td></td>
</tr>
<tr>
<td>1 Claustrophobic</td>
<td></td>
</tr>
<tr>
<td>2 Scan data not clean</td>
<td></td>
</tr>
<tr>
<td>17 Analyzed CT completers and noncompleters, received at least 7 sessions of CT and had clean data</td>
<td></td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
<td>381 Considered for CT/SSRI preference trial via initial evaluation</td>
</tr>
<tr>
<td>319 Excluded at initial evaluation</td>
<td></td>
</tr>
<tr>
<td>143 Axis I rule-outs</td>
<td></td>
</tr>
<tr>
<td>63 Other</td>
<td></td>
</tr>
<tr>
<td>43 Age</td>
<td></td>
</tr>
<tr>
<td>29 Not interested</td>
<td></td>
</tr>
<tr>
<td>18 HRSD &lt;14</td>
<td></td>
</tr>
<tr>
<td>14 Medical disorder</td>
<td></td>
</tr>
<tr>
<td>9 Drug abuse</td>
<td></td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td>52 Depressed cohort 2 CT</td>
</tr>
<tr>
<td>1 Scan data not clean</td>
<td></td>
</tr>
<tr>
<td>15 Failed to enter CT treatment phase after scan</td>
<td></td>
</tr>
<tr>
<td>3 Disinterest or inconvenience</td>
<td></td>
</tr>
<tr>
<td>2 Substance abuse</td>
<td></td>
</tr>
<tr>
<td>1 Too large for IMRI</td>
<td></td>
</tr>
<tr>
<td>1 Task issues</td>
<td></td>
</tr>
<tr>
<td>1 Prior head injury</td>
<td></td>
</tr>
<tr>
<td>1 Bipolar</td>
<td></td>
</tr>
<tr>
<td>1 Psychotic</td>
<td></td>
</tr>
<tr>
<td>1 Age</td>
<td></td>
</tr>
<tr>
<td>2 Claustrophobic</td>
<td></td>
</tr>
<tr>
<td>1 Unreachable</td>
<td></td>
</tr>
<tr>
<td>1 Metal in body</td>
<td></td>
</tr>
<tr>
<td>4 Failed to complete at least 7 sessions of CT, refused further treatment</td>
<td></td>
</tr>
<tr>
<td>32 Analyzed CT</td>
<td></td>
</tr>
<tr>
<td>Received at least 7 sessions of CT and had clean data</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. CONSORT diagram. CT indicates cognitive therapy; fMRI, functional magnetic resonance imaging; HRSD, Hamilton Rating Scale for Depression; and SSRI, selective serotonin reuptake inhibitor.
Table 1. Demographic and Clinical Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Depressed Participants/CT</th>
<th>Control Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cohort 1</td>
<td>Cohort 2</td>
</tr>
<tr>
<td></td>
<td>Cohort 1</td>
<td>Cohort 2</td>
</tr>
<tr>
<td>No. of participants</td>
<td>17</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Male sex, No.</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>White race, No.</td>
<td>14</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Age, y</td>
<td>24-51</td>
<td>19-55</td>
</tr>
<tr>
<td></td>
<td>25-54</td>
<td>21-52</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>37.41 (9.17)</td>
<td>36.13 (11.11)</td>
</tr>
<tr>
<td></td>
<td>38.93 (9.74)</td>
<td>32.95 (8.37)</td>
</tr>
<tr>
<td>Educational level, mean (SD), y</td>
<td>15.35 (2.23)</td>
<td>15.06 (2.60)</td>
</tr>
<tr>
<td></td>
<td>15.67 (2.26)</td>
<td>16.30 (2.13)</td>
</tr>
<tr>
<td>No. of depressive episodes, median (SD)</td>
<td>5.59 (3.52)</td>
<td>4.03 (4.83)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAART VIQ equivalent, mean (SD)</td>
<td>110.31 (8.68)</td>
<td>108.42 (9.92)</td>
</tr>
<tr>
<td></td>
<td>111.61 (7.97)</td>
<td>109.21 (8.96)</td>
</tr>
<tr>
<td>Emotion ratings for negative words, mean (SD)</td>
<td>2.55 (0.66)</td>
<td>2.11 (0.62)</td>
</tr>
<tr>
<td></td>
<td>2.99 (0.65)</td>
<td>2.49 (0.72)</td>
</tr>
<tr>
<td>RT to negative words, mean (SD)</td>
<td>1403.82 (269.41)</td>
<td>1637.34 (526.16)</td>
</tr>
<tr>
<td></td>
<td>1411.31 (380.26)</td>
<td>1242.20 (374.90)</td>
</tr>
<tr>
<td>Personal relevance ratings for negative words, mean (SD)</td>
<td>3.07 (0.66)</td>
<td>3.03 (0.57)</td>
</tr>
<tr>
<td></td>
<td>2.33 (0.72)</td>
<td>2.07 (0.49)</td>
</tr>
</tbody>
</table>

Clinical change in cohort 1 completers

Pretreatment BDI score, mean (SD)         33.0 (8.8)
Posttreatment BDI score, mean (SD)       10.7 (10.2)
BDI score change, mean (SD)              22.3 (13.4)
Responders, No. (%)                     9 (69)
Remitters, No. (%)                      8 (62)

Clinical change in whole sample: No. completing BDI

Pretreatment BDI score, mean (SD)         34.4 (8.7)
Posttreatment BDI score, mean (SD)       14.1 (12.2)
Change in BDI score, mean (SD)           20.3 (13.4)
Responders with BDI score, No. (%)       10 (63)
Remitters with BDI score, No. (%)        8 (50)

Clinical change in HRSD for whole sample, No.

Pretreatment HRSD score, mean (SD)        24.6 (6.7)
Posttreatment HRSD score, mean (SD)      9.6 (5.5)
Change in HRSD score, mean (SD)          15.0 (7.7)
Responders with HRSD, No. (%)            10 (59)
Remitters with HRSD, No. (%)             7 (47)

Abbreviations: BDI, Beck Depression Inventory; CT, cognitive therapy; ellipses, not applicable; HRSD, Hamilton Rating Scale for Depression; NAART, North American Adult Reading Test; RT, rating task; VIQ, Verbal IQ.

The f tests confirmed that no groups or cohorts were different from any other on any demographic variable (sex, ethnicity, age range, age, or educational level). Cohorts did not differ in depressive severity or median number of depressive episodes. Group and cohort differences in emotion ratings and reaction times are described in the “Demographic and Behavioral Data” subsection of the “Results” section and eMaterial VI. The first lines of the clinical response section, using only CT completers, replicates the analysis in Siegle et al. Remaining lines generalize that result.

7 (as in the National Institute of Mental Health–sponsored Se- quenced Treatment Alternatives to Relieve Depression).

IMRI TASK AND PROCESSING

Apparatus

Twenty-nine 3.2-mm slices were acquired parallel to the anterior commissure–posterior commissure line using a poste- rior to anterior echoplanar imaging pulse sequence to mini- mize susceptibility artifacts in the amygdala and orbitofrontal regions (3T Siemens Trio [Siemens Medical Solutions], T2*- weighted images depicting blood oxygenation level–dependent (BOLD) contrast; repetition time, 1500 milliseconds; echo time, 27 milliseconds; field of view, 24 cm; flip angle, 80°), yielding 8 whole-brain images per 1-second trial. Stimuli were displayed in black on a white background via a back- projection screen (0.88° visual angle). Responses were re- corded using a data glove (Psychology Software Tools).

Personal Relevance Rating Task

As in our previous publications, in 60 slow-event related trials, participants viewed a fixation cue (1 second; row of Xs with prongs around the center X) followed by a positive, a negative, or a neutral word (200 milliseconds; only negative words analyzed herein), followed by a mask (row of Xs; 10.8 seconds). Participants pushed a button to indicate whether the word was relevant, somewhat relevant, or not relevant to them or their lives (button orders balanced across participants), as quickly and accurately as they could. Participant-generated and normed words from Bradley and Lang were used as in our
previous studies of depression4,20,30-41 (procedures in eMaterial III; reaction time preparation in eMaterial VI-B).

**fMRI Data Preparation**

We followed standard preprocessing similar to that used in our previous study8 with slight modernizations (slice time correction, motion correction, linear detrending, voxelwise outlier rescaling, conversion to percent change, temporal smoothing [7-point gaussian filter], 32-parameter nonlinear warping to the Montreal Neurological Institute Colin-27 brain data set, spatial smoothing [6-mm full width half maximum], and response time series variability normalization across scanners; these methods are fully described and compared with those of our previous study14 in eMaterial IV). The same “reactivity” index used in our previous study yielded peak and sustained responses to negative words as the mean of the fourth through the seventh images of each negative-word trial minus the trial’s first (prestimulus) scan acquired while the fixation cue was on the screen.

Primary hypotheses were examined for mean reactivity in a 24-voxel 20-mm-radius sphere centered at Talairach coordinates 6, 17, -6, the centroid of the sgACC region prognostic for outcome in our previous study.8 Our broader network included right and left DLPFC regions, the left amygdala, and Brodmann area 24 (BA24) in the rostral cingulate (criteria are described in eMaterial V); to explicitly replicate methods from the previous study,14 we also examined regions from voxelwise analyses, subject to empirical type I error control (eMaterial IV-B).

**DATA ANALYSIS**

**Associations Between Pretreatment Brain Function and Clinical Change**

Clinical outcome was operationalized using residual severity, calculated as final severity controlling for initial severity. Thus, to replicate the methods of the previous study,8 we examined associations of pretreatment activity with residual severity, computed by regressing final BDI scores on initial scores and retaining the residuals in participants who completed treatment in cohort 1. To increase generalizability, we then included participants who did not complete treatment in this cohort as well as cohort 2. To predict response and remission, a within-sample grid search found the cutoff that maximized percentage correct discrimination. Standard indices of signal detection are reported (sensitivity, specificity, d’), and receiver operating characteristic (ROC) curve statistics are given in eMaterial VIII). One thousand permutation tests using the same algorithm (randomly permuting associations of response or remission with fMRI indices) assessed the significance of discrimination (percentage correct and d’; tests of d’ generally agreed with the percentage correct and thus are only reported when results diverged on significance). To increase generalizability outside clinical trials in which residual severity can be calculated, we examined z scores of sgACC activity (previously described reactivity index) normalized with respect to the same index computed for the controls’ initial assessment (ie, sgACC = [X−mean_controls]/SD_(controls)) and severity change scores (final−initial). Multiple regions were associated with residual severity using multiple regression.

Robustness was examined via random forest regression and classification,44 using as features activity from all a priori structures (criteria in eMaterial V) and their partial mutual information via the measures integrated by Zhou et al45 to capture systemic effects involving functional connectivity. Regression and classification forests were trained on cohort 1 using random resampling (bagged); a subset of the activation and mutual information measures was selected by maximizing out-of-bag estimates of prediction and classification accuracy on cohort 1. We tested generalizability by computing prediction and classification accuracy of the cohort 1–trained algorithms to cohort 2. We obtained P values by permuting responses and computing the proportion of times the permuted regression, and classification results were as good as or better than the unpermuted results.

**Change**

We examined associations of pretreatment and posttreatment activity via correlation and change in the time series of responses for participants with each combination of predicted and actual response. That is, to infer whether CT addressed the mechanisms of low sgACC, we examined whether sgACC activity normalized in participants with low and high pretreatment activity who strongly responded to CT. We similarly considered change in those who did not respond strongly to treatment.

**DEMOGRAPHIC AND BEHAVIORAL DATA**

Patients and controls did not differ significantly on demographic variables between cohorts or groups (Table 1). Group × cohort analyses of variance revealed that patients rated negative words as more negative and more personally relevant and reacted more slowly than did controls (eMaterial VI). Cohort 2 rated negative words as more negative (t98 = 3.27 [P = .002]; Cohen d = 0.73) with no other cohort main effects. There were no group × cohort interactions on valence or meaning ratings. Because group × cohort effects on reaction times (eMaterial VI) were small and the hemodynamic window began after their likely effects, they were not used in fMRI analyses. There were no significant associations of reaction time, valence ratings, or personal relevance ratings with residual BDI scores or sgACC activity (r < 0.15 [P > .4]). Pretreatment depressive severity (BDI score) was uncorrelated with pretreatment sgACC activity (r = -0.13). Depressed participants were moderately sad and anxious and minimally happy before and slightly less so after the task, whereas controls were comparably minimally sad and anxious and moderately happy before and after the task (eMaterial VI-C and eTable 1).

**CLINICAL OUTCOME**

Cognitive therapy was successful at rates at least as high as those observed in the literature with enough outcome variability (>5 patients in all response/remission cells) to allow further analysis (Table 1, individual trajectories, and eMaterial VII).

**Goal 1: Efficacy Replication**

For the efficacy replication, sgACC activity was tested as a prognostic indicator of clinical change in CT. As shown in Figure 2 and Table 2 and in our previous study,4 decreased sgACC activity was strongly corre-
tantly, prediction of responder and remitter status was owing to clinical and therapist heterogeneity. Impor-
what weaker (Figures 2 and 3 and Table 2), potentially
geneous sample with minimally supervised community
We examined effectiveness using a more clinically hetero-
tive specificity to the sgACC.
BDIresidual revealed sgACC to be associated with even more
area under the curve (AUC), reflecting good discrimi-
was collected on a slightly larger s ubset of participants.
cluded. This prediction generalized to the HRSD, which
the outcome variable. E, Waveforms for hemodynamic responses for
participants with nonremitting depression (nonremitters [n=21]; final BDI
score <10) compared with controls (n=35) and participants with remitting
depression (remitters [n=22]; final BDI score <10). Areas significant at
P <.05 by means of analyses of variance at each scan are highlighted in
gray.

Figure 2. Correlation of decreased anatomically defined subgenual anterior
cingulate cortex (sgACC) activity with a stronger clinical response (decreased
residual Beck Depression Inventory [BDI]). A, A strong correlation is seen in
cohort 1 (green squares indicate noncompleters whose final BDI scores were
interpolated). B, A moderate association is seen in cohort 2. C, A stronger
association is seen in the combined cohort. D, A stronger association is also
seen in the combined cohort using z scores for sgACC and change in BDI as
the outcome variable. E, Waveforms for hemodynamic responses for
participants with nonremitting depression (nonremitters [n=21]; final BDI
score <10) compared with controls (n=35) and participants with remitting
depression (remitters [n=22]; final BDI score <10). Areas significant at
P <.05 by means of analyses of variance at each scan are highlighted in gray.

Goal 2: Effectiveness Generalization

We examined effectiveness using a more clinically hetero-
geneous sample with minimally supervised community
therapists. The same general pattern of results held
significantly in cohort 2, although effect sizes were some-
what weaker (Figures 2 and 3 and Table 2), potentially
owing to clinical and therapist heterogeneity. Import-
tantly, prediction of responder and remitter status was
still strong, with more than 75% correct predictions for
response across all samples and measures and more than
65% correct predictions for remission, preserving strong
specificity. When data from cohorts 1 and 2 were com-
bined, yielding a mix of highly supervised and not as
strongly supervised therapists and a wide range of the
clinical phenotype, prediction remained strong. The ROC
analyses (eMaterials VIII and IX) were significant, re-
reflecting fair discrimination. Remitters were character-
ized by downward hemodynamic responses, whereas non-
remitters were characterized by upward hemodynamic
responses (Figure 2E), with significant differences in a
3-group (control, remitter, and nonremitter) analysis of
variance from 6 to 12 seconds (F2,70= 5.56 [P =.06]).

Goal 3: Additional Variance Explained by Other
Brain Regions and Measures

Other Regions. A voxelwise mega-analysis combining data
across cohorts (hierarchical regression in which the scanner
[cohort] and initial BDI score were entered in step 1 to
simultaneously covary them, with the final BDI score
entered in step 2) confirmed a 2-voxel region across both
samples (Talairach coordinates 5, 16, −9) that was sta-
tistically significant at P <.001 (P <.05 corrected) and
clinically significant, explaining more than 40% of the
variation in BDIresidual and located in the a priori mask
centered on the region detected in our previous study
(Figure 3). No other significant regions of more than 24
voxels explaining at least 40% of the variance were de-
tected. Similarly, voxelwise meta-analytic conjunction
analyses revealed voxels in the sgACC at P <.01 in each
cohort but no other regions of more than 24 voxels pre-
dictive of outcome in both cohorts. Thus, sgACC asso-
ciations were confirmed, and no new regions were added
to the a priori list.

Associations were similarly strong across all a priori
regions (Table 3 and eMaterial X), suggesting that the
relationship of the sgACC to outcome is not qualita-
tively unique. Whereas the sgACC explained 29% of the
variation, the other 3 regions explained only 12% addi-
tional variation, yielding a nonsignificant increment
(F3,38 = 2.65 [P =.06]), although the final model with all a
priori regions as predictors was significant. The prediction
equation is as follows:

\[
\text{BDI}_{\text{residual}} = -1.80 + 43.13 \times \text{sgACC} + 17.2 \times \text{Amygdala} + 35.7 \times \text{DLPFC} - 12.0 \times \text{BA24}
\]

Thus, decreased sgACC and possibly regulated DLPFC
activity explains better CT outcome.

Other Measures. Adding common non-fMRI predictors
of response, including scanner, rumination, pupillary
mortality during the task, and demographic variables, did not
explain additional significant variance, and sgACC ac-
tivity remained a significant predictor (eMaterial XI-D).

Other Valences. Positive and neutral words did not pre-
dict residual symptomatology in the combined cohort.
With positive, negative, and neutral words in the same
model, only negative words were a significant predictor.
of residual symptomatology (eMaterial XIII). Thus, prediction was specific to negative words.

Robust Classification. As shown in Table 4, classification for response and remission status in cohort 2 was strong based on an sgACC cutoff solely derived from cohort 1. That is, by relying on the region from a previous study with a cutoff determined in cohort 1, we achieved 74% correct prediction of response and 78% correct prediction of remission in cohort 2. These data suggest that assessment on one scanner yielded a predictive algorithm applicable on another scanner. Adding indices of activity among multiple regions and their connectivity improved classification and continuous prediction of BDI/HRSD residual. Suggesting that accounting for a broader a priori network robustly adds to predictive power; had additional terms not statistically improved classification, they would not have been retained in the final model. The ROC curve analyses (eMaterial VIII; ROC AUC, eMaterial XIC) further showed that initial severity yielded poor classification (AUC < 0.7). Multivariate classifications on the test data set yielded significant estimates reflecting fair (AUC > 0.7) to good (AUC > 0.8) discriminability with just decreased severity and decreased sgACC and good discriminability with the full model, including decreased sgACC-DLPFC connectivity, predicting outcome.

Table 2. Associations of Anatomical sgACC Activity With Clinical Change Using Each Cohort as a Sample of Interest

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Measure</th>
<th>Residual</th>
<th>Response Prediction</th>
<th>Remission Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort 1 CT completers, in Siegle et al† (n = 14)</td>
<td>BDI</td>
<td>R² = 0.41; F₁,15 = 7.78 (P &lt; .02)</td>
<td>84% Correct (P &lt; .06)</td>
<td>84% Correct (P &lt; .08)</td>
</tr>
<tr>
<td>Cohort 1 CT completers and noncompleters, estimated (n = 16)</td>
<td>BDI</td>
<td>R² = 0.36; F₁,15 = 7.79 (P &lt; .01)</td>
<td>81% Correct (P &lt; .07)</td>
<td>75% Correct (P &lt; .14)</td>
</tr>
<tr>
<td>Participants with valid data (n = 17)</td>
<td>HRSD</td>
<td>R² = 0.34; F₁,16 = 7.77 (P &lt; .01)</td>
<td>76% Correct (P &lt; .41)</td>
<td>82% Correct (P &lt; .045)</td>
</tr>
<tr>
<td>Cohort 2 (n = 27)</td>
<td>BDI</td>
<td>R² = 0.17; F₁,26 = 4.91 (P &lt; .04)</td>
<td>78% Correct (P &lt; .04)</td>
<td>74% Correct (P &lt; .04)</td>
</tr>
<tr>
<td>Participants with valid data (n = 32)</td>
<td>HRSD</td>
<td>R² = 0.04; F₁,31 = 1.24 (P &lt; .27)</td>
<td>78% Correct (P &lt; .02)</td>
<td>66% Correct (P &lt; .38)</td>
</tr>
<tr>
<td>Cohorts 1 and 2 participants with pretreatment/posttreatment BDI scores (n = 43)</td>
<td>BDI</td>
<td>R² = 0.29; F₁,42 = 16.61 (P &lt; .001)</td>
<td>79% Correct (P &lt; .003)</td>
<td>72% Correct (P &lt; .01)</td>
</tr>
<tr>
<td>Participants with valid data (n = 49)</td>
<td>HRSD</td>
<td>R² = 0.12; F₁,48 = 6.29 (P &lt; .02)</td>
<td>78% Correct (P &lt; .001)</td>
<td>69% Correct (P &lt; .06)</td>
</tr>
</tbody>
</table>

Abbreviations: BDI, Beck Depression Inventory; CT, cognitive therapy; HRSD, Hamilton Rating Scale for Depression; sgACC, subgenual anterior cingulate cortex.

Goal 4: Clinically Applicable Formulations: Psychometrics, z and Symptom-Change Scores

Reliability and Discriminant Validity. Subgenual anterior cingulate z scores and sgACC reactivity had moderate test-retest reliability in controls undergoing testing approximately 16 weeks apart (27 participants; r = 0.39 [P < .04]), which was similar for each scanner (cohort 1, r = 0.44; cohort 2, r = 0.35) and did not change when the scanner was covaried (r = 0.39). Controls’ mean sgACC activity did not change between the scans (t₁₀₀ = 1.36 [P = .19]; Cohen d = 0.26). All but 1 had a pretest z score less than 0.5, and all but 2 had a posttest z score less than 0.5, suggesting stability within a restricted range, decreasing the estimated reliability. Reliability for the composite predictor associated with the sgACC, amygdala, DLPFC, and BA24 was no higher (r = 0.36). Subgenual anterior cingulate cortex z-scored reactivity was minimally correlated with initial depressive severity among depressed patients (49 patients; r_HRSD = −0.10 [P = .50]; r_DLPFC = −0.21 [P = .13]).

Symptom Change Scores. Associations of sgACC with individually interpretable BDI_post - pretest were nearly identical to associations of percentage change with BDI_residual (Figure 2D and eMaterials VIII and X). For example, in cohort 1, treatment completers’ sgACC, predicted 54%
of the variance in BDI\textsuperscript{posttest-pretest} ($F_{1,15}=16.32$ [P = 0.001]), and $R^2=0.29$ in the combined sample ($F_{1,30}=16.87$ [P < 0.001]). Cutoffs for sgACC in the range of 0.46 to 0.74 strongly predicted response across measures for both samples (76\%-81\% correct response classification), suggesting that high levels of sgACC activity predicted non-response. Similarly, cutoffs closer to 0 predicted remission, so participants with average or higher levels of sgACC responses compared with controls were unlikely to experience remission in CT.

Goal 5: Change in sgACC Activity With Treatment

Forty depressed participants had pretreatment and posttreatment fMRI and BDI scores. Participants with the lowest pretreatment sgACC activity (primarily CT remitters) also had the lowest posttreatment activity ($r_{\text{pretreatment-posttreatment}}=0.39; F_{1,39}=6.82$ [P = 0.01]) (Figure 4A). Depressed participants with pretreatment activity below the predicted response threshold who experienced remission had pretreatment and posttreatment sgACC activity below that of controls throughout the trial (Figure 4B) and did not increase significantly (eTable 4, t test, pretreatment vs posttreatment). In contrast, controls with low pretreatment activity increased significantly (eTable 4 and eFigure 7B), and depressed nonremitters with low pretreatment activity nearly so (Figure 4C, for statistics, see eTable 4). Moreover, depressed remitters with low pretreatment activity had a nonsignificantly smaller proportion of participants who increased and a lower mean level of increase than did controls or nonremitters with low pretreatment activity (eMaterial XII-B). Thus, we cannot conclude that sgACC activity increased as a function of treatment. Qualitatively, in contrast, 5 of 6 nonremitters with low pretreatment activity (Figure 4A, green squares) showed increased sgACC activity after treatment. Similarly, 3 of 5 remitters who had high pretreatment sgACC activity had decreased posttreatment activity (Figure 4A, blue squares; Figure 4D shows the average BOLD response), whereas 8 of the 12 nonremitters with high pretreatment sgACC ac-

![Figure 3.](https://archpsyc.jamanetwork.com/)

**Figure 3.** Decreased empirically defined subgenual anterior cingulate activity associated strongly with response. A, Pretreatment regions associated with decreased depressive severity from Siegle et al.\textsuperscript{4} B, The new cognitive therapy samples ($n=40; R^2=0.40$ [P < 0.001]). Green indicates the anatomical region of interest used as a mask; orange, regions only in the new data set; and red, voxels that overlap. C and D, These regions were reflected in predictive regions in cohorts 1 ($R^2=0.43-0.79$ [P < 0.01]) and 2 ($R^2=0.26$ [P < 0.01]), respectively.

### Table 3. Prediction of Residual Symptoms Using an A Priori Network With Combined Cohort as a Sample of Interest\textsuperscript{a}

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Measure</th>
<th>Region</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-order relationships,</td>
<td>BDI</td>
<td>sgACC</td>
<td>$R^2=0.29$</td>
</tr>
<tr>
<td>$P&lt;.05$ for all comparisons \textsuperscript{b}</td>
<td>Right amygdala</td>
<td>$R^2=0.20$</td>
<td></td>
</tr>
<tr>
<td>Multivariate relationships, full model</td>
<td>BDI</td>
<td>Constant</td>
<td>$t_{\text{st}}=0.47$ (P = .71)</td>
</tr>
<tr>
<td>$R^2=0.43$; $F_{1,30}=7.39$ (P &lt; .001) \textsuperscript{c}</td>
<td>sgACC</td>
<td>$t_{\text{st}}=0.24$ (P = .10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>amygda</td>
<td>$t_{\text{st}}=0.30$ (P = .05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BA24 in the</td>
<td>$t_{\text{st}}=−0.10$ (P = .52)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VMPFC</td>
<td>$t_{\text{st}}=−0.10$ (P = .52)</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BA, Brodmann area; BDI, Beck Depression Inventory; DLPFC, dorsolateral prefrontal cortex; sgACC, subgenual anterior cingulate cortex; $t_{\text{st}}$, standardized beta; VMPFC, ventromedial prefrontal cortex.

\textsuperscript{a} Classification not evaluated in the multivariate model without robust estimation given the potential for type I error; rather, evaluations in Table 4 reflect robust estimations from the multivariate model (n = 43).

\textsuperscript{b} See eMaterial VIII and eTable S2A for complete statistics.

\textsuperscript{c} See eMaterial VIII and eTable S2A for complete statistics.
tivity increased (Figure 4A, purple squares). Thus, the emerging picture is that participants with high posttreatment sgACC activity did not experience remission.

**COMMENT**

This study replicates and extends previous results, suggesting that decreased sgACC reactivity to negative words is prognostic for response, remission, and clinical change in CT for depression. The study was prognostic in an efficacy sample and a more clinically diverse sample who received more clinically representative care, although effect sizes were somewhat reduced, possibly because of increased patient and therapist heterogeneity and nonstandard administration of CT. Using measures of brain function interpretable at a single-subject level (z scores) and simple change scores preserved prognostic utility. Response, remission, and change in severity were robustly predicted on the basis of a few a priori regions when using thresholds derived in a different sample on a different scanner above and beyond pretreatment severity. These data suggest that the proposed assessment could have scientific and practical utility. However, remitters with low pretreatment sgACC activity did not generally demonstrate increased sgACC activity, suggesting that successful treatment did not operate by normalizing this mechanism but rather by remaining more like healthy individuals from pretreatment measurements.

Although failure to observe an increase in sgACC activity after treatment could be the result of the index's

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Measure</th>
<th>Residual</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CT: training set, all brain regions</td>
<td>BDI</td>
<td>$R^2 = 0.28$ ($P &lt; .02$)</td>
<td>75% Correct</td>
<td>69% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 75%</td>
<td>Sensitivity, 62%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 90%</td>
<td>Specificity, 75%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 1.28$</td>
<td>$d' = .99$ ($P = .06$)</td>
</tr>
<tr>
<td></td>
<td>HRSD</td>
<td>$R^2 = 0.21$ ($P &lt; .01$)</td>
<td>76% Correct</td>
<td>65% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 92%</td>
<td>Sensitivity, 57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 40%</td>
<td>Specificity, 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 1.13$</td>
<td>$d' = 0.7$ ($P = .12$)</td>
</tr>
<tr>
<td>2. CT: generalization set, initial severity alone</td>
<td>BDI</td>
<td>$R^2 = 0.0$ ($P &lt; .74$)</td>
<td>48% Correct</td>
<td>38% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 53%</td>
<td>Sensitivity, 23%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 40%</td>
<td>Specificity, 54%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 0.18$</td>
<td>$d' = 0.64$ ($P = .03$)</td>
</tr>
<tr>
<td></td>
<td>HRSD</td>
<td>$R^2 = 0.0$ ($P = .63$)</td>
<td>56% Correct</td>
<td>47% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 60%</td>
<td>Sensitivity, 38%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 50%</td>
<td>Specificity, 53%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 0.25$</td>
<td>$d' = 0.23$ ($P = .29$)</td>
</tr>
<tr>
<td>2. CT: generalization set, initial severity + sgACC</td>
<td>BDI</td>
<td>$R^2 = 0.0$ ($P &lt; .07$)</td>
<td>74% Correct</td>
<td>78% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 100%</td>
<td>Sensitivity, 86%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 30%</td>
<td>Specificity, 69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 1.04$</td>
<td>$d' = 1.57$ ($P = .003$)</td>
</tr>
<tr>
<td></td>
<td>HRSD</td>
<td>$R^2 = 0.02$ ($P &lt; .07$)</td>
<td>81% Correct</td>
<td>62% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 95%</td>
<td>Sensitivity, 69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 58%</td>
<td>Specificity, 58%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 1.86$</td>
<td>$d' = 0.7$ ($P = .08$)</td>
</tr>
<tr>
<td>2. CT: generalization set, initial severity + all brain regions, retained: baseline BDI, left DLPFC, sgACC, $\phi_{\text{L DLPFC, sgACC}}$</td>
<td>BDI</td>
<td>$R^2 = 0.15$ ($P &lt; .003$)</td>
<td>81% Correct</td>
<td>63% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 100%</td>
<td>Sensitivity, 79%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 44%</td>
<td>Specificity, 46%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 1.43$</td>
<td>$d' = 0.7$ ($P = .03$)</td>
</tr>
<tr>
<td></td>
<td>HRSD</td>
<td>$R^2 = 0.07$ ($P = .02$), (points off) = 3.8</td>
<td>75% Correct</td>
<td>66% Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity, 95%</td>
<td>Sensitivity, 85%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Specificity, 42%</td>
<td>Specificity, 53%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$d' = 1.43$</td>
<td>$d' = 1.09$ ($P = .02$)</td>
</tr>
</tbody>
</table>

Abbreviations: BDI, Beck Depression Inventory; CT, cognitive therapy; DLPFC, dorsolateral prefrontal cortex; HRSD, Hamilton Rating Scale for Depression; M, mean; phi, spectral correlation; $\phi_{\text{L DLPFC, sgACC}}$, spectral correlation between the left DLPFC and subgenual anterior cingulate (sgACC).

$^a$P values for random forests were estimated via permutation tests on percentage correct using the same random forest methodology. Regression and classification accuracy estimates are unbiased, computed on out-of-bag samples for robust generalization. A more elaborated version of this Table, including full receiver operating characteristic analyses, is included in eMaterial VIII-A, eTable S2A.
Figure 4. Relationships of pretreatment subgenual anterior cingulate cortex (sgACC) reactivity to posttreatment sgACC activity (low and high).

A. Continuous change for the temporal region of interest. B-E. Waveforms in participants who were predicted to remit (pretreatment percentage change < −0.02) or not remit. Areas significant at \( P < .05 \) after analyses of variance at each scan are highlighted in gray.

low reliability or regression to the mean, these explanations would not explain the association of initial negative-going responses with the smallest sgACC change. Thus, we suggest that increasing sgACC function is not a mechanism of clinical change. Rather, sgACC activity could interfere with voluntary emotion regulation essential for CT, possibly because of its roles in sadness upregulation or limbic monitoring, possibly via corticocortical connectivity, or possibly because of automatic down-regulation of emotion\(^4\) preventing the use of considered reappraisal emphasized in CT. These interpretations support the continued development of interventions that decrease sgACC activity\(^3\) and do not suggest that CT or other interventions (eg, neurofeedback) should work to increase ventromedial function. Alternately, low sgACC activity could facilitate CT response if it is easiest to learn to challenge thoughts when emotion-monitoring/generation processes naturally disengage. That multiple regions associated with limbic reactivity and regulation explain overlapping and independent variance suggests that outcome reflects functioning of a wider network.

More practically, because we used a priori regions with a single 12-minute fMRI and 7-minute structural acquisition, automated preprocessing, and scores that can be calculated on single subjects, our algorithm is feasible and costs significantly less than $1000 at most scanning centers. It is cost-effective for use in clinics in a 30-minute MRI appointment without radiologist interpretation. Cross-scanner generalizability is likely because sgACC activity patterns were qualitatively different for remitters (activity decreased from a prestimulus baseline) and nonremitters (activity increased) (Tables 2 to 4 and Figure 2E). Thus, interscanner scaling differences did not affect prediction. Because sensitivity using just the sgACC was low for practical use (Table 2) but adequate when other regions were accounted for (100% sensitivity in the Table 4 generalization set), it may be useful to assess a network of regions in prediction or to consider the sgACC primarily as a method to help patients decide what interventions not to try first. Because predictions were stronger in the highly controlled cohort, therapists who adhere strongly to a single treatment model may make the most valid use of predictive neuroimaging.

Increased activity in anatomically proximal regions including the rostral cingulate has been shown to positively predict outcome to antidepressant therapy using resting state positron emission tomography\(^13,45-51\) and electroencephalography.\(^52\) Thus, ventral cingulate activity could, on prospective replication, yield an algorithm for selection into CT vs antidepressant treatment.\(^51\) Assessing sgACC function during multimodal assessments could also help to validate mechanisms for less interpretable prognostic indicators, such as psychophysiological assessments.\(^53\) Nevertheless, because fMRI-derived sgACC activity explained variance independent of other common predictive measures (eMaterial XI-D), it may represent a more unique construct in the literature.

This study had multiple limitations. Although they constitute one of the largest prognostic fMRI studies in psychiatry, our samples were small by the standards of clinical trials. Randomized trials are needed to make testable inferences regarding the relationship of sgACC function specifically to CT outcome. Future examination of depressed participants receiving medication therapy will help to generalize to clinically representative conditions. If healthy individuals’ moderate reliability (\( r < 0.5 \)) is representative of patients, refinement in multiple-baseline patient studies is needed before inferring clinical applicability to single patients. Amygdala results were not consistent with the findings of Siegle et al,\(^4\) possibly reflecting our use of an anatomically defined amygdala region vs the empirically derived smaller region in the earlier study. Our index captured sustained processing of emotional information but could have missed preparatory or initial aspects of reactivity.

These limitations notwithstanding, our data suggest that a simple, automated, and theoretically motivated neuroimaging assessment yields a marker of whether an in-
individual is likely to respond to a specific validated treat-
ment for unipolar depression, CT. Although the data do not
suggest that neuroimaging should be used clinically at this
time, they do indicate that neuroimaging is ready for “next-step” investigations, including larger random-
ized clinical trials, creating age, sex, and scanner norms
and considerations regarding dissemination.

Submitted for Publication: May 9, 2011; final revision
received December 7, 2011; accepted December 9, 2011.
Correspondence: Greg J. Siegle, PhD, Western Psychi-
atric Institute and Clinic, University of Pittsburgh School
of Medicine, 3811 O’Hara St, Pittsburgh, PA 15213 (gsiegle
@pitt.edu).

Financial Disclosure: Dr Siegle is an unpaid consultant
for Trial IQ and Neural Impact. During the past 3 years,
Dr Thase reported having served as an advisor/consultant
for Alkermes, AstraZeneca, Bristol-Myers Squibb Com-
pany, Eli Lilly and Company, GlaxoSmithKline, Janssen
Pharmaceutica, MedAvante Inc, Merck Inc, Neuronetics
Inc, Novartis Inc, Otsuka Inc, FamLab Inc, Pfizer Inc, Phar-
maneuroboost Inc, Shire US Inc, Supernus Pharmacetical-
s, and Wyeth Pharmaceutilicals. In 2009 and 2010, Dr
Thase had speakers bureau affiliations with AstraZeneca,
Bristol-Myers Squibb Company, Eli Lilly and Company,
Merck Inc, and Wyeth Pharmaceutilicals. Dr Thase has eq-
uity holdings in MedAvante Inc; has received royalties from
American Psychiatric Publishing Inc, Guilford Publica-
tions, and Herald House; and has a family relationship
with senior staff at Embryon Inc (formerly Cardinal Health
and Advogen). During the past 3 years, Dr Thase has
received research funding from the Agency for
Healthcare Research and Quality, Eli Lilly and Company,
GlaxoSmithKline, the National Institute of Mental Health,
Otsuka Inc, Pfizer Inc, Pharmaneuroboost, and Roche,
Inc. Dr Friedman has speakers bureau or advisory board
affiliations with AstraZeneca, Eli Lilly and Company,
GlaxoSmithKline, and Pfizer Wyeth-Ayerst and has
obtained grant/research support from Aspect Medical Sys-
tems, Indevus, AstraZeneca, Bristol-Myers Squibb, Pfizer,
sanofi-aventis, Wyeth-Ayerst, Cyberonics, Novartis,
NorthStar/St Jude Medical, Medtronic, and Respironics.

Funding/Support: This study was supported by grants
MH074807, MH082998, MH58356, MH58397, and
MH69618 from the National Institutes of Health; by the
Pittsburgh Foundation; and by the Emmerling Fund.

Role of the Sponsor: The sponsors had no role in the de-
sign and conduct of the study; in the collection, analy-
is, and interpretation of the data; or in the preparation,
review, or approval of the manuscript.

Previous Presentation: Portions of this manuscript were
presented at the 2010 annual meeting of the World Con-
gress for Behavioral and Cognitive Therapies; June 5, 2010;
Boston, Massachusetts.

Online-Only Material: The eMaterials cited in the text
and the eFigures, eTables, and eReferences are available
at http://www.pitt.edu/~gsiegle/Siegle-fMRI-Prediction-
Archives12-AuthorMaterial.pdf.

Additional Contributions: We thank Agnes Haggerty, BS,
Mauri Cesare, BA, our therapists, and the staff of the Mood
Disorders Treatment and Research Program at Western
Psychiatric Institute and Clinic; the interped commu-
nity therapists who worked on the study; and our pa-
ients, without whom none of this work would have been
possible.

REFERENCES

2. Hollon SD, Thase ME. Markowitz JC. Treatment and prevention of depression.
3. Kemp AH, Gordon E, Rush AJ, Williams LM. Improving the prediction of treat-
ment response in depression: integration of clinical, cognitive, psychophysio-
logical, neuroimaging, and genetic measures. CNS Spectr. 2008;13(12):1066-
1088.
4. Siegle GJ, Carter CS, Thase ME. Use of fMRI to predict recovery from unipolar
735-738.
5. Fu CH, Williams SC, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, Don-
aldson C, Suckling J, Andrew CM, Steinert H, Murray RM. Neural responses to
sad facial expressions in major depression following cognitive behavioral therapy.
Outcome of cognitive-behavioral therapy for depression: relation to hemi-
spheric dominance for verbal processing. J Abnorm Psychol. 1997;106(1):
138-144.
7. Mayberg HS. Modulating dysfunctional limbic-cortical circuits in depression: to-
wards development of brain-based algorithms for diagnosis and optimised treat-
8. Mayberg HS, Liotti M, Brannan SK, McGinnis S, Maharun RK, Jerabek PA, Silva
JA, Tekeli JL, Martin CC, Lancaster JL, Fox PT. Reciprocal limbic-cortical func-
tion and negative mood: converging PET findings in depression and normal
8(8):813-829.
EM, Williams PM, Reed LJ, Mitterschiffthaler MT, Suckling J, Bullmore ET.
Attenuation of the neural response to sad faces in major depression by antide-
pressant treatment: a prospective, event-related functional magnetic resonance
lation of cortical-limbic pathways in major depression: treatment-specific ef-
12. Mayberg HS, Braman SK, Maharun RK, Jerabek PA, Brickman JS, Tekeli JL, Silva
JA, McGinnis S, Glass TG, Martin CC, Fox PT. Cingulate function in depression:
Subgenual cingulate and visual cortex responses to sad faces predict clinical out-
come during antidepressant treatment for depression. J Affect Disord. 2010;
120(1-3):120-125.
Amygdala and ventral anterior cingulate activation predicts treatment response
to cognitive behaviour therapy for post-traumatic stress disorder. Psychosom
studies of emotion, personality, and social cognition. Perspect Psychol Sci
16. Kriegeskorte N, Simmons WK, Belsling PS, Baker CJ. Circular analysis in sys-
535-540.
17. DeRubeis RJ, Siegle GJ, Hollon SD. Cognitive therapy versus medication for de-
9(10):788-796.
Gabriel S, Gabrieli JD, Gotlib IH. Amygdala reactivity to emotional faces pre-
19. Davidson RJ. Asymmetric brain function, affective style and psychopathology:
20. Siegle GJ, Thompson W, Carter CS, Steinhauser SR, Thase ME. Increased amygd-
al and decreased dorsolateral frontal BOLD responses in unipolar depres-
activity and abnormal responses to errors in subjects with elevated depressive
22. Lemogne C, le Bastard G, Mayberg H, Vollie E, Bergouignan L, Leheny S, Al-

©2012 American Medical Association. All rights reserved.


