

# 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

**Author Affiliations:** Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania (Drs Siegle and Friedman, Mss Collier and Berman, and Mr Feldmiller); and Departments of Psychiatry, University of California, San Diego (Dr Thompson), and University of Pennsylvania School of Medicine, Philadelphia (Dr Thase).

**C**OGNITIVE THERAPY (CT)<sup>1</sup> is a common empirically supported intervention that addresses systematic negative thinking and is effective for 40% to 60% of patients with unipolar depression.<sup>2</sup> Knowing which patients are likely to benefit from CT could increase response rates and decrease costs by targeted referrals. However, biological predictors of CT outcome<sup>3-6</sup> have not been adopted clinically. This gap in translation may result from the lack of reliability and validity data for these measures or from the inability to overcome variability across scanners/laboratories. Herein we examined whether our previously observed association of CT outcome with pretreat-

ment activity in the subgenual anterior cingulate cortex (sgACC)<sup>4</sup> replicates in multiple new samples and scanners.

The sgACC is an intuitive target to examine as a treatment predictor because it has connections to limbic regions such as the amygdala and has been suggested to serve as a proximal regulator of limbic function.<sup>7</sup> The sgACC further has abnormalities in activity in depressed individuals,<sup>8,9</sup> changes in CT and treatment with medications,<sup>10,11</sup> and associations with symptom change in depression.<sup>12,13</sup> In addition, the sgACC is cytoarchitectonically uniform and has easily anatomically identifiable boundaries, making results likely to be generalizable across studies. The sgACC is prognostic for clini-

cal change in multiple CT studies using different paradigms in depression<sup>4,5</sup> and other disorders, such as post-traumatic stress disorder.<sup>14</sup>

To pave the way for clinical adoption of this prognostic 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.<sup>4</sup> 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).<sup>15,16</sup> 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,<sup>17</sup> including the left amygdala, as in previous studies,<sup>4,18</sup> and in regions associated with regulatory control that have decreased functioning in depression, including the dorsolateral prefrontal cortex (DLPFC),<sup>8,19,20</sup> the rostral anterior cingulate,<sup>21-23</sup> and their functional corticolimbic/corticocortical relationships.<sup>24,25</sup> 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,<sup>26-28</sup> 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.<sup>4</sup> 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 increases in sgACC activity yield increased sadness,<sup>29</sup> (2) trait-related increased sgACC activity is associated with higher sadness<sup>30</sup> and increased depressive severity,<sup>31</sup> and (3) sgACC inhibition via deep brain stimulation facilitates recovery.<sup>32</sup> 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.

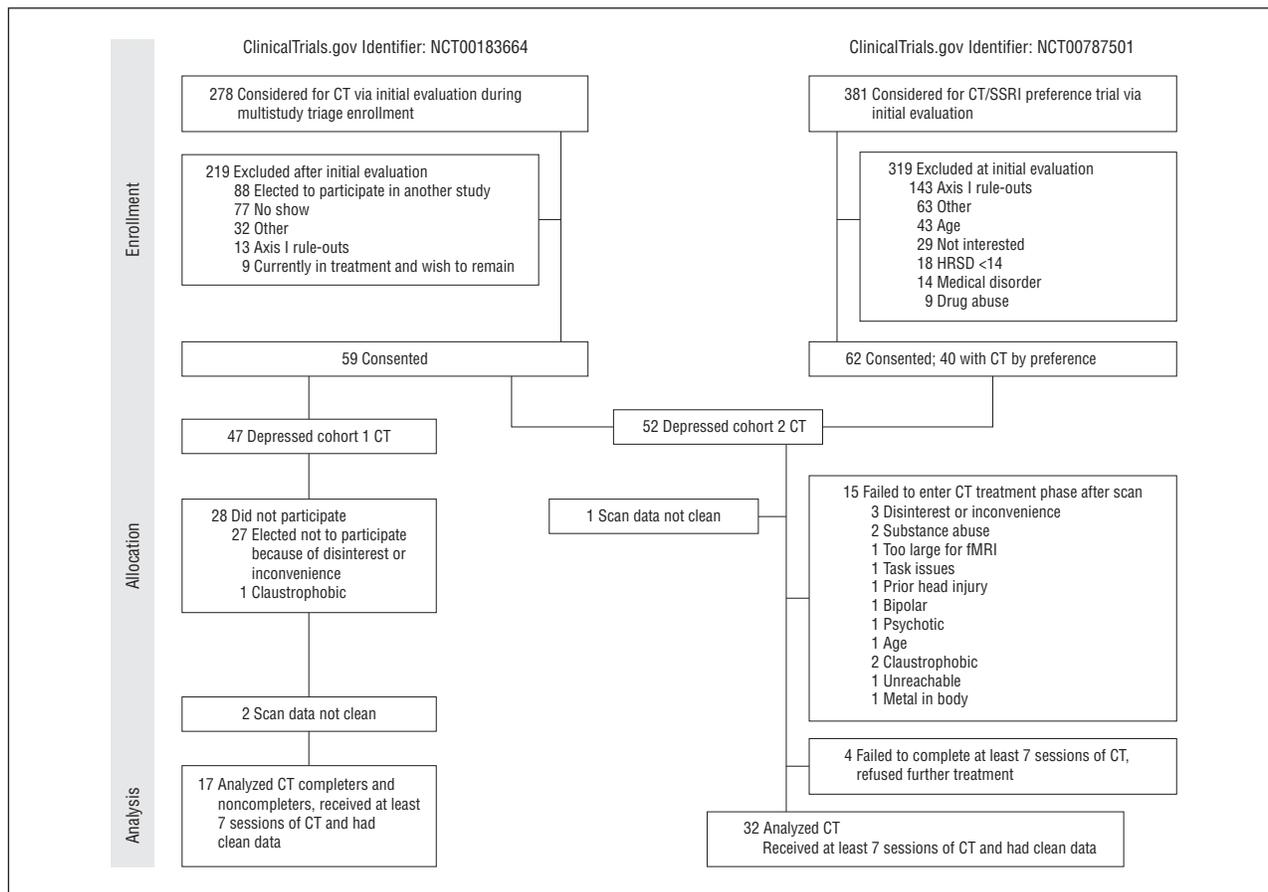
To evaluate whether sgACC activity normalizes in treatment, a sample of healthy never-depressed control participants were recruited to establish a normative baseline. The controls also afforded us the ability to examine pretreatment activity with respect to normative function (ie, as clinically interpretable 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<sup>33</sup>). 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/~gsiegle/Siegle-fMRI-Prediction-Archives12-AuthorMaterial.pdf>]). After attrition and data cleaning (fully described in Figure 1), cohort 1 (the efficacy cohort; from M.E.T., principal 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]<sup>34</sup>) treated by the same 3 therapists (including E.S.F.) from the study by Siegle et al<sup>4</sup> 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  $\geq 2$  tapes per therapist by other therapists or outside raters) receiving high marks on the Cognitive Therapy Scale<sup>35</sup> (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



**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.

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<sup>36</sup> (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),<sup>37</sup> 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<sup>1</sup> (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 fMRI session (after completing CT or approximately 12 weeks from their first fMRI 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 1. Demographic and Clinical Characteristics<sup>a</sup>**

Characteristic	Depressed Participants/CT		Control Participants	
	Cohort 1	Cohort 2	Cohort 1	Cohort 2
No. of participants	17	32	15	20
Male sex, No.	4	5	7	5
White race, No.	14	29	12	17
Age, y				
Range	24-51	19-55	25-54	21-52
Mean (SD)	37.41 (9.17)	36.13 (11.11)	38.93 (9.74)	32.95 (9.37)
Educational level, mean (SD), y	15.35 (2.23)	15.06 (2.60)	15.67 (2.26)	16.30 (2.13)
No. of depressive episodes, median (SD)	5.59 (3.52)	4.03 (4.83)	0	0
NAART VIQ equivalent, mean (SD)	110.31 (8.68)	108.42 (9.92)	111.61 (7.97)	109.21 (8.96)
Emotion ratings for negative words, mean (SD) <sup>b,c</sup>	2.55 (0.66)	2.11 (0.62)	2.99 (0.65)	2.49 (0.72)
RT to negative words, mean (SD), ms <sup>d</sup>	1403.82 (269.41)	1637.34 (526.16)	1411.31 (390.26)	1242.20 (374.90)
Personal relevance ratings for negative words, mean (SD) <sup>e,f</sup>	3.07 (0.66)	3.03 (0.57)	2.33 (0.72)	2.07 (0.49)
Clinical change in cohort 1 completers <sup>g</sup>				
Pretreatment BDI score, mean (SD)	33.0 (8.8)	...	...	...
Posttreatment BDI score, mean (SD) <sup>g</sup>	10.7 (10.2)	...	...	...
BDI score change, mean (SD)	22.3 (13.4)	...	...	...
Responders, No. (%) <sup>h</sup>	9 (69)	...	...	...
Remitters, No. (%) <sup>i</sup>	8 (62)	...	...	...
Clinical change in whole sample: No. completing BDI	16	27	15	20
Pretreatment BDI score, mean (SD)	34.4 (8.7)	30.7 (9.2)	1.87 (3.58)	1.26 (1.63)
Posttreatment BDI score, mean (SD)	14.1 (12.2)	11.3 (9.6)	...	...
Change in BDI score, mean (SD)	20.3 (13.4)	19.4 (9.7)	...	...
Responders with BDI score, No. (%) <sup>h</sup>	10 (63)	17 (63)	...	...
Remitters with BDI score, No. (%) <sup>i</sup>	8 (50)	14 (51)	...	...
Clinical change in HRSD for whole sample, No.	17	32	...	...
Pretreatment HRSD score, mean (SD)	24.6 (5.7)	20.6 (4.8)	...	...
Posttreatment HRSD score, mean (SD)	9.6 (6.5)	8.1 (5.8)	...	...
Change in HRSD score, mean (SD)	15.0 (7.7)	12.5 (5.6)	...	...
Responders with HRSD, No. (%) <sup>h</sup>	10 (59)	20 (63)	...	...
Remitters with HRSD, No. (%) <sup>i</sup>	7 (47)	13 (41)	...	...

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.

<sup>a</sup>The *t* 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.<sup>4</sup> Remaining lines generalize that result.

<sup>b</sup>Rating of 1 indicates very negative; 7, very positive.

<sup>c</sup>For depressed participants vs controls,  $t_{82} = 2.89$  ( $P = .005$ );  $d = 0.64$ .

<sup>d</sup>For depressed participants vs controls,  $t_{82} = -2.52$  ( $P = .01$ );  $d = 0.56$ .

<sup>e</sup>Rating of 1 indicates not relevant; 5, very relevant.

<sup>f</sup>For depressed participants vs controls,  $t_{82} = 6.57$  ( $P < .001$ );  $d = 1.43$ .

<sup>g</sup>Indicates CT completers only ( $n = 13$ ) as defined in Goldapple et al.<sup>11</sup>

<sup>h</sup>Response was defined as a 50% decrease in score.

<sup>i</sup>Remission was defined as a score less than 10.

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

## fMRI TASK AND PROCESSING

### Apparatus

Twenty-nine 3.2-mm slices were acquired parallel to the anterior commissure–posterior commissure line using a posterior to anterior echoplanar imaging pulse sequence to minimize 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 12-second trial. Stimuli

were displayed in black on a white background via a back-projection screen (0.88° visual angle). Responses were recorded using a data glove (Psychology Software Tools).

### Personal Relevance Rating Task

As in our previous publications,<sup>4,20</sup> 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<sup>38</sup> were used as in our

previous studies of depression<sup>4,20,39-41</sup> (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 study<sup>4</sup> 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 study<sup>4</sup> in eMaterial IV). The same “reactivity” index used in our previous study<sup>4</sup> 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.<sup>4</sup> 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,<sup>4</sup> 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,<sup>4</sup> 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,  $sgACC_z = [X - \text{mean}_{D_{\text{controls}}}] / SD_{\text{controls}}$ ) 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,<sup>42</sup> 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 al<sup>43</sup> to capture systemic effects involving functional connectivity. Regression and classification forests were trained on cohort 1 using ran-

dom 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.

## RESULTS

### DEMOGRAPHIC AND BEHAVIORAL DATA

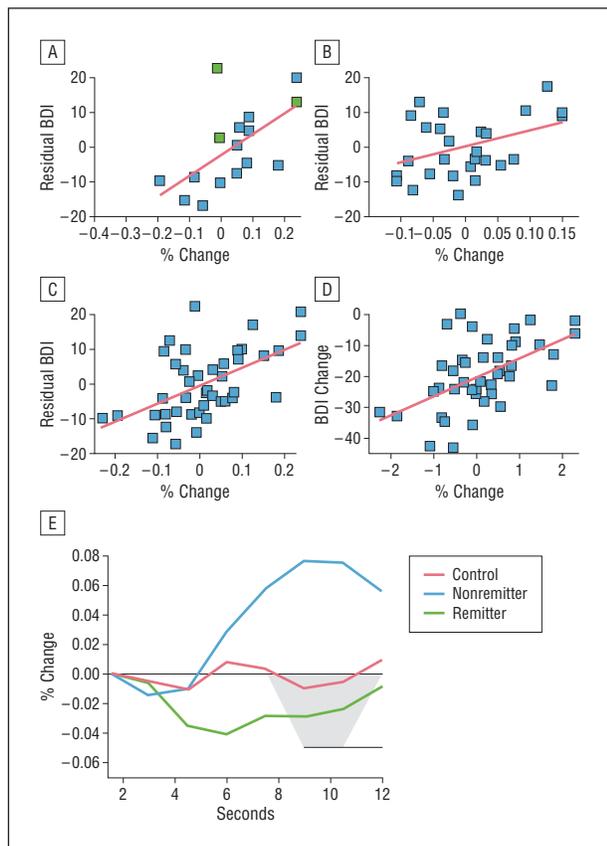
Patients and controls did not differ significantly on demographic variables between cohorts or groups (Table 1). Group  $\times$  cohort analyses of variance revealed that patients rated negative words as more negative and more personally relevant and reacted more slowly than did controls (eMaterial VI). Cohort 2 rated negative words as more negative ( $t_{82} = 3.27$  [ $P = .002$ ]; Cohen  $d = 0.73$ ) with no other cohort main effects. There were no group  $\times$  cohort interactions on valence or meaning ratings. Because group  $\times$  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,<sup>4</sup> decreased sgACC activity was strongly corre-



**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  $\geq 10$ ) 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.

lated with stronger clinical change (more negative  $BDI_{residual}$ ) in cohort 1, whether or not completers were included. This prediction generalized to the HRSD, which was collected on a slightly larger subset of participants. The ROC curves (eMaterial IX) yielded a significant area under the curve (AUC), reflecting good discrimination (eMaterial VIII-A). Voxelwise associations with  $BDI_{residual}$  revealed sgACC to be associated with even more of the variance ( $>80\%$  in some voxels; **Figure 3C**), suggesting possible optimization and, as in Siegle et al,<sup>4</sup> relative specificity to the sgACC.

### Goal 2: Effectiveness Generalization

We examined effectiveness using a more clinically heterogeneous sample with minimally supervised community therapists. The same general pattern of results held significantly in cohort 2, although effect sizes were somewhat weaker (Figures 2 and 3 and Table 2), potentially owing to clinical and therapist heterogeneity. Importantly, 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 combined, 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, reflecting fair discrimination. Remitters were characterized by downward hemodynamic responses, whereas nonremitters 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 ( $F_{2,76}=5.56$  [ $P=.006$ ]).

### 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 statistically significant at  $P < .001$  ( $P < .05$  corrected) and clinically significant, explaining more than 40% of the variation in  $BDI_{residual}$  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 detected. 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 predictive of outcome in both cohorts. Thus, sgACC associations 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 XI), suggesting that the relationship of the sgACC to outcome is not qualitatively unique. Whereas the sgACC explained 29% of the variation, the other 3 regions explained only 12% additional variation, yielding a nonsignificant increment ( $F_{3,38}=2.65$  [ $P=.06$ ]), although the final model with all a priori regions as predictors was significant. The prediction equation is as follows:

$$BDI_{residual} = -1.80 + 43.13 \times \text{sgACC} + 17.2 \times \text{Amygdala} + 35.7 \times \text{DLPFC} - 12.0 \times \text{BA24}.$$

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

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

**Other Valences.** Positive and neutral words did not predict residual symptomatology in the combined cohort. With positive, negative, and neutral words in the same model, only negative words were a significant predictor

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

Cohort	Measure	Residual	% Correct, Permutation P Value	
			Response Prediction	Remission Prediction
Cohort 1 CT completers, in Siegle et al <sup>4</sup> (n = 14)	BDI	$R^2 = 0.41$ ; $F_{1,12} = 7.78$ ( $P = .02$ )	84% Correct ( $P = .1$ ) [.04 based on $d'$ ] Sensitivity, 100% Specificity, 86% $d' = 2.2$	84% Correct ( $P = .08$ ) [.04 based on $d'$ ] Sensitivity, 100% Specificity, 75% $d' = 2.1$
Cohort 1 CT completers and noncompleters, estimated (n = 16)	BDI	$R^2 = 0.36$ ; $F_{1,15} = 7.79$ ( $P = .01$ )	81% Correct ( $P = .07$ ) Sensitivity, 83% Specificity, 80% $d' = 1.8$	75% Correct ( $P = .14$ ) Sensitivity, 100% Specificity, 50% $d' = 1.4$
Participants with valid data (n = 17)	HRSD	$R^2 = 0.34$ ; $F_{1,16} = 7.77$ ( $P = .01$ )	76% Correct ( $P = .41$ ) Sensitivity, 80% Specificity, 67% $d' = 1.3$	82% Correct ( $P = .045$ ) Sensitivity, 100% Specificity, 57% $d' = 1.7$
Cohort 2 (n = 27)	BDI	$R^2 = 0.17$ ; $F_{1,26} = 4.91$ ( $P = .04$ )	78% Correct ( $P = .04$ ) Sensitivity, 40% Specificity, 100% $d' = 1.5$	74% Correct ( $P = .04$ ) Sensitivity, 54% Specificity, 93% $d' = 1.6$
Participants with valid data (n = 32)	HRSD	$R^2 = 0.04$ ; $F_{1,31} = 1.24$ ( $P = .27$ )	78% Correct ( $P = .02$ ) Sensitivity, 42% Specificity, 100% $d' = 1.61$	66% Correct ( $P = .38$ ) Sensitivity, 32% Specificity, 100% $d' = 1.32$
Cohorts 1 and 2 participants with pretreatment/posttreatment BDI scores (n = 43)	BDI	$R^2 = 0.29$ ; $F_{1,42} = 16.61$ ( $P < .001$ )	79% Correct ( $P = .003$ ) Sensitivity, 50% Specificity, 96% $d' = 1.79$	72% Correct ( $P = .01$ ) Sensitivity, 38% Specificity, 95% $d' = 1.39$
Participants with valid data (n = 49)	HRSD	$R^2 = 0.12$ ; $F_{1,48} = 6.29$ ( $P = .02$ )	78% Correct ( $P = .001$ ) Sensitivity, 47% Specificity, 93.8% $d' = 1.46$	69% Correct ( $P = .06$ ) Sensitivity, 34% Specificity, 100% $d' = 1.58$

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

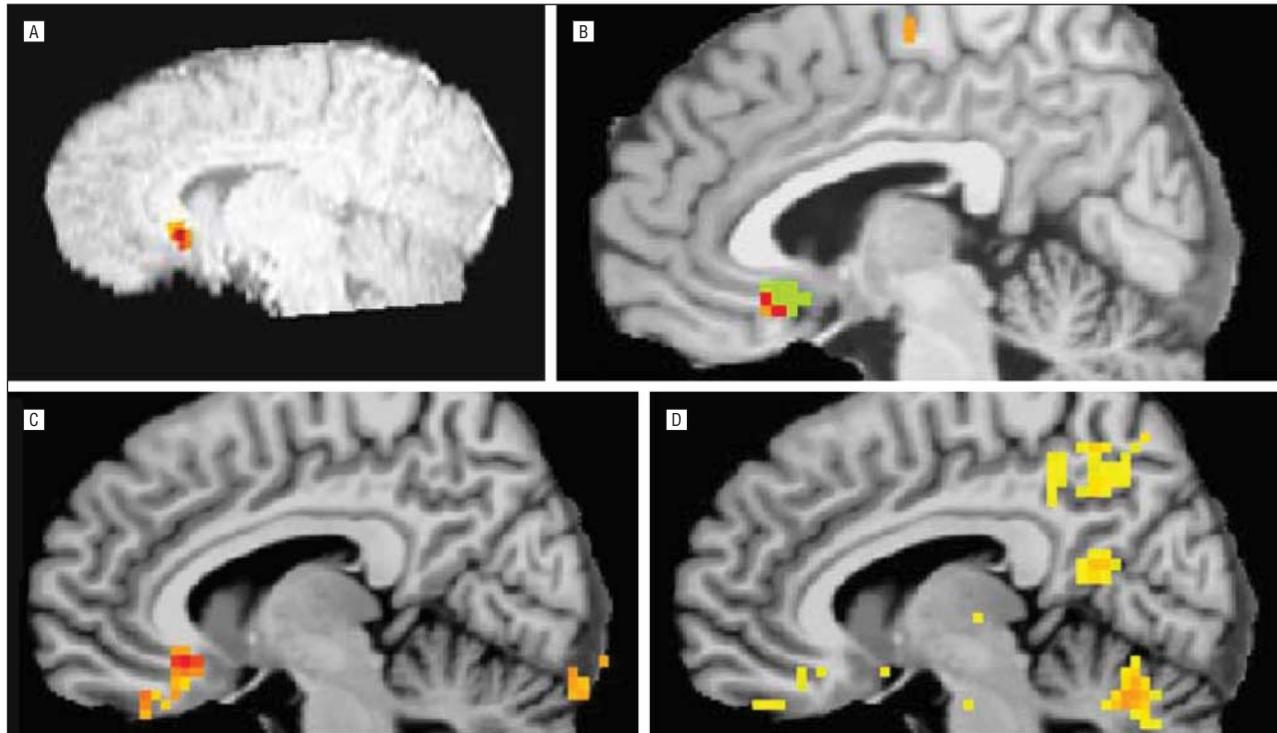
of residual symptomatology (eMaterial XIII). Thus, prediction was specific to negative words.

**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<sub>residual</sub>, 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 XI-C) 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.

#### 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 ( $t_{26} = 1.36$  [ $P = .19$ ]; 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;  $r_{HRSD} = -0.10$  [ $P = .50$ ];  $r_{BDI} = -0.21$  [ $P = .13$ ]).

**Symptom Change Scores.** Associations of sgACC <sub>$z$</sub>  with individually interpretable BDI<sub>posttest-pretest</sub> were nearly identical to associations of percentage change with BDI<sub>residual</sub> (Figure 2D and eMaterials VIII and X). For example, in cohort 1, treatment completers' sgACC <sub>$z$</sub>  predicted 54%



**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.<sup>4</sup> B, The new cognitive therapy samples (n=40;  $R^2 > 0.40$  [ $P < .001$ ]). 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 ( $R^2 = 0.43-0.79$  [ $P < .01$ ]) and 2 ( $R^2 = 0.26$  [ $P < .01$ ]), respectively.

**Table 3. Prediction of Residual Symptoms Using an A Priori Network With Combined Cohort as a Sample of Interest<sup>a</sup>**

Relationship	Measure	Region	Statistic
Zero-order relationships, $P < .05$ for all comparisons <sup>b</sup>	BDI	sgACC	$R^2 = 0.29$
		Right amygdala	$R^2 = 0.16$
		Left DLPFC	$R^2 = 0.20$
		BA24 in the VMPFC	$R^2 = 0.11$
Multivariate relationships, full model $R^2 = 0.43$ ; $F_{1,38} = 7.39$ ( $P < .001$ ) <sup>c</sup>	BDI	Constant	( $P = .21$ )
		sgACC	st $\beta = 0.47$ ( $P < .001$ )
		Right amygdala	st $\beta = 0.24$ ( $P = .10$ )
		Left DLPFC	st $\beta = 0.30$ ( $P = .05$ )
		BA24 in the VMPFC	st $\beta = -0.10$ ( $P = .52$ )

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

<sup>a</sup>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).

<sup>b</sup>See eMaterial VIII and eTable S2A for complete statistics.

<sup>c</sup>See eMaterial VIII and eTable S2A for complete statistics.

of the variance in  $BDI_{\text{posttest-prettest}}$  ( $F_{1,15} = 16.32$  [ $P = .001$ ]), and  $R^2 = 0.29$  in the combined sample ( $F_{1,42} = 16.87$  [ $P < .001$ ]). Cutoffs for sgACC<sub>z</sub> 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 ( $r_{\text{pretreatment,posttreatment}} = 0.39$ ;  $F_{1,39} = 6.82$  [ $P = .01$ ]) (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-

**Table 4. Generalizable Classification From Just sgACC or All 4 Regions and Their Connectivity Using Random Forest Methodology With Cohort 1 as the Training Set and Cohort 2 as the Test Set<sup>a</sup>**

Cohort	Measure	Residual	% Correct, Permutation P Value	
			Response Prediction	Remission Prediction
1, CT: training set, all brain regions	BDI	$R^2 = 0.28 (P < .02)$	75% Correct Sensitivity, 75% Specificity, 90% $d' = 1.28$ ( $P = .02$ )	69% Correct Sensitivity, 62% Specificity, 75% $d' = .99 (P = .06)$
	HRSD	$R^2 = 0.21 (P < .01)$	76% Correct Sensitivity, 92% Specificity, 40% $d' = 1.13$ ( $P = .03$ )	65% Correct Sensitivity, 57% Specificity, 70% $d' = 0.7 (P = .12)$
2, CT: generalization set, initial severity alone	BDI	$R^2 = 0 (P < .74)$	48% Correct Sensitivity, 53% Specificity, 40% $d' = -0.18$ ( $P = .63$ )	38% Correct Sensitivity, 23% Specificity, 54% $d' = -0.64$ ( $P = .92$ )
	HRSD	$R^2 = 0 (P = .63)$	56% Correct Sensitivity, 60% Specificity, 50% $d' = 0.25$ ( $P = .29$ )	47% Correct Sensitivity, 38% Specificity, 53% $d' = -0.23$ ( $P = .7$ )
2, CT: generalization set, initial severity + sgACC	BDI	$R^2 = 0 (P = .07)$	74% Correct Sensitivity, 100% Specificity, 30% $d' = 1.04$ ( $P = .03$ )	78% Correct Sensitivity, 86% Specificity, 69% $d' = 1.57$ ( $P = .003$ )
	HRSD	$R^2 = 0.02 (P = .07)$	81% Correct Sensitivity, 95% Specificity, 58% $d' = 1.86$ ( $P < .001$ )	62% Correct Sensitivity, 69% Specificity, 58% $d' = 0.7 (P = .08)$
2, CT: generalization set, initial severity + all brain regions, retained: baseline BDI, left DLPFC, sgACC, $\phi_{L\ DLPFC, sgACC}$	BDI	$R^2 = 0.15 (P < .003)$	81% Correct Sensitivity, 100% Specificity, 44% $d' = 1.43$ ( $P < .001$ )	63% Correct Sensitivity, 79% Specificity, 46% $d' = 0.7 (P = .09)$
	HRSD	$R^2 = 0.07 (P = .02)$ , (points off) = 3.8	75% Correct Sensitivity, 95% Specificity, 42% $d' = 1.43$ ( $P = .002$ )	66% Correct Sensitivity, 85% Specificity, 53% $d' = 1.09$ ( $P = .02$ )

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

<sup>a</sup> 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.

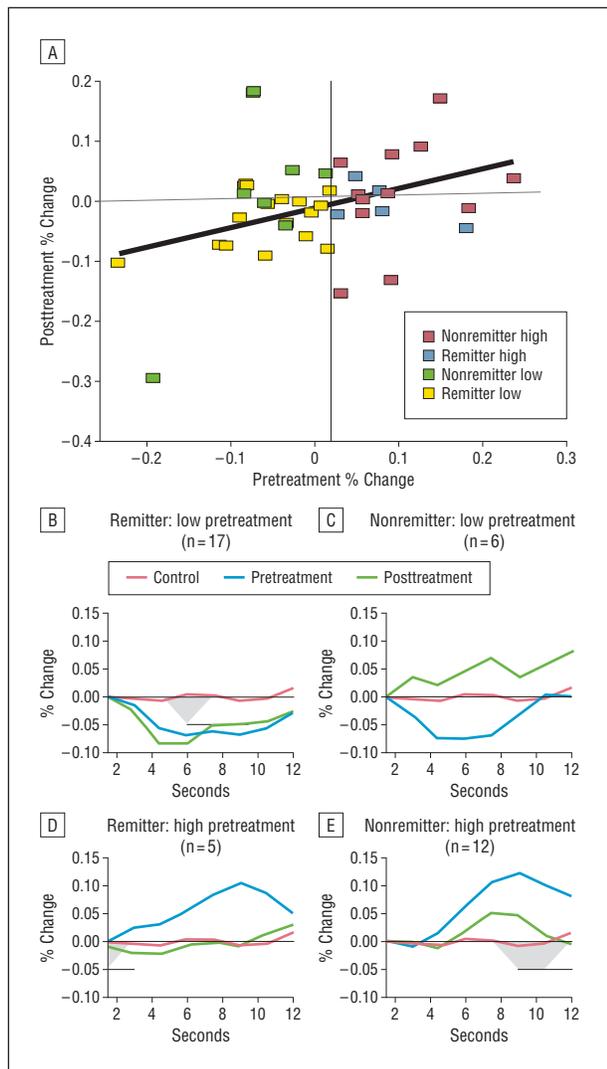
tivity increased (Figure 4A, purple squares). Thus, the emerging picture is that participants with high posttreatment sgACC activity did not experience remission.

#### COMMENT

This study replicates and extends previous results,<sup>4,5</sup> 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 func-

tion 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 to posttreatment measurements.

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



**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<sup>29</sup> or limbic monitoring, possibly via corticocortical connectivity, or possibly because of automatic downregulation of emotion<sup>44</sup> preventing the use of considered reappraisal emphasized in CT. These interpretations support the continued development of interventions that decrease sgACC activity<sup>32</sup> 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, BDI prediction), 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,<sup>12,45,46</sup> fMRI reactivity more analogous to the current design,<sup>13,47-51</sup> and electroencephalography.<sup>52</sup> Thus, ventral cingulate activity could, on prospective replication, yield an algorithm for selection into CT vs antidepressant treatment.<sup>17</sup> Assessing sgACC function during multimodal assessments could also help to validate mechanisms for less interpretable prognostic indicators, such as psychophysiological assessments.<sup>53</sup> 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,<sup>4</sup> 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-

dividual is likely to respond to a specific validated treatment 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 randomized 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 Psychiatric 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 Company, Eli Lilly and Company, GlaxoSmithKline, Janssen Pharmaceutica, MedAvante Inc, Merck Inc, Neuronetics Inc, Novartis Inc, Otsuka Inc, PamLab Inc, Pfizer Inc, Pharmed Neuroboost Inc, Shire US Inc, Supernus Pharmaceuticals, and Wyeth Pharmaceuticals. In 2009 and 2010, Dr Thase had speakers bureau affiliations with AstraZeneca, Bristol-Myers Squibb Company, Eli Lilly and Company, Merck Inc, and Wyeth Pharmaceuticals. Dr Thase has equity holdings in MedAvante Inc; has received royalties from American Psychiatric Publishing Inc, Guilford Publications, and Herald House; and has a family relationship with senior staff at Embryon Inc (formerly Cardinal Health and Advogent). 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, Pharmed Neuroboost, 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 Systems, Indevus, AstraZeneca, Bristol-Myers Squibb, Pfizer, sanofi-aventis, Wyeth-Ayerst, Cyberonics, Novartis, North-Star/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 design and conduct of the study; in the collection, analysis, 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 Congress 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 intrepid commu-

nity therapists who worked on the study; and our patients, without whom none of this work would have been possible.

## REFERENCES

1. Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. New York, NY: Guilford Press; 1979.
2. Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychol Sci Public Interest*. 2002;3(2):39-77.
3. Kemp AH, Gordon E, Rush AJ, Williams LM. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, 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 depression with cognitive behavior therapy. *Am J Psychiatry*. 2006;163(4):735-738.
5. Fu CH, Williams SC, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, Donaldson C, Suckling J, Andrew C, Steiner H, Murray RM. Neural responses to sad facial expressions in major depression following cognitive behavioral therapy. *Biol Psychiatry*. 2008;64(6):505-512.
6. Bruder GE, Stewart JW, Mercier MA, Agosti V, Leite P, Donovan S, Quitkin FM. Outcome of cognitive-behavioral therapy for depression: relation to hemispheric dominance for verbal processing. *J Abnorm Psychol*. 1997;106(1):138-144.
7. Mayberg HS. Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. *Br Med Bull*. 2003;65:193-207.
8. Mayberg HS, Liotti M, Brannan SK, McGinnis S, Mahurin RK, Jerabek PA, Silva JA, Tekell JL, Martin CC, Lancaster JL, Fox PT. Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am J Psychiatry*. 1999;156(5):675-682.
9. Drevets WC. Neuroimaging studies of mood disorders. *Biol Psychiatry*. 2000;48(8):813-829.
10. Fu CH, Williams SC, Cleare AJ, Brammer MJ, Walsh ND, Kim J, Andrew CM, Pich EM, Williams PM, Reed LJ, Mitterschiffthaler MT, Suckling J, Bullmore ET. Attenuation of the neural response to sad faces in major depression by antidepressant treatment: a prospective, event-related functional magnetic resonance imaging study. *Arch Gen Psychiatry*. 2004;61(9):877-889.
11. Goldapple K, Segal Z, Garson C, Lau M, Bieling P, Kennedy S, Mayberg H. Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Arch Gen Psychiatry*. 2004;61(1):34-41.
12. Mayberg HS, Brannan SK, Mahurin RK, Jerabek PA, Brickman JS, Tekell JL, Silva JA, McGinnis S, Glass TG, Martin CC, Fox PT. Cingulate function in depression: a potential predictor of treatment response. *Neuroreport*. 1997;8(4):1057-1061.
13. Keedwell PA, Drapier D, Surguladze S, Giampietro V, Brammer M, Phillips M. Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. *J Affect Disord*. 2010;120(1-3):120-125.
14. Bryant RA, Felmingham K, Kemp A, Das P, Hughes G, Peduto A, Williams L. Amygdala and ventral anterior cingulate activation predicts treatment response to cognitive behaviour therapy for post-traumatic stress disorder. *Psychol Med*. 2008;38(4):555-561.
15. Vul E, Harris C, Winkelman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci*. 2009;4(3):274-290.
16. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*. 2009;12(5):535-540.
17. DeRubeis RJ, Siegle GJ, Hollon SD. Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nat Rev Neurosci*. 2008;9(10):788-796.
18. Canli T, Cooney RE, Goldin P, Shah M, Sivers H, Thomason ME, Whitfield-Gabrieli S, Gabrieli JD, Gotlib IH. Amygdala reactivity to emotional faces predicts improvement in major depression. *Neuroreport*. 2005;16(12):1267-1270.
19. Davidson RJ. Asymmetric brain function, affective style and psychopathology: the role of early experience and plasticity. *Dev Psychopathol*. 1994;6:741-758.
20. Siegle GJ, Thompson W, Carter CS, Steinhauer SR, Thase ME. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: related and independent features. *Biol Psychiatry*. 2007;61(2):198-209.
21. Pizzagalli DA, Peccoralo LA, Davidson RJ, Cohen JD. Resting anterior cingulate activity and abnormal responses to errors in subjects with elevated depressive symptoms: a 128-channel EEG study. *Hum Brain Mapp*. 2006;27(3):185-201.
22. Lemogne C, le Bastard G, Mayberg H, Volle E, Bergouignan L, Lehericy S, Al-

- lilaire JF, Fossati P. In search of the depressive self: extended medial prefrontal network during self-referential processing in major depression. *Soc Cogn Affect Neurosci*. 2009;4(3):305-312.
23. Dichter GS, Felder JN, Smoski MJ. Affective context interferes with cognitive control in unipolar depression: an fMRI investigation. *J Affect Disord*. 2009;114(1-3):131-142.
  24. Davidson RJ. Affective neuroscience and psychophysiology: toward a synthesis. *Psychophysiology*. 2003;40(5):655-665.
  25. Anand A, Li Y, Wang Y, Wu J, Gao S, Bukhari L, Mathews VP, Kalnin A, Lowe MJ. Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study. *Biol Psychiatry*. 2005;57(10):1079-1088.
  26. Cooney RE, Joormann J, Atlas LY, Eugène F, Gotlib IH. Remembering the good times: neural correlates of affect regulation. *Neuroreport*. 2007;18(17):1771-1774.
  27. Campbell-Sills L, Simmons AN, Lovero KL, Rochlin AA, Paulus MP, Stein MB. Functioning of neural systems supporting emotion regulation in anxiety-prone individuals. *Neuroimage*. 2011;54(1):689-696.
  28. Drevets WC, Savitz J, Trimble M. The subgenual anterior cingulate cortex in mood disorders. *CNS Spectr*. 2008;13(8):663-681.
  29. Hamilton JP, Glover GH, Hsu JJ, Johnson RF, Gotlib IH. Modulation of subgenual anterior cingulate cortex activity with real-time neurofeedback. *Hum Brain Mapp*. 2011;32(1):22-31.
  30. Liotti M, Mayberg HS, McGinnis S, Brannan SL, Jerabek P. Unmasking disease-specific cerebral blood flow abnormalities: mood challenge in patients with remitted unipolar depression. *Am J Psychiatry*. 2002;159(11):1830-1840.
  31. Matthews S, Simmons A, Strigo I, Gianaros P, Yang T, Paulus M. Inhibition-related activity in subgenual cingulate is associated with symptom severity in major depression. *Psychiatry Res*. 2009;172(1):1-6.
  32. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C, Schwab JM, Kennedy SH. Deep brain stimulation for treatment-resistant depression. *Neuron*. 2005;45(5):651-660.
  33. Siegle GJ. Beyond depression commentary: wherefore art thou, depression clinic of tomorrow? *Clin Psychol Sci Pract*. 2011;18(4):305-310.
  34. First MB, Spitzer RL, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition*. Vol 20. New York: Biometrics Research Dept, New York State Psychiatric Institute; 1996.
  35. Young J, Beck AT. *Cognitive Therapy Scale: Rating Manual*. Philadelphia, PA: Center for Cognitive Therapy; 1980.
  36. Nelson HE, Willison J. *The National Adult Reading Test*. Berkshire, England: Nfer-Nelson Publishing Co; 1991.
  37. Beck AT, Steer RA, Brown G. *Beck Depression Inventory*. 2nd ed. San Antonio, TX: Psychological Corp; 1996.
  38. Bradley MM, Lang PJ. *Affective Norms for English Words ANEW Technical Manual and Affective Ratings*. Gainesville: Center for Research in Psychophysiology, University of Florida; 1997.
  39. Siegle GJ, Granholm E, Ingram RE, Matt GE. Pupillary and reaction time measures of sustained processing of negative information in depression. *Biol Psychiatry*. 2001;49(7):624-636.
  40. Siegle GJ, Steinhauer SR, Carter CS, Rameil W, Thase ME. Do the seconds turn into hours? relationships between sustained pupil dilation in response to emotional information and self reported rumination. *Cognit Ther Res*. 2003;27(3):365-383.
  41. Siegle GJ, Steinhauer SR, Thase ME, Stenger VA, Carter CS. Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biol Psychiatry*. 2002;51(9):693-707.
  42. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324.
  43. Zhou D, Thompson WK, Siegle G. MATLAB toolbox for functional connectivity. *Neuroimage*. 2009;47(4):1590-1607.
  44. Phillips ML, Ladouceur CD, Drevets WC. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol Psychiatry*. 2008;13(9):829, 833-857.
  45. Brannan SK, Mayberg HS, McGinnis S, Silva JA, Tekell J, Mahurin RK, Jerabek PA, Fox PT. Cingulate metabolism predicts treatment response: a replication [abstract]. *Biol Psychiatry*. 2000;47:107S.
  46. Brody AL, Saxena S, Mandelkern MA, Fairbanks LA, Ho ML, Baxter LR. Brain metabolic changes associated with symptom factor improvement in major depressive disorder. *Biol Psychiatry*. 2001;50(3):171-178.
  47. Davidson RJ, Irwin W, Anderle MJ, Kalin NH. The neural substrates of affective processing in depressed patients treated with venlafaxine. *Am J Psychiatry*. 2003;160(1):64-75.
  48. Chen CH, Ridler K, Suckling J, Williams S, Fu CH, Merlo-Pich E, Bullmore E. Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biol Psychiatry*. 2007;62(5):407-414.
  49. Nitschke JB, Sarinopoulos I, Oathes DJ, Johnstone T, Whalen PJ, Davidson RJ, Kalin NH. Anticipatory activation in the amygdala and anterior cingulate in generalized anxiety disorder and prediction of treatment response. *Am J Psychiatry*. 2009;166(3):302-310.
  50. Langenecker SA, Kennedy SE, Guidotti LM, Briceno EM, Own LS, Hooven T, Young EA, Akil H, Noll DC, Zubieta JK. Frontal and limbic activation during inhibitory control predicts treatment response in major depressive disorder. *Biol Psychiatry*. 2007;62(11):1272-1280.
  51. Roy M, Harvey PD, Berlim MT, Mamdani F, Beaulieu MM, Turecki G, Lepage M. Medial prefrontal cortex activity during memory encoding of pictures and its relation to symptomatic improvement after citalopram treatment in patients with major depression. *J Psychiatry Neurosci*. 2010;35(3):152-162.
  52. Pizzagalli D, Pascual-Marqui RD, Nitschke JB, Oakes TR, Larson CL, Abercrombie HC, Schaefer SM, Koger JV, Benca RM, Davidson RJ. Anterior cingulate activity as a predictor of degree of treatment response in major depression: evidence from brain electrical tomography analysis. *Am J Psychiatry*. 2001;158(3):405-415.
  53. Siegle GJ, Steinhauer SR, Friedman ES, Thompson WS, Thase ME. Remission prognosis for cognitive therapy for recurrent depression using the pupil: utility and neural correlates. *Biol Psychiatry*. 2011;69(8):726-733.