Remission Prognosis for Cognitive Therapy for Recurrent Depression Using the Pupil: Utility and Neural Correlates

Greg J. Siegle, Stuart R. Steinhauer, Edward S. Friedman, Wesley S. Thompson, and Michael E. Thase

**Background:** Although up to 60% of people with major depressive disorder respond to cognitive therapy (CT) in controlled trials, clinicians do not routinely use standardized assessments to inform which patients should receive this treatment. Inexpensive, noninvasive prognostic indicators could aid in matching patients with appropriate treatments. Pupillary response to emotional information is an excellent candidate, reflecting limbic reactivity and executive control. This study examined 1) whether pretreatment assessment of pupillary responses to negative information were associated with remission in CT and 2) their associated brain mechanisms.

**Methods:** We examined whether pretreatment pupillary responses to emotional stimuli were prognostic for remission in an inception cohort of 32 unipolar depressed adults to 16 to 20 sessions of CT. Twenty patients were then assessed on the same task using functional magnetic resonance imaging. Pupillary responses were assessed in 51 never-depressed controls for reference.

**Results:** Remission was associated with either low initial severity or the combination of higher initial severity and low sustained pupillary responses to negative words (87% correct classification of remitters and nonremitters, 93% sensitivity, 80% specificity: 88% correct classification of high-severity participants, \( p < .01 \), 90% sensitivity, 92% specificity). Increased pupillary responses were associated with increased activity in dorsolateral prefrontal regions associated with executive control and emotion regulation.

**Conclusions:** For patients with higher severity, disruptions of executive control mechanisms responsible for initiating emotion regulation, which are indexed by low sustained pupil responses and targeted in therapy, may be key to remitting in this intervention. These mechanisms can be measured using inexpensive noninvasive psychophysiological assessments.

**Key Words:** Affect, cognitive therapy, depression, emotion, psychophysiology, psychotherapy, pupillometry, remission

The most effective treatments for major depressive disorder (MDD), including cognitive therapy (CT) (1), yield response among 40% to 60% of depressed patients (2). Patients with low severity are likely to benefit (3–5). Reliable methods to determine which moderately severe patients will most strongly benefit from CT do not exist, leading to delivery of unwarranted treatments, long times to remission, and a high societal toll. Identifying which patients will remit in CT could decrease suffering, health care burden, and costs by improving the precision of referrals. Thus, we examined prognosis for remission in depressed patients entering CT using a quick, inexpensive physiologic measure.

CT involves, among other skills, learning to interrupt automatic sustained emotional processing (the “downward spiral” of automatic negative thoughts and dysphoric affects). Neuroimaging studies demonstrate associations of symptom change in CT with sustained limbic activity, potentially reflecting automatic emotional engagement, and decreased prefrontal function, often associated with decreased emotion regulation, with recovery (6–8). Potentially, CT helps depressed participants address neurocognitive features that most strongly characterize their depression. However, neuroimaging is lengthy, cumbersome, and expensive and therefore unlikely to be employed clinically in the near future. Valid, brief, and inexpensive psychophysiological proxies for such measures would thus be useful. Although many behavioral and psychophysiological prognostic indicators of symptom change have been observed, including heart rate variability (9,10), electroencephalographic asymmetry (11,12), and dichotic listening performance (13), none have been routinely adopted clinically. Potential reasons include lack of links to psychologic processes (i.e., face validity) and having not been examined in patients with at least moderate severity in whom their effort would be more strongly warranted. If true, it would be useful to show that a quick, inexpensive psychophysiological measure is 1) prognostic among more severe participants, 2) intuitively reflects psychologic processes, and 3) reflects expected brain mechanisms.

Accordingly, our goal was to associate clinical remission (arguably, the standard patients would use for engaging in a treatment) with an intuitive, validated, easily acquired physiologic measure with previously demonstrated differences between depressed and healthy individuals. Thus, we examined whether pupillary responses to negative words are prognostic for remission in CT in depressed individuals with moderate or higher levels of symptomatology. We chose pupillary responses because 1) they can be measured in depressed individuals. Pupillary responses to emotional words are increased and sustained in depressed compared with healthy individuals (14,15), potentially reflecting increased limbic engagement or increased need for regulatory control. 2) Furthermore, pupillary responses reflect cognitive and emotional processing. Pupils dilate in response to cognitive load (16) and in response to emotional stimuli (17) and remain dilated throughout expenditure of cognitive load (18). To assess both processes, we used a task in which trials alternated between demands previously associated with reliable pupillary responses, including attention to emotional...
features of stimuli (emotion naming) (14), and executive control (digit sorting) (19). Finally, pupillary responses reflect relevant brain mechanisms. Responses such as dilation are well known to reflect both limbic activity (20) and mechanisms of cognitive control (19). That said, as with other unidimensional psychophysioligic measures, they are unable to distinguish between different neural inputs, thus not allowing directional inferences regarding association with constructs such as reactivity and regulation.

We hypothesized that if pupil dilation following a negative word primarily reflects limbic hyperactivity, it would be positively associated with remission because it would indicate that specific abnormalities addressed in CT were present. Alternatively, if sustained pupil dilation is more reflective of prefrontal regulation, decreased sustained pupil dilation may be associated with better remission, again because CT is presumed to address the underlying mechanism of deficient regulatory control. Thus, as a secondary and more exploratory goal, limbic and prefrontal brain mechanisms were examined in a subset of participants who completed the same tasks again during functional magnetic resonance imaging to aid interpretability. To understand whether patterns associated with remission were consistent with decreased initial disease mechanism load (psychophysioligic reactivity more like healthy individuals compared with nonremitters) or increased presence of the types of mechanisms addressed in treatment, secondary analyses compared remitters and nonremitters to healthy, never-depressed participants.

Methods and Materials

Participants

Participants included 36 outpatients with recurrent MDD (via Structured Clinical Interview for DSM (21) from a CT trial (Thase, principal investigator; http://ClinicalTrials.gov: NCT00183664), and 53 controls with no history of Axis I disorder. Participants reported no significant eye problems, interfering health problems, psychoactive drug or alcohol abuse within the past 6 months (one control participant’s alcohol use was ambiguous, and the individual could not be recontacted), antidepressant use within 2 weeks (6 for fluoxetine), history of psychosis, or manic episodes. Participants passed a participant’s alcohol use was ambiguous, and the individual could not be recontacted), antidepressant use within 2 weeks (6 for fluoxetine), history of psychosis, or manic episodes. Participants passed a cognitively screen (North American Adult Reading Test—II Verbal Intelligence Quotient > 80) (22). Figure S1 in Supplement 1 shows patient flow, yielding 51 control and 32 depressed participants who received CT. Of these participants, 20 completed functional magnetic resonance imaging (fMRI) assessment as part of a later-funded protocol.

Ethics

This study and the parent trial were approved by the University of Pittsburgh Institutional Review Board (IRB). This study was also approved by the VA Pittsburgh IRB. After being fully informed regarding the nature of the study, all participants gave written consent to participate in the study by signing University of Pittsburgh and, if necessary, VA Pittsburgh IRB-approved consent forms.

Procedures

As in the Consolidated Standards of Reporting Trials (CONSORT) Diagram (Figure S1 in Supplement 1), after signing IRB-approved consent forms, a diagnostic interview, and word-list generation, participants completed the Beck Depression Inventory II (BDI-II) to assess depressive severity on the testing day and information processing tasks during pupillary assessment (details in Supplement 1, Data—Method: Apparatus). A digit sorting task (repeatedly putting digits in numerical order) (19) to decrease novelty effects was followed by other tasks in a counterbalanced order. Eligible patients were assessed again with fMRI (34 or 30 3.2-mm slices acquired parallel to the anterior commissure–posterior commissure line (3T GE scanner, T2*-weighted images, reverse spiral pulse sequence to increase signal strength in the amygdala and orbitofrontal regions; repetition time = 1500 msec, echo time = 5 msec, field of view = 24 cm, flip = 60°) within 2 weeks.

The analyzed task involved 60 27-sec compound emotion-identification/digit sorting trials as shown in Figure 1. Emotion-identification consisted of an instruction (“What’s the emotion?”) 1 sec), fixation mask (1 sec), word (200 msec; 10 positive, 10 negative, and 10 neutral personally relevant and 10 positive, 10 negative, and 10 neutral-normed words (14, 15) (see also Supplement 1, Data—Method: Word List Generation) presented in a differing pseudo-random order for each participant, and mask (8.5 sec) during which participants were instructed to name the emotionality of the word using buttons for “Positive” (+), “Negative” (−), or “Neutral” (N) as quickly and accurately as they could. The digit-sorting portion contained a cue (“Order the numbers”; 1 sec), mask (1 sec), five digits (2 sec), mask (5 sec), and target digit from the set (7.5 sec). Participants were told that when the digits appeared, they should read them from left to right, put them in numerical order in memory, and remember the middle digit in the sorted list; when the target appeared, they should push buttons for “yes” (Y) or “no” (N) to indicate whether the target was the middle digit from the previous set, as quickly and accurately as they could. Participants were told they should not use special strategies such as sorting the digits by moving their eyes on the screen or sorting only the first few digits because we were examining the process of sorting items in memory. Pupil dilation and prefrontal activity were parametrically associated with the number of digits to be sorted on the digit sorting task in healthy participants from this sample (19,23).

Depressed participants then received 16 sessions (20 for those with Hamilton Depression Rating Scale reduction < 40% at Session 9) of CT following Beck et al.’s (1) guidelines. At each session, participants completed the BDI to assess depressive severity. (The BDI-I was used at clinical assessments as part of the larger associated clinical protocol. The BDI-II was used at physiologic testing as part of the physiologic protocol and was thus used only to separate participants into high and low severity groups at the assessment day. The protocols were linked post hoc for the current analysis. In our de-
pressed sample, BDI-I could be estimated as 4.7 + .83 * BDI-II with high reliability, r = .78.)

Data Selection and Analysis

Clinical Status. Final severity scores (BDI) were obtained at patients’ blind evaluation, approximately 1 to 2 weeks following their last treatment session. For patients missing the blind evaluation, severity from their last-attended appointment was used, allowing a last-observation-carried-forward analysis. Remission was defined as having a final depressive severity of BDI < 10 and no MDD diagnosis following acute treatment. Although there is no agreed-on threshold for remission (24) based on the BDI, criteria have ranged from greater than 9 (25) to less than 12 (26); our criterion was weighted toward specificity as the probable use of this measure is likely to be in deciding to opt out of rather than into therapy.

Pupillary Data. Data were cleaned using our laboratory’s standard procedures (23) including interpolation through blinks and removal of trials with over 50% blinks, reaction times less than 150 msec, and incorrect digit sorting responses (Md = 6, M[SD] = 9.6 [10.7] trials eliminated per subject), and subtraction of a 100-msec prestimulus baseline mean to produce a response index. Data from the second pupillometer were scaled and offset to match the controls (classification only improved when this procedure was not implemented).

Because we were interested primarily in sustained pupillary responses, conventional statistics such as peak response were not appropriate. Rather, contrasts were examined via statistical tests at each sample along pupillary waveforms. To control Type I error, 35 consecutive samples (.56 sec) of tests in a row significant at p < .1, considered replications, were considered significant at p = .05 (27,23) (see also Supplement 1, Data: Method: Type I Error Control). Results report tests of the mean pupillary response in significant windows.

fMRI. The fMRI data were preprocessed via motion and outlier correction, linear trend removal, temporal smoothing (5-point filter), linear cross-registration to a reference brain, and spatial smoothing (6 mm full width at half maximum). Relationships of brain activity to pupillary response were examined in regions of interest previously implicated in sustained processing of emotion in depression (28) and treatment outcome, including the left dorsolateral prefrontal cortex (DLPFC) that differentiates depressed and control participants on digit sorting (28), a right DLPFC region from the same data set that reflected parametric activation with task difficulty in both depressed and control participants on that task via conjunction analysis (within group analysis of variance thresholded at p < .05 for each component map, yielding p < .0025 for the conjunction, thresholded using an empirically determined contingency threshold for this threshold via the Analysis of Functional Neurimages [AFNI] AlphaSim (29]), the prognostic Brodmann’s area (BA) 25 region previously examined in a subset of this sample (6), a hand-traced left amygdala region as in Siegle et al. (28), and an anatomically defined regions of interest from AFNI’s Talairach Atlas for BA 24 in the rostral cingulate gyrus, which has been consistently associated with outcome in other studies (reviewed in DeRubeis [30]). Raw mean condition-related time courses represented as percent change from the first scan of a trial were extracted.

We used partial least squares to capture modes of variation in the fMRI signal associated with maximally discriminant pupillary motility (11.78 sec) across depressed participants (n = 20 had fMRI data). Subject was a random factor, and each brain region was examined separately.

Results

Sociodemographic and Behavioral Data

Sociodemographic and behavioral data are presented in Table 1. Participants rated positive words as positive, negative words as negative, and neutral words as neither positive nor negative. Performance was excellent on digit-sorting (means > 90% correct for both groups following all types of words). Depressed participants rated positive words as less positive than control participants and were slower to name the valence of positive words. No other reliable group or condition-related differences in behavioral data (ratings, reaction times to valence identification, percent correct on digit-sorting) were detected.

Clinical Change and Initial Severity

Depressive symptoms reliably decreased following therapy, BDI M[SD]pre-CT = 29.6 (12.9), BDI M[SD]post-CT = 11.7 (11.9), t(31) = 8.87, p < .005, standardized difference (Cohen’s d) = 1.6. Seventeen patients met remission criteria; the remaining 15 patients were nonremitters. Logistic regression of initial severity on remission suggested, as in previous studies, that higher pretreatment severity was associated with a lower likelihood of remission (β = −0.1, R² = .41, p < .0005). A standard high-end pretreatment BDI cut score of 20 (31–34) yielded 59% correct classification. For participants with low initial severity (BDI ≤ 20), there was one nonremitter and six remitters. Thus, most participants with low levels of initial severity remitted in CT. To aid prognosis of clinical remission for more severe participants, subsequent pupil analyses concentrated on discriminating remitters with high severity scores on pupil assessment day (BDI-II > 20; henceforth “high pretreatment severity”; n = 25; 14 nonremitters and 11 remitters) from nonremitters with severity scores on pupil assessment day below this threshold (n = 7; one nonremitter; six remitters—similar to the use of pretreatment BDI scores).

Associations of Pupillary Responses to Negative Stimuli with Remission in CT

Figure 2 shows pupillary responses to negative words followed by digit-sorting on the alternating task for control participants, and for high-pretreatment-severity remitters and nonremitters. High-pretreatment-severity remitters had lower pupil dilation in response to emotional words than the low-pretreatment-severity participants in multiple significant temporal windows: .82 to 1.93 sec: t(23) = 2.43, p = .02, D = .04 mm, D = .98, 11.15 to 12.83 sec: t(23) = 2.03, p = .05, D = .10 mm, D = .82, 13.02 to 15.55 sec: t(23) = 2.14, p = .04, D = .10 mm, D = .86, 19.80 to 20.83 sec: t(23) = 2.10, p = .05, D = .10 mm, D = .84, 26.32 to 27.00 sec: t(23) = 2.43, p = .02, D = .07 mm, D = .98.

Maximal discrimination between high-pretreatment-severity remitters and non-remitters was achieved in multiple intervals from 11.5 to 12.25 sec after trial onset (9.5–10.25 sec after stimulus onset), yielding 88% correct classification, p < .0005 via permutation tests. Figure 3 shows a scatterplot in which final severity is on the x axis and the representative median pupillary prediction at 11.78 sec is on the y axis. As shown in Figure 3, of the high initial-severity participants, 9 of 11 remitters had pupillary responses less than −.05 (82% sensitivity), whereas 13 of 14 nonremitters had pupillary responses greater than −.05 (93% specificity). Monte Carlo simulations accounting for the number of employed comparisons and temporal smoothness of relevant pupillary waveforms (r = .998; 150 iterations) further suggested that prognosis associated with pupillary responses was better than expected by chance, for which a classification rate of 88% correct was never achieved (p < .007).

www.sobp.org/journal
and for which a classification rate of .8 was achieved less than 95% of the time. Leave-one-out cross-validation using the same method obtained similar intervals and thresholds for all participants (11.7–12.2 sec, thresholds between −.09 and −.05) with a 76% correct classification rate, suggesting that results are relatively robust. To demonstrate robustness further, when participants were further restricted to those who displayed high severity at both the initial therapy and pupil-assessment days (BDI-I > 19, BDI-II > 20; n = 22), leave-one-out cross-validation yielded 86% correct classification.

Sensitivity Analyses: Alternate Remission Criteria, Continuous Severity, and Effects of Valence, Personal Relevance, Rumination, and Pupilometer

As described in Supplement 1 (Data), when response was defined using a variety of alternate outcome measures statistics were

---

**Table 1. Sociodemographic, Clinical, and Behavioral Data**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Depressed</th>
<th>Control</th>
<th>Significant Difference if Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>32</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>No. Male</td>
<td>15</td>
<td>15</td>
<td>ns</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>25 white, 7 African American</td>
<td>39 white, 5 African American, 3 Asian, 4 biracial</td>
<td>ns</td>
</tr>
<tr>
<td>Age Range</td>
<td>20–56</td>
<td>19–55</td>
<td></td>
</tr>
<tr>
<td>Age, M (SD)</td>
<td>39.84 (10.0)</td>
<td>31.7 (10.8)</td>
<td>t(81) = −3.4, p = .001</td>
</tr>
<tr>
<td>Education, M (SD)</td>
<td>14.94 (3.1)</td>
<td>16.1 (2.7)</td>
<td>t(81) = 1.82, p = .07</td>
</tr>
<tr>
<td>Severity, M (SD)b</td>
<td>29.5 (12.0)</td>
<td>3.6 (4.3)</td>
<td>t(36.1) = −11.71, p &lt; .0005</td>
</tr>
<tr>
<td>No. Depressive Episodes</td>
<td>1 &gt; 10, MD = 4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>NAART VIQ, M (SD)</td>
<td>111.5 (8.5)</td>
<td>109.9 (8.1)</td>
<td>ns</td>
</tr>
<tr>
<td>Postscan Emotion Ratings of Employed Words&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive words, M (SD)</td>
<td>5.68 (.60)</td>
<td>6.12 (.41)</td>
<td>t(79) = 3.89, p &lt; .005, D = .44</td>
</tr>
<tr>
<td>Negative words, M (SD)</td>
<td>2.04 (.59)</td>
<td>2.06 (.45)</td>
<td>ns</td>
</tr>
<tr>
<td>Neutral words, M (SD)</td>
<td>4.05 (.22)</td>
<td>4.15 (.39)</td>
<td>ns</td>
</tr>
<tr>
<td>Reaction Times for Valence Identification in msec&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive words, M (SD)</td>
<td>1723.3 (466.9)</td>
<td>1297.8 (546.7)</td>
<td>t(81) = 3.64, p &lt; .005, D = 425 msec</td>
</tr>
<tr>
<td>Negative words, M (SD)</td>
<td>1643.1 (415.5)</td>
<td>1492.0 (597.1)</td>
<td>ns</td>
</tr>
<tr>
<td>Neutral words, M (SD)</td>
<td>1630.1 (504.7)</td>
<td>1453.8 (558.2)</td>
<td>ns</td>
</tr>
<tr>
<td>Digit-Sorting Following Emotion-Identification Task</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Correct&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive words, M (SD)</td>
<td>.91 (.16)</td>
<td>.90 (.22)</td>
<td></td>
</tr>
<tr>
<td>Negative words, M (SD)</td>
<td>.92 (.14)</td>
<td>.92 (.20)</td>
<td></td>
</tr>
<tr>
<td>Neutral words, M (SD)</td>
<td>.93 (.14)</td>
<td>.92 (.19)</td>
<td></td>
</tr>
</tbody>
</table>

D, Difference; M, mean; MD, median; NAART, North American Adult Reading Test—II; VIQ, Verbal Intelligence Quotient.

<sup>a</sup>One participant was 55 when consented but 56 when tested.

<sup>b</sup>BDI-II Depression Inventory score at pupil assessment.

<sup>c</sup>Valence x group ns, valence main effect ns.

---

**Figure 2.** Pupillary responses to negative words followed by digit sorting on the alternating task for control participants, and both responders and nonresponders with high-initial-severity (Beck Depression Inventory—II [BDI-II] > 20). Significant differences between responders and nonresponders are highlighted below the x axis; yellow = p < .1, red = p < .05. Black lines show regions with enough consecutive tests to be considered significant.

**Figure 3.** A scatterplot showing initial severity on the x axis and pupillary response at 11.78 sec on the y axis for participants with initial Beck Depression Inventory (BDI) greater than 20. Remitters were defined as participants with final estimated BDI less than 10 and no depression diagnosis. The threshold of maximal discrimination between remitters and nonremitters (−.07 mm) is plotted as a blue line.
at least as strong or stronger. When remission was defined as having a final Hamilton Depression Rating Scale score less than 12, among participants with high initial severity, there were 15 remitters and 8 nonremitters; using the same methods described earlier, 88% prognostic accuracy was achieved, and 22 of 25 participants were correctly classified. When the initial-severity cutoff threshold and pupillary response were maximized simultaneously, results were similar: having a score on pupil assessment day below 23 was strongly associated with remission; for those above BDI 23, pupillary responses below —0.07 were significantly prognostic for remission.

As described in Supplement 1 (Data), the same qualitative pattern of differences in pupillary response between nonremitters and remitters was observed for using initial BDI-I scores as well as when neutral words were examined but not for positive words. Remitters with low BDI scores also appeared to display sustained pupillary responses to neutral words. Prognostic power was stronger for the idiosyncratically generated negative words than the normed words. The threshold that maximally discriminated high BDI remitters from nonremitters was similar regardless of pupilometer.

Also, as shown in Supplement 1 (Data), although pupillary responses in the region maximally predictive for remission were not strongly associated with continuous change in symptomatology in the high-initial-symptomatology group or the entire cohort, relationships with pupillary motility in the last part of the trial (~30 sec following stimulus onset) were moderate. We thus examined whether sustained pupillary responses were associated with rumination as we have previously observed (15). As described in Supplement 1, pupillary responses were moderately but not significantly associated with self-reported rumination in the period of maximal discrimination.

**Associations of Pupillary Responses with Brain Activity Measured Using fMRI**

We employed fMRI data to examine whether pupillary responses strongly reflect activity in limbic regions (e.g., amygdala) or regions more strongly associated with regulatory control (e.g., DLPFC) among depressed participants who received treatment. **Figure 4** shows the correlation at each scan within trials from regressions of blood oxygen level–dependent (BOLD) responses to negative words in the a priori regions with pupil responses at 11.78 sec with statistically significant scans highlighted below the y axis, and scatterplots for the scans with the strongest correlations. Because the BOLD response lags neural activity by up to 6 sec whereas pupillary responses lag neural function by only up to 1 sec, BOLD activity at 17 sec would be expected to be associated with pupillary function at 11.78 sec proximally. As shown in the figure, positive zero-order correlations between pupillary responses and brain activity before/during the pupillary response were observed only for putative regulatory regions, the left DLPFC, and BA 24. Partial least-squares regressions jointly accounting for the time course of brain reactivity and its relationship to pupillary reactivity, confirmed that a significant relationship with the predictive aspects of pupil dilation was observed only for the DLPFC \( R^2 = .33 \), adjusted \( R^2 = .25 \), \( F(2,17) = 4.2, p = .03 \); significant modes of variation were positively associated, peaking at 7.5 sec, \( p = .009 \), and peaking at 10.5 and 25.5 secs, \( p = .01 \). As shown in Figure S11 in Supplement 1, zero-order relationships were similar for control participants in the
DLPFC but not for amygdala and BA 25. As shown in Figure S12 in Supplement 1, results were qualitatively similar when only the high BDI depressed participants were included rather than all treated patients except for the negative relationships in the right DLPFC, which was not robust. Finally, as described in the supplementary data, pupillary motility was moderately but not significantly related to functional connectivity across these structures.

Discussion

This experiment examined prognosis for remission in CT using pupillary response during cognitive and emotional information processing. Consistent with previous work, individuals with lower pretreatment depressive severity had better outcomes (3–5), whereas remission was more variable among those with higher pretreatment severity. In this group, increased sustained pupillary responses were associated with lower remission rates. As in healthy individuals (19), brain imaging on our depressed participants suggested that increased pupillary responses were associated with increased dorsolateral prefrontal function, which has been linked to executive control and emotion regulation. In support of this idea, an overlapping subset of depressed participants displayed decreased sustained pupillary responses on an executive control (Stroop) task without explicit emotional content (35). Similarities in prefrontal influences for control and depressed participants could reflect common influences of voluntary regulatory control, and dissimilarities for the amygdala and BA25 could reflect differential recruitment of these structures in depression. If these data are replicable, there are strong practical and theoretical implications.

Implications

Practically, results suggest an algorithm to help patients to decide whether to expend time and resources in treatment with CT. Consistent with other literature, we suggest that patients with low depressive severity are likely remitters and thus may not benefit from further assessment to aid their decision. Among more severely depressed patients, assessment of pupillary responses could provide a quick, inexpensive, sensitive, and specific pretreatment prognostic indicator. Given the strong effects using just 10 idiosyncratic negative words, the assessment could take less than 10 min.

Theoretically, some CT techniques may be conceptualized as fostering the depressed person’s ability to recruit executive control in response to negative stimuli or thoughts. To the extent that sustained pupil dilation reflects prefrontally mediated executive control, more severely depressed individuals who do not have this capability may benefit from what CT has to offer. Those who already have this capability may benefit from treatments that do not act by teaching executive control, such as antidepressant medications (29). Like our more severely depressed treatment-remitters, many depressed children display decreased pupillary responses to negative words on a similar task (36), potentially suggesting that they would be good candidates for CT.

This algorithm applies only to CT and cannot help decide for which of multiple treatments patients are best suited. Additional studies are necessary to determine whether this profile is predictive of remission in other interventions, particularly those that do not target executive control. If effects are specific, patients with decreased sustained pupillary responses may also respond well to targeted interventions specifically designed to increase executive control (37).

These data qualify our observations of sustained pupil diameter in depressed individuals using similar tasks (14,15), suggesting there may be multiple relevant subgroups of depressed individuals. The qualitative similarity of control participants to nonremitting depressed (Figure 2) and mechanistic correlates (Figure S9 in Supplement 1) could reflect increased mobilization of prefrontal control in each of these groups. Whether healthy individuals “need” such control as a function of the emotional impact of the stimuli or habitually engage these mechanisms either after an emotional stimulus or before the subsequent cognitive task is not clear from the current design.

Limitations

This study has several limitations. Without a no-treatment arm, it is unclear whether predicted remitters would remit without CT—thus pupillary responses cannot yet be used in decisions of whether to require CT. In addition, our sample was small compared with most treatment-moderator studies, requiring replication. The task was long, and part of a 3-hour battery. Many facets of the task were analyzed only in an exploratory context (e.g., responses to positive words; see supplementary data) because our goal involved understanding the association of physiologic responses to negative stimuli with clinical change. This relationship was strong enough that adding other design facets would not have helped quantitatively. That said, the effects of fatigue due to the long task on observed results are difficult to gauge; replication without positive and neutral words and only one task could be useful. Control participants were younger, on average, than depressed participants. For this study, this was not a major confound because primary comparisons did not regard controls. The clinical sample was entirely composed of recurrently depressed participants limiting generalizability. Predictions using pupillary responses to neutral words were similar, potentially suggesting either 1) the operative feature of the task was not negative words but the cognitive control portion (digit sorting) or anticipation thereof or 2) negative elaboration on neutral stimuli. Brain imaging data were acquired in participants who had already undergone pupillary assessment on the same task, potentially introducing practice effects. The low number of subjects with brain imaging data precludes reliable conclusions regarding mechanistic specificity. Rather, these analyses provide initial exploratory evidence useful in interpreting the most strongly contributing underlying mechanisms. Our data do not say whether examined phenomena were trait- or state-related.

Summary and Future Directions

The foregoing limitations notwithstanding, our results suggest a sensitive, specific, inexpensive, quick, and mechanistically interpretable method to aid in treatment planning for depression. The assessment setup could be implemented for under $12,000 in virtually any clinic environment. Costs per assessment are trivial, requiring 10 to 30 minutes from a minimally trained technician. The software for implementing this experiment can be obtained, for free, from the principal investigator. Thus, upon replication, this type of psychophysiologic assessment could become an integral part of the emerging trend toward personalized treatment for depression.

This work was supported by National Institute of Mental Health Grant Nos. MH082998, MH064159, MH60473, MH55762, MH58397, and MH58356; National Alliance for Research on Schizophrenia and Depression, the Veteran’s Administration, and the Veteran’s Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. This material is the result of work supported with resources and the use of facilities at the Pittsburgh VA Healthcare System.
Greg Siegle is a consultant for Trial IQ. Michael Thase reported having served as an adviser/consultant for AstraZeneca; Bristol-Myers Squibb Company; Dey Pharma LP; Eli Lilly and Co.; Forest Laboratories; Gerson Lehman Group; GlaxoSmithKline; Guidepoint Global; H. Lundbeck A/S; MedAvante, Inc.; Merck and Co., Inc.; Neuronetics, Inc.; Novartis; Otsuka; Ortho-McNeil Pharmaceuticals; Pamlab, LLC; Pfizer (formerly Wyeth Ayerst Pharmaceuticals); Schering-Plough (formerly Organon, Inc.); Shire US, Inc.; Supernus Pharmaceuticals; Takeda (Lundbeck); and Transcept Pharmaceuticals. Dr. Thase has been a member of the Speakers Bureaus for AstraZeneca; Bristol-Myers Squibb Company; Eli Lilly and Co.; Merck and Co., Inc.; and Pfizer (formerly Wyeth Ayerst Pharmaceuticals). He receives grant funding from the Agency for Healthcare Research and Quality, Eli Lilly and Co., GlaxoSmithKline, National Institute of Mental Health, and Sepracor, Inc. Dr. Thase has equity holdings in MedAvante, Inc. and receives royalty income from American Psychiatric Publishing, Inc., Guilford Publications, Herald House, Oxford University Press, and W.W. Norton and Company, Inc. His wife is employed as the group scientific director for Embryon (formerly Advogenet; Embryon does business with Bristol-Myers Squibb and Pfizer/Wyeth). Dr. Friedman has obtained grant/research support from Aspect Medical Systems, Indevus, AstraZeneca, Bristol-Myers Squibb, Sanofi-Aventis, Wyeth-Ayerst, Cyberonics, Novartis, North Star/St. Jude Medical, Medtronic, Respironics, National Institute of Mental Health, and Agency for Healthcare Research and Quality. Dr. Friedman also has received other financial or material support (expert testimony, patents, royalties) from Springer. The work was conducted at the University of Pittsburgh School of Medicine and VA Pittsburgh Healthcare System. All other authors reported no biomedial financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.


reaction activity differentially predict concurrent and prospective func-
27. Guthrie D, Buchwald JS (1991): Significance testing of difference poten-

www.sobp.org/journal