IMPORTANCE  Patients' previous experience with performance-based cognitive tests in clinical trials for cognitive impairment associated with schizophrenia can create practice-related improvements. Placebo-controlled trials for cognitive impairment associated with schizophrenia are at risk for these practice effects, which can be difficult to distinguish from placebo effects.

OBJECTIVES  To conduct a systematic evaluation of the magnitude of practice effects on the Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive Battery (MCCB) in cognitive impairment associated with schizophrenia and to examine which demographic, clinical, and cognitive characteristics were associated with improvement in placebo conditions.

DESIGN, SETTING, AND PARTICIPANTS  A blinded review was conducted of data from 813 patients with schizophrenia who were treated with placebo in 12 randomized placebo-controlled clinical trials conducted mostly in outpatient clinics in North America, Europe, Asia, and Latin America from February 22, 2007, to March 1, 2014. A total of 779 patients provided data for the primary outcome measure at baseline and at least 1 follow-up. Seven trials had prebaseline assessments wherein the patients knew that they were not receiving treatment, allowing a comparison of practice and placebo effects in the same patients.

INTERVENTIONS  Placebo compared with various experimental drug treatments.

MAIN OUTCOMES AND MEASURES  Composite score on the MCCB.

RESULTS  Of the 813 patients in the study (260 women and 553 men; mean [SD] age, 41.2 [11.5] years), the mean MCCB composite score at baseline was 22.8 points below the normative mean, and the mean (SEM) total change in the MCCB during receipt of placebo was 1.8 (0.2) T-score points (95% CI, 1.40-2.18), equivalent to a change of 0.18 SD. Practice effects in the 7 studies in which there was a prebaseline assessment were essentially identical to the postbaseline placebo changes. Baseline factors associated with greater improvements in the MCCB during receipt of placebo included more depression/anxiety ($F_{1,438} = 5.41; P = .02$), more motivation ($F_{1,272} = 4.63; P = .03$), and less improvement from screening to baseline ($F_{1,421} = 59.32; P < .001$).

CONCLUSIONS AND RELEVANCE  Placebo effects were minimal and associated with the number of postbaseline assessments and several patient characteristics. Given that the patients performed 2.28 SDs below normative standards on average at baseline, a mean placebo-associated improvement of less than 0.2 SD provides evidence that ceiling effects do not occur in these trials. These minimal changes in the MCCB could not be responsible for effective active treatments failing to separate from placebo.
Cognitive impairment associated with schizophrenia is severe and a primary cause of poor functional outcomes. Extensive resources have been devoted to developing pharmacologic,1-6 psychosocial,7,8 and cognitive remediation9-13 treatments to reduce cognitive and functional impairment in patients with schizophrenia. To our knowledge, no treatments to date have had sufficient benefit to warrant regulatory approval. Because there are no effective medical treatments for cognitive impairment associated with schizophrenia, it is challenging to determine whether the lack of success of previous treatment development programs has been due to inadequacies of the treatments being tested or to weaknesses in the methods used.

Cognitive enhancement studies have typically used a regulatory pathway that was developed during the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) initiative.14,15 Central to this method has been the assessment of cognition with a set of performance-based tests referred to as the MATRICS Consensus Cognitive Battery (MCCB).16-18 which was selected on the basis of several criteria, including usefulness as a repeated measure, defined as high test-retest reliability combined with minimal practice effects. However, the MCCB validation study used a single test-retest assessment,17 which does not reflect the complexity of standard clinical trials that have more than 2 assessments performed with the MCCB1,2,19,20 and multiple treatment conditions.1,4

Patients with schizophrenia manifest consistent cognitive performance over time with a variety of different tests17,21,22 and may generate practice effects when they are reassessed.21 However, most clinical trials have used parallel group designs with a placebo control; thus, it is difficult to determine if improved performance over time in patients receiving placebo is due to simple learning or practice effects or to the expectation bias inherent in a placebo-controlled design. A comparison of the improvement in patients receiving placebo with those who were aware that they were not receiving any treatment would help to determine if expectation bias increases the known improvement due to practice effects.

Regardless of their source, improvements on cognitive tests in patients receiving placebo may be so large that treatment effects are impossible to detect. The opposite speculation has equivalent face validity and has been supported by single studies1,19,20. If an individual is familiar with being assessed and with the assessment materials, additional assessments are less likely to lead to gains in performance based solely on test exposure. Furthermore, the magnitude of expected practice and placebo response of the MCCB, which is used in most large multisite clinical trials for cognitive impairment associated with schizophrenia, is unknown. Finally, while certain patients with schizophrenia and depression who are receiving placebo demonstrate large symptom responses that can be anticipated based on their baseline characteristics,21 it has not been determined if patients receiving placebo who respond in schizophrenia cognition trials can be similarly identified.

There are clear implications in these questions for the interpretation of previous results and the design of future clinical trials. If individual differences among patients, including differences in their performance change from screening to baseline, are associated with postbaseline retest changes, these measures could be used as factors to be considered in statistical analysis or in patient randomization and stratification.

In this study, we merged data from 12 separate clinical trials,1-3,20,24-29 including 813 patients who were randomized to receive placebo during the course of a double-blind pharmacologic intervention study, and investigated the characteristics of retest-associated improvements in performance on the MCCB. We compared prerandomization improvements with postrandomization improvements in patients receiving placebo, and we assessed the effect of a number of measures on the magnitude of postbaseline improvement, including the number of retest assessments, the spacing of the retest assessments, the time elapsed since baseline while receiving placebo, and the number of treatment arms of the study. Finally, we examined a variety of patient demographic and clinical characteristics to determine if they are associated with cognitive placebo response.

Methods

Sample Characteristics
This review included 813 patients with schizophrenia who were treated with placebo in randomized clinical trials from February 22, 2007, to March 1, 2014. The Table describes the demographic and baseline characteristics of the merged data set, with mean (SD) values for continuous variables and numbers (percentages) for categorical variables.30 As is typical of clinical trial data on patients with schizophrenia, approximately two-thirds of the patients (553 [68.0%]) were men. Most of the patients were evaluated in North America, with samples also collected in Europe, Asia, and Latin America. eTable 1 in the Supplement describes the characteristics of the 12 trials from which the data were merged, including ClinicalTrials.gov identification number, total number of individuals randomized and assigned to receive placebo, dates of the trial, the number of treatment groups,
and the number of MCCB assessments performed. Each clinical trial was approved by central and/or local institutional review boards, and all study participants provided written informed consent.

Thirty-one patients had 1 or more missing MCCB domain values at the baseline visit, resulting in 782 patients with a cognitive composite score derived from all 7 MCCB domains at baseline. Three patients did not have at least 1 follow-up score, resulting in a final sample size of 779. An additional 54 patients had test results missing in a domain that had additional tests, so a domain score was able to be calculated. For all of the trials, patients were treated with a stable dose of antipsychotic medication, and most patients were receiving only 1 antipsychotic. The mean (SD) MCCB cognitive composite score at first assessment was 27.2 (12.78), consistent with similarly timed assessment scores from a larger data set of more than 4000 patients. This score reflects performance that is more than 2 SDs below normative standards. An improvement of 1 point on the MCCB cognitive composite score reflects a 0.1-SD improvement in normative scores, which means that the patients were a mean (SD) 23.0 (2.3) points below the normative population mean and at least a mean (SD) of 73.0 (7.3) points below a maximum score on the MCCB composite.

MCCB Assessments

The MCCB subtests are organized into the following 7 domains: (1) Speed of Processing: Trailmaking Test, Brief Assessment of Cognition Symbol Coding, and Category Fluency; (2) Attention-Vigilance: Continuous Performance Test-Identical Pairs; (3) Working memory: Wechsler Memory Scale-III Spatial Span and Letter-Number Span; (4) Verbal learning: Hopkins Verbal Learning Test-Revised; (5) Visual learning: Brief Visuospatial Memory Test-Revised; (6) Reasoning and Problem Solving: Neuropsychological Assessment Battery Mazes; and (7) Social Cognition: Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) Managing Emotions.

Data Quality Assurance

All testers who interacted with patients were first trained on the administration and scoring of the MCCB using video and group training sessions and were individually certified by an MCCB expert. All MCCB data were scored locally or sent to a central site, where they were scored or rescored. Training, data collection, and data quality assurance were implemented or supervised by an experienced psychologist (R.S.E.K. or A.S.A.) as per the guidelines outlined in the MCCB manual.17

MCCB Composite Scores

For the 10 subtests, 7 domains, and 2 composite scores of the MCCB, the scoring program yields T scores that are standardized and corrected for age and sex, based on previously collected normative data matched for age, sex, and race/ethnicity to the US Census. Each standardized measure has a mean (SD) of 50 (10). The cognitive composite is the standardized total of the 7 domains. The neurocognitive composite is calculated similarly but does not include Social Cognition.32

<table>
<thead>
<tr>
<th>Table. Demographic and Baseline Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristic</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Age, mean (SD), y (n = 813)</td>
</tr>
<tr>
<td>Sex (n = 813)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Race (n = 813)</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Geographical region (n = 813)</td>
</tr>
<tr>
<td>North America</td>
</tr>
<tr>
<td>Eastern Europe</td>
</tr>
<tr>
<td>Asia</td>
</tr>
<tr>
<td>Latin America</td>
</tr>
<tr>
<td>Western Europe</td>
</tr>
<tr>
<td>Postsecondary education (n = 552)</td>
</tr>
<tr>
<td>Current smoker (n = 603)</td>
</tr>
<tr>
<td>Age at onset of illness, mean (SD), y (n = 233)</td>
</tr>
<tr>
<td>Duration of illness &lt;10 y (n = 626)</td>
</tr>
<tr>
<td>Baseline antipsychotic (n = 488)</td>
</tr>
<tr>
<td>Risperidone/paliperidone</td>
</tr>
<tr>
<td>Olanzapine</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Quetiapine</td>
</tr>
<tr>
<td>Aripipazole</td>
</tr>
<tr>
<td>Baseline MCCB composite T score, mean (SD) (n = 782)</td>
</tr>
<tr>
<td>Baseline UPSA-2 total, mean (SD) (n = 225)</td>
</tr>
<tr>
<td>Baseline SCoRS total, mean (SD) (n = 158)</td>
</tr>
<tr>
<td>Baseline PANSS total, mean (SD) (n = 669)</td>
</tr>
<tr>
<td>Baseline NSA-16 total, mean (SD) (n = 489)</td>
</tr>
</tbody>
</table>

Abbreviations: MCCB, Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive Battery; NSA-16, Negative Symptom Assessment; PANSS, Positive and Negative Syndrome Scale; SCoRS, Schizophrenia Cognition Rating Scale; UPSA-2, University of California at San Diego Performance-based Skills Assessment, 2nd edition.30

*Data are presented as number (percentage) of patients unless otherwise indicated.*

Symptom Assessments

Positive and Negative Syndrome Scale

Data from the Positive and Negative Syndrome Scale (PANSS)33 were available in 7 studies.2,3,24-26 The PANSS symptoms were grouped into 5 factors (positive symptoms, negative symptoms, disorganized thought, hostility/excitement, and depression/anxiety), based on a previously established model.34

Negative Symptom Assessment

In 5 studies,3,25,26 the Negative Symptom Assessment35,36 was used to examine negative symptoms in 4 factors: communication, emotion, motivation, and sociality.

Statistical Analysis

Data were combined from the intent-to-treat populations of 12 MCCB clinical trials. Visit schedules varied considerably among the studies, ranging from 2 to 6 MCCB assessments.
(1-4 postbaseline reassessments) during a period of 4 to 24 weeks. Change from baseline in the MCCB cognitive composite score was investigated using a basic linear mixed model for repeated measures, controlling for baseline value and week of assessment nested within a random study effect. The ability of patient-level and study-level variables to estimate the magnitude of the placebo effect was analyzed with separate mixed models for each predictor. Change in MCCB composite score between screening and baseline (“pure” practice effect) was compared between studies with analysis of covariance adjusting for the screening value. This analysis included the 7 studies that administered the MCCB prior to and again at patients’ baseline assessment, during a period in which patients knew that they were not receiving treatment. P < .05 was considered significant. Complete information on the statistical plan is in the eAppendix in the Supplement.

**Results**

The mean MCCB composite score at baseline was 22.8 points below the normative mean (Table). Figure 1 describes the mean MCCB cognitive composite score over time in all 12 studies. The full 56-week data for 3 studies are available in the eFigure in the Supplement. The overall mean (SEM) change in the MCCB cognitive composite during receipt of placebo, after adjusting for baseline score, week of assessment, and study, was 1.8 (0.2) T-score points (95% CI, 1.40-2.18), with a range across studies of 0.6 to 3.3 points (Figure 2).1-3,20,24-29 The effects of baseline score ($F_{1,642} = 7.52$), week of assessment ($F_{11,180} = 3.43$), and study ($F_{11,180} = 2.82$) were all statistically significant ($P < .01$). Mean (SEM) change in scores for the 10 subtests comprising the MCCB are shown in the eFigure in the Supplement.

---

**Figure 1.** Mean Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive Battery (MCCB) Cognitive Composite T Scores by Study and Visit for Patients Receiving Placebo

**Figure 2.** Change in Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive Battery Cognitive Composite T Score by Study for Patients Receiving Placebo
MCCB were all small, ranging from 0.2 (0.3) (MSCEIT managing emotions) to 2.2 (0.3) (Trail Making A) T-score points (eTable 2 in the Supplement).

Prerandomization Practice Effects
Seven of the studies had a prebaseline screening assessment (n = 558). The changes from screening to baseline assessment for the cognitive composite score varied from 0.7 to 3.1 MCCB T-score points (mean [SEM], 2.2 [0.25] T-score points). The analysis of covariance for the practice effects found a significant effect of retesting (F1,502 = 10.98; P = .001). The effect of the study was nonsignificant (F6,502 = 1.69; P = .12), indicating that the retest effects were not significantly different across studies.

Postbaseline Changes in Patients Receiving Placebo
To examine the effects of a prebaseline screening assessment on postbaseline placebo effects, the postbaseline changes in the 7 studies with a screening assessment were compared with the changes for the 5 studies without such an assessment. When a dichotomous variable indicating the presence or absence of a screening MCCB assessment was added to the model as a covariate for the postbaseline analyses, its effect was not statistically significant (F1,695 = 1.64; P = .20), indicating that performing prebaseline screening assessments was not associated with differences in postbaseline changes while the patients were receiving placebo.

We also explored the association between post baseline changes and the numbers of weeks and assessments that had occurred since baseline, adjusting for baseline scores and whether the study had a screening visit. The effect of time since baseline was statistically significant (F1,649 = .67; P < .001) overall and for all pair-wise comparisons. Patients were classified into 3 groups based on the magnitude of the change in their screening to baseline practice effect: negative (change of ≤−5 points), small or neutral (change of >−5 to 5 points), and positive (>5-point change). Forty-two patients without composite scores at screening, baseline, and at least 1 follow-up were excluded from the analysis (n = 516). Test for practice effect-by-week interaction was nonsignificant (P = .67).

Factors Associated With Placebo Response
The following variables were associated with placebo response: the Negative Symptoms Assessment motivation factor score (F1,727 = 4.63; P = .03), the PANSS Depression/Anxiety Marder Factor score (F1,418 = 5.41; P = .02), the practice effect between screening and baseline MCCB (F1,421 = 5.93; P < .001), 4 vs 3 treatment groups (F1,452 = 5.14; P = .03), and the study sample size (F1,725 = 4.65; P = .03). A larger placebo response was associated with more motivation, greater depression/anxiety, lower prebaseline practice effects, 4 vs 3 treatment groups, and a larger study sample.

Given the robustness of the screening to baseline MCCB composite score change as a factor associated with response, we further investigated the effect of classifying patients into the following 3 groups based on the magnitude of the change in their screening to baseline practice effect: negative (≤−5-point change), small or neutral (>−5 to 5-point change), and positive (>5-point change). As demonstrated in Figure 3, there were robust differences between the groups. Patients who manifested negative practice effects (mean [SEM], 4.8 [0.68] T-score points) demonstrated larger placebo effects (t425 = 3.43; P < .001) than those with small practice effects (mean [SEM], 2.2 [0.26] T-score points), who in turn demonstrated larger placebo effects (t410 = 4.13; P < .001) than those who had positive practice effects (mean [SEM], −0.1 [0.51] T-score points).

Discussion
The data from this study were generated from the placebo condition in 12 separate randomized, placebo-controlled clinical trials using the MCCB as the primary end point to

Figure 3. Change in Measurement and Treatment Research to Improve Cognition in Schizophrenia Consensus Cognitive Battery Cognitive Composite T Score by Level of Practice Effect in Studies With a Screening Assessment

Least squares mean change from baseline through week 24 is plotted by level of practice effect as calculated by the T-score change in the cognitive composite from screening to baseline. Results are from a mixed-effect repeated measures model with a continuous fixed covariate for the baseline score and a categorical effect for visit nested within a random study effect, assuming an unstructured covariance matrix. This analysis was limited to the 7 studies that assessed cognition at screening (n = 558) and revealed a significantly increasing placebo effect with decreased practice effect between screening and baseline (P < .001 overall and for all pairwise comparisons). Patients were classified into 3 groups based on the magnitude of the change in their screening to baseline practice effect: negative (change of ≤−5 points), small or neutral (change of >−5 to 5 points), and positive (>5-point change). Forty-two patients without composite scores at screening, baseline, and at least 1 follow-up were excluded from the analysis (n = 516). Test for practice effect-by-week interaction was nonsignificant (P = .67).
assess response to treatment with a pharmacologic intervention in patients with schizophrenia. The mean amount of total improvement during treatment in patients receiving placebo was approximately 2 T-score points on the MCCB composite score, consistent with a Cohen $d$ effect size of 0.18, which is slightly smaller than the Cohen definition of small. The amount of change due to the practice effect from prebaseline to baseline was almost identical to the mean amount of change reported in placebo conditions. Studies using a prebaseline assessment did not have smaller postbaseline placebo treatment effects than did studies with the first MCCB assessment at baseline. The size of the placebo effect was more influenced by the number of postbaseline retests than by the duration of placebo treatment, with the number of reassessments associated with an increase in scores. Key additional factors associated with the magnitude of placebo response in these trials included smaller improvement from prebaseline to baseline assessment with the MCCB, less severe negative symptoms, more severe depression/anxiety symptoms, 4 vs 3 treatment groups in the trial, and larger sample size.

There are several important implications of these findings. The most obvious conclusion from these data is that the concept of a substantial placebo effect in schizophrenia clinical trials using the MCCB—one that may obscure the detection of meaningful clinical change—is clearly not substantiated. The data in this report suggest that cognitive performance assessed with the MCCB is 2.28 SDs below the mean of the healthy general population at baseline. Thus, the 0.18-SD change reported with placebo is exceeded by the size of the baseline cognitive impairment in patients with schizophrenia by a factor of 12. This amount of postbaseline improvement is almost identical to the amount of prebaseline improvement in these trials and is consistent with the amount of prebaseline improvement reported in a larger sample of 1814 patients from 10 trials. Thus, the amount of improvement in the placebo group in cognition clinical trials is consistent with the expected amount of improvement due to learning effects and is small enough to be easily overwhelmed by a legitimate treatment effect. The notion that this magnitude of retest improvement accounts for the negative results in previous cognitive enhancement trials seems unlikely. For the cognitive changes associated with a pharmacologic agent to be of real benefit to patients with schizophrenia, they must exceed the small learning effect represented in the placebo condition in these trials, which is also a part of the cognitive response in patients receiving active treatment.

This amount of improvement with placebo in performance-based cognitive assessments is clearly less than that seen in clinical trials across different neuropsychiatric conditions, including symptom rating scales such as the PANSS and the Hamilton Depression Rating Scale. Because the amount of cognitive improvement in patients with cognitive impairment associated with schizophrenia receiving placebo is almost identical to the learning effects that are seen prior to baseline, the increases are almost certainly due to practice, comfort, or learning effects and not to the expectation biases that can have a tremendous effect on rating scales.

Our results can be interpreted to provide recommendations regarding the strategy of testing patients once prior to baseline to expose them to the testing process or to minimize a learning effect. A simple analysis comparing patients who were tested prior to baseline with patients who were not suggests that they had similar improvements in placebo conditions and that this strategy does not systematically lead to reduced postbaseline changes. Also, a larger number of assessments after baseline was associated with slightly larger MCCB improvements, although the magnitude of between-assessment improvement may diminish as the number of postbaseline assessments increases.

These results have implications for study design. First, researchers will need to weigh the value of additional assessments during clinical trials against the potential for an additional learning effect in patients receiving placebo and active treatment, although this effect is small. Second, important information may be gained from a prebaseline assessment. Patients whose cognitive performance worsened from screening to baseline had the largest placebo effect, and those with the largest improvements from screening to baseline had no placebo effect at all. These results, consistent with the principle of regression to the mean, suggest that, for schizophrenia cognition studies, the concept of eliminating “placebo responders” who would then go on to have large placebo effects during a trial is not a viable approach. However, a screening assessment may produce information that can facilitate study design, such as stratification of patients based on potential placebo effects. It may also allow statistical approaches that adjust for screening values (along with the customary baseline values) in the primary efficacy model as a method for reducing the noise associated with large fluctuations in patients who may have had exceptionally good or bad days at the baseline visit. Preliminary analyses using this method suggest that it may increase precision of assessing treatment effects in cognitive treatment trials. Other individual difference characteristics found in our analyses to be associated with placebo group improvements, including more severe depression/anxiety and more motivation, could be combined with prebaseline changes to equalize factors that appear to increase the placebo response.

The present results raise a more general question of whether cognitive impairments measured by neuropsychological tests are indeed mutable with pharmacologic intervention in schizophrenia, and whether other measures of brain function are better candidates to serve this purpose. First, neuropsychological tests have demonstrated sensitivity to changes under a variety of different conditions in schizophrenia. They are highly sensitive to the effects of alcohol, nicotine, anticholinergic medications, and stimulants, as well as other drugs. Second, because they have consistently demonstrated a clinically significant magnitude of improvement in response to cognitive remediation strategies in schizophrenia, the possibility that neuropsychological tests could be a plastic attribute in response to behavioral intervention but a stable trait in response to pharmacologic intervention does not seem reasonable.
Third, although it is often assumed that the effects of drugs on other, more biologically proximal measures of brain function, such as brain imaging or electrophysiology, are more powerful than on neuropsychological tests, there is little evidence to support this assumption.\textsuperscript{45} Finally, several early-phase studies have demonstrated improvements on the MCCB with pharmacologic intervention\textsuperscript{1,2} that were not replicated in larger programs. It is not at all clear whether the inability of later studies to replicate these findings were due to inefficacy of the pharmacologic intervention, insensitivity of the cognitive tests, or other weaknesses of study design and implementation.\textsuperscript{2,40} Only one phase 3 program has used the MCCB as a primary end point, and the results of that program were inconclusive.\textsuperscript{40}

Conclusions

In 12 double-blind, placebo-controlled, clinical trials for cognitive impairment associated with schizophrenia, the magnitude of placebo effects on the MCCB was small and not different from the retesting effects during the pretreatment screening period when patients knew that they were not receiving treatment. The magnitude of the placebo effect was not large enough to obscure a treatment effect with a small to medium effect size. There are patient and study characteristics that have a modest influence on postbaseline retest changes, and the number of reassessments is minimally associated with changes in performance on the MCCB.

REFERENCES


40. Brannan S. The difficulties in going from P2 to P3 in CNS trials: red flags from a recent CIS program. Paper presented at: 13th Annual Scientific Meeting of the International Society of CNS Clinical Trials; February 23, 2017, Washington, DC.


