

Review

Circumstances Under Which Practice Does Not Make Perfect: A Review of the Practice Effect Literature in Schizophrenia and Its Relevance to Clinical Treatment Studies

Terry E Goldberg^{*1}, Richard SE Keefe², Robert S Goldman³, Delbert G Robinson¹ and Philip D Harvey⁴

¹Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY, USA; ²Duke University; ³Pfizer; ⁴Emory University

In this article, we review the literature on practice effects in schizophrenia, an underappreciated confound in interpreting cognitive improvement in clinical trials. We first examine claims regarding first- and second-generation antipsychotic medications as cognitive enhancers, and follow it with a discussion of recent studies demonstrating how practice or placebo effects may drive 'positive' findings. Thus, this review suggests that many previous findings can be reinterpreted in this light. Critically, we also make several suggestions about test construction, study design, and statistical analyses that the field might use to overcome this potential confound. Our suggestions may also have implications for drug discovery and regulatory approval of cognitive-enhancing adjunctive agents, in terms of study design and/or test psychometric characteristics, including the development of tests that are relatively insensitive to practice-related changes. Such advances might be important for improving the methodology involved in the assessment of cognitive change in treatment studies.

Neuropsychopharmacology (2010) **35**, 1053–1062; doi:10.1038/npp.2009.211; published online 20 January 2010

Keywords: cognition; schizophrenia/antipsychotics; clinical pharmacology/trials; learning and memory; practice effects

INTRODUCTION

Schizophrenia affects multiple domains of life functioning. Currently, treatment studies and drug discovery efforts focus on positive symptoms, negative symptoms, and neurocognitive impairments as treatment targets. Many of the limitations of evaluating symptom change (eg, subject's denial of symptoms owing to lack of insight or desire of being discharged) are well known to clinicians, as they are similar to the problems faced when evaluating their patients. Clinicians and researchers may be much less familiar with difficulties arising while assessing changes in performance-based measures, such as neuropsychological tests. This review will focus on the underappreciated difficulties (eg, practice effects and placebo effects) often encountered while interpreting the results of serial cognitive testing designs typically used in current treatment studies. In our review, we differentiate practice effects from placebo effects. We consider practice effects to be based on item-specific learning, development of test-taking strategies (eg, chunking or deep encoding), and/or procedural

learning that might include stimulus–response mappings. We consider placebo effects to be the result of increases in motivation, decreases in anxiety, and generalized positive effects of being in a closely monitored treatment study. Critically, it is possible that both practice and placebo effects can be confounded with cognitive enhancement associated with drug treatments. We will review the possibility that practice or placebo effects are present in clinical trials assessing cognitive change. Second, we will discuss various approaches that might minimize practice or placebo effects at levels of test construction, study design, and statistical analysis. Lastly, we will discuss the implications of the findings.

COGNITION AS A TREATMENT TARGET

Cognitive deficits are important targets for medication development given that neuropsychological impairments (in particular, impairments of executive function, episodic memory, and speed of processing or attention) account for a large share of social and vocational morbidity associated with schizophrenia (Goldberg and Green, 2002). Assessment of neuropsychological performance has become the norm in clinical trials of antipsychotic medications, in terms of identifying both potentially beneficial effects and potentially deleterious side effects. Their importance has become better understood largely because research has indicated that

*Correspondence: Dr Terry Goldberg, Division of Psychiatry Research, Zucker Hillside Hospital, 75-59 263rd Street, Glen Oaks, NY 11004, USA, Tel: +1 718 470 8151, Fax: +1 718 343 1659, E-mail: TGoldber@NSHS.edu

Received 11 September 2009; revised 10 November 2009; accepted 10 November 2009

cognitive impairments are stronger predictors of functional disability than psychotic symptoms (ie, delusions and hallucinations), which form the cornerstone of the diagnosis of schizophrenia. For instance, in an influential meta-analysis, Green (1996) demonstrated that several domains of cognitive function, including attention, working memory, and episodic memory, were significant predictors of functional outcome. In contrast, psychotic symptoms (hallucinations and delusions) have generally been found to be weak predictors and correlates of functional outcome. The relative contributions of symptoms and neurocognition to functional outcome have only rarely been directly compared using appropriate statistical analyses, including multiple regression or path modeling. In studies in which this comparison was carried out, contributions of neurocognition to outcome were stronger than those of positive symptoms. For example, Bowie *et al* (2006) demonstrated that a composite cognitive score was the strongest predictor of a performance-based measure of everyday living skills and it showed substantial correlations with everyday outcomes, whereas neither positive nor negative symptoms predicted functional capacity and only weakly predicted everyday function. Mohamed *et al* (2008) reached similar conclusions, especially in the domains of work and instrumental, goal-directed activities. Negative symptoms have higher correlations with functional outcome than positive symptoms, but across studies it has been observed that relationships are neither stronger nor more consistent than those for neurocognitive deficits (Harvey *et al*, 1998; Velligan *et al*, 1997). Although negative symptoms covaried to, at least, a modest extent with neurocognition (Velligan *et al*, 1997), their relationship with function seems to be mediated through statistical overlap, ie, they did not make independent contributions to explain the outcome variance (see the study by Harvey *et al* (2006) for a demonstration of this possibility).

SERIAL COGNITIVE TESTING AND PRACTICE EFFECTS

One property of many of the cognitive tests used in clinical trials that is not widely considered is the possibility that subjects may demonstrate practice- or placebo-related improvements after repeated exposure to the same test. Although this issue has been raised in the literature (Gold *et al*, 2000); there has been little empirical examination of its possible role until recently. As discussed below, such effects can make it difficult to distinguish between treatment-related versus practice- or placebo-related improvements in cognitive functioning. This practice-related improvement could be due to a number of factors, including increased familiarity with and recall of specific task content, instructions, or equipment; improvements in test taking strategy; or procedural learning of stimulus-response mapping; whereas the placebo-related improvement could be the result of positive expectations or biases for change, as well as of emotional factors such as increased motivation and decreased anxiety. All of these factors, other than the very first, apply even across alternate versions (ie, 'forms') of assessment devices and must be managed in such a way that

the utility of the tests, as measures of treatment-induced cognitive change, is not compromised.

PRACTICE EFFECTS AS A FUNCTION OF GENERATION OF ANTIPSYCHOTIC MEDICATION

First Generation Antipsychotic Medications

In a series of reviews on first-generation antipsychotic compounds (Weickert and Goldberg, 2005), it was generally considered that they provided little benefit to cognitive function, but did not exact much of a cost (with the possible exception of motor function early in treatment). However, when approximately 20 studies (involving over 800 subjects) were subjected to a meta-analysis, a small, statistically significant effect size (ES) of about 0.22 (with a confidence interval that excluded zero) was observed (Mishara and Goldberg, 2004). Many of the studies that were reviewed assessed cognition serially at baseline and again after several months of treatment, although compelling evidence for practice effects was not found.

Studies that directly examined practice effects have found inconclusive results on examining retesting during naturalistic treatment with conventional antipsychotic medications. For instance, Harvey *et al* (2005) examined a sample of 45 older, community-dwelling patients with schizophrenia, who were treated with stable doses of first-generation antipsychotics. These patients were examined at baseline and retested after 8 weeks in a 'simulated clinical trial' design. They reported that out of a 22-test neuropsychological assessment battery, only three tests showed significant retest effects. Of the tests, two were administered with alternate forms and 20 were administered with the same form. Interestingly, the test with the greatest change from time 1- to 8-week follow-up was a test with two alternate forms presented in a fixed order, suggesting that form effects may have been greater than practice effects. These data also suggested that, for patients treated with conventional antipsychotic medications, practice effects were clearly less than those that would be expected in healthy control populations, and were also less than retesting effects previously reported in similar samples treated with second-generation antipsychotics. In a similarly remarkable study that demonstrated the opposite result, Heaton *et al* (2001) found large, statistically significant practice effects on composite and individual measures of cognition in both schizophrenia and healthy control groups ($N=142$ and $N=206$) irrespective of whether time between two assessments was shorter (approximately 3 months) or longer (approximately 18 months), or whether global cognition was high or low. Furthermore, patients on first-generation antipsychotics were assessed. The magnitude of the improvement in the schizophrenia group ranged from 0.33 to 0.50 for full scale IQ, overall impairment rating, and global neuropsychological score. The change score was not related to clinical state, baseline cognition, or tardive dyskinesia. This set of results suggests two important points: first, as a similar testing battery was used in both studies (including digit span, digit symbol, Wisconsin card sort test, trail making, finger tapping, and verbal list learning), it is unlikely that the tests themselves were

somehow 'immune' to practice effects. Second, and more important, cohort effects may be present from study to study.

In an important and recent meta-analysis, Woodward *et al* (2007) observed that haloperidol-treated patients had less cognitive gain after repeated testing in comparison with healthy controls on two of six measures. Digit symbol substitution and verbal fluency demonstrated blunting; the trail making test, pegboard speed, and global cognitive scores did not show such blunting. The haloperidol data were based on 4–11 studies and *N* values of 185–384; the healthy control data were based on 4–18 studies and *N* values of 144–981. Other less speed-dependent tests were less affected by haloperidol treatment. This finding is consistent with the idea that haloperidol, a first-generation antipsychotic, might suppress practice effects and particularly on tests requiring speed and sustained effort. Such a finding would be consistent with the results of Harvey *et al* (2005), but does not address the Heaton *et al* (2001) practice effect findings.

Second Generation Antipsychotic Medications

The availability of second-generation antipsychotic medications, beginning in the 1990s, led to a renaissance in scientific interest in the pharmacological treatment of schizophrenia. Numerous studies on the effects of these drugs on cognition were undertaken. Influential meta-analyses of these studies have been conducted. Keefe *et al* (1999) observed that second-generation antipsychotics seemed to have an advantage of about 0.25 ES units over first-generation antipsychotic on a wide range of cognitive measures. Woodward *et al* (2005) in a large meta-analysis encompassing 1513 patients, 14 studies, and domains of cognitive function that included learning, attention, speed, and fluency, came to very similar conclusions. The difference between these two ESs (for first-generation antipsychotics and second generation antipsychotics) could, therefore, either be due to a beneficial effect of second-generation antipsychotic treatment or a suppressive effect of first-generation treatment. Many of the studies used in the two reviews used naturalistic or parallel group designs in which subjects were randomized to the second-generation antipsychotic or comparator group (often haloperidol) after a brief washout phase or with no washout, and were tested repeatedly over 1–6-month intervals. In most instances, the same version of the test was administered multiple times (eg, four administrations in a 12-month period).

Few studies were conducted that directly addressed the possibility that some of these effects may have been due to practice, resulting from multiple exposures to a given test. Although randomized, direct comparisons of conventional and atypical treated populations typically found benefits for the atypical group, there are many additional confounds in these studies. These include nonrandomized treatments, problems in the dosing of the conventional comparators, failure to consider previous treatments and associated carry-over effects, and widely different selection of tests and administration procedures across studies.

Thus, one possible interpretation of these results is that second-generation antipsychotic treatment in patients with

schizophrenia is associated with a normalization of practice effects, bringing them closer to what healthy controls demonstrate. As noted in detail above, retest effects with first-generation antipsychotics are associated with changes that range from zero to effects smaller than those seen with second-generation antipsychotics (Harvey and Keefe, 2001; Mishara and Goldberg, 2004; Woodward *et al*, 2005). As a result, it is possible to hypothesize, but probably impossible to determine, that there is a gradient of practice effects in people with schizophrenia, with the smallest seen in unmedicated patients and the largest in patients treated with second-generation antipsychotics. However, even this view may be arguable, given recent evidence that first-generation antipsychotic medications seem to produce the same magnitude of improvement as second-generation antipsychotics when compared directly in both first episode and more chronic groups, as in CATIE or EUFEST (Keefe *et al*, 2007; Davidson *et al*, 2009).

RECENT WORK

Studies on individuals in their first episode of schizophrenia offer certain unique research advantages. As the duration of psychotic symptoms has often been relatively short, issues associated with chronicity, such as patient role, institutionalization, interactions with aging, and disease processes, are minimized. Second, long and complicated medication treatment histories with unknown effects on neurobiology are also avoided. Third, treatment response in terms of psychiatric symptoms is generally relatively substantial early in the course, affording an opportunity to determine how symptomatic clinical improvement is related to the improvement in other domains.

Large industry-sponsored controlled trials examining risperidone or olanzapine in first episode patients found significant improvement from baseline with the second-generation antipsychotics after patients underwent multiple assessments; ES values ranged from about 0.35 to 0.55 on composite measures of cognition. Furthermore, low-dose treatments with first-generation medications were found to be significantly inferior to the effects of the second-generation antipsychotics. Critically, these studies did not include either untreated or healthy comparison groups, making it impossible to determine whether improvements were due to practice effects and whether differences across drug classes with treatment are differences in practice effects or true treatment differences.

In a recent study on first episode patients, we sought to determine the effects of two second-generation antipsychotic medications, risperidone and olanzapine, on cognition (Goldberg *et al*, 2007). This study also included a large group of healthy controls to directly compare the magnitude of cognitive change in first episode patients and healthy controls; the first study to report such data in the context of a clinical trial. In the latter group, improvement could only be reasonably attributed to practice or exposure. Of the 104 first episode patients, 80 were never previously exposed to antipsychotic medication and 14 had less than 1 week of antipsychotic exposure; thus changes over the following weeks could not be attributed to a switch in medication, withdrawal from medication, or long and/or

complex histories of treatment. All patients were actively psychotic when they entered the study. A total of 84 healthy controls were recruited from the community by advertisement or word-of-mouth.

The FE patients were assessed at baseline and randomly assigned to treatment with olanzapine ($N=51$) or risperidone ($N=54$) for 16 weeks. First episode patients and the healthy controls group received cognitive assessments at baseline (when most first episode patients were drug free) and 6 and 16 weeks thereafter. The cognitive tests included measures of processing speed, episodic memory, working memory, executive function, and motor speed/dexterity.

Briefly, there were no differential effects of olanzapine and risperidone on cognition. We therefore combined the patients into a single psychotic group and compared it to the healthy control group. For nearly all measures, we found that there was improvement over time (ie, a main effect of time) but no group time interactions. Thus, the majority of variables did not demonstrate rates of improvement above and beyond practice effects: verbal episodic memory, visual spatial processing, card sorting and set shifting, and digit symbol coding speed. It is also sobering to note that the cognitive composite ES in the psychotic group (0.35) would be considered moderate and could be attributed to treatment; only when it is compared with the ES in the healthy control group (0.33) does it become clear that the magnitude of the effect is in keeping with practice-related phenomena in healthy controls. Goldberg *et al* (2007) concluded that gains in the first episode group were consistent with practice-related phenomena. Although the interpretation was inferential, the authors were able to largely rule out some possible indications of a drug effect (eg, dose effects, differences between the drugs in cognitive profile). It should be noted that there was no untreated or first-generation treatment comparison samples in this study.

Several studies outside the context of clinical trials used a healthy control group while following changes in the cognitive status of first episode patients (Albus *et al*, 2006; Hill *et al*, 2004; Hoff *et al*, 1999). In these designs, patients and healthy individuals were tested serially over equivalent intervals. Results were remarkably similar to those described above, in that over time patients generally demonstrated improvements, but these were no greater than those demonstrated by the healthy control group, who also underwent serial cognitive assessment. In some cases, a group-time interaction on cognition was found, which favored the healthy control group. As noted, medications in these studies were not rigorously controlled for the full duration of the follow-up; however, although medication regimens of the studies were naturalistic, most patients were treated with second-generation antipsychotics (Albus *et al*, 2006; Hill *et al*, 2004). Keefe *et al* (2006) compared first episode patients, all of whom were treated with olanzapine, to healthy controls. Although the sample sizes at 12 months (after three to four assessments) were not large, it is nevertheless interesting to note that both groups improved on an extensive neurocognitive battery, and to a significant and strikingly similar degree. Crespo-Facorro *et al* (2009) also studied a group of first episode patients assigned initially to haloperidol, risperidone, and olanzapine treatment, and compared their performance on neurocognitive tests at baseline (which occurred after 10 weeks of

treatment), 6 months thereafter, and 12 months thereafter with that of a healthy control group also tested serially. Sample sizes ranged from 30 to 39. All groups improved on multiple measures and to a very similar extent. There were no group-time interactions.

Recent data collected from a large clinical trial demonstrate that practice effects are not restricted to 'rarefied' first episode groups but may be present in middle-aged chronic patients as well (Keefe *et al*, 2008). In this trial, patients remained on a single second-generation antipsychotic medication over a 12-week period while participating in a double-blind cognitive enhancement study of placebo versus donepezil, during which they were cognitively assessed three times. Moderate improvements on repeated testing were observed (the composite ES was 0.45) that could be attributed only to practice or placebo effects. Interestingly, improvements with repeated testing were found even on tests in which alternate forms were used (eg, verbal list learning; these tests also use taxonomic categories that could further practice effects because of strategy-driven semantic encoding of categories). In total, these findings suggest retesting effects can be quite large, even in the absence of re-exposure to the same content, and are not restricted to a first episode sample, but can be observed in the older, multi-episode patients typically recruited for clinical trials.

MAGNITUDE OF PRACTICE EFFECTS

A number of factors may be thought to influence the magnitude of practice effects, including characteristics of the test under study (see section below on test design) and intertest time interval (very long intervals are generally thought to be associated with smaller effects, as in the study by Salthouse *et al* (2004)). A meta-analysis that examined schizophrenia patients and 'internal' healthy controls found that for five of nine cognitive measures, ESs of improvement over time were highly similar (Szöke *et al*, 2008), whereas for the remaining measures (fluency, trails, a logical memory, and card sort categories), improvements were somewhat larger in the healthy control group. Interestingly, several studies directly examined the effects of age, IQ, and diagnostic groups in both the schizophrenia and healthy control literature and did not find significant effects of these variables (Basso *et al*, 1999; Heaton *et al*, 2001). Larger practice effects are generally observed between the initial and second assessments, with smaller incremental benefits with subsequent reassessments thereafter. Interested readers may obtain additional information in the 556-page monograph of McCaffrey *et al* (2000), which is entirely comprised of tables displaying change scores at retest assessments for various tests over differing intervals in groups of healthy controls, psychiatric and neurologic patients, and medical controls.

In Table 1, we list ES values for practice effects on the first test-retest interval of about 1-3 months in very recent studies that included both schizophrenia patients and healthy controls that were not included in the meta-analysis by Szöke *et al* (2008) (Goldberg *et al*, 2007; Ahn *et al*, 2009; Crespo-Facorro *et al*, 2009). In general practice effects were comparable between controls and patients in the Goldberg

Table 1 Magnitude of Cognitive Improvement in Recent Studies of First Episode Schizophrenia and Internal Healthy Control Groups

Ahn et al.

Effect sizes of baseline to week 8 comparisons

	Intention-to-treat analysis		Per-protocol analysis	
	Patient	Control	Patient	Control
<i>K-WAIS: general intelligence^a</i>				
Full-scaled IQ	0.08	0.16	0.18	0.19
Digit span	0.07	0.08	0.08	0.01
Vocabulary	-0.02	-0.07	0.08	-0.08
Arithmetic	0.00	0.05	0.06	0.04
Picture arrangement	0.10	0.23	0.14	0.17
Block design	0.17	0.32	0.21	0.37
<i>COWAT: working memory and executive function</i>				
Correct response ^a	0.38	0.27	0.41	0.34
Perseverative response ^b	0.17	-0.04	0.17	0.00
<i>Rey-Kim memory test: verbal and nonverbal memory^a</i>				
Full-scaled MQ	0.37	0.88	0.45	0.84
AVLT sum of trails 1-5 (words)	0.38	0.95	0.50	0.97
AVLT delayed recall	0.25	0.72	0.34	0.64
AVLT delayed recognition	0.07	0.28	0.10	0.30
RCFT copy	-0.02	-0.11	0.11	-0.04
RCFT immediate recall	0.39	0.59	0.41	0.58
RCFT delayed recall: 20 min	0.42	0.56	0.41	0.59
<i>Conners' CPT: attention</i>				
Omission error ^b	-0.05	-0.11	-0.06	-0.15
Commission error ^b	-0.31	-0.20	-0.17	-0.26
Reaction time sum, ms ^b	0.04	-0.05	0.04	-0.02
SE of reaction ^b	-0.08	-0.13	-0.02	-0.18
<i>d'</i> ^a	0.17	0.18	0.03	0.28
<i>TMT: psychomotor speed^b</i>				
Trail A time, s	-0.27	-0.25	-0.18	-0.32
Trail A error	-0.17	0.13	-0.44	0.24
Trail B time, s	-0.01	-0.26	0.02	-0.34
Trail B error	-0.21	-0.10	-0.17	-0.34

Crespo-Facorro et al.

Effect sizes of baseline to 6-month comparisons

	Haloperidol	Olanzapine	Risperidone	Controls
CPT degraded stimulus ^a	0.07	0.25	0.17	-0.17
Brief test of attention ^a	0.31	0.09	0.21	0.45
Grooved pegboard ^b	-0.33	-0.37	-0.12	-0.50
Finger tapping ^a	0.39	0.03	0.19	0.03
Rey AVLT (learning) ^a	0.50	0.42	0.50	0.48
Rey AVLT (long-term recall) ^a	0.43	0.48	0.30	0.48
WAIS-III backward digits ^a	0.34	0.29	0.04	0.04
Rey complex figure test ^a	0.62	0.48	0.59	0.43

Crespo-Facorro et al., Cont.

Effect sizes of baseline to 6-month comparisons

	Haloperidol	Olanzapine	Risperidone	Controls
Trail making test B ^b	-0.19	-0.47	-0.40	-0.38
WAIS-III digit symbol ^a	0.33	0.59	0.32	0.74
FAS verbal fluency ^a	0.08	0.32	0.12	0.45
Iowa gambling ^a	-0.07	0.36	0.20	0.69

Goldberg et al.

Effect sizes of baseline to week 6 comparisons

	Patients with schizophrenia	Healthy controls
<i>Mental status</i>		
MMSE: total score	0.70	0.01
<i>Speed</i>		
Verbal fluency: words produced	-0.21	0.04
Trail making A and B: total time	0.43	0.24
WAIS-R digit symbol: scaled score	0.43	1.15
<i>Executive function</i>		
WCST: % perseveration	0.37	0.42
WCST: loss of set	-0.23	-0.13
<i>Episodic memory</i>		
CVLT: trails 1-5	0.40	0.64
CVLT: recognition errors	0.37	0.42
WMS-R: logical memory	0.74	0.79
WMS-R: visual reproduction	0.64	0.12
<i>Spatial processing</i>		
Judgment of line orientation	0.39	0.17
<i>Working memory</i>		
WMS-R digit span: sequences recalled	0.24	0.42
<i>Motor</i>		
Finger tapping: number of taps	0.12	0.13
Grooved pegboard: time to complete	0.20	0.51
<i>Symptom dimensions</i>		
SANS	0.38	N/A
Hallucinations	1.30	N/A
Delusions	0.96	N/A
Severity of illness	1.21	N/A
Disorganization	0.67	N/A

Symptom dimensions

SANS	0.38	N/A
Hallucinations	1.30	N/A
Delusions	0.96	N/A
Severity of illness	1.21	N/A
Disorganization	0.67	N/A

^aPositive indicates improvement.^bNegative indicates improvement.

et al (2007) and Crespo-Facorro *et al* (2009) studies, whereas in the Ahn *et al* (2009) study, practice effects were somewhat larger in the healthy control group. In addition, it can be seen that practice effects were ubiquitous and in the moderate range.

To summarize, there is a robust set of findings showing that practice effects are detectable, substantial, and possibly not different from healthy samples, at least, in schizophrenia patients treated with second-generation antipsychotics. In addition, we noted that in several studies in which improvement can be directly attributed to practice (eg, Keefe *et al*, 2008; Crespo-Facorro *et al*, 2009; Heaton *et al*, 2001), neuropsychological tests overlapped with those used in studies in which practice effects were minimal (eg, Harvey *et al*, 2005). As noted, this suggests that the tests may not be immune to practice effects, but that some cohorts of patients may be.

MECHANISMS OF PRACTICE EFFECTS

At first glance, practice effects may be clinically advantageous. Many activities in daily life rely on practice or repetition for optimizing performance. However, there is little evidence that improvement of this type or magnitude will generalize or transfer to other tasks. This is because a practice effect may be paradigm specific (eg, familiarity with testing instructions and demands) or content specific (eg, words on a list). For instance, massive amounts of practice on a specific action resulted in great improvement for the practiced skill (foul shooting in basketball), but not for other similar skills in the same class (Keetch *et al*, 2005). Various computational accounts of cognitive architecture are also compatible with this idea (Logan, 2002). Thus, practice effects may not reflect change in the compromised neurobiology of schizophrenia, which would then effect improvement in broad domains of cognition. Furthermore, even moderate practice effects may not compensate for baseline differences, as patients will 'start lower and end lower' than controls (who are also practicing) despite improvement. Indeed, in most studies in which patients were retested two or three times, the end point scores for the patient sample did not reach the baseline scores for the healthy control sample.

Although we appreciate the possibility that individuals may attain adequate functional ability even if they are not completely 'normal,' we believe this may be the exception rather than the rule. Bowie *et al* (2006) observed linear relationships between cognition and various measures of everyday function. Second, Goldberg *et al* (in press) demonstrated that even in a group of patients with mild cognitive impairment (amnestic subtype) the relationship between cognition and function was not sigmoidal, as had been assumed, but was linear when psychometrically appropriate performance-based measures of function were used. Lastly, we note speculatively that not all types of practice may yield the same real world benefits. Self-generated skill development that occurs in the subject's environment may have broader or larger effects.

There are several reasons to believe that patients should be able to evidence a practice effect. First, patients

demonstrate near-normal retention over delays during episodic memory (Gold *et al*, 2000; Heaton *et al*, 1994) particularly when encoding is facilitated through various input manipulations. Thus, once an item is encoded successfully, it is not subject to rapid forgetting, translating into relatively normal savings. Second, patients have relatively intact procedural learning and probabilistic learning that may be responsible for stimulus response mappings (Weickert *et al*, 2002). To the extent that some practice-related improvement may be sub-served by such learning systems, patients may be expected to benefit from setting up more efficient responding patterns even if the core abilities indexed by the test are unaffected. It has also been demonstrated that general familiarity with items, knowledge about solutions in problem-solving tasks, improvements in strategy or monitoring of responses, and reduction in load of context memory for instructions irrespective of item differences can result in practice effects, as individuals become more efficient in task-related processing.

From a theoretical perspective, while examining neurophysiological studies on practice and automatization in healthy controls, Kelly and Garavan (2005) described a variety of neurophysiological signatures of practice, which were different from initial learning. In several studies, regions engaged after practice of a task were different from those involved in initial learning (eg, for verb generation, practice reduced activation in the anterior cingulate and prefrontal cortex and increased activation in the insular and sylvian cortex (Raichle *et al*, 1994)). This suggests that the neural systems relevant for practice may be different and dissociable from those engaged by initial learning. If these systems are used in schizophrenia, they may be relatively intact and result in practice-related benefits.

PLACEBO EFFECTS

In the context of the Keefe *et al* (2008) study, the degree of observed improvement in cognitive performance with repeated assessments in a treatment trial may be a function of three factors: treatment effect, practice effect, and placebo effect. With respect to the latter, when a patient enters into a trial or is treated with a medication that is believed to contribute beneficially to cognitive performance, expectation bias can have strong effects on performance (de la Fuente-Fernández *et al*, 2002). Patients who are told that their cognitive abilities may improve may be able to perform better on test batteries used in the study because their expectations become more positive and they become more motivated, confident, and less anxious. These same factors may have an impact on a patient receives in his or her community/living situation. Future trials of cognitive-enhancing compounds could be designed in such a way to distinguish practice effects from placebo effects. In addition to an active medication group and a placebo group, the trial could include a group that receives treatment as usual without placebo; this 'practice effect only' group could be compared with the placebo group to determine whether placebo effects are active in addition to practice effects in these trials.

REDUCING PRACTICE EFFECTS

Test Construction

It is possible that tests can be constructed using certain principles from the cognitive science literature, that will substantially attenuate practice effects. A combination of multiple items, a restricted set of stimuli that serve to induce interference, and alternative and equivalent forms with different items and sequences in tests of attention, working memory, and executive function might serve to reduce practice effects to a marked degree. This view derives from findings that performance in several tests in the study by Goldberg *et al* (2007) did not improve significantly in either group. These were: CPT-identical pairs, delayed match to sample, digit span, and verbal fluency. The CPT-identical pairs test involves dozens of trials with a restricted set of stimuli (numbers). The delayed match to sample test also involves dozens of trials and a stimulus set consisting of similar nonverbalizable shapes. The digit span test involves multiple trials of numbers between one and nine (Harvey *et al* (2000) found practice-related improvements in the CPT-identical pairs test, but only after multiple daily practice sessions). One possible criticism of this approach is that it is heavily reliant on interference-based tests for assessing executive functions. However, recent study has suggested that interference suppression is a prominent feature of prefrontal cortex in managing representations (Durstewitz *et al*, 2000; Miyake and Shah, 1999). Furthermore, interference (due to similarities among trials) makes it difficult to remember specific instances (ie, items are not distinctive). For episodic memory, obligatory common encoding of items that minimize *intra-individual* changes in encoding strategy over time (a potentially important source of uncontrolled variance), followed by recognition to minimize retrieval strategies and alternate forms may minimize practice effects. However, alternate test forms in and of themselves may not be a panacea because subjects may develop 'learning to learn' strategies, as they construct strategy-based approaches in which they use semantic encoding methods or have increasing familiarity with presentation (eg, recall after delay), and test context (Beglinger *et al*, 2005; Uchiyama *et al*, 1995).

Study Design

One approach to reducing practice effects involves serial testing during a lead-in period to the trial. Underlying this approach is the assumption that practice effects involving familiarity, reduced anxiety, and procedural stimulus-response mapping will reach an asymptote, after which any gains could be attributed to the active treatment. In a study of this type (Boulay *et al*, 2007), a small number of schizophrenia patients underwent four assessments in a 4-day period while in a drug washout phase. Cognitive gains were quite large and were present in measures of short-term memory (eg, digit span forward), reaction times, attention, and executive function or cognitive control (eg, letter number span and Stroop test). In the post-randomization phase (when patients were treated with olanzapine or haloperidol), no further changes were

observed. Mozley *et al* (2008) and Falleti *et al* (2006) also used this approach. One problem with the approach is that ceiling effects could theoretically occur. However, this risk may be small in most samples of people with schizophrenia, whose performance after practice is typically not near that of healthy controls' performance at the first assessment. It is also possible that certain nonspecific sources of improvement, including those related to sculpting an efficient response at the cognitive and presumably neural level may also be diminished, and that in executive tests, problem solving or adaptation to novelty demands may be reduced such that the test no longer measures what it was designed to measure. It is also unclear psychometrically whether all tests would undergo stabilization at the same rate over multiple testings during a lead-in period (see Mozley *et al*, 2008).

Use of crossover with counterbalancing to reduce practice effects may not be without pitfalls. Crossover studies may be prone to complex carry-over effects, drug withdrawal effects, and time one-time two differences (Weickert *et al*, 2003).

A third method would employ the use of surrogate tests to match groups at baseline followed by testing at end point using the primary cognitive outcome. For instance, active and comparator groups might be matched on current IQ at baseline, given its correlation with a wide range of cognitive measures, and cognitive measures of interest (eg, memory and speed) would then be assessed only once, at the study's end point. Intelligence quotient would in effect serve as a surrogate for speed and memory tests at baseline. However, it might be difficult to conclusively rule out pre-existing group differences and analyses of repeated measures could not be performed. Furthermore, the approach may be subject to the vagaries of correlations between IQ and other cognitive domains.

A more extreme solution would be to routinely use a healthy control group in comparisons of antipsychotic medication effects on cognition in which serial testing is conducted. Nevertheless, the financial and logistical burdens of this design would probably make this approach impractical for industry.

Reliable Change Analyses

Another pragmatic approach might be the development of comprehensive norms for change with reassessment (Heaton *et al*, 2001). This would require reassessment of healthy individuals who are demographically similar to the expected characteristics of clinical trial participants with schizophrenia. They would need to be reassessed with the same assessment battery and in the same time frame as schizophrenia patients. As a consensus battery already exists for treatment studies (eg, Nuechterlein and Green, 2006), such a norming process would not be a major challenge.

This procedure would allow for the development of a 'reliable change index' to identify level of change that would exceed those expected by reassessment alone, which could be applied at the individual case level. The prevalence of subjects exceeding such a confidence interval, one that takes into account practice effects and other test variables could be compared across treatment arms. This procedure would

be simplified in the case of a standardized assessment battery, such as the MATRICS consensus cognitive battery, in which a single large-scale study could conceivably develop these norms.

Nevertheless this approach is not without problems. Practice effects may vary across retest intervals and the number of assessments, raising the issues that the precision of expected practice effect benefits may be dependent on various matching procedures. It would be desirable to have adjustments for demographic factors and to collect norms for many possible reassessments.

To make the point that change above and beyond practice effects can be large at the level of individual cases, we analyzed data provided in Table 3 of Nuechterlein *et al* (2008), which displayed Time 1 and Time 2 results of MCCB tests in a sample of multi-episode schizophrenia patients who remained on antipsychotic medication over the course of the study. We used MCCB published data because of the care taken during the data collection, the large sample, the clear tabular format, and the use of commonly administered clinical neurocognitive tests, not because we believed that that the MCCB was in any way uniquely prone to these effects. To do this, we first computed reliable change index confidence intervals using the SD of the Time 1–Time 2 difference scores in the formula (reliable change index + practice formula) advocated by Heaton *et al* (2001) using a 90% confidence interval (ie, 5% for each tail of the distribution). Exceeding the resulting confidence interval would be necessary for an individual subject to demonstrate a reliable gain above and beyond simple practice. We then used the new score to determine the magnitude of improvement using ES statistics for a given subject to compare practice-related change across various tests using a common metric. The results of this re-analysis, as shown in Table 2, suggest that most ES values for conclusively nonrandom changes on the part of individual patients retested on the MCCB tests were between 1.0 and 1.35 ES units. Thus, nonrandom cognitive enhancement detected on the individual case level would be associated with an ES gain of more than 1.0 unit. Importantly, even on those tests in which the original mean Time 1–Time 2 differences were small (eg, category fluency and CPT-IP), required ESs for nonrandom changes could be large, presumably because the SD of the difference score was large. However, even with any practice effect detected, scores on these neuropsychological tests are not at or even close to ceiling. These data also raise a theoretical and pragmatic point: high test-retest reliability in and of itself does not militate against a practice effect (eg, in the case in which all subjects improve and the rank order of subjects at Time 1 and at Time 2 is maintained).

IMPLICATIONS

At a time when the NIMH has allocated tens of millions of dollars for projects designed to assess the efficacy of adjunctive cognitive-enhancing drugs to ameliorate cognitive impairments in schizophrenia (in CATIE, MATRICS, CNTRICS, and TURNS), the possibility that the cognitive enhancement observed in clinical trials of second-generation antipsychotic medications in schizophrenia reflects practice

Table 2 Effect Sizes Based on RCI Confidence Intervals from the MCCB

Test	Change in raw score performance to exceed RCI 90% confidence interval	Effect size for definite improvement for an individual patient excluding practice effects
Category fluency	6.6	1.17
Trails A	17.16	1.36
Digit symbol	12.87	1.06
Symbol coding	10.56	1.18
CPT	1.52	1.54
CPT-IP	0.76	1.01
Digit sequence	5.12	1.26
Letter number	3.14	1.21
Spatial span	3.14	0.98
Spatial delay	13.53	1.21
NAB memory	8.09	1.16
HVLT	7.1	1.36
NAB shape learning	6.77	1.55
BVMT	10.07	1.28
Block design	11.55	1.09
Tower	7.59	1.65
Mazes	6.6	1.03
Perceived emotion	18.15	1.08
Managing emotion	13.2	1.28

effects is sobering. Therefore, we believe that the practical implications of this area are substantial. It is well known that cognitive impairment is an enduring and central feature of schizophrenia, and accounts for much of the social and vocational disability associated with the disorder. Cognitive tests now occupy a key place in many clinical trials of drugs for the treatment of schizophrenia. If the proper tools are not developed to measure cognitive change in a precise manner independent of practice effects, ie, if we ‘don’t get the tools right,’ it is possible that results of clinical trials involving cognitive enhancement may be routinely misinterpreted. This would not be ideal for the field or the consumer/patient as it could result in the registration of ineffective compounds or exclusion of medications with suitable benefits.

Our findings may also have implications for drug discovery and regulatory approval of new antipsychotic medications. We believe that the wealth of findings reviewed here will increase awareness of practice effects as potential source for cognitive change in clinical trials and that our findings can be used heuristically in the development of study designs and tests that are relatively insensitive to practice-related changes, as proposed here. Such advances might be important for improving the methodology involved in the assessment of cognitive change in clinical trials. Although we are sensitive to the issue of creating barriers to the development of cognitive-enhancing drugs, we do not believe that it is anyone’s interest to generate ambiguous or spurious results.

Hence, we recommend improvements in the psychometric aspects of the test themselves (see above on manipulations that reduce test sensitivity to practice effects), use of surrogate tests at baseline or a period of lead-in testing, or statistical analyses of change at the case level.

We recognize that the proposals for minimizing or interpreting practice or placebo effects set a higher bar for drug trials assessing cognition. Thus, studies we have reviewed cannot fully disambiguate contributions to cognitive change due to practice effects, placebo effects, pseudospecificity, and drug-induced cognitive enhancement. Nevertheless, we hope that our interpretation increases awareness of practice effects as potential source for cognitive change in clinical trials and that our suggestions can be used heuristically in the development of study designs, statistical approaches, and tests that are relatively insensitive to practice related changes. Such advances might be important for improving methodology involved in the assessment of cognitive change in clinical trials.

DISCLOSURES

Drs TE Goldberg, PD Harvey, and RSE Keefe receive royalties from the BACS, a neurocognitive instrument used in clinical trials. Dr TE Goldberg has a research grant from Pfizer. During the past year, Dr PD Harvey has served as a consultant for Merck, Dainippon Sumitomo America, Shire Pharma, Eli Lilly, and Solvay Pharma. He has a research grant from Astra Zeneca. Dr RS Goldman is employed by Pfizer. Dr DG Robinson receives grant support from Bristol-Myers Squibb and Janssen. He has received compensation from Astra Zeneca, Lundbeck, and MedAvante. Dr Keefe reports that he currently or in the past 12 months has received investigator-initiated research funding support from the National Institute of Mental Health, Allon, Novartis and the Singapore National Medical Research Council, and an unrestricted educational grant from Astra-Zeneca. He currently or in the past 12 months has received honoraria or served as a consultant or advisory board member for Abbott, Astra-Zeneca, BiolineRx, Bristol Myers Squibb, Cephalon, Dainippon Sumitomo Pharma, Eli Lilly, Johnson & Johnson, Lundbeck, Memory Pharmaceuticals, Merck, Neurosearch, Orion, Orexigen, Otsuka, Pfizer, Roche, Targacept, Sanofi/Aventis, Shire, Wyeth, and Xenoport.

REFERENCES

Ahn YM, Lee KY, Kim CE, Kim JJ, Kang DY, Jun TY *et al* (2009). Changes in neurocognitive function in patients with schizophrenia after starting or switching to amisulpride in comparison with the normal controls. *J Clin Psychopharmacol* **29**: 117–123.

Albus M, Hubmann W, Mohr F, Hecht S, Hinterberger-Weber P, Seitz NN *et al* (2006). Neurocognitive functioning in patients with first-episode schizophrenia. *Eur Arch Psychiatry Clin Neurol* **256**: 442–451.

Basso MR, Bornstein RA, Lang JM (1999). Practice effects on commonly used measures of executive function across twelve months. *Clin Neuropsychol* **13**: 283–292.

Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J *et al* (2005). Practice effects and the use of alternate

forms in serial neuropsychological testing. *Arch Clin Neuropsychol* **20**: 517–529.

Boulay LJ, Labelle A, Bourget D, Robertson S, Habib R, Tessier P *et al* (2007). Dissociating medication effects from learning and practice effects in a neurocognitive study of schizophrenia: olanzapine versus haloperidol. *Cogn Neuropsychiatry* **12**: 322–338.

Bowie CR, Reichenberg A, Patterson TL, Heaton RK, Harvey PD (2006). Determinants of real world functional performance in schizophrenia subjects: correlations with cognition, functional capacity, and symptoms. *Am J Psychiatry* **163**: 418–425.

Crespo-Facorro B, Rodríguez-Sánchez JM, Pérez-Iglesias R, Mata I, Ayesa R, Ramirez-Bonilla M *et al* (2009). Neurocognitive effectiveness of haloperidol, risperidone, and olanzapine in first-episode psychosis: a randomized, controlled 1-year follow-up comparison. *J Clin Psychiatry* **70**: 717–729.

Davidson M, Galderisi S, Weiser M, Werbeloff N, Fleischhacker WW, Keefe RS *et al* (2009). Cognitive effects of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: a randomized, open-label clinical trial (EUFEST). *Am J Psychiatry* **166**(6): 675–682.

de la Fuente-Fernández R, Schulzer M, Stoessl AJ (2002). The placebo effect in neurological disorders. *Lancet Neurol* **1**: 85–91.

Durstewitz D, Seamans JK, Sejnowski TJ (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J Neurophysiol* **83**: 1733–1750.

Falletti MG, Maruff P, Collie A, Darby DG (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol* **28**: 1095–1112.

Gold JM, Rehkemper G, Binks SW, Carpenter CJ, Fleming K, Goldberg TE *et al* (2000). Learning and forgetting in schizophrenia. *J Abnorm Psychol* **109**: 534–538.

Goldberg TE, Goldman RS, Burdick KE, Malhotra AM, Lencz T, Patel RC *et al* (2007). Cognitive improvements after treatment with second-generation antipsychotic medications in first episode schizophrenia: is it a practice effect? *Arch Gen Psychiatry* **64**: 1115–1122.

Goldberg TE, Green MF (2002). Neurocognitive functioning in patients with schizophrenia: an overview. In: Davis KL, Charney D, Coyle JT, Nemeroff C (eds). *Neuropsychopharmacology: The Fifth Generation of Progress*. Raven Press: New York. pp 657–670.

Goldberg TE, Koppel J, Kechlisen L, Christen E, Werringloer U, Conejero-Goldberg C *et al*. Performance-based measures of everyday function in mild cognitive impairment. *Am J Psychiatry* (in press).

Green MF (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry* **153**: 321–330.

Harvey PD, Howanitz E, Parrella M, White L, Davidson M, Mohs RC *et al* (1998). Symptoms, cognitive functioning, and adaptive skills in geriatric patients with lifelong schizophrenia: a comparison across treatment sites. *Am J Psychiatry* **155**: 1080–1086.

Harvey PD, Keefe RS (2001). Studies of cognitive change in patients with schizophrenia following novel antipsychotic treatment. *Am J Psychiatry* **158**(2): 176–184.

Harvey PD, Koren D, Reichenberg A, Bowie CR (2006). Negative symptoms and cognitive deficits: what is the nature of their relationship? *Schizophr Bull* **32**: 250–258.

Harvey PD, Moriarty PJ, Serper MR, Schnur E, Lieber D (2000). Practice-related improvement in information processing with novel antipsychotic treatment. *Schizophr Res* **46**: 139–148.

Harvey PD, Palmer BW, Heaton RK, Mohamed S, Kennedy J, Brickman A (2005). Stability of cognitive performance in older patients with schizophrenia: an 8-week test-retest study. *Am J Psychiatry* **162**: 110–117.

- Heaton RK, Gladsjo JA, Palmer BW, Kuck J, Marcotte TD, Jeste DV (2001). Stability and course of neuropsychological deficits in schizophrenia. *Arch Gen Psychiatry* 58: 24–32.
- Heaton RK, Paulsen JS, McAdams LA, Kuck J, Zisook S, Braff D et al (1994). Neuropsychological deficits in schizophrenics. Relationship to age, chronicity, and dementia. *Arch Gen Psychiatry* 51: 469–476.
- Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD et al (2001). Detecting change: a comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychol* 16: 75–91.
- Hill SK, Shuepbach D, Herbener ES, Keshavan MS, Sweeney JA (2004). Pretreatment and longitudinal studies of neuropsychological deficits in antipsychotic-naïve patients with schizophrenia. *Schizophr Res* 68: 49–63.
- Hoff AL, Sakuma M, Wieneke M, Horon R, Kushner M, DeLisi LE (1999). Longitudinal neuropsychological follow-up study of patients with first-episode schizophrenia. *Am J Psychiatry* 156: 1336–1341.
- Keefe RS, Bilder RM, Davis SM, Harvey PD, Palmer BW, Gold JM et al (2007). Neurocognitive effects of antipsychotic medications in patients with chronic schizophrenia in the CATIE Trial. *Arch Gen Psychiatry* 64: 633–647.
- Keefe RS, Malhotra AK, Meltzer H, Kane JM, Buchanan RW, Murthy A et al (2008). Efficacy and safety of donepezil in patients with schizophrenia or schizoaffective disorder: significant placebo/practice effects in a 12-week, randomized, double-blind, placebo-controlled trial. *Neuropsychopharmacology* 33: 1217–1228.
- Keefe RS, Perkins DO, Gu H, Zipursky RB, Christensen BK, Lieberman JA (2006). A longitudinal study of neurocognitive function in individuals at-risk for psychosis. *Schizophr Res* 88: 26–35.
- Keefe RS, Silva SG, Perkins DO, Lieberman JA (1999). The effects of atypical antipsychotic drugs on neurocognitive impairment in schizophrenia: a review and meta-analysis. *Schizophr Bull* 5: 201–222.
- Keetch KM, Schmidt RA, Lee TD, Young DE (2005). Especial skills: their emergence with massive amounts of practice. *J Exp Psychol Hum Percept Perform* 31: 970–978.
- Kelly AM, Garavan H (2005). Human functional neuroimaging of brain changes associated with practice. *Cereb Cortex* 15: 1089–1102.
- Logan GD (2002). An instance theory of attention and memory. *Psychol Rev* 109: 376–400.
- McCaffrey RJ, Duff K, Westervelt HJ (2000). *Practitioner's Guide to Evaluating Change with Neuropsychological Testing Instruments*. Kluwer Academic: New York.
- Mishara A, Goldberg TE (2004). A meta-analysis of the effects of typical neuroleptic medications on cognition: reopening the book. *Biol Psych* 55: 1013–1022.
- Miyake A, Shah P (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press: New York.
- Mohamed S, Rosenheck R, Swartz M, Stroup S, Lieberman JA, Keefe RS (2008). Relationship of cognition and psychopathology to functional impairment in schizophrenia. *Am J Psychiatry* 165: 940–943.
- Mozley LH, Verma A, Vogt R, Gargano C, Potter W, Egan MF (2008). Comparison of cognition batteries for use in clinical trials of schizophrenia. ACNP Annual Meeting. Scottsdale, AZ, USA.
- Nuechterlein KH, Green MH (2006). *MATRICES Consensus Cognitive Battery*. MATRICS Assessment: Los Angeles.
- Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD et al (2008). The MATRICS consensus cognitive battery, Part 1: test selection, reliability, and validity. *Am J Psychiatry* 165: 203–213.
- Raichle ME, Fiez JA, Videen TO, MacLeod AM, Pardo JV, Fox PT et al (1994). Practice related changes in human brain functional anatomy during nonmotor learning. *Cereb Cortex* 4: 8–26.
- Salthouse TA, Schroeder DH, Ferrer E (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Dev Psychol* 40: 813–822.
- Szöke A, Trandafir A, Dupont ME, Méary A, Schürhoff F, Leboyer M (2008). Longitudinal studies of cognition in schizophrenia: meta-analysis. *Br J Psychiatry* 192: 248–257.
- Uchiyama CL, D'Elia LF, Dellinger AM, Becker JT, Selnes OA, Wesch JE et al (1995). Alternate forms of the auditory-verbal learning test: issues of test comparability, longitudinal reliability, and moderating variables. *Arch Clin Neuropsychol* 10: 133–145.
- Velligan DI, Mahurin RK, Diamond PL, Hazleton BC, Eckert SL, Miller AL (1997). The functional significance of symptomatology and cognitive function in schizophrenia. *Schizophr Res* 25: 21–31.
- Weickert TW, Goldberg TE (2005). First- and second-generation antipsychotic medication and cognitive processing in schizophrenia. *Curr Psychiatry Rep* 7: 304–310.
- Weickert TW, Goldberg TE, Marengo S, Bigelow LB, Egan MF, Weinberger DR (2003). Comparison of cognitive performances in a placebo period and an atypical neuroleptic treatment period in schizophrenia. *Neuropsychopharmacology* 28: 1491–1500.
- Weickert TW, Terrazas A, Bigelow LB, Malley JD, Hyde T, Egan MF et al (2002). Habit and skill learning in schizophrenia: evidence of normal striatal processing with abnormal cortical input. *Learn Mem* 9: 430–442.
- Woodward ND, Purdon SE, Meltzer HY, Zald DH (2005). A meta-analysis of neuropsychological change to clozapine, olanzapine, quetiapine, and risperidone in schizophrenia. *Int J Neuropsychopharm* 8: 457–472.
- Woodward ND, Purdon SE, Meltzer HY, Zald DH (2007). A meta-analysis of cognitive change with haloperidol in clinical trials of atypical antipsychotics: dose effects and comparison to practice effects. *Schizophr Res* 89: 211–224.