Published in final edited form as:

Schizophr Res. 2014 October ; 159(1): 182–187. doi:10.1016/j.schres.2014.07.032.

Detecting reliable cognitive change in individual patients with the MATRICS Consensus Cognitive Battery

Bradley E. Gray, M.A.^a, Robert P. McMahon, Ph.D.^a, Michael F. Green, Ph.D.^b, Larry J. Seidman, Ph.D.^c, Raquelle I. Mesholam-Gately, Ph.D.^c, Robert S. Kern, Ph.D.^b, Keith H. Nuechterlein, Ph.D.^b, Richard S. Keefe, Ph.D.^d, and James M. Gold, Ph.D.^a

^aUniversity of Maryland School of Medicine, Department of Psychiatry, Maryland Psychiatric Research Center. PO Box 21247, Baltimore, Maryland 21228, USA

^bSemel Institute for Neuroscience and Human Behavior at UCLA, and VA Greater Los Angeles Healthcare System, Los Angeles, CA 90073, USA

^cMassachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, and Harvard Medical School Department of Psychiatry, Boston, Massachusetts, 02115, USA

^dDuke University School of Medicine, Department of Psychiatry, Translational Neuroscience. Durham, NC 27710, USA

Abstract

Objective—Clinicians often need to evaluate the treatment response of an individual person, and to know that observed change is true improvement or worsening beyond usual week to week changes. This paper gives clinicians tools to evaluate individual changes on the MATRICS Consensus Cognitive Battery (MCCB). We compare three different approaches: a descriptive analysis of MCCB test-retest performance with no intervention, a reliable change index (RCI) approach controlling for average practice effects, and a regression approach.

Method—Data were gathered as part of the MATRICS PASS study (Nuechterlein et al., 2008). A total of 159 people with schizophrenia completed the MCCB at baseline and four weeks later. Data were analyzed using an RCI and a regression formula establishing confidence intervals.

Results—The RCI and regression approaches agree within one point when baseline values are close to the sample mean. However, the regression approach offers more accurate limits for expected change at the tails of the distribution of baseline scores.

© 2014 Elsevier B.V. All rights reserved.

Corresponding Author: James Gold. +1410-402-7871. jgold@mprc.umaryland.edu. MPRC, PO Box 21247, Baltimore, Maryland 21228, USA.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Contributors.

Authors Green, Seidman, Kern, Nuechterlein, Keefe and Gold designed the original protocol. Authors Gray and Gold did the literature review. Authors Gray, Gold, and McMahon did the statistical analyses. Author Gray wrote the first draft and all authors contributed to and approved the final manuscript.

Conclusions—Though both approaches have their merits, the regression approach provides the most accurate measure of significant change across the full range of scores. As the RCI does not account for regression to the mean and has confidence limits which remain constant across baseline scores, the RCI approach effectively gives narrower confidence limits around an inaccurately predicted average change value. Further, despite the high test-retest reliability of the MCCB, a change in an individual's score must be relatively large to be confident that it is beyond normal month to month variation.

Introduction

The MATRICS Consensus Cognitive Battery (MCCB) was developed for use in clinical trials to evaluate the effectiveness of treatments for cognitive impairment in people with schizophrenia (Nuechterlein et al., 2008). Randomized clinical trials inform clinicians about average improvements in cognitive function at the group level, and therefore provide guidance about treatments that are most likely to be effective. However, in everyday practice, clinicians need to evaluate the cognitive changes of an individual person, whether improvement or worsening, and determine if the observed change can be reliably attributed to introduction of a new treatment or worsening of underlying health status, as opposed to day to day performance variations or practice effects.

A variety of statistical approaches have been developed to assist clinicians in evaluating the impact of an intervention on an individual patient. One commonly used method of testing for significant cognitive change is the reliable change index (RCI), which analyzes the test-retest performance of a group of subjects in the absence of any intervention (Temkin et al, 1999). Any changes observed in an untreated group are thought to reflect measurement error (i.e., imperfect reliability). Beyond measurement error, people often score significantly higher upon retesting as a function of practice. Practice effects come about from a familiarity with test instructions, demands, or content, and it is common for individuals to obtain higher scores on many measures upon repeated testing. Such practice effects can potentially confound the interpretation of results from interventions designed to enhance cognitive performance, as one cannot separate the extent of improvement due to treatment from improvement that would occur without treatment. Further, in interpreting testing from a single individual, one does not have an untreated control group to help separate practice from intervention effects.

The distinction between practice effects and treatment effects has been an important issue in the schizophrenia literature regarding the cognitive effects of second generation antipsychotics. Goldberg et al. (2007) showed that first-episode patients treated with second generation antipsychotics had a composite effect size (Cohen's d) gain of 0.36 upon repeat testing. However, a comparison sample of healthy control had an observed effect size gain of 0.33. Thus, the effect of treatment did not exceed expected practice effects (although it is possible that treatment facilitated such practice effects). Practice effects are also important in the interpretation of stable performance after the introduction of an intervention. If practice-related gains are expected without intervention, observing stable performance upon retesting might actually be evidence of intervention-related suppression of practice effects. To

account for this, Temkin et al. (1999) suggested adjustments to the RCI to account for practice effects, which is discussed in more detail below.

The RCI formula accounting for practice effects in addition to lack of reliability provides a good representation of the magnitude, on average, that an individual's score at time 2 is expected to differ from his or her score at time 1. However, this formula assumes that the expected change and confidence limits from baseline to Time 2 are the same across all levels of baseline performance. Importantly, the RCI approach fails to account for the effect of regression to the mean, such that especially high and especially low baseline scores are more likely to be closer to the average score at Time 2. This effect results in larger changes for individuals whose baseline scores are further from the mean. Those who are far below the mean are likely to improve more and those who are far above the mean are likely to improve less (or even decrease). The effect of regression to the mean may be particularly important for interpreting change in everyday clinical practice as clinicians may be more motivated to try cognitive enhancing treatments with their most impaired patients.

An alternative to the RCI that does account for these issues is linear regression, which uses baseline scores to predict changes from baseline and to calculate confidence limits around that prediction (Crawford & Howell, 1998). This method accounts for learning effects and regression to the mean. Unlike the RCI, which produces the same confidence limits regardless of the baseline score, the regression confidence limits become wider the further a person's baseline value is from the overall mean, reflecting the greater uncertainty that exists when predicting extreme values.

The goal of the analyses presented below is to provide clinicians with tools that can be used to evaluate change on the MCCB at the level of an individual patient, using data from a large sample of representative patients on stable medication as a comparison group to predict future scores from past performance. We provide information about the use of 3 approaches: 1) first, we present summary data on descriptive analyses of MCCB test-retest performance (in the absence of any intervention), establishing the overall frequency of observed points of rarity for the amount of change seen in this particular study population; 2) second, we present the results from the RCI approach where the expected effects of practice are used to increase the precision of that approach; and 3) we present the results from the regression approach.

Method

Participants

These data were gathered as part of the MATRICS PASS study (Green et al., 2008, Nuechterlein et al., 2008). One hundred seventy-six people with Schizophrenia (PSZ) were tested at baseline across five study sites, and 167 were reassessed four weeks later (a 95% retention rate), with all subtests administered in the same order each time. Due to some PSZ having incomplete data, 8 subjects were removed, providing a final sample size of 159 (58% White, 29% African-American, 13% other). Eighty-six percent of participants received a diagnosis of schizophrenia, with the remaining 14% receiving a diagnosis of schizoaffective

disorder- depressed type, as confirmed by the Structured Clinical Interview for DSM-IV (SCID; First et al., 2002).

Measures

The data and methods for this study are described in detail by Nuechterlein et al. (2008).

Procedure

After receiving a description of the study, participants provided written informed consent based on the guidelines of each study site's institutional review board and the coordinating site. Participants then provided medical history and were given the SCID to confirm diagnosis. Following this, participants completed the MCCB, with testing taking approximately 1 to 1.5 hours. Participants were then asked to return after a 4-week period for a retest.

Data Analyses

The RCI formula that accounts for practice effects and lack of reliability is based on computing an average practice effect over all subjects who completed a given neuropsychological test. An individual is considered to have improved if the change from baseline on their T-score at Time 2 is greater than the average change + $(1.64 \times SD_{diff})$, where SD_{diff} is the standard deviation of test-retest differences (6). Alternatively, an individual is considered to have worsened if the change from baseline to Time 2 is less than the average change - $(1.64 \times SD_{diff})$. The constant 1.64 is the normal z-score corresponding to a two-sided 90% confidence limit (i.e., 5% at the top and bottom of the distribution) for changes in a new individual with schizophrenia, based on data from the current sample. Note that we used age-and gender-corrected T scores, which are the default in the MCCB computer scoring program. We calculated the RCI of each domain by first computing the average change from baseline to Time 2 for each domain. We then calculated the 90% confidence interval for individual changes of the RCI using the formula (average change $\pm (1.64 \times SD_{diff})$).

For the regressions, we used the linear regression formula (predicted change from baseline = intercept + $(b \times \text{baseline value})$) to predict the expected value of change from baseline, given an individual's baseline value, and then calculated appropriate upper and lower 90% confidence limits for new individual predictions around it (see online supplements, with formulas provided in S.2). An individual would be declared improved if their performance was above the upper limit and worsened if it was below the lower limit. In Table 2, we also report 80% confidence limits but focus our discussion on the 90% limits as these are standard in the literature.

Statistical calculations of confidence limits treat the variables as continuous numbers, but many cognitive test scores, like the MCCB T scores, are reported as integers. Due to this, Crawford and Howell (1998) recommend using "operating limits" for change by rounding confidence bounds to the greatest integers that do not exceed the continuous limits for "no change" (7). For example, if the upper RCI 90% confidence boundary, calculated from (mean score $\pm 1.64 \times \text{standard deviation}$), is equal to 11.7, a score of 11 would be accepted

as “no change,” but a score of 12 would be rejected. In this instance, 11 is the upper operating limit for no change. The same reasoning applies to regression-based confidence limits. In the analyses below we calculated operating limits for the RCI and regression-based confidence intervals, to reflect actual possible scores on the MCCB and simplify their use in practice.

Results

One hundred fifty-nine PSZ had full MCCB data at both visits; their mean (s.d.) age was 44.3 (11.0); their mean education was 12.4 (2.4) years, and average maternal and paternal education was 12.2 (3.2) and 12.0 (4.2) years, respectively. Seventy-four percent were male, and 59% were White. In Table 1, MCCB Composite and domain T-scores are presented at Times 1 and 2. Mean Time 2 - Time 1 differences were positive for all T scores except for Social Cognition (Mean difference = -0.77). The mean difference was statistically significant for the Composite score and all domains except Visual Learning, $t(158) = -0.82$, $p = .42$ and Social Cognition, $t(158) = 1.05$, $p = .30$. The Pearson correlation between the Composite score at Time 1 and Time 2 was $r = 0.91$, $p < .001$, and the intra-class correlation (ICC) was 0.91 , $p < .001$, evidence that the MCCB Composite score is a highly reliable measure (1).

Figure 1 shows the cumulative percentages of PSZ with MCCB change scores less than or equal to a given value. Sixty-one percent of participants had difference scores between 0 and 10 points. Five percent of participants had change scores -10 or $+12$, and 10% of participants had change scores -7 or $+9$. This empirical distribution of change scores is a simple non-parametric method of estimating “reliable change,” which does not assume an underlying normal distribution as the RCI does. Thus, from the empirical distribution in this large sample, new individual cases with retest scores that are 7 points or more below baseline are potential evidence of cognitive decline, while retest scores that are 9 points or more above baseline might suggest an intervention-related benefit.

Table 1 presents the mean (s.d.) change for the Composite score and each domain, as well as the 90% confidence intervals of the RCI. The mean change is used as an estimate of the expected practice effect from repeating a test. As seen in the table, the width of the 90% confidence intervals varies substantially among the different domain scores, reflecting differences in the test-retest reliabilities of these domains. These RCI confidence limits can be applied in a straightforward manner. For example, for the MCCB Composite score, declines of greater than 7 points or increases of more than 10 points are unlikely to be attributable to measurement error or practice effects and, in the context of an intervention, may reflect the impact of the intervention. We note that the 90% RCI limits are nearly identical to the corresponding limits from the empirical distribution function displayed in Figure 1.

Table 2 provides regression-based operating limits for the MCCB Composite T-score (rounding the upper 90% confidence limit to the next lowest integer, the lower limit to the next highest integer) for baseline values going from 1 to 60 (See Supplementary Tables 1 and 2 for operating limits for the individual domain scores which may be useful if there is

reason to suspect a heterogeneous cognitive response to an intervention). Unlike the RCI 90% limits, the width of these operating limits depends on how far an individual's baseline score is above or below the mean baseline MCCB Composite score (28.92) in the data set used to derive these estimated limits, reflecting greater uncertainty predicting future observed change scores farther from the center of the distribution. The regression-based operating limits also assume that, due to regression to the mean, individuals with low scores at baseline are more likely to increase than decrease, and individuals with high scores at baseline are more likely to decrease than increase.

In figure 2, a scatterplot of individual MCCB Composite T change and baseline scores is displayed with the RCI and regression-based confidence limits for individual change superimposed on top. The two approaches for calculating confidence limits for individual change agree closely near the baseline mean, but diverge away from that point. The regression has a slight negative slope indicating that larger, positive difference scores are more frequent without intervention at lower levels of baseline performance while smaller positive and even negative difference scores are more frequent at higher levels of baseline performance. Thus, at the tails of the distribution there are differences between the RCI and regression approaches. In essence, by not taking account of regression to the mean, the RCI approach will “underdetect” evidence of worsening in cognitive performance and “overdetect” cognitive improvement among individuals with lower scores at baseline, and the RCI approach will “underdetect” improvement and “overdetect” cognitive decline among individuals with higher scores at baseline.

Differences in which change scores the two approaches would consider “reliable” are more easily discernible for some of the domain T-scores that have lower ICCs than for the MCCB Composite T (see Supplementary Tables 1 and 2 for a comparison of RCI and regression based confidence limits for individual domains). An example is displayed in the graphical representation of the Social Cognition domain (Figure 3), where the regression confidence limits clearly follow the overall pattern of the data more closely than the RCI confidence limits.

Discussion

The MCCB composite score is a highly reliable test, with a high correlation ($r > 0.9$) between repeated measurements, making it a sensitive instrument for detecting cognitive changes when comparing treatment groups. However, even with such a reliable test, deciding whether an individual with schizophrenia has improved or worsened in cognition is more difficult due to the remaining random variation in individual change scores, practice effects, and regression to the mean.

We used test-retest scores from 159 PSZ to examine the behavior of three statistical approaches to setting confidence limits for identifying “reliable” changes: 1) using the upper and lower 5% of the empirical distribution of change scores; 2) using the approximately normal distribution of the T-scores to calculate a Reliable Change Index (RCI) from the mean change score $\pm (1.64 \times SD_{diff})$; and 3) using linear regression to obtain confidence limits on predicted changes for new individuals, given their baseline scores. Only the third

approach takes account of regression to the mean. For individual participants, substantial improvements, on the order of 10 T score units are needed for a change in score to be considered clearly beyond measurement error. While a more lenient confidence limit would accept a smaller change score as exceeding likely measurement error, the fact remains that a surprising degree of individual change is needed to be considered reliable even for a test with excellent reliability.

In the current data set, the two-sided 90% confidence limits from the empirical distribution function agreed closely with those set by the RCI. However, because the empirical distribution upper and lower limits are set primarily by a few extreme values in the tails, they are likely to be less stably estimated than the RCI. For many test subjects, there is a good deal of agreement across the different statistical approaches. For the majority of Time 1 MCCB Composite scores, the RCI and regression approaches would identify a Time 2 change score in the same way (i.e., as either unchanged or changed). This level of agreement comes, in part, from the high test-retest reliability of the MCCB Composite score. Test-retest reliability varies among the individual domains of the MCCB, and for individual domains with reduced reliability, effects of regression to the mean -- and consequent disagreements with the RCI over whether an individual score is significantly changed or not -- will be greater.

The regression approach is more useful than the RCI approach as scores get further from the baseline mean T score (approximately 29 points on the MCCB Composite score). Regression to the mean causes a significant shift in the expected Time 2 score for both low and high Time 1 scores, and ignoring this influence could lead to faulty conclusions in patients with initially high or low scores. At the low end of the distribution, practice effects and regression to the mean are pushing scores in the same direction, such that an individual with a low Time 1 score is likely to improve by more T-score points than a person with an average-high range Time 1 score. To account for this, the confidence limit range must be shifted up at the low end of the distribution so that cognitive improvement is not “overdetected”. At the high end of the distribution, practice effects and regression to the mean begin to counter each other, so it is not the case that individuals with high scores at Time 1 will always achieve a higher score at Time 2. To account for this conflict, the confidence limit range must be shifted down to not “underdetect” improvement or “overdetect” decline. For individuals with baseline scores not very close to the overall mean (of the patient sample) baseline MCCB Composite score of PSZ, the limits in Table 2 provide a more accurate way than the RCI to determine whether an individual with schizophrenia is helped (by exceeding the upper confidence limit) or harmed (by exceeding the lower confidence limit) by an intervention. We also encourage clinicians to use the data in the supplementary tables, particularly when evaluating the retest performance of subjects who have scores more than one standard deviation below or above the patient mean (i.e., MCCB Composite T scores of 20 and 40).

One limitation of these results is that the confidence limits given for both RCI and regression-based approaches may not generalize to situations such as clinical trials or physician treatment choices in which people are selected for a study or a treatment based on severity of cognitive deficit (e.g., inclusion criteria of a baseline MCCB Composite T-score

1 s.d. below the healthy control mean T-score of 50, which would likely result in lower overall mean scores than observed in this sample). In that setting, regression to the mean effects would be increased and the standard deviation of the change scores would be larger than what would be seen in an unselected sample of PSZ, such as that used in the current study (Davis, 1976, Follman, 1991)).

These data are most applicable to the first and second administrations of the MCCB and one would expect smaller practice effects with subsequent testing occasions. Smaller improvements might be reliable upon further repeat testing, but changes of the magnitude described here will certainly be reliable. The MATRICS PASS study cohort was recruited from 5 sites around the country. Each group of patients had mild-moderate levels of symptom severity and was thought to be representative of the types of patients who might be candidates for cognitive enhancing treatments. Our estimated practice effect of 1.4 points for the Composite score closely resembles the 2.2 point gain reported by Keefe et al. (2011) in a sample of 323 clinical trial participants tested twice over 7–21 days, and their ICC of .88 is very similar to the .91 observed in our group, increasing confidence that our data will likely generalize. Notably, our patients were not tested in the context of a treatment trial and possible placebo effects. However, placebo responses of only 2.4 and 3.2 points were observed in the studies by Buchanan et al.,(2011). and Javitt et al., (2012) respectively. While the same precise confidence intervals might not be replicated in another similar sample, or a sample with different clinical features, we suspect that the broad guidelines provided here will prove to be reliable. The present results are a function of the reliability of the MCCB, the observed distribution of change scores, and the operation of regression to the mean. While the current data from the PASS study were obtained on two occasions, one month apart, the reliability of the MCCB composite score remained similarly high (ICC=0.93) on 3 test occasions over 12 weeks in the placebo group in the Javitt et al trial (2012). Thus, we think clinicians can use these data with a fair degree of confidence even when evaluating patients who are somewhat different from those studied in this cohort (such as first episode patients once they have been stabilized clinically). As in all clinical decision-making, one can adjust statistical thresholds given the types of risks involved in the decision. For example, one might accept a smaller negative change score as evidence of likely harm if the intervention is a drug where a negative effect is plausible given the mechanism of action. Similarly, one might accept a smaller positive change score, using the 80% confidence interval, as evidence of benefit if there were no risks associated with the treatment. The data provided here can serve as useful guidelines in making such decisions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Role of Funding Source

This work was made possible by NIMH contract N01MH22006 provided to the University of California, Los Angeles (Drs. Marder, Green, and Fenton); and an option (Drs. Green and Nuechterlein) to the NIMH MATRICS initiative.

Conflict of Interest

Mr. Gray reports no support from commercial interests, and has no competing interests. Dr. McMahon has been a consultant for Amgen, Inc. within the past 36 months. Dr. Green has been a consultant to AbbVie, Biogen, DSP, and Roche, and he is on the scientific advisory board of Mnemosyne. He has received research funds from Amgen, Inc. Dr. Seidman reports no support from commercial interests. He receives research support from the Massachusetts Department of Mental Health as Principal Investigator (SCDMH82101008006), from NIMH as Principal Investigator (PI) or site PI, or PI of site subcontract (1 U01 MH081928-01A1, 2 R01 MH065571, 1 R21 MH091461-01A1, 1R21 MH092840-01A1, 1R21 MH093294-01A1, R01 MH092440, R01 MH096027-01, 1R01MH092380-01A1, R01 MH 101052-01, R01 MH103831), or Investigator (1R01MH096942-01A, R01 HD067744-01A1), and from the Sidney R. Baer, Jr. Foundation for clinical program development. Dr. Mesholam-Gately reports no support from commercial interests, and has no competing interests. Dr. Kern is an officer for MATRICS Assessment, Inc. and receives financial compensation for his role in that non-profit organization. Dr. Nuechterlein has research grants from Janssen Scientific Affairs, Genentech, and Posit Science and has served as a consultant to Genentech and Otsuka. Dr. Richard Keefe currently or in the past 3 years has received investigator-initiated research funding support from the Department of Veteran's Affairs, Feinstein Institute for Medical Research, GlaxoSmithKline, National Institute of Mental Health, Novartis, Psychogenics, Research Foundation for Mental Hygiene, Inc., and the Singapore National Medical Research Council. He currently or in the past 3 years has received honoraria, served as a consultant, or advisory board member for Abbvie, Akebia, Amgen, Astellas, Asubio, AviNeuro/ChemRar, BiolineRx, Biomarin, Boehringer-Ingelheim, Eli Lilly, EnVivo, GW Pharmaceuticals, Helicon, Lundbeck, Merck, Mitsubishi, Novartis, Otsuka, Pfizer, Roche, Shire, Sunovion, Takeda, Targacept. Dr. Keefe receives royalties from the BACS testing battery and the MATRICS Battery (BACS Symbol Coding). He is also a shareholder in NeuroCog Trials, Inc. and Sengenix. Dr. Gold has been a consultant for Amgen and Pfizer, served on an advisory board for Hoffman LaRoche, and receives royalty payments from the BACS.

References

- Buchanan RW, Keefe RS, Lieberman JA, Barch DM, Csernansky JG, Goff DC, Gold JM, Green MF, Jackson LF, Javitt DC, Kimhy D, Kraus MS, McEvoy JP, Mesholam-Gately RI, Seidman LJ, Ball MP, McMahon RP, Kern RS, Robinson J, Marder SR. A randomized clinical trial of MK-0777 for the treatment of cognitive impairments in people with schizophrenia. *Biol Psychiatry*. 2011; 69:442–449. [PubMed: 21145041]
- Crawford JR, Howell DC. Regression equations in clinical neuropsychology: An evaluation of statistical methods for comparing predicted and obtained scores. *J Clin Exp Neuropsychol*. 1998; 20:755–762. [PubMed: 10079050]
- Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol*. 1976; 104:493–498. [PubMed: 984023]
- First, MB.; Spitzer, RL.; Miriam, G.; Williams, JBW. Structured Clinical Interview for DSM–IV–TR Axis I Disorders, Research Version, Patient Edition With Psychotic Screen. New York, NY: New York State Psychiatric Institute, Biometrics Research; 2002.
- Follman DA. The effect of screening on some pretest-posttest test variances. *Biometrics*. 1991; 1(47): 763–771.
- Goldberg TE, Goldman RS, Burdick KE, Malhotra AK, Lencz T, Patel RC, Woerner MG, Schooler NR, Kane JM, Robinson DG. Cognitive improvement after treatment with second-generation antipsychotic medications in first-episode schizophrenia: is it a practice effect? *Arch Gen Psychiatry*. 2007; 64:1115–1122. [PubMed: 17909123]
- Green MF, Nuechterlein KH, Kern RS, Baade LE, Fenton WS, Gold JM, Keefe RS, Mesholam-Gately RI, Seidman LJ, Stover E, Marder SR. Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *Am J Psychiatry*. 2008; 165:221–228. [PubMed: 18172017]
- Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcott TD, Grant I. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychol*. 2001; 16:75–91. [PubMed: 14590193]
- Javitt DC, Buchanan RW, Keefe RS, Kern R, McMahon RP, Green MF, Lieberman J, Goff DC, Csernansky JG, McEvoy JP, Jarskog F, Seidman LJ, Gold JM, Kimhy D, Nolan KS, Barch DS, Ball MP, Robinson J, Marder SR. Effect of the neuroprotective peptide davunetide (AL-108) on cognition and functional capacity in schizophrenia. *Schizophr Res*. 2012; 136:25–31. [PubMed: 22169248]

- Keefe RS, Fox KH, Harvey PD, Cucchiaro J, Siu C, Loebel A. Characteristics of the MATRICS consensus cognitive battery in a 29-site antipsychotic schizophrenia clinical trial. *Schizophr Res.* 2011; 125:161–168. [PubMed: 21075600]
- Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, Essock S, Fenton WS, Frese FJ, Gold JM, Goldberg TE, Heaton RK, Keefe RS, Kraemer H, Mesholam-Gately RI, Seidman LJ, Stover E, Weinberger DR, Young AS, Zalcman S, Marder SR. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry.* 2008; 165:203–213. [PubMed: 18172019]
- Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc.* 1999; 5:357–369. [PubMed: 10349298]

Figure 1.

Frequency distribution of change scores. An estimated 90% confidence interval on individual changes using the empirical distribution would run from -7 (percentile = 5.1) to $+9$ (percentile = 94.9).

Figure 2.

Actual (MCCBChange) and regression-predicted (MCCBPredChange) changes in MCCB Composite Score from Time1 to Time 2 by score at Time 1, together with two-sided 90% confidence limits for individual changes for the Reliable Change Index (RCI) and regression-based confidence limits. For low values of the Composite score at Time 1, observed changes at Time 2 are on average higher than the mean change (used to calculate the RCI), but observed changes at Time 2 are lower on average than the mean change for high values of the Composite score. Regression based confidence limits reflect this behavior than the RCI.

Figure 3.

Actual (SCChange) and regression-predicted changes (SCPredChange) in MCCB Social Cognition score from Time1 to Time 2 by score at Time 1, together with two-sided 90% confidence limits for individual changes for the Reliable Change Index (RCI) and regression-based confidence limits. For low values of the Social Cognition score at Time 1, observed changes at Time 2 are on average higher than the mean change (used to calculate the RCI), but observed changes at Time 2 are lower on average than the mean change for high values of the Social Cognition score. Both RCI and regression-based confidence limits for changes in Social Cognition are wider than the limits for the MCCB Composite score (Figure 2), reflecting the lower correlation between repeated measures of Social Cognition ($r=0.72$) compared to the Composite score ($r=0.91$).

Table 1

Mean (SD) of MCCB Composite T Score and Domains T scores, Estimates of Practice Effects, and Two-Sided 90% Confidence Limits for Individual Change

MCCB Performance	R	Time 1		Time 2		Change score		Practice Effect	RCI (90%)
		Mean	SD	Mean	SD	Mean	SD		
MCCB Composite	0.91	28.92	12.55	30.33	12.76	1.41**	5.36	+1.41	-7, 10
Attentional Vigilance	0.85	38.35	12.36	39.74	12.65	1.39**	6.71	+1.39	-9, 12
Processing Speed	0.84	33.62	11.50	35.04	10.87	1.42**	5.83	+1.42	-8, 10
Reasoning and Problem Solving	0.75	38.96	8.04	39.92	8.80	0.96*	6.01	+0.96	-8, 10
Social Cognition	0.73	36.69	12.68	35.92	12.33	-0.77	9.24	-0.77	-15, 14
Verbal Learning	0.69	37.91	8.41	39.14	9.47	1.23*	7.23	+1.23	-10, 13
Visual Learning	0.67	38.21	13.51	38.88	12.38	0.67	10.40	+0.67	-16, 17
Working Memory	0.81	35.82	12.13	37.01	11.19	1.19*	7.05	+1.19	-10, 12

** Change score significant at the $p < .01$ level

* Change score significant at the $p < .05$ level

R = Pearson correlation between time 1 and time 2

Table 2
Regression 90% and 80% and 80% “operating,” confidence intervals for MCCB Composite organized by ranges of baseline scores

Baseline Score Range	LICI 90%	UICI 90%	LICI 80%	UICI 80%	N per range	Cumulative N
1 to 8	-5	12	-3	10	6	3.77%
9 to 11	-5	11	-3	9	10	10.06%
12	-5	11	-4	9	1	10.69%
13 to 19	-6	11	-4	9	21	23.90%
20	-6	10	-4	9	3	25.79%
21 to 24	-6	10	-4	8	12	33.33%
25	-6	10	-5	8	6	37.11%
26 to 31	-7	10	-5	8	37	60.38%
32	-7	9	-5	8	6	64.15%
33 to 36	-7	9	-5	7	17	74.84%
37	-7	9	-6	7	2	76.10%
38 to 44	-8	9	-6	7	17	86.79%
45 to 48	-8	8	-6	6	9	92.45%
49 to 57	-9	8	-7	6	11	99.37%
58 to 59	-9	7	-7	5	1	100%
60	-10	7	-8	5	0	100%