Interpreting T Scores and Percentile Equivalents Using the MATRICS Consensus Cognitive Battery in Schizophrenia Trials

TO THE EDITOR: In the February 2008 issue of the *Journal*, Robert S. Kern, Ph.D. et al. (1) presented normative data for the National Institute of Mental Health's Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery. The authors are to be congratulated for conducting a rigorous study that facilitated conorming and standardization for MATRICS Consensus Cognitive Battery (MCCB) test scores. The resulting normative-derived T scores are useful metrics for assessing the degree of normality or abnormality of any given test score. However, it should be noted that raw test scores need to follow a symmetrical and asymptotic normal distribution curve. Even for a cognitive test with normal distribution in the general population, normality assumptions may still be violated in a schizophrenia population known for significant cognitive deficits. Extreme standard T scores may be observed in the latter population, making the assessment of percentile equivalents and valid interpretations difficult. For example, extreme standard T scores in a schizophrenia population can be as low as 36 (i.e., 1.4 standard deviations below normal) or 20 for some tests (i.e., 3 standard deviations below the healthy comparison mean) (2, 3). Because these values lie at the tail of the normal distribution curve, a 10-point difference between T scores of 35 and 25 produces only 6.4 percentile points relative to 34 percentile points derived from a difference between T scores of 50 and 40. Thus, interpretation of the differences between two standard T scores and their percentile equivalents can be misleading, depending on where the two T scores fall on the normal curve.

An additional point to consider when using T scores is the requirement for extrapolation when the normative sample is not large enough to include data points that are close to the actual scores. For example, the estimated prevalence for a T score of 25 (2.5 standard deviations below the mean) is only 2 in 320 subjects. It would have been useful if the authors had provided the range, in addition to the standard deviation, of the raw test scores in the normative sample. Recognizing these issues in using the T score and its percentile equivalent is important for the analysis and valid interpretation of these normative-derived standard scores.

References

1. Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JM, Keefe RSE, Mesholam-Gately R, Mintz J, Seidman LJ, Stover E, Marder SR: The MATRICS Consensus Cognitive Battery, Part 2: co-norming and standardization. Am J Psychiatry 2008; 165:214–220
2. Harvey PD, Keefe RSE: Cognitive impairment in schizophrenia and implications of atypical neuroleptic treatment. CNS Spectr 1997; 2:1–11
3. Keefe RSE, Bilder RM, Harvey PD, Davis SM, Palmer BW, Gold JM, Meltzer HY, Green MF, Miller DD, Canive JM, Adler LW, Manschreck TC, Swartz M, Rosenheck R, Perkins DO, Walker TM, Stroup TS, McEvoy JP: Baseline neurocognitive deficits in the CATIE schizophrenia trial. Neuropsychopharmacology 2006; 31:2033–2046

CYNTHIA O. SIU, PH.D.
Ringoes, N.J.

Dr. Siu has served as a consultant for Dainippon Sumitomo Pharma America, Organon, Memory Pharmaceuticals, Prescient, Pfizer, and Wyeth.

This letter (doi: 10.1176/appi.ajp.2008.08030379) was accepted for publication in April 2008.

Dr. Kern and Colleagues Reply

TO THE EDITOR: We thank Dr. Siu for her letter commenting on our article. Dr. Siu identifies two psychometric issues that should be considered when interpreting standardized T scores for individuals with schizophrenia. She notes that even when raw test scores follow symmetrical and asymptotic normal distributions for the normative and target patient samples, there can be difficulties in interpreting T scores and their percentile equivalents. Additionally, she states that extrapolation is necessary for patient T score values that are not captured within the range of normative sample raw test scores and asks for the range of scores by test in our study.

Regarding the distribution of scores in the normative sample, each MCCB test was examined for normality of score distribution. The primary measures from the Trail Making Test, Part A; the Hopkins Verbal Learning Test—Revised; and the Neuropsychological Assessment Battery Mazes subtest required log transformation to address problems in skewness. The MCCB scoring program uses these transformed scores in calculating the T scores and percentiles for each test and, hence, provides a reference that is normally distributed. Parenthetically, in the patient sample in our study, only the Trail Making Test, Part A from the final MCCB required transformation to normalize the distribution. Dr. Siu's point regarding the relationship between standard score values and their corresponding percentiles applies to any neuropsychological assessment of individuals with clinical disorders. As a result of the way that these scores are derived, percentile scores change most in relation to standard scores when they fall near the median of the normative distribution, and they change least when they fall at the tails. The proportion of individuals represented within the tail of the normal curve was necessarily small.

Dr. Siu is correct that relatively few subjects in the community normative sample scored at extremely low levels. However, the accuracy of the T scores for these low performance levels was not dependent on the few subjects who scored low, but rather on the extent to which the distribution of the entire normative sample met the assumption of a normal distribution of scores. To be fair, the range in which one is most likely to see deviations from normality and, hence, risks to accuracy falls within the tails, but this is also the range in which one is least able to determine the normality, since it provides the least amount of data. To the extent that patient scores overlapped the community sample and the normality assumption was met, the percentiles were reasonably accurate and meaningful. As requested by Dr. Siu, we provide the following range

of raw scores from the community sample for each MCCB test in our study (means and standard deviations appear in the online data supplement according to age group): Trail Making Test: Part A (seconds): range=12–80; Brief Assessment of Cognition in Schizophrenia: Symbol Coding (total correct): range=21–86; Hopkins Verbal Learning Test-Revised (total recall trials 1–3): range=9–36; Wechsler Memory Scale-III: Spatial Span (total correct): range=5–25; Letter-Number Span (total correct): range=5–24; Neuropsychological Assessment Battery: Mazes (total raw score): range=2–26; Brief Visuospatial Memory Test-Revised (total recall trials, 1–3): range=5–36; Category fluency: animal naming (total correct): range=8–39; Mayer-Salovey-Caruso Emotional Intelligence Test: Managing Emotions (branch score): range=68–117; Continuous Performance Test-Identical Pairs Version (mean d-prime across 2-, 3-, and 4-digit conditions): range=0.45–4.24.

It is worth noting that within a clinical trials context to assess the efficacy of promising new compounds, investigators who utilize the MCCB will be primarily interested in assessing change in cognition over time, which is best measured on the T score rather than percentile scale. The MCCB provides a standardized method of measurement to assess test performance across multiple cognitive domains and evaluate the efficacy of new compounds for this purpose. Furthermore, the use of T scores based on a community sample as the primary metric allows investigators to consider the extent to which a potential cognitive enhancer improves the cognitive performance of a clinical group relative to typical performance levels in the community.

ROBERT S. KERN, PH.D.
MICHAEL F. GREEN, PH.D.
KEITH H. NUECHTERLEIN, PH.D.
CATHERINE A. SUGAR, PH.D.
Los Angeles, Calif.

The authors' disclosures accompany the original article.

This letter (doi: 10.1176/appi.ajp.2008.08030379r) was accepted for publication in April 2008.

Work-Hour Regulation: Collateral Damage to Consultation Psychiatry

TO THE EDITOR: We read with interest the commentary by Patrick A. Rabjohn, M.D., Ph.D. and Joel Yager, M.D., published in the March 2008 issue of the *Journal*, which discussed the effects of resident work-hour regulation on psychiatry (1). We wish to highlight the specific repercussions of work-hour regulation on consultation psychiatry.

As consultation psychiatrists who work in academic general hospitals, we have noticed a seismic shift in the "ownership" of patients by residents in other specialties (2). Since Accreditation Council for Graduate Medical Education (ACGME) regulations were adapted, a team-based shift-work approach has replaced the traditional model of work hours. The most immediate effect of this replacement model on the consultant psychiatrist has been to render more difficult clarification of the reason for psychiatric consultations. Attempts at such clarification are frequently met with an assortment of buck-passing responses, such as "I'm just covering"; "It's not in my sign-out"; "I was just told to call it in"; or "Just ask the team in the morning." This diffusion of clinical responsibility across the breadth of a multimember team leads to an absurdist interaction of the patient and psychiatrist sitting together in a hospital room, with neither knowing why the other is there.

Another, perhaps more disturbing, result of work-hour regulation has been the farming out of the doctor-patient relationship to psychiatric consultants. The team approach does not allow for the development of a relationship between the patient and any single physician because patients are typically seen by several different residents each day. If a psychiatric consultation is placed, the psychiatrist that provides the initial evaluation usually returns on subsequent days to provide follow-up, allowing for a constancy of interaction that fosters the familiarity, comfort, and support that has been traditionally provided by the primary physician.

This outsourcing of the doctor-patient relationship bodes poorly for two chief reasons. First, it engenders the risk of producing a generation of non-psychiatrist physicians who are inexperienced in the human aspects of medicine. When these trainees complete their education, some of them will practice in settings where psychiatric consultation is not easily available and then realize the forfeited opportunities to learn vital elements of doctoring. Second, our own trainees have already begun to express frustration with consultation work, since they are not often asked to diagnose or treat a patient but instead to participate in a relationship that they (and often the patient) identify as one that belongs to someone else. This dissatisfaction may portend future difficulty in recruiting residents to careers in consultation psychiatry or psychosomatic medicine.

The aforementioned is not intended to castigate non-psychiatrist physicians, who often attempt to address psychosocial issues through consultation, nor is it intended to blame the medical education system. Medicine has become more complex, and team-based approaches are here to stay. However, the time has come to acknowledge the disconnect between the humanistic rhetoric of medical training and the real-life obstacles to its practical implementation. The relationship between hour reduction team-based care and psychiatric consultation might be a good place to start.

References

1. Rabjohn PA, Yager J: The effects of resident work-hour regulation on psychiatry. Am J Psychiatry 2008; 165:308–311
2. Kontos N, Freudenreich O, Querques J: Ownership, responsibility and hospital care: lessons for the consultation psychiatrist. Gen Hosp Psychiatry 2008; 30:257–262

JASON P. CAPLAN, M.D.
Phoenix, Ariz.
JOHN QUERQUES, M.D.
OLIVER FREUDENREICH, M.D.
Boston, Mass.
NICHOLAS KONTOS, M.D.
Cambridge, Mass.

Dr. Freudenreich has received research support from Cephalon; he has also received education funding from Primedia. Drs. Caplan, Querques, and Kontos report no competing interests.

This letter (doi: 10.1176/appi.ajp.2008.08030404) was accepted for publication in April 2008.