Evidence for Added Value of Baseline Testing in Computer-Based Cognitive Assessment

Tresa M. Roebuck-Spencer, PhD; Andrea S. Vincent, PhD; Robert E. Schlegel, PhD; Kirby Gilliland, PhD

University of Oklahoma, Norman

Context: Large-scale baseline cognitive assessment for individuals at risk for concussion is a common part of the protocol for concussion-surveillance programs, particularly in sports. Baseline cognitive testing is also being conducted in US military service members before deployment. Recently, the incremental validity of large-scale baseline cognitive assessment has been questioned.

Objective: To examine the added value of baseline cognitive testing in computer-based neuropsychological assessment by comparing 2 methods of classifying atypical performance in a presumed healthy sample.

Design: Cross-sectional study.

Setting: Military base.

Patients or Other Participants: Military service members who took the Automated Neuropsychological Assessment Matrix (ANAM) before and after deployment (n = 8002).

Main Outcome Measure(s): Rates of atypical performance in this healthy, active-duty sample were determined first by comparing postdeployment scores with a military normative database and then with each individual's personal baseline performance using a reliable change index.

Results: Overall rates of atypical performance were comparable across these 2 methods. However, these methods were highly discordant in terms of which individuals were classified as atypical. When norm-referenced methods were used, 2.6% of individuals classified as normal actually demonstrated declines from baseline. Further, 65.7% of individuals classified as atypical using norm-referenced scores showed no change from baseline (ie, potential false-positive findings).

Conclusions: Knowing an individual's baseline performance is important for minimizing potential false-positive errors and reducing the risks and stresses of misdiagnosis.

Key Words: Automated Neuropsychological Assessment Metrics, computerized neuropsychological testing, concussions, mild traumatic brain injuries, military athletes

Key Points

- When service members' predeployment and postdeployment performances on the Automated Neuropsychological Assessment Metrics test were compared using norm-referenced scores and their personal baselines, the absolute rates of atypical performance were similar.
- However, for those individuals whose performance was classified as atypical, a high degree of discordance was noted between the methods. Using norms alone resulted in a high level of false-positive errors.
- Mistakenly classifying an individual as cognitively impaired should be avoided whenever possible because it can lead to overuse of medical resources and undue emotional stress.

omputer-based cognitive testing is fast becoming standard practice in the assessment and management of mild traumatic brain injury (mTBI) in sports and military concussion-management programs. Recommended sport-concussion protocols typically include baseline assessment of cognitive performance for individuals at high risk for concussion, against which postconcussion performance is compared.^{1,2} Because of increasing concern for the risk of cognitive injury during military deployment, Congress mandated in 2008 that all US military service members receive computer-based predeployment and postdeployment neuropsychological assessments.³ As with sport-concussion management programs, the purpose of the baseline assessment is to archive the service member's predeployment neurocognitive performance, so that it can be compared with performance after an injury or other neurologic insult.

Such models presume that baseline information is beneficial in determining the presence and severity of cognitive insult after injury by documenting an individual's level of cognitive functioning before engaging in activities that increase the risk of concussion (eg, sport activity, military deployment). Guskiewicz et al¹ stated that the goal of baseline testing is to provide the most reliable benchmark against which to compare postinjury performance. Without baseline information, clinicians must rely solely on norm-referenced postinjury scores. This practice may increase the risk of false-positive errors in healthy individuals whose premorbid cognitive functioning falls at the lower end of the normal curve. Similarly, this practice may also increase the risk of false-negative errors in injured individuals whose premorbid abilities fall at the higher end of the normal curve, for whom "average" performance relative to norms at a postinjury time point may actually represent a decline.

Recently, the empirical validity of concussion-monitoring and management programs for preventing or mitigating risk associated with concussion has been called into

question.⁴ Although it is unclear whether these programs mitigate serious risks at the population level, monitoring individuals who have sustained a concussion or other injury is valuable to detect serious cognitive insult, prevent poor outcomes, and optimize return to activity while lowering the risk of future and potentially more damaging injuries. This argument is even more compelling for those individuals who display an atypical recovery course from concussion or other injury and those who have sustained multiple past injuries. In this vein, the utility and added value of large-scale baseline cognitive testing requires demonstration. That is, more information is needed regarding the incremental validity of comparing postinjury cognitive testing with baseline performance over and above the more traditional comparison of postinjury scores with normative values.

Although this question is important at both the community and school levels, it becomes even more important in a military context, where population-wide cognitive assessment is standard and false-positive errors have implications for potential overuse of medical and financial resources; in addition, military personnel being screened for possible cognitive impairment may experience undue stress. Thus, we sought to determine the added value of baseline computerized cognitive assessment to decrease potential false-positive errors within a military sample.

METHODS

Participants

Deidentified data were obtained for a sample of activeduty military service members $(n = 10\,869)$ who were administered the Automated Neuropsychological Assessment Metrics Traumatic Brain Injury Battery-Military version (ANAM4 TBI-MIL) as part of their routine predeployment and postdeployment medical processing. Individuals who self-reported a history of concussion within 4 years of deployment, during their most recent deployment, or at both time points were excluded to prevent potential confounding effects of prior concussion on the current analyses (n = 2111). Most of these cases involved concussions that occurred before deployment (n =1346). Additional exclusions included other nonspecific injuries (n = 268), extreme outliers at either time point (n =169), incomplete data on potential injury history (n = 259), and a predeployment-postdeployment testing interval of less than 90 days (n = 60). The typical length of deployment for this sample is approximately 1 year, so these short testretest intervals were considered outliers and therefore excluded. Extreme outliers were defined as observations of performance in the top 1% for speed (fast reaction times) while simultaneously in the bottom 1% for accuracy. This performance pattern was rare overall (<2% of sample) and suggestive of someone pressing a response button extremely quickly without attending to test stimuli, consistent with either poor effort or misunderstanding of test directions. Equal proportions of individuals meeting these criteria were found at the baseline and postdeployment testing time points.

After exclusions, a subset of 8002 individuals remained with available data at both predeployment and postdeployment. Because these individuals were on active duty and did not report concussion or other serious injuries during either testing time point, they were presumed to be healthy and free of major medical or psychiatric illness that might negatively affect cognition. Available demographic information indicated that the sample consisted mostly of men (91%), with a mean age of 26.5 \pm 6.4 years. Army personnel constituted the overwhelming majority of this sample (99.9%). The average test-retest interval was 396 \pm 78 days (minimum = 90 days, maximum = 489 days), which is consistent with the typical length of deployment for Army personnel.

Measure

Automated Neuropsychological Assessment Metrics, version 4 (ANAM4). The ANAM4 Traumatic Brain Injury (TBI) Battery is an instrument designed to aid in the assessment of general cognitive function after suspected brain injury or other cognitive insult. The ANAM has a long history of use in medication trials, assessment of the cognitive effects of extreme conditions and neurologic disorders, military research, and sport concussion (see review⁵). It has good construct validity with traditional tests of attention, processing speed, and working memory.^{6,7} Test-retest reliability for the individual tests varies between 0.4 and 0.8, with tests requiring higher cognitive processing showing better reliability and simple reaction time showing lower reliability, most likely due to the restricted response range. For this analysis, we used an ANAM composite score, which has been shown to have test-retest reliability coefficients ranging from 0.62 to 0.74.8 This reliability is lower than ideal but likely compromised by a large testretest interval (greater than 1 year) and is generally consistent with other traditional neuropsychological tests of information-processing speed.9 A large normative database (n > 107000) of military service members was available for standardization of raw scores, with corrections for the effects of age and sex on performance.¹⁰ Service members were administered a version of the ANAM4 TBI battery, titled ANAM4 TBI-MIL, which includes the same tests as the nonmilitary version along with a customized military-specific demographic data-collection module and TBI questionnaire modeled after the Brief Traumatic Brain Injury Survey.¹¹ The battery takes approximately 20 minutes to complete via personal computer. Brief descriptions of each test in the order of administration are provided in Table 1. Detailed descriptions of ANAM tests can be found elsewhere.^{8,10,12} It should be noted that this battery includes all but 2 tests from the version of ANAM currently distributed for use in sports medicine applications (ie, Spatial Processing and Sternberg Memory Test). The cognitive domains assessed by these tests (ie, visual-spatial skills and working memory) are well represented within other tests in the battery used for this study. Thus, we believe that results from this ANAM battery would easily generalize to the ANAM battery used in previous and ongoing sport-concussion research.

Administration Procedures

The ANAM4 TBI-MIL battery was administered in a standardized manner in a group setting by trained test proctors who were present at all times. Predeployment test administration occurred in conjunction with the readiness

Table 1.	Automated	Neuropsychologica	A Assessment Metrics	(ANAM) Batter	y Description,	Version 4
----------	-----------	-------------------	----------------------	---------------	----------------	-----------

Test Name	Description		
Traumatic brain injury questionnaire	Self-report of the user's head injury history and related symptomatology; adaptation of the Brief Traumatic Brain Injury Survey ¹¹		
Sleepiness scale	Self-assessment of the user's level of sleepiness; modification of the Stanford Sleepiness Scale ^a		
Mood scale	Self-assessment of the user's mood state in 7 categories: vigor, happiness, depression, anger, fatigue, anxiety, and restlessness		
Simple reaction time	Requires the user to respond as quickly as possible to a target stimulus		
Code substitution—learning	Measures visual scanning, processing speed, attention, and learning by asking the user to compare a single symbol-digit pairing with a set of defined symbol-digit pairs presented at the top of the screen; the user is instructed to learn the symbol-digit pairing for a memory test to follow later in the battery		
Procedural reaction time	Measures attention and processing speed by having the user respond as quickly as possible to different sets of stimuli based on simple rules (eg, "Press the left mouse button if you see a 2 or 3 and the right mouse button if you see a 4 or 5.")		
Mathematical processing	Measures attention, basic computational skills, and working memory by asking the user to solve a single-digit arithmetic problem (eg, $(5 - 2 + 3 = ")$) involving 2 operations		
Matching to sample	Measures visual-spatial discrimination and working memory by presenting the user with a visual pattern for a specified period of time and then, after a brief delay, asking the user to select the previously seen pattern from 2 choices		
Code substitution— delayed memory	Measures visual recognition memory by asking the user to compare a single displayed symbol-digit pair with the previously learned symbol-digit pairs presented earlier in the test battery (ie, during the code substitution—learning test)		
Simple reaction time (R)	Identical to the simple reaction test administered earlier and designed to measure fatigue		

^a Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. *Psychophysiology*. 1973;10(4):431–436.

processing required for service members. Postdeployment testing assessment was conducted on day 6 of a 7-day reintegration process. Test administration followed a standardized briefing in which service members were advised of the nature and importance of the testing, provided procedural instructions, and given the chance to ask questions. According to standardization rules, test proctors clarified instructions when needed for individual examinees. Additionally, as part of the standard protocol for the military baseline-testing program, data were screened immediately after testing for unusual patterns (eg, accuracy rates below the chance level), and examinees were instructed to retake flagged tests. More detailed descriptions of administration procedures are offered elsewhere.^{8,10,12}

Data Analysis

Performance was examined using an ANAM composite score (ACS) summarizing overall performance with the throughput (TP) variable of each test. The TP is a ratio of reaction time and accuracy that represents correct responses per minute of available response time and is considered a good measure of cognitive efficiency.¹³ This variable was chosen over other ANAM variables of reaction time and accuracy because it more closely follows a normal distribution for parametric analyses while still capturing the information provided by these 2 important variables in a single metric. To create the ACS, raw TP scores for each ANAM test were converted to T-scores relative to an ageand sex-matched normative group. The TP T-scores were then summed across the tests and converted to a Z-score relative to the summed T-score of a normative control group.

Significant declines in performance relative to baseline were identified using a reliable change index (RCI). An RCI statistically takes into account possible fluctuations in performance due to test-retest reliability and standard error of measurement, thereby allowing change beyond that expected from chance alone to be determined. The RCI formula we chose also incorporated expected change due to practice effects as described by Chelune et al¹⁴ and Maassen et al.¹⁵ Specifically, the following formula for RCIp was used:

$$\text{RCIp} = \frac{(Y_2 - Y_1) - X_{\text{normchg}}}{\sqrt{(s_{X_1}^2 + s_{X_2}^2)(1 - r_{xx})}}, \quad (1)$$

where $Y_2 - Y_1$ is the observed change between postdeployment score Y_2 and baseline score Y_1 , $X_{normchg}$ is the average change in the control group, $s_{X_1}^2$ and $s_{X_2}^2$ are the time 1 and time 2 variances in the control sample, and r_{xx} is the test-retest reliability.

Because population estimates to calculate RCIs for military data do not exist outside of these data, the overall sample of 8002 available individuals was randomly divided into a derivation sample of 4001 participants to calculate the RCI values. These RCI values were then applied to the remaining 4001 participants (validation sample) to determine the frequency with which individuals demonstrated significant declines from baseline. The derivation sample (n = 4001) and the validation sample (n = 4001) did not differ in age, sex, or performance variables.

Atypical postdeployment performance was classified in the validation sample (n = 4001) for each individual in 2 ways: (1) For the baseline-referenced comparison method, postdeployment performance was compared with baseline performance to assess a possible performance decline. Performance was classified as atypical, or potentially impaired, when the postdeployment decline exceeded the RCI using a 90% confidence interval (CI; corresponding to a Z-score of -1.64). (2) For the norm-referenced method, postdeployment performance was compared with age- and sex-matched military-specific norms.¹⁰ In order to be

Table 2. Classification Rates of Performance on the Automated Neuropsychological Assessment Metrics (ANAM)

	Clas	eline	
Performance	Normal	Atypical	Row Totals
Classification based on norm-	referenced postdeployment scores		
Normal ^a	3764 (97.4%)	100 (2.6%)	3864 (96.6%)
Atypicalª	90 (65.7%)	47 (34.3%)	137 (3.4%)
Column totals	3854 (96.3%)	147 (3.7%)	4001

^a Percentage of cases relative to normative classification methods (ie, row totals).

statistically comparable with the baseline classification method, performance was classified as atypical when the ACS score was more than 1.64 standard deviations below the mean of the normative group. Data analyses were conducted in accordance with approval from the University and Department of Defense Human Subject Review Boards. This was a retrospective analysis of archival deidentified data; thus, informed consent was not required.

RESULTS

On comparing postdeployment scores with baseline performance, we found that 3.7% of the entire sample (n = 147) demonstrated atypical performance (based on declines in performance using the RCI criteria described earlier). Similarly, comparing postdeployment scores with military norms indicated that 3.4% (n = 137) showed atypical performance. Thus, when the sample was considered as a whole, these 2 methods classified similar percentages of individuals as having atypical performance, Z = 0.60, P = .55.

Given that an 80% CI interval has been suggested as more sensitive and more appropriate for detecting concussion,^{16,17} percentages of atypical performance were examined again at this threshold with a statistically comparable Z-score cut-point of -1.28 for norms-based comparisons. Although the percentages of atypical performance were somewhat higher, as expected, they did not differ across the norms-based (6.45%) or baseline-based (7.4%) classification methods, Z = 1.67, P = .09.

Although absolute classification rates of atypical performance were comparable across the norms-based and baseline-based classification methods, they were highly inconsistent in identifying individuals as atypical. (See Table 2 for classification rates of atypical performance across methods at the 90% CI.) For instance, of the 147 individuals classified as atypical by baseline-based methods, 68.0% (n = 100) were classified as normal by normsbased methods. This subgroup demonstrated an average ACS score of +0.92 at baseline, which declined almost 1.5 standard deviations to -0.51 at postdeployment. Similarly, of the 137 individuals classified as atypical by norms-based methods, 65.7% (n = 90) showed no change in performance from baseline (ACS = -2.15) to postdeployment (ACS = -2.13). In other words, if norms-based comparisons alone were used, 2.6% of those considered normal actually showed a decline based on baseline methods, and 65.7% of those considered atypical actually showed no decline from baseline. (See the Figure for illustrations of baseline and postdeployment performance across classification methods.) Notably, this overall pattern of results did not change when an 80% CI and Z-score cut-point of -1.28 were examined.

DISCUSSION

The value and utility of large-scale baseline cognitive testing initiatives within concussion-management programs and military programs have been questioned. In an effort to demonstrate the added value of baseline cognitive data, we compared rates of atypical (or potentially impaired) performance using norm-referenced classification methods and baseline-referenced methods in a healthy sample. Although the absolute rates of atypical performance were comparable across these methods, the identification of which individuals were atypical was highly discordant. If norm-referenced postdeployment scores were considered in isolation, 100 cases (2.6%) of the sample classified as normal (n = 3864) actually demonstrated declines in performance from baseline to postdeployment of almost 1.5 standard deviations. Whether these cases represent regression to the mean in healthy individuals or actual impairment in individuals who started out with higher-thanaverage performance that declined to average is unclear. Of perhaps more significance, 90 individuals (65.7%) who would have been classified as atypical (n = 137) on normreferenced methods actually showed no change from baseline. Notably, this group may represent false-positive errors when using norm-referenced methods for individuals with baseline low cognitive performance. Thus, we have empirically demonstrated the added value of having baseline data to avoid potential false-negative errors in premorbidly high-functioning individuals and potential false-positive errors in premorbidly low-functioning individuals.

Classification rates vary significantly according to how cut-points are set to identify atypical performance and also vary in the real world based on differences in clinical decision-making practices. For the purposes of this analysis, we chose a cut-point comparable to Z = -1.64(or 90% CI) for both classification methods to ensure statistical comparability. When we examined other statistically equated thresholds (eg, 80% CI: Z = -1.28 and 95% CI: Z = -1.96), the overall pattern of findings did not change. However, in clinical practice, decision rules for determining atypical performance in different testing scenarios may vary significantly and may not always ensure statistical equivalence. For instance, in the preliminary analyses of these data, we used classification rules believed to be similar to decision-making strategies used by clinicians.18 Data were examined on a test-by-test basis with the cut-point for atypical performance being more than 2 standard deviations below normative values on 1 ANAM test or more than 1 standard deviation below normative values on 2 or more ANAM tests. Significant declines in performance were defined as postdeployment scores exceeding RCI using a 95% CI. Using these rules, 21.8% of the entire sample would have been flagged as potentially



Figure. Mean baseline and postdeployment Automated Neuropsychological Assessment Metrics (ANAM) composite scores (ACS) in Z-score units by classification group. Abbreviation: RCI, reliable change index.

impaired when the performance was not different from baseline. Thus, discrepancies between normative- and baseline-classification methods may be magnified when different decision rules are used, particularly if these decision rules are not comparable statistically.

We examined presumed healthy individuals. Knowing the prevalence of atypical scores for healthy individuals for a given test battery is an essential component of test interpretation because unusually low scores are commonly seen in healthy individuals across neuropsychological batteries. In fact, base rates of atypical scores are now commonly included in neuropsychological test manuals and do not appear to vary as a function of any particular battery, age group, or cognitive domain.¹⁹ On various traditional neuropsychological test batteries, up to 37% of healthy individuals demonstrate low scores, or "atypical" performance, on at least 1 or 2 tests.^{19–21} Thus, the baseline rates of atypical performance for ANAM we found are entirely consistent with those on other traditional neuropsychological test batteries. Understanding these baseline rates is essential in clinical decision making to avoid overdiagnosis of cognitive impairment.

One limitation of our study is that we did not have access to detailed medical histories of our participants with documentation of potential comorbid emotional and psychiatric factors. Because the sample comprised activeduty service members who had not been medically discharged, they are presumed to be healthy and free of medical or psychiatric conditions that would significantly impair performance on neuropsychological testing. Cognitive impairment would not be expected for this sample, so scores in a range consistent with clinical impairment (ie,

"atypical" performance) are assumed to represent falsepositives. However, some individuals in this sample may have actual cognitive impairment due to unreported concussions, other medical or neurologic factors, or unknown reasons. The rate of reported concussions in this sample (before exclusions) is lower $(7.4\%)^{22}$ than that reported in other postdeployment studies (7.6% to 22.8%),^{11,23–26} raising the possibility that unreported concussions may have occurred. Also given the high rates of posttraumatic stress disorder (PTSD) in returning service members,²⁶ it is possible that some individuals in this sample may have experienced PTSD or another form of emotional distress that could have affected cognitive functioning. To the extent that these factors are present, atypical performance may reflect true cognitive impairment and, thus, would not be a false-positive.

This specific analysis was limited to the examination of a presumed healthy sample with the goal of identifying potential false-positive errors. To increase support for incremental validity of baseline testing, future authors should also examine how RCI-based comparisons with baseline performance reduce the potential for false-negative errors in individuals with a known injury or neurologic condition. Of note, individuals reporting a deploymentrelated concussion who were excluded from the current analyses were not examined because of the lack of information about the date of injury and the presumption that any acute effects of deployment-related concussions might have resolved by the time of testing. Future investigators should, however, examine rates of cognitive decline in this subset of individuals and in individuals with known acute injuries.

Another potential limitation of this study is that classification using the RCI may be more conservative than Z-scores, given that lower-than-optimal test-retest reliabilities for some computer-based cognitive tests may increase the RCI interval, resulting in the need for a larger observed difference to qualify as a meaningful change. This possibility was mitigated by the fact that we used statistically comparable cut-points across classification methods and a composite score, which demonstrated overall better test-retest reliability than the individual tests alone.⁸

When generalizing these findings to sport-concussion applications, we note the many similarities between military and sport-concussion samples (eg, both are at considerable risk for concussion and are relatively young, healthy, and physically active) and in testing procedures, protocols, and purposes. That is, individuals in both settings are commonly tested with computerized reaction-timebased tests, often in group settings, with the goal of using baseline testing as a comparison point should a concussion occur in the future. However, some sample-specific differences should be noted. First, military participants are at risk for exposure to blast injuries and have high rates of PTSD,²⁵ conditions not typically seen in sport-concussion patients. These differences are mitigated in this study by an active-duty sample without known selection bias (ie, individuals were tested consecutively and routinely and were not drawn from clinic samples) and excluding individuals who reported injuries that resulted in signs and symptoms of concussion, including blast exposures. Although some service members may have underreported these experiences, we do not expect that those who chose not to report them would have lasting cognitive impairments as a result of them. Also, multiple studies indicate that deployment alone in large representative military samples does not result in cognitive impairment as measured by ANAM^{8,24} and that, when present, symptoms of PTSD affect cognitive performance at longer intervals (ie, 1 year) after return from deployment but not at the shorter intervals seen in this study. Thus, particularly in light of the noted similarities, we believe that sample differences are unlikely to strongly affect the ability to broadly generalize these results to a sport-concussion context.

These findings are important to consider when using computer-based testing for large-scale screening programs in which false-positive errors are reasonably expected for a certain percentage of the group being monitored. Further, these findings are particularly relevant for implementing cognitive testing within the military, given the high rates of mTBI being reported in service members returning from deployment.^{11,23,25} In this setting, incorrectly identifying cognitive impairment has significant implications, including the potential for overuse of financial and medical resources. Although cognitive testing is valuable in documenting the presence of cognitive symptoms after a concussion, cognitive testing should not be used in isolation as a diagnostic tool but rather in combination with other clinical indicators of concussion, including the clinical history, symptom report, and balance testing, to make clinical decisions. Regardless of setting or testing methods, misdiagnosis of mTBI can have significant and potentially devastating effects on individuals who may come to believe

they have true cognitive impairment and then misattribute everyday cognitive errors or cognitive difficulties related to an emotional cause as being the result of a neurologic injury.

CONCLUSIONS

We examined the added value of baseline cognitive testing in computer-based neuropsychological assessment by comparing 2 methods of classification of atypical performance in a presumed healthy, active-duty military sample tested before and after deployment. Postdeployment performance was classified as either normal or atypical based on comparison with military norms versus comparison with personal baseline performance using RCIs. Although the absolute rates of atypical performance were comparable with these methods, a high degree of discordance existed. Specifically, a high percentage of individuals classified as atypical by normative standards actually showed no change from baseline, indicating that using norms alone may result in a large number of falsepositive errors. These data came from a military sample, so the purposes of evaluation and testing procedures and protocols are similar to those in sport-concussion monitoring; thus, the added value of baseline testing can arguably be broadly generalized to the sport-concussion literature. Misclassification of cognitive impairment, regardless of setting, can have serious effects on an individual who may be misinformed that he or she has cognitive impairment related to an injury and may ultimately lead to overuse of financial and medical resources in concussion-surveillance programs.

ACKNOWLEDGMENTS

This project was conducted by the Cognitive Science Research Center (formerly the Center for the Study of Human Operator Performance: C-SHOP) at the University of Oklahoma and is made possible by a Department of Defense Grant that was awarded and administered by the US Army Medical Research & Materiel Command (USAMRMC) and the Telemedicine & Advanced Technology Research Center (TATRC), at Fort Detrick, MD, under contract number W81XWH-07-2-0097. We acknowledge the significant contributions of the many individuals, staff, and service members at the following organizations who made data collection for this analysis possible: Fort Campbell, KY; Neurocognitive Assessment Branch, Rehabilitation and Reintegration Division, US Army Office of the Surgeon General; and the Defense and Veterans Brain Injury Center.

FINANCIAL DISCLOSURE

The University of Oklahoma holds the exclusive license for the Automated Neuropsychological Assessment Metrics (ANAM). The Cognitive Science Research Center (formerly C-SHOP) at the University of Oklahoma is responsible for research and development of ANAM. VistaLifeSciences holds the exclusive license for ANAM commercialization. Authors Gilliland and Schlegel have standard university royalty agreements for the sale of ANAM. No other authors of this manuscript receive funds or salary support from ANAM sales.

NONENDORSEMENT DISCLAIMER

The views, opinions and/or findings contained in this manuscript are those of the author(s)/company and do not necessarily reflect the views of the Department of Defense and

should not be construed as an official Department of Defense/ Army position, policy, or decision unless so designated by other documentation. No official endorsement should be made.

REFERENCES

- Guskiewicz KM, Bruce SL, Cantu RC, et al. Recommendations on management of sport-related concussion: summary of the National Athletic Trainers' Association position statement. *Neurosurgery*. 2004;55(4):891–895.
- Guskiewicz KM, Bruce SL, Cantu RC, et al. National Athletic Trainers' Association position statement: management of sportrelated concussion. J Athl Train. 2004;39(3):280–297.
- 3. United States House of Representatives H.R. 4986. National Defense Authorization Act for Fiscal Year 2008, Sec. 1618, Comprehensive plan on prevention, diagnosis, mitigation, treatment, rehabilitation of, and research on, traumatic brain injury, post traumatic stress disorder, and other mental health conditions in members of the armed forces. http://thomas.loc.gov/cgi-bin/query/C?r110:./temp/~r110itMI1I. Accessed January 6, 2010.
- Randolph C, Kirkwood MW. What are the real risks of sport-related concussion, and are they modifiable? *J Int Neuropsychol Soc.* 2009; 15(4):512–520.
- McCaffrey RJ, Kane RL, eds. DoD contributions to computerized neurocognitive assessment: the ANAM test system. *Arch Clin Neuropsychol.* 2007;22(supp 1):1–144.
- Bleiberg J, Kane RL, Reeves DL, Garmoe WS, Halpern E. Factor analysis of computerized and traditional tests used in mild brain injury research. *Clin Neuropsychol.* 2000;14(3):287–294.
- Kabat MH, Kane RL, Jefferson AL, DiPino RK. Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures. *Clin Neuropsychol.* 2001;15(4):498–507.
- Vincent AS, Roebuck-Spencer TM, Lopez M, et al. The effects of military deployment on cognitive functioning. *Mil Med.* 2012;177(3): 248–255
- Strauss E, Sherman EMS, Spreen O. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. New York, NY: Oxford University Press; 2006.
- Vincent AS, Roebuck-Spencer T, Gilliland K, Schlegal R. Automated Neuropsychological Assessment Metrics (v4) traumatic brain injury battery: military normative data. *Mil Med.* 2012;177(3):256– 269.
- Schwab KA, Ivins B, Cramer G, et al. Screening for traumatic brain injury in troops returning from deployment in Afghanistan and Iraq: initial investigation of the usefulness of a short screening tool for traumatic brain injury. *J Head Trauma Rehabil*. 2007;22(6):377–389.
- 12. Vincent AS, Bleiberg J, Yan S, et al. Reference data from the automated Neuropsychological Assessment Metrics for use in

traumatic brain injury in an active duty military sample. *Mil Med.* 2008;173(9):836–852.

- 13. Thorne DR. Throughput: a simple performance index with desirable characteristics. *Behav Res Methods*. 2006;38(4):569–573.
- Chelune GJ, Naugle RI, Luders H, Sedlack J, Awad IA. Individual change after epilepsy surgery: practice effects and base rate information. *Neuropsychology*. 1993;7(1):41–52.
- Maassen GH, Bossema ER, Brand N. Reliable change assessment with practice effects in sport concussion research: a comment on Hinton-Bayre. Br J Sports Med. 2006;40(10):829–833.
- Barr WB. Neuropsychological testing of high school athletes: preliminary norms and test-retest indices. *Arch Clin Neuropsychol*. 2003;18(1):91–101.
- Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol.* 2003; 17(4):460–467.
- Roebuck-Spencer T, Vincent AS, Schlegal R, Gilliland K. Evidence of added value of baseline testing in computerized testing [poster]. Paper presented at: National Academy of Neuropsychology; October 13–16, 2010; Vancouver, BC, Canada.
- 19. Brooks BL, Iverson GL. Comparing actual to estimated base rates of "abnormal" scores on neuropsychological test batteries: implications for interpretation. *Arch Clin Neuropsychol.* 2010;25(1):14–21.
- Schretlen DJ, Testa SM, Winicki JM, Pearlson GD, Gordon B. Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *J Int Neuropsychol Soc.* 2008;14(3):436– 445.
- Heaton RK, Miller SW, Taylor MJ, Grant I. Revised Comprehensive Norms for an Expanded Halstead-Reitan Battery: Demographically Adjusted Neuropsychological Norms for African Americans and Caucasian Adults. Lutz, FL: Psychological Assessment Resources, Inc; 2004.
- 22. Roebuck-Spencer TM, Vincent AS, Twillie DA, et al. Cognitive change associated with self-reported mild traumatic brain injury sustained during the OEF/OIF conflicts. *Clin Neuropsychol.* 2012; 26(3):473–489.
- Terrio H, Brenner LA, Ivins BJ, et al. Traumatic brain injury screening: preliminary findings in a US Army Brigade Combat Team. *J Head Trauma Rehabil.* 2009;24(1):14–23.
- Vasterling JJ, Proctor SP, Amoroso P, Kane R, Heeren T, White RF. Neuropsychological outcomes of Army personnel following deployment to the Iraq War. *JAMA*. 2006;296(5):519–529.
- Hoge CW, McGurk D, Thomas JL, Cox AL, Engel CC, Castro CA. Mild traumatic brain injury in U.S. soldiers returning from Iraq. N Engl J Med. 2008;358(5):453–463.
- Tanielian T, Jaycox LH, eds. Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery. Santa Monica, CA: RAND Corporation; 2008.

Address correspondence to Tresa M. Roebuck-Spencer, PhD, Cognitive Science Research Center, University of Oklahoma, 3200 Marshall Avenue, Suite 260, Norman, OK 73072. Address e-mail to tresa@ou.edu.