APPLIED NEUROPSYCHOLOGY: ADULT
https://doi.org/10.1080/23279095.2017.1330749

Routledge
Taylor & Francis Group

Computerized cognitive testing norms in active-duty military personnel: Potential for contamination by psychologically unhealthy individuals

Ian Coffman, Helaine E. Resnick, James Drane and Corinna E. Lathan

AnthroTronix, Inc., Silver Spring, Maryland, USA

ABSTRACT

Normative reference data used for clinical interpretation of neuropsychological testing results are only valid to the extent that the sample they are based on is composed of "normal" individuals. Accordingly, efforts are made to exclude individuals with histories and/or diagnoses that might bias test performance. In this report, we focus on these features in active-duty military personnel because published data on computerized neurocognitive testing norms for this population have not explicitly considered the consequences of neurobehavioral disorders (e.g., PTSD, depression), which are prevalent in this population and known to affect performance on some cognitive assessments. We administered DANA, a mobile, neurocognitive assessment tool, to a large sample of active-duty military personnel and found that scores on self-administered psychological assessments negatively impacted a number of neurocognitive tests. These results suggest that neurobehavioral disorders that are relatively common in this population should be controlled for when establishing normative datasets for neurocognitive outcomes.

KEYWORDS

Active-duty military; cognitive assessment; normative data; PTSD

Introduction

Normative data are used extensively in the clinical interpretation of neurocognitive testing results since a single score on an assessment is difficult to interpret without knowing how it compares to others in the test-taking population. For example, an individual's performance might be compared to a reference distribution centered on a sample mean that is derived from a set of "normal" subjects. Because there is variability in these results—even among normally functioning subjects—cutoff points at each end of the distribution are established as thresholds for classifying patients as potentially "nonnormal." If the reference empirical distribution is approximately symmetric, then any patient falling more than +/−2 standard deviations away from the mean may be considered "nonnormal."

It is widely recognized that normative data are only useful to the extent that (a) they can be applied to individuals with similar characteristics to the sample from which the data were collected (e.g., Heaton et al., 1986; Ross & Lichtenberg, 1998) and (b) that the reference data comprise observations from "normal," that is, unimpaired, subjects. Appreciation of the latter point has resulted in efforts to exclude from normative data individuals with histories and/or diagnoses that can be reasonably expected to affect test performance and, by extension, the larger distribution. Exclusion criteria can include general, domain-specific features (e.g., history of memory complaints if the normative data are for scores on a memory-based test) as well as features that might be expected to occur with greater incidence in a particular population (e.g., presence of dementia in a geriatric sample). For example, Schneider et al. (2015) prescreened subjects for history of dementia and other age-related neurologic issues when they collected normative data for a number of neuropsychological tests from a sample of older adults.

Demographic characteristics of the sample, e.g., age and gender, may also affect normative data. If the population that the sample is meant to represent includes multiple levels or values of these characteristics (e.g., both males and females for gender) and if differences in these values are thought to yield substantial effects on the measure of interest, then these effects must also be controlled for. Typically, this is accomplished by presentation of stratified tables of means conditioned on each value of the demographic feature. For example, after statistical testing revealed significant effects of age and education, Ganguli et al. (2010) presented normative data for a number of cognitive assessments stratified by both of these features.

CONTACT Ian Coffman ian.coffman@atinc.com 8737 Colesville Rd. Ste. L-203, Silver Spring, MD, 20910, USA.

In this report, we consider the task of establishing a normative dataset for a neurocognitive assessment tool (NCAT) in the context of the active-duty military population. In particular, we focus on which population-specific features should be accounted for in the process of defining a normative dataset. Beyond controlling for basic demographic factors such as age and gender, most published normative data on NCATs for military use cases (e.g., Roebuck-Spencer, Vincent, Schlegel, & Gilliland, 2013; Vincent, Roebuck-Spencer, Gilliland, & Schlegel, 2012) do not consider the consequences of neurobehavioral disorders that likely affect active duty service members at a rate greater than that of the general population (e.g., posttraumatic stress disorder [PTSD]). Although Roebuck-Spencer and colleagues identify this issue as a potential limitation of their approach, they note that because their sample "comprised active duty 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" (p. 503). This report explores this basic assumption.

Despite the implications of service members' discharge status or classification as "active-duty," there is evidence that a subset of this population may be affected by neurobehavioral challenges. For example, Hoge et al. (2004) estimated the rate of probable PTSD at 9% among predeployed service members, and 12 and 18% at postdeployment among participants in Operations Enduring Freedom and Iraqi Freedom, respectively. Those findings are relevant because PTSD is known to negatively impact performance on certain neurocognitive tests (e.g., Horner & Hamner, 2002; Swick, Honzel, Larsen, Ashley, & Justus, 2012). These results suggest that it is prudent to explicitly test for the presence of various psychological disorders to determine whether their neurocognitive consequences, if any, are extreme enough to substantially bias data that would otherwise be mistakenly classified as "normative."

Using a large sample of active duty service members aged 18–64, we examined performance on eight cognitive tests, and then studied the impact of self-reported sleep disturbance, depression, and PTSD on these distributions. We hypothesized that independent of age and gender, active duty military personnel with sleep disturbances, depression, and PTSD would perform less favorably on computerized cognitive tests than their counterparts who do not report these conditions. If this hypothesis proves true, it has direct implications for how normative data are used to evaluate cognitive efficiency in active duty military personnel.

Methods

DANA, a hand-held, computerized neurocognitive assessment tool, was administered to 814 active duty service members (71% male) aged 18–64 stationed at the Fort Hood military post near Killeen, Texas. A data collection error resulted in three instances in which two participants' data were assigned to a single unique identifier. These six records were excluded from analysis, yielding a final sample size of 808.

Service members were recruited via distribution of fliers at locations around the Ft. Hood post and through direct briefings by the site PI after unit/company formation. Consent materials stated that the research goal was to collect a large database of cognitive data on active-duty military, and, as a result, participants were naïve to any potential theoretical comparisons. It should be noted that since 59% of our sample had been previously deployed, depending on when deployment occurred, some may have been exposed to the Automated Neuropsychological Assessment Metrics (ANAM) battery (e.g., Vincent et al., 2012). A 2008 Congressional mandate required administration of this battery prior to deployment. ANAM's subtests are comparable to DANA's; therefore, it is possible that some participants entered the experimental setting having prior experience with computerized cognitive testing.

Military personnel were eligible to take part in the study if they were classified as active duty and between the ages of 18 and 64 (inclusive). Potential participants were excluded if they had consumed alcohol within the last eight hours, regularly used mind-altering medications (e.g., anti-psychotic medications, benzodiazepines, Benadryl), or had sustained a concussion within the month prior to testing. DANA was administered on Samsung Galaxy S4 smartphones, which the developers of DANA have found to be technically suitable for this purpose.

DANA contains a battery of tests designed to examine cognitive performance on several tasks, and it also includes several psychological tests. Its favorable psychometric properties and test–retest reliability has been documented (Lathan, Spira, Bleiberg, Vice, & Tsao, 2013; Russo & Lathan, 2015). Russo & Lathan demonstrate that the test–retest reliability coefficients for DANA's Simple Reaction Time and Procedural Reaction Time subtests (intraclass correlation coefficients of 0.81 and 0.75, respectively) are comparable to other assessment batteries that contain these tests. A summary of the neurocognitive tests examined in this study is provided in Table 1.

A score of less than 66% correct on any DANA subtest is considered an invalid administration and excluded from analysis. This criterion is evaluated on a per-subtest basis; if a participant scored less than

Table 1. Description of DANA subtests.

Test name	Task description
Simple Reaction Time (SRT)	The subject taps an orange target symbol as quickly as possible each time it appears. The location and shape of the stimulus does not vary from trial to trial.
Procedural Reaction Time (PRT)	The screen displays one of four numbers (1, 2, 3, or 4) for 2 seconds. The subject taps the left button ("2 or 3") or right button ("3 or 4") as quickly as possible to indicate which category corresponds to the number displayed.
Go/No-Go (GNG)	A building is presented on the screen with several windows. Either a "friend" (green) or "foe" (gray) appears in a window. The subject must tap the "BLAST" button as quickly as possible only when a "foe" appears.
Code Substitution – Learning (CSL)	Subjects refer to a key of 9 symbol-digit pairs that are shown across the upper portion of the screen. Single symbol-digit pairs are presented in succession below the key, and the subject indicates whether or not the single pair matches the code by tapping "Yes" or "No" As quickly as possible.
Code Substitution – Recall (CSR)	After a delay of several intervening tests, the same symbol-digit pairs from the earlier Code Substitution – Learning task are presented without the key. The subject indicates whether or not the pairing was included in the code that was presented in the earlier Code Substitution – Learning section by tapping "Yes" or "No" as quickly as possible.
Spatial Processing (SP)	Pairs of four-bar histograms are displayed on the screen, one pair at a time and simultaneously, with one histogram rotated 90 degrees (either clockwise or counterclockwise). The subject is required to determine whether the two histograms would be identical if no rotation was applied by tapping either the "Same" or "Different" button as quickly as possible.
Matching to Sample (MTS)	A single 4 × 4 checkerboard pattern is presented on the screen for a study period of 3000 ms. It then disappears for 5 seconds, after which two patterns are presented side-by-side. The subject indicates which of these two patterns was displayed during the study period by tapping on the checkerboard that they believe is identical to the originally presented stimulus as quickly as possible.

66% correct on a given subtest or set of subtests, only those observations are excluded, but the remainder of their record is included in analysis. The main outcome variable in this study is "throughput," a speed-accuracy product that quantifies the number of correct responses per minute (Thorne, 2006):

$$\text{Accuracy} \times \text{Speed} \times 60,000$$

where accuracy is the proportion of correct responses, speed is the reciprocal of mean correct reaction time, and the scaling factor of 60,000 converts the quantity to units of min^{-1}.

Participants were also administered three psychological assessments: the Posttraumatic Stress Disorder Checklist-Military (PCL-M), a military-specific posttraumatic stress disorder assessment (McDonald & Calhoun, 2010), the Patient Health Questionnaire 8 (PHQ-8), a depression diagnostic (Kroenke et al., 2009), and the Pittsburgh Sleep Quality Index (PSQI), a measure of sleep disturbance/insomnia (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989).

Analysis

The goal of this report is to examine the effects of PCL-M (PTSD), PHQ-8 (depression), and PSQI (insomnia) scores on cognitive performance, and to understand the impact of these measures on normative data in active duty military personnel.

We assess relationships between throughput and scores on the three psychological assessments with regression models that control for age and gender. Given this strategy, an immediate issue to consider is the expected pattern of comorbidities among the disorders that these measures assess. For example, the relationships between PTSD and depression (e.g., Brady, Killeen, Brewerton, & Lucerini, 2000; O'Donnell, Creamer, & Pattison, 2004) and PTSD and disturbed sleep (e.g., Leskin, Woodward, Young, & Sheikh, 2002; Mysliwiec et al., 2013) have received considerable attention in the literature. If present in our sample, evidence of these associations would have important implications for both the conceptual and data-analytic components of our study. Accordingly, we first examined correlations among PCL-M, PHQ-8, and PSQI scores to inform our general strategy for modeling the data.

Statistical analyses were carried out under R version 3.3.1 (R Core Team, 2016). Regression models were fit via ordinary least squares, and visual inspection of fitted vs. residual and normal quantile-quantile plots indicated that model assumptions were adequately met, and consideration of Cook's distance revealed that no data points had undue influence on parameter estimates. Gender and age were considered potential confounders and included as covariates, and they were treated as categorical and dummy coded. For age, the 18–19 band serves as the reference level, and the female gender category serves as the reference for gender. Reported coefficients are unstandardized so that effect sizes can be interpreted in terms of throughput, the unit of interest.

Results

Table 2 provides a breakdown with marginal totals of the final sample age-gender distribution using only PCL-M scores as predictor variables.[1]

DANA throughput values are approximately normally distributed, but this is not true of scores on the administered psychological assessments. Table 3 provides descriptive statistics for these outcomes. The

[1]These age groups are hard-coded into the demographic questionnaire included in DANA and were thus not devised with reference to the present analysis.

Table 2. Sample age and gender distribution.

	18–19	20–24	25–29	30–34	35–44	45–54	55–64	Total
Male	99	100	100	94	91	58	29	571
Female	24	94	51	29	23	13	3	237
Total	123	194	151	123	114	71	32	**808**

Table 3. Descriptive statistics for administered psychological assessments.

Assessment	Possible range	Min	1st quartile	Median	3rd quartile	Max	Mean
PCL-M	17–85	17.00	17.00	20.00	26.00	85.00	24.07
PHQ-8	0–24	0.00	0.00	2.00	4.00	22.00	2.80
PSQI	0–21	0.00	2.00	4.00	7.00	16.00	4.79

Table 4. Pairwise correlations between psychological assessments.

	PCL-M	PHQ-8
PCL-M		
PHQ-8	.74***	
PSQI	.66***	.66***

Note. ***$p < .001$.

distributions of scores for all three assessments are highly right-skewed, with extreme scores on the high end of the distribution suggesting outliers. We take this asymmetry to reflect the general incidence of the disorders they measure and note that including all scores in the regression analyses described below yielded models that display an adequate linear fit to the data, suggesting that while relatively few in number, these extreme scores are nonetheless principled.

As Table 4 indicates, scores on these assessments are highly correlated. Given these correlations, multicollinearity is likely to prevent isolation of the unique effect of each psychological assessment score on throughput outcomes. We therefore employed a hierarchical procedure in which two regression models were fit for each DANA subtest. The first regresses throughput on the control variables (age and gender) and PCL-M scores.[2] The second is identical but was further specified to include covariates for PHQ-8 and PSQI scores. We used the first model to examine the coefficient associated with PCL-M scores, and the second was compared to the first, allowing assessment of whether PHQ-8 and PSQI scores significantly impact throughput beyond the effect of PCL-M scores alone.[3]

Results of the PCL-M-only models reveal a significant negative effect of PCL-M scores for five of the eight DANA subtests: SRT1, PRT, GNG, CSL and SRT2 (Table 5). Consistent with findings from other normative studies of neurocognitive assessment (e.g., Vincent et al., 2012), significant effects of age and gender are also observed in the expected direction for a number of subtests as well.[4] We also point out that our validity criterion of greater than 66% of trials correctly completed results in different numbers of exclusions depending on the subtest. A particular issue concerns the Code Substitution – Recall (CSR) subtest, where 13% of the observations were excluded as invalid. This is likely due to this subtest's greater difficulty relative to others.

Figure 1 plots age- and gender-adjusted PCL-M slopes to facilitate a comparison of effect sizes across the DANA subtests where the PCL-M coefficient reached significance.

For each subtest, the PCL-M-only model was compared to a model including covariates for PHQ-8 and PSQI scores via ANOVA. The results of these analyses show that inclusion of PHQ-8 and PSQI covariates provides minimal additional explanatory power over inclusion of the PCL-M covariates alone (Table 6), showing that the PCL-M covariates account for a large majority of the variance among psychological assessment measures.

Discussion

This report focused on the task of establishing a normative dataset of neurocognitive performance for the active-duty military population. This work was driven in part by previous research describing the incidence and consequences of behavioral issues in active-duty military samples. For example, Spira et al. (2014) documented negative relationships between PCL-M and PHQ-8 scores and concussion history and

[2]We chose to include PCL-M scores as a predictor in our reduced models under the assumption that it would explain the most variance in the data relative to PHQ-8 and PSQI scores. This assumption is based on the implicational relationships present among the three disorders these measures assess: in terms of symptoms, PTSD can imply depression and sleep issues, but the reverse is not necessarily true. PCL-M scores were scaled from their original range (17–85) to 0–68 to facilitate interpretation of the intercept term.

[3]We adopted this strategy because multicollinearity issues are only problematic for inference on individual coefficients within a model and not inference on model comparison as summarized by an ANOVA F-ratio.

[4]Treating PCL-M scores as a continuous variable renders the presentation stratified normative tables impossible. However, an individual's mean throughput values can be predicted from the regression equations. For example, the expected SRT1 throughput value for a 32 year-old male with a PCL-M score of 30 can be calculated as follows: $182.34 - 0.38*(30-17) + 9.42 - 3.69 = 183.13$. See Van Breukelen & Vlaeyen (2005) for more detail on the regression-based approach to normative data.

Table 5. Coefficient estimates and (standard errors) by DANA subtest.

β	SRT1 (N = 804)	PRT (N = 805)	GNG (N = 761)	CSL (N = 803)
Intercept	182.34 (3.20)***	95.51 (1.55)***	116.30 (2.14)***	45.64 (0.87)***
PCL-M	−0.38 (0.11)***	−0.22 (0.05)***	−.015 (0.07)*	−0.11 (0.03)***
Male	9.42 (2.30)***	2.66 (1.11)	4.95 (1.53)**	1.12 (0.62)
Age: 20–24	5.00 (3.37)	−1.18 (1.63)	−3.31 (2.23)	−0.03 (0.92)
Age: 25–29	3.17 (3.50)	0.52 (1.69)	−2.92 (2.32)	0.49 (0.95)
Age: 30–34	−3.69 (3.66)	−2.63 (1.77)	−4.14 (2.41)	−2.26 (0.99)*
Age: 35–44	−7.00 (3.73)	−2.73 (1.80)	−7.40 (2.45)**	−4.17 (1.02)***
Age: 45–54	−15.32 (4.28)***	−11.11 (2.07)***	−21.37 (2.87)***	−9.68 (1.17)***
Age: 55–64	−23.27 (5.75)***	−16.23 (2.78)***	−25.40 (3.87)***	−12.60 (1.62)***
β	CSR (N = 701)	MTS (N = 759)	SP (N = 798)	SRT2 (N = 793)
Intercept	46.62 (1.27)***	33.43 (0.87)***	33.55 (0.81)***	172.06 (3.46)***
PCL-M	−0.07 (0.04)	−0.04 (0.03)	−0.02 (0.03)	−0.61 (0.12)***
Male	1.65 (0.91)	1.96 (0.62)**	2.18 (0.58)***	9.65 (2.50)***
Age: 20–24	1.62 (1.32)	−0.80 (0.91)	−1.69 (0.58)*	1.92 (3.65)
Age: 25–29	1.01 (1.36)	0.69 (0.95)	−1.39 (0.87)	3.00 (3.80)
Age: 30–34	−0.29 (1.43)	0.20 (0.99)	−2.60 (0.93)**	2.87 (3.96)
Age: 35–44	−4.45 (1.45)**	−1.09 (1.00)	−4.23 (0.94)***	1.06 (4.03)
Age: 45–54	−10.54 (1.71)***	−4.95 (1.16)***	−7.57 (1.08)***	−3.68 (4.65)
Age: 55–64	−13.00 (2.39)***	−5.49 (1.60)***	−8.50 (1.45)***	−9.55 (6.20)

Note. The *N* for each subtest varies due to excluding administrations where less than 66% of trials were correctly completed.
*$p < .05$. **$p < .01$. ***$p < .001$.

neurocognitive performance, a result that suggests behavioral issues may impact neurocognitive abilities in this population in a more direct fashion. The potential implications of these questions can be appreciated in the context of results from a meta-analysis of 25 studies that estimated the prevalence of DSM-IV major depression among U.S. military personnel. That study showed that the prevalence of depression was 12.0% among current deployed personnel and 13.1% among previously-deployed personnel (Gadermann et al., 2012). Behavioral issues such as depression negatively correlate with cognitive function (e.g., slowed reaction time; Azorin, Benhaim, Hasbroucq, & Possamaï, 1995); therefore, the relatively high prevalence of these psychological risk factors suggests that they may play a potentially large role in influencing the distribution of "normative" cognitive function measures in active-duty military personnel.

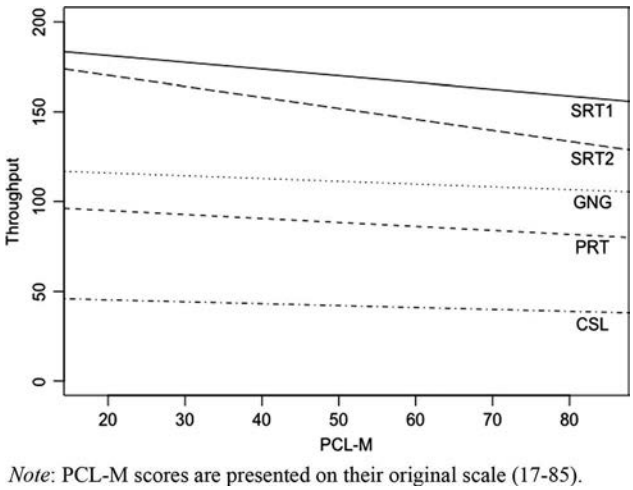

Figure 1. Age- and gender-adjusted slope effects of PCL-M score on throughput for subtests where the effect was significant. *Note.* PCL-M scores are presented on their original scale (17–85).

Our data showed that scores on the PCL-M, an instrument that assesses PTSD severity, were also negatively associated with performance on a number of neurocognitive tests in the DANA battery. This basic observation extends the implications of our findings concerning establishment of normative data in the active duty military population. We also observed strong correlations among PCL-M, PHQ-8, and PSQI scores. These correlations reflect an expected comorbidity pattern among these factors and suggest that these factors tend to cluster among affected individuals. We also found that PHQ-8 and PSQI scores did not account for a significant portion of the variance in throughput beyond what is contributed by PCL-M scores alone. Thus, although PTSD severity, sleep disturbance, and depression are associated with one another, in our sample, the impact of these factors on neurocognitive performance can be explained adequately using only PCL-M.

An immediate application of these findings concerns the establishment of normative datasets for neurocognitive performance among active-duty military personnel. The

Table 6. Comparisons of PCL-M-only models and fully specified models.

	Adjusted r^2			
	PCL-M only	PCL-M + PSQI and PHQ-8	ANOVA	*p*
SRT1	0.08	0.08	F(2, 793) = 1.56	0.21
PRT	0.10	0.10	F(2, 794) = 0.94	0.39
GNG	0.12	0.12	F(2, 750) = 0.15	0.86
CSL	0.18	0.18	F(2, 792) = 0.41	0.67
CSR	0.13	0.12	F(2, 690) = 0.36	0.70
MTS	0.05	0.05	F(2, 748) = 0.22	0.80
SP	0.10	0.10	F(2, 787) = 2.32	0.10
SRT2	0.05	0.05	F(2, 782) = 0.50	0.61

consequences of neurobehavioral disorders on testing results have not been adequately examined in this setting. We suspect this gap in the literature is due in part to the assumption than an "active-duty" designation implies a population free of psychiatric disorders that would affect performance on neurocognitive assessments (e.g., Roebuck-Spencer et al., 2013). To our knowledge, this is the first study to this issue. Our findings challenge the basic assumption that an active-duty designation is sufficient to define the population from which normative data can be derived for military personnel.

In particular, our models show that increasing PCL-M scores correlate negatively with throughput for some DANA subtests. If the effect of these scores is not controlled for, then the false negative rate for psychologically healthy individuals will increase. This is because any "nonnormal" threshold (e.g., greater than two standard deviations below the mean, etc.) itself depends on the location of the distribution. If the mean is downwardly biased, then more individuals who are "nonnormal" in the unbiased distribution will be classified as "normal" when the biased distribution is utilized for comparison. The extent of misclassification due to this bias depends on its effect size, with larger effects resulting in a greater number of misclassifications.

Extending beyond the issue of what population should be used to define normative neuropsychological data among active-duty military personnel, results of this study have implications for a more basic issue concerning the utility of normative neuropsychological data: specifically, the fundamental challenge of knowing which features of a population to measure and control for to ensure that a normative data set truly represents the performance of normally functioning individuals. While it is possible to identify many features by examining both domain- and population-specific factors that might reasonably be expected to affect cognitive performance, others will invariably be missed.

A final issue concerns the interpretation of our results in terms of what constitutes a "normal" reference sample. If the incidence of PTSD and related comorbidities is relatively high in a population, then can a sample that includes affected individuals be considered "normal?" If the complete sample were utilized, it would be necessary to control for the effects of these disorders (e.g., via regression-based norms with appropriate covariates or via stratified tables of conditional means). On the other hand, it can be argued that it would be appropriate to exclude these individuals if their psychological features are thought to be generally uncharacteristic of the target population under consideration.

In our data, 13% of participants scored within clinical range on the PCL-M (≥34 for "moderate PTS"), 6% within clinical range on the PHQ-8 (≥10 for "major depression") and 48% within clinical range on the PSQI (≥5 for "poor sleep quality"). If the definition "normal" for this population is based in part on these frequencies, then it might be difficult to justify excluding individuals with evidence of poor sleep quality since they comprise nearly half of the sample. On the other hand, relatively few participants scored within the clinical range for depression; therefore, they might be considered "outside the norm" and excluded from the sample with minimal loss of power related to the reduction in sample size. Although the purpose of this report was not to argue a position on this issue, we highlight this particular practical consequence of our findings, which reflects the original purpose for pursuing this line of investigation.

Two important limitations of this study should be addressed. First, evidence of neurobehavioral issues (PTSD, depression, and disturbed sleep) were obtained via self-report, and no formal diagnoses were obtained. Although the instruments used to assess these disorders have been validated, they are not perfectly predictive of clinical diagnosis. However, if the scores obtained on these assessments are in some cases not truly reflective of the underlying construct they are meant to measure, this issue only affects the interpretation of our results and not their utility. The absence of clinical diagnosis does not alter the fact that participants' self-assessments correlate negatively with neurocognitive performance; only the link between these measures and the construct they assess is under question.

A related issue is that we considered only a limited number of neurobehavioral issues and a limited number of instruments for their assessment. It is possible that other assessments, or a combination of assessments, would provide a more accurate diagnostic of psychological disorder. Furthermore, there are other disorders, such as anxiety, which we did not assess that may also negatively impact neurocognitive testing results. We expect that future research efforts will consider this issue in more detail.

Conclusion

This report examined key underlying assumptions concerning the use of normative data for cognitive testing in active-duty military personnel. We show that scores on assessments for PTSD, depression, and disturbed sleep—psychological issues that occur with relatively high frequency among active-duty personnel—correlate negatively with quantitative measures of cognitive efficiency. These psychological factors may therefore skew the distributions of

cognitive efficiency measures in large samples of seemingly healthy military personnel in ways that could affect "nonnormal" classifications.

Declaration of interest

AnthroTronix, Inc. is the developer of DANA.

Funding

This study was funded by The Geneva Foundation (award # W81XWH-09-2-0057) and a Research Innovation Fund awarded by the United States Army Medical Research Acquisition Activity (award # W81XWH-15-C-0141).

References

Azorin, J. M., Benhaim, P., Hasbroucq, T., & Possamaï, C. A. (1995). Stimulus preprocessing and response selection in depression: A reaction time study. *Acta Psychologica*, *89*(2), 95–100. doi:10.1016/0001-6918(94)00024-b

Brady, K. T., Killeen, T. K., Brewerton, T., & Lucerini, S. (2000). Comorbidity of psychiatric disorders and posttraumatic stress disorder. *Journal of Clinical Psychiatry*, *61*, 22–32.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. doi:10.1016/0165-1781(89)90047-4

Gadermann, A. M., Engel, C. C., Naifeh, J. A., Nock, M. K., Petukhova M., Santiago, P. N., … Kessler, R. C. (2012). Prevalence of DSM-IV major depression among U.S. military personnel: Meta-analysis and simulation. *Military Medicine*, *177*(80), 47–59. doi:10.7205/milmed-d-12-00103

Ganguli, M., Snitz, B. E., Lee, C. W., Vanderbilt, J., Saxton, J. A., & Chang, C. C. H. (2010). Age and education effects and norms on a cognitive test battery from a population-based cohort: The Monongahela–Youghiogheny healthy aging team. *Aging and Mental Health*, *14*(1), 100–107. doi:10.1080/13607860903071014

Heaton, R. K., Grant, I., & Matthews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. *Neuropsychological assessment of neuropsychiatric disorders*, 100–120.

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, *351*(1), 13–22. doi:10.1056/nejmoa040603

Horner, M. D., & Hamner, M. B. (2002). Neurocognitive functioning in posttraumatic stress disorder. *Neuropsychology review*, *12*(1), 15–30.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1), 163–173. doi:10.1016/j.jad.2008.06.026

Lathan, C., Spira, J. L., Bleiberg, J., Vice, J., & Tsao, J. W. (2013). Defense automated neurobehavioral assessment (DANA)-psychometric properties of a new field-deployable neurocognitive assessment tool. *Military Medicine*, *178*(4), 365–371. doi:10.7205/milmed-d-12-00438

Leskin, G. A., Woodward, S. H., Young, H. E., & Sheikh, J. I. (2002). Effects of comorbid diagnoses on sleep disturbance in PTSD. *Journal of Psychiatric Research*, *36*(6), 449–452. doi:10.1016/s0022-3956(02)00025-0

McDonald, S. D., & Calhoun, P. S. (2010). The diagnostic accuracy of the PTSD checklist: A critical review. *Clinical Psychology Review*, *30*(8), 976–987. doi:10.1016/j.cpr.2010.06.012

Mysliwiec, V., McGraw, L., Pierce, R., Smith, P., Trapp, B., & Roth, B. J. (2013). Sleep disorders and associated medical comorbidities in active duty military personnel. *Sleep*, *36*(2), 167–174. doi:10.5665/sleep.2364

O'Donnell, M. L., Creamer, M., & Pattison, P. (2004). Posttraumatic stress disorder and depression following trauma: understanding comorbidity. *American Journal of Psychiatry*, *161*(8), 1390–1396. doi:10.1176/appi.ajp.161.8.1390

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://R-project.org.

Roebuck-Spencer, T. M., Vincent, A. S., Schlegel, R. E., & Gilliland, K. (2013). Evidence for added value of baseline testing in computer-based cognitive assessment. *Journal of Athletic Training*, *48*(4), 499–505. doi:10.4085/1062-6050-48.3.11

Ross, T. P., & Lichtenberg, P. A. (1998). Expanded normative data for the Boston Naming Test for use with urban, elderly medical patients. *The Clinical Neuropsychologist*, *12*(4), 475–481.

Russo, C. R., & Lathan, C. E. (2015). An evaluation of the consistency and reliability of the defense automated neurocognitive assessment tool. *Applied Psychological Measurement*, *39*(7), 566–572. doi:10.1177/0146621615577361

Schneider, A. L., Sharrett, A. R., Gottesman, R. F., Coresh, J., Coker, L., Wruck, L., … Mosley, T. H. (2015). Normative data for eight neuropsychological tests in older blacks and whites from the atherosclerosis risk in communities (ARIC) study. *Alzheimer Disease and Associated Disorders*, *29*(1), 32–44. doi:10.1097/wad.0000000000000042

Spira, J. L., Lathan, C. E., Bleiber, J., & Tsao, J. W. (2014). The impact of multiple concussions on emotional distress, post-concussive symptoms, and neurocognitive functioning in active duty United States marines independent of combat exposure or emotional distress. *Journal of Neurotrauma*, *31*, 1823–1834. doi:10.1089/neu.2014.3363

Swick, D., Honzel, N., Larsen, J., Ashley, V., & Justus, T. (2012). Impaired response inhibition in veterans with post-traumatic stress disorder and mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *18*(05), 917–926. doi:10.1017/s1355617712000458

Thorne, D. R. (2006). Throughput: A simple performance index with desirable characteristics. *Behavior Research Methods*, *38*(4), 569–573. doi:10.3758/bf03193886

Van Breukelen, G. J., & Vlaeyen, J. W. (2005). Norming clinical questionnaires with multiple regression: The pain cognition list. *Psychological Assessment*, *17*(3), 336–344. doi:10.1037/1040-3590.17.3.336

Vincent, A. S., Roebuck-Spencer, T., Gilliland, K., & Schlegel, R. (2012). Automated neuropsychological assessment metrics (v4) traumatic brain injury battery: Military normative data. *Military Medicine*, *177*(3), 256–269. doi:10.7205/milmed-d-11-00289