Journal of the International Neuropsychological Society (2018), **24**, 156–162.

doi:10.1017/S1355617717001187

Intraindividual Cognitive Variability: An Examination of ANAM4 TBI-MIL Simple Reaction Time Data from Service Members with and without Mild Traumatic Brain Injury

Wesley R. Cole,[1,2,3] Emma Gregory,[1,3] Jacques P. Arrieux,[1,2,3] and F. Jay Haran[4]

[1]Defense and Veterans Brain Injury Center, Silver Spring, Maryland and Fort Bragg, North Carolina
[2]Womack Army Medical Center, Fort Bragg, North Carolina
[3]General Dynamics Health Solutions, Fairfax, Virginia
[4]Uniformed Service University of the Health Sciences, Bethesda, Maryland

(Received February 24, 2017; Final Revision May 26, 2017; Accepted July 18, 2017; First Published Online November 23, 2017)

Abstract

Objectives: The Automated Neuropsychological Assessment Metrics 4 TBI-MIL (ANAM4) is a computerized cognitive test often used in post-concussion assessments with U.S. service members (SMs). However, existing evidence remains mixed regarding ANAM4's ability to identify cognitive issues following mild traumatic brain injury (mTBI). Studies typically examine ANAM4 using standardized scores and/ or comparisons to a baseline. A more fine-grained approach involves examining inconsistency within an individual's performance (i.e., intraindividual variability). **Methods:** Data from 237 healthy control SMs and 105 SMs within seven days of mTBI who took the ANAM4 were included in analyses. Using each individual's raw scores on a simple reaction time (RT) subtest (SRT1) that is repeated at the end of the battery (SRT2), we calculated mean raw RT and the intraindividual standard deviation (ISD) of trial-by-trial RT. Analyses investigated differences between groups in mean RT, RT variability (i.e., ISD), and change in ISD from SRT1 and SRT2. **Results:** Using regression residuals to control for demographic variables, analysis of variance, and pairwise comparisons revealed the control group had faster mean RT and smaller ISD compared to the mTBI group. Furthermore, the mTBI group had a significant increase in ISD from SRT1 to SRT2, with effect sizes exceeding the minimum practical effect for comparisons of ISD in SRT2 and change in ISD from SRT1 to SRT2. **Conclusions:** While inconsistencies in performance are often viewed as test error, the results suggest intraindividual cognitive variability may be more sensitive than traditional metrics in detecting changes in cognitive function after mTBI. Additionally, the findings highlight the utility of the ANAM4's repeating a RT subtest at two points in the same session for exploring within-subject differences in performance variability. (*JINS*, 2018, *24*, 156–162)

Keywords: Concussion, Military, Computerized assessment, Intraindividual variability cognitive function, Reaction time, Neuropsychology

INTRODUCTION

Every year thousands of service members (SMs) in the U.S. military are diagnosed with a mild traumatic brain injury (mTBI), also known as concussion (Defense and Veterans Brain Injury Center [DVBIC], 2016). These injuries can take place in a variety of settings due to several causes, including those similar to sports-related concussion in the civilian sector. Regardless of where or how concussion occurs, there is a need for timely and effective evaluation of an individual's cognitive functioning (Kelly, Coldren, Parish, Dretsch, & Russell, 2012). Assessment of cognitive abilities via neuropsychological (NP) tests is considered the cornerstone of concussion management (McCrory et al., 2013). However, these tests are time consuming and require particular expertise for administration and interpretation of results. In more recent years, computerized neurocognitive assessment tools (NCATs) have been increasingly used as a quicker and more feasibly administered alternative to NP tests (Friedl et al., 2007; McCrory et al., 2013).

The Automated Neuropsychological Assessment Metrics 4 TBI-MIL (ANAM4) is an NCAT developed by the U.S. Army (Friedl et al., 2007) and widely used in the military (Defense Health Board, 2016). ANAM4 is regularly

---

Correspondence and reprint requests to: Wesley R. Cole, Intrepid Spirit, Womack Army Medical Center, Fort Bragg, NC 28310. E-mail: wesley.r.cole.ctr@mail.mil

administered before a deployment as a means to generate a neurocognitive baseline for post-deployment and post-injury comparison (DoDi 6490.13). Despite the goal of NCATs, including ANAM4, existing evidence is inconclusive regarding the ability to identify cognitive issues following concussion (see Arrieux, Cole, & Ahrens, 2017; Resch, McCrea, & Cullum, 2013).

Typically findings from ANAM4 are based on analyses comparing post-injury scores either to individual baseline measurements or normative databases (see Haran et al., 2016; McCrea et al., 2008). The current analyses focus on within-person inconsistent performance, or intraindividual neurocognitive variability, within a single test session, as a metric potentially better suited to detect the cognitive effects of mTBI. Variability has been described in multiple ways, but often relates to three principles: persons, measures, and occasions (Hultsch, MacDonald, & Dixon, 2002). Inter-individual variability, or "diversity," often measures differences between *persons* or groups. An individual's variability across multiple *measures* can be thought of as intraindividual differences or "dispersion." The focus of the current analyses is intraindividual variability (IIV), or an individual's variability on the same test across multiple *occasions*, referred to as "inconsistency."

Although IIV is often viewed as noise or test error, it may in fact reflect fluctuation in cognitive processing and reveal cognitive deficits that a mean or standard score is attempting, but failing, to capture. For example, research in aging populations has shown IIV on various behavioral and neurophysiological measures to be associated with decline in cognitive performance (Fjell, Rosquist & Walhovd, 2009; Hultsch et al., 2002; Lovden, Li, Shing, & Linderberger, 2007). Although the literature base is relatively small, IIV in acute and post-acute concussion populations has been studied for more than two decades using both traditional NP and reaction time (RT) tests (Rabinowitz & Arnett, 2013; Sosnoff, Broglio, Hillman, & Ferrara, 2007; Stuss et al., 1989).

Using NP tests, Hill, Rohling, Boettcher, and Meyers (2013) analyzed IIV using means from the Meyers Neuropsychological Battery in individuals reporting a history of mTBI and found that overall performance is negatively correlated with variability. Similarly, in a study using RT-based stimulus discrimination and flanker tests, history of concussion was shown to be associated with increased IIV (Parks et al., 2015). Beyond behavioral measures, Segalowitz, Dywan, and Unsal (1997) demonstrated for a TBI group, and not for a control group, RT variability was related to electrophysiological measures of attentional allocation and sustainment (the P300 amplitude and the preresponse component of the contingent negative variation E-Wave), supporting the idea that RT variability reflects this attentional processing.

Studies have also examined IIV in TBI using NCATs. Bleiberg, Garmoe, Halpern, Reeves, and Nadler (1997) demonstrated participants with mild to moderate TBI performed more inconsistently in same-day and across multiple day sessions than a healthy control group. Makdissi et al. (2001) investigated a simple RT test in a different NCAT, CogState, in athletes and found greater standard deviation in RT in acutely concussed *versus* never concussed athletes at follow-up, although not at baseline. However, longer RT in concussed participants as compared to controls could account for greater standard deviation in RT. Sosnoff et al. (2007) adjusted for mean RT in a group of individuals tested within 72 hr of concussion and found that after this adjustment, concussed individuals did not have greater RT standard deviation than healthy age- and gender-matched individuals.

The above studies, most of which demonstrate an ability to differentiate TBI and control group performance using IIV measures, all compare an individual's performance on a test (i.e., *measures*) or whole battery across test sessions (i.e., *occasions*). In contrast, the present investigation explores potential differences in IIV *within* a single test session by comparing performance on one subtest repeated within a battery, in patients with acute concussion and healthy controls. Our approach allows examination of the use of IIV analyses within an abbreviated window and without a need for repeat testing of an entire battery. The ANAM4 is an ideal test to examine IIV in this way, as unlike most NCATs, the ANAM4 includes an identical simple RT (SRT) task at the beginning and the end of the battery.

Although the ANAM4 standard output generates the RT standard deviation on each subtest, our approach differs because it examines the standard deviation of the difference between the trial-by-trial RT data. This approach allows for a more fine-grained measure of IIV and an individual's change in RT and RT variability over a brief period of time. In addition to looking at trial-by-trial raw RT data, the current study investigated acutely concussed individuals, as previous research suggests ANAM4 has limited clinical utility more than eight days following concussion, as well as healthy controls (e.g., Nelson et al., 2016). We hypothesize that this alternative trial-by-trial approach to interpreting RT on ANAM4 will reveal differences in IIV (i.e., differences in "inconsistency") across the two groups. As a secondary objective, we use interindividual differences (i.e., "diversity") to investigate the potential impact demographic variables may have on any differences identified in IIV.

METHODS

Sample

A total sample of 356 individuals was selected from a larger study's sample of SMs from Fort Bragg with and without mTBI where ANAM4 was administered (Cole, Arrieux, Dennison, & Ivins , 2017; Cole, Arrieux, Ivins, Schwab, & Qashu, 2017). Informed consent was obtained from all subjects and data were collected in compliance with Womack Army Medical Center's Institutional Review Board regulations and requirements. The sample included 240 healthy controls (CTRL) and 116 participants within 7 days of a medically documented mTBI. The mean time since injury was 4.8 days (range, 0–7 days). All injuries were sustained on or around Fort Bragg (i.e., no combat related injuries) with most injuries sustained due to hard landings during parachute training jumps (85.3%), and the remaining injuries (all <5%) due to motor vehicle crashes, falls, assaults, sports-related injury, or blast exposures during training exercises.

Instrumentation

The ANAM4 (CSRC, 2014) is an automated, computerized neurocognitive test battery that includes a sleepiness scale, mood scale, a self-report TBI questionnaire, and seven core subtests: Code Substitution Delayed (CDD), Code Substitution (CDS), Matching-to-Sample (M2S), Mathematical Processing (MTH), Procedural Reaction Time (PRO), Simple Reaction Time (SRT1), and Simple Reaction Time Repeated (SRT2). Due to the larger study's procedures, an additional battery of questionnaires was administered before testing, with the seven core ANAM4 subtests administered per usual procedures following completion of questionnaires (Cole, Arrieux, Ivins, et al., 2017). Validity of the data was evaluated by an embedded effort index (EI), which flags atypical scores based on accuracy and discrepancy of responses (Roebuck-Spencer, Vincent, Gilliland, Johnson, & Cooper, 2013). Specifically, the ANAM4 EI assesses accuracy and RT discrepancy on four of the battery's subtests, which are transformed to weighted scores based on the infrequency of those scores. Weighted scores range from 0–48, and scores above 14 are considered invalid (Roebuck-Spencer et al., 2013). For the purposes of this manuscript, only the EI and the raw data from the SRT1 and SRT2 were used in the analyses. The raw data for the SRT1 and SRT2 consisted of 40 trials for each subtest.

Data Processing

To prepare the SRT1 and SRT2 data for the analyses we first removed any participant who was deemed to have invalid performance by the ANAM4 EI. Fourteen participants, 11 in the mTBI group (9.5%) and 3 in the CTRL group (1.3%), were flagged by the ANAM4 EI and removed from the sample. The resulting sample size was 342 participants (13,680 trials) with 237 participants (9480 trials) in the CTRL group and 105 participants (4200 trials) in the mTBI group. Second, extremely fast and slow responses, potentially indicating common key press errors (e.g., accidental key presses or interruption of the task) were trimmed from the dataset per commonly used procedures (Batterham, Bunce, Mackinnon, & Christensen, 2014; Dixonet al., 2007; Garret, MacDonald, & Craik 2012; Hultsch et al., 2002; Hultsch, Strauss, Hunger, & MacDonald 2008). Specifically, a lower bound for responses was set at 150 ms, with a total of 275 (2.0%) and 450 (3.3%) trials trimmed for the SRT1 and SRT2, respectively. An upper bound was set for each individual, with any trials exceeding a within-subject subtest mean of plus three standard deviations trimmed, for a total of 258 (1.9%) and 292 (2.1%) trials for the SRT1 and SRT2, respectively. To maintain a complete dataset, trimmed values were imputed using a linear interpolation procedure from the relationships among all trials from all participants in the dataset (Hultsch et al., 2002).

Statistical Analyses

Four distinct statistical analyses were conducted to address our study objectives. The first analysis investigated group differences on demographic variables and established residuals to control for any differences in subsequent between-groups analyses. The second analysis aimed to identify differences between the CTRL and mTBI groups on SRT1 and SRT2 performance (i.e., diversity). The third analysis investigated if there were differences between the CTRL and mTBI groups in terms of variability in RT on both SRT1 and SRT2 (i.e., dispersion). The fourth analysis, addressing the primary aim of this study, investigated if there were between-groups differences in within-persons change in RT and RT variability from SRT1 to SRT2 (i.e., inconsistency).

Group differences for demographic data were examined using Mann-Whitney *U* Tests and Chi-Square tests. There were minor violations of the Lilliefors test of normality for the simple reaction subtest data; however, the potential for a familywise type I error due to multiple comparisons was accounted for with sample sizes sufficient enough (i.e., $n > 30$) for the central-limit theorem to apply.

Levene's test was used to assess between-groups variability (i.e., diversity) in SRT1 and SRT2 performance. Any observed differences between groups on variability may be an artifact of group differences in mean performance, as larger standard deviations tend to be associated with larger means (Hale, Myerson, Smith, & Poon, 1988). Factors such as sex and age have been reported to confound RT mean performance and variability (Der & Deary, 2006). To control for these possible effects of sex and military rank, which is highly correlated with age, four separate linear regression procedures were used to calculate residuals for SRT1 and SRT2 (for both the CTRL and mTBI groups). The results were interpreted using the following criteria for squared association indices: recommended minimum practical effect size ($r^2 = 0.04$), moderate effect ($r^2 = 0.25$), and strong effect ($r^2 = 0.64$) (Ferguson, 2009).The absolute values of the resulting residuals were used in within-persons variability (i.e., IIV) calculations.

IIV can be used to examine both dispersion on a task (i.e., across the 40 trials within each of the ANAM4 SRT subtests) and to examine inconsistency across time (i.e., change across SRT1 and SRT2). While there are numerous indices that can be used to compute IIV, the simplest is the intraindividual standard deviation (ISD). ISDs are standard deviations of RT variability within each of SRT1 and SRT2, calculated for each individual using the purified residual scores.

To examine group differences in ISD dispersion, two separate general linear model (1 × 2) analyses [analysis of variance (ANOVA)] were performed, one each for SRT1 and SRT2, with group membership (two levels) as the between-subjects variable. Effect size (ES) for group differences was calculated using the partial eta squared (η_p^2) and the results were interpreted using the aforementioned criteria for squared association indices.

To examine group differences in ISD inconsistency, a general linear model (2 × 2) ANOVA with repeated measures was performed, with group membership (two levels; CTRL *vs*. mTBI) as the between-subjects variable and subtest (two levels; SRT1 *vs*. SRT2) as the within-subjects variable. Pairwise comparisons were conducted to follow-up significant main effects. Significant interaction effects were explored using *post hoc* comparisons (one-way ANOVAs with repeated measures). ES were calculated using the partial eta squared (η_p^2) and the results were interpreted using the aforementioned criteria for squared association indices.

All analyses were performed with Matlab 2015b (Mathworks, Natick, MA) and SPSS Version 22 (IBM, Armonk, NY).

RESULTS

There were significant differences between the CTRL and mTBI groups for all demographic variables (Table 1). These differences are believed to be due to the larger number of officers in the CTRL group. That is, there are more female officers than enlisted personnel, and officers are generally older and more educated than the enlisted population. It is believed that officers were over-represented in the CTRL group due to their greater ability to control and dictate their daily schedules, allowing them to take time off to volunteer in a research study. Sex and rank, likely accounting for all other demographic differences, were controlled for in subsequent analyses, as described above.

Table 2 reports the standard deviations for SRT1 and SRT2 as a function of injury status. It should be noted that the standard deviations for the mTBI group were nearly two and nearly three times those for the CTRL group for SRT1 and SRT2, respectively. Levene's test for the homogeneity of variance indicated significant group differences in variability for both the CTRL ($F_{(1,13,678)} = 848.65$; $p < .0001$) and mTBI ($F_{(1,13,678)} = 1{,}815.71$; $p < .0001$) groups. Taken together these results indicate increased diversity in the RT subtest performance with injury status.

Table 3 reports the results of the regression analyses. For SRT1, there was a significant linear trend for rank and sex in both groups indicating that increased diversity for females and increased diversity with higher rank. For SRT2, there was a significant linear trend for rank, but not sex, in both the CTRL and mTBI groups, indicating increased diversity with increasing rank. In general, the magnitude of the significant trends were modest, all were less than 2% of the variance, and none reached the recommended minimum practical effect size for squared association indices.

The results of the ANOVA performed to examine group differences in dispersion for SRT1 revealed that there was a significant main effect for group membership ($F_{(1,340)} = 10.00$; $p = .002$; $\eta_p^2 = .03$), although it did not reach the recommended minimum practical effect size. For SRT2, the ANOVA revealed a significant main effect that did exceed the recommended minimum practical effect size for group membership ($F_{(1,340)} = 30.72$; $p < .001$; $\eta_p^2 = .08$). Pairwise comparisons revealed that the ISD mean for the CTRL group was significantly lower than the ISD mean for the mTBI group.

The results of the ANOVA performed to examine group differences in inconsistency (i.e., IIV) revealed that there was

Table 1. Participant characteristics for CTRL and mTBI groups.

	CTRL (n = 237)	mTBI (n = 105)	p-value
Characteristic			
Sex, male, n (%)	191 (81%)	103 (98%)	0.0001[b]
Rank, n (%)			
Enlisted	125 (53%)	99 (94%)	0.0001[b]
Officer	112 (47%)	6 (6%)	
Age, years			
Mean	33	26	0.0001[a]
Range	19–58	19–48	
Years Active Duty			
Mean	5	10	0.000[a]
Range	0–24	0–34	
No. of Deployments			
Mean	2	1	0.037[a]
Range	0–11	0–7	
Education Level[c], n (%)			
High School or Less	27 (11%)	45 (43%)	0.0001[a]
Some College	64 (27%)	37 (35%)	
Associate Degree	24 (10%)	10 (10%)	
Bachelor's Degree or Higher	122 (52%)	12 (11%)	

Abbreviations: CTRL = control group; mTBI = mTBI group.

[a] Two-Tailed Mann-Whitney U.

[b] Chi-square test

[c] Data missing for one mTBI participant

Table 2. Standard deviations and means of raw reaction time (RT) and RT intraindividual standard deviation (ISD).

			Group	
Data Type	**ANAM4 Subtest**		**CTRL** *n* = 237	**mTBI** *n* = 105
Raw RT	SRT1	SD	82.27	148.97
		M	288.40	321.98
	SRT2	SD	73.71	187.15
		M	273.79	326.59
ISD	SRT1	SD	40.93	51.61
		M	37.29	53.77
	SRT2	SD	33.28	72.57
		M	39.02	70.72

Abbreviations: CTRL = control group; mTBI = mTBI group; SRT1 = Simple Reaction Time; SRT2 = Simple Reaction Time Repeated

Table 3. Summary table for regression of residuals on linear trends.

Group		**Predictor**	**β**	***R***	**Δ*R*2**	**# of Trials**
CTRL	SRT1	Model	—	0.107	0.012	9480
		Sex	21.89	—	—	
		Rank	−5.92	—	—	
	SRT2	Model				
		Sex	—	0.090	0.008	4200
		Rank	−6.20	—	—	
			−72.81	—	—	
mTBI	SRT1	Model	—	0.13	0.018	9480
		Sex	104.22	—	—	
		Rank	−57.44	—	—	
	SRT2	Model	—	0.090	0.008	4200
		Sex	−6.20	—	—	
		Rank	−72.81	—	—	

Abbreviations: CTRL = control group; mTBI = mTBI group; SRT1 = Simple Reaction Time; SRT2 = Simple Reaction Time Repeated

a significant interaction of group and time ($F_{(1,340)} = 15.87$; $p = .001$; $\eta_p^2 = .03$). The main effects for group ($F_{(1,340)} = 23.75$; $p = .001$; $\eta_p^2 = .07$) and time ($F_{(1,340)} = 15.87$; $p = .001$; $\eta_p^2 = .05$) were also significant and exceeded the recommended minimum for a practical effect. The *post hoc* one-way ANOVA with repeated measures revealed a significant main effect for time that exceeded the recommended minimum for a practical effect for the mTBI group only ($F_{(1,340)} = 11.49$; $p = .001$; $\eta_p^2 = 0.10$). Pairwise comparisons revealed that the ISD mean for the CTRL group was significantly lower (i.e., less IIV) than the ISD mean for the mTBI group, and within the mTBI group the mean ISD for SRT1 was significantly lower than the mean ISD for SRT2.

DISCUSSION

The current study investigated differences in mean RT and RT IIV between CTRLs and those with acute concussion using raw trial-by-trial RT data from the repeated SRT subtests in the ANAM4. This approach was relatively unique as most previous studies have focused on the use of standardized scores and cognitive efficiency metrics (e.g., ANAM4 throughput scores) to investigate group differences. Moreover, prior studies examining differences in RT variability have almost exclusively done so across test sessions rather than using a subtest repeated within the same battery and test session. Our hypotheses were largely supported, as those with acute concussion had slower RTs and greater RT IIV than CTRLs.

While it is not surprising that there were differences in the RT means and IIV between the CTRL and mTBI groups, of interest, the CTRL group demonstrated improved (i.e., faster) RT as well as less IIV in SRT2 compared to SRT1. Although the mTBI group demonstrated longer, but relatively stable mean RT across subtests, the most important finding was that those in the mTBI group had more IIV in both subtests than controls, with IIV actually increasing in the second, repeated subtest. This suggests an apparent practice effect observed in the CTRL group that is not observed in the mTBI group. The lack of a demonstrated practice effect in those recovering from neurological insult could be clinically meaningful when a practice effect is otherwise expected (Lezak, Howieson, Bigler, & Tranel, 2012).

Additionally, RT and its variability have been shown to provide information about the allocation of attentional resources in those with neurological insult such as mTBI. Specifically, it is thought that attention allocation can be measured by RT latency in healthy controls, whereas in those with mTBI attention allocation is more related to RT variability than RT latency (Bleiberg et al., 1997; Segalowitz et al., 1997). As such, the current finding of greater IIV in ANAM4 SRT performance, with increasing IIV across trials, in an acute mTBI group provides additional evidence to the body of literature.

In general, these results reveal greater trial-to-trial fluctuations in performance for the mTBI group as compared to the CTRL group. Based on the central tendency theory, these fluctuations are often viewed as noise, instability, or error. However, they may be indicative of low scores due to the acute effects of concussion that may otherwise be missed by more traditional metrics. That is, analyses of variability in raw RT and ISD of RT trials appear to be a potentially valuable alternative metric for NCATs. That is, these alternative metrics may offer greater clinical utility than metrics commonly used in cognitive testing. Given the computerized nature of NCATs, metrics such as raw RT and RT ISD can be more quickly and feasibly calculated. However, inclusion of these score calculations would require changes to the NCAT's scoring output, as otherwise the responsibility to calculate would fall on the clinician.

Additionally, norms with clinically meaningful cut points would need to be established before such metrics could be applied to clinical decision making. Lastly and of note, ANAM4 presents a conceivable advantage over other NCATs by including a repeated simple reaction time test, allowing comparison of RT and RT variability across time although still within one testing session, potentially tapping into "cognitive fatigue." Other NCATs may benefit from a repeated RT subtest in their battery.

Limitations

The current study was derived from data from a larger study, and, therefore, procedures not relevant to the current analyses surrounded the collection of the data used in this study. These procedures sometimes included other testing before the ANAM4, which could have increased fatigue. However, any potential fatigue would be relatively equitable across groups and relatively controlled for by comparing SRT2 to SRT1, which occurred within the same testing session. Additionally, recent studies demonstrated that, when administering multiple NCATs in one session, performance was not affected by the order of administration (Cole, Arrieux, Dennison, et al., 2017; Nelson et al., 2016).

The sample included in the current analyses was primarily male and mostly enlisted, especially in the mTBI group; moreover the CTRL group was not a matched CTRL group. We attempted to control for potential differences in group composition in our analyses. However, as with any study involving cognitive assessment, there are many additional factors that could potentially impact testing results, and specifically IIV, for which we did not control. Such factors include premorbid functioning, sleep, emotional status (e.g., acute stress reaction or post-traumatic symptoms), the nature of injury, time since injury, ongoing symptomatology, potential medication with cognitive side effects (e.g., stimulants or sedatives), among others.

We believed it was beyond the scope of the current study to attempt to control for the innumerable confounding variables. Additionally, as with all studies of NCATs, there are technological and environmental considerations that could impact testing results, such as the hardware and software configurations and the testing environment. All efforts were taken to administer the tests with a computer platform as close to the ANAM4 manual specifications. Additionally, testing was done in a quiet room with a trained test proctor, in an environment similar to how baseline or post-injury testing would likely occur, likely rendering the results ecologically valid despite the potential for other sources of error.

CONCLUSIONS AND FUTURE DIRECTIONS

The results from this study support a small but growing body of literature that raw RT scores and RT variability may be much more sensitive to the cognitive decline seen during the acute period after concussion. The findings suggest mTBI participants can temporarily perform similarly to normal controls on RT latency, but repeated RT assessments at multiple time points throughout a battery demonstrate a lack of improvement in RT and increased variability. Interpreting these metrics rather than the traditionally reported standardized scores (e.g., throughput), where effects may otherwise be "washed out, appears to hold promise for the use of ANAM4 in acute concussion populations.

However, additional work is needed to fully clarify the clinical utility (e.g., diagnostic and prognostic capabilities) of these metrics and to determine if they do indeed offer advantages over traditional metrics obtained from traditional NP tests and NCATs, especially when controlling for other potential confounding variables. There is some existing evidence that shorter ANAM4 SRT is predictive of recovery in those acutely concussed (Norris, Carr, Herzig, Labrie, & Sams, 2013). Thus, it may be that improved (i.e., faster) raw RT from SRT1 to SRT2, less RT variability within both subtests, and stable RT variability across SRT1 and SRT2 could be predictive of faster and/ or better recovery after concussion and, therefore, incorporated into return to duty or return to play decisions. Given the military and many sports leagues' baseline testing procedures, it will also be important to determine if baseline assessments are valuable with regard to such metrics for diagnostic and prognostic purposes.

ACKNOWLEDGMENTS

The authors thank Karen Schwab, Brian Ivins, Felicia Qashu, Mary Alice Dale, James Wes McGee, Katie Toll, and Alex Fender for their contributions to this research. This material is published by permission of the Defense and Veterans Brain Injury Center, operated by General Dynamics Information Technology for the U.S. Defense Health Agency under Contract No.W91YTZ-13-C-0015. The authors have no financial interests to disclose. The views expressed in this manuscript are those of the author(s) and do not necessarily reflect the official policy or position of the Department of the Navy, Department of the Army, Department of Defense, or the U.S. Government.

REFERENCES

Arrieux, J.P., Cole, W.R., & Ahrens, A.P. (2017). A review of the validity of computerized neurocognitive assessment tools in mild traumatic brain injury assessment. *Concussion*.

Batterham, P.J., Bunce, D., Mackinnon, A.J., & Christensen, H. (2014). Intra-individual reaction time variability and all-cause mortality over 17 years: A community-based cohort study. *Age and Ageing*, *43*, 84–90.

Bleiberg, J., Garmoe, W.S., Halpern, E.L., Reeves, D.L., & Nadler, J.D. (1997). Consistency of within-day and across-day performance after mild brain injury. *Cognitive and Behavioral Neurology*, *10*(4), 247–253.

Cognitive Science Research Center, CSRC. (2014). ANAM Military Battery, Administration Manual. Norman, OK.

Cole, W.R., Arrieux, J.P., Dennison, E.M., & Ivins, B.J. (2017). The impact of administration order in studies of computerized neurocognitive assessment tools (NCATs). *Journal of Clinical and Experimental Neuropsychology*, *39*(1), 35–45.

Cole, W.R., Arrieux, J.P., Ivins, B.J., Schwab, K.A., & Qashu, F.M. (2017). A comparison of four computerized neurocognitive assessment tools to a traditional neuropsychological test battery in service members with and without mild traumatic brain injury. *Archives of Clinical Neuropsychology*, 1–18. doi: 10.1093/arclin/acx036

Defense Health Board. (2016). *Review of the scientific evidence of using population normative values for post-concussive computerized neurocognitive assessments*. Report prepared for the Undersecretary of Personnel and Readiness, Falls Church, VA: Defense Health Board.

Defense and Veterans Brain Injury Center. (2016). *DoD worldwide numbers for TBI worldwide totals*. Retrieved from http://dvbic.dcoe.mil/files/tbi-numbers/DoD-TBI-Worldwide-Totals_2000-2016_Q1_May-16-2016_v1.0_2016-06-24.pdf.

Department of Defense Instruction (DoDI) 6490.13. (2013). *Comprehensive policy on traumatic brain injury-related neurocognitive assessments by the military services*. Washington, DC: Department of Defense.

Der, G., & Deary, I.J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging*, *21*, 62–73.

Dixon, R.A., Garrett, D.D., Lentz, T.L., MacDonald, S.W.S., Strauss, E., & Hultsch, D.F. (2007). Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*, *21*, 381–399.

Fjell, A.M., Rosquist, H., & Walhovd, K.B. (2009). Instability in the latency of P3a/P3b brain potentials and cognitive function in aging. *Neurobiology of Aging*, *30*(12), 2065–2079.

Friedl, K.E., Grate, S.J., Proctor, S.P., Ness, J.W., Lukey, B.J., & Kane, R.L. (2007). Army research needs for automated neuropsychologial tests: Monitoring soldier health and performance status. *Archives of Clinical Neuropsychology*, *22*, 7–14.

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538.

Garrett, D.D., MacDonald, S.W.S., & Craik, F.I.M. (2012). Intraindividual reaction time variability is malleable: Feedback- and education-related educations in variability with age. *Frontiers in Human Neuroscience*, *6*, 1–10. doi: 10.3389/fnhum.2012.00101

Hale, S., Myerson, J., Smith, G.A., & Poon, L.W. (1988). Age, variability and speed: Between-subjects diversity. *Psychology and Aging*, *3*, 407–410.

Haran, F.J., Dretsch, M.N., Slaboda, J.C., Johnson, D.E., Adam, O.R., & Tsao, J.W. (2016). Comparison of baseline-referenced versus norm-referenced analytical approaches for in-theatre assessment of mild traumatic brain injury neurocognitive impairment. *Brain Injury*, *30*(3), 280–286. doi: 10.3109/02699052.2015.1118766

Hill, B.D., Rohling, M.L., Boettcher, A.C., & Meyers, J.E. (2013). Cognitive intra-individual variability has a positive asso ciation with traumatic brain injury severity and suboptimal effort. *Archives of Clinical Neuropsychology*, *28*(7), 640–648.

Hultsch, D.F., MacDonald, S.W.S., & Dixon, R.A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences*, *57B*(2), 101–115.

Hultsch, D.F., Strauss, E., Hunter, M.A., & MacDonald, S.W.S. (2008). Intraindividual variability, cognition and aging. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (3rd ed.), pp 491–556). New York: Psychology Press.

Kelly, M.P., Coldren, R.L., Parish, R.V., Dretsch, M.N., & Russell, M.L. (2012). Assessment of acute concussion in the combat environment. *Archives of Clinical Neuropsychology*, *27*(4), 375–388.

Lezak, M.D., Howieson, D.B., Bigler, E.D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.), New York, NY: Oxford University Press.

Lovden, M., Li, S.C., Shing, Y.L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, *45*, 2827–2838.

Makdissi, M., Collie, A., Maruff, P., Darby, D.G., Bush, A., McCrory, P., … Bennell, K. (2001). Computerised cognitive assessment of concussed Australian Rules footballers. *British Journal of Sports Medicine*, *35*(5), 354–360.

McCrea, M., Pliskin, N., Barth, J., Cox, D., Fink, J., French, L., … Powell, M. (2008). Official position of the military TBI task force on the role of neuropsychology and rehabilitation psychology in the evaluation, management, and research of military veterans with traumatic brain injury: APPROVED by: American Academy of Clinical Neuropsychology (AACN) American Psychological Association Division 40 (Neuropsychology) American Psychological Association Division 22 (Rehabilitation Psychology) National Academy of Neuropsychology (NAN). *The Clinical Neuropsychologist*, *22*(1), 10–26.

McCrory, P., Meeuwisse, W.H., Aubry, M., Cantu, B., Dvořák, J., Echemendia, R.J., … Sills, A. (2013). Consensus statement on concussion in sport: The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine*, *47*(5), 250–258.

Nelson, L.D., LaRoche, A.A., Pfaller, A.Y., Lerner, E.B., Hammeke, T.A., Randolph, C., … McCrea, M.A. (2016). Prospective, head-to-head study of three computerized neurocognitive assessment tools (CNTs): Reliability and validity for the assessment of sport-related concussion. *Journal of the International Neuropsychological Society*, *22*(1), 24–37.

Norris, J.N., Carr, W., Herzig, T., Labrie, D.W., & Sams, R. (2013). ANAM4 TBI reaction time-based tests have prognostic utility for acute concussion. *Military Medicine*, *178*(7), 767–774.

Parks, A.C., Moore, R.D., Wu, C.T., Broglio, S.P., Covassin, T., Hillman, C.H., & Pontifex, M.B. (2015). The association between a history of concussion and variability in behavioral and neuroelectric indices of cognition. *International Journal of Psychophysiology*, *98*(3), 426–434.

Rabinowitz, A.R., & Arnett, P.A. (2013). Intraindividual cognitive variability before and after sports-related concussion. *Neuropsychology*, *27*(4), 481–490.

Resch, J.E., McCrea, M.A., & Cullum, C.M. (2013). Computerized neurocognitive testing in the management of sport-related concussion: An update. *Neuropsychology Review*, *23*(4), 335–349.

Roebuck-Spencer, T.M., Vincent, A.S., Gilliland, K., Johnson, D.R., & Cooper, D.B. (2013). Initial clinical validation of an embedded performance validity measure within the automated neuropsychological metrics (ANAM). *Archives of Clinical Neuropsychology*, *28*(7), 700–710.

Segalowitz, S.J., Dywan, J., & Unsal, A. (1997). Attentional factors in response time variability after traumatic brain injury: An ERP study. *Journal of the International Neuropsychological Society*, *3*(2), 95–107.

Sosnoff, J.J., Broglio, S.P., Hillman, C.H., & Ferrara, M.S. (2007). Concussion does not impact intraindividual response time variability. *Neuropsychology*, *21*(6), 796–802.

Stuss, D.T., Stethem, L.L., Hugenholtz, H., Picton, T., Pivik, J., & Richard, M.T. (1989). Reaction time after head injury: Fatigue, divided and focused attention, and consistency of performance. *Journal of Neurology, Neurosurgery, & Psychiatry*, *52*(6), 742–748.