Exploring Neurocognitive Performance Differences in Military Aviation Personnel

Ashley Maltez-Laurienti; Adam Minniear; Rich Moore; Tanya McGovern; Paul Newman; Timothy Brearly

INTRODUCTION: Military aeromedical evaluations are common, but specialized neuropsychological norms for aviation personnel are sparse, resulting in a need to rely on norms from the general population. Little has been published regarding aviation subpopulations and how their neuropsychological profiles may differ from general population normative data. This study investigated neuropsychological test results of aeromedical service members to evaluate consistency with general population norms, and to delineate differences between aviation subpopulations.

METHODS: Analyses were conducted on demographic variables and test scores of military aviators ($N = 26$) and nonaviator crewmembers ($N = 36$) referred for evaluation due to a clinical problem requiring a waiver for flight status. Performance differences between subsamples were investigated with general linear modeling. Base rates for low scores were described.

RESULTS: Mean test scores in both subsamples were 0 to 1 standard deviation (SD) above the general population's means, with the largest discrepancies being found on measures of visuospatial ability (crewmembers) and verbal learning (aviators). Modeling revealed a significant difference between aviators and crewmembers on Trail Making Test Part B, after accounting for education. Aviators produced fewer low scores than crewmembers, even when using education adjusted normative data.

DISCUSSION: Results suggest the cognitive profile of aviators is uniquely strong in specific domains, with fewer low scores. The development of aviator-specific norms may enhance sensitivity to cognitive decrements in this population. Future studies might separately assess crewmember roles to further assess cognitive performance standards across specialties.

KEYWORDS: Military, neurocognitive performance, cognition, aeromedical, aviator.

Maltez-Laurienti A, Minniear A, Moore R, McGovern T, Newman P, Brearly T. ***Exploring neurocognitive performance differences in military aviation personnel.*** **Aerosp Med Hum Perform. 2021; 92(9):702–709.**

The medical status of military aviators and crewmembers is routinely assessed to ensure their medical fitness is maintained. Certain medical or neurobehavioral conditions have the potential to adversely affect cognitive functioning and so disqualify aviation personnel from full flight status. This necessitates an aeromedical evaluation if the service member wishes to obtain a waiver to pursue or regain flight status. In such cases, flight surgeons often refer the service member for a neuropsychological evaluation to help discern the appropriateness of a waiver. Mild traumatic brain injury (mTBI), psychiatric conditions, symptoms of attention-deficit hyperactivity disorder (ADHD), or treatment with psychiatric medications are examples of disqualifying clinical problems. Neuropsychologists must make recommendations in accordance with existing aeromedical policies that differ by branch of service and by specific job duties.[15] However, the aeromedical policies do not stipulate the cognitive profile characteristics that deem someone fit or unfit for a waiver. Psychiatric and neurological conditions that typically lead to a determination of unsuitability have been previously characterized,[18] but there has been little investigation into the neurocognitive characteristics of military aviation personnel. Aeromedically trained neuropsychologists

From Walter Reed National Military Medical Center, Bethesda, MD, USA.
This manuscript was received for review in September 2020. It was accepted for publication in June 2021.
Address correspondence to: Ashley Maltez-Laurienti, Bayne-Jones Army Medical Center, Department of Behavioral Health, 1585 Third St. Fort Polk, LA 71459; ashley.l.maltezlaurienti.mil@mail.mil.

DOI: https://doi.org/10.3357/AMHP.5788.2021

must interpret their examination results with an understanding of what makes aviation personnel neurocognitively unique before recommending for or against a waiver.

To assume that aviation personnel's neurocognitive characteristics are similar to the general population is to disregard both physiological and neuropsychological evidence to the contrary. Researchers have discovered specific neurophysiological processes associated with aviation tasks and have described patterns of cortical processing specific to career aviators.[1,15,29] Chen at al.[8] used fMRI to show aviators had greater functional connectivity within the default mode network (DMN) when compared to nonaviators, including greater functional integration of the middle occipital gyrus when the DMN is active, implicating a specific enhancement in not only arousal and attention but also spatial processing. This unique neurophysiological characteristic is congruent with prior neurocognitive research showing stronger spatial judgment and mental rotation abilities in aviators compared to nonaviators.[12] Interestingly, one study found that gender differences typically seen in the general population on mental rotation tasks were not present in an aviator population.[30] Given the evidence that aviators' brains are unique in how they are functionally specialized, it stands to reason that their cognitive abilities should be measured against a different standard than the general population.

The value of specific neuropsychological norms in select populations is established (e.g., evaluation of physicians),[9] but specialized norms for aviation personnel are sparse. The development and publication of the Cogscreen Aeromedical Edition (CogScreen-AE) was a major step forward in aviation neuropsychology, as it provided a collection of brief cognitive tests with aviator-specific normative data.[20] The ability to compare cognitive test scores to normal healthy aviators makes the CogScreen-AE a very useful instrument to include in aeromedical neuropsychological exams, but not all aeromedically trained neuropsychologists have made the financial investment in this test, and the neuropsychologist who uses the CogScreen-AE may need to administer additional tests depending on the clinical problem under investigation. Unfortunately, very little has been published about how healthy career aviators perform on other cognitive tests. Kay[21] presented normative data for commercial airline pilots using the Wechsler Adult Intelligence Scale-Revised (WAIS-R), the Expanded Halstead-Reitan Battery (EHRB), the Rey Complex Figure Test, the California Verbal Learning Test (CVLT; first edition), the Paced Auditory Serial Addition Test, and the Wisconsin Card Sorting Test. Not only did this provide a selection of aviator-specific norms, Kay's analysis of the data supported that aviators' cognitive profiles are different from the general population; he demonstrated that no more than 5% of the pilots obtained a WAIS-R IQ score at or below 100, which is the average score in the general population. However, the drawback of this set of aviator norms is that the WAIS-R and the CVLT are outdated, and many tests that make up the EHRB have dropped out of common use as the flexible battery approach gained popularity among practicing neuropsychologists. Our literature review found no other sets of normative data based on career aviators' performances on traditional neuropsychological tests. Without aviator-specific normative data, it can be difficult for an examiner to determine whether an aviator's mild cognitive decline constitutes an unacceptable risk to safely performing aviation duties. While the safety risk may be obvious for an aviator whose sustained attention scores (for example) are impaired relative to the general population, it cannot be assumed that any scores falling within the general population's normal range would necessarily predict adequate performance in the cockpit, where the demand for vigilance and attentional resources is especially high. In other words, what is considered normal cognitive performance for the general population might not be normal for aviators. Knowing that an aviator's scores fall either within or below what is considered normal for healthy aviators would give the examiner more confidence when recommending for or against a waiver.

Compared to civilian aviators, even less is known about the neurocognitive characteristics of military aviators. Kay[20] found that aviators flying for large commercial carriers, who were more likely to have prior military aviation experience, had stronger cognitive performances than did aviators flying for smaller carriers. Another study found that a small sample of 15 U.S. Army aviators demonstrated relative strengths in semantic knowledge, mental math, fine-motor speed, and concentration, although their performances were not substantially different from those of education-matched nonaviators.[16] A literature review by the authors found no other studies describing normal cognitive performances in military career aviators, but there were a limited number of studies describing results of intelligence testing in military aviators-in-training. A study using Armed Services Vocational Aptitude Battery (ASVAB) scores to predict the intelligence (IQ) scores of U.S. Army aviators reported higher mean estimated IQ scores relative to the general population.[24] Published normative data for U.S. Air Force (USAF) cadets and officers selected for aviator training found IQ scores of 1.33 to 1.5 standard deviations above the mean for adults in the general population.[7,23,27] A similar study conducted with USAF aviators-in-training yielded similar IQ estimates, and also found a variety of bidirectional performance differences between USAF aviator candidates and commercial aviators on neurocognitive tasks considered important to aviation duties.[6] Yet another study found similar IQ estimates in USAF cadets using the Multidimensional Aptitude Battery.[22] Other studies have found that military aviators-in-training perform better than nonaviators on visuospatial tasks.[14] The data on aviators-in-training support the idea that there are unique cognitive aptitudes in the military aviator population. However, the focus on cadets in many of these studies narrows the scope to include only younger samples, some of whom may not have completed training to become aviators. Additionally, generalizing findings from aviators-in-training to career aviators may impose some erroneous assumptions.

Beyond the different cognitive standards that exist between aviators and the general population, there is also a need to explore whether differences in baseline cognitive functioning exist between subpopulations in the aviation community.

Limiting subjects largely to USAF service members risks overlooking possible cognitive differences among aviation personnel in different service branches (Army, USAF, Navy, USMC, Coast Guard), similar to differences noted between commercial and military aviators.[6] The differences in training between service branches may strengthen different aspects of cognitive skills. Further, one service branch's selection criteria may differentially select for aviators with specific baseline cognitive characteristics. Beyond possible aviator differences between service branches, are there important differences between aviators who fly different aircraft? Perhaps there are differences in cognitive strengths between rotary and fixed-wing aviators rather than variations that are service branch-specific. If there are differences between various military service aviation communities, is this difference meaningful in the context of an aeromedical evaluation? For example, if a USAF aviator with a history of mild TBI performs below normal limits for aviators in her service branch on a test of divided attention but within normal limits relative to aviators grouped across all service branches, should she be denied a waiver to return to full flight status? The exclusion of nonaviator crewmembers from aviation neuropsychology research leaves a particularly limited understanding of expected performance among this subpopulation that comprises a substantial portion of aeromedical waiver referrals in the military. There is a need to further examine the neuropsychological profiles of military aviation personnel and to ultimately develop a robust set of norms for neuropsychological tests commonly used when assessing members of this unique population.

The current study sought to examine the cognitive performances of military aviation personnel referred for aeromedical neuropsychological evaluation in order to examine consistency with general population normative data and to delineate performance differences between subpopulations of aviation personnel. It was hypothesized that military aviators would perform above standard normative expectations on a range of cognitive functions including measures of simple and complex attention, processing speed, and visuospatial processing. It was also hypothesized that aviation personnel would yield a lower frequency of low scores compared to previously published base rates of low scores in the general population. Additionally, the current study sought to explore differences between aviator and crewmember cognitive performances as, to our knowledge, this has not been done before.

METHODS

Subjects

The Institutional Review Board at Walter Reed National Military Medical Center (WRNMMC) determined this study to be exempt research. This study used archival data from a convenience sample of military aviation personnel (N = 114) comprised of aviators and nonaviator crewmembers referred to the WRNMMC Neuropsychology Assessment Service for aeromedical evaluation from 1/1/2012 through 1/18/2019. Subjects were all pursuing a waiver due to a disqualifying clinical problem, which included a current or past disqualifying diagnosis (e.g., psychiatric, medical, mTBI, ADHD) and/or use of a disqualifying medication (e.g., an SSRI). Subjects known to have abnormal neuroimaging findings (N = 2), those who did not undergo a comprehensive neuropsychological test battery (N = 1), those whose specific job was not recorded in clinic archives (N = 39), and cadets not yet serving in an aviation capacity (N = 10) were excluded from the study. No subjects were excluded on the basis of waiver outcome; this information was not available for all subjects. All subjects performed within acceptable limits on performance validity tests. The final sample used for analysis was divided into two subsamples: aviators (N = 26) and crewmembers (N = 36). The aviator subsample included any aircraft or unmanned aerial vehicle pilots, and the crewmember subsample included any flight status personnel other than pilots (e.g., crew chief, gunner, flight surgeon, air traffic controller, loadmaster). Chi-squared and independent-sample t-test analyses revealed that the groups did not differ significantly in terms of age, gender, or race/ethnicity. Aviators had significantly higher education attainment than crewmembers. The majority of both groups sought waivers due to a psychiatric diagnosis or use of a disqualifying psychotropic medication. Demographic characteristics are reported in **Table I**.

Measures

The use of a retrospective convenience sample resulted in some differences in which neuropsychological tests were administered to each subject, but those selected for analyses have well-established psychometric properties and adequately cover the range of cognitive domains. They included five Wechsler Adult Intelligence Scale-IV (WAIS-IV) subtests (Coding, Digit Span, Arithmetic, Block Design, and Visual Puzzles), Logical Memory I&II from the Wechsler Memory Scale-IV (WMS-IV),

Table I. Demographic Characteristics of Aviators (***N*** = 28) and Crewmembers (***N*** = 36).

CHARACTERISTIC	AVIATORS (*N* = 26) M (SD) OR *N*	CREWMEMBERS (*N* = 36) M (SD) OR *N*	t/X^2	*P*
Age	39.15 (7.24)	32.31 (8.98)	0.507	0.479
Education (years)	16.46 (1.84)	14.47 (2.57)	6.701	**0.012**
Race/Ethnicity			5.112	0.164
Black	1	6		
Asian	0	2		
Hispanic	1	3		
White	24	25		
Gender			0.699	0.404
Male	23	29		
Female	3	7		
Condition			9.353	0.053
Psychiatric	16	20		
Medical	5	1		
ADHD	1	7		
mTBI	2	7		
Subjective complaints	1	1		
In-flight mishap	1	0		

three tests from the Expanded Halstead-Reitan Battery (EHRB; Trail Making Test Part B, Category Fluency-Animals, and Letter Fluency-FAS), and the California Verbal Learning Test-II (CVLT-II). Many subjects also completed Trail Making Test Part A, but it was not included in this study because it did not meet the homogeneity of slopes assumption for ANCOVA. A description of the psychometric properties and the general population normative data can be found in the respective test manuals. Test score data used in this study included raw scores as well as standardized scores. According to the general population norms published in their respective test manuals, the WAIS-IV,[31] WMS-IV,[32] and EHRB[17] standardized scores were adjusted for age, education, gender, and race. CVLT-II[10] standardized scores were adjusted for age and gender.

Analyses

All statistical analyses were run using IBM SPSS Statistics, version 24.[19] Selected measures met assumptions for Analysis of Covariance (ANCOVA). Descriptive statistics within each subsample were obtained for raw scores and standardized scores. One-way ANCOVA analyses were used to examine performance raw score differences (means and SDs) between the aviators and crewmembers on each test, allowing for the control of differences in educational attainment between the two groups. Raw scores were selected for this purpose to help eliminate the influence of education level and age on the scores. Alpha levels of ≤ 0.05 were used to define statistical significance. Effect sizes of group differences were measured using partial eta squared.

The base rates of low scores across the 12 scores used in the above ANCOVA analysis were calculated under two score transformation conditions, but first subjects were excluded from the calculation if their data included fewer than 10 of the 12 scores. The first condition used the test publishers' demographically adjusted norms to transform subject's raw scores. For the WAIS-IV and WMS-IV subtests and the EHRB tests, transformations included adjustments for age, education, gender, and race.[26,17] CVLT-II score transformations included adjustments for age and gender only.[10] In the secosnd condition, aviators' raw scores were transformed to *t*-scores using the aviator subsample means and standard deviations, and crewmembers' raw scores were transformed using the crewmember subsample means and standard deviations obtained in this study. No demographic adjustments were applied in the second condition. Base rate calculations for each condition were carried out 3 times according to 3 different ways of defining a low score – first defined as ≥ 1 SD below the mean, then ≥ 1.5 SD below the mean, and then ≥ 2 SD below the mean.

RESULTS

Table II summarizes the cognitive test score means and SDs for aviators and crewmembers and ANCOVA results. In both subsamples, all mean test scores fell within approximately 0 to 1 SDs above the general population means. The greatest discrepancies from the normative mean were on tests of verbal learning (aviators) and visuospatial processing (crewmembers). In examining the raw score differences between aviators and crewmembers, ANCOVA results showed that aviators performed significantly better than crewmembers after controlling for education attainment on Trail Making Test Part B, $F(1,47) = 4.74$, $P \leq 0.04$, with a medium effect size ($\eta^2 = 0.09$).[11] There were no other significant test score differences between aviators and crewmembers.

Table III presents the base rates of low scores. Only the 12 scores used in the above ANCOVA analysis were considered for base rate calculations. Subjects with fewer than 10 of these scores were excluded for the calculations. For the aviator subsample, 11 (47.8%) had all 12 scores, 6 (26.1%) had 11, and 6 (26.1%) had 10. For the crewmember subsample, 16 (55.2%) had all 12 scores, 6 (20.7%) had 11, and 7 (24.1%) had 10. The following findings are for the condition using the test publishers' demographically adjusted normative data to transform subjects' scores: when defining a low score as ≥ 1 SD below the mean, 39.1% of aviators had 1 or more low score, and 13.0% had 2 or more, while 51.7%

Table II. Test Performance: Means for Aviators and Crewmembers, Standard Deviations, and Analysis of Covariance Results for Cognitive Tests.

	AVIATORS		CREWMEMBERS					
TEST	RAW M (SD)	STANDARD M (SD)	RAW M (SD)	STANDARD M (SD)	F	df	P	η_p^2
Animal Naming	26.96 (4.85)	56.56 (9.27)	25.40 (5.55)	54.57 (11.31)	1.57	1,57	0.215	0.027
COWA (FAS)	52.04 (10.97)	55.12 (8.93)	45.34 (12.13)	50.43 (10.21)	1.86	1,57	0.178	0.032
CVLT-II LDFR	13.17 (2.13)	-	12.65 (2.81)	-	0.62	1,46	0.617	0.013
CVLT-II Total Learning	56.91 (8.89)	58.83 (9.71)	54.65 (9.89)	54.35 (10.07)	0.31	1,46	0.583	0.007
Trail Making Test B	44.95 (8.64)	56.63 (8.08)	53.65 (17.66)	53.35 (9.33)	4.74	1,47	**0.035**	0.092
WAIS-IV Arithmetic	16.42 (2.55)	48.62 (9.06)	15.56 (3.32)	51.14 (11.22)	1.26	1,57	0.267	0.022
WAIS-IV Block Design	53.92 (6.76)	54.76 (6.48)	54.68 (7.56)	57.25 (9.25)	0.10	1,53	0.920	0.000
WAIS-IV Coding	81.62 (6.45)	56.00 (6.18)	78.86 (13.73)	55.69 (8.64)	0.01	1,58	0.905	0.000
WAIS-IV Digit Span	33.31 (4.27)	54.50 (9.44)	30.14 (4.35)	51.22 (8.09)	3.85	1,58	0.055	0.062
WAIS-IV Visual Puzzles	19.43 (3.37)	56.71 (6.84)	17.61 (4.61)	54.67 (11.19)	1.89	1,34	0.178	0.053
WMS-IV Logical Memory I	30.28 (6.07)	53.92 (8.86)	27.70 (6.06)	52.68 (9.18)	1.59	1,55	0.213	0.028
WMS-IV Logical Memory II	26.84 (7.12)	54.44 (9.45)	24.39 (6.80)	52.50 (9.03)	0.87	1,55	0.36	0.015

NOTE: COWA = Controlled Oral Word Association Test; CVLT-II = California Verbal Learning Test, Second Edition; LDFR = Long Delay Free Recall; WAIS-IV = Wechsler Adult Intelligence Scale-IV; WMS-IV = Wechsler Memory Scale-IV. Statistical tests performed using raw scores; standard scores (*t*-scores) are provided for reference only. LDFR standard scores not included as continuous data is not reported by CVLT-II scoring software.

of crewmembers had 1 or more low score, and 34.5% had 2 or more. When defining a low score as ≥ 1.5 SD below the mean, 17.4% of aviators had 1 or more, and 4.3% had 2 or more, while 31.0% of crewmembers had 1 or more, and 10.3% had 2 or more. When defining a low score as ≥ 2 SD below the mean, none of the aviators had any low scores, while 17.2% of crewmembers had 1 or more, and 3.4% had 2 or more. On average, 5.2%, 2.0%, and 0.0% of an aviator's scores were low when using the ≥ 1 SD, ≥ 1.5 SD, and ≥ 2 SD criteria, respectively, and the corresponding figures for an average crewmember were 9.2%, 3.8%, and 2.0%.

The following findings are for the condition using the sub-samples' respective means and standard deviations to transform subjects' scores: when defining a low score as ≥ 1 SD below the mean, 87.0% of aviators had 1 or more low score, and 47.8% had 2 or more, while 69.0% of crewmembers had 1 or more low score, and 51.7% had 2 or more. When defining a low score as ≥ 1.5 SD below the mean, 52.2% of aviators had 1 or more, and 17.4% had 2 or more, while 37.9% of crewmembers had 1 or more, and 10.3% had 2 or more. When defining a low score as ≥ 2 SD below the mean, 26.1% of aviators had 1 or more, and 8.7% had 2 or more, while 17.2% of crewmembers had 1 or more, and 6.9% had 2 or more. On average, 15.1%, 7.3%, and 3.3% of the aviators' scores were low when using the ≥ 1 SD, ≥ 1.5 SD, and ≥ 2 SD criteria, respectively, and the corresponding figures for crewmembers were 16.5%, 5.1%, and 2.3%.

DISCUSSION

While some authors have argued against the need for aviator-specific norms on the basis that education-based norms sufficiently account for performance differences between aviators and non-aviators,[16] the current study supports the need for such norms and is the first to argue for a separate set of norms specifically for crewmembers. Crewmembers, to our knowledge, have not been included in previous studies despite comprising a significant number of aeromedical evaluation referrals.

Despite the fact that subjects in our study were all drawn from a clinical convenience sample, we found that these aviation personnel outperformed the general population on multiple cognitive measures, and aviators in particular yielded proportionally fewer low scores. On average, military aviators' performances on a measure of verbal learning curve approached 1 SD above age-and-gender-adjusted general population means. They also scored over 0.5 SD above full demographically adjusted means on tests of verbal fluency, complex speeded attention, processing speed, and mental rotation/visuospatial reasoning. While we expected relatively stronger performances among military aviators in many of these cognitive domains, the magnitudes of the differences were not as large as might be expected given previous findings of IQ scores approaching 1.33 SD above general population means among USAF aviation cadets and trainees.[6,7,27] One possible reason for this discrepancy is the fact that our sample was selected from a clinical population, which may have diluted the mean scores. Still, this highlights the point that measures of IQ alone are insufficient in aeromedical evaluations where examinees are expected to have strengths in specific cognitive domains.

Crewmember mean performance was comparable to the general population on most cognitive tasks, with performance on a processing speed and a visuospatial reasoning task falling over 0.5 SD above the general population mean. While their

Table III. Low Score Base Rates: Prevalence of Aviators and Crewmembers with Abnormal Scores on Aeromedical Evaluation, and Percent Low Scores Per Case.

SUBSAMPLE	SCORE TRANSFORMATION SOURCE	LOW SCORE DEFINITION	NUMBER OF LOW SCORES: 1 OR MORE	NUMBER OF LOW SCORES: 2 OR MORE	PERCENT LOW SCORES
Aviators (*N* = 23; Mean total scores = 11.2; range = 10-12)	Test publishers' demographically adjusted norms	≥ 1.0 SD Below Mean	39.1%	13.0%	Mean = 5.24 SD = 7.84
		≥ 1.5 SD Below Mean	17.4%	4.3%	Mean = 2.02 SD = 4.96
		≥ 2.0 SD Below Mean	0.0%	0.0%	-
	Aviator subsample Means and SD	≥ 1.0 SD Below Mean	87.0%	47.8%	Mean = 15.14 SD = 10.59
		≥ 1.5 SD Below Mean	52.2%	17.4%	Mean = 7.26 SD = 9.06
		≥ 2.0 SD Below Mean	26.1%	8.7%	Mean = 3.25 SD = 6.13
Crewmembers (*N* = 29; Mean total scores = 11.3; range = 10-12)	Test publishers' demographically adjusted norms	≥ 1.0 SD Below Mean	51.7%	34.5%	Mean = 9.15 SD = 13.96
		≥ 1.5 SD Below Mean	31.0%	10.3%	Mean = 3.84 SD = 7.77
		≥ 2.0 SD Below Mean	17.2%	3.4%	Mean = 2.04 SD = 5.33
	Crewmember subsample Means and SD	≥ 1.0 SD Below Mean	69.0%	51.7%	Mean = 16.47 SD = 16.76
		≥ 1.5 SD Below Mean	37.9%	10.3%	Mean = 5.05 SD = 8.85
		≥ 2.0 SD Below Mean	17.2%	6.9%	Mean = 2.30 SD = 5.85

performances were more consistent with the general population than were the aviators', this provides some initial evidence that crewmembers may also have specific cognitive strengths as a group, which warrants further study. Moreover, in examining the significant differences between aviators and crewmembers, aviators performed significantly better only on a task of complex speeded attention, although they also tended to produce fewer low scores than the crewmembers, which again highlights the need for aviator and crewmember normative data and interpretation guidelines. Further, these data suggest that in an aeromedical context, it may be prudent to give more weight to subtle weaknesses in test scores (e.g., a score that is in the low end of the normal range in an otherwise solidly normal or better cognitive profile) by administering additional tests within the same domain to confirm the repeatability and magnitude of the potential deficit.

While a test's population-specific normative data can help a clinician determine whether individual test scores are so low as to raise suspicion of a clinical problem, it is common for neurocognitively healthy individuals to have one or more low scores. Therefore, considering the number of low scores someone has relative to his specific population is relevant to clinical decision making and should be considered in the aeromedical context. The base rates of low scores among neurocognitively healthy people in the general population have been reviewed[3,4,28] and Kay[20] addressed the need for such analyses in the aviation community. In this study, we used test publishers' demographically adjusted norms to show that base rates for low scores in our aviator sample were lower than in the general population. When defining low as ≥ 1 SD below the mean, 39% of aviators undergoing aeromedical evaluation had 1 or more low scores on a battery yielding 10 to 12 scores, and 13% had 2 or more. Compare this to findings within the general population showing 36% had 2 or more scores ≥ 1 SD below the mean on a comprehensive battery of 10 tests.[28] For crewmembers, the base rates of low scores looked comparable to the general population. Again using ≥ 1 SD below the mean to define a low score, 52% of crewmembers had 1 or more low score on a battery yielding 10 to 12 scores, and 35% had 2 or more.

According to the review by Binder et al.[2] of several base rate studies, the expected number of low scores (when defined as greater than one SD below the mean) can be roughly estimated at 10–15% of the scores in a test battery. This basically holds true for crewmembers in our aeromedical sample, where on average 9% of an individual's scores were low. For aviators, however, only 5% of an individual's scores were low on average. When using the test publishers' demographically adjusted normative data, our preliminary findings suggest the clinician should expect fewer low scores for aviators than the general population, but crewmember base rates are roughly commensurate with the general population. However, since the base rate figures presented here were derived from a clinical convenience sample and likely overestimate the rates of low scores in healthy aviators and crewmembers, they should not be relied upon for clinical decision-making. Future studies are needed to establish reliable base rates for healthy aviation personnel.

Ideally, clinicians who examine aviation personnel should use population-specific norms, and over time it is hoped that such normative data will be collected and made available for a wide range of common neuropsychological tests. Should this be the case one day, it would be helpful to examine the base rates of low scores when using aviator-specific norms. While this study's aeromedical data set should not be considered normative data, we used the raw score means and standard deviations to transform subjects' raw scores into *t*-scores in order to describe the base rates of low scores when measuring aviator and crewmember cognitive test performances against their respective peers. When defining low as ≥ 1 SD below the mean, 87% of aviators undergoing aeromedical evaluation had 1 or more low score on a battery yielding 10 to 12 scores, and 48% had 2 or more; 69% of crewmembers had 1 or more low score, and 52% had 2 or more. Looking at Table III, it is clearly common for the majority (87%) of aviators undergoing aeromedical evaluation to have 1 or more scores ≥ 1 SD below the population-specific mean, and about half (52%) had 1 or more score ≥ 1.5 SD below the mean. However, it is much less common for aviators to have 2 or more low scores, especially when using ≥ 1.5 SD below the mean to define low. Our preliminary findings suggest that a cut-off of 2 low scores may be useful for discriminating between normal and abnormal profiles when using aviator-specific norms, as a cut-off of 1 low score may increase the likelihood of a false-positive error. Additional studies, ideally incorporating testing normed in aviators (e.g., CogScreen-AE), are needed to establish whether this cut-off is useful for clinical decision-making.

Our study has a few notable limitations that must be taken into consideration when interpreting the findings. First, our small sample size raises the possibility that the moderate effect sizes of between-group differences in speeded complex attention may be over-estimated. Secondly, because our sample consisted of aviation personnel undergoing aeromedical evaluation it is reasonable to assume that some subjects' clinical problems (e.g., mTBI, ADHD, sleep apnea) impacted their test scores and potentially confounded our data. However, the group differences in condition were not significant, and attempts were made to minimize the overall effect of clinical problems on the cognitive test scores by excluding subjects with known brain lesions (e.g., subarachnoid hemorrhage and ischemia). A related limitation is that some of our subjects demonstrated cognitive weaknesses that were judged by the examining clinical neuropsychologists to be severe enough to recommend against returning to full flight status. Given these limitations, our results should be interpreted with caution.

Unfortunately, not all aeromedical cases in the data archive documented the clinician's recommendation, but based on the subjects whose recommendations were documented the majority of our subjects had been judged cognitively fit for full flight status. Additionally, we examined a limited number of tests due to subjects receiving differing test batteries in our clinic. Thus, there may exist differences between groups on other cognitive measures that were not examined. Another limitation of this exploratory study is that our crewmember sample included a mix of every occupation requiring flight status other than aviators, as the sample was not large enough to further delineate occupational subtypes such as aircraft maintainers, ground support personnel, and medical support personnel. There is a possibility that members of specific occupations have specific cognitive skills that would be better delineated if subgroups were examined independently. Pertinent to a concern that was previously mentioned, this study includes predominantly U.S. Army personnel, and is not representative of the aviation community across all services.

Future normative studies should strive to include a large sampling of healthy aviation personnel with a representative sample from each service branch and major job duty (e.g., mechanic, loadmaster, crew chief, etc.) as well as different types of aviators (e.g., rotary versus fixed-wing). In addition to publishing raw score means and standard deviations, we recommend that normative studies also present data on the frequencies of low scores.

Because there tend to be variations in what the different service branches consider appropriate for a given assessment, the community of aeromedical neuropsychologists should strive to develop a robust set of norms for all of the most widely used cognitive tests (e.g., continuous performance tests, WAIS-IV, WMS-IV, CVLT-II, Trail Making Test, verbal fluency tests, etc.) to ensure aeromedical neuropsychologists have a variety of tests at their disposal to assess a range of possible conditions along with the ability to detect subtle decrements in cognitive function that may adversely impact flight performance. Our findings lend some support to the recommended core aeromedical battery proposed by Graver et al.[15] in that the Trail Making Test Part B subtest showed significant differences between our two groups with at least a medium effect size. Thus, this measure should be included in batteries used in future studies. Also, while our battery included many of the tests used as part of the Federal Aviation Administration's core battery,[13] future studies should strive to include all of these to help determine if there may be utility in using a similar core battery within the military services.

Future research should further clarify cognitive abilities and differences between various occupations within the crewmember population. Of particular interest would be examination of differences between crewmember occupations such as crew chiefs, door gunners, aircraft maintainers, ground support personnel, loadmasters, flight medics, flight physicians, navigators, etc. These jobs all require flight status (with some variation between branches of service) and each serve very different roles within the aviation environment which may require different sets of cognitive skills.[15] Our preliminary findings suggest there may be specific cognitive strengths in this group overall, yet some subgroups (e.g., flight physicians and navigators) may have more advanced education and correspondingly higher neuropsychological test performances than other subgroups (e.g., enlisted crew chiefs, aircraft maintainers, etc.).

ACKNOWLEDGMENTS

The identification of specific products or scientific instrumentation is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the author, DoD, or any component agency. The views expressed in this manuscript are those of the authors and do not necessarily reflect the official policy of the Department of Defense or the U.S. Government.

Financial Disclosure Statement: The authors have no conflicts to disclose.

Authors and Affiliations: Ashley Maltez-Laurienti, Psy.D., Adam Minniear, Psy.D., Rich Moore, Psy.D., Tanya McGovern, Psy.D., Paul Newman, Ashley Maltez-Laurienti, Psy.D., Timothy Brearly, Psy.D., Walter Reed National Military Medical Center, Bethesda, MD, USA; and Ashley Maltez-Laurienti, Psy.D., Bayne-Jones Army Medical Center, Department of Behavioral Health, Fort Polk, LA, USA; and Timothy Brearly, Salisbury Veteran Affairs Health Care System, Salisbury, NC, USA.

REFERENCES

1. Adamson MM, Taylor JL, Heraldez D, Khorasani A, Noda A, et al. Higher landing accuracy in expert pilots is associated with lower activity in the caudate nucleus. PLoS One. 2014; 9(11):e112607.
2. Binder LM, Iverson GL, Brooks BL. To err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. Arch Clin Neuropsychol. 2009; 24(1):31–46.
3. Brooks BL, Holdnack JA, Iverson GL. Advanced clinical interpretation of the WAIS-IV and WMS-IV: prevalence of low scores varies by level of intelligence and years of education. Assessment. 2011; 18(2):156–167.
4. 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.
5. Callan DE, Terzibas C, Cassel DB, Callan A, Kawato M, Sato MA. Differential activation of brain regions involved with error-feedback and imitation based motor simulation when observing self and an expert's actions in pilots and non-pilots on a complex glider landing task. Neuroimage. 2013; 72:55–68.
6. Callister JD, King RE, Retzlaff PD. Cognitive assessment of USAF pilot training candidates: Multidimensional Aptitude Battery and CogScreen-Aeromedical Edition. Technical Reports NTIS#199614. Brooks Air Force Base (TX): Armstrong Laboratory; 1995. 22p. [Accessed July 19, 2021.] Available from: https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/ADA303388.xhtml.
7. Carretta TR, King RE, Ree MJ, Teachout MS, Barto E. Compilation of cognitive and personality norms for military aviators. Aerosp Med Hum Perform. 2016; 87(9):764–771.
8. Chen X, Xu K, Yang Y, Wang Q, Jiang H, et al. Altered default mode network dynamics in civil aviation pilots. Front Neurosci. 2020; 13:1406.
9. Del Bene VA, Brandt J. Identifying neuropsychologically impaired physicians. Clin Neuropsychol. 2020; 34(2):318–331.
10. Delis DC, Kramer JH, Kaplan E, Ober BA. California Verbal Learning Test– Second Edition– Adult Version: manual. San Antonio (TX): Psychological Corporation; 2000.
11. Draper S. Effect Size [Internet]. Glasgow: University of Glasgow; 2011. [Accessed 2020 Apr 08]. Available from http://www.psy.gla.ac.uk/~steve/best/effect.html.
12. Dror IE, Kosslyn SK, Waag WL. Visual-spatial abilities of pilots. J Appl Psychol. 1993; 78(5):763–773.
13. Federal Aviation Administration. Specifications for neuropsychological evaluations for potential neurocognitive impairment [Internet]. 2018. [Accessed 2021 Feb 19]. Available from: https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/ame/guide/media/npevalspecs.pdf.
14. Gordon HW, Leighty R. Importance of specialized cognitive function in the selection of military pilots. J Appl Psychol. 1988; 71(1):38–45.
15. Graver CJ, Armistead-Jehle P, Fritch AM. Neuropsychologist's guide to aeromedical examinations in the military. Mil Behav Health. 2020. [Accessed 2020 Apr 08] Available from: https://www.tandfonline.com/doi/full/10.1080/21635781.2020.1742823.

16. Guilmette TJ, Treanor JJ. Baseline and comparative neuropsychological data on U.S. Army aviators. Aviat Space Environ Med. 1986; 57(10, Pt 1): 950–953.
17. Heaton RK, Miller SW, Taylor MJ, Grant I. Revised comprehensive norms for an expanded Halstead-Reitan Battery: demographically adjusted neuropsychological norms for African American and Caucasian adults: professional manual. Lutz (FL): PAR; 2004.
18. Heaton JE, Heaton JN, Merriweather TV, McDonald KD. Analysis of the U.S. Air Force flying class waiver outcomes of the aeromedical consultation service and major command units in neuropsychiatry. 2011. Report of the Air Force Research Laboratory, Case Number: 88ABW-2011-6400.
19. Corp IBM. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk (NY): IBM Corp.
20. Kay GG. CogScreen Aeromedical Edition professional manual. Odessa (FL): Psychological Assessment Resources; 1995.
21. Kay GG. Guidelines for the psychological evaluation of air crew personnel. Occup Med. 2002; 17(2):227–245, iv.
22. King RE, Barto E, Ree MJ, Teachout MS, Retzlaff P. Compilation of pilot cognitive ability norms. Technical Report, USAF School of Aviation Medicine, Wright Patterson AFB, OH.
23. King RE, Caretta TR, Retzlaff P, Barto E, Ree MJ, Teachout MS. Standard cognitive psychological tests predict military pilot training outcomes. Aviat Psychol Appl Hum Factors. 2013; 3(1):28–38.
24. Kratz K, Poppen B, Burroughs L. The estimated full-scale intellectual abilities of U.S. Army aviators. Aviat Space Environ Med. 2007; 78(5, Suppl): B261–B267.
25. Oltra-Cucarella J, Sánchez-SanSegundo M, Rubio-Aparicio M, Arango-Lasprilla JC. Ferrer- Cascales R. The association between the number of neuropsychological measures and the base rate of low scores. Assessment. 2021; 28(3):955–963.
26. Pearson. Advanced Clinical Solutions for WAIS-IV and WMS-IV: clinical and interpretive manual. San Antonio (TX): NCS Pearson; 2009.
27. Retzlaff PD, Gibertini M. Air Force Pilot Personality: Hard Data on the "Right Stuff". Multivariate Behav Res. 1987; 22(4):383–399.
28. 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.
29. Toppi J, Borghini G, Petti M, He EJ, De Giusti V, et al. Investigating Cooperative Behavior in Ecological Settings: An EEG Hyperscanning Study. PLoS One. 2016; 11(4):e0154236.
30. Verde P, Piccardi L, Bianchini F, Trivelloni P, Guariglia C, Tomao E. Gender effects on mental rotation in pilots vs. nonpilots. Aviat Space Environ Med. 2013; 84(7):726–729.
31. Wechsler D. Wechsler Adult Intelligence Scale – Fourth Edition: technical and interpretive manual. San Antonio (TX): NCS Pearson; 2008.
32. Wechsler D. Wechsler Memory Scale – Fourth Edition: technical and interpretive manual. San Antonio (TX): NCS Pearson; 2009.