Age and Test Setting Affect the Prevalence of Invalid Baseline Scores on Neurocognitive Tests

Jonathan D. Lichtenstein,*† PsyD, MBA, Rosemarie Scolaro Moser,† PhD, and Philip Schatz,‡ PhD
Investigation performed at the Sports Concussion Center of New Jersey, Lawrenceville, New Jersey

Background: Prevalence rates of invalid baseline scores on computerized neurocognitive assessments for high school, collegiate, and professional athletes have been published in the literature. At present, there is limited research on the prevalence of invalid baseline scores in pre–high school athletes.

Hypothesis: Pre–high school athletes assessed with baseline neurocognitive tests would show higher prevalence rates of invalidity than older youth athletes, and those athletes, regardless of age, who were tested in a large group setting would show a higher prevalence rate of invalidity than athletes tested in a small group setting.

Study Design: Cross-sectional study; Level of evidence, 3.

Methods: A total of 502 athletes between the ages of 10 and 18 years completed preseason baseline neurocognitive tests in “large” or “small” groups. All athletes completed the online version of ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing). Baseline test results that were “flagged” by the computer software as being of suspect validity and labeled with a “1” symbol were identified for analysis. Participants were retrospectively assigned to 2 independent groups: large group or small group. Test administration of the large group occurred off-site in groups of approximately 10 athletes, and test administration of the small group took place at a private-practice neuropsychology center with only 1 to 3 athletes present.

Results: Chi-square analyses identified a significantly greater proportion of participants obtaining invalid baseline results on the basis of age; younger athletes produced significantly more invalid baseline scores (7.0%, 17/244) than older athletes (2.7%, 7/258) ($\chi^2 (1) = 4.99; P = .021$). Log-linear analysis revealed a significant age (10-12 years, 13-18 years) × size (small, large) interaction effect ($\chi^2 (4) = 66.1; P < .001$) on the prevalence of invalidity, whereby younger athletes tested in larger groups were significantly more likely to provide invalid results (11.9%) than younger athletes tested in smaller groups (5.4%), older athletes tested in larger groups (2.7%), and older athletes tested in smaller groups (2.7%).

Conclusion: Younger athletes tend to exhibit a greater prevalence of invalid baseline results on neurocognitive computerized tests than older youth athletes; the prevalence increases when testing is conducted in a large group and nonclinical setting.

Keywords: concussion testing; baseline testing; validity; neurocognitive testing; group administration; pediatric concussion; ImPACT

A sports-related concussion, especially in youth, has been identified as a serious public health concern, as the incidence of concussions in the pediatric population has increased at an alarming rate over the past 15 years. Clinical data suggest that younger athletes require a longer period of recovery time than older athletes. Further, within the recent proliferation of research into the effects of concussions in youth, preliminary studies are demonstrating significant alterations in brain mechanisms such as cerebral blood flow and white matter density.

Neuropsychological testing has emerged as the “cornerstone” in the evaluation of concussions, as it provides an objective method of evaluating cognitive changes related to concussions. Given a true representation of an athlete’s baseline cognitive skills before a head injury, comparison to postinjury testing can help determine when that athlete’s brain has recovered. The baseline and postinjury assessment model of concussion management, promoted and administered by neuropsychologists, began at the college level in the 1980s and is considered an important component of concussion management. However, consensus experts have recently suggested that there is insufficient evidence to recommend the widespread use of baseline neurocognitive testing. Although the recent literature has held that postconcussion test
data can be validly compared with normative data (eg, rather than baseline data), these findings are preliminary and have not been replicated, and these procedures have not been adopted by any professional organizations.

Where traditional paper and pencil–based neuropsychological tests were once used in the assessment of athletes, now computerized neurocognitive testing is employed to save time and resources. Widespread baseline “preinjury” testing programs have been implemented by a variety of sports medicine professionals at the high school, collegiate, and professional levels of competition. However, with greater attention on the effects of concussions in youth, there is now an increased interest in implementing such baseline testing programs at the pre–high school level and in nonscholastic (recreational and club) sports. In fact, some states in the United States are beginning to propose legislation to mandate concussion management programs for nonscholastic sports.

For baseline tests to be useful for comparison with postinjury performance, the data must reflect a valid assessment of an athlete’s “true” baseline ability. Unfortunately, administrators of baseline testing programs are often inadequately trained in neurocognitive testing and may not be cognizant of the factors that may render a test invalid. For example, athletic trainers often administer cognitive tests in schools but commonly fail to examine baseline test results for suspect validity. There are myriad factors that can negatively influence test performance and yield suboptimal results, ranging from fatigue and anxiety to poor effort and motivation. Additionally, an athlete’s performance on testing may be affected by environmental factors, such as group or individual settings, confusion about test directions, noises and distractions, mechanical issues with the computer device, or other individual factors. It has also been suggested that the lack of a serious, controlled, and standardized approach to baseline testing (such as that afforded to standardized testing in schools) may be a large factor in the prevalence of suspect validity of test results. Regardless, baseline test results should always be examined for validity.

The scientific literature on ImPACT (Immediate Post-Concussion Assessment and Cognitive Testing) has shown varying prevalence rates of suspect validity on baseline testing. Results have ranged from 4.1% to 11% in collegiate athletes and from 4.5% to 6.3% to 7.9% in high school athletes; researchers using traditional paper and pencil–based baseline measures also found 11% of high school athletes to have suspect validity. Currently, there are no published rates of suspect validity on ImPACT in athletes younger than 13 years old. Thus, the purpose of this article was to further examine the prevalence of suspect, invalid baseline test results in a neuropsychology center where test administration was performed in a controlled manner both on-site and off-site, with particular attention to possible differences in (1) younger versus older youth athletes and (2) small versus large group test settings.

**MATERIALS AND METHODS**

**Participants**

Participants were 502 youth athletes, aged 10 to 18 years, involved primarily in nonscholastic sports who completed baseline neurocognitive testing using the online version of ImPACT under the administration and supervision of a neuropsychology center during the time span of 2010 to early 2013. The sample was primarily male (85% male, 15% female participants), with ice hockey (59%) and lacrosse (28%) being the most common sports represented. The prevalence of an attention deficit disorder/learning disorder (ADD/LD) was 7.6% and was normally distributed across age groups. Thirteen percent of athletes reported 1 previous concussion, and 4.9% reported ≥2 previous concussions. As expected, older youth athletes reported more concussions. Most of the athletes belonged to athletic organizations that recommended baseline testing for their teams, while others were self-referred. Group demographic data, by age group, are presented in Table 1.

All participants who had completed baseline testing since the inception of the center’s implementation of the online version of ImPACT were included. Thus, the study was retrospective in nature, with no opportunity for randomization or sampling. Baseline testing was performed during daytime hours, during the school or work week, as well as on weekends. Participants were retrospectively assigned to 2 independent groups: large group or small group.

- The large group consisted of 207 athletes (mean age, 13.9 ± 2.26 years) who completed baseline neurocognitive testing off-site at an athletic facility (not at the neuropsychology center). For this large group, test administration was conducted with approximately 10 participants at a time in a room, where at least 1 doctoral-level neuropsychologist and 1 master’s-level neuropsychology technician were present throughout testing. Test administrators oriented and seated athletes at their testing stations, discussed the importance of effort and minimizing distractions, described the test, and assisted in the preliminary modules of ImPACT.
(demographics, symptom checklist). At least 2 test administrators were present in the room to observe, supervise, and help maintain a controlled test environment.

- The small group consisted of 295 athletes (mean age, 12.1 ± 1.72 years) who completed baseline neurocognitive testing at the neuropsychology center. For the small group, test administration was conducted with 1 to 3 participants at a time in a room. At least 1 doctoral-level neuropsychologist and 2 master’s-level neuropsychology technicians were on-site to orient, seat, speak to, and instruct athletes in the same manner as in the large group, but only interfacing with 1 to 3 students at a time (eg, 1:1, 1:2, or 1:3 staff:athlete ratio). Once athletes were set up for testing, test administrators were present outside the testing room(s) checking on athletes or sitting in the testing room to provide assistance if needed.

Materials

All participants completed the online version of ImPACT, which consists of 6 neurocognitive subtests, each designed to target distinct aspects of cognitive functioning, including attention, memory, visual-motor processing speed, and reaction time. The results generate 5 composite scores and a speed:accuracy ratio; comprehensive descriptions are available in the literature.20,23,35 The online version of ImPACT identifies baseline scores of suspect validity based on an algorithm of certain composite and subscale score indicators. When such criteria are met, the baseline results are “flagged” by the computer software and labeled with a “+” symbol. In accordance with the ImPACT’s clinical interpretation manual,22 invalid baseline scores were defined by the following criteria: Impulse Control Composite score >30, Word Memory Learning percentage <69%, Design Memory Learning percentage <50%, X’s and O’s Total Correct Interference score >30, and Three Letters Total Letters Correct score <8.

Procedures

Institutional review board approval was obtained for retrospective analysis of de-identified data. General assent and/or informed consent for the de-identified use of test results in research were obtained from athletes (or parents of minor athletes) upon admission to the center. There was no common script of instructions used to orient participants to the test situation. There was, however, a basic protocol that was utilized in all baseline tests administered including the following: education about concussions; a discussion on the use and utility of baseline testing; and an emphasis on reading, following all subtest directions and prompts, and maintaining good effort and a quiet test environment. Testing was always conducted under close administrator supervision (see “Participants” above). In some cases (notably 10- and 11-year-old children), instructions were read to participants to ensure an understanding of the test directions and reduce error due to confusion or variable reading skills.

As part of the baseline testing procedure, although not a part of the present study analysis, all athletes with suspect baseline test results were identified, and those athletes were requested to retake the test. Athletes and parents in both groups were provided with printed educational materials before the test session. Some parents and athletes in both groups were also provided with oral presentations at the sites as well as an audio recording via e-mail. In addition, parents and athletes who completed testing in the neuropsychology center setting were afforded a greater opportunity to speak individually with the neuropsychology staff.

Analyses

The prevalence of invalid baseline results was documented for each age group and each group administration type. Chi-square analyses were conducted to identify the likelihood of invalid results by age (10-12 years, 13-18 years) and administration type (small, large) as well as sex, history of concussion (0, 1, ≥2), and history of ADD/LD. Log-linear analysis (also referred to as multiway frequency analysis43) was conducted to identify interaction effects among age (10-12 years, 13-18 years), administration type (small, large), and validity results (valid, invalid). All analyses were conducted using SPSS version 19 (SPSS Inc, Armonk, New York).43

RESULTS

Chi-square analyses identified a significant difference in the proportion of participants obtaining invalid baseline results on the basis of age and history of ADD/LD; a greater proportion of younger athletes provided invalid baseline
scores (7.0%) than did older athletes (2.7%) ($\chi^2 (1) = 4.99; P = .021$), and a greater proportion of athletes with a history of ADD/LD provided invalid baseline scores (13.2%) than did those without a history of ADD/LD (4.1%). However, $\chi^2$ analyses identified no significant difference in the proportion of participants obtaining invalid baseline scores on the basis of group size (5.3% large vs 4.4% small; $P = .64$), sex (5.2% male vs 2.6% female; $P = .34$), or history of concussion (4.5% >1 previous concussions vs 4.9% no previous concussions; $P = .88$). With respect to sex, $\chi^2$ analysis identified a significantly lower proportion of male athletes in the younger age group of 10 to 12 years (80.7%) than the older age group of 13 to 18 years (88.8%) ($\chi^2 (1) = 6.2; P = .02$).

Log-linear analysis revealed a significant age (10-12 years, 13-18 years) $\times$ size (small, large) interaction effect ($\chi^2 (4) = 66.1; P < .001$) on the prevalence of invalid results, whereby younger athletes tested in larger groups were significantly more likely to provide invalid results (11.9%) than younger athletes tested in smaller groups (5.4%), older athletes tested in larger groups (2.7%), and older athletes tested in smaller groups (2.7%) (Table 2). Of note, including only athletes with no history of ADD/LD, those with a history of concussion, and only male athletes did not alter these findings ($\chi^2 (4) = 52.2; P < .001$).

### DISCUSSION

The results of the present study document that a greater proportion of younger youth athletes, 10 to 12 years of age, obtained invalid results on baseline testing than did older youth athletes, aged 13 to 18 years. Until now, the statistical prevalence of such invalid results had not been available for this younger population. With the push to provide concussion management programs and baseline testing for all youth athletes, the present findings have implications for greater care and vigilance in the test administration and review of baseline test results for younger youth athletes.

Indeed, high school athletic programs often have limited resources required to execute baseline testing programs, with many high schools lacking athletic trainers for oversight. This makes the review of baseline test validity and readministration a daunting task, especially when students are tested in large groups of $\geq$20 per room. Interestingly, in the present study, there was no significant difference in the prevalence of suspect valid baseline scores for adolescents (13-18 years) whether they were tested in a large group setting (10 per room) at an athletic facility or in a small group setting (1-3 per room) at a neuropsychology center. For older youth athletes (13-18 years), whether the testing is performed off-site or in a neuropsychology center may not be as important as the standardization and oversight of the administration by trained neurocognitive professionals. In fact, in this study, the prevalence of invalid results for both small and large group test settings for the 13- to 18-year age group was less than what has been previously reported in the literature. Thus, a lower frequency of invalidity was observed for high school students when there was consistent accountability, instruction, supervision, and observation in the test administration.

However, even with significant care in the test administration for the 10- to 12-year-old children, the small group test setting resulted in a prevalence rate that was, at best, similar to that reported in the literature, whereas the large group test setting rendered a prevalence rate that was higher than that documented in the literature. It is important to note that while the prevalence of invalid baseline scores has been systematically documented in high school and collegiate samples, the utility or validity of the indicators contributing to the identification of invalid test results is not yet fully understood. In this regard, it is not clear if “invalid” test results are the product of intentional malingering, poor test-taking strategies, interference/distraction, equipment failure, history of ADD/LD, cumulative effects of previous concussions, or other factors negatively affecting performance. Further, while the prevalence rates for invalid results (approximately 5% for the total sample: 7% for 10-12 years and 3% for 13-18 years) are generally low, there is no a priori level of invalidity that is considered “acceptable” or thought to be “chance.” Given that the normal curve would allow for 2.5% of athletes to fall in the tails of $\pm$2 standard deviations or 5% beyond a 1-sided tail, one might expect an invalid rate of 5% to be within normal limits.

In the current study, the large group test setting appeared to introduce distractions as well as decreased individualized attention for the younger athletes. Younger, pre–high school athletes are less developmentally mature and independent than adolescents. They may require more attention and supervision when completing tasks and may not intellectually grasp the importance and utility of baseline testing because of their less mature cognitive development. Additionally, their reading skills may be less advanced. Not surprisingly, all these factors likely contribute to a greater prevalence of suspect baseline test results in the younger group. Finally, while there was no greater proportion of athletes obtaining invalid baseline scores on the basis of sex, there was a higher proportion of invalid baseline results obtained by younger athletes. That there was a significantly higher proportion of male athletes in the older age group may suggest some contributing role of sex.

The current study demonstrated that even under strict supervision in a neuropsychology center, younger athletes still delivered more invalid baseline scores than did their
adolescent counterparts. Therefore, it seems that even when there is a strong attempt to control for external environmental factors, younger athletes should be expected to have a higher rate of suspect valid results on ImPACT than adolescent athletes. These results bring attention to the need for strict administration procedures with younger athletes and, more importantly, the need to design baseline tools that are tailored to younger athletes.

As concussive awareness, litigation, and legislation are on the rise,10,21,34,42 there will be increased pressure upon academic institutions and youth sports organizations to provide concussion management programs that include baseline neurocognitive testing. Elementary schools and recreational leagues will face the same financial challenges as high school, collegiate, and professional programs when it comes to providing concussion management services, but with even fewer economic resources.

This study is not without its limitations. First, the study is retrospective in nature and thus lacks blinding, randomization, and comparison with a control group. While administrators followed a similar protocol, there was no control of test administrator assignment, a script of test instructions, or designated seating arrangements. Further, test performance was not controlled for specific variables that may influence cognitive performance, such as intelligence quotient, reading level, or number of previous concussions. As such, the generalizability of these findings is decreased. Second, the sex distribution was not equal, as the current sample contained more boys than girls, which may have influenced the study results.

There are additional elements, not encompassed in the current study, that will require further research. This sample was composed of mainly ice hockey and lacrosse players, so future studies may wish to examine athletes from a wider array of sports. Also, future studies will need to examine the effect of specific environmental and test administration factors to better understand which elements play significant roles in test performance. Such elements may include distracting sounds in the test setting, the interaction and rapport developed with a test administrator, the test administrator’s presence in the room, seating arrangements, and the number of test takers per room.

Overall, these findings support the need for standardized, better controlled test administration procedures by trained professionals who understand neurocognitive testing. These findings also provide a warning to those who may contemplate implementing computerized baseline testing programs for younger athletes. Such programs will require even greater vigilance, caution, individualized attention, and resources for administration.

REFERENCES


For reprints and permission queries, please visit SAGE's Web site at http://www.sagepub.com/journalsPermissions.nav