

Henry Ford Health System

Henry Ford Health System Scholarly Commons

Behavioral Health Articles

Behavioral Health Services / Psychiatry

6-1-2018

Prevalence of Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator

Christopher A. Abeare

Isabelle Messa

Brandon G. Zuccato

Bradley Merker

Henry Ford Health System, bmerker1@hfhs.org

Laszlo Erdodi

Follow this and additional works at: https://scholarlycommons.henryford.com/behavioralhealth_articles

Recommended Citation

Abeare CA, Messa I, Zuccato BG, Merker B, Erdodi L. Prevalence of Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator. *JAMA Neurol* 2018; 75(6):697-703.

This Article is brought to you for free and open access by the Behavioral Health Services / Psychiatry at Henry Ford Health System Scholarly Commons. It has been accepted for inclusion in Behavioral Health Articles by an authorized administrator of Henry Ford Health System Scholarly Commons.

Prevalence of Invalid Performance on Baseline Testing for Sport-Related Concussion by Age and Validity Indicator

Christopher A. Abeare, PhD; Isabelle Messa, MSc; Brandon G. Zuccato, BA; Bradley Merker, PhD; Laszlo Erdodi, PhD

IMPORTANCE Estimated base rates of invalid performance on baseline testing (base rates of failure) for the management of sport-related concussion range from 6.1% to 40.0%, depending on the validity indicator used. The instability of this key measure represents a challenge in the clinical interpretation of test results that could undermine the utility of baseline testing.

OBJECTIVES To determine the prevalence of invalid performance on baseline testing and to assess whether the prevalence varies as a function of age and validity indicator.

DESIGN, SETTING, AND PARTICIPANTS This retrospective, cross-sectional study included data collected between January 1, 2012, and December 31, 2016, from a clinical referral center in the Midwestern United States. Participants included 7897 consecutively tested, equivalently proportioned male and female athletes aged 10 to 21 years, who completed baseline neurocognitive testing for the purpose of concussion management.

INTERVENTIONS Baseline assessment was conducted with the Immediate Postconcussion Assessment and Cognitive Testing (ImPACT), a computerized neurocognitive test designed for assessment of concussion.

MAIN OUTCOMES AND MEASURES Base rates of failure on published ImPACT validity indicators were compared within and across age groups. Hypotheses were developed after data collection but prior to analyses.

RESULTS Of the 7897 study participants, 4086 (51.7%) were male, mean (SD) age was 14.71 (1.78) years, 7820 (99.0%) were primarily English speaking, and the mean (SD) educational level was 8.79 (1.68) years. The base rate of failure ranged from 6.4% to 47.6% across individual indicators. Most of the sample (55.7%) failed at least 1 of 4 validity indicators. The base rate of failure varied considerably across age groups (117 of 140 [83.6%] for those aged 10 years to 14 of 48 [29.2%] for those aged 21 years), representing a risk ratio of 2.86 (95% CI, 2.60-3.16; $P < .001$).

CONCLUSIONS AND RELEVANCE The results for base rate of failure were surprisingly high overall and varied widely depending on the specific validity indicator and the age of the examinee. The strong age association, with 3 of 4 participants aged 10 to 12 years failing validity indicators, suggests that the clinical interpretation and utility of baseline testing in this age group is questionable. These findings underscore the need for close scrutiny of performance validity indicators on baseline testing across age groups.

JAMA Neurol. 2018;75(6):697-703. doi:10.1001/jamaneurol.2018.0031
Published online March 12, 2018.

Author Affiliations: Department of Psychology, University of Windsor, Windsor, Ontario, Canada (Abeare, Messa, Zuccato, Erdodi); Department of Behavioral Health, Henry Ford Health System, Detroit, Michigan (Merker).

Corresponding Author: Christopher A. Abeare, PhD, Department of Psychology, University of Windsor, 401 Sunset Ave, Windsor, ON N9B 3P4, Canada (cabeare@uwindsor.ca).

In the United States, it is estimated that between 1.1 million and 1.9 million sport- and recreation-related concussions occur annually in children aged 18 years or younger.¹ This statistic, when combined with concerns about microstructural damage in the white matter tracts² of the brain and the potential long-term effects of multiple concussions,³ is leading to a focus on the management of concussion as a public health concern. The heterogeneity in clinical presentation and natural history,⁴ with the absence of a reliable biomarker or other medical test to identify concussion,⁵ forces examiners to diagnose concussion based on the clinical assessment of medical history, balance testing, neurocognitive functioning, and postconcussion symptoms.

In an attempt to improve the diagnostic accuracy of concussion, baseline neurocognitive testing was introduced in the 1980s to assess preinjury neurocognitive functioning against which postinjury functioning could be compared.⁶ In contrast to the traditional methods of neuropsychological assessment in which the premorbid level of functioning is estimated and the test results are compared with normative data, baseline testing was implemented to reduce the inherent error in estimating premorbid functioning. Theoretically, this practice increases the validity of decision making because individuals serve as their own healthy controls.⁷ However, the utility of baseline testing compared with the use of normative data in postinjury evaluations has long been debated.^{8,9} The most recent Consensus Statement on Concussion in Sport⁴ concluded that baseline testing can be useful in the management of concussion, but there is not enough evidence to suggest that baseline testing should be mandatory.

During the past decade, computerized neurocognitive testing has largely replaced traditional paper-and-pencil methods and has become widespread at all levels of sport¹⁰ because of the ease of administration and scoring, which allows for the baseline testing of groups and reduces barriers for use by a wider range of health care professionals.¹¹ This expansion of the user base supports the need for monitoring the validity of assessment practices. Computerized baseline testing has the potential to contribute to this goal. However, computer-administered evaluations also introduce new concerns about the validity of the data they produce.

Performance validity is defined as the extent to which test-taking behavior provides an accurate reflection of the underlying cognitive ability that the instrument was designed to measure. Performance validity tests are objective measures designed to identify response sets that are unlikely to accurately reflect the true ability of the test taker. Valid performance is a basic assumption of neuropsychological testing, and it is necessary to arrive at valid and useful clinical decision making.¹² Noncredible responding in neurologically intact populations can be attributed to inattentiveness, poor task comprehension, lack of incentive to perform well, or incentive to perform poorly. The influence of these factors varies as a function of age, with children and adolescents more likely to have invalid performance associated with inattentiveness, poor task comprehension, and a lack of appreciation for the importance of performing their best,^{13,14} whereas adults have been shown to also be influenced by external incentives.¹⁵⁻¹⁸

Key Points

Questions What is the prevalence of invalid neurocognitive performance at baseline in the management of sport-related concussion and does it vary by age and validity indicator?

Findings In this cross-sectional study of 7897 participants who completed baseline neurocognitive testing for the management of sport-related concussion, 56% failed at least 1 of 4 published validity indicators. Base rates of failure varied considerably across age groups from 84% in those aged 10 years to 29% in those aged 21 years.

Meaning Base rates of failure were surprisingly high overall, suggesting a need for a critical examination of performance validity assessment practices on baseline testing in concussion management programs.

The base rate of failure (BRF), also known as the base rate of invalid performance, varies as a function of metric, population, and situational variables. Although there is limited research on the BRF during baseline testing, a systematic review by Gaudet and Weyandt¹⁹ focused specifically on the Immediate Postconcussion and Cognitive Testing (ImPACT),²⁰ the most widely used computerized neurocognitive test for the management of concussion. An estimated 75% of National Collegiate Athletic Association member schools use the ImPACT for baseline assessments,²¹ with BRF ranging from 2.7%²² to 27.9%²³ (weighted mean, 6.1%).¹⁹ These numbers are likely underestimates, as they are based on embedded validity indicators (EVIs) that are less sensitive. These BRFs tend to be lower than those found in neurologically intact young adults who participate in academic research (BRF range, 18.3%-36.7%).²⁴

Although children and adolescents typically pass performance validity tests designed for and normed on adults, elevated BRF in young examinees are commonly reported in the pediatric population and are typically attributed to the stage of cognitive development.¹³ The BRF for children in clinical settings is highly variable, ranging from 0% to 70% (weighted mean, 15.5%).¹³ Within studies, young children tend to have higher BRF,^{25,26} prompting some researchers to suggest that examinees younger than 10 years should be exempt from performance validity testing.²⁷ Other studies found a more complex association between age and BRF that was mediated by the combined choice of performance validity tests and the cut-off used.^{25,26,28} Associations with age have also been found in athletes, with younger athletes (aged 10-12 years) having higher BRF on baseline testing than older athletes (aged 13-18 years).¹⁴

Four EVIs have been published for the ImPACT; 2 are included in the ImPACT Clinical Manual²⁹ and the others were developed by independent research teams. In addition to the default ImPACT EVI that automatically flags invalid performance, the ImPACT Clinical Manual²⁹ also provides a second EVI, "Red Flags," as a more liberal index of suboptimal performance. Two alternative EVIs have been introduced by Schatz and Glatts¹⁶ and Higgins et al³⁰ based on experimental malingered paradigms. These new EVIs produce higher BRFs than in the ImPACT EVIs (Table 1).

Table 1. BRF for Athletes by Age and EVI

EVI and Component	Cutoff	Age, y											
		10 (n = 140)	11 (n = 142)	12 (n = 197)	13 (n = 574)	14 (n = 3503)	15 (n = 1368)	16 (n = 816)	17 (n = 508)	18 (n = 388)	19 (n = 131)	20 (n = 82)	21 (n = 48)
Default ImPACT													
Xs and Os, total incorrect, No. ^a	>30	2.9	2.1	3.0	0.7	0.9	0.7	0.9	1.0	0.3	0.8	1.2	0.0
Impulse control composite, No. ^b	>30	3.6	2.1	3.0	0.9	1.1	0.7	0.9	1.0	0.3	0.8	1.2	0.0
Word memory learning, % correct	<69	1.4	0.0	0.0	0.2	0.3	0.1	0.4	0.8	0.0	0.0	0.0	0.0
Design memory learning, % correct	<50	0.7	3.5	1.0	2.4	2.2	2.9	3.1	2.6	2.1	3.1	0.0	0.0
3 Letters, total letters correct, No.	<8	13.6	6.3	4.1	4.2	3.5	3.4	3.1	1.8	2.1	3.1	2.4	2.1
ImPACT Red Flags													
Reaction time composite, s ^c	>0.80	29.3	23.2	7.6	6.6	4.5	3.1	2.3	3.1	2.8	1.5	1.2	0.0
Verbal memory, No. ^d	<70	29.3	14.1	17.3	14.8	12.9	13.7	11.6	9.3	6.4	7.6	9.8	8.3
Visual memory, No. ^e	<60	32.9	34.5	17.8	19.5	20.3	24.4	21.7	17.5	14.4	16.8	17.1	12.5
Processing speed	<25	52.1	32.4	14.7	7.5	6.1	5.1	2.9	3.7	3.6	2.3	0.0	2.1
Schatz and Glatts,¹⁶ 2013													
Reaction time, s	>0.80	29.3	23.2	7.6	6.6	4.5	3.1	2.3	3.1	2.8	1.5	1.2	0.0
Visual motor speed, No. ^f	<25	52.1	32.4	14.7	7.5	6.1	5.1	2.9	3.7	3.6	2.3	0.0	2.1
Word memory CD, No. ^g	<22	25.0	21.8	17.8	17.2	20.0	21.5	20.8	19.5	15.2	24.4	14.6	12.5
Design memory CD, No. ^h	<16	36.4	32.4	33.5	35.0	35.3	41.5	38.6	38.6	30.9	29.0	26.8	16.7
Higgins et al,³⁰ 2016													
Word memory learning, % ^{i,j}	≤0.90	10.7	10.6	7.6	6.1	6.5	8.0	8.5	7.9	4.1	9.9	4.9	6.3
Word memory delayed, % correct ^k	≤81	26.4	15.5	24.9	15.7	19.2	21.6	21.0	22.6	13.7	20.6	19.5	12.5
Memory, %^l													
Design memory	≤64.75	22.9	20.4	19.3	21.8	22.3	25.7	25.6	21.7	18.8	18.3	19.5	16.7
XO ^j	≤107.5	97.1	90.8	73.1	48.3	39.4	33.1	28.7	23.8	23.2	20.6	23.2	22.9
LRE cutoff	≤0.23	62.9	62.0	46.2	35.0	35.0	36.3	33.2	29.5	21.6	28.2	20.7	20.8

Abbreviations: BRF, base rate of failure; CD, correct distractors; EVI, embedded validity indicator; ImPACT, Immediate Postconcussion and Cognitive Testing; LRE, logistic regression equation.

^a Xs and Os is a memory task that measures visual working memory and visual processing/visual motor speed.

^b The impulse control composite is a measure of the total number of errors made on the interference phase of X's and O's as well as errors made on the color match task.

^c Reaction time composite score is comprised of the average of the following: (1) average correct reaction time of interference stage of Xs and Os, (2) symbol match average correct reaction time visible/3, and (3) color match average correct reaction time.

^d The verbal memory composite score represents the average performance on (1) word memory total percent correct, (2) symbol match total correct hidden/9 multiplied by 100, and (3) 3 letters, percent total letters correct.

^e The visual memory composite score is comprised of the average of (1) design memory, total percent correct, and (2) Xs and Os (total correct memory)/12 multiplied by 100.

^f Visual motor speed composite score is comprised of the average of (1) total number correct/4 during interference of Xs and Os, and (2) average counted correctly multiplied by 3 from countdown phase of 3 Letters.

^g Number of correctly identified distractor items.

^h Number of correctly identified incorrect distractor items.

ⁱ Hits + correct distractors/24 multiplied by 100.

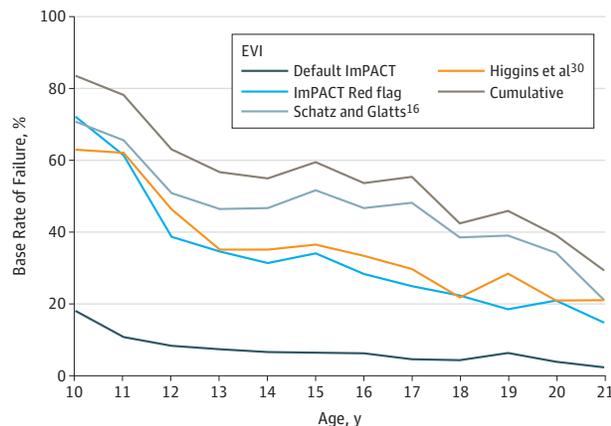
^j These indicators were not included in the calculation of the overall BRF; they were included in the LRE cutoff. Indicators are displayed to assist with interpretation of the LRE results.

^k Delay hits + correct distractor delay items.

The combination of the experimental and observational evidence suggests that true BRFs could be as high as 40% in high school and collegiate athletes and even higher in younger athletes. Group-based computerized testing may be associated with a higher BRF compared with 1-on-1 evaluations because of the lack of close monitoring of test-taking behavior;³¹ however, other factors, such as adherence to standardized administration and the quality of supervision, may be better pre-

dictors of BRF than group size.³² The lack of consensus regarding the most appropriate EVI for determining the validity of ImPACT scores, in combination with wide discrepancies in BRFs across samples, indicators, and research designs, necessitates a direct comparison of the available EVIs in a naturalistic setting. The present study was designed to compare the 4 existing ImPACT-based EVIs across age groups in a large sample of athletes undergoing baseline testing.

Figure. Base Rate of Failure for 7897 Athletes by Age and Embedded Validity Indicator (EVI)



ImPACT indicates Immediate Postconcussion and Cognitive Testing.

The default ImPACT EVI was expected to produce a BRF comparable to the 6.1% reported by Gaudet and Weyandt,¹⁹ and the ImPACT Red Flags EVI was expected to have a BRF between 20% and 30%, as reported in previous studies.^{23,33} Based on previous reports, we predicted that the BRF would be the highest for the EVIs introduced by Schatz and Glatts¹⁶ and Higgins et al.³⁰ However, the paucity of research on these 2 EVIs in naturalistic settings precluded a more specific prediction. We hypothesized that the BRF on all 4 EVIs would vary across age group, producing a higher BRF in younger athletes partially because of the reliance on raw scores uncorrected for developmental changes in the underlying cognitive ability. We also predicted that BRFs would be elevated as athletes became more acculturated to playing at the more competitive levels of sport, such as the collegiate level, when the incentive to underperform would be greatest.

Methods

Participants

Participants included 7897 consecutively evaluated athletes aged 10 to 21 years who completed baseline neurocognitive testing for the management of concussion. Most of the participants were English speaking (7820 [99.0%]) and right-handed (6848 [86.7%]). The most commonly reported preexisting diagnoses were attention-deficit/hyperactivity disorder (869 participants [11.0%]), dyslexia (145 [1.8%]), and autism (27 [0.3%]). Most athletes played football (1652 [20.9%]), followed by soccer (1252 [15.9%]), volleyball (770 [9.8%]), basketball (733 [9.3%]), hockey (726 [9.2%]), and field hockey (680 [8.6%]).

Procedure

Data were collected during baseline testing for concussion between January 1, 2012, and December 31, 2016, through the concussion management program of a large hospital in the Midwestern United States. The online version of the ImPACT,³⁴ version 2.1 (ImPACT Applications Inc), was administered to all

participants. Testing lasted approximately 45 minutes and was conducted in groups of approximately 20 athletes at schools, at community centers, or in a hospital-based setting. Examinees were overseen by either an athletic trainer or a licensed clinical neuropsychologist. This study was approved by the institutional research ethics boards of Henry Ford Health System and the University of Windsor, which also waived the need for participant informed consent because of the retrospective use of deidentified data. Ethical guidelines regulating research with human participants were followed throughout the study.

Measure

The ImPACT is a computerized neurocognitive test designed for baseline and postconcussion assessment. The test includes 5 performance-based cognitive indices: reaction time, visual motor speed, impulse control, and visual and verbal memory as well as an inventory of postconcussion symptoms. One-year test-retest reliabilities are variable and range from low (intraclass correlation coefficient = 0.22) to high (0.85).^{35,36} Convergent validity has been established against traditional neuropsychological measures.³⁷ The ImPACT demonstrated high sensitivity (0.82) and specificity (0.89) for concussion, with an overall classification accuracy of 86%.³⁸

Statistical Analysis

Data were analyzed with IBM SPSS, version 24.0 (IBM Corporation). No data were missing for any variables. The BRFs were calculated for the entire sample and separately for each age group (10 through 21 years), for each of the 4 EVIs and cumulative BRF, representing the percentage of athletes who failed at least 1 of the 4 EVIs. Risk ratio (RR) was calculated between the 10-year-old group and the 21-year-old group to summarize the association of age. Mean ImPACT composite scores were calculated for each age group.

Results

Of the 7897 study participants, 4086 (51.7%) were male, the mean (SD) age was 14.71 (1.78) years, 7820 (99.0%) were primarily English speaking, and the mean (SD) educational level was 8.79 (1.68) years. Across the sample of 7897 participants, the BRF was 6.4% (505 participants) for the default ImPACT EVI, 31.8% (2509) for the ImPACT Red Flags, 34.9% (2759) for the Higgins et al³⁰ logistic regression equation, and 47.6% (3757) for the Schatz and Glatts¹⁶ EVI. The cumulative BRF was 55.7% (4400). Examination of the BRF by age demonstrated a strong age association (Figure). The age with the highest cumulative BRF was the 10-year-old group at 83.6% (117 of 140 participants), whereas the 21-year-old group had the lowest at 29.2% (14 of 48 participants) (RR, 2.86; 95% CI, 2.60-3.16; $P < .001$) (Table 1). Table 2 provides ImPACT mean raw composite scores and percentiles by age.

Discussion

Baseline cognitive testing is a common practice for concussion management programs and was introduced to enhance

Table 2. Mean Raw ImpACT Composite Scores for Athletes by Age

Measure	Sex	Mean Raw (Percentile) Score by Age, y																							
		10	11	12	13	14	15	16	17	18	19	20	21	10	11	12	13	14	15	16	17	18	19	20	21
Verbal memory	Male	75.12 (31.4)	79.33 (40.8)	80.19 (47.2)	80.99 (48.9)	80.64 (47.3)	80.59 (47.2)	81.88 (46.4)	82.41 (46.9)	85.68 (55.8)	83.77 (50.2)	85.00 (55.1)	84.12 (50.4)	81.88 (46.4)	82.41 (46.9)	85.68 (55.8)	83.77 (50.2)	85.00 (55.1)	84.12 (50.4)	85.68 (55.8)	83.77 (50.2)	85.00 (55.1)	84.12 (50.4)		
	Female	79.42 (44.7)	79.73 (42.3)	80.53 (44.5)	84.44 (47.0)	83.71 (44.8)	83.09 (43.2)	84.81 (48.4)	84.96 (49.1)	86.56 (54.9)	86.78 (59.0)	85.32 (54.1)	83.73 (50.1)	84.81 (48.4)	84.96 (49.1)	86.56 (54.9)	86.78 (59.0)	85.32 (54.1)	83.73 (50.1)	86.56 (54.9)	86.78 (59.0)	85.32 (54.1)	83.73 (50.1)		
Visual memory	Male	65.67 (52.6)	66.86 (65.4)	71.85 (66.6)	71.70 (66.8)	71.04 (65.4)	70.00 (65.0)	71.40 (42.9)	72.38 (43.8)	75.82 (53.0)	75.92 (71.7)	77.24 (74.4)	75.24 (72.0)	71.04 (65.4)	70.00 (65.0)	71.40 (42.9)	72.38 (43.8)	75.82 (53.0)	75.92 (71.7)	77.24 (74.4)	75.24 (72.0)	75.82 (53.0)	75.92 (71.7)	77.24 (74.4)	75.24 (72.0)
	Female	65.21 (64.4)	64.73 (66.3)	69.85 (67.1)	72.92 (75.1)	71.75 (73.9)	70.23 (72.3)	71.47 (77.9)	72.89 (77.4)	73.65 (79.9)	72.8 (84.0)	75.27 (80.5)	78.73 (78.1)	71.75 (73.9)	70.23 (72.3)	71.47 (77.9)	72.89 (77.4)	73.65 (79.9)	72.8 (84.0)	75.27 (80.5)	78.73 (78.1)	72.8 (84.0)	75.27 (80.5)	78.73 (78.1)	
Visual motor speed	Male	23.70 (16.6)	26.33 (28.6)	30.69 (47.2)	32.12 (33.3)	33.20 (38.3)	34.56 (43.6)	35.96 (36.6)	37.46 (42.9)	39.40 (50.8)	38.82 (43.5)	39.90 (47.0)	40.79 (51.6)	33.20 (38.3)	34.56 (43.6)	35.96 (36.6)	37.46 (42.9)	39.40 (50.8)	38.82 (43.5)	39.90 (47.0)	40.79 (51.6)	39.40 (50.8)	38.82 (43.5)	39.90 (47.0)	40.79 (51.6)
	Female	26.18 (22.7)	27.72 (30.7)	29.52 (43.3)	35.39 (38.5)	35.74 (39.6)	37.10 (46.5)	38.84 (44.9)	38.91 (45.4)	39.16 (47.5)	41.07 (56.5)	41.05 (56.2)	42.34 (61.7)	29.52 (43.3)	35.39 (38.5)	35.74 (39.6)	37.10 (46.5)	38.84 (44.9)	38.91 (45.4)	39.16 (47.5)	41.07 (56.5)	41.05 (56.2)	42.34 (61.7)		
Reaction time	Male	0.78 (27.7)	0.74 (38.0)	0.67 (55.9)	0.66 (34.4)	0.64 (41.8)	0.63 (45.8)	0.62 (41.2)	0.61 (43.6)	0.60 (46.7)	0.61 (36.5)	0.59 (41.3)	0.62 (31.4)	0.67 (55.9)	0.66 (34.4)	0.64 (41.8)	0.63 (45.8)	0.62 (41.2)	0.61 (43.6)	0.60 (46.7)	0.61 (36.5)	0.59 (41.3)	0.62 (31.4)		
	Female	0.76 (33.5)	0.73 (37.4)	0.688 (53.4)	0.64 (37.7)	0.62 (40.4)	0.61 (44.5)	0.6 (37.8)	0.60 (39.8)	0.60 (37.7)	0.60 (40.5)	0.59 (44.1)	0.58 (44.8)	0.688 (53.4)	0.64 (37.7)	0.62 (40.4)	0.61 (44.5)	0.6 (37.8)	0.60 (39.8)	0.60 (37.7)	0.60 (40.5)	0.59 (44.1)	0.58 (44.8)		
Impulse control	Male	11.78 (32.1)	10.23 (43.2)	10.31 (54.2)	8.26 (45.8)	7.59 (48.2)	6.96 (50.4)	6.86 (41.8)	6.35 (44.3)	5.77 (51.6)	5.73 50.5	5.38 (54.5)	4.91 (51.3)	10.31 (54.2)	8.26 (45.8)	7.59 (48.2)	6.96 (50.4)	6.86 (41.8)	6.35 (44.3)	5.77 (51.6)	5.73 50.5	5.38 (54.5)	4.91 (51.3)		
	Female	8.31 (41.3)	8.66 (44.2)	10.34 (52.1)	6.66 (49.6)	6.61 (49.7)	6.50 (51.6)	5.92 (52.2)	6.40 (52.9)	5.53 (55.0)	4.98 60.0	5.35 (58.7)	5.4 (58.7)	6.66 (49.6)	6.61 (49.7)	6.50 (51.6)	5.92 (52.2)	6.40 (52.9)	5.53 (55.0)	4.98 60.0	5.35 (58.7)	5.4 (58.7)			
No. of athletes	Male	73	69	118	286	1643	719	486	289	234	91	45	33	118	286	1643	719	486	289	234	91	45	33		
	Female	67	73	79	288	1860	649	330	219	154	40	37	15	79	288	1860	649	330	219	154	40	37	15		

Abbreviation: ImpACT, Immediate Postconcussion and Cognitive Testing.

the clinical utility of postinjury ImpACT data for making return-to-play decisions for athletes. However, given the reports of high and fluctuating BRFs across validity indicators and cut-offs, the validity of baseline data has become a source of concern. To empirically evaluate the legitimacy of this concern and to characterize it across development, we compared the BRF on the 4 previously published EVIs for the ImpACT. As expected for the EVIs, the BRF on the default ImpACT EVI was the lowest at 6.4%, comparable with the 6.1% rate found by Gaudet and Weyandt.¹⁹ The ImpACT Red Flags identified 31.8% of the sample as invalid, and the Higgins et al³⁰ logistic regression equation identified 34.9%. The EVI with the highest BRF was developed by Schatz and Glatts¹⁶ at 47.6%. The cumulative BRF (individuals with ≥1 EVI failure) was 55.7%.

Consistent with our prediction, the youngest athletes (aged 10-11 years) had the highest BRF across all 4 EVIs. Age and invalid performance had a strong negative association; older athletes were less likely to fail EVIs. However, contrary to our prediction, the collegiate age groups did not show an increase in BRF, with the isolated exception of the 19-year-old group. The BRF continued to decrease across ages 20 and 21.

Although the literature on performance validity at baseline focuses largely on deliberate suppression of performance (ie, “sandbagging”) in high school and collegiate athletes, invalid performance is multifactorial and includes age-related fluctuations in attention and comprehension characteristic to the stage of cognitive development¹³ as well as varying ability to appreciate the importance of demonstrating one’s true ability level during cognitive testing. The very young athletes are most likely to be negatively affected by such confounding variables, which could explain the high BRF. It must be noted that these factors are likely to be independent of the conscious downward manipulation of test scores.^{27,28} Nevertheless, these confounding factors place examinees at higher cumulative risk of failing EVIs, particularly given that EVIs are distributed throughout the test battery as opposed to free-standing performance validity tests, which typically assess validity at discrete points in time.

The BRFs varied considerably across different EVIs. The default ImpACT EVI and ImpACT Red Flags rely on some of the same subtest scores but use different cutoffs, with default ImpACT EVI based on 2 SDs and ImpACT Red Flags based on 1.5 SDs below the mean. On the other hand, the 2 newest EVIs were calibrated empirically using experimental malingering paradigms in high school athletes and nonathlete college students. Although research suggests that performance validity tests designed for adults can be applied to children,¹³ the downward extension of adult cutoffs may violate important trajectories in cognitive development and increase the likelihood of false-positive errors in younger age groups.

Experimental malingering studies are at risk for overfitting their detection model to an atypical (ie, artificially exaggerated) manifestation of invalid responding. Namely, they were calibrated to detect the most egregious forms of invalid responding in examinees who had little incentive to avoid detection, limiting its generalizability to real-life situations with potentially substantial (perceived) reward for successful malingering.²⁴ This suggests that (1) these EVIs may not apply

to athletes below the high school level and (2) the current BRF in the collegiate athletes, and possibly the high school athletes, may be underestimated, as these EVIs may not detect more subtle or sophisticated forms of noncredible responding.

The age association may partly be an artifact of norming practices within the ImpACT. Specifically, given the strong age associations on BRF, the use of wide age bands in the normative sample puts the youngest child within an age band at higher risk to fail the EVIs compared with the oldest child within the same age band. Consistent with this explanation, the greatest changes in the slope of BRF occur at the ages of 12 and 18 years, which mark the upper limit of their age band. Further evidence of this explanation comes from the pattern of percentile scores (intended to measure relative standing in terms of cognitive ability) across ImpACT scales represented in Table 2, showing a consistent pattern of lower percentile scores for the 10- and 11-year-old groups. These youngest examinees are statistically at higher risk of failing 1 of the EVIs, suggesting that the relative contribution of ability and test-taking effort cannot be separated at the low end of the age distribution. The particularly high BRF for the 10- and 11-year-olds may be resolved by the recent publication of the ImpACT Pediatric,³⁹ which is intended for children aged 5 to 11 years. However, this hypothesis is yet to be tested empirically. The association between age and BRF may reflect a cohort effect between the current sample and the cohort for which the normative data were gathered, which was more than 17 years ago.

In the absence of data on independent, well-established performance validity tests, the true meaning of EVI failures and the unusually high BRF in the youngest athletes is ultimately unknown. At the descriptive level, the age and instrumentation artifacts (ie, varying BRF as a function of examinee age and EVI cutoff) are compelling and warrant follow-up investigations. Although the very high BRFs in 10- to 12-year-old children presents a serious challenge in clinical interpretation and demands a sensible explanation, the available evidence precludes any definitive conclusion. Both extremes must be considered: the BRFs could represent either false-positive errors (ie, undeveloped cognitive skills that were mistaken for invalid responding by validity cutoffs designed for adults) or true-positives (ie, most young athletes produce invalid data). Future research should examine this issue. To the extent that the high BRFs represents false-positives, existing ImpACT EVIs should be recalibrated by applying more conservative cutoffs for failure to account for developmental factors. If the high BRFs represent true-positives, that would render the majority of baseline data meaningless, calling into question the utility of the practice in this age group.

These findings have several practical implications. The high BRF in the youngest athletes and the unexpectedly lower BRF in collegiate-age athletes necessitates differential age-specific approaches to the management of concussion. Given that more than 4 of 5 children aged 10 years fail EVIs, the clinical interpretation and utility of baseline testing in this age range is highly questionable. Until the high BRF in young children is better

understood, this group could be exempt from group-administered, computerized baseline testing. This finding raises concerns about the validity of postinjury test results in young children, suggesting that their postinjury data should be interpreted with caution. Baseline assessment of athletes aged 13 to 18 years should include procedures aimed at reducing BRF, such as strictly adhering to standardized administration practices with specific consideration to extra-test factors that may increase the BRF (limiting group size, close supervision, and minimizing distractions during testing). For the collegiate-age athletes, determinations about the validity of the baseline test scores should make use of the cumulative EVI because of the potential insensitivity of individual cutoffs in this age group.

Limitations

Providing BRF as a function of age helps clinicians contextualize the performance of athletes under their care. Knowing how common a given EVI failure is can guide diagnostic decision making because BRF foreshadows classification accuracy.⁴⁰ However, the high overall BRF signals a potential confound in the measurement model. In the absence of objective, well-validated criterion measures, the degree to which the high BRF reflects false-positive errors or truly invalid response sets has yet to be determined. The standard practice of validating performance validity tests through comparison with other, well-established performance validity tests as part of a neuropsychological test battery, administered in the traditional one-on-one fashion, is warranted. This approach has the potential to isolate situational artifacts within the ImpACT (computerized group administration of a battery designed for a specific purpose). Depending on the findings, EVIs may need to be recalibrated to account for the association of examinee age and assessment context. These recommendations, combined with the complex nature of the clinical diagnosis of concussion, highlight the necessity of concussion programs to include members who have advanced knowledge of psychometric testing and performance validity assessment in addition to knowledge of management of concussion.

Conclusions

Results converge on a number of conclusions: (1) the high BRF, particularly for younger athletes, poses significant concerns regarding the validity of the baseline ImpACT data; (2) the factors that may contribute to invalid performance vary by age, suggesting that adopting age-appropriate test administration strategies (ie, item content and test instructions that match the examinee's stage of cognitive development, individual administration, improved supervision, and strict adherence to standardized administration) may lower BRF; and (3) clinicians should routinely consider performance validity, as measured by all 4 EVIs, as well as the age-specific BRF when making return-to-play decisions based on postinjury evaluations.

ARTICLE INFORMATION

Accepted for Publication: December 8, 2017.

Published Online: March 12, 2018.
doi:10.1001/jamaneurol.2018.0031

Author Contributions: Drs Abeare and Merker had full access to all of the data in the study and take

responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Abeare, Messa, Zuccato, Erdodi.

Acquisition, analysis, or interpretation of data: Abeare, Zuccato, Merker, Erdodi.

Drafting of the manuscript: Abeare, Messa, Zuccato, Erdodi.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Abeare, Zuccato, Erdodi.

Administrative, technical, or material support: Abeare, Zuccato, Merker, Erdodi.

Study supervision: Abeare, Merker.

Conflict of Interest Disclosures: None reported.

REFERENCES

- Bryan MA, Rowhani-Rahbar A, Comstock RD, Rivara F; Seattle Sports Concussion Research Collaborative. Sports- and recreation-related concussions in US youth. *Pediatrics*. 2016;138(1):e20154635.
- Gardner A, Kay-Lambkin F, Stanwell P, et al. A systematic review of diffusion tensor imaging findings in sports-related concussion. *J Neurotrauma*. 2012;29(16):2521-2538.
- Gardner A, Iverson GL, McCrory P. Chronic traumatic encephalopathy in sport: a systematic review. *Br J Sports Med*. 2014;48(2):84-90.
- McCrory P, Meeuwisse W, Dvořák J, et al. Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. *Br J Sports Med*. 2017;51(11):838-847. doi:10.1136/bjsports-2017-097699
- West TA, Marion DW. Current recommendations for the diagnosis and treatment of concussion in sport: a comparison of three new guidelines. *J Neurotrauma*. 2014;31(2):159-168.
- Barth JT, Alves WM, Ryan TV, et al. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin HS, Eisenberg HM, Benton AL, eds. *Mild Head Injury*. New York, NY: Oxford University Press; 1989:257-275.
- Iverson GL, Schatz P. Advanced topics in neuropsychological assessment following sport-related concussion. *Brain Inj*. 2015;29(2):263-275.
- Randolph C, McCrea M, Barr WB. Is neuropsychological testing useful in the management of sport-related concussion? *J Athl Train*. 2005;40(3):139-152.
- Schatz P, Kontos A, Elbin R. Response to Mayers and Redick: "clinical utility of ImPACT assessment for postconcussion return-to-play counseling: psychometric issues". *J Clin Exp Neuropsychol*. 2012;34(4):428-434.
- Resch JE, McCrea MA, Cullum CM. Computerized neurocognitive testing in the management of sport-related concussion: an update. *Neuropsychol Rev*. 2013;23(4):335-349.
- De Marco AP, Broshek DK. Computerized cognitive testing in the management of youth sports-related concussion. *J Child Neurol*. 2016;31(1):68-75.
- Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR; Conference Participants. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol*. 2009;23(7):1093-1129.
- DeRight J, Carone DA. Assessment of effort in children: a systematic review. *Child Neuropsychol*. 2015;21(1):1-24.
- Lichtenstein JD, Moser RS, Schatz P. Age and test setting affect the prevalence of invalid baseline scores on neurocognitive tests. *Am J Sports Med*. 2014;42(2):479-484.
- Erdal K. Neuropsychological testing for sports-related concussion: how athletes can sandbag their baseline testing without detection. *Arch Clin Neuropsychol*. 2012;27(5):473-479.
- Schatz P, Glatts C. "Sandbagging" baseline test performance on ImPACT, without detection, is more difficult than it appears. *Arch Clin Neuropsychol*. 2013;28(3):236-244.
- Marvez A. Players may try to beat concussion tests. *Fox Sports*. <http://www.foxsports.com/nfl/story/nfl-players-could-try-to-beat-concussion-tests-042111>. Published April 21, 2011. Accessed May 10, 2017.
- Reilly R. Talking football with Archie, Peyton, Eli. *ESPN*. <http://www.espn.com/espn/news/story?id=6430211>. Published April 27, 2011. Accessed May 10, 2017.
- Gaudet CE, Weyandt LL. Immediate Post-Concussion and Cognitive Testing (ImPACT): a systematic review of the prevalence and assessment of invalid performance. *Clin Neuropsychol*. 2017;31(1):43-58.
- Lovell MR, Collins MW, Podell K, Powell J, Maroon J. *ImPACT: Immediate Post-Concussion Assessment and Cognitive Testing*. Pittsburgh, PA: NeuroHealth Systems; 2000.
- Kerr ZY, Snook EM, Lynall RC, et al. Concussion-related protocols and preparticipation assessments used for incoming student-athletes in National Collegiate Athletic Association member institutions. *J Athl Train*. 2015;50(11):1174-1181.
- Nelson LD, Pfaller AY, Rein LE, McCrea MA. Rates and predictors of invalid baseline test performance in high school and collegiate athletes for 3 computerized neurocognitive tests: ANAM, Axon Sports, and ImPACT. *Am J Sports Med*. 2015;43(8):2018-2026.
- Szabo AJ, Alosco ML, Fedor A, Gunstad J. Invalid performance and the ImPACT in national collegiate athletic association division I football players. *J Athl Train*. 2013;48(6):851-855.
- An KY, Kaploun K, Erdodi LA, Abeare CA. Performance validity in undergraduate research participants: a comparison of failure rates across tests and cutoffs. *Clin Neuropsychol*. 2017;31(1):193-206.
- Lichtenstein JD, Erdodi LA, Rai JK, Mazur-Mosiewicz A, Flaro L. Wisconsin Card Sorting Test embedded validity indicators developed for adults can be extended to children. *Child Neuropsychol*. 2018;24(2):247-260.
- Welsh AJ, Bender HA, Whitman LA, Vasserman M, Macallister WS. Clinical utility of reliable digit span in assessing effort in children and adolescents with epilepsy. *Arch Clin Neuropsychol*. 2012;27(7):735-741.
- Constantinou M, McCaffrey RJ. Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychol*. 2003;9(2):81-90.
- Lichtenstein JD, Erdodi LA, Linnea KS. Introducing a forced-choice recognition task to the California Verbal Learning Test-Children's Version. *Child Neuropsychol*. 2017;23(3):284-299.
- Lovell MR. *Clinical Interpretation Manual: Online ImPACT 2007-2012*. Pittsburgh, PA: ImPACT Applications; 2011.
- Higgins KL, Denney RL, Maerlender A. Sandbagging on the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) in a high school athlete population. *Arch Clin Neuropsychol*. 2017;32(3):259-266.
- Moser RS, Schatz P, Neidzowski K, Ott SD. Group versus individual administration affects baseline neurocognitive test performance. *Am J Sports Med*. 2011;39(11):2325-2330.
- Vaughan CG, Gerst EH, Sady MD, Newman JB, Gioia GA. The relation between testing environment and baseline performance in child and adolescent concussion assessment. *Am J Sports Med*. 2014;42(7):1716-1723.
- Rabinowitz AR, Merritt V, Arnett PA. A pilot investigation of the Motivation Behaviors Checklist (MBC): an observational rating scale of effort towards testing for baseline sports-concussion assessment. *J Clin Exp Neuropsychol*. 2016;38(6):599-610.
- Immediate Postconcussion and Cognitive Testing (ImPACT). ImPACT Applications Inc. <https://impactconcussion.com/>. Accessed February 5, 2018.
- Bruce J, Echemendia R, Meeuwisse W, Comper P, Sisco A. 1 Year test-retest reliability of ImPACT in professional ice hockey players. *Clin Neuropsychol*. 2014;28(1):14-25.
- Elbin RJ, Schatz P, Covassin T. One-year test-retest reliability of the online version of ImPACT in high school athletes. *Am J Sports Med*. 2011;39(11):2319-2324.
- Maerlender A, Flashman L, Kessler A, et al. Examination of the construct validity of ImPACT computerized test, traditional, and experimental neuropsychological measures. *Clin Neuropsychol*. 2010;24(8):1309-1325.
- Schatz P, Pardini JE, Lovell MR, Collins MW, Podell K. Sensitivity and specificity of the ImPACT Test Battery for concussion in athletes. *Arch Clin Neuropsychol*. 2006;21(1):91-99.
- Lovell MR. *ImPACT Pediatric: Administration and Interpretation Manual*. San Diego, CA: ImPACT Applications; 2016. <https://impacttest.app.box.com/v/Peds-Clinical-Manual-US-Public>. Accessed on May 18, 2017.
- Erdodi LA, Lichtenstein JD. Invalid before impaired: an emerging paradox of embedded validity indicators. *Clin Neuropsychol*. 2017;31(6-7):1029-1046.