This is an accepted manuscript. The original article has been published in Journal of Science and Medicine in Sport. 2017, 20(5), pp. 424-431.
http://dx.doi.org/10.1016/j.jsams.2016.09.009

a. Title:
Interpreting Change on the SCAT3 in Professional Ice Hockey Players

b. Authors:
Timo Hänninen (A)
Jari Parkkari (A)
Markku Tuominen (B,C,D)
Grant L. Iverson (E, F, G)
Juha Öhman (H)
Matti Vartiainen (I)
Teemu M. Luoto (H)

c. Institution and affiliations:
(A) UKK Institute for Health Promotion Research
Kaupinpuistonkatu 1, FI-33500 Tampere, Finland
Tampere Research Centre of Sports Medicine

(B) International Ice Hockey Federation (IIHF)
Brandschenkestrasse 50, Postfach 1817, 8027 Zurich, Switzerland
Medical Committee

(C) Liiga / Finnish Ice Hockey Association
Veturitie 13 H, 00240 Helsinki, Finland

(D) Medisport Ltd
Keskisenkatu 5, Tampere 33710, Finland

(E) Harvard Medical School, Spaulding Rehabilitation Hospital
300 First Avenue, Boston, MA, 02129 USA
Department of Physical Medicine and Rehabilitation

(F) MassGeneral Hospital for Children Sport Concussion Program
175 Cambridge Street, Boston, MA 02114, USA

(G) Home Base, A Red Sox Foundation and Massachusetts General Hospital Program,
101 Merrimac Street, Boston, MA 02215, USA

(H) Tampere University Hospital, Finland.
P.O. Box 2000, FI-33521 Tampere, Finland
Department of Neurosurgery

(I) University of Helsinki
Siltavuorenpenger 5, FI-00100 Helsinki, Finland
Institute of Behavioural Sciences, Division of Cognitive Psychology and Neuropsychology

d. Corresponding author
Timo Hänninen, timo.hanninen@fimnet.fi
UKK institute, P.O. Box 30, FI-33501, Tampere, Finland.
Tel: +358405882621 Fax: +35832829200

e. Word count: 2907; Word count for abstract: 227; Number of figures: 6 Number of tables: 3

Interpreting change on the SCAT3 in professional ice hockey players

Abstract

Objectives: To examine test-retest reliability of the SCAT3 for two consecutive seasons using a large sample of professional male ice hockey players, and to make recommendations for interpreting change on the test.

Design: A cross-sectional descriptive study.

Methods: Preseason baseline testing was administered in the beginning of the seasons 2013-2014 and 2014-2015 to 179 professional male hockey players in rink side settings.

Results: The test-retest reliabilities of the SCAT3 components were uniformly low. However, the majority of athletes remained grossly within their own individual performance range when two pre-season SCAT3 baseline scores were compared to published normative reference values. Being tested by the same person or a different person did not influence the results. It was uncommon for the Symptom score to worsen by $\geq$3 points, the Symptom Severity score to worsen by $\geq$5 points, SAC total score to worsen by $\geq$3 points, M-BESS total error points to increase by $\geq$3, or the time to complete Tandem Gait to increase by $\geq$4 seconds; each occurred in less than 10% of the sample.

Conclusions: The SCAT3 has low test-retest reliability. Change scores should be interpreted with caution, and more research is needed to determine the clinical usefulness of the SCAT3 for diagnosing concussion and monitoring recovery. Careful examination of the natural distributions of difference scores provides clinicians with useful information on how to interpret change on the test.

Keywords: Brain Concussion, Head injuries, Baseline Survey, Ice Hockey

Introduction

Head and brain injuries, especially concussions, are common and important health issues in collision sports. Ice hockey is a sport characterized by high velocity, rapid changes in direction, and injuries caused by collision with other players, boards, sticks, or pucks. The systematic collection of injury reports from team medical staff shows that the most commonly injured body region in professional male ice hockey is a player's head.[1,2]

The rink side or sideline recognition of sport-related concussion relies on a clinician's evaluation. Injury mechanics, visible signs, reported symptoms, changes in cognitive and physical performance related to concussion, and exclusion of spinal injury are the key points of assessment. International guidelines for sport-related concussion recommend the use of the Sport Concussion Assessment Tool – Third edition (SCAT3) as a supportive instrument in concussion diagnostics.[3-5] Post-injury SCAT3 scores are best interpreted when compared with either an accurate and reliable individual baseline or to age- and sport-specific normative data.[6,7]

Annual pre-season concussion baseline testing (e.g., computer-based neuropsychological assessment) is common practice in many professional contact sports. However, there are very few published studies on how often baseline testing should be administered. For example, the SCAT3 is a widely used concussion assessment instrument that has no evidence-based guidelines regarding baseline testing frequencies or intervals. For accurate comparisons between post-injury and baseline performance, it is essential to know how consistent the test-retest results are (i.e. the reliability and stability of the baseline SCAT3 assessment over time). One factor that could influence reliability is learning effect. It is not known if SCAT3 performance is improved by learning when repeatedly done, and if so how long this learning effect lasts.

The purpose of this study was to examine the long-term test-retest reliability of SCAT3 assessments in a realistic clinical setting to better understanding normal variation of the scores. We also aimed to describe if there is a significant difference between intra- and interrater reliability and whether the common practice of administering SCAT3 baseline on an annual basis is an ideal time frame or not. Suggestions for interpreting change on the SCAT3 are offered.

Methods

This study is a part of a larger research project that strives to translate international recommendations regarding diagnosis and management of concussions into practice in Finnish professional ice hockey. Ethics approval for the study was obtained from the Ethical Committee of Pirkanmaa Hospital District, Tampere, Finland (code: R13070), and each participating subject signed written informed consent according to the Declaration of Helsinki. This study was financially supported by the Finnish Ministry of Education and Culture, the Finnish Hockey League, the Finnish Medical Foundation, and the Maire Taponen Foundation. There was no involvement with any commercial sponsor for this study regarding the study design; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the paper for publication.

SCAT3 baseline testing became mandatory for all players in the highest Finnish professional male ice hockey league before the season 2013-2014, but there was not a requirement to do this annually. The total number of athletes playing in the league in two consecutive seasons (2013-2014 and 2014-2015) was 309. Only annually completed preseason SCAT3 baseline tests administered for seasons 2013-2014 and 2014-2015 were included in this study. The number of players who completed both pre-season SCAT3 baseline tests was 179 (58%). Most of the players who were not included in the study completed only one baseline. A small number of athletes were not included for various reasons (e.g., being injured in the time of preseason baseline testing).

In an effort to replicate how the SCAT3 assessment is given clinically, every player was tested individually, at least ten minutes after physical exertion, by the teams' current medical staff, who were trained to administer SCAT3 in accordance with the SCAT3 instructions in regional training sessions led by the authors before the season 2013-2014. If a player had sustained a concussion prior to testing, he had to have been asymptomatic and participated at least one month in normal game play after the concussion and before the SCAT3 baseline was administered. Demographic variables and medical history were obtained at the time of testing using the Background section of the SCAT3 form. Due to language difficulties with non-Finnish and non-English speakers, this subgroup (n=8) was excluded from the statistical analysis of the symptom evaluation and the SAC components. The Finnish translation of the SCAT3 was accomplished by a professional translator and reviewed by the authors to maintain the original denotation and connotation of items instead of exact literal or syntactical equivalence.

Descriptive statistics [mean (M), median (Md), standard deviation (SD), interquartile range (IQR)] for both seasons and the individual differences between the test-retest results of the SCAT3 components were calculated. The relationships between five categorical background variables and test-retest differences were examined. Categorical background variables included: (i) examiner: same/different, (ii) age under 20-years: yes/no (iii) self-reported history of concussions during seasons 2012-2013 and 2013-2014: yes/no, (iv) language of testing: native/non-native, (v) history of headache or migraine: yes/no. The data related to learning or attention problems (n=1) and psychiatric problems (n=1) could not be meaningfully analyzed due to small sample sizes.

The normality of the data was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. The distribution of the scores in every component of the baseline SCAT3 were skewed so the correlations between two continuous variables were measured using the Spearman rho coefficient, Kendall's tau b, and Wilcoxon Signed-Rank Test. Categorical variables in relation to continuous variables (individual test-retest absolute difference scores) were tested with the Mann-Whitney U-test (MWU). The level of statistical significance was set at 0.05. IBM SPSS Statistics 21.0 (IBM Corp. Armonk, NY, USA) was used to perform the analyses.

Results

The athletes were between the ages of 16 and 38 (M=25.4, SD=5.1) years and 19 (10.6%) players were under 20 years before the first SCAT3 baseline test. The Finnish version of SCAT3 was used with 164 (95.9%) Finnish players; the others were tested with the English version. All athletes were Caucasian. The total number of medical staff who served as examiners was 33. More than one-third (35.8%, n=64) were tested by the same person before both seasons. The average time between athlete's two baseline tests was 367 days (SD=24.2, IQR=360-378). A minority (n=25, 14%) of the players reported history of headache or migraine. A history of concussion was reported by 56.4% of the players, and 17.9% of all athletes reported having been hospitalized or undergone neuroimaging following a head trauma before the first SCAT3 baseline test. Eleven (6.6%) of them reported sustaining a concussion during the season (2012-2013) preceding the first SCAT3 baseline test in 2013. The number of athletes reporting a concussion between the two baselines was 31 (17.3%; i.e., during the 2013-2014 season or prior to preseason testing in 2014).

The descriptive statistics and test-retest correlations for the SCAT3 components in two consecutive seasons are presented in Table 1. In general, at the group level, most of the SCAT3 mean baseline scores remained stable within the one-year interval. The test-retest correlations, however, were uniformly low, with 8/11 scores having a Spearman coefficient of 0.3 or lower. We have previously published normative reference values for the SCAT3 components.[8] Those normative reference values were based on the pre-season SCAT3 test results of season 2013-2014 (n=304 athletes), and they are reprinted in Table 2. The percentages of the players who were categorized in the same normative classification range in both preseason baseline tests are presented in Table 3. As seen in column two, most of the players scored in the same normative classification range at both test and retest, and the large majority scored either in the same classification or a higher classification.

The distributions of the individual test-retest absolute difference scores are presented in Figures 1-6. The absolute difference scores of the SCAT3 components: Symptom score and severity, SAC, M-BESS, and Tandem gait had no statistically significant association with examiner (same/different), age (under/over 20 years), history of headache or migraine (yes/no), or self-reported history of concussion between baselines (yes/no). Better scores on concentration, a subcomponent of the SAC, were obtained by athletes that were tested by the same person in both baselines (positive ranks 45.3% vs. 29.5%; MWU=2817.5, p=0.01). All other subcomponents of SAC and M-BESS did not differ. Only three (1.7%) athletes failed the Coordination test on the first baseline test, and only four (2.2%) players failed the Coordination test during the second baseline. None of these players made errors on this test during both seasons. Over the two season baseline testing, none of the athletes made errors in the double leg stance of the M-BESS.

Figure 1 illustrates the distribution of test-retest difference scores for all subcomponents of the SCAT3. By examining the values in the grey shaded regions to the left of each figure, it is possible to identify unusual worsening in performance (i.e., difference scores that are found in only 10% or 5% of uninjured athletes). Worsening means greater symptoms, greater error points on the M-BESS, greater time on the Tandem Gait, or lower scores on the SAC. Most athletes (75.6%) do not show test-retest changes of those magnitudes in any subcomponent of the SCAT3. More refined analyses of changes scores, for each SCAT3 component, are provided in Figures 2-6. As seen in Figure 2, an increase (worsening) of two or more symptoms at retest occurred in 14.7% of players, and an increase of three or more symptoms occurred in only 10%. As seen in Figure 3, a total symptom severity score that increases (worsens) by three or more points occurred in 14.1% of athletes, and an increase by five or more points occurred in only 8.8%. As seen in Figure 4, worsening of two or more points on the SAC occurs in 15.5% of professional athletes, and worsening of three or more points occurs in only 7.2%. As seen in Figure 5, a worsening on the M-BESS (i.e., an increase in raw scores) by two or more points occurred in 16.7% of players, and a worsening by three or more points was uncommon, occurring in only 8.6% of players. As seen in Table 6, performing the Tandem Gait test more slowly, by three or more seconds, occurred in 18.2% of the players. Performing four or more seconds slower was uncommon, occurring in only 6.8% of players.

Discussion

This large-scale study of the one-year test-retest reliability of the SCAT3 revealed several important findings for researchers and clinicians. First, the test-retest reliabilities of each component were uniformly low and mostly considered weak according to conventional standards for interpreting stability of human performance tests (see Table 1). The symptom scores had the largest test-retest correlations. These low correlations are related, in part, to the skewed distributions of the test scores. The limited number of options in the scoring of each SCAT3 component results in limited variability and ceiling effects (accumulation of the scores for a large percentage of people, usually to minimum and/or maximum score), which causes bias to reliability estimations and may reduce the magnitude of correlations. When tests, such as the SCAT3, are used for clinical decision making, it is important for the test to have adequate reliability and validity for the intended purpose and with the specific clinical population with which it is being used.[9] The problems with reliability, illustrated in this study, are partially mitigated by having normative reference values (Table 2) and natural distributions of change scores (the Figures) for the SCAT3 components in professional hockey players. Second, the athletes' level of performance, as a group (as reflected by mean and median scores), remained stable from test to retest across the components of the SCAT3. Nearly equal number of athletes showed improvement and declines over the test-retest interval for the Symptom score, Symptom Severity score, and Tandem Gait; in contrast, performance on the SAC and M-BESS was more likely to improve on the second baseline test. Third, there was no statistically significant difference related to whether the SCAT3 was administered by the same or a different examiner. Fourth, there were no differences in test-retest difference scores in those who sustained a concussion between the two baseline assessments and those who did not. Finally, most players (i.e., 71-92%) obtained scores that were in the same normative classification range at both test and retest, based on the norms published by Hanninen and colleagues[8]. Moreover, retest scores were in the same or better normative classification range in 86-96% of athletes. This means that when an athlete is tested a second time, it is very likely that his score will be similar to or better than it was previously.

The information presented in Figures 1-6 is very useful for clinicians and researchers who want to better understand the natural distribution of test-retest difference scores on the SCAT3. This information can also be used to determine an unusual amount of change (i.e., worsening or improving) in SCAT3 performance in Finnish professional ice hockey players. Based on the current results, 10% or fewer of the athletes showed the following worsening of SCAT3 retest scores: an increase of three or more points on the Symptoms Score, an increase of five or more points on the Symptom Severity score, a worsening of three or more points on the SAC, an increase of three or more error points on the M-BESS, or an increase of four or more seconds on the Tandem Gait. Clinicians should note that professional hockey players perform perfectly or nearly perfectly on the Coordination test and the double-leg stance of the M-BESS, so errors on those tests should be considered abnormal. Clinicians can use the normative classification ranges in Table 2 in combination with these change scores. For example, as seen in Table 2, an increase of three symptoms or five points on symptom severity will usually result in a worsening in the normative classification, too. For the SAC, a decline by three or more points will always result in a worsening in the normative classification range, unless the person scores nearly perfectly at baseline (i.e., a score of 29 or 30). For the M-BESS, an increase of three error points on retesting will often, but not always, result in a change in the normative classification range. The change scores presented in this paper might prove to be particularly useful for identifying deficits in athletes who perform nearly perfectly on baseline SAC and M-BESS testing, because those athletes could worsen in performance but still have scores that are considered broadly normal.

It is important to note, however, that there are no validated rules or guidelines for interpreting change in performance on the SCAT3 in professional or amateur athletes. This requires clinical judgment. For example, an athlete who reports headache and dizziness following a hard check into the boards, and who scores two points lower on the SAC (base rate = 15.5% in uninjured athletes) and obtains two more error points on the M-BESS (base rate = 16.6% in uninjured athletes) compared to his baseline might, in fact, be experiencing the acute effects of concussion even though his change scores are not in the grey area of Figures 4 and 5, respectively. The grey areas in the figures demarcate change scores that are statistically uncommon in uninjured athletes. The sensitivities of those change scores to the acute effects of concussion, however, are unknown.

This study design was implemented in practical everyday life of professional ice hockey teams in order to maximize the generalizability and the applicability of the results. For this reason, we did not use independent external examiners. Additionally, we explored individual baseline performance changes and not only group level statistics. The strength of our study was the large sample size and the pragmatic study design.

There are several limitations to this study. The athletes' medical history was based on the SCAT3 form and therefore some relevant disease/injury history (e.g., lower limb injuries, sleep history) was not included. The information on previous concussions was solely based on self-report and it is known that athletes may underestimate their past concussions.[10] The possibility of selection bias in player recruitment exists. Finally, we focused on a very narrow demographic group of professional athletes; additional research should be conducted on other sports, levels, genders, and age cohorts.

Conclusions

The SCAT3 is designed to be used on the day-of-injury and in the initial days following injury. It is not designed to measure post-acute or long-term effects of concussion. There are no evidence-based guidelines regarding whether or not baseline preseason testing is necessary, and how often to do baseline testing (e.g., yearly or less frequently). More research is needed to determine the optimal frequency of baseline testing with the SCAT3, such as studies comparing reliability over different time periods and studies comparing post-injury scores to baseline scores after varying time intervals. It is reasonable to assume that the best way to interpret SCAT3 scores is a combination of comparing an athlete's post-injury scores to a reliable personal baseline and to quality normative data. However, the SCAT3 has low test-retest reliability, making test-retest comparisons challenging. Careful examination of the natural distributions of difference scores provides clinicians and researchers with useful information on how to interpret change on the test. It is important for clinicians and researchers to appreciate that symptom scores can increase as a result of multiple factors separate from concussion, and some variability in test-retest performance is common on the performance-based measures (i.e., SAC, M-BESS, and Tandem Gait) in uninjured athletes.

Practical Implications

- It is important to appreciate that SCAT3 symptom reporting can be affected by several factors separate from concussion, and some variability in the balance and cognition measures is common.
- Despite low test-retest reliability of the SCAT3, most players have scores that fall within a similar normative classification range across a one-year test-retest interval.
- Careful examination of the natural distributions of difference scores provides clinicians and researchers with useful information on what should be considered unusual or rare changes in performance in uninjured athletes.

Acknowledgements

We greatly appreciate the cooperation of all the team physicians and physiotherapists (Davies Paul, Helenius Petri, Huttunen Jussi, Hyde Jacob, Hämäläinen Jari, Ikonen Juhani, Isotupa Juha, Kivelä Mika, Kuosmanen Mika, Kuutamoinen Ville, Lahtinen Teppo, Lampinen Juha, Lehtonen Petri, Lukkari Janne, Mali Hero, Mustapää Sampo, Nakari Juha, Nissilä M., Partanen Joonas, Piekkola Ville, Pihkoluoma Tomi, Pohjonen Veli-Matti, Puustinen Vesa, Ruokkeinen Mikko, Suomalainen Ville, Tikanto Jukka, Toivanen Tero, Tähkäpää Mikko, Uotila Toni, Vahtera Ville, Virtala Mika, Väisänen Jonne, Yrttiaho Harri) who volunteered their time to collect the data for this project. We sincerely thank Mr. Kari Tokola, M.Sc., for the statistical advice, Ms. Birgitta Järvinen, M.Sc for help in conducting literature searches, and research assistant Anne Simi for her contribution to data collection. We would like to thank the Finnish Hockey League, Finnish Ice Hockey Association (FIHA), and Finnish Hockey Player Association (FHPA) for all their support. This study was financially supported by the Finnish Ministry of Education and Culture, The Finnish Medical Foundation; Maire Taponen Foundation, and the Finnish Hockey League. GLI acknowledges support from the Mooney-Reed Charitable Foundation and ImPACT Applications, Inc.

Disclosures

Grant Iverson has been reimbursed by the government, professional scientific bodies, and commercial organizations for discussing or presenting research relating to mild TBI and sport-related concussion at meetings, scientific conferences, and symposiums. He has a clinical and consulting practice in forensic neuropsychology involving individuals who have sustained mild TBIs (including professional athletes). He has received research funding from several test publishing companies, including ImPACT Applications, Inc., CNS Vital Signs, and Psychological Assessment Resources (PAR, Inc.).

References

1 Tuominen M, Stuart MJ, Aubry M, et al. Injuries in men's international ice hockey: a 7-year study of the International Ice Hockey Federation Adult World Championship Tournaments and Olympic Winter Games. *Br J Sports Med* 2015;49:30-36.

2 McKay CD, Tufts RJ, Shaffer B, et al. The epidemiology of professional ice hockey injuries: a prospective report of six NHL seasons. *Br J Sports Med* 2014;48:57-62.

3 Harmon KG, Drezner JA, Gammons M, et al. American Medical Society for Sports Medicine position statement: concussion in sport. *Br J Sports Med* 2013;47:15-26.

4 Giza CC, Kutcher JS, Ashwal S, et al. Summary of evidence-based guideline update: evaluation and management of concussion in sports: report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* 2013;80:2250-2257.

5 McCrory P, Meeuwisse WH, Aubry M, et al. Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport, Zurich, November 2012. *J Athl Train* 2013;48:554-575.

6 Putukian M, Echemendia R, Dettwiler-Danspeckgruber A, et al. Prospective clinical assessment using Sideline Concussion Assessment Tool-2 testing in the evaluation of sport-related concussion in college athletes. *Clin J Sport Med* 2015;25:36-42.

7 Glaviano NR, Benson S, Goodkin HP, et al. Baseline SCAT2 Assessment of Healthy Youth Student-Athletes: Preliminary Evidence for the Use of the Child-SCAT3 in Children Younger Than 13 Years. *Clin J Sport Med* 2015;25:373-379.

8 Hanninen T, Tuominen M, Parkkari J, et al. Sport concussion assessment tool - 3rd edition - normative reference values for professional ice hockey players. *J Sci Med Sport* 2016;19:636-641.

9 Mann AH. The clinical relevance of psychometric testing. *Eur J Clin Pharmacol* 1985;28 Suppl:31-34.

10 Valovich McLeod TC, Bay RC, Heil J, et al. Identification of sport and recreational activity concussion history through the preparticipation screening and a symptom survey in young athletes. *Clin J Sport Med* 2008;18:235-240.

Table 1. Descriptive statistics, effect sizes, and stability coefficients for the SCAT3.

Test	n	Baseline 2013-2014 (test)						Baseline 2014-2015 (retest)								
		M	Md	SD	IQR	Skew	Kurt.	M	Md	SD	IQR	Skew	Kurt.	r_s	tau b	sig-r
Symptom																
Score	170	1.6	1	3.0	2	3.7	16.8	1.4	0	2.1	2	1.8	3.3	0.41 $p<0.05$	0.36 $p<0.05$	p=0.59
Severity	170	2.3	1	4.5	3	3.6	15.0	2.0	0	3.5	3	3.2	15.0	0.38 $p<0.05$	0.33 $p<0.05$	p=0.56
SAC																
Total Score	169	27.0	27	1.7	2	-1.0	2.9	27.3	27	1.7	3	-0.4	-0.2	0.34 $p<0.05$	0.27 $p<0.05$	p=0.02
Orientation	170	4.9	5	0.4	0	-3.1	9.7	4.9	5	0.3	0	-3.4	9.6	-0.03 p=0.73	-0.03 p=0.73	p=0.08
Immediate Memory	170	14.6	15	0.6	1	-1.4	1.6	14.7	15	0.7	0	-2.8	9.3	0.25 $p<0.05$	0.23 $p<0.05$	p=0.06
Concentration	170	3.8	4	0.9	2	-0.1	-1.0	4.0	4	0.9	2	-0.4	-0.9	0.46 $p<0.05$	0.40 $p<0.05$	p=0.02
Delayed Recall	171	3.7	4	1.0	1	-0.7	0.6	3.7	4	1.2	2	-0.5	-0.7	0.33 $p<0.05$	0.27 $p<0.05$	p=0.91
M-BESS																
Total	176	2.1	1.5	2.7	3	3.0	13.3	1.8	1	2.7	2	3.4	15.5	0.25 $p<0.05$	0.21 $p<0.05$	p=0.02
Single leg stance	178	1.4	1	1.7	2	2.5	9.6	1.2	1	1.7	2	3.1	12.4	0.19 $p<0.05$	0.16 $p<0.05$	p=0.01
Tandem stance	176	0.7	0	1.7	1	4.3	20.7	0.6	0	1.6	1	4.6	23.8	0.25 $p<0.05$	0.24 $p<0.05$	p=0.17
Tandem Gait	44	10.8	11.2	1.7	2.5	-0.7	-0.4	10.8	11.0	1.5	2.7	-0.2	-1.1	0.04 p=0.78	0.03 p=0.81	p=0.96

Note: n = sample size, M= mean, Md = median, SD = standard deviation, IQR = interquartile range, skew = skewness, Kurt. = kurtosis, r_s = Spearman's r, tau-b = Kendall's tau b, and sig-r= Wilcoxon signed-rank test. Athletes did not make any errors in double leg stance for the M-BESS (legs together) so those values are not included in the table.

Table 2. Cutoff scores and classification ranges for the SCAT3 components for healthy professional male ice hockey players (n=304, from Hanninen et al., 2016[8])

	Broadly Normal		**Below Average/ Above Average**		**Unusually Low/ Unusually High**		**Extremely Low/ Extremely High**	
	Cutoff	**% in this Range**	**Cutoff**	**% at or Below**	**Cutoff**	**% at or Below**	**Cutoff**	**% at or Below**
Symptom Score (0-22p)	0-2	80.3%	3	19.7%	4-10	9.9%	11+	1.9%
Symptom Severity (0-132p)	0-3	82.7%	4-5	17.3%	6-18	9.5%	19+	1.8%
SAC (0-30p)	26-30	83.5%	25	16.5%	24	8.6%	23-	1.8%
Orientation (0-5p)	5	92.2%	N/A	N/A	4	7.8%	3	1.4%
Immediate memory (0-15p)	14-15	93.7%	N/A	N/A	13	6.3%	12-	0.7%
Concentration (0-5p)	3-5	95.7%	N/A	N/A	2	4.3%	0-1	0%
Digits backward (0-4p)	3-4	97.5%	N/A	N/A	1	2.5%	0	0%
Delayed recall (0-5p)	3-5	88.1%	2	11.9%	1	4.6%	0	0.4%
M-BESS (0-30 errors)	0-3	83.6%	4-5	16.6%	6-10	5.8%	11+	2.0%
Single leg stance (0-10 errors)	0-2	82.4%	3	17.6%	4+	9.3%	N/A	N/A
Tandem stance (0-10 errors)	0-1	90.9%	N/A	N/A	2-4	9.1%	5+	2.0%
Tandem gait (seconds)	12.1	76.6%	12.2-12.8	23.4%	12.9-13.9	9.6%	14.0+	1.1%

Classification ranges are based on the natural distribution of scores because the distributions were not normal. The goal was to select a below/above average cutoff that corresponded with the 25th and 75th percentile ranks, but this usually was not possible given the score distributions. Unusually low/high scores correspond with approximately the 10th and 90th percentile ranks, and extremely low/high scores correspond with approximately the 2nd and 98th percentile ranks. The classifications are worded differently based on the direction of the scoring for the SCAT3 component. Symptom scores and number of errors on the M-BESS are referred to as high and performance on cognitive testing and tandem gait are referred to as low. The months in reverse were stated correctly by 94.0% (n=265; not included as a row in Table 2). Abbreviations: p = points, SAC = Standardized Assessment of Concussion, and M-BESS = Modified Balance Error Scoring System. Adapted and reproduced with permission from: *Hanninen T, Tuominen M, Parkkari J, et al. Sport concussion assessment tool - 3rd edition - normative reference values for professional ice hockey players. J Sci Med Sport 2016;19:636-641. (To journal: we are in the process of requesting to adapt and reproduce this table).*

Table 3. The percentages of the players who were categorized in the same (same or better) normative classification range as the previous preseason baseline test.

	Total Sample	Concussion between baselines		Baseline tests performed by	
		No	Yes	Different examiner	Same examiner
Symptom Score	72.4 (87.1)	70.7 (86.4)	80.0 (90.0)	72.5 (87.2)	72.1 (86.9)
Symptom Severity	70.6 (85.9)	69.3 (85.0)	76.7 (90.0)	71.6 (86.2)	68.9 (85.2)
SAC Total Score	76.8 (88.1)	76.8 (90.6)	76.7 (80.0)	72.0 (85.0)	85.2 (95.1)
SAC Orientation	82.9 (89.4)	82.1 (93.6)	86.7 (93.3)	78.9 (90.8)	90.1 (98.4)
SAC Immediate Memory	89.4 (94.7)	90.7 (96.4)	83.3 (86.7)	88.0 (93.5)	91.9 (96.8)
SAC Concentration	92.3 (95.9)	92.1 (96.4)	93.3 (96.7)	91.6 (96.8)	93.5 (100)
SAC Delayed Recall	74.2 (91.2)	75.9 (84.4)	66.7 (76.7)	75.2 (84.4)	72.6 (80.6)
M-BESS Total errors	75.3 (90.8)	78.3 (93.7)	61.3 (77.4)	76.8 (99.0)	72.6 (91.9)
M-BESS Single leg stance	77.4 (91.0)	79.4 (92.5)	67.7 (83.9)	77.4 (92.2)	77.4 (88.7)
M-BESS Tandem stance	82.3 (92.6)	84.0 (95.1)	74.2 (80.6)	85.0 (92.0)	77.4 (93.5)
Tandem Gait	72.7 (88.1)	70.3 (81.0)	85.7 (100)	66.7 (83.3)	80.0 (85.0)

Note: Normative reference values from Hanninen and colleagues[8], presented in Table 2, were used. Percentages who were in the same and same or better (in parentheses) classification range on retesting are presented for the total sample, for those who sustained a concussion between the two baselines, and for those tested by the same or a different examiner.

Figure 1. The distributions of individual test-retest absolute difference scores.

Note: Athletes who had the exact same score (test-retest difference score = 0), better score (fewer symptoms; more points on the SAC; fewer errors on the M-BESS; faster time in Tandem Gait), and worse score during the second baseline. Worse 10% and 5% in gray. Midmost 90% in dotted line box.

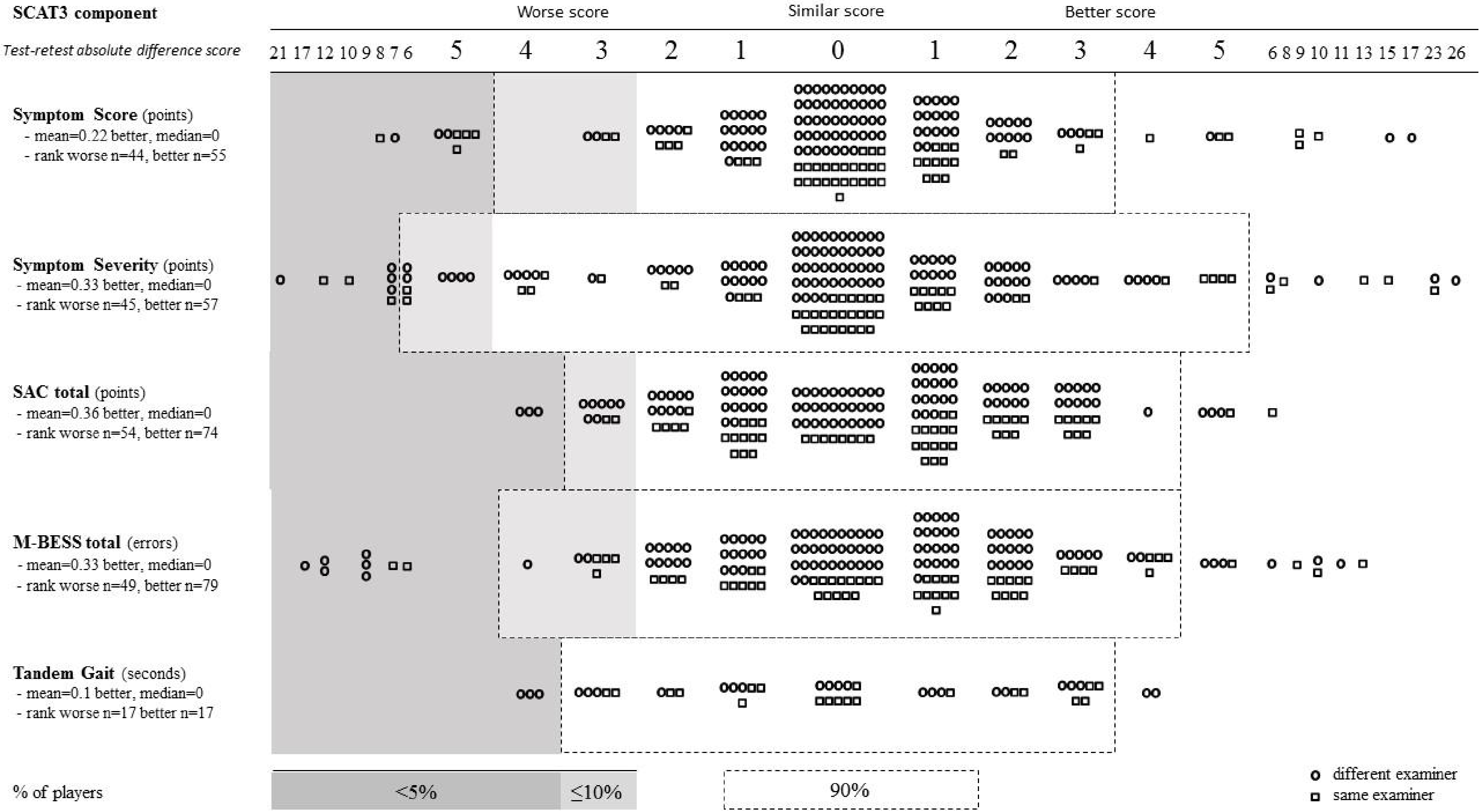

Figure 2. The distribution of individual test-retest absolute difference scores (Symptom Score)

Note: Athletes who had the exact same score (test-retest difference score = 0), better score, and worse score in the second baseline. A worse score indicates an increase in number of symptoms reported.

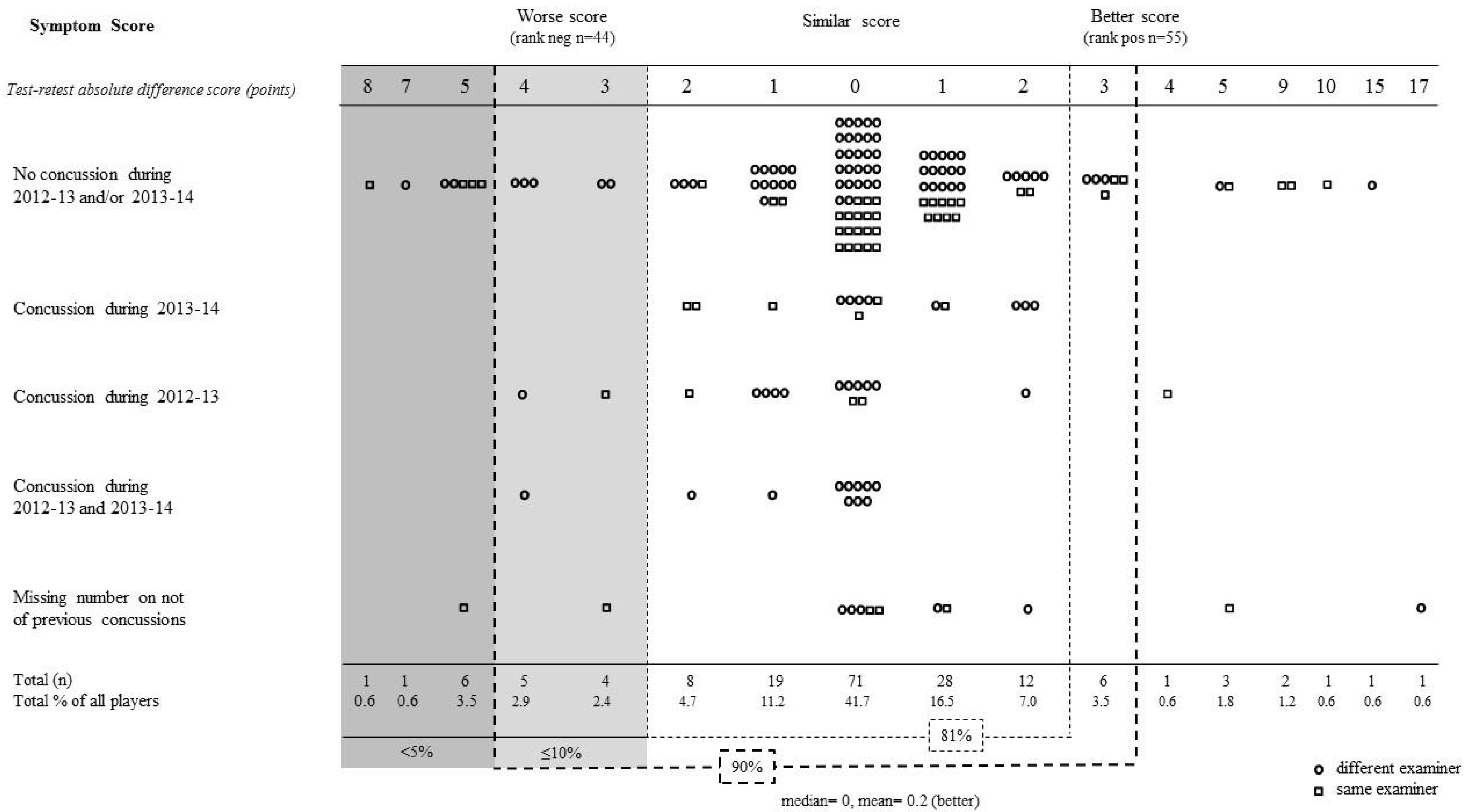

Figure 3. The distribution of individual test-retest absolute difference scores (Symptom Severity)

Note: Athletes who had the exact same score (test-retest difference score = 0), better score, and worse score in the second baseline. A worse score indicates an increase in severity of symptoms reported.

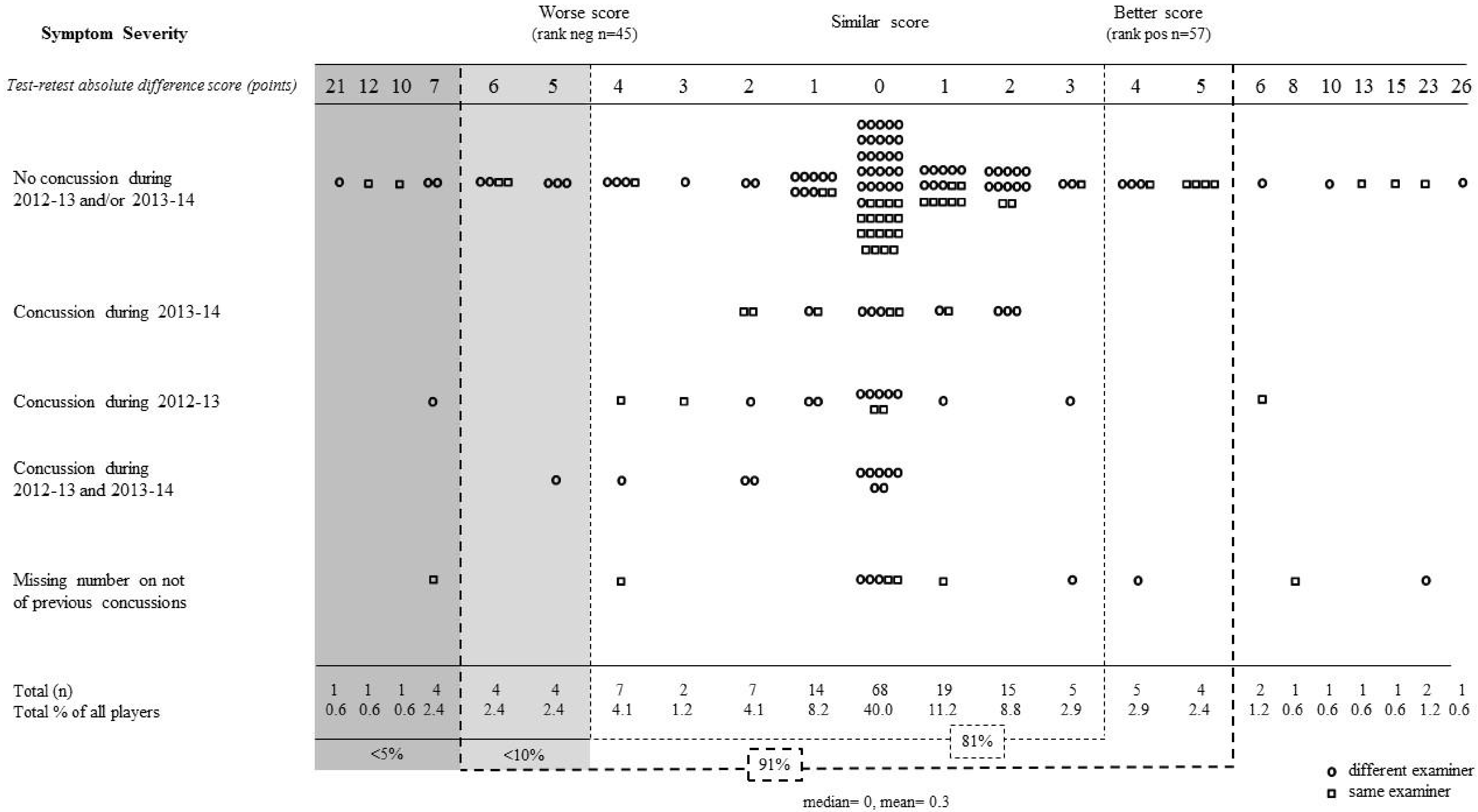

Figure 4. The distribution of individual test-retest absolute difference scores (SAC total)

Note: Athletes who had the exact same score (test-retest difference score = 0), better score, and worse score in the second baseline.

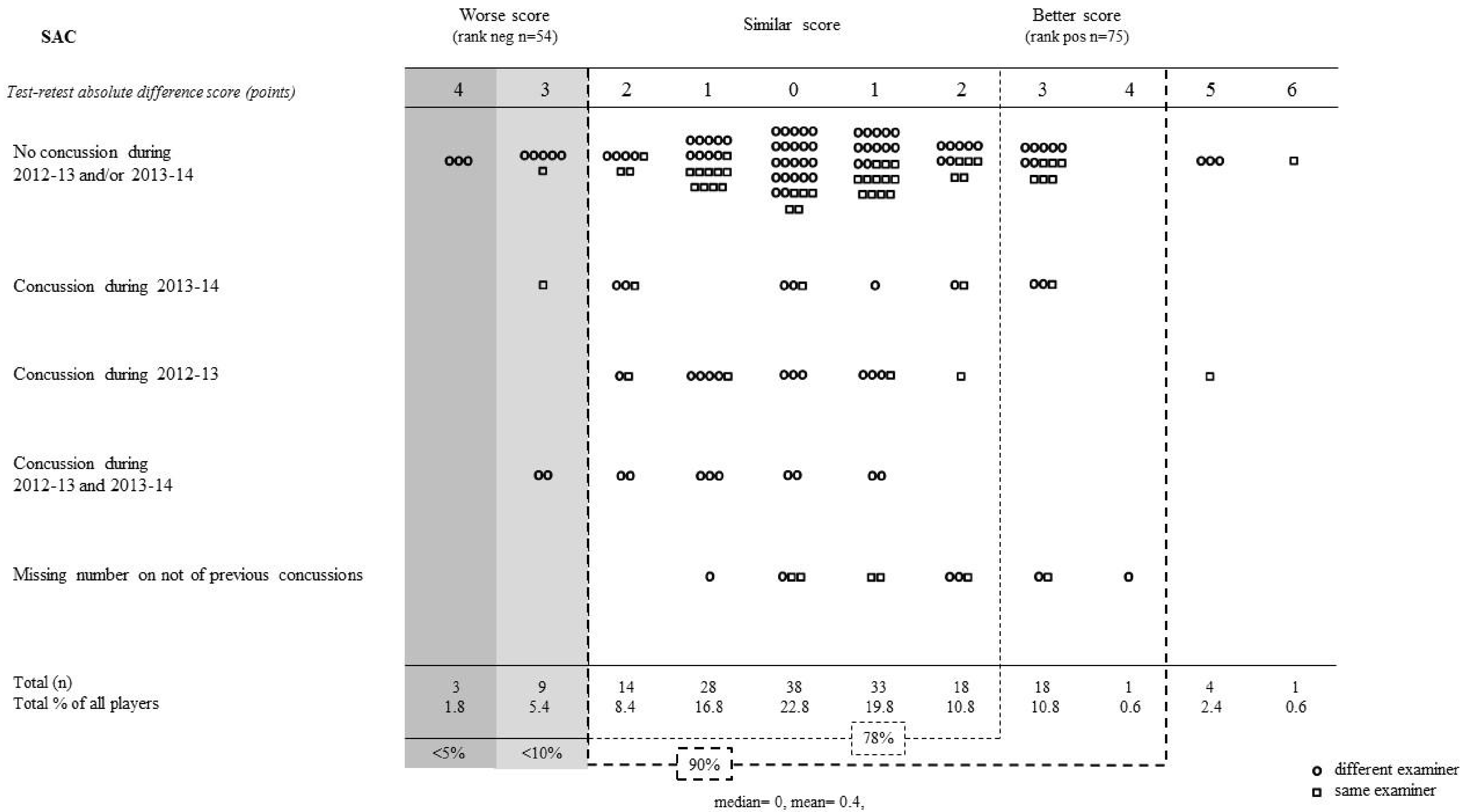

Figure 5. The distribution of individual test-retest absolute difference errors (M-BESS total)

Note: Athletes who had exact as many errors (test-retest difference errors = 0), less errors (better performance), and more errors (worse performance) in the second baseline.

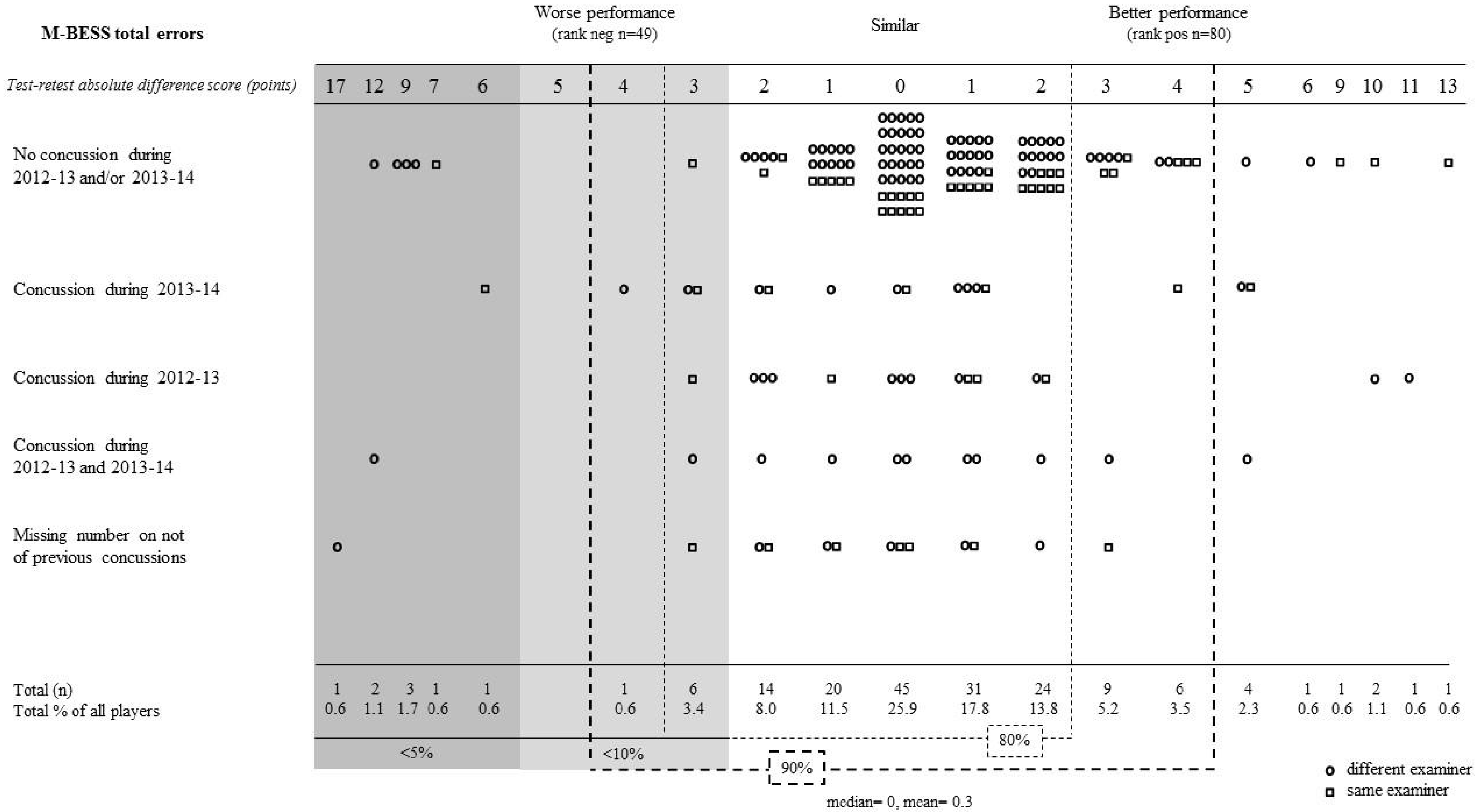

Figure 6. The distribution of individual test-retest absolute difference time in seconds (Tandem Gait)

Note: Athletes who had the exact same time in seconds (test-retest time difference = 0), better time, and worse time in the second baseline.