Advances in Sport and Exercise Medicine

Raghav Trivedi

Advances in Sport and Exercise Medicine

Advances in Sport and Exercise Medicine

Raghav Trivedi

VIDYA BOOKS

Published by Vidya Books,
305, Ajit Bhawan,
21 Ansari Road,
Daryaganj, Delhi 110002

Raghav Trivedi
ISBN: 978-93-5429-041-1

Contents

1

A sport-specific wearable jump monitor for figure skating

Dustin A. Bruening[1] *, **Riley E. Reynolds**[2], **Chris W. Adair**[2], **Peter Zapalo**[3], **Sarah T. Ridge**[1]

1 Exercise Sciences Department, Brigham Young University, Provo, Utah, United States of America,
2 Mechanical Engineering Department, Brigham Young University, Provo, Utah, United States of America,
3 Athlete High Performance Department, United States Figure Skating Association, Colorado Springs, CO, United States of America

* dabruening@byu.edu

Abstract

Advancements in wearable technology have facilitated performance monitoring in a number of sports. Figure skating may also benefit from this technology, but the inherent movements present some unique challenges. The purpose of this study was to evaluate the feasibility of using an inertial measurement unit (IMU) to monitor three aspects of figure skating jumping performance: jump count, jump height, and rotation speed. Seven competitive figure skaters, outfitted with a waist-mounted IMU, performed a total of 59 isolated multi-revolution jumps and their competition routines, which consisted of 41 multi-revolution jumps along with spins, footwork, and other skills. The isolated jumps were used to develop a jump identification algorithm, which was tested on the competition routines. Four algorithms to estimate jump height from flight time were then evaluated using calibrated video as a gold standard. The identification algorithm counted 39 of the 41 program jumps correctly, with one false positive. Flight time and jump height errors under 7% and 15% respectively were found using a peak-to-peak scaling algorithm. Rotation speeds up to 1,500˚/s were noted, with peak speeds occurring just over halfway between takeoff and landing. Overall, jump monitoring via IMUs may be an efficient aid for figure skaters training multi-revolution jumps.

Editor: Bijan Najafi, Baylor College of Medicine, UNITED STATES

1. Introduction

Figure skating has evolved throughout its history from an artform to an athletic sport. Notably, in 1990 the discipline of figures, to which the sport owes its name, was eliminated from competitions. This event accelerated an already ongoing shift in training time from figures to freestyle, and specifically to jumping. Judging criteria also evolved to reward more technically demanding programs. Today's competitive skaters train extensively, engaging in several hours daily of both on-ice and off-ice workouts. The number of jumps that elite skaters perform each day has been roughly estimated at between 50 and 100 [1], but to our knowledge has not been formally tracked. With a potential link between increased jumping and increased injury rates [2–8], there is a need to track jump training volume.

Elite figure skaters often face a potential conflict between performance improvement and injury prevention. Numerous jump repetitions are performed to learn and hone technique, yet

Competing interests: This study was partially funded by a grant from the United States Figure Skating Association to DAB and STR. Peter

Zapolo is a member of the funding body and a co-author on this paper.

repetitive impacts from jump landings are likely linked to chronic overuse injuries. For many skaters and coaches, short-term desires for performance improvement may supersede any concerns of long-term injury risks. This dilemma is not unique to figure skating. Many sports have increased monitoring and even regulations in an attempt to balance injuries and performance. For example, in baseball, pitch count limits have been instituted [9, 10], while in American football there has been an increase in monitoring head impacts related to concussions [11]. Advancements in wearable technology have increased the ability to measure and monitor various parameters related to both injury risk and performance in a variety of sports. This has the potential to increase understanding of risk factors, raise awareness of training volume, and increase training efficiency. Figure skating may also benefit from this technology.

Existing wearable technology, in the form of an accelerometer and gyroscope (called an inertial measurement unit, or IMU), could be used to monitor three important parameters for figure skating: jump count, jump height, and rotation speed. Jump count could be used to document, monitor, and regulate training volume while jump height could be used to measure jumping performance as well as monitor fatigue or performance degradation. Rotation speed could similarly be used as a performance feedback aid and a measure of performance degradation. The combination of the three parameters has the potential to help skaters train more efficiently, which may be the key to balancing performance and injury conflicts. However, there are a number of challenges in deriving these quantities from IMU signals in figure skating.

These three quantities depend on the ability to identify a jump from an accelerometer signal. For jumping in general, acceleration peaks are found near take-off and landing events. Yet, figure skating performances often contain footwork and other movements that could contain acceleration peaks that mimic those from jumps. Determining optimal peak detection tolerances is therefore important in avoiding false positives. Once a jump is identified, jump height can be calculated from flight time using a Newtonian projectile motion relationship (see Methods 2.3.2). A number of accelerometer-based methods to determine flight time have been applied to vertical jumping, with good accuracy [12–14]. Additional challenges are inherent in figure skating due to the rotation that accompanies jumps as well as potential differences between take-off and landing positions.

The purpose of this study was to determine the feasibility and initial accuracy of using an IMU to measure multi-revolution jump count, jump height, and rotation speed in figure skating. We theorized that the combined use of an accelerometer and gyroscope could successfully capture the number of multi-revolution jumps performed despite the presence of numerous confounding movements inherent in this sport.

2. Methods

2.1 Participants

Seven healthy competitive figure skaters between the ages of 12 and 27 years participated in the study (6 F, 1 M). Skaters' competition levels ranged from Juvenile to Senior (as determined by their participation in U.S. Figure Skating sanctioned events), with all skaters able to perform at least double revolution jumps. All skaters were volunteers and they and their parents (when necessary) signed assent/consent forms approved by the Brigham Young University Institutional Review Board. Subject mean ± standard deviation height and mass were 1.57±0.06 m and 49.2±6.3 kg.

2.2 Data collection

Data collection consisted of two parts, the order of which was determined by the skater's preference: isolated jumps and a competition routine. The isolated jumps were used as a training set for the jump-finding algorithm and as the primary analysis for the jump height calculations. The competition routine, which included jumps, spins, and footwork, was used as a testing set to evaluate potential acceleration signals that could confound jump identification. Jumps were classified as either successful or unsuccessful. Successful jumps were landed cleanly on one foot, while falls, under-rotated jumps, and two-footed landings were considered unsuccessful. Classification was performed from video by two experienced skaters.

Prior to collection, each skater was outfitted with an IMU (Opal model, APDM, Inc., Portland OR, USA), affixed to the lower back at approximately the L4-L5 spinal level with elastic therapeutic tape (KT Health, LLC, American Fork UT, USA). The IMU axes were aligned with the anatomical cardinal planes. As video recording was used for jump height validation, a high contrast color elastic strap was wrapped around the waist at approximately the same level as the IMU and used as a visual surrogate for the skater's center of mass. A location on the ice was marked for the isolated jumps. A high-speed camera (Sony nex-fs700U with a Canon EFS 17–55 mm lens), mounted on a tripod, was positioned approximately 8 meters from this location. A range pole was used as a distance calibration reference. Snapshots of the pole were collected with it held on the jump location as well as two other depths: approximately 0.5 m in front of and behind the location. During jump testing, the closest of the three reference positions to the jump mid-flight location was noted.

Each skater performed between six and ten isolated multi-revolution jumps. These consisted primarily of single axels (1.5 revolutions) and various double revolution jumps, with a few of the simpler triple revolution jumps (Fig 1A). The numbers and types of jumps performed by each skater were not controlled, as the goal was simply to achieve a sufficiently broad data set to perform a group analysis without overburdening each skater. Competition routines lasted between 2.0 and 3.5 minutes, depending on the level of the skater, and contained between 4 and 11 jumps (Fig 1B) interspersed with other movements. IMU data was recorded for both the isolated jumps and the competition routine at 128 frames/s. Video was recorded at 240 frames/s with a resolution of 1920 by 1080 pixels for the isolated jumps. Competition routines were recorded with a cell phone camera, which was used only as a reference.

2.3 Data analysis

The IMU data was analysed using custom MATLAB software (Mathworks, Inc., Natick MA, USA). All accelerometer signals were processed with a low pass filtered (6th order Butterworth with 10Hz cutoff) prior to analysis. MATLAB's findpeaks() function was then used to identify jumps and implement jump height algorithms. Video analysis was performed using Dartfish software (Dartfish USA, Inc. Alpharetta GA, USA).

2.3.1 Jump identification. Video and IMU data were matched by manually identifying the two acceleration peaks (from the IMU data) that corresponded with the beginning and end of each jump (from the video). These acceleration peak magnitudes were extracted from the IMU data along with the time between peaks and the peak rotational velocity occurring between them (Fig 2). The algorithm for automated jump detection was developed from this manual identification, using a three-step process. The appropriate IMU signal thresholds to be used in each step were determined from the isolated jumps, with some additional tolerances added to slightly broaden the characterization. In order to be counted, a jump had to include the following:

1. Two acceleration peaks (take-off and landing), both greater than 25 m/s^2.
2. Acceleration peaks separated by 0.3 s to 0.85 s.
3. A rotation peak greater than 688 deg/s (12 rad/s) between the two acceleration peaks.

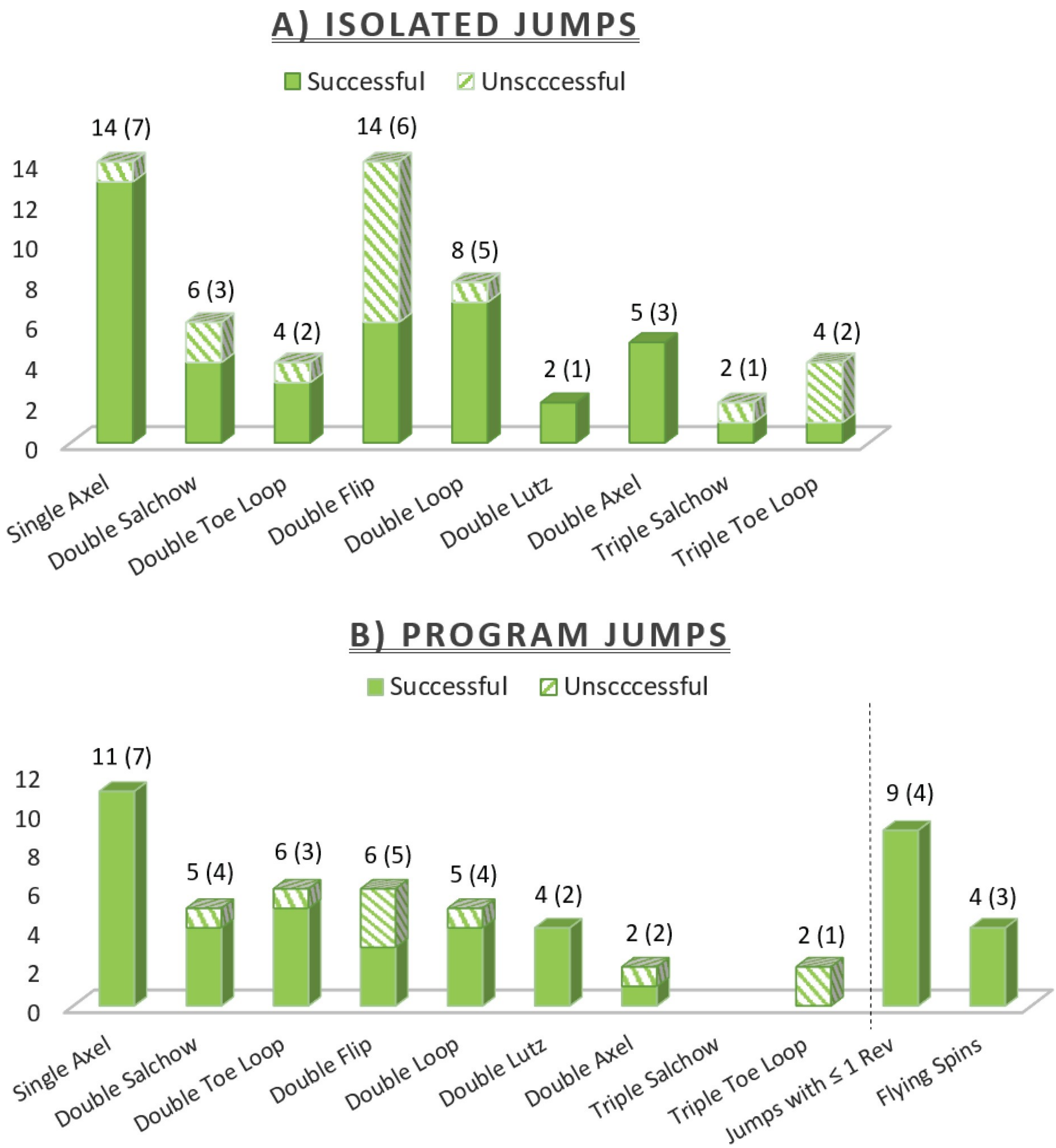


Fig 1. Tally of all jump types performed. The numbers at the top of each stacked bar represent the total number of jumps performed and number of skaters who contributed to that tally in parentheses. All jumps were marked as being landed successfully or unsuccessfully, the latter designation given to falls, under-rotations, and two-footed landings. For the isolated jumps (A), a total of 59 multi-revolution jumps were recorded from seven skaters. Forty-two (71%) of the jumps were successfully landed and 17 (29%) were unsuccessful. From the competition routines (B), a total of 41 multi-revolution jumps were recorded from the seven skaters. Thirty-two (78%) of the jumps were successfully landed and nine (22%) were unsuccessful. In addition, nine jumps of one rotation or less and 4 flying spins were tallied.

Following characterization of the isolated jumps, the three-part algorithm, with the above thresholds, was applied to the competition routines to determine its effectiveness in identifying multi-revolution jumps while avoiding false positives. The number of jumps identified by the

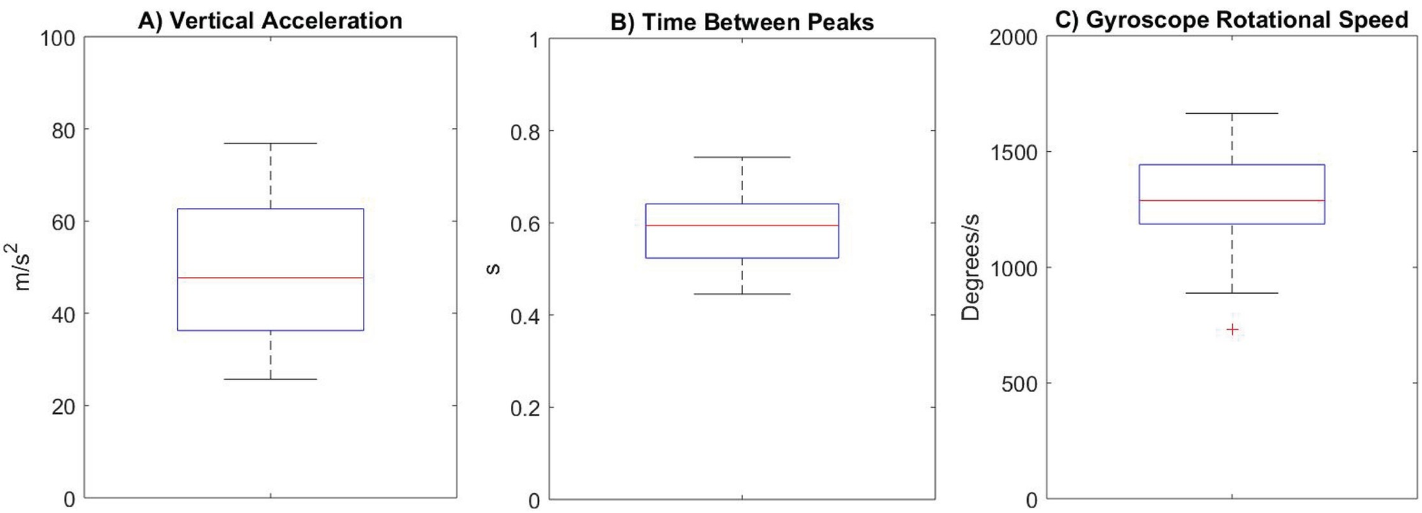


Fig 2. Boxplots characterizing all isolated jumps. A) Peak vertical accelerations during take-off and landing; B) Time between peak accelerations; and C) Peak rotation speed between peak accelerations (as measured by the gyroscope). Boxplots are median and quartiles with range whiskers, outlier represented by red pluses.

IMU algorithm, along with the number of correct jumps (tallied from reference video) were quantified at each algorithm step.

2.3.2 Jump height. IMU based jump height estimation was based on determining flight time (t). Assuming that the center of mass height at take-off and landing are the same and that air resistance is negligible, jump height can be determined from the physics of a free falling body: Height = $gt^2/8$. However, peak accelerations in most jumps typically occur slightly before the instant of take-off and slightly after the instant of landing; therefore jump height estimation focused on potential algorithms to either determine take-off and landing events in the accelerometer signal, or adjust the time between accelerometer peaks. Four potential flight time algorithms were considered (Fig 3):

1. *Gravitational threshold (GT)* (Fig 3B, triangles): Take off was marked as the time following the first peak when acceleration dropped below 1.5*g. Landing was marked as the time when it again rose above 1.5*g just before the second peak. This algorithm was adapted from Monnet et al. [13].
2. *Peak to peak scaling (PPS)* (Fig 3B, circles): The time between accelerometer peaks was multiplied by a scale factor. This scale factor was calculated by dividing the mean peak to peak time (from IMU) by the mean flight time (from video). The resulting scale factor was 0.761.
3. *Valley to valley scaling (VVS)* (Fig 3B, diamonds): The time between the first valley following the take-off peak and the last valley before the landing peak was multiplied by a scale factor. This scale factor was calculated by dividing the mean valley to valley time (from IMU) by the mean flight time (from video). The resulting scale factor was 0.887.
4. *Vertical / horizontal acceleration intersection (VHI)* (Fig 3B, squares): Take-off and landing were marked at the intersection of the vertical and horizontal accelerations (just after vertical take-off peak and just before vertical landing peak).

Flight time and jump height were obtained from the isolated jump videos and used to determine algorithm accuracy (Fig 4). First, the events of take-off (the frame when the toe of the take-off leg first lifted off the ground) and landing (the frame when the landing toe first

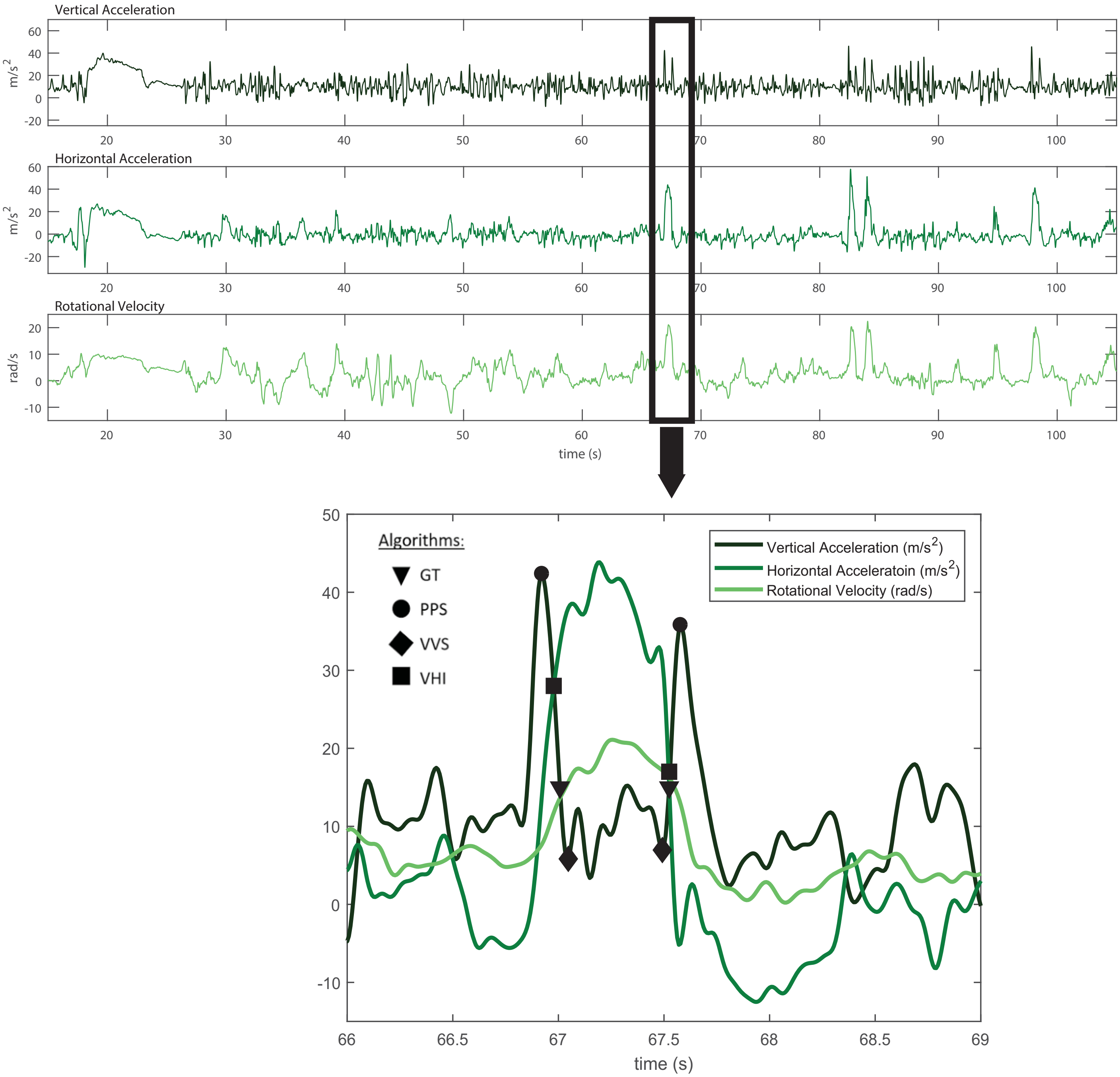


Fig 3. Example IMU signals. The first half of a roughly 3.5 minute long Novice level men's competition routine is shown (Top). The top signal represents vertical acceleration, the middle represents radial acceleration, and the bottom represents rotation around the vertical axis (from the gyroscope). A three-second sub-section surrounding a double flip jump is highlighted and expanded (bottom). Each of the three signals are plotted on the same axis to show their interactions. Take-off and landing events from the four algorithms are also shown using symbols. Take-off events are on the left, just following the first peak, while landing events are on the right, just prior to the last peak.

touched the ground) were marked. The time between the two events was considered the gold standard for flight time. Next, the positions of the high contrast waistband were measured at take-off, peak height, and landing positions. The snapshot of the ranging pole, in the appropriate reference position, was used to convert pixels to centimeters. Video analysis was performed by two separate researchers, with values averaged between them.

The temporal information from the video was considered more accurate than the spatial information, so accuracy of the algorithm was first assessed by comparing IMU algorithm based flight times to video based flight times using mean absolute errors (MAEs). Jump heights derived from these flight times were similarly compared using MAEs. The spatial video analysis was used primarily to determine whether there were any differences between take-off and landing height positions. To do this, ascent height (vertical position at peak minus takeoff) was compared to descent height (peak minus landing).

2.3.3 Rotation speed. Analysis of rotation speed focused on tabulating peak rotation speeds along with the timing of these peaks relative to take-off and landing events. This was done across all isolated and competition routine jumps. Peak rotation speeds were tabulated for all multi-revolution jumps, as well as separated between double and triple jumps.

3. Results

3.1 Jump identification

Applying the three-step jump identification algorithm (developed on the isolated jumps) to the competition routines successfully captured 39 of the 41 multi-revolution jumps. The two missed jumps were a double loop that was the second jump in a two-jump combination (no steps between jumps), and a double flip in which the skater fell on the landing. One of the single revolution jumps also met the criteria. This was a single loop that was originally intended to be a double, but the rotation was aborted in the air (i.e. popped). The algorithm did not capture the other nine ≤ one revolution jumps. None of the flying spins, footwork, or connecting movements were captured (Fig 5).

3.2 Jump height

PPS was the most accurate method, with MAEs of 0.03 s in flight time or 3.3 cm in derived jump height (Table 1, Fig 6). The worst performing method was GT, which had MAEs of 0.098 s in flight time and 7.81 cm in jump height. VHI tended to slightly overestimate jump height, while GT tended to underestimate it (PPS and VVS were scaled from the means and thus inherently matched mean values). When algorithms were applied only to the successfully landed jumps, accuracy improved slightly for PPS and VVS (but not GT and VHI), with PPS MAEs reaching 0.024 s in flight time and 2.62 cm in jump height.

Jump height measured from spatial video analysis was on average 1.5 cm higher (25.9 ± 7.1 cm) than that derived from video flight time (24.4 ± 6.5). Spatially measured landing positions were on average 1.5 cm (± 4.2 cm) higher than take-off positions (Fig 7).

3.3 Rotation speed

Peak rotation speed across all jumps ranged from 889 to 1,665 deg/s, with a mean of 1,289 deg/s (see also Fig 1C). Separating double and triple jumps, the mean speed for the doubles was 1,273 deg/s and the mean speed for the triples was 1,465 deg/s. The peak rotation speed occurred 64 ± 16% (mean ± SD) of the way between the two acceleration peaks.

4. Discussion

The purpose of this study was to determine the feasibility of using an IMU to characterize multi-revolution jumping performance in the sport of figure skating. We performed a

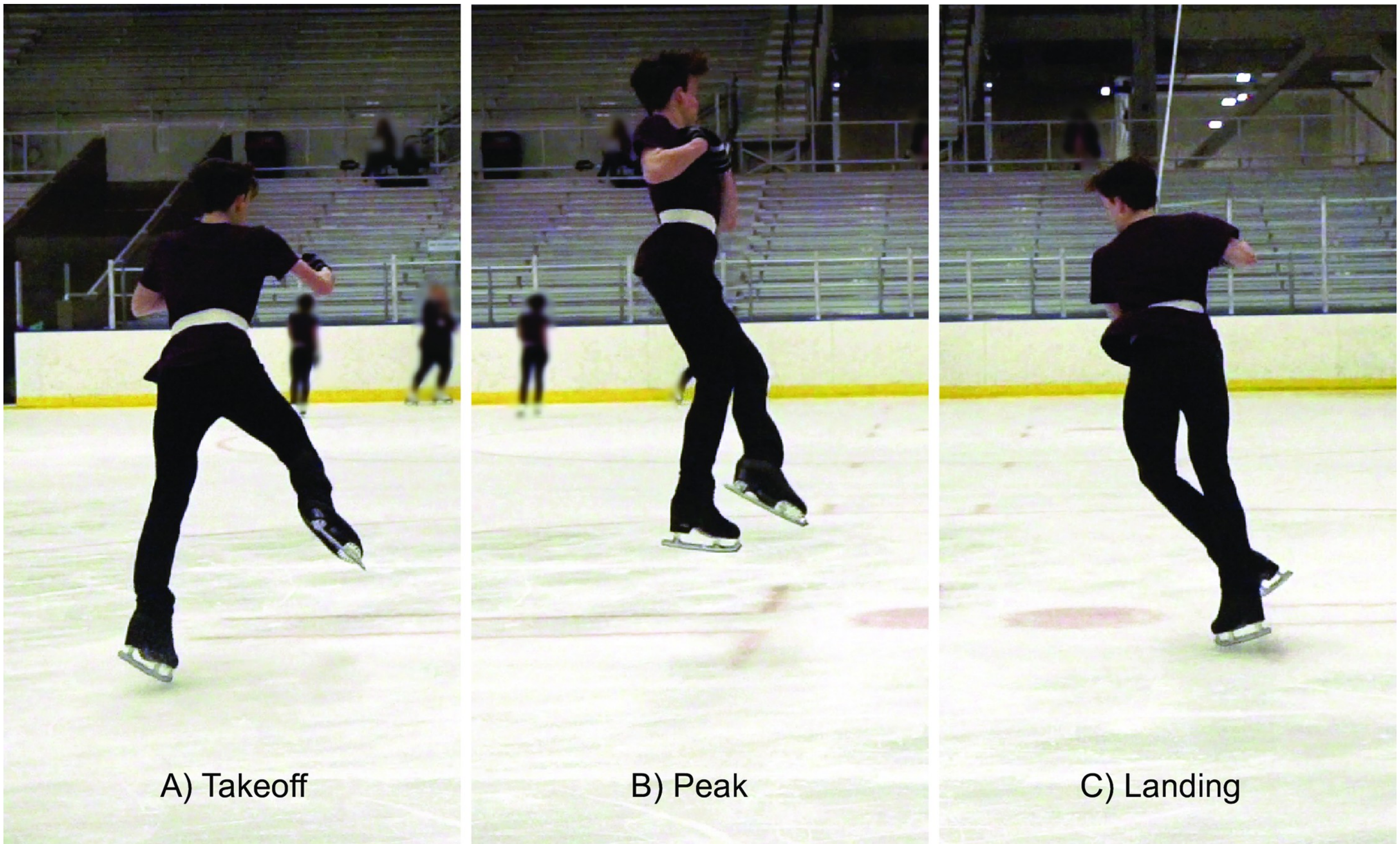


Fig 4. Video identification of jumping events. Snapshots of the take-off (A), peak (B), and landing (C) events from a subject performing an isolated double axel jump. The high contrast waistband is shown, from which spatial jump height measurements were taken. The individual in this picture has given written informed consent to publish this photograph.

systematic analysis on 59 isolated training jumps as well as competition routines containing 41 multi-revolution jumps and confounders such as spins, footwork, and other skills. The results are promising for jump identification and rotation speed quantification, but some accuracy hurdles remain for accurate jump height estimation.

4.1 Jump identification

Out-of-sample testing on the competition routines showed the potential for accurate multi-revolution jump detection despite the presence of numerous confounding movements. These non-jump movements also demonstrated the importance of all three steps in the algorithm. For example, a number of movements met the first two algorithm criteria. Jump count was reduced from 123 to 40 only after applying a rotational requirement. There were only two false negatives. One was a fall in which the skater did not have a strong landing impact peak due to off-axis positioning and continued rotation after landing contact. The other was the second jump in a two-jump combination, having a small take-off acceleration (and low jump height). There was also one false positive, a jump that was meant to be a double jump but was aborted in mid-flight. This jump displayed initial acceleration and rotation characteristics more similar to a double jump than a single jump. All other single jumps were excluded in our algorithm due either to insufficient acceleration or insufficient rotation, and usually both.

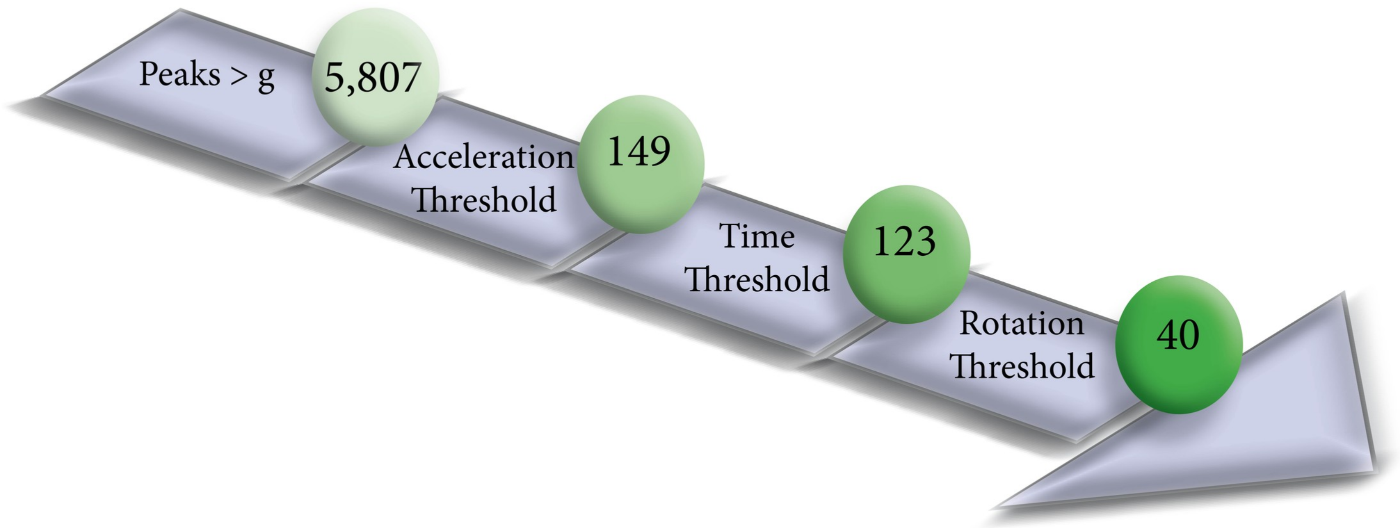


Fig 5. Jump identification flow chart. This shows the effect of each step of the jump identification algorithm on the competition routines. A total of 5,807 peak pairs were present, indicating potential jumps. This was reduced to 40 jumps with the three-step algorithm. There were 41 total multi-revolution jumps performed, 39 of these jumps were captured by the algorithm (missing 2), and one ≤ one revolution jump was captured.

We targeted jumps greater than one revolution as these are generally considered more dangerous than single jumps, perhaps due to reduced landing preparation time [15]. However, studies on impacts and injury development are still lacking; the proposed device could be a first step towards empirically determining these relationships. If the supposed link between jumps and injuries is related primarily to peak accelerations, it may be that single jumps and other movements could also contribute to the cumulative toll that jump impacts apply to the body. Many of the non-jump movements, such as footwork and flying spins, had accelerations similar to single jumps, and sometimes doubles. Lowering thresholds to capture single jumps is certainly possible, but would also capture many of these other movements. More research is needed to determine the extent to which single jumps should be monitored. Our dataset also consisted of intermediate to upper level skaters, which specifically train for double and triple jumps; it is possible that lower level skaters may have greater distinction between single jumps and other skills. Subject specific tuning of algorithm parameters may address these issues, but would require a larger sample to fully implement and evaluate.

Table 1. Flight time and jump height.

Method	All Isolated Jumps		Successful Jumps Only	
	Flight time (s)	Jump Height (cm)	Flight time (s)	Jump Height (cm)
Video	0.44 ± .060	24.4 ± 6.5	0.44 ± .058	24.4 ± 6.2
***Errors compared to video*:**				
GT	0.098 ± 0.154	7.81 ± 10.8	0.103 ± 0.165	8.00 ± 11.4
PPS	0.031 ± 0.025	3.33 ± 2.75	0.024 ± 0.015	2.62 ± 1.72
VVS	0.047 ± 0.046	7.81 ± 10.83	0.040 ± 0.041	4.36 ± 4.84
VHI	0.165 ± 0.053	4.87 ± 3.87	0.167 ± 0.056	4.85 ± 4.13

Flight time and derived jump height from video, followed by mean absolute errors (MAEs) from the four IMU-based algorithms (as compared to video). MAEs were calculated as the absolute differences between each method and the gold standard video (expressed as means ± standard deviations). Evaluation consisted of all isolated jumps followed by only the successfully landed jumps.

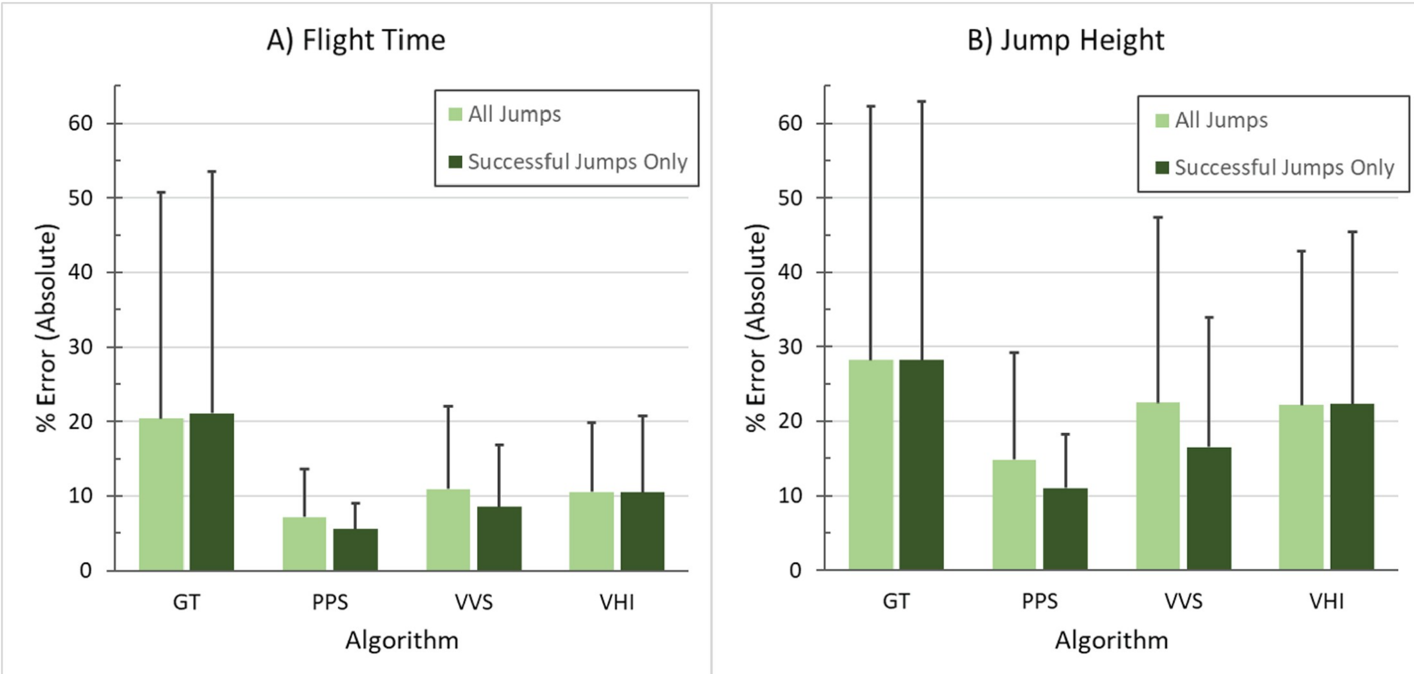


Fig 6. Algorithm accuracy expressed as % error. All four algorithms were tested on the isolated jumps to determine accuracy in flight time (A) and jump height (B).

4.2 Jump height

PPS was the most accurate algorithm for calculating jump height. However, PPS had a slight inherent advantage as its scaling factor was tuned directly to the data (note that VVS was also tuned to the data and it performed poorly). GT performed the worst but may have some future

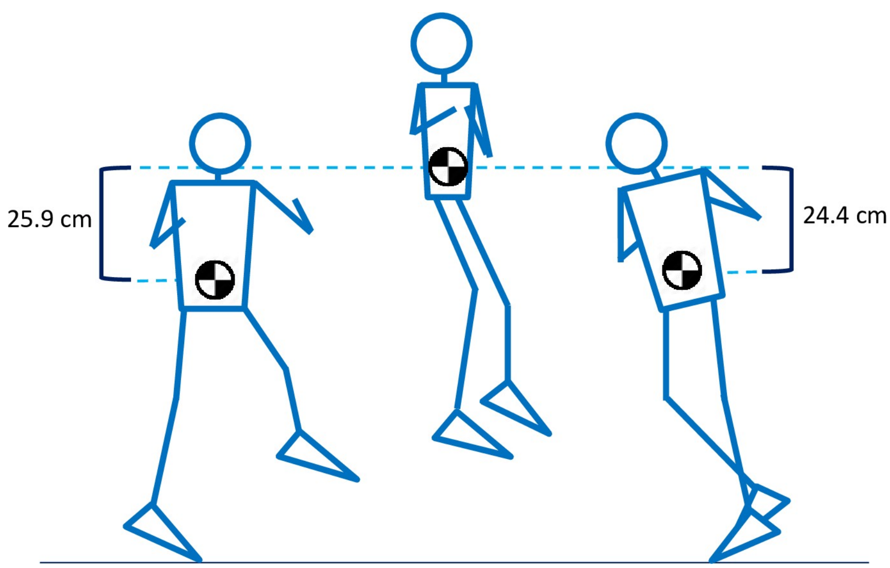


Fig 7. Spatial analysis of isolated video jump heights. Ascent height was calculated as the difference between vertical positions at peak and take-off events. Descent height was similarly the difference between peak and landing events. Landing position was on average 1.53 cm higher than take-off position.

potential. Its poor performance was due primarily to fluctuations in vertical acceleration during air time, resulting in delays in crossing the 1.5g threshold in some subjects. Thirteen of the 59 isolated jumps had obvious, large delays in threshold crossing. If these 13 are removed, accuracy improves to rival that of PPS, with mean errors of 5.7% in flight time and 11.5% in jump height (compared to 5.5% and 11.1% for PPS on successful jumps). Some of the acceleration variability during air time may be due to off-axis rotations, which could be improved with a more complex algorithm that incorporates orientation estimates. For further discussion of this specific algorithm limitation, see 4.4 below. VHI utilized the crossing of the vertical and radial acceleration signals, which we originally thought could have some physiological relevance in signifying a transition from upwards linear acceleration to rotational motion. However, the variability in the signals makes this crossing an unreliable indicator of take-off or landing. It is possible that thresholds in radial acceleration or gyroscope signals could be used to indicate how take-off and landing events relate to rotation—this could be jump type specific and would require a larger data set to fully evaluate.

While temporal information from the video analysis was considered a priori to be more accurate than spatial information, both appeared to be in general agreement, allowing us to draw some preliminary conclusions regarding future jump height accuracy. Spatial video analysis was used primarily to look for differences between take-off and landing positions. We found that landing was on average 1.5 cm higher than take-off (i.e. subjects were more extended at landing)—this would result in a slightly shorter flight time (compared to landing at an equal height) and therefore a slight underestimation of true jump height when derived from flight time. Using the mean values for flight time (0.44s) and height difference (1.5 cm), this should result in a mean underestimation of approximately 1.0 cm (24.47 cm—23.7 cm). This is close to the spatially measured jump height, which was on average 1.5 cm higher than that determined from flight time. It may therefore be possible to improve overall jump height accuracy by incorporating a fixed, or subject specific, height difference into the algorithm. Alternately, it may be sufficient to simply focus training programs around within-subject changes in flight time itself, rather than trying to derive accurate measures of jump height. For many skaters, focusing only on these changes may still be extremely beneficial in identifying fatigue onset or documenting performance improvements.

4.3 Rotation speed

Rotation speed may be the most desirable monitoring parameter for increasing skating jump performance. To accomplish more difficult jumps, skaters can either increase jump height or increase rotation speed. It is generally easier to lower rotational inertia by adjusting air position than it is to gain jumping height, as evidenced by research showing no differences in jump height between double and triple jumps [15]. A few previous studies have measured peak rotation speeds derived from video analysis [15–17]. Our peak rotation speeds from the IMU gyroscope compare favorably with these studies. In addition to peak rotation speed, other aspects of the rotation speed/ time profile may be beneficial to training; for example, in determining how quickly skaters achieve peak angular velocity. In our study, peak rotation speeds occurred on average just over halfway (64%) between the take-off and landing acceleration peaks. In order to complete more rotations, this may need to occur earlier in the jump.

4.4 Limitations and future opportunities

This was a feasibility study, and as such, we employed a relatively small sample size. There were also some discrepancies between the difficulty of the isolated jumps and the jumps contained in the competition routines, with slightly lower difficulty in the competition routines because skaters opted out of performing many of their harder jumps within the constraints of the routine. However, these differences had no detrimental effect on the jump count algorithm itself, which performed well across all competition routines. The main advantage of a larger sample size would be the ability to perform additional analyses; for example, distinguishing jump types (e.g. edge jumps from toe jumps) or investigating subject-specific threshold tuning. For example, with a larger sample size machine learning methods could be used to help classify jumps within competition routines.

Additional complexity could also be added to the algorithms to overcome potential drift and orientation hurdles. Gyroscopes have inherent drift, which we did not control for in our short-duration testing, but this may be a factor in longer duration monitoring. We also utilized the IMU sensor components independently, aligning the IMU axes to the cardinal body planes. Axis misalignment errors are therefore possible due both to placement and changes in body posture. Two approaches to overcome misalignment errors have been used in other studies. The first simply combines the individual acceleration components into a single resultant acceleration [18], but this does not appear to be feasible in skating due primarily to high radial accelerations when rotating. The second relies on a sensor fusion algorithm to calculate orientation and transform the sensor to body specific axes [19, 20]. Orientation accuracy has not been tested in figure skating, and our positive results suggest that only minor improvements are likely with the additional complexity. A few instances warrant note, however. Two of the isolated double axel jumps contained an extra acceleration peak just after take-off. This was caused by radial acceleration that bled into the vertical component because the skater had substantial off-axis body lean. This only occurred twice, but could negatively affect jump identification in some skaters. Additionally, some of the flying spins have substantial changes in body posture that could similarly confound jump identification.

5. Conclusions

In this study we developed a prototype jump monitor for figure skating. Overall, our results suggest that accurate identification of multi-revolution jumps and quantification of rotation speeds can be accomplished using a single waist-mounted IMU. Further algorithm development could increase jump height estimation accuracy. A fully integrated jump monitor which incorporates these capabilities may increase training efficiency and help skaters and coaches balance injury and performance conflicts.

Acknowledgments

The authors would also like to thank Hannah Leon and Kevin Jurgensmeier for their help in data collection and processing.

Author Contributions

Conceptualization: Dustin A. Bruening, Chris W. Adair, Peter Zapalo, Sarah T. Ridge.

Formal analysis: Dustin A. Bruening, Riley E. Reynolds, Sarah T. Ridge.

Funding acquisition: Dustin A. Bruening, Peter Zapalo, Sarah T. Ridge.

Investigation: Dustin A. Bruening, Chris W. Adair, Sarah T. Ridge.

Methodology: Dustin A. Bruening, Riley E. Reynolds, Chris W. Adair, Sarah T. Ridge.

Writing – original draft: Dustin A. Bruening, Riley E. Reynolds, Sarah T. Ridge.

Writing – review & editing: Dustin A. Bruening, Riley E. Reynolds, Chris W. Adair, Peter Zapalo, Sarah T. Ridge.

References

1. Bruening DA, Richards JG. The effects of articulated figure skates on jump landing forces. Journal of applied biomechanics. 2006; 22(4):285–95. PMID: 17293625
2. Caine CG. Figure Skating. Epidemiology of Injury in Olympic Sports, Volume XVI. 2009:393–410.
3. Campanelli V, Piscitelli F, Verardi L, Maillard P, Sbarbati A. Lower extremity overuse conditions affecting figure skaters During daily training. Orthopaedic journal of sports medicine. 2015; 3 (7):2325967115596517.
4. Fortin JD, Roberts D. Competitive figure skating injuries. Pain Physician. 2003; 6(3):313–8. PMID: 16880878
5. Porter EB. Common injuries and medical problems in singles figure skaters. Current sports medicine reports. 2013; 12(5):318–20. https://doi.org/10.1249/JSR.0b013e3182a4b94e PMID: 24030306
6. Porter EB, Young CC, Niedfeldt MW, Gottschlich LM. Sport-specific injuries and medical problems of figure skaters. WMJ-MADISON-. 2007; 106(6):330.
7. Quinn BJ. Spine Injuries in the Aesthetic Athlete. Spinal Injuries and Conditions in Young Athletes: Springer; 2014. p. 89–97.
8. Han J, Geminiani E, Micheli L. Epidemiology of Figure Skating Injuries: A Review of the Literature. Sports Health: A Multidisciplinary Approach. 2018. Epub May 1, 2018.
9. Andrews J, Fleisig G, ElAttrache N, Ciccotti M, Ahmad C, Romeo A, et al. USA Baseball Pitch Smart Program 2006. Available from: http://m.mlb.com/pitchsmart.
10. Fleisig GS. Editorial Commentary: Changing Times in Sports Biomechanics: Baseball Pitching Injuries and Emerging Wearable Technology. Elsevier; 2018.
11. Post A, Oeur A, Hoshizaki B, Gilchrist MD. An examination of American football helmets using brain deformation metrics associated with concussion. Materials & Design. 2013; 45:653–62.
12. Magnúsdóttir Á, Karlsson B. Comparing three devices for jump height measurement in a heterogeneous group of subjects. The Journal of Strength & Conditioning Research. 2014; 28(10):2837–44.
13. Monnet T, Decatoire A, Lacouture P. Comparison of algorithms to determine jump height and flight time from body mounted accelerometers. Sports Engineering. 2014; 17(4):249–59.
14. Setuain I, Martinikorena J, Gonzalez-Izal M, Martinez-Ramirez A, Gómez M, Alfaro-Adrián J, et al. Vertical jumping biomechanical evaluation through the use of an inertial sensor-based technology. Journal of sports sciences. 2016; 34(9):843–51. https://doi.org/10.1080/02640414.2015.1075057 PMID: 26256752
15. King DL, Arnold AS, Smith SL. A kinematic comparison of single, double, and triple axels. Journal of Applied Biomechanics. 1994; 10(1):51–60.
16. King D, Smith S, Higginson B, Muncasy B, Scheirman G. Figure Skating: Characteristics of triple and quadruple toe-loops performed during the salt lake city 2002 winter olympics. Sports Biomechanics. 2004; 3(1):109–23. https://doi.org/10.1080/14763140408522833 PMID: 15079991
17. Knoll K, Härtel T, editors. Biomechanical Conditions for Stabilizing Quadruple Figure Skating Jumps as a Process of Optimization. ISBS-Conference Proceedings Archive; 2008.
18. Quagliarella L, Sasanelli N, Belgiovine G, Moretti L, Moretti B. Evaluation of standing vertical jump by ankles acceleration measurement. The Journal of Strength & Conditioning Research. 2010; 24 (5):1229–36.
19. Milosevic B, Farella E, editors. Wearable Inertial Sensor for Jump Performance Analysis. Proceedings of the 2015 workshop on Wearable Systems and Applications; 2015: ACM.
20. Picerno P, Camomilla V, Capranica L. Countermovement jump performance assessment using a wearable 3D inertial measurement unit. Journal of sports sciences. 2011; 29(2):139–46. https://doi.org/10.1080/02640414.2010.523089 PMID: 21120742

2

Bayesian Estimation of Small Effects in Exercise and Sports Science

Kerrie L. Mengersen[1,2]*, Christopher C. Drovandi[1,2], Christian P. Robert[3], David B. Pyne[4,5], Christopher J. Gore[4,5,6]

1 Science and Engineering Faculty, Mathematical Sciences, and Institute for Future Environments, Queensland University of Technology, Brisbane, Australia, **2** Australian Research Council Centre of Excellence in Mathematical and Statistical Frontiers in Big Data, Big Models and New Insights, Brisbane, Australia, **3** Ceremade, Universite Paris Dauphine, Paris, France, **4** Australian Institute of Sport, Canberra, Australia, **5** Research Institute for Sport and Exercise, University of Canberra, Bruce, ACT, Australia, **6** Exercise Physiology Laboratory, Flinders University of South Australia, Bedford Park, South Australia

* k.mengersen@qut.edu.au

Abstract

The aim of this paper is to provide a Bayesian formulation of the so-called magnitude-based inference approach to quantifying and interpreting effects, and in a case study example provide accurate probabilistic statements that correspond to the intended magnitude-based inferences. The model is described in the context of a published small-scale athlete study which employed a magnitude-based inference approach to compare the effect of two altitude training regimens (live high-train low (LHTL), and intermittent hypoxic exposure (IHE)) on running performance and blood measurements of elite triathletes. The posterior distributions, and corresponding point and interval estimates, for the parameters and associated effects and comparisons of interest, were estimated using Markov chain Monte Carlo simulations. The Bayesian analysis was shown to provide more direct probabilistic comparisons of treatments and able to identify small effects of interest. The approach avoided asymptotic assumptions and overcame issues such as multiple testing. Bayesian analysis of unscaled effects showed a probability of 0.96 that LHTL yields a substantially greater increase in hemoglobin mass than IHE, a 0.93 probability of a substantially greater improvement in running economy and a greater than 0.96 probability that both IHE and LHTL yield a substantially greater improvement in maximum blood lactate concentration compared to a Placebo. The conclusions are consistent with those obtained using a 'magnitude-based inference' approach that has been promoted in the field. The paper demonstrates that a fully Bayesian analysis is a simple and effective way of analysing small effects, providing a rich set of results that are straightforward to interpret in terms of probabilistic statements.

Editor: Cathy W.S. Chen, Feng Chia University, TAIWAN

Competing Interests: The authors have declared that no competing interests exist.

Introduction

A key interest in sports science is the estimation and evaluation of small effects, such as the difference in finishing times between world-class athletes, or the impact of exercise training and/or lifestyle interventions such as dietary changes or sleep behaviors on performance [1]. While such an interest is not confined to this context [2], there are some features of sports science that make accurate and relevant estimation of small effects particularly challenging. Two such challenges are small sample sizes when dealing with international-standard, elite-level athletes and frequent small true between-individual differences in competitive performance. The issue of dealing with small sample sizes in studies has drawn comment in the fields of both medicine [3, 4] and sports science [5].

These issues have been addressed by a number of sports science researchers. For example, Batterham and Hopkins (2006) challenged the traditional method of making an inference based on a p-value derived from a hypothesis test, arguing that it is confusing, potentially misleading and unnecessarily restrictive in its inferential capability [6]. The authors suggested alternative is to focus on the confidence interval as a measure of the uncertainty of the estimated effect, and examine the proportion of this interval that overlaps pre-defined magnitudes that are clinically or mechanistically relevant. As illustration, Batterham and Hopkins identify 'substantially positive', 'trivial' and 'substantially negative' magnitudes, as well as more finely graded magnitudes. The authors then translate these proportions to a set of likelihood statements about the magnitude of the true effect.

Batterham and Hopkins justify their suggested approach and corresponding inferences by drawing an analogy between their method and a Bayesian construction of the problem. In particular, they claim that their approach is approximately Bayesian based on no prior assumption about the distribution of the true parameter values. This has drawn criticism by a number of authors, such as Barker and Schofield (2008) who–rightly–point out that the approach is *not* Bayesian, and that the assumed priors in an analogous Bayesian approach may indeed be informative [7]. More recently, Welsh and Knight (2014) further criticised the approach of Batterham and Hopkins and suggested that relevant statistical approaches should use either confidence intervals or a fully Bayesian analysis [8].

The aim of this paper is to provide a Bayesian formulation of the method proposed by Batterham and Hopkins (2006) and provide a range of probabilistic statements that parallel their intended magnitude-based inferences. The models described here can be expanded as needed to address other issues. For further exposition, the model is described in the context of a small-scale athlete study authored by Humberstone-Gough and co-workers [9], which employed Batterham and Hopkins' approach to compare the effect of two altitude training regimens (live high train low, and intermittent hypoxic exposure) on running performance and blood measurements of elite triathletes.

Methods

General model

Both Bayesian and frequentist approaches require specification of a statistical model for the observed data, which contains a number of parameters that need to be estimated. Bayesian methods are different from frequentist approaches in that the parameters are treated as random variables. That is, they are considered as having true, but unknown, values and are thus described by a (posterior) probability distribution that reflects the uncertainty associated with how well they are known, based on the data. The posterior distribution is obtained by multiplying the likelihood, which describes the probability of observing the data given specified values of the parameters, and the prior distribution(s), which encapsulates beliefs about the probability of obtaining those parameter values independently of the data. These priors may be developed using a range of information sources including previous experiments, historical data and/or expert opinion. Alternatively, they may be so-called uninformative or vague distributions, to allow inferences to be driven by the observed data.

This study describes a simple statistical model that might be considered in the context of examining small effects in sports science and also some possible prior distributions that might be placed on the parameters of this model. Some extensions to the model are considered in a later section.

Suppose that there are G treatment groups. For the gth group ($g = 1,\ldots,G$), let n_g denote the total number of individuals in the group, $y_{i(g)}$ denote an observed effect of interest for the ith individual in the group ($i = 1,\ldots,n_g$), y_g denote the set of observations in the group, $\bar{y}_g$ and s_g^2 denote respectively the sample mean and sample standard deviation of all the observed responses from the group, and $v_g = n_g-1$ denote the degrees of freedom. For example, in the following case study, there are $G = 3$ groups (training regimens); $y_{i(g)}$ is the difference between the post- and pre-treatment measurements for a selected response for the ith athlete in the gth training regimen, and $\bar{y}_g$ is the average difference for that group.

Assume that an observation $y_{i(g)}$ is Normally distributed around a group mean μ_g, with a group-specific variance σ_g^2, i.e.:

$$y_{i(g)} \sim \text{Normal}(\mu_g, \sigma_g^2) \tag{1}$$

A vague prior density is adopted for the pair of parameters (μ_g, σ_g^2) [10] so that:

$$p(\mu_g, \sigma_g^2) \propto \sigma_g^{-2} \tag{2}$$

(where $\propto$ denotes proportional to). Based on [1] and [2], the posterior conditional distributions for μ_g and σ_g^2 are given by

$$\mu_g|\sigma_g^2, y_g \sim \text{N}(\bar{y}_g, \frac{\sigma_g^2}{n_g}) \tag{3}$$

$$\sigma_g^2|y_g \sim \text{Inverse}\chi^2(\upsilon_g, s_g^2). \tag{4}$$

The marginal posterior distribution for μ_g can be shown to have a t distribution on v_g degrees of freedom: [10]

$$\mu_g|y_g \sim t_g(\bar{y}_g, \frac{s_g^2}{n_g}) \tag{5}$$

so that

$$(\mu_g - \bar{y}_g)/(s_g^2/\sqrt{n_g})|y_g \sim t_{\upsilon g} \tag{6}$$

Relationship with frequentist results

The marginal posterior distributions for σ_g^2 and μ_g, based on the data, are given by Eqs (4) and (5), respectively. Because of the choice of the vague prior (Eq (2)), these distributions can be shown to be closely related to analogous distributions for the (appropriately scaled) sufficient statistics, given μ_g and σ_g^2, based on frequentist sampling theory: [10]

$$\upsilon_g s_g^2 | \sigma_g^2 \sim \chi_{\upsilon g}^2 \tag{7}$$

$$(\bar{y}_g - \mu_g)/(s_g^2/\sqrt{n_g}) \sim t_{\upsilon g}. \tag{8}$$

Estimation of values of interest

A range of posterior estimates (conditional on the available data) arising from the model may be of interest, including:

1. the mean and standard deviation for each group (e.g., each training regimen in the study), given by μ_g and σ_g^2, respectively
2. the difference between the group means: $\delta_{kl} = \mu_k - \mu_l$ and the associated standard deviation of this difference, σ_{kl}
3. a (1−α)% credible interval (CI) for a measure of interest, θ, say, such that there is a posterior probability (1−α) that θ lies in this interval (e.g., θ could be the mean of group 2, i.e., $\theta = \mu_2$, and a 95% CI of (3.1, 4.2), for instance, indicates that the probability that μ_2 is between 3.1 and 4.2, given the data, is 0.95), which is a much more direct statement than is possible under a frequentist approach
4. Cohen's *d* [11] for the difference between two groups, given by $d_{kl} = \delta_{kl}/\sigma_{kl}$ when comparing groups k and l, $k \neq l$
5. the probability that Cohen's *d* exceeds a specified threshold such as a 'smallest worthwhile change' (*SWC*,[6]), given by $\Pr(d_{kl} > \text{SWC})$ or $\Pr(d_{kl} < -\text{SWC})$, depending on whether d_{kl} is positive or negative, respectively
6. the predicted outcome of each individual under each training regimen, regardless of whether or not they have participated in that training, obtained from Eq (1), with an estimate of the corresponding uncertainty of this prediction
7. the ranks of each individual under each training regimen, with corresponding uncertainty in these orderings.

Given the data y_g for each group (and hence the sufficient statistics $\bar{y}_g$ and s_g^2), it is straightforward to use Eqs (4) and (5) to compute posterior estimates μ_g and σ_g^2, and other probabilities of interest. An alternative, simple approach is to simulate values of interest using Eqs (3) and (4) iteratively, employing a form of Markov chain Monte Carlo (MCMC) [12]. A more technical explanation of this approach including the Gibbs sampling techniques is provided by Geman and Geman [13]. At each iteration, a value of σ_g^2 is simulated from Eq (4) and then a value of μ_g given that value of σ_g^2 is simulated from Eq (3). This process is repeated a large number of times. The simulated values can be used to compute other measures (e.g. $\exp(\mu_1 - \mu_2)$ if this is of interest), indicators $I(\mu_1 > c)$ or $I(\mu_1 > \mu_2)$ and so on. Then $E(\exp(\mu_1 - \mu_2))$, $\Pr(\mu_1 > c)$ and $\Pr(\mu_1 > \mu_2)$ can be estimated (where E denotes expectation) as the respective averages of these values over all of the iterations. Similarly, at each iteration, the simulated parameter values can be input into Eq (1) to obtain predicted values of y for each individual under each regimen, and the individuals can be ranked with respect to their predicted outcome. The posterior distributions for individual predicted outcomes, and the probability distribution for the ranks, are computed from the respective values obtained from the set of iterations.

The Cohen's *d* is a standardized effect size estimate, calculated as the difference between two means divided by the corresponding standard deviation. While there are many effect size estimators, Cohen's *d* is one of the most common since it is appropriate for comparing between the means of distinctly different group and it has appealing statistical properties; for example it has a well known distribution and is maximum likelihood estimator [14].

Model extensions

The model described above can be easily extended in a range of ways. Three such extensions are considered here. The first extension is that other prior distributions can be considered instead of Eq (2) above. For example, another common approach is to assign a normal distribution for the group means,

$$\mu_g \sim \text{Normal}(\text{M}, \text{V}) \tag{9}$$

and a Uniform distribution for the standard deviations,

$$\sigma_g \sim \text{Uniform}(0, \text{R}), \tag{10}$$

where M and V denote the mean and variance of the normal distribution, respectively, and R is the upper bound of the uniform distribution. Alternatives to the uniform are the half-normal or half-Cauchy. If the sample sizes within groups are small and little is known *a priori* about the comparative variability of measurements within and between the groups, then σ_g^2 can be imprecisely estimated; to avoid this, the individual variances be replaced by a common variance, σ^2 say.

There are many ways of setting the values of M, V and R. For example, if there is no prior information about these values and if the groups are considered to be independent, this can be reflected by specifying very large values of V and R, relative to the data. This means that the priors in Eqs (9) and (10) will have negligible weight in the posterior estimates of the group means μ_g and variances σ_g^2. If V is sufficiently large, the value of M will not matter, so it is commonly set to 0 in this case. Alternatively, the groups can be perceived as having their own characteristics (described by μ_g and σ_g^2) but also being part of a larger population with an overall mean M and variance V. This random effects model is very common as it helps to accommodate outliers and improve estimation of small groups. Another alternative is to use other information to set the values of M, V and R. This information can include results of previous similar experiments, published estimates, expert opinion, and so on. Depending on the problem and the available information, different values of M, V and R can be defined for the different groups. The Bayesian framework can be very helpful in providing a mechanism for combining these sources of information in a formal manner.

The second extension is that the model described in Eq (1) can be expanded to include explanatory variables that can help to improve the explanation or prediction of the response. This is the model that is adopted in the case study described below, where the explanatory variables comprise the group label and a covariate reflecting training-induced changes. For this purpose, Eq (1) is extended as follows:

$$y_i = x_i'\beta + \varepsilon \tag{11}$$

where the explanatory variables and their regression coefficients are denoted by x and β, respectively, and ε_i describes the residual between the observation y_i and its predicted value $x_i'\beta$. Note that the superscript $'$ denotes the transpose. It is common to assume that $\in_i$~Normal

$(0,\sigma^2)$. Normally distributed priors are placed on the parameters in this regression model:

$$\beta \sim \text{Normal}_k(b_0, \text{B}_0^{-1}); \sigma^2 \sim \text{Gamma}(c_0/2,\ \ d_0/2) \tag{12}$$

where k represents the number of parameters, Normal$_k$ indicates a k-dimensional Gaussian distribution and Gamma indicates a Gamma distribution described by shape and scale parameters, in this case given by constants c_0 and d_0.

An uninformative prior specification can be defined for β by setting zero values for the mean vector b_0 and precision matrix B_0. Similarly, negligible prior information about the magnitude of the residuals is reflected by setting small values for c_0 and d_0 in the distribution for σ^2 [15].

An alternative, popular formulation is to use Zellner's g-prior, whereby the variance of the prior for β is defined in terms of the variance for the data. More explicitly, b is specified as a multivariate normal distribution with a covariance matrix that is proportional to the inverse Fisher information matrix for β, given by $g(x^T x)^{-1}$. This is an elegant way of specifying the 'information' contained in the prior, relative to that contained in the data: the value of g is analogous to the 'equivalent number of observations' that is contributed to the analysis by the prior [16, 17].

The third extension is the choice of the response y. This depends on the aim of the analysis, biological and other contextual knowledge of the problem, and the available data. The residuals are assumed to have a normal distribution with a mean of zero, and normally distributed priors can be defined as the difference between an individual's post-training and pre-training measurements, the difference of the logarithms of these measurements, the relative difference between the pairs of measurements (i.e. (post-pre)/pre) or some other context-relevant transformation.

Case Study

The Bayesian approach described above was applied to a study by Humberstone-Gough *et al.* [9] who used a two-period (pre-post) repeated measures design to compare the effects of three training regimens 'Live High Train Low' altitude training (LHTL), 'Intermittent Hypoxic Exposure' (IHE) and 'Placebo' on running performance and blood characteristics. The study comprised eight subjects (elite male triathletes) in each regimen, and had one dropout in the LHTL group. Although ten running and blood variables were considered in the original study; three variables with the most complete data are selected here for illustration: hemoglobin mass (Hbmass, units of grams), submaximal running economy (RunEcon, units of L O_2.min^{-1}) and maximum blood lactate concentration (La-max, units of mmol/L). The authors also employed a covariate reflecting training-induced changes, namely the percent change in weekly training load from pre- to during-camp for each individual athlete. The data used for the analyses are shown in S1 Table (data extracted from original study of Humberstone-Gough et al (2013 and provided by co-author Gore).

Casting this study in terms of the models described above, there are $G = 3$ groups denoting the training regimens (Placebo by $g = 1$; IHE by $g = 2$; LHTL by $g = 3$). Letting pre$_i$ and post$_i$ denote respectively the pre- and post-treatment measurements for the ith individual, an (unscaled) effect of interest, y_i, was defined in terms of the difference between the pairs of measurements:

$$y_i = \text{post}_i - \text{pre}_i. \tag{13}$$

A log transformation was adopted in the original analysis by Humberstone-Gough *et al.* [9] but was not performed in the analysis described below, as there was insufficient information in the observed data to strongly motivate a transformation of the measurements, particularly after adjusting for the covariate reported by Humberstone-Gough *et al.* (comparative summary plots not shown). However, it is acknowledged that this decision was based purely on the available data and there may be compelling biological or experimental reasons for choosing the log (or other) scale; for example, under this transformation covariates can be considered to have multiplicative rather than additive effects on the original response. On the one hand, retaining the original scale allows for more direct interpretation of the results. On the other, if the underlying assumptions are not met, the inferences based on the results must be treated with caution. In this study, the premise was adopted of not transforming unless there is a compelling domain-specific or statistical reason to do so. Hence the decision was made not to take a log transformation of the data as other authors have suggested–a statistical decision–and to consider a relative change in performance as well as an absolute difference–a domain-based decision since this measure is of interest to sports scientists. A similar issue arises about the inclusion of covariates in a small sample analysis. In this case, the associated regression parameters may be estimated with substantial uncertainty and the usual model comparison methods are often inadequate in determining any associated improvement in model fit. Again, the decision may be more domain-based than statistical. In the study considered here, results were reported with and without a covariate that was considered to be important for sports scientists, and a deliberate decision was made to avoid formal model comparison. These issues of data transformation and model comparison for small samples merit further research.

Here we consider instead an analogous scaled effect defined in terms of the relative difference between the pairs of measurements:

$$y_i = (\text{post}_i\text{-pre}_i)/\text{pre}_i. \quad (14)$$

For both the unscaled response given by Eq (13) and the relative response given by Eq (14), the list of posterior estimates of interest were:

- the differences between the two experimental training regimens (IHE, LHTL) and the Placebo group, given by δ_{12} and δ_{13}, respectively, and the difference between the two training regimens IHE and LHTL, given by δ_{23};
- Cohen's *d* for each of the two experimental regimens compared with the Placebo regimen, given by $d_{12} = \delta_{12}/\sigma_{12}$ for IHE and $d_{13} = \delta_3/\sigma_{13}$ for LHTL;
- Cohen's *d* for the standardized difference between LHTL versus IHE, given by $d_{23} = \delta_{23}/\sigma_{23}$;
- the probabilities that the standardized difference between the IHE training regimen and the Placebo exceed the 'smallest worthwhile change' (*SWC*, specified as a standardised change of *0.2* based on previous recommendations [18]), denoted by $SWCU_{12} = \Pr(d_{12} > 0.2)$ and $SWCL_{12} = \Pr(d_{12} < -0.2)$;
- analogous probabilistic comparisons with the *SWC* for the difference between the LHTL training regimen and the Placebo, and the LHTL and IHE training regimens,
- the posterior distributions of the expected outcome $E(y_{ij}) = \beta_0 + \beta_1 X + \beta_2 I_{j=1} + \beta_3 I_{j=2}$ for the *i*th individual under the *j*th training regimen (where $I_{j=1} = 1$ if the treatment is IHE and = 0 otherwise, and $I_{j=2} = 1$ if the treatment is LHTL and = 0 otherwise); the expected outcome, obtained by substituting the simulated parameter values ($\beta_0, \beta_1, \beta_2, \beta_3$) into this equation at each MCMC simulation,
- the analogous posterior predicted outcome for each individual under each training regimen, which allows for within-subject variation around the expected outcome, i.e., $y_{ij}^{\text{pred}} = y_{ij}^{\text{pred}} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2)$, which is obtained in the same manner as above,

- the ranks of the individuals based on their expected and predicted outcomes under each of the treatment regimens; again, this is a probability distribution, reflecting the fact that rankings may change depending on the precision of the estimated treatment effects and within-subject variation.

Note that although the denominator of the Cohen's d_{kl} values can be calculated using the traditional equation, i.e., $\sigma_{kl} = \sqrt{\mathrm{Var}(\delta_l - \delta_k)} = \sqrt{((\nu_l \mathrm{Var}(\delta_l) + \nu_k \mathrm{Var}(d_k))/(\nu_l + \nu_k))}$, this can also be directly calculated using the simulated values of d_{kl} obtained from the MCMC iterations, i.e., $\sigma_{kl} = \sqrt{\mathrm{Var}(d_{kl})}$.

Based on exploratory plots of the relationships between the observed pre- and post-training values of Hbmass, RunEcon and La-max among the three groups, and with the covariate, two analyses of the data were undertaken. In the first analysis, the covariate was excluded and the model was fit using Eqs (3) and (4). In the second analysis, the covariate was included given previous work showing that training load can influence the hemopoietic response [19] and the model was fit using Eq (11). The models were implemented using the statistical software R, with packages BRugs and R2WinBugs, which call WinBUGS [15, 20, 21], and MCMCregress in the MCMCpack library [22]. Estimates were based on 150,000 MCMC iterations, after discarding an initial burn-in of 50,000 iterations. For comparability with Humberstone-Gough *et al* [9], the results of the second analysis are reported below. The R code for this model is presented as a text file in S1 Text.

As described above, the primary analyses for the case study utilized an uninformative prior specification for β in Eq (12), which was obtained by setting the values of the prior mean vector b_0 and prior precision matrix B_0 to zero. The impact of informative priors was evaluated by considering a range of non-zero values for these terms, with Hbmass as the response measure. The values were motivated by the results of a recent meta-analysis of training regimens on Hbmass [23], which reported a mean response of 1.08% increase in Hbmass per 100 hours of LHTL training. Based on the study of Humberstone-Gough with 240 hours of exposure, the prior expectation is thus that the mean increases for the LHTL and IHE groups would be 2.6% and 0% respectively. The latter figure is also supported by a report that 3 h/day at 4000–5,500 m was inadequate to increase Hbmass at all [24]. This literature also provides a prior expectation of 0% increase in Hbmass of the Placebo group. The 2013 meta-analysis [23] also provided an estimate of 2.2% for the within-subject standard deviation of Hbmass.

Results

The distribution of the covariate X (representing the % change in weekly training load from pre- to post-camp) within and among the three training regimens (Placebo, IHE, LHTL) is displayed in Fig 1. The plots show that there is non-negligible variation between individuals within a regimen with respect to this variable and substantive differences between the regimens. It is clear that adjustment needs to be made for X before evaluating the comparative impact of the three regimens. This is accommodated in the regression model described in Eq (11).

Scatterplots of the unscaled differences given by Eq (13) and scaled differences given by Eq (14) are presented in Figs 2–4. Based on these plots, there is no clear visual association between the three measurements under consideration in this case study (Hbmass, RunEcon and La-max), or between these measurements and the covariate.

Plots of the posterior distributions of the differences between the training regimens, IHE vs Placebo, LHTL vs Placebo, LHTL vs IHE, given by δ_{12}, δ_{13} and δ_{23}, respectively, are shown in Fig 5. Corresponding posterior estimates of the effects (mean, s.d., 95% and 90% credible intervals) are given in Table 1.

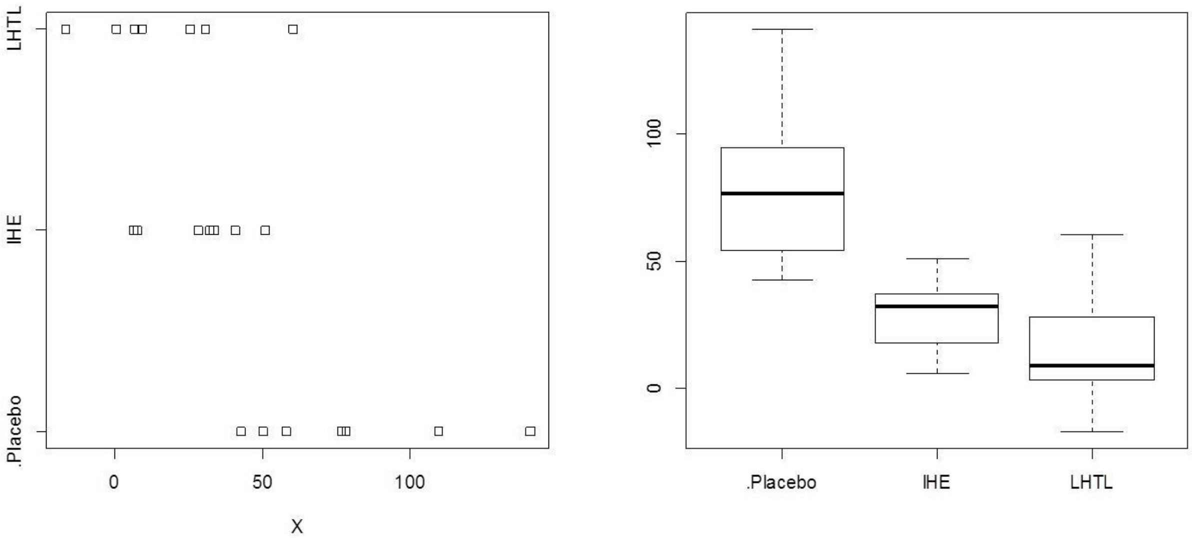


Fig 1. Exploratory analyses comprising stripcharts (left) and boxplots (right) for the covariate X in the three training regimens (Placebo, Intermittent Hypoxic Exposure (IHE), Live High Train Low (LHTL)), where X is a measure of the percent change in training load for each of the 23 individuals in the study. (See text for details.).

From Fig 5 and Table 1, it can be seen that for Hbmass and RunEcon, although there is a slight detrimental effect of IHE and a slight beneficial effect of LHTL compared with the Placebo, these are not substantive: a difference of 0 is reasonably well supported by the posterior distributions. However, this slight differential in response results between IHE and LTHL: a difference of 0 appears to have less support in the posterior densities; the 90% credible interval does not include 0 and the posterior probability that Cohen's *d* exceeds the SWC is 0.96 and 0.93 for Hbmass and RunEcon respectively. These outcomes strongly indicate that LHTL is substantively better than IHE for both of these outcome measures.

In contrast, for La-max, both IHE and LHTL show a substantive beneficial effect compared with the Placebo, with the corresponding 95% (and hence 90%) credible intervals excluding 0

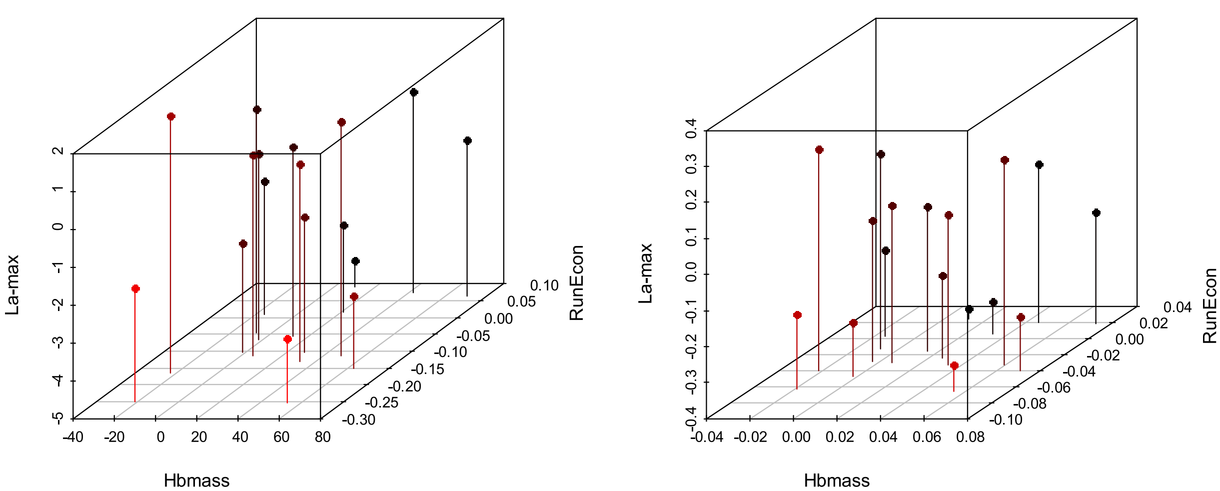


Fig 2. Three-dimensional scatterplot of the three measurements, Hemoglobin Mass (Hbmass), Running Economy (RunEcon) and maximum blood lactate concentration (La-max), unscaled data (left) and scaled data (right). Unscaled data are calculated as $post_i–pre$, and scaled data are calculated as $(post_i–pre_i) / pre_i$.

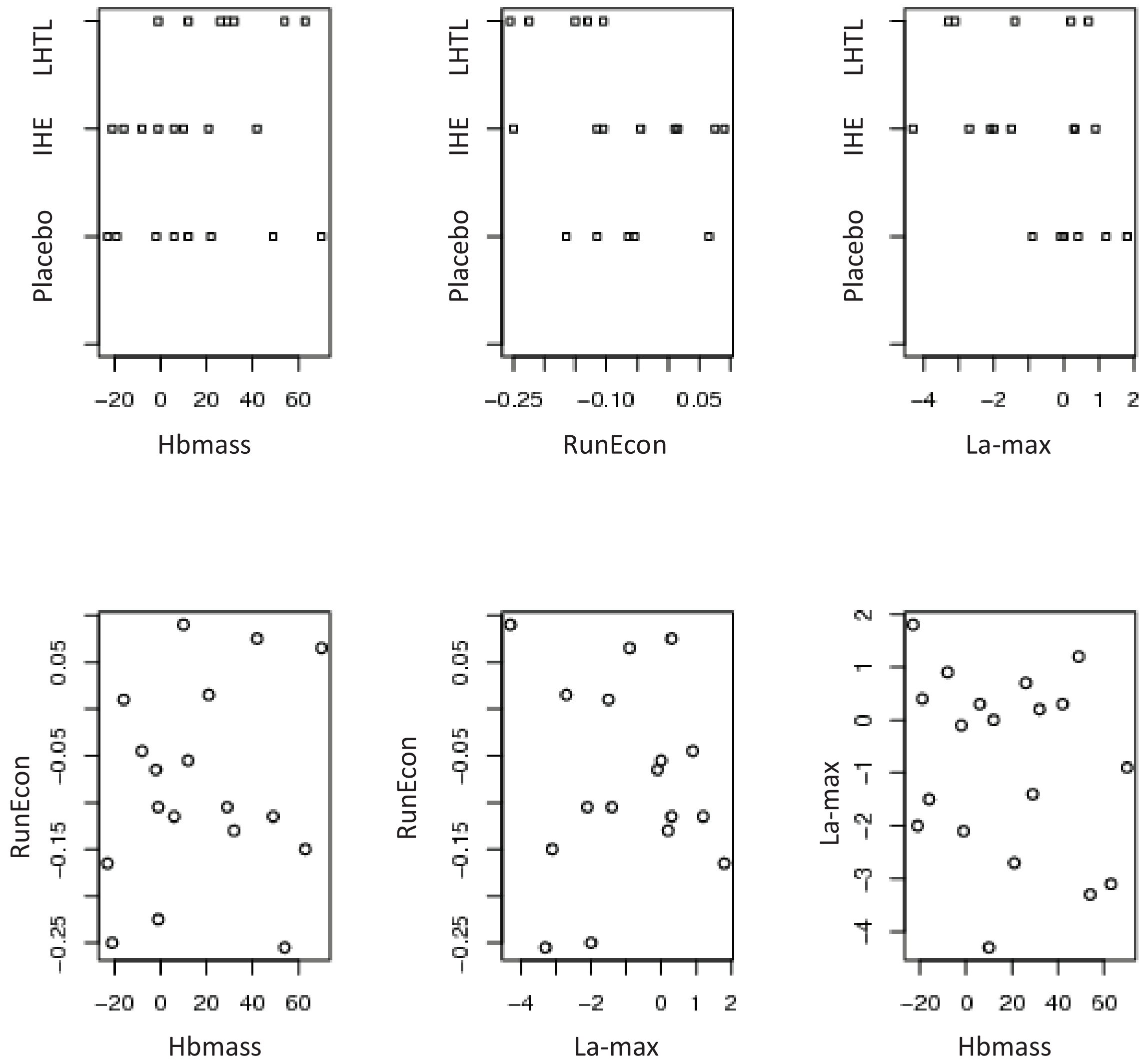


Fig 3. Two-dimensional scatterplots of the three measurements of Hbmass, RunEcon and La-max, under three regimes Placebo, Intermittent Hypoxic Exposure (IHE) and Live High Train Low (LHTL), unscaled data.

and a probability of 0.97 that Cohen's *d* exceeds the SWC. As a consequence, the difference between LHTL and IHE is thus attenuated for this outcome measure.

Posterior estimates of parameters of interest for the scaled (relative) measures are shown in Fig 6 and Table 2. The figures and table confirm the above results. Similar to the unscaled effects, there is no clear visual association between two of the measurements under consideration in this case study (Hbmass and RunEcon), or between these measurements and the covariate of change in weekly training load. However, there is a clear difference in the values of the covariate among individuals in the Placebo group compared with the two training regimens (LHTL and IHE). The two training regimens both appear to substantively improve La-max, even after accounting for training-induced changes in the individual athletes. The direct probabilistic comparisons with the SWC provide more complete information about these treatments based on these data.

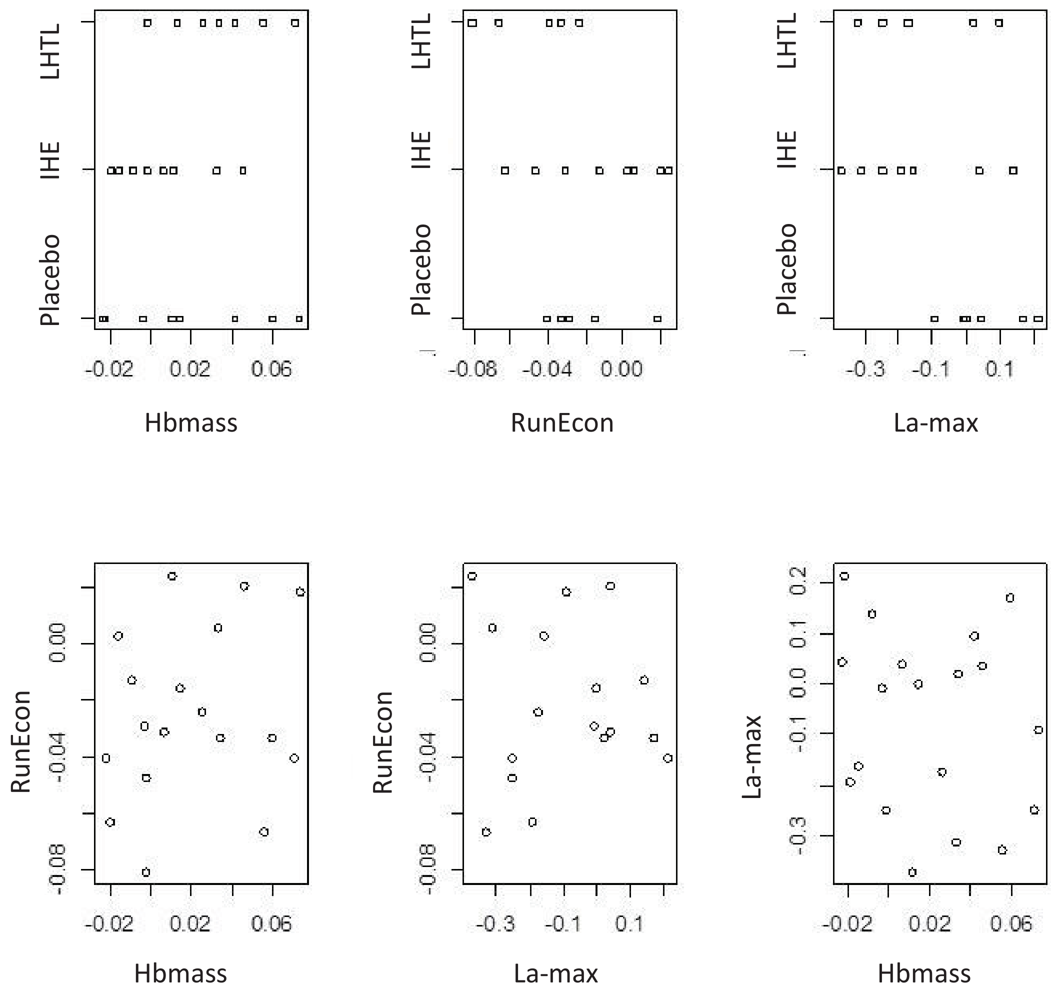


Fig 4. Two-dimensional scatterplots of the three measurements of Hbmass, RunEcon and La-max, under three regimes Placebo, Intermittent Hypoxic Exposure (IHE) and Live High Train Low (LHTL), scaled data.

The posterior expected outcome of Hbmass for each individual under each regimen is illustrated in Fig 7, for the unscaled data. The boxplots indicate the distribution of possible outcomes, with the box corresponding to the middle 50% of values and the limits of the bars corresponding to the minimum and maximum values. The corresponding expected rank and associated interquartile range for the 23 individuals are reported in Table 3. It is noted that the predictions and ranks are substantively driven by the covariate values in this model, with comparatively much less influence from the effect of the training regimens. Hence Table 3 displays only a selection of results.

A comparison of two of the primary outcome measures Hbmass and RunEcon based on the Bayesian and magnitude-based inference approach is presented in Table 4. Note that the two sets of results differ slightly not only because of differences in analytic method, but also because of differences in modelling. For example, the magnitude-based inferences are based on a log-transformed response forecast to a covariate value (a 44% increase in weekly training load), with covariate adjustment undertaken within each treatment group; in contrast, the Bayesian inferences are based on the unadjusted and relative responses forecast to the mean covariate

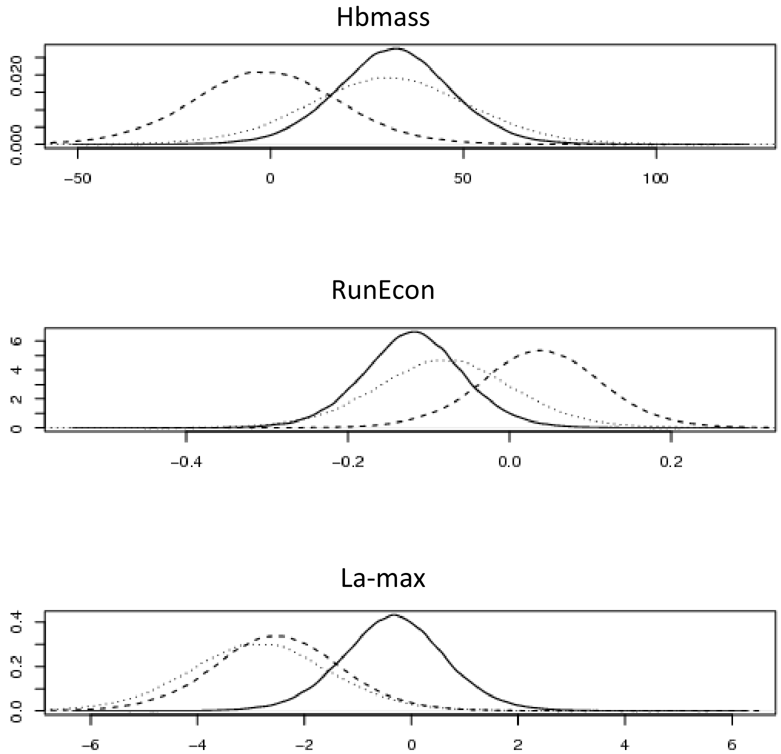


Fig 5. Posterior densities of the three measurements, Haemoglobin Mass, Running Economy and Running Maximum Lactate, comparing Live High Train Low (LHTL) vs Intermittent Hypoxic Exposure (IHE) (solid line), LHTL vs Placebo (dotted line) and IHE vs Placebo (dashed line), unscaled data.

value and adjustment is undertaken using all of the data for reasons of small sample size. Furthermore, as described above, the method of computation of the denominator of the standardized values is not based on asymptotics in the Bayesian analysis, which makes a difference for small samples. Notwithstanding these differences, the overall conclusions are similar for the two sets of analyses. For Hbmass, the Bayesian analysis indicated a substantially higher increase for LHTL with both unscaled and scaled data, with magnitude-based analysis indicating possibly higher for LHTL with unscaled data, and likely higher with scaled data. Similarly for RunEcon the outcomes were similar between the analytical approaches—the Bayesian analysis indicated a substantial improvement (lower oxygen cost) with both unscaled and scaled data, while magnitude-based analysis indicated possibly lower oxygen cost in both cases.

Comparison of the expected values of Hbmass and La-max under each of the training regimens is further illustrated in Fig 8. The diagonal line indicates no treatment effect. The cloud of points represents the values obtained from the MCMC simulations in the Bayesian analysis. Displacement of the cloud from the line indicates that that there is an expected improvement

Table 1. Posterior estimates based on unscaled data.

Hbmass				
Posterior parameter estimates (units of grams)				
Effect	Mean	s.d.	95% CI	90% CI
X	0.25	0.25	-0.25, 0.74	-0.17, 0.66
IHE	-1.4	19.8	-40.5, 37.8	-33.8, 30.9
LHTL	30.7	21.7	-12.4, 73.4	-4.9, 66.2
LHTL-IHE	32.0	15.3	1.9, 62.2	7.04, 57.2
Cohen's d				
Effect	Mean	s.d.	95% CI	90% CI
IHE	-0.07	1.0	-2.1, 1.9	-1.7. 1.6
LHTL	1.4	1.0	-0.57, 3.4	-0.23, 3.0
LHTL-IHE	2.1	1.0	0.12, 4.1	0.46, 3.7
Prob. Cohen's d <> 0.2				
Parameter	Prob. *d*<-0.2	Prob. *d*>0.2		
IHE	0.45	0.39		
LHTL	0.052	0.89		
LHTL-IHE	0.013	0.97		
RunEcon				
Posterior parameter estimates (unit of L/min)				
Effect	Mean	s.d.	95% CI	90% CI
X	0.00045	0.0010	-0.0016, 0.0025	-0.0012, 0.0021
IHE	0.039	0.079	-0.12, 0.20	-0.09, 0.17
LHTL	-0.080	0.090	-0.26, 0.097	-0.23, 0.065
LHTL-IHE	-0.12	0.064	-0.25, 0.0094	-0.22, -0.014
Cohen's d				
Effect	Mean	s.d.	95% CI	90% CI
IHE	0.50	1.0	-1.5, 2.5	-1.1, 2.1
LHTL	-0.89	1.0	-2.9, 1.1	-2.5, 0.73
LHTL-IHE	-1.85	1.0	-3.8, 0.15	-3.5, -0.22
Prob. Cohen's d <> 0.2				
Parameter	Prob. *d*<-0.2	Prob. *d*>0.2		
IHE	0.23	0.62		
LHTL	0.77	0.13		
LHTL-IHE	0.95	0.023		
La-max				
Posterior parameter estimates (units of mmol/L)				
Effect	Mean	s.d.	95% CI	90% CI
X	-0.018	0.015	-0.050, 0.013	-0.044, 0.0076
IHE	-2.5	1.3	-5.0, -0.06	-4.6, -0.50
LHTL	-2.8	1.3	-5.6, -0.10	-5.10-, -0.59
LHTL-IHE	-0.32	1.0	-2.3, 1.66	-1.2, 1.3
Cohen's d				
Effect	Mean	s.d.	95% CI	90% CI
IHE	-2.0	1.0	-4.0, -0.054	-3.7, -0.40
LHTL	-2.1	1.0	-4.0, -0.070	-3.7, -0.48
LHTL-IHE	-0.32	1.0	-2.3, 1.7	-2.0, 1.3

(*Continued*)

Table 1. (*Continued*)

Prob. Cohen's d <> 0.2		
Parameter	Prob. *d*<-0.2	Prob. *d*>0.2
IHE	0.97	0.015
LHTL	0.98	0.015
HTL-IHE	0.55	0.29

or decline in the outcome measure associated with the respective treatment, and the range of values for which this is anticipated to take effect.

The alternative priors that were motivated by the available external information are shown in Table 5. The consequent changes in the parameter values arising from the incorporation of these priors in the model are also shown in this table. It is clear that although the parameter estimates change slightly, the inferences reported above are generally robust to relatively small changes in the priors. However, the posterior estimates start to differ in a natural manner when the priors become more informative with respect to either the mean or variance. It is also noted that, reassuringly, the original (vague prior) setting yielded a posterior estimate of a relative increase of 2.6% in Hemoglobin mass under the LHTL regimen, which is equivalent to the anticipated value based on the (independent) prior information.

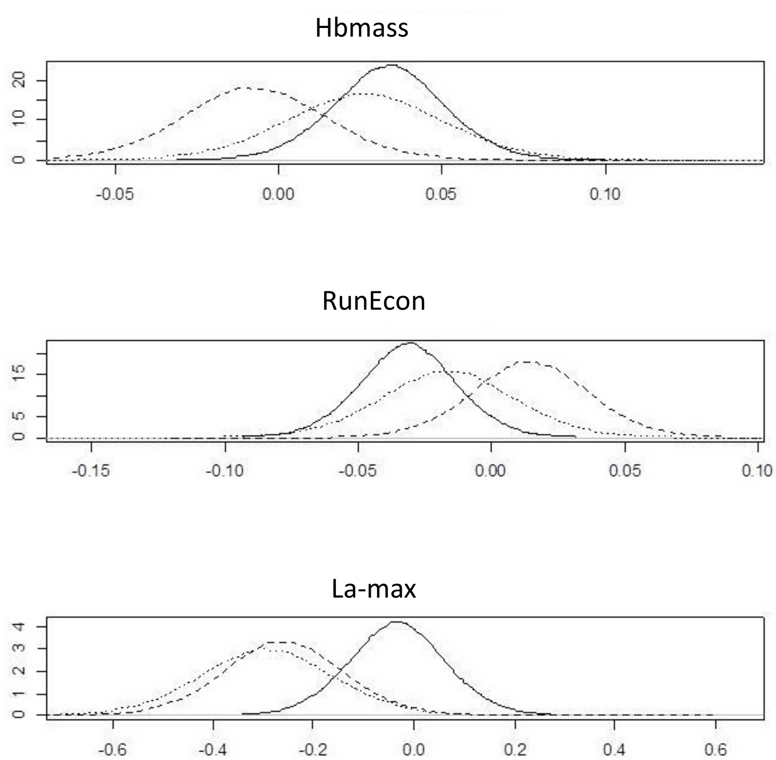


Fig 6. Posterior densities of the three measurements, Haemoglobin Mass, Running Economy and Running Maximum Lactate, comparing Live High Train Low (LHTL) vs Intermittent Hypoxic Exposure (IHE) (solid line), LHTL vs Placebo (dotted line) and IHE vs Placebo (dashed line), scaled data.

Table 2. Posterior estimates based on scaled data.

Hbmass				
Posterior parameter estimates (units of percent / 100)				
Effect	Mean	s.d.	95% CI	90% CI
X	0.00020	0.00029	-0.00038, 0.00077	-0.00028, 0.00067
IHE	-0.0075	0.023	-0.053, 0.038	-0.045, 0.030
LHTL	0.026	0.025	-0.023, 0.076	-0.015, 0.068
LHTL-IHE	0.034	0.018	-0.0011, 0.069	0.0050, 0.063
Cohen's d				
Effect	Mean	s.d.	95% CI	90% CI
IHE	-0.33	1.0	-2.3, 1.7	-1.2, 1.3
LHTL	1.1	1.0	-0.93, 3.0	-0.60, 2.7
LHTL-IHE	1.9	1.0	-0.059, 3.9	0.28, 3.6
Prob. Cohen's d <> 0.2				
Parameter	Prob. *d*<-0.2	Prob. *d*>0.2		
IHE	0.55	0.29		
LHTL	0.10	0.81		
LHTL-IHE	0.019	0.96		
RunEcon				
Posterior parameter estimates (units of percent / 100)				
Effect	Mean	s.d.	95% CI	90% CI
X	0.00023	0.00031	-0.00038, 0.00083	-0.00027, 0.00072
IHE	0.015	0.023	-0.032, 0.061	-0.024, 0.053
LHTL	-0.016	0.027	-0.069, 0.037	-0.060, 0.027
LHTL-IHE	-0.031	0.019	-0.069, 0.0071	-0.062, 0.00016
Cohen's d				
Effect	Mean	s.d.	95% CI	90% CI
IHE	0.63	1.00	-1.4, 2.6	-1.0, 2.3
LHTL	-0.61	1.00	-2.6, 1.4	-2.2, 1.0
LHTL-IHE	-1.62	1.00	-3.6, 0.37	-3.3, 0.0085
Prob. Cohen's d <> 0.2				
Parameter	Prob. *d*<-0.2	Prob. *d*>0.2		
IHE	0.19	0.68		
LHTL	0.67	0.20		
LHTL-IHE	0.93	0.035		
La-max				
Posterior parameter estimates (units of percent / 100)				
Effect	Mean	s.d.	95% CI	90% CI
X	-0.0019	0.0016	-0.0051, 0.0013	-0.0045, 0.00072
IHE	-0.26	0.13	-0.51, -0.0094	-0.47, -0.054
LHTL	-0.29	0.14	-0.58, -0.014	-0.52, -0.065
LHTL-IHE	-0.034	0.10	-0.24, 0.17	-0.20, 0.13
Cohen's d				
Effect	Mean	s.d.	95% CI	90% CI
IHE	-2.6	1.0	-4.0, -0.074	-3.7, -0.43
LHTL	-2.1	1.0	-4.1, -0.10	-3.7, -0.46
LHTL-IHE	-0.33	1.0	-2.3, 1.7	-2.0, 1.3

(*Continued*)

Table 2. (*Continued*)

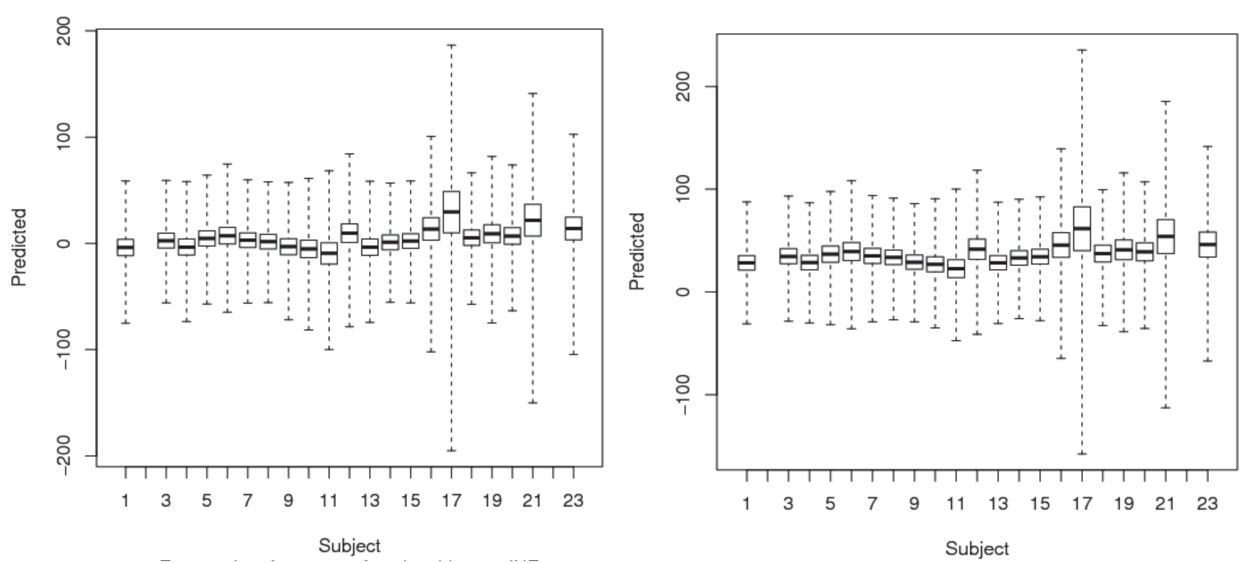
Prob. Cohen's d <> 0.2		
Parameter	Prob. *d*<-0.2	Prob. *d*>0.2
IHE	0.97	0.014
LHTL	0.97	0.014
LHTL-IHE	0.56	0.29

Discussion

In 2008, Barker and Schofield [7] suggested that "to correctly adopt the type of inference advocated by Batterham and Hopkins [6], sport scientists need to use fully Bayesian methods of analysis". They also noted that most sport scientists are not trained in Bayesian methods, likely because this approach has only become commonplace as a statistical technique in approximately the last 20 years. To help make the Bayesian approach more accessible for those working in exercise science and sports medicine, we have provided here both a worked example (using statistical software) together with a description of the underlying models. We hope that this template will encourage those who deal with small samples and small effects to explore the full Bayesian method, which is well suited to the analysis of small samples. Other supporting information, where available, can be represented via the prior and hence formally and transparently incorporated with the data. In the absence of such information, the uncertainty induced by small samples is properly incorporated in the posterior estimates and inferences. In both of these situations, the analytical decision-making is enhanced, in support of the ultimate practical/clinical decision-making undertaken by sports practitioners.

Case study re-interpreted with Bayesian inferences

An experimental study by Humberstone-Gough and colleagues reported changes (mean ± 90% confidence interval) in Hbmass of -1.4 ± 4.5% for IHE compared with Placebo and 3.2 ± 4.8%



Expected outcomes for Hbmass for each subject on IHE

Expected outcomes for Hbmass for each subject on LHTL

Fig 7. Boxplots of the posterior expected outcomes for Hbmass for each individual in the study, under each of the two training regimens Intermittent Hypoxic Exposure (left) and Live High Train Low (right).

Table 3. Expected rank and associated interquartile range for the 23 individuals in the study.

ID	Hbmass		RunEcon		La-max	
	Mean	IQR	Mean	IQR	Mean	IQR
1	3	3–19	3	3–19	19	3–19
2	NA	NA	NA	NA	NA	NA
3	10	10–12	10	10–12	12	10–12
4	5	5–17	5	5–17	17	5–17
5	12	10–12	12	10–12	10	10–12
6	15	7–15	15	7–15	7	7–15
7	11	11–11	11	11–11	11	11–11
8	8	8–14	8	8–14	14	8–14
9	6	6–16	6	6–16	16	6–16
10	2	2–20	2	2–20	20	2–20
11	1	1–21	1	1–21	21	1–21
12	17	5–17	17	5–17	5	5–17
13	4	4–18	4	4–18	18	4–18
14	7	7–15	7	7–15	15	7–15
15	9	9–13	9	9–13	13	9–13
16	18	4–18	18	4–18	4	4–18
17	21	1–21	21	1–21	1	1–21
18	13	9–13	13	9–13	9	9–13
19	16	6–16	16	6–16	6	6–16
20	14	8–14	14	8–14	8	8–14
21	20	2–20	20	2–20	2	2–20
22	NA	NA	NA	NA	NA	NA
23	19	3–19	19	3–19	3	3–19

doi:10.1371/journal.pone.0147311.t003

for LHTL compared with Placebo [9]. For RunEcon the authors reported 'no beneficial changes' for IHE compared with Placebo, and a change of 2.8 ± 4.4% for LHTL compared with Placebo. Although the analyses were undertaken using different outcome measures and a slightly different analytical model, the conclusions based on the posterior estimates and probabilities obtained from the Bayesian analysis reported above are broadly consistent with those reported by Humberstone-Gough *et al*. Importantly, the Bayesian approach allows a much more direct probabilistic interpretation of credible intervals and posterior probabilities; for example, the probability that the mean change in Hbmass after LHTL compared with the change after IHE is greater than the smallest worthwhile change (0.2) is 0.96.

Cohen's effect size magnitudes are well established [11] but the selection of a small effect ($d = 0.2$) as the threshold value for a worthwhile change or difference has been questioned. In the sporting context, worthwhile changes in competition performance, which can alter medal rankings, have been derived [25] as approximately 0.3 times the within-subject standard deviation [26, 27], or ~0.3–1% of performance time in a range of sports [28–30]. Empirical evidence confirms that small effects (on competitive performance) are worthwhile for elite athletes and of practical relevance for coaches and scientists attempting to understand the likely benefit or harm of training regimen, lifestyle intervention or change in technique. The full Bayesian approach provides a robust and acceptable method of estimating the likelihood of a small effect. For instance, in the Humberstone-Gough et al. case study Hbmass increased ~21 g (or by 2.3%) more in LHTL vs IHE. Given that every gram of hemoglobin can carry ~4 mL O_2, [31], it is

Table 4. Analysis of pre- to post-training measurements for LHTL vs IHE–outcomes for Bayesian and Magnitude-based Inferences for both unscaled and scaled data. SD = standard deviation, CL = confidence limits, CI = credible interval.

Analysis	Measure	Hemoglobin Mass (g)	Running Economy ($L.min^{-1}$)
Bayesian Unscaled	Mean ± SD	21 ± 17	-0.17 ± 0.052
	90% CI	-6, 48	-0.25, -0.08
	Cohen's d; 90% CI	1.26; -0.37, 2.90	-3.20; -4.84, -1.57
	Probability |d|>0.2	0.931	0.998
	Qualitative inference	Higher	Lower
Magnitude-based Inference	Mean; 90% CL	36; -5, 78	-0.13; -0.22, 0.04
	Cohen's d; 90% CL	0.18; -0.02, 0.39	-0.20; -0.34, -0.07
	Qualitative inference	Possibly Higher	Possibly Lower
Bayesian Scaled	Mean ± SD (% / 100)	0.023 ± 0.019	-0.042 ± 0.017
	90% CI	-0.008, 0.054	-0.069, -0.015
	Cohen's d; 90% CI	1.21; -0.42, 2.85	-2.51; -4.14, -0.88
	Probabilty |d|>0.2	0.926	0.993
	Qualitative inference	Higher	Lower
Magnitude-based Inference	Smallest worthwhile difference (% / 100)	0.016	0.019
	Difference ± SD	0.047 ± 0.035	-0.028 ± 0.044
	Cohen's d; 90% CL	0.20; 0.05, 0.35	-0.14; -0.34, 0.07
	Qualitative inference	Likely Higher	Possibly Lower

reasonable to infer that this small increase in Hbmass is likely beneficial to overall oxygen transport capacity. The corresponding 95% credible interval for this comparison of absolute change ranged from -11.8 to +53.8 g, but on balance the probability is >0.8 that the true increase in Hbmass is substantial (worthwhile), which should be sufficient encouragement for most

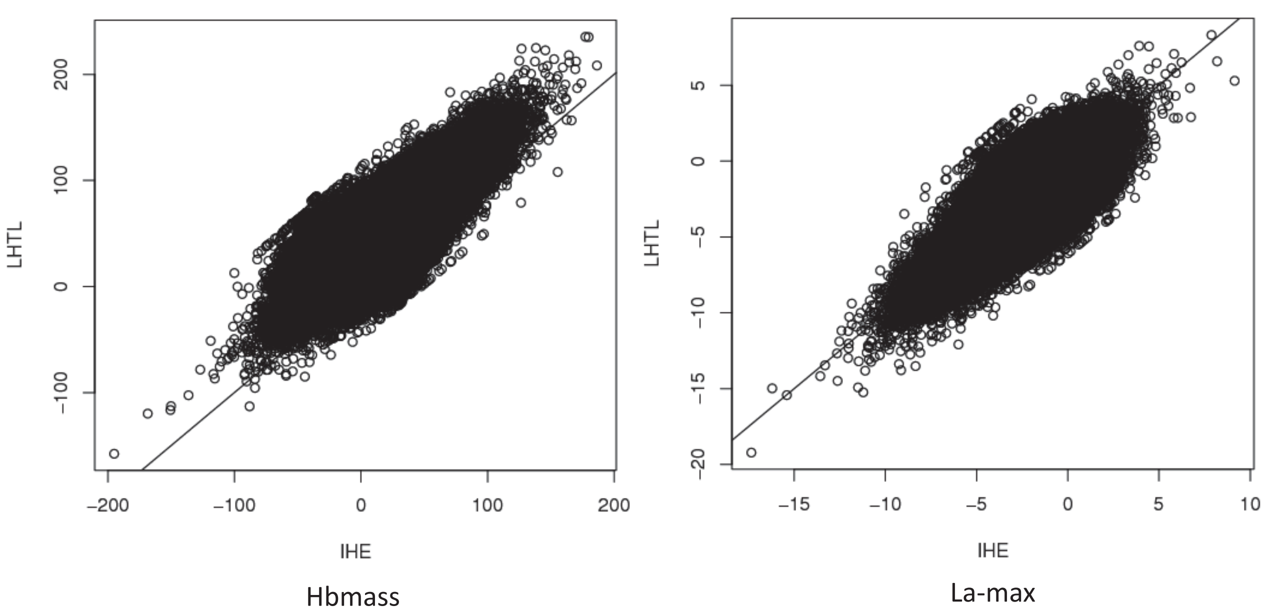


Fig 8. Comparison of the posterior distributions of the expected measurements of Hbmass (left) and La-max (right) under each of the training regimens Intermittent Hypoxic Exposure (IHE) and Live High Train Low (LHTL), unscaled data.

Table 5. Configurations of hyperparameter values for informative priors in the Bayesian model [Eqs (11 and 12)]. Here b_0 and B_0 denote respectively the prior mean vector and precision matrix for the regression coefficients, and $c_0/2$ and $d_0/2$ denote respectively the shape parameter and scale parameter for the inverse Gamma prior on σ^2 (the variance of the disturbances). These latter two parameters can be respectively interpreted as indicating the amount of information, and the sum of squared errors, from c_0 pseudo-observations, for the inverse Gamma prior on σ^2 (the variance of the residuals) [16]. Note that (a) depicts the baseline uninformative priors used in the primary analyses, whereas (b) to (h) illustrate seven alternate priors.

Setting	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
b_0	(0,0,0,0)	(0,0,0,2.6)	(0,0,0,2.6)	(0,0,0,2.6)	(0,0,0,0)	(0,0,0,0)	(0,0,0,0)	(0,0,0,2.6)
diag(B_0)	(0,0,0,0)	(0,0,.2,.2)	(5,5,5,5)	(0,0,5,5)	(0,0,0,0)	(0,0,0,0)	(0,0,5,5)	(0,0,0,0)
c_0	0.0001	0.0001	0.0001	0.0001	20	20	20	20
d_0	0.0001	0.0001	0.0001	0.0001	100	5	100	100
Int.	0.00047	0.0044	-0.0027	-0.0028	0.011	0.0060	-0.64	0.0060
	(0.026)	(0.026)	(0.026)	(0.026)	(1.4)	(0.30)	(0.28)	(0.30)
X	0.00020	0.00020	0.00026	0.00026	0.00018	0.00019	0.0062	0.00019
	(0.00029)	(0.00029)	(0.00029)	(0.00029)	(0.015)	(0.0034)	(0.0033)	(0.0034)
IHE	-0.0075	-0.0073	-0.0021	-0.0021	-0.017	-0.0096	0.41	-0.0096
	(0.023)	(0.023)	(0.023)	(0.023)	(1.2)	(0.27)	(0.23)	(0.27)
LHTL	0.026	0.027	0.035	0.035	0.024	0.026	0.79	0.26
	(0.025)	(0.025)	(0.025)	(0.025)	(1.3)	(0.29)	(0.27)	(0.29)
σ^2	0.0011	0.0011	0.0011	0.0011	2.9	0.14	0.17	0.14
	(0.00042)	(0.00042)	(0.00043)	(0.00043)	(0.71)	(0.035)	(0.047)	(0.045)

scientists and coaches to utilize altitude training to increase Hbmass–a position also supported by a meta-analysis of Hbmass and altitude training [23]. Likewise in the Humberstone-Gough et al. case study RunEcon improved (was lower) by ~0.17 $L.min^{-1}$ (or lower by 4.2%) more in LHTL vs IHE. The associated 95% credible interval for this comparison of relative change ranged from –0.9 to -7.5%, with a probability of ~0.99 that the true decrease in submaximal oxygen consumption is substantial (worthwhile). Although contentious [32], an improved running economy after altitude training is advantageous to distance running performance because it reduces the utilization of oxygen at any given steady-state running speed [33, 34].

Limitations of quasi-Bayesian approaches

Batterham and Hopkins (2006) have challenged the frequentist approach as being too conservative, and provided a useful, if somewhat unconventional, framework for interpreting small effects. The so-called magnitude-based approach emerging in sports science [18, 26, 35] is based on defining and justifying clinically, practically or mechanistically meaningful values of an effect. Confidence intervals are then used to interpret uncertainty in the effect in relation to these reference or threshold values. Much discussion has centred on the legitimacy of using vague priors in the magnitude-based approach and whether prior knowledge is actually useful in all cases [36]. There are inferential limitations to their approach [7, 8] which can be circumvented by using a full Bayesian approach that we have elaborated here.

A major criticism of the approach suggested by Batterham and Hopkins (2006) is that, contrary to the authors' claims, their method is not (even approximately) Bayesian and that a Bayesian formulation of their approach would indeed make prior assumptions about the distribution of the true parameter values. Barker and Schofield (2008) suggest that the underlying prior distribution would be uniform, which makes a clear assumption about the parameter values (that any parameter value in the defined range is equally likely) and which can be influenced by transformations of the parameter. As demonstrated in our paper, a Bayesian formulation of the problem considered by Batterham and Hopkins (2006) can quite easily be constructed, using a reference prior which is arguably vague (often referred to as the Jeffreys prior [10]). Moreover, there are clear and natural links between the frequentist distributions based on sampling theory and the Bayesian posterior distributions under these prior assumptions. The use of the reference prior for the estimation and comparison problem considered in this paper is well-founded, theoretically sound and very commonly employed [10]. As discussed in the Methods section, however, other priors can also be considered, particularly if there is other information available to complement the analysis.

Another criticism levelled at Batterham and Hopkins (2006) by Barker and Schofield (2008) is their choice and use of an expanded set of categories, based on a non-standard choice of the thresholds used to define the categories, the use of different thresholds for different problems (e.g., sometimes 0.025 and 0.975 instead of 0.05 and 0.95), and the descriptors used to label the categories, namely 'almost certainly not,. . .almost certainly'). However, while the expanded set of categories proposed by Batterham and Hopkins is not 'standard' in classical statistics, this does not mean that it is wrong, misleading or not useful. Indeed, such categorizations can be very useful *if* they are clearly justified, interpreted properly and provide additional decision support for clinical (or, in this case, sporting) interventions. Even in 'traditional' statistics, some statisticians suggest that a p-value less than 0.10 indicates 'substantive' evidence against the null hypothesis, while other statisticians would not counsel this. Similarly, although a p-value of 0.05 is almost overwhelmingly taken as the 'significance level', many statisticians strongly advise against its unconsidered use and suggest that other levels (such as 0.01 or 0.10) may be more appropriate for certain problems and desired inferences. A number of commentators in the sports science field have made similar observations [1, 37, 38]. The overwhelming advice is that the probabilities obtained as a result of statistical analysis must be useful in providing decision support for the problem at hand, and different probabilities can indeed be used if they are well justified, transparently reported and correctly interpreted.

The technical interpretation of a (frequentist) confidence interval is poorly understood by many practitioners. This has caused, and will continue to lead to, clumsy statements about the inferences that can be made on its basis. In contrast, an analogous Bayesian interval is directly interpretable: for example, a 95% credible interval indicates that the true parameter lies within this interval with an estimated probability of 0.95. Moreover, the analysis can be used to obtain other decision support statements such as a set of meaningful probabilities; for example, as demonstrated in the case study, one can obtain the probability that a particular parameter exceeds an objectively-derived threshold of clinical/practical/sporting interest. Of course, the particular decisions that are made on the basis of these probabilities remain the prerogative of the decision-maker. For example, the outcome of an intervention to improve athletic performance (e.g. a new experimental therapeutic treatment) may be classified as 'possible' in some cases (acceptable probability of improving performance, within minimal adverse effects, low cost, readily available, and legal in terms of anti-doping regulations), and hence lead to a decision of using, whereas in another context it may be deemed too risky (unacceptable risk of impairing performance, adverse effects on health and well-being, high cost and limited availability, and some uncertainty in meeting anti-doping regulations) and lead to no action. In practice, these decisions may not coincide with the traditional statement of a statistically significant effect at a 5% level [36]. In both cases, however, the decisions are enhanced by the richer probabilistic and inferential capability afforded by the Bayesian analysis.

In the context of small samples such as those encountered in this study, it is important to understand the nature and implications of the statistical assumptions underlying the adopted models and inferences. For example, in a standard linear regression model a common assumption is that the residuals (the differences between the observed and predicted values) are normally distributed. Note that this only applies to the residuals, not the explanatory or response variables. This assumption was also adopted in the model and analysis presented in this paper. There is a rich literature about the appropriateness of this assumption for small sample sizes. Importantly, if the residuals are indeed normally distributed then the regression estimates will possess all three desirable statistical characteristics of unbiasedness, consistency, and efficiency among all unbiased estimators; however, even if they are not normally distributed they will still be unbiased (accurate) and consistent (improve with increasing sample size) but will only be most efficient (i.e. have smallest variance) among a smaller class of (linear unbiased) estimators [39]. The most obvious implication of non-normal residuals is that the inferences may not be as sharp, but by virtue of the central limit theorem the sampling distribution of the coefficients will approach a normal distribution as the sample size increases, under mild conditions. In our study, this was achieved by employing a single residual term across all groups which effectively increased the sample size. Feasible alternatives would have been to allow different residual variances for each group or to employ a robust regression approach, for example using a *t* distribution for the errors. It is also noted that the Bayesian estimates avoid some of the inferential concerns, since the credible intervals and probabilistic rankings are obtained from the MCMC samples, i.e., from the posterior distributions themselves, as opposed to relying on stronger asymptotic assumptions that are required for frequentist inferences

Another topical issue that has substantive implications for small sample analysis is reproducibility [40]. Indeed, the very measure of reproducibility arguably faces similar challenges as those reported here, and a Bayesian approach is arguably preferable over measures based on p-values or confidence intervals [41–43]. See also a recent blog article that discusses this topic (http://alexanderetz.com/2015/08/30/the-bayesian-reproducibility-project/). The current debates are often conducted in the context of large samples, so the challenge is much greater for studies such as the one presented here. This is another topic for future research.

Conclusion

We have demonstrated that a Bayesian analysis can be undertaken for small scale athlete studies and can yield comparable, but more directly interpretable and theoretically justified probabilistic outcomes compared with the so-called magnitude-based (quasi-Bayesian) approach. The model described here is one of the simplest Bayesian formulations, and can be expanded as needed to address other issues. Analytical approaches for small sample studies using full Bayesian, quasi-Bayesian, and frequentist decisions must be well justified, reported transparently and interpreted correctly.

Supporting Information

S1 Table. Data used in the case study.
(DOCX)

S1 Text. R code used in the analysis of the case study.
(DOCX)

Author Contributions

Conceived and designed the experiments: KM CD CG. Performed the experiments: KM CD. Analyzed the data: KM CD CR DP CG. Contributed reagents/materials/analysis tools: KM CD. Wrote the paper: KM CD CR DP CG.

References

1. Atkinson G, Batterham AM, Hopkins WG. Sports performance research under the spotlight. Int J Sports Med. 2012; 33: 949. doi: 10.1055/s-0032-1327755 PMID: 23165647
2. Ploutz-Snyder RJ, Fiedler J, Feiveson AH. Justifying small-n research in scientifically amazing settings: challenging the notion that only "big-n" studies are worthwhile. J Appl Physiol. 2014; 116: 1251–2. doi: 10.1152/japplphysiol.01335.2013 PMID: 24408991
3. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Medicine. 2010; 8: 17. doi: 10.1186/1741-7015-8-17 PMID: 20307281
4. Bacchetti P, Deeks SG, McCune JM. Breaking free of sample size dogma to perform innovative translational research. Sci Transl Med. 2011; 3(87): 87sp24.
5. Beck TW. The importance of a priori sample size estimation in strength and conditioning research. J Str Cond Res. 2013; 27: 2323–37.
6. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. Int J Sports Physiol Perf. 2006; 1: 50–7.
7. Barker RJ, Schofield MR. Inference about magnitudes of effects. Int J Sports Physiol Perf. 2008; 3: 547–57.
8. Welsh AH, Knight EJ. "Magnitude-Based Inference": A Statistical Review. Med Sci Sports Exerc. 2014; 47: 874–84.
9. Humberstone-Gough C, Saunders PU, Bonetti DL, Stephens S, Bullock N, Anson JM et al. Comparison of live high: train low altitude and intermittent hypoxic exposure. J Sports Sci Med. 2013; 12: 394–401. PMID: 24149143
10. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. *Bayesian Data Analysis*. 3rd ed.: Chapman and Hall.; 2013, 64–9 p.
11. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988, 1–17 p.
12. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. *Bayesian Data Analysis*. 3rd ed.: Chapman and Hall; 2013, 275–92 p.
13. Geman S, Geman D. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984; 6: 721–41. PMID: 22499653
14. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Orlando: Academic Press; 1985.
15. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBugs—a Bayesian modelling framework: concepts, structure and extensibility. Stats Computing. 2000; 10: 325–37.
16. Marin J, Robert C. *Bayesian Core*: *A Practical Approach to Computational Bayesian Statistics*. Springer; 2007.
17. Marin J, Robert C. *Bayesian Essentials with R*. Springer; 2014.
18. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 2009; 41: 3–13. doi: 10.1249/MSS.0b013e31818cb278 PMID: 19092709
19. Garvican LA, Martin DT, McDonald W, Gore CJ. Seasonal variation of haemoglobin mass in internationally competitive female road cyclists. Eur J Appl Physiol. 2010; 109: 221–31. doi: 10.1007/s00421-009-1349-2 PMID: 20058020
20. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. J Stats Software. 2005; 12: 1–16.
21. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS Open. R News. 2006; 6: 12–7.
22. Martin A, Quinn K, Park J-H. MCMCpack: Markov chain Monte Carlo in R. J Stats Software. 2011; 429: 1–21.
23. Gore CJ, Sharpe K, Garvican-Lewis LA, Saunders PU, Humberstone CE, Robertson EY et al. Altitude training and haemoglobin mass from the optimised carbon monoxide rebreathing method determined by a meta-analysis. Br J Sports Med. 2013; 47: i31–9.
24. Gore CJ, Rodriguez FA, Truijens MJ, Townsend NE, Stray-Gundersen J, Levine BD. Increased serum erythropoietin but not red cell production afer 4 wk of intermittent hypobaric hypoxia (4,000–5,500 m). J Appl Physiol. 2006; 101: 1386–91. PMID: 16794028
25. Hopkins WG, Hawley JA, Burke LM. Design and analysis of research on sport performance enhancement. Med Sci Sports Exerc. 1999; 31: 472–85. PMID: 10188754
26. Hopkins W. How to interpret changes in an athletic performance test. Sportscience. 2004; 8: 1–7.

27. Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med. 2001; 31: 211–34. PMID: 11286357

28. Bonetti DL, Hopkins WG. Variation in performance times of elite flat-water canoeists from race to race. Int J Sports Physiol Perf. 2010; 5: 210–7.

29. Pyne DB, Trewin C, Hopkins WG. Progression and variability of competitive performance of Olympic swimmers. J Sports Sci. 2004; 22: 613–20. PMID: 15370491

30. Smith TB, Hopkins WG. Variability and predictability of finals times of elite rowers. Med Sci Sports Exerc. 2011; 43: 2155–60. doi: 10.1249/MSS.0b013e31821d3f8e PMID: 21502896

31. Schmidt W, Prommer N. Effects of various training modalities on blood volume. Scand J Med.Sci. Sports 2008; 18: 57–69. doi: 10.1111/j.1600-0838.2008.00833.x PMID: 18665953

32. Lundby C, Calbert JA, Sander M, van Hall G, Mazzeo RS, Stray-Gundersen J et al. Exercise economy does not change after acclimatization to moderate to very high altitude. Scand J Med.Sci.Sports. 2007; 17: 281–91. PMID: 17501869

33. Conley DL, Krahenbuhl GS. Running economy and distance running performance of highly trained athletes. Med Sci Sports Exerc. 1980; 12: 357–60. PMID: 7453514

34. Daniels JT. A physiologist's view of running economy. Med Sci Sports Exerc. 1985; 17: 332–8. PMID: 3894870

35. Wilkinson M. Distinguishing between statistical significance and practica/clinical meaningfulness using statistical inference. Sports Med. 2014; 44.

36. Burton PR, Gurrin LC, Campbell MJ. Clinical significance not statistical significance: a simple Bayesian alternative to p values. J Epidemiol Comm Health. 1998; 52: 318–23.

37. Stang A, Poole C, Kuss O. The ongoing tyranny of statistical significance testing in biomedical research. Eur J Epidemiol. 2010; 25: 225–30. doi: 10.1007/s10654-010-9440-x PMID: 20339903

38. Stapleton C, S M.A., Atkinson G. The 'so what' factor: statistical versus clinical significance. Int J Sports Med. 2012; 30: 773–4.

39. Williams MN, Grajales GAG, Kurkiewiez D. Assumptions of multiple regression: correcting two misconceptions. Practical Assessment, Research and Evaluation. 2013; 18: 1–14.

40. Open_Science_Collaboration. Estimating the reproducibility of psychological science. Science. 2015; 349: 6251: aac4716.

41. Dienes Z. Using Bayes to get the most out of non-significant results. Front. Psych. 2014; 5: 781.

42. Gelman A, Stern H. The difference between "significant" and "not significant" is not itself statistically significant. Am Stat. 2006; 60: 328–31.

43. Verhagen J, Wagenmakers EJ. Bayesian tests to quantify the result of a replication attempt. J Exp Psych Gen. 2014; 143: 1457–75.

3

Repeatability and predictive value of lactate threshold concepts in endurance sports

Jules A. A. C. Heuberger[1] *, **Pim Gal**[1], **Frederik E. Stuurman**[1], **Wouter A. S. de Muinck Keizer**[1,2], **Yuri Mejia Miranda**[1], **Adam F. Cohen**[1,3]

1 Centre for Human Drug Research, Leiden, the Netherlands, **2** Free University of Amsterdam, Amsterdam, the Netherlands, **3** Department of Internal Medicine, Leiden University Medical Centre, Leiden, the Netherlands

* jheuberger@chdr.nl

Abstract

Introduction

Blood lactate concentration rises exponentially during graded exercise when muscles produce more lactate than the body can remove, and the blood lactate-related thresholds are parameters based on this curve used to evaluate performance level and help athletes optimize training. Many different concepts of describing such a threshold have been published. This study aims to compare concepts for their repeatability and predictive properties of endurance performance.

Methods

Forty-eight well-trained male cyclists aged 18–50 performed 5 maximal graded exercise tests each separated by two weeks. Blood lactate-related thresholds were calculated using eight different representative concepts. Repeatability of each concept was assessed using Cronbach's alpha and intra-subject CV and predictive value with 45 minute time trial tests and a road race to the top of Mont Ventoux was evaluated using Pearson correlations.

Results

Repeatability of all concepts was good to excellent (Cronbach's alpha of 0.89–0.96), intra-subject CVs were low with 3.4–8.1%. Predictive value for performance in the time trial tests and road race showed significant correlations ranging from 0.65–0.94 and 0.53–0.76, respectively.

Conclusion

All evaluated concepts performed adequate, but there were differences between concepts. One concept had both the highest repeatability and the highest predictability of cycling performance, and is therefore recommended to be used: the Dmax modified method. As an easier to apply alternative, the lactate threshold with a fixed value of 4 mmol/L could be used as it performed almost as well.

Editor: Laurent Mourot, University of Bourgogne France Comté, FRANCE

Competing interests: The authors have declared that no competing interests exist.

Trial registration

Dutch Trial Registry NTR5643

Introduction

The measurement of blood lactate is extensively used in sports medicine, although there is debate on how lactate affects fatigue in endurance athletes. [1] Nevertheless, the concentration of lactate in the blood relative to the exercise intensity is a relevant marker of endurance performance. [2–5] This can be visualized in a blood lactate curve (BLC) using a maximal graded exercise test (GXT): as the workload on the athlete increases over time, blood lactate concentrations (bLa) are measured at defined intervals. During high intensity contractions lactate is formed along with H^+ in the muscles, [6] followed by an increased elimination of lactate from plasma. [7, 8] When elimination becomes saturated, bLa will start to rise when production exceeds clearance. This (exponential) rise in bLa in the BLC is of importance, as the corresponding exercise intensity is associated with endurance performance since it correlates with the transition from aerobic to anaerobic workout. [9] Since the 1960's BLCs have been analysed trying to accurately determine a point in this curve that predicts endurance performance. Although many terms have been used for this point, in this work they will be termed lactate threshold (LT) concepts. BLCs and LT concepts can be used to assess 'endurance fitness' in athletes, [10] and to evaluate the effects of and to prescribe training exercises for individual athletes. [4, 5] Therefore these measures are relevant in sports medicine, both in amateur and professional sports. But as LT is based on a maximal exercise test protocol that does not directly mimic endurance exercise, finding a single point in the resulting BLC that has a strong relation to endurance performance is challenging. Moreover, determining where this single point lies in the relatively smooth curve, that is the result of a complex system of factors, can prove difficult as well. On the other hand, the more accurate method of determining maximum lactate steady state (MLSS), using several sessions with different workloads takes more time, which is the reason why an approximation of MLSS using lactate threshold concepts was developed. [11]

A previous literature review showed that there are many methods used to analyse the BLCs, with approximately 25 different concepts identified in literature to describe some form of LT. [9] These different concepts are used interchangeably throughout scientific studies and in sports and show variable repeatability and predictive value. Moreover, populations that were included in different studies often differed in training status, age and category of sport. For these reasons there is debate about these LT concepts. [9] The aim of this study is to evaluate the repeatability and predictive value of representative concepts using a large dataset of BLCs from a group of well-trained cyclists who performed multiple GXTs, time trials and an uphill road race in the setting of a clinical study.

Materials and methods

Study design and participants

Blood lactate curve data in this paper were generated in a previously published study. [12] Briefly, the study was a double-blind, randomized, placebo controlled, parallel, single centre trial to evaluate the effects of recombinant human erythropoietin (rHuEPO) in forty-eight healthy male cyclists aged 18 to 50. Informed consent was obtained from all individual participants included in the study. The study was approved by the Independent Ethics Committee of the Foundation Evaluation of Ethics in Biomedical Research (Stichting Beoordeling Ethiek Biomedisch Onderzoek, Assen, Netherlands). The study is registered in the Dutch Trial Registry (Nederlands Trial Register), number NTR5643. For inclusion, participants had to be well-trained, as evaluated by a maximum power-to-weight ratio during the GXT at screening that should exceed 4 W/kg. During the eleven week study duration, twenty-four participants received weekly rHuEPO injections and twenty-four received placebo injections for eight weeks. Participants had to maintain their regular training schedule during the study.

Procedures

Maximal exercise tests. Five GXTs were performed on a Monark LC4r ergometer (COSMED, Rome, Italy) with approximately 2-week intervals between each test, see Fig 1. After a two-minute warm-up at 75 Watts, the GXT dictated an increase in pedalling resistance to 175 Watts, which increased by an additional 25 Watts every five minutes. Between 4:15 and 4:45 into each step and immediately after termination of the exercise test, blood was drawn to measure bLa. Gas exchange was measured using a Quark CPET system (COSMED, Rome, Italy) and breath-by-breath sampling technology. During the test cadence had to be maintained between 70 and 90 rpm. The test terminated when cadence could not be maintained above 70 rpm or when a participant stopped the test.

Lactate determination. During the GXTs blood for lactate determination was drawn via an IV cannula (Venflon 7 Pro Safety, BD, Switzerland) with a 30 cm extension set between the cannula and a three way stopcock for blood sampling in the antecubital vein. Before the first and after every sampling the stopcock and extension set were flushed with 2 mL saline. Before blood sampling 0.5 mL was withdrawn from the stopcock to remove any remaining saline. Next, 1 mL of blood was taken from the stopcock. Within ten seconds from withdrawal the blood was placed on the Lactate Pro 2 (Arkray, Kyoto, Japan) strip which was then inserted in the Lactate Pro 2 device. The same device was used throughout the whole study and was given at least 20 minutes to adjust to the room temperature before sampling.

Time trial tests. The time trial tests were performed twice on the same ergometer used for the GXT, with the first (TT1) 3–8 days after the first GXT and the second (TT2) one week after GXT four. Participants were instructed to produce the highest mean power output during a 45-minute period at a cadence of 70–90 rpm, attempting to mimic competitive cycling time trials. At the start of the test pedalling resistance was set at 80% of the maximal power reached during GXT1. Participants could adjust the power by indicating to increase or decrease in power by steps of 10 Watts. They were informed of the remaining time on a regular basis during the test.

Mont Ventoux race. Approximately one week after the last GXT participants competitively climbed the Mont Ventoux (Vaucluse département, France) via Bédoin, a climb of approximately 21.5 km with an average gradient of 7.5%. The race was preceded by a stage of 110 km in the French Provence (total elevation gain 1524 m) that was completed collectively.

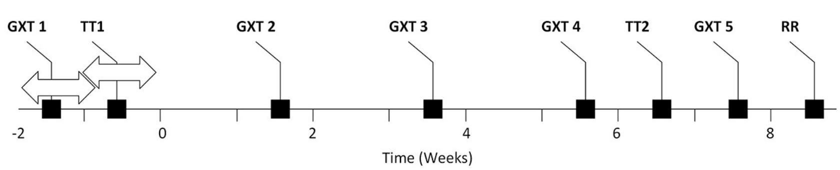


Fig 1. Study design. Study design showing timing of different tests. Time point 0 weeks indicates start of treatment (rHuEPO or placebo) for all participants. GXT, graded exercise test; TT, time trial test; RR, road race.

Racing bikes of participants were equipped with a Single Leg Power Meter SGY-PM910H2 (Pioneer Europe, Antwerp, Belgium) with Shimano Ultegra 6800 crank (Shimano, Osaka, Japan) to log power data on the bicycle during the race. Data were uploaded to the dedicated database Cyclo-Sphere.

Lactate threshold concepts. The BLCs from the GXTs were then used to calculate several representative LT concepts. Concepts were selected as follows: First, published concepts were retrieved from a review by Faude *et al.* [9] and by a literature research within the PubMed database. The database was searched for the search terms 'lactate threshold', 'aerobic threshold', 'anaerobic threshold', 'endurance performance' or 'maximal lactate steady state' or similar terms in different combinations. The references of the selected articles were searched for further relevant articles. Secondly, retrieved concepts were divided into seven different categories, see S1 Table. A few retrieved concepts could not be implemented, reasons being lacking lactate concentrations in the recovery phase after exercise and no availability of the full text article describing the method of the concept despite various efforts obtaining it. (S1 Table, listed under "not selected categories"). From each remaining category, concepts that were representative and were used frequently in other research were selected. If there were multiple concepts in one category that were commonly used and fundamentally different in methodology, more than one concept of that category was included in the analysis. Selecting multiple commonly used, but very similar concepts from one category was not deemed useful for the purpose of this study. This resulted in a final selection of eight concepts from the five implementable categories for analysis in our study.

Implementation of lactate threshold concepts. All selected concepts were implemented according to the articles that described the concept (S1 Table). When exact reproduction of the method was not feasible due to the use of different parameters (e.g. running velocity was used), we approximated the description as close as possible (e.g. we used power output). For concepts that required data fitting of the blood lactate curve a third-order polynomial was chosen, based on the shape of the blood lactate curve data and given that it is a proven method, although there is no generally accepted method for data fitting. [9] An example of a blood lactate curve with a depiction of all lactate threshold concepts is shown in Fig 2.

LT1. Similar to what Tanaka described [13] we plotted bLa (mmol/L) *versus* power (W). Three authors (JH, WdMK and PG) were asked to independently select the first point in the BLC that marks a substantial increase above resting level. LT1 was defined as the power value corresponding to the point selected by at least two researchers, or in cases without consensus, the three researchers discussed until consensus was reached.

LT2. Coyle *et al.* [14] determined LT as 1 mmol/L above a visually determined baseline in the BLC. We took the lactate measurement chosen as LT1 and calculated the mean of the measurements preceding this point to create an average baseline value. The power value belonging to the first measured lactate value after baseline that supersedes the baseline value plus 1 mmol/L was considered LT2.

LT3. As Dickhuth *et al.*, [15] we determined the minimum lactate equivalent (the lowest value when bLa is divided by work intensity) using third-order polynomial fitting and added 1.5 mmol/L to the corresponding bLa, termed individual anaerobic threshold in the paper, to find the power value on the fitted polynomial of the BLC and termed it LT3.

LT4. As described by Amann *et al.*, [16] we calculated the first rise of 1 mmol/L or more between two bLa measurements where the next rise was similar or larger than 1 mmol/L. The measurement that preceded this first increase was considered LT4.

LT5. Based on the method described by Dickhuth *et al.*, [17] we divided bLa (mmol/L) by the 30 second average VO_2 (mL/min/kg) and plotted it against power. These values were

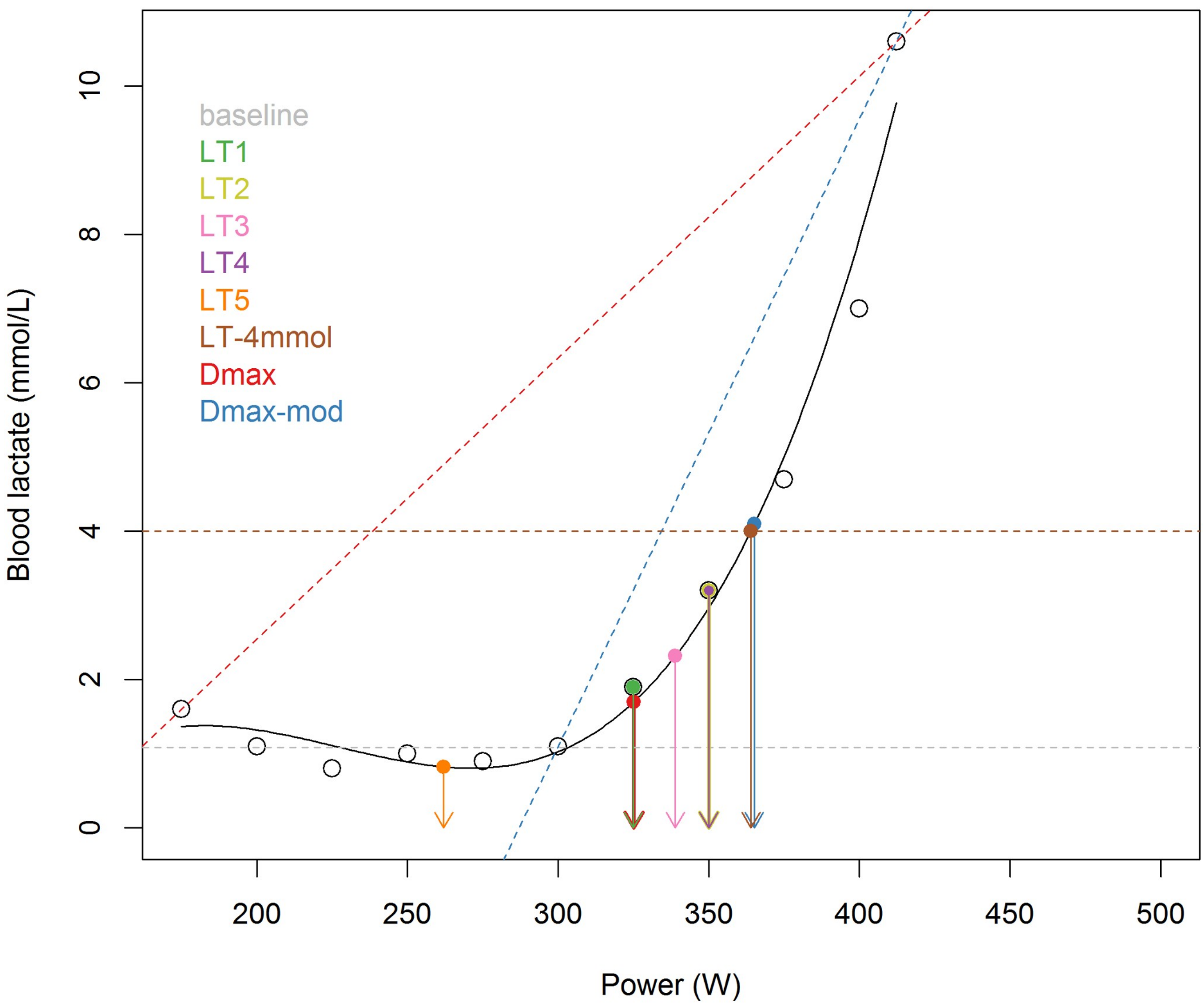


Fig 2. Graphical representation of lactate threshold concepts. Example of a blood lactate curve with the location of the different lactate threshold concepts for this particular curve. Open circles: observed blood lactate values at each exercise intensity; Black curve: third-order polynomial; Grey dashed line: baseline; Green circle and arrow: LT1, observer-determined first rise in blood lactate; Yellow circle and arrow: LT2, first observed blood lactate value more than 1 mmol/L above baseline; Pink circle and arrow: LT3, minimum lactate equivalent (blood lactate divided by power) plus 1.5 mmol/L; Purple circle and arrow: LT4, first blood lactate value that shows an increase of at least 1 mmol/L; Orange circle and arrow: LT5, minimum lactate equivalent (blood lactate divided by VO2); Brown circle and arrow and dashed line: LT-4mmol, value at 4 mmol/L; Red circle and arrow and dashed line: Dmax, value with the maximum perpendicular distance to the polynomial from the dashed line; Blue circle and arrow and dashed line: Dmax-mod, value with the maximum perpendicular distance to the polynomial from the dashed line.

interpolated with a third-order polynomial and the power value at the lowest point in this curve was considered LT5.

LT-4mmol. A widely used concept is the LT-4mmol method, as described for example by Sjodin *et al.* [18] The power in the interpolated third-order polynomial BLC that corresponds to a bLa of 4 mmol/L was considered LT-4mmol.

Dmax and Dmax modified. Similar to the method proposed by Cheng *et al.*, [19] we plotted bLa versus power, interpolated with a third-order polynomial and plotted a line from the first measurement to the last measurement. The point in the interpolated BLC that has the maximum perpendicular distance with that line was considered Dmax. A modified version as described by Bishop *et al.*, [20] uses the measurement that precedes an increase of at least 0.4 mmol/L instead of the first bLa measurement to draw the line to the last measurement, which is termed Dmax modified (Dmax-mod).

Data management

Data was stored in a validated database system (Promasys, Omnicomm Inc., Fort Lauderdale, USA) and checked for accuracy and completeness. Blinded data review before code-breaking and analysis was performed according to a standard procedure at our unit. This included evaluating whether the GXT was performed to maximal ability, which was based on power, VO_2 and bLa values and report by the subject.

Statistical analysis

We used statistical software R version 3.4.0 [21] to plot measurements, calculate the third-order polynomial that best fits the data using polynomial regression with the R-function lm (y~poly(3)), implement the LT concepts and perform the statistical testing. R was used with the following packages: dplyr 0.5.0, [22] psych 1.7.5, [23] tidyr 0.6.3. [24] Data of all subjects enrolled in the study were used in the analysis.

Repeatability. To measure repeatability we determined the weighted intra-subject coefficient of variation (CV) and the Cronbach's alpha based on the five GXT results for each LT concept. Weighted intra-subject CV was calculated correcting for missing values (CV based on the sum of the variance per subject multiplied by the amount of measurements, divided by the total amount of measurements). Both the weighted intra-subject CV and Cronbach's alpha were calculated only using data from participants receiving placebo, as there might have been longitudinal effects of rHuEPO treatment on the GXTs.

Predictive properties. For the predictive properties we calculated the Pearson correlation between each LT concept and the mean power of the corresponding relevant endurance parameter. The LT concept from the GXT closest in time to the endurance tests TT1 and TT2 and road race (see Fig 1), namely GXT 1, 4 and 5 respectively, were used for correlations between the LT concept and corresponding average power output. In addition, the difference between each measurement pair was calculated and averaged to create the mean difference between the LT concept and endurance test power. This value indicates how the power at the LT concept translates to average endurance power in a time trial or race. For these Pearson correlation and mean difference analyses both subjects receiving rHuEPO and placebo were included. This was done as LT concepts are designed to be a predictive parameter for endurance exercise, which should be irrespective of a subject being treated with rHuEPO or not. In addition, given that the measurements of each pair are at most a week apart, no changes in the LT concept or endurance performance are expected due to rHuEPO. Moreover, GXT1 and TT1 were performed before starting the treatment period, and no rHuEPO administrations took place between GXT5 and the race. For these analyses therefore no treatment effect was expected and pooling was considered appropriate.

Results

In total 49 subjects entered the study, of which 47 were completers (Fig 3); one subject dropped out after having performed the first GXT and time trial test and was replaced. One other subject dropped out after completing two GXTs and one time trial test and was not replaced. Subject characteristics can be found in Table 1. Of the remaining 238 planned GXTs, five were not performed due to illness or injury. An additional 22 were excluded from analysis, five due to having less than four bLa samples for the GXT, most others due to the fact that subjects

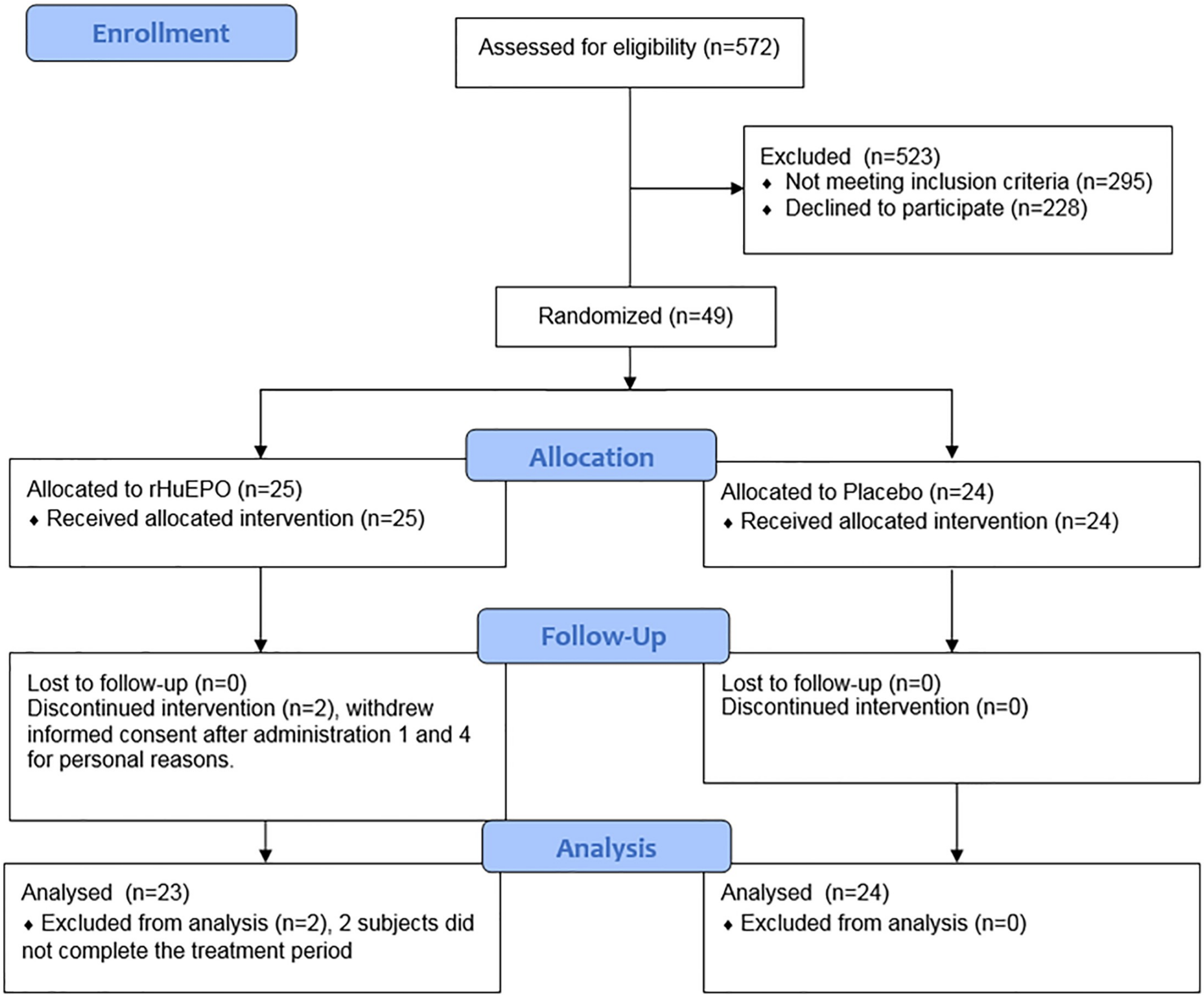


Fig 3. CONSORT flowchart.

Table 1. Subject disposition.

	All subjects	Placebo subjects
N	48	24
Age (years)	33.6 (20.0–50.0)	33.8 (20.0–50.0)
Weight (kg)	76.9 (9.0; 59.2–95.6)	76.9 (8.9; 59.2–95.6)
Height (cm)	186 (7.3; 172–203)	186 (6.7; 174–203)
Maximal Power output per kg (W/kg)	4.36 (4.03–5.18)	4.36 (4.03–4.94)
$VO_{2,max}$ (mL/min/kg)	55.7 (4.6; 45.3–67.5)	56.0 (4.1; 47.0–62.8)

Values are presented as mean (standard deviation (SD) where appropriate; range where). $VO_{2,max}$: maximal oxygen consumption.

indicated having physical problems (e.g. illness/injury, sore legs from recent exercise) potentially affecting test results, leaving 211 GXTs (of which 109 from placebo subjects) with analysable lactate threshold data. A total of 96 time trial tests were performed and used in the analysis, and power data of 37 subjects was available for the road race. Out of the 47 subjects that completed the study, three could not participate in the road race, four did not reach the finish line due to exhaustion, and three did not have a power meter on their bike, therefore lacking power data for the road race.

Lactate threshold concepts and endurance

All eight LT concepts were successfully implemented on the GXT data; for LT1 which was determined visually by three researchers, a unanimous decision about the lactate threshold was reached in 56.8% of the tests, in 40.0% of the cases two out of three researchers agreed and there was originally no consensus in 3.2% of the tests. Several concepts were based on the third-order polynomial data fitting, mean r-squared values of all individual curves were 0.978 (SD = 0.032, range 0.716–1.000). Mean values for each LT concept of the placebo group can be found in Table 2. Mean (SD) power output for TT1 was 268 W (28 W) in the placebo group and 271 W (29 W) in the rHuEPO group, and estimated mean for TT2 was 277 W and 283 W for the placebo and rHuEPO groups respectively. Estimated mean power during RR were 266 W and 257 W for the two groups, during a mean race time of 1 h 37 min 45 s (SD = 12 min 40 s) and 1 h 38 min 23 s (SD = 14 min 9 s), respectively.

Repeatability

The overall intra-subject CV of each LT concept is indicated in Table 2, and shows some minor differences between concepts, with LT3, LT-4mmol, Dmax and Dmax-mod having CVs $< 5\%$ and LT5 having the highest intra-subject CV with 8.1%. The Cronbach's alpha values for all LT concepts in the placebo group are between 0.89 and 0.97 and although 95% CIs largely overlap, the same four concepts as observed for intra-subject CVs perform best with Cronbach's alpha values >0.95 (Table 3).

Predictive properties

Pearson correlation coefficients and the mean difference between each correlation pair are listed in Table 4. All correlations are highly significant ($p<0.0002$), indicating the null

Table 2. Mean lactate threshold concept power output.

GXT number	LT1 (W)	LT2 (W)	LT3 (W)	LT4 (W)	LT5 (W)	LT-4mmol (W)	Dmax (W)	Dmax-mod (W)
1	283.3 (29.9; 225–350)	292.9 (37.2; 250–375)	286.1 (32.9; 219–352)	275.0 (41.1; 200–350)	225.0 (31.2; 175–283)	301.8 (41.0; 222–381)	275.7 (24.6; 222–323)	299.5 (35.3; 225–367)
2	283.0 (22.3; 225–375)	293.2 (31.0; 250–375)	288.7 (29.2; 231–373)	276.1 (34.0; 175–375)	231.8 (25.7; 175–300)	305.0 (33.0; 234–389)	280.0 (23.6; 233–339)	301.2 (28.7; 237–369)
3	281.0 (29.5; 225–400)	290.5 (27.9; 250–400)	285.7 (26.3; 240–390)	272.6 (33.5; 225–375)	224.5 (29.5; 175–318)	300.8 (28.9; 253–411)	278.7 (20.9; 250–343)	297.5 (29.6; 257–413)
4	283.7 (35.8; 225–400)	292.4 (38.0; 225–425)	291.6 (29.1; 240–392)	272.8 (36.9; 200–400)	229.8 (29.5; 175–323)	307.0 (34.2; 249–415)	284.0 (22.7; 232–338)	308.7 (35.6; 251–396)
5	278.3 (28.5; 225–325)	290.2 (37.5; 225–350)	285.0 (31.6; 216–339)	271.7 (37.9; 200–350)	230.4 (30.5; 175–274)	297.2 (38.9; 204–364)	280.3 (20.3; 245–325)	298.9 (29.2; 253–365)
Overall	282.1 (5.7%)	292.2 (5.0%)	287.7 (3.6%)	274.1 (5.6%)	228.4 (8.1%)	302.7 (3.8%)	280.0 (3.4%)	301.5 (4.3%)

Weighted mean power output (SD; range) for the placebo group at every exercise test. Overall combined (based on 109 GXTs) for each lactate threshold concept (CV). CV is weighted intra-subject CV.

Table 3. Cronbach's alpha for each lactate threshold concept.

Lactate threshold concept	Cronbach's alpha	Lower 95% CI	Upper 95% CI
LT1	0.91	0.85	0.96
LT2	0.95	0.92	0.98
LT3	0.97	0.94	0.99
LT4	0.94	0.91	0.98
LT5	0.89	0.82	0.96
LT 4_mmol	0.96	0.94	0.99
Dmax	0.96	0.93	0.98
Dmax-mod	0.96	0.94	0.98

Cronbach's alpha for the placebo group for each lactate threshold concept with 95% confidence interval (CI).

hypothesis that the correlation is equal to zero can be rejected. The strength of the relationship differs for different concepts. Correlation with TT1 was very strong for Dmax-mod and strong for all other concepts except LT5, which showed a moderate correlation. Correlation with TT2 was strong for all concepts except LT5, which showed a moderate correlation. Correlation with RR was strong for Dmax and Dmax-mod, and moderate for all other concepts. Dmax-mod has the highest correlation with time trial test 1 (r = 0.94), LT-4mmol with time trial test 2 (r = 0.85) and Dmax-mod with road race power (r = 0.76). The mean difference with the endurance parameters differs substantially between concepts, ranging from the lactate threshold on average being up to 45.3 W lower than the related endurance parameter for LT5 to 36.6 W higher for LT-4mmol. Linear regression between each LT concept and average race power, including accompanying r^2 values, is plotted in Fig 4.

Discussion

All LT concepts that were included in this analysis performed good on repeatability and reasonable to good on predicting a lab-based time trial and a real-life road race. Nevertheless, this

Table 4. Predictive value of lactate threshold concepts.

Lactate threshold concept	Pearson correlation			Mean difference (SD)		
	TT1	*TT2*	*RR*	*TT1*	*TT2*	*RR*
	n = 42	*n = 46*[a]	*n = 34*[b]	*n = 42*	*n = 46*[a]	*n = 34*[b]
LT1	0.78	0.74	0.54	-11.3 (18.3)	-8.7 (22.1)	-9.8 (37.2)
LT2	0.87	0.80	0.53	-23.2 (16.0)	-18.5 (19.3)	-27.4 (39.3)
LT3	0.88	0.84	0.64	-16.2 (14.3)	-16.1 (18.1)	-21.9 (33.8)
LT4	0.78	0.82	0.61	-3.5 (23.7)	-0.6 (26.2)	-13.5 (36.1)
LT5	0.67	0.65	0.58	43.7 (21.7)	45.3 (23.5)	39.1 (35.9)
LT-4mmol	0.88	0.85	0.61	-31.7 (19.4)	-32.3 (23.0)	-36.6 (36.1)
Dmax	0.89	0.82	0.73	-4.4 (12.1)	-3.8 (15.8)	-13.4 (32.4)
Dmax-mod	0.94	0.84	0.76	-27.3 (11.8)	-29.9 (16.6)	-33.7 (29.1)

Pearson correlation between each lactate threshold concept in GXT 1 and time trial test 1 (TT1), GXT 4 and time trial test 2 (TT2) and GXT 5 and average road race (RR) power for all subjects combined. All correlations are significant ($p<0.0002$). To determine potential differences in power output between the LT concept and time trial power or race power, mean difference (SD) between each measurement pair is calculated. Negative values indicate lactate threshold power is higher than exercise test average power.

[a] For LT5 n = 44;

[b] for LT5 n = 32.

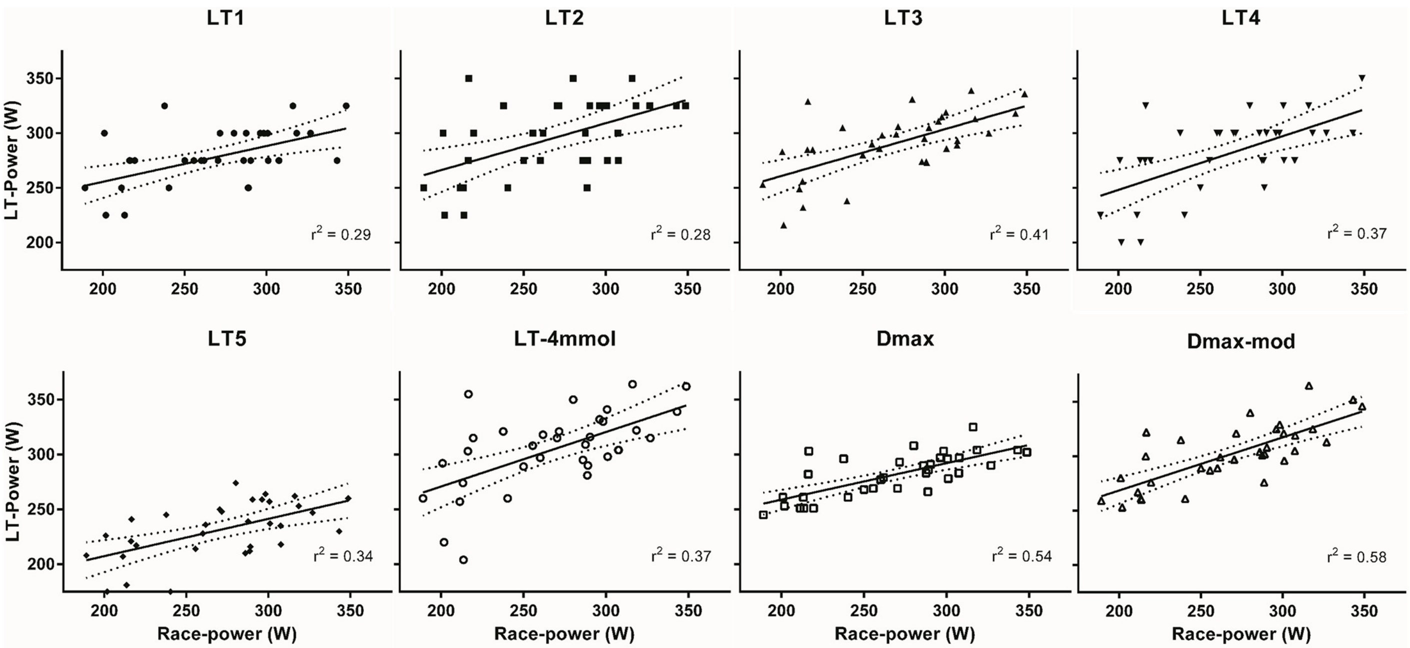


Fig 4. Linear regression lactate threshold concept power and average race power. Linear regression of lactate threshold power and average race power per LT concept for all subjects depicting linear regression line (solid line) and 95% confidence interval (dotted lines). r^2: R-squared or coefficient of determination is the proportion of the variance in the dependent variable that is predictable from the independent variable.

study identified several LT concepts that outperformed the others in the setting of this trial. The best method being Dmax-mod, but Dmax, LT-4mmol and LT3 performed well too.

Methodology

The design of the exercise protocol, for example stage duration, is known to impact blood lactate curves. [25, 26] We selected an exercise protocol with five minute stages and 25 W increments because it takes 3–4 minutes for the body to reach steady state and lactate accompanying that effort level can be measured accurately. [27] In addition, longer protocols may be more sensitive to performance changes. [25] As described in more detail elsewhere, [12] GXT results show our subjects were well-trained with maximal power output and VO2 max values comparable to elite cyclists and triathletes when using longer exercise protocols. [28, 29] All evaluated concepts were applied to data from the same exercise tests, with the same sampling and assay method, and the same fitting procedure was used for those applicable concepts. As a result such factors could not have affected the comparison between concepts within this study. The current study was designed in that way to give the most accurate estimate of performance parameters and its controlled set-up seems to be the most robust and valuable way to determine differences between concepts. Nevertheless, when any of these factors are changed (e.g. using a different exercise test protocol) it is possible the outcomes might not translate perfectly. With regards to data fitting, the third-order polynomial in the applicable concepts performed well given the high mean r-squared values observed.

Selection of concepts

After inspection of all identified lactate concepts, it became clear that there were similarities between quite some of them. For this reason, the concepts were grouped into categories, and a selection was made of concepts to be analysed to have at least one representative per category and thereby ensuring that results from this study would be informative for all regularly used lactate concepts. The selection includes concepts such as a fixed lactate value (usually at 4 mmol/L) and the visually determined LT concepts that have been used since the conception of the LT, and more recent concepts such as LT3, LT4 and Dmax and Dmax-mod. [9]

Mean threshold

The mean power output (Table 2) is relatively constant over time for each concept. These results confirm there was no placebo-effect on any of the LT concepts, although such an impact would already theoretically be improbable. What can also be seen is that not all concepts seem to identify the same point in the blood lactate curve: LT5 gives the lowest estimate of LT (228.4 W), much lower than other concepts (274.1–302.7 W). LT-4mmol and Dmax-mod have the highest estimates (302.7 and 301.5 W), indicating these concepts identify different intensities of performance and have different physiological meanings. Applying the terminology as described in Faude *et al*, [9] based on mean threshold and mean difference with TT and RR (Table 4), some concepts seem to be more related to the aerobic threshold (LT5), others to the aerobic/anaerobic transition (e.g. LT1, LT4, Dmax) or the anaerobic threshold (LT-4mmol, Dmax-mod).

Repeatability

Intra-subject CV's over all five measurements were low (3.4–8.1%) and Cronbach's alphas high (0.89–0.97), indicating repeatability of all concepts over the study period of approximately 8 weeks was good. This corresponds well to previous findings of repeatability for power or speed at different lactate concepts, both in terms of CV, determined at 1.3–5.9% in a meta-analytic review, [30] and in terms of Pearson correlations 0.88–0.96 [31, 32] or ICC 0.98–0.99. [33] One study applied different LT concepts to the same dataset from two exercise tests and showed that intra-subject CV's and correlation was good for LT2, LT-4mmol and a concept similar to LT4 (CV 3–4% and $r \geq 0.85$), but not for Dmax (10.3% and 0.57). [34] Our data, based on more subjects (24 versus 14) and more measurements per subject (5 versus 2), disputes this relatively poor repeatability for Dmax. However, our study does show differences between concepts, with LT3, LT-4mmol, Dmax and Dmax-mod having the lowest intra-subject CV (<5%) and the highest Cronbach's alpha (>0.95).

Correlation with performance

As we have established that CV and repeatability for all LT concepts was good, the most relevant question is whether these concepts correlate to actual endurance performance. As previously indicated, it is highly unlikely that the rHuEPO treatment impacted this particular correlation analysis. When analysing the groups separately, some differences in correlation coefficients could be observed between the two groups (data not shown), but these differences were already present for the correlation between GXT1 and TT1 when treatment had not yet started, indicating that this was not due to rHuEPO treatment. Because combining all subjects generates more informative and robust results being based on a bigger population, pooling the groups was considered justified.

Data in Table 4 show that for all concepts correlations with time trial tests were higher compared to the road race (based on all subjects median of all concepts r = 0.875 for TT1 and 0.82 for TT2, *versus* 0.61 for RR). This is most likely partly due to additional variability in the road race due to the circumstances (e.g. weather, uphill racing with changes in steepness over the course, and differences in race duration (range 72–126 minutes)). Possibly there was also a minor impact of using different equipment for power measurement during the RR, as it was not measured on the ergometer but on the subjects' bike. What can also be seen is that correlation of the LT concepts with TT1 in general is slightly higher than with TT2. More importantly however, correlation for both time trials show that the ranking among different concepts is very similar, confirming the results are robust. It seems that in general, using a technique of interpolation for the BLC has superior performance, as LT concepts that were based on the third-order polynomial derived from the individual lactate concentration measurements (LT3, LT5, LT-4mmol, Dmax, Dmax-mod) performed better than the ones that used actual measured bLa values without interpolation (LT1, LT2 and LT4), with the exception of LT5. This poor performance of LT5 is most likely due to the fact that it is conceptually different from the other concepts; it is the power at the minimum lactate equivalent, in this case the lowest value for the lactate-VO2 ratio. In contrast, LT3 also uses a form of the minimum lactate equivalent, but it adds 1.5 mmol/L to this value. As can be seen in Table 2 and Fig 2, this leads to LT5 on average determining a point even before the first rise in lactate concentration as determined by LT1. This concept therefore relates to much lower (aerobic) work intensities than the other concepts. Additionally it is less repeatable (see Table 3). From all tested concepts LT5 correlates least with 45 min TT performance, but for the longer RR performance relative to the other concepts it performs somewhat better than for TT. This could mean that is this concept is more related to long-term exercise efforts.

Many studies previously evaluated correlations of LT concepts with endurance performance, of which most used running performance. An overview of these studies by Faude *et al* shows a median r = 0.84–0.92 for several different LT concepts for endurance distances (>5km), [9] comparable to our results. There are fewer studies that have compared LT concepts and their correlation with different types of cycling endurance performance, [16, 20, 26, 35–38] but correlation with endurance performances (30–90 minutes) for each concept seem to vary between these studies, see Table 5. In addition, the comparison between concepts within these studies shows varying conclusions about which is the best concept. This could partially be due to differences between studies, for example study populations differ (mean VO2max ranges from 48 to 68 mL/kg/min, and some studying female, others male cyclists and/or triathletes). However, they are more or less as heterogeneous as our population with an SD of 4–8 mL/kg/min on VO2max. The applied exercise protocols all used long stages similar to ours (3–5 minutes), although the increases in workload differ (20-50W). Finally, correlation to endurance exercise was based on time trials that lasted between approximately 30 to 90 minutes (our TT of 45 min at the lower end and RR of on average 98 min at the higher end), a difference that might impact the correlation to different LT concepts. Nevertheless, taking these differences into account, comparison is possible, albeit with some caution. Moreover, a robust

Table 5. Reported correlations between LT concepts and endurance performance.

Correlation reported in publication	Lactate threshold concept						
	LT1	LT2	LT4	LT-4mmol	Dmax	Dmax-mod	LTlog
Amann [16]	-	0.72	0.59	0.60	-	-	-
Bentley [37]	-	-	-	0.54	0.77	-	0.91
Bishop [20]	0.81	-	0.61	0.81	0.84	0.83	0.69
Borszcz [35]	-	0.31	-	0.56	0.75	-	-
McNaughton [38]	-	-	-	0.90	0.91	-	0.86
Nichols [36]	-	-	0.88	0.67	-	-	-

Literature data for LT concepts and correlation with 30–90 minute-during performances. LTlog: the power output at which bLa starts to increase when log(bLa) is plotted against log(power output).

and valid LT concept should perform well in any of these datasets. What can be observed is that all these concepts except LT1, Dmax and Dmax-mod have shown correlations below 0.75, and that in all four direct comparisons that evaluated both Dmax and LT-4mmol, Dmax showed a higher correlation. This latter finding could be due to the fact that LT-4mmol is less robust to changes in settings such as exercise protocol duration, sampling site and lactate analyser because of its fixed nature. Our study expands on this information, and compared to previous studies as reviewed in Table 5, is based on approximately 2–4 times more subjects, therefore allowing for more robust conclusions. This is especially true since our population is a heterogeneous well-trained, and therefore relevant, group (range maximal power output at baseline 256–425 W). Similar to what can be extracted from the literature, our study too shows that Dmax and Dmax-mod have highest correlations with time trial performance, although LT-4mmol and LT3 show a similarly high correlation in our study. For the correlation with RR, there are slightly larger differences between concepts. Correlation is highest for Dmax and Dmax-mod, mainly because for the other concepts correlation for a few subjects is very poor, as visualized in Fig 4 (e.g. for LT-4mmol). These findings combined, we conclude that Dmax, and even more so Dmax-mod, have the best correlation with endurance performance. One recent study evaluated correlation between MLSS, which could be considered to be the gold standard for the physiological endurance threshold, and different LT concepts generated from GXTs with different protocol durations. [26] This study concluded that for a GXT with 4-minute steps (most similar to our GXT), correlation was high for many of the concepts, but validity was highest for LT-2.5mmol, Dmax-mod, and two modified versions of Dmax-mod. In contrast, LT2, LT-4mmol and Dmax showed much higher mean differences with MLSS and therefore were designated as invalid estimates of MLSS. Combining these findings with our own results, Dmax-mod determined in a GXT with approximately 5-minute stages is both a valid estimate of MLSS and has a high correlation with actual endurance performance.

Absolute power difference

The mean difference of each concept with the endurance parameter gives an indication of how the absolute power of the LT concept corresponds to the average power produced during TT and RR. On average, power is higher compared to the endurance test for each concept (except the poorest performing concept LT5). This difference in power between LT concepts and endurance test is possibly due to having to sustain the power for a much longer time during the endurance tests, needing a systematic lower power in order to cope with the effort. Interestingly, Dmax-mod and LT-4mmol, concepts that show among the highest correlations, have the largest difference in absolute power (approximately 30 W). Given the high correlation with performance this should not disqualify these concepts, but one should take into account that there is a systematic difference with endurance performance of approximately 30 W.

Conclusions

LT concepts are correlated with endurance performance, but a review showed that many different concepts are used in literature, which is undesirable. [9] Also for cycling performance, there is no consensus on which LT concept should be applied and results vary highly. [16, 20, 35–38] In this study we compared eight different representative LT concepts on the same large cycling performance dataset to evaluate repeatability and predictive properties. All concepts showed high repeatability, and correlated with endurance performance. However, LT3, LT-4mmol, Dmax and Dmax-mod showed the best repeatability, and had the highest correlation with time trial performance. As correlation with performance was consistently high for Dmax and Dmax-mod, also with the uphill road race, the latter performing slightly better on each criterion, and because Dmax-mod was previously shown to be a valid estimate of MLSS, we would recommend using Dmax-mod when analyzing the blood lactate curve.

Supporting information

S1 Table. Lactate threshold concept categories.
(DOCX)

S1 Protocol. Study protocol CHDR1514.
(PDF)

Author Contributions

Conceptualization: Jules A. A. C. Heuberger, Adam F. Cohen.

Data curation: Jules A. A. C. Heuberger, Pim Gal, Frederik E. Stuurman.

Formal analysis: Jules A. A. C. Heuberger, Wouter A. S. de Muinck Keizer, Yuri Mejia Miranda.

Investigation: Jules A. A. C. Heuberger, Pim Gal, Frederik E. Stuurman.

Methodology: Jules A. A. C. Heuberger.

Supervision: Jules A. A. C. Heuberger, Adam F. Cohen.

Writing – original draft: Jules A. A. C. Heuberger.

Writing – review & editing: Pim Gal, Frederik E. Stuurman, Wouter A. S. de Muinck Keizer, Yuri Mejia Miranda, Adam F. Cohen.

References

1. Cairns SP. Lactic acid and exercise performance: culprit or friend? Sports Med. 2006; 36(4):279–91. https://doi.org/10.2165/00007256-200636040-00001 PMID: 16573355
2. Atkinson G, Davison R, Jeukendrup A, Passfield L. Science and cycling: current knowledge and future directions for research. J Sports Sci. 2003; 21(9):767–87. https://doi.org/10.1080/0264041031000102097 PMID: 14579871
3. Kindermann W, Simon G, Keul J. The significance of the aerobic-anaerobic transition for the determination of work load intensities during endurance training. Eur J Appl Physiol Occup Physiol. 1979; 42 (1):25–34. PMID: 499194
4. Londeree BR. Effect of training on lactate/ventilatory thresholds: a meta-analysis. Med Sci Sports Exerc. 1997; 29(6):837–43. PMID: 9219214
5. Antonutto G, Di Prampero PE. The concept of lactate threshold. A short review. J Sports Med Phys Fitness. 1995; 35(1):6–12. PMID: 7474995
6. Robergs RA, Ghiasvand F, Parker D. Biochemistry of exercise-induced metabolic acidosis. Am J Physiol Regul Integr Comp Physiol. 2004; 287(3):R502–16. https://doi.org/10.1152/ajpregu.00114.2004 PMID: 15308499
7. MacRae HS, Dennis SC, Bosch AN, Noakes TD. Effects of training on lactate production and removal during progressive exercise in humans. J Appl Physiol (1985). 1992; 72(5):1649–56.
8. Stanley WC, Gertz EW, Wisneski JA, Neese RA, Morris DL, Brooks GA. Lactate extraction during net lactate release in legs of humans during exercise. J Appl Physiol (1985). 1986; 60(4):1116–20.
9. Faude O, Kindermann W, Meyer T. Lactate threshold concepts: how valid are they? Sports Med. 2009; 39(6):469–90. https://doi.org/10.2165/00007256-200939060-00003 PMID: 19453206
10. Coyle EF, Coggan AR, Hopper MK, Walters TJ. Determinants of endurance in well-trained cyclists. J Appl Physiol (1985). 1988; 64(6):2622–30.
11. Billat VL, Sirvent P, Py G, Koralsztein JP, Mercier J. The concept of maximal lactate steady state: a bridge between biochemistry, physiology and sport science. Sports Med. 2003; 33(6):407–26. https://doi.org/10.2165/00007256-200333060-00003 PMID: 12744715

12. Heuberger J, Rotmans JI, Gal P, Stuurman FE, van't Westende J, Post TE, et al. Effects of erythropoietin on cycling performance of well trained cyclists: a double-blind, randomised, placebo-controlled trial. Lancet Haematol. 2017; 4(8):e374–e86. https://doi.org/10.1016/S2352-3026(17)30105-9 PMID: 28669689

13. Tanaka H. Predicting running velocity at blood lactate threshold from running performance tests in adolescent boys. Eur J Appl Physiol Occup Physiol. 1986; 55(4):344–8. PMID: 3758032

14. Coyle EF, Martin WH, Ehsani AA, Hagberg JM, Bloomfield SA, Sinacore DR, et al. Blood lactate threshold in some well-trained ischemic heart disease patients. J Appl Physiol Respir Environ Exerc Physiol. 1983; 54(1):18–23. https://doi.org/10.1152/jappl.1983.54.1.18 PMID: 6826403

15. Dickhuth H-H, Yin L, Niess A, Rocker K, Mayer F, Heitkamp HC, et al. Ventilatory, lactate-derived and catecholamine thresholds during incremental treadmill running: relationship and reproducibility. Int J Sports Med. 1999; 20(2):122–7. https://doi.org/10.1055/s-2007-971105 PMID: 10190774

16. Amann M, Subudhi AW, Foster C. Predictive validity of ventilatory and lactate thresholds for cycling time trial performance. Scand J Med Sci Sports. 2006; 16(1):27–34. https://doi.org/10.1111/j.1600-0838.2004.00424.x PMID: 16430678

17. Dickhuth H.-H.; Huonker M. MT, Drexler H., Berg A., Keul J. Individual anaerobic threshold for evaluation of competitive athletes and patients with left ventricular dysfunctions. Advances in ergometry. 1991.

18. Sjodin B, Jacobs I. Onset of blood lactate accumulation and marathon running performance. Int J Sports Med. 1981; 2(1):23–6. https://doi.org/10.1055/s-2008-1034579 PMID: 7333732

19. Cheng B, Kuipers H, Snyder AC, Keizer HA, Jeukendrup A, Hesselink M. A new approach for the determination of ventilatory and lactate thresholds. Int J Sports Med. 1992; 13(7):518–22. https://doi.org/10.1055/s-2007-1021309 PMID: 1459746

20. Bishop D, Jenkins DG, Mackinnon LT. The relationship between plasma lactate parameters, Wpeak and 1-h cycling performance in women. Med Sci Sports Exerc. 1998; 30(8):1270–5. PMID: 9710868

21. Chambers J. Project R The R Project for Statistical Computing [3.4.0:[https://www.r-project.org/.

22. Hadley Wickham RF, Lionel Henry, Kirill Müller. dplyr: A Grammar of Data Manipulation dplyr: A Grammar of Data Manipulation2017 [https://cran.r-project.org/web/packages/dplyr/.

23. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research psych: Procedures for Psychological, Psychometric, and Personality Research2017 [https://cran.r-project.org/web/packages/psych/.

24. Hadley Wickham LH. tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions2017 [https://cran.r-project.org/web/packages/tidyr/.

25. Bentley DJ, Newell J, Bishop D. Incremental exercise test design and analysis: implications for performance diagnostics in endurance athletes. Sports Med. 2007; 37(7):575–86. https://doi.org/10.2165/00007256-200737070-00002 PMID: 17595153

26. Jamnick NA, Botella J, Pyne DB, Bishop DJ. Manipulating graded exercise test variables affects the validity of the lactate threshold and [Formula: see text]. PLoS One. 2018; 13(7):e0199794. https://doi.org/10.1371/journal.pone.0199794 PMID: 30059543

27. Thoden JS. Testing aerobic power. In: MacDougall JD, Wenger HA, Green HJ, editors. Physiological testing of the high-performance athlete. Champaign: Human Kinetics; 1991. p. 107–74.

28. San Millan I, Bing K, Brill C, Hill JC, Miller LE. Randomized controlled trial of Micro-Mobile Compression (R) on lactate clearance and subsequent exercise performance in elite male cyclists. Open Access J Sports Med. 2013; 4:221–7. https://doi.org/10.2147/OAJSM.S51956 PMID: 24379728

29. Bentley DJ, McNaughton LR. Comparison of W(peak), VO2(peak) and the ventilation threshold from two different incremental exercise tests: relationship to endurance performance. J Sci Med Sport. 2003; 6(4):422–35. PMID: 14723392

30. Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med. 2001; 31(3):211–34. https://doi.org/10.2165/00007256-200131030-00005 PMID: 11286357

31. Weltman A, Snead D, Stein P, Seip R, Schurrer R, Rutt R, et al. Reliability and validity of a continuous incremental treadmill protocol for the determination of lactate threshold, fixed blood lactate concentrations, and VO2max. Int J Sports Med. 1990; 11(1):26–32. https://doi.org/10.1055/s-2007-1024757 PMID: 2318561

32. Grant S, McMillan K, Newell J, Wood L, Keatley S, Simpson D, et al. Reproducibility of the blood lactate threshold, 4 mmol.l(-1) marker, heart rate and ratings of perceived exertion during incremental treadmill exercise in humans. Eur J Appl Physiol. 2002; 87(2):159–66. https://doi.org/10.1007/s00421-002-0608-2 PMID: 12070627

33. Pfitzinger P, Freedson PS. The reliability of lactate measurements during exercise. Int J Sports Med. 1998; 19(5):349–57. https://doi.org/10.1055/s-2007-971929 PMID: 9721059

34. Pallares JG, Moran-Navarro R, Ortega JF, Fernandez-Elias VE, Mora-Rodriguez R. Validity and Reliability of Ventilatory and Blood Lactate Thresholds in Well-Trained Cyclists. PLoS One. 2016; 11(9): e0163389. https://doi.org/10.1371/journal.pone.0163389 PMID: 27657502

35. Borszcz FK, Tramontin AF, de Souza KM, Carminatti LJ, Costa VP. Physiological Correlations With Short, Medium, and Long Cycling Time-Trial Performance. Research quarterly for exercise and sport. 2018; 89(1):120–5. https://doi.org/10.1080/02701367.2017.1411578 PMID: 29334005

36. Nichols JF, Phares SL, Buono MJ. Relationship between blood lactate response to exercise and endurance performance in competitive female master cyclists. Int J Sports Med. 1997; 18(6):458–63. https://doi.org/10.1055/s-2007-972664 PMID: 9351693

37. Bentley DJ, McNaughton LR, Thompson D, Vleck VE, Batterham AM. Peak power output, the lactate threshold, and time trial performance in cyclists. Med Sci Sports Exerc. 2001; 33(12):2077–81. PMID: 11740302

38. McNaughton LR, Roberts S, Bentley DJ. The relationship among peak power output, lactate threshold, and short-distance cycling performance: effects of incremental exercise test design. J Strength Cond Res. 2006; 20(1):157–61. https://doi.org/10.1519/R-15914.1 PMID: 16506862

4

Sport-Induced Substance Use—An Empirical Study to the Extent within a German Sports Association

Monika Frenger*, Werner Pitsch, Eike Emrich

Department Economics and Sociology of Sport, Faculty of Human Sciences and Economics, Saarland University, Saarbrücken, Germany

* m.frenger@mx.uni-saarland.de

Abstract

In cooperation with the Sports Association of the Palatinate (SBP), a survey was conducted on substance use by recreational and amateur athletes. Distribution of the online questionnaire took place by means of chain-referral sampling, and questions on substance use were presented using the randomized response technique (RRT) to protect the anonymity of respondents and prevent socially desirable answers. The estimated lowest limit for the population share for use of prohibited substances during the last season (4%) and for lifetime use (3.6%) did not differ significantly. Approximately 21% of respondents had used substances for training or competitions that were taken for a purpose other than performance enhancement (e.g., to improve their mood or to help with recuperation from a minor injury or illness) in the last year. 49% had done so at some point in their life.

Editor: Fabio Lucidi, University of Rome, ITALY

Competing Interests: The authors have declared that no competing interests exist.

Introduction

In recent years, there has been a multitude of studies on doping in recreational and in amateur sport. As a result, the extent of this phenomenon was clarified for special populations (like e.g. bodybuilders [1–2]) and for special substances (mostly for anabolic steroids [3–9]). Nevertheless, the prevalence in the total population of amateur and recreational sportsmen and -women remained unclear due to several analytical, methodological, and empirical issues. Despite the uncertainty of prevalence data, doping in recreational and in amateur sport has been labelled a public health issue [10–12]. The following article aims to supplement this argument with a sound estimation of the prevalence of doping in a general population of sportsmen and -women.

Theoretical Assessment of the Problem Area

State of research

It is immediately evident that the large number of studies on the prevalence of doping in recreational and amateur sport taking into consideration only certain substance classes or special sport populations can only provide limited estimates of the prevalence within the total population. Therefore, we will first focus on studies that address either unspecific or general substance (ab-)use, or that address general populations with only minor limitations in socio-economic parameters, especially those studies without limitations by sport discipline.

A survey among students at four universities in Belgium, Demark, Germany, and Switzerland (population: well educated young European adults) on the use of enhancing substances for both mental performance (neuro-enhancement) and sports performance (doping) with the Randomized Response Technique (RRT) (see methods section in this article) showed that the prevalence of doping in that population was 5%, while the prevalence of (neuro-) enhancement ranged between 3 and 8% [13]. It should be noted here that this proportion only represented the reliably estimated lower boundary of "honest yes respondents" in the Randomized Response estimate (see methods section). The existence of respondents who failed to follow survey instructions may have led to an inaccurate estimate of the potential number of users. Another study, conducted at the German Sport University Cologne, revealed a higher prevalence of substance use at 11.2% among students (population: young, well educated German adults, highly interested in sports). This figure resulted from analyses of urine test samples and is not affected by any voluntary bias. However, the prevalence rate may be elevated due to the specific nature of the university, at which only sports students were questioned [14].

Concerning the substance use of young male adults in school and university environments, unspecified substance use has also been studied using direct questioning. With a wide definition of performance enhancing substances, which also include dietary supplements, the prevalence was found to be about 31% [15]. This rate suggested a high level of readiness to enhance performance among the sample. Among those who indicated they had used performance enhancing substances, 31% confessed to having used illicit substances according to the NCAA guidelines. This resulted in a doping prevalence of 9.6% for this population. For a similar population from Sweden, the overall proportion of doping was reported at 1.6% [16]. A study among Italian high school attendees found a prevalence rate of 1.5% for using illegal performance enhancing substances, while the total rate of substance use for performance enhancement was 6.8% [17]. As the focus of the study was substance use during the last three months, the reported prevalence might be low in comparison to other studies considering a longer period for reporting use.

To critically assess the evidence from these studies, it is important to note that the first two studies used methods which were either impervious to social desirability bias [14] or that were designed to minimize it [13]. This was not the case for the latter studies that were conducted using direct questioning. As a result of these survey findings on relatively large populations for general substance use to enhance physical performance, we estimated the overall prevalence between approximately 5% and 10% in the total population.

Now we turn to studies that additionally identified risk factors for smaller populations or for single substances or substance groups. In order to structure the large number of studies, we will categorize the discussion based on the different determinants studied.

The most important and highly stable social background variable is gender, with female respondents generally showing a lower propensity to dope. Its significance as a determinant of doping prevalence in amateur and recreational sport was shown in studies among gym users [1–2, 18] as well as for the use of anabolic androgenic steroids among university students ([3–9, 16], 14 more studies from the United States, cited in [5], 25 studies and three periodically conducted national youth surveys in the U.S., cited in [11], and three studies from Great Britain and Germany, cited in [10]). This effect was also shown in a special mixed sample of at-risk individuals [9], as well as in studies on doping substances in general among university students [17]. Few studies where the influence of gender was investigated offered no support for this determinant [19] or provided contradictory results [20]. Although this pattern is stable, there have been only few attempts to link these results to theoretical issues such as gender specific health-related behaviour or gender specific risk behaviour.

Additionally studies have identified further social background variables such as social status [4, 7, 9, 12, 20, 21] and the intensity of involvement in sport [2, 7, 17, 19, 22] as determinants of doping. As we were not able to test hypotheses on these parameters due to the limited number of responses to our survey (see results section), we will not discuss these variables in detail.

Besides these social background variables, an increased prevalence among users of other legal or illegal substances has been reported in several studies [2, 6–9, 15–16, 22–24], though there were some inconsistent results concerning various drugs [22].

Drawing on this research, we hypothesized that doping prevalence in recreational and in amateur sport differs by gender. In general, this study did not aim to identify new determinants for doping, but to 1.) estimate the overall prevalence, and 2.) to determine if gender is a determinant for doping. So far, gender has only been shown as a determinate in limited populations and mostly for single substances.

The concept of doping in recreational and amateur sport

The concept of doping has evolved in the context of professional elite sport and it is unclear if this concept can be adopted in sporting activities outside the professional realm. It is important to note that the definition of substance use relevant to this study is not necessarily that promulgated by the World Anti-Doping Agency (WADA). To clarify the concept of "substance use in recreational and amateur sports", we must differentiate between three definitions.

1. Substance use, or doping, according to the World Anti-Doping Code (WADC): This code defines doping in terms of a list of prohibited substances or methods. Whether or why the athlete intends to use these substances or methods does not matter, due to the strict liability principle. Additionally, under this definition doping does not require intent to enhance one's performance, because there are additional reasons why substances are prohibited (e.g., the "spirit of sport" argument).
2. The European Union's (quoted in [25]) definition of substance use: The core concept is the intention to unfairly enhance one's athletic performance; the specific techniques and methods used therefore do not matter at all.
3. Lay persons' understanding of substance use in sports contexts, which can be derived, for example, from discussions on the internet: In this context, neither the intention to increase one's performance nor a specific list of substances matter. For example, some pain-killers are discussed as "doping substances", while pharmaceuticals that are included on lists of banned substances are not necessarily seen as doping as long as they are used for (self-) therapy.

If participants in mass sport events are members of a sports club, they accept the rules of the WADC and of the sports organization, especially those of fair play, with their membership. Athletes in recreational and amateur sport, too, are expected to perform at their best by using only their talent and by training hard, even if there is no organized surveillance for doping violations. Doping surveillance and sanctioning can, at least in principle, be applied to those who participate in organized competitions. However, as there are typically no doping tests at the amateur level of sport there is no practical need for participants to be informed about the legal meaning of doping (WADA's definition) or the list of prohibited substances.

Therefore, substance use in recreational and amateur sport is better understood in the sense of the Thomas theorem: "If people think that something is real, it is real in its consequences"

[26]. In recreational and amateur sport, the lack of doping tests and their results mean that there is no institutional agent for defining behaviours as doping. This means the labelling of a behaviour by the participants in this field is most relevant, whether or not tests have been conducted. If the behaviour is labelled as doping, consequences such as loss of reputation and social marginalization are equally effective sanctions, regardless of whether the behaviour was doping in a legal sense.

In recreational and amateur sport, it is also important to consider the intent behind substance use. For athletes who are under surveillance in the WADA Anti-Doping Testing regime, the relevant question is only whether or not a substance is in the athlete's body; it does not matter why this substance has been used. Applying this standard to recreational and amateur sport could have major consequences for the use of over-the-counter medicine. Taking a pill to treat nasal congestion caused by a cold may constitute an anti-doping rule violation if the pill contains pseudoephedrine hydrochloride. Under WADA's definition, this is a prohibited substance. However, using it for self-treatment in a non-sports-related context would hardly be considered doping (see [27] for the differentiation between doping/enhancement and treatment in the general public). Using a medication to facilitate training participation when an athlete feels like he may have caught a cold might not be labelled prohibited substance use by athletes in recreational and amateur sport. On the other hand, using creatine, caffeine, or l-carnithin (none of which are prohibited) in marathon training [28] or using pain-killers in sport [29] has already been colloquially labelled doping by some in these populations.

Taking these observations together, substance use in recreational and amateur sport is understood by participants in this field as a behaviour

1. that is regarded as performance-enhancing, and
2. that is (self-) perceived as illegal.

Therefore, measuring prohibited substance use prevalence in this sense requires measuring the proportion of athletes who believe they have done something prohibited. This prevalence is an indicator of an openness to transgress between substance categories (i.e., from acceptable to prohibited) that is accepting of doping. Cases where prohibited substance use acceptance exists but has not (yet) led to substance use behaviour due to the absence of necessity or opportunity are known systematic measurement errors, which lead to an underestimation of the prevalence of a doping-tolerant mind-set.

Method

Study's definitions–doping and sport induced self-medication

The operationalization of substance use has one important implication when measuring its prevalence: to rule out other reasons for substance use. When measuring the concept of "doping", one has to emphasize three essential elements of our definition: "athlete", "prohibited substances or methods", and "the intention to enhance your performance in sport". Our questionnaire asked explicitly about the use of substances that the respondent believes to be prohibited. Therefore, using substances that are not known to be prohibited would lead to the answer "no", while substances that are thought to be prohibited would lead to the answer "yes", irrespective of whether these substances are, in fact, prohibited. The defining element "athlete" ensures that respondents will refer to a behaviour that they perceive as related to their sports participation. Substances used for marginal health problems, which are not part of the social phenomenon "illicit substance use in recreational and amateur sports", are beyond the scope of this definition.

In addition to these performance related behaviour, we were interested in the substance use for purposes other than performance enhancement. The differences to these two behaviours are 1) the intention (other than performance enhancement, such as pain reduction, mood, etc.), 2) that the reference to a specific sport is only relevant for doping but not for sport induced self-medication, and 3) that the substances which are used for doping are seen as illicit by the respondents, while in the case of self-medication this issue is irrelevant. By taking the substance, the athletes could, for example, try to reduce pain, accelerate recuperation, or improve their mood. Although any of these attempts can be seen as (also) performance enhancing, the main difference to doping is that medication use was not primarily intended to enhance performance.

We asked the entire sample about substance use for sport-induced self-medication (in tables and figure, we will use the abbreviation medi) over their whole sport life and the persons who were still active in the last year. This means we only asked about substance use linked to active sports participation. The original question (translated into English) is: “Have you ever (last year) used substances in training or competition which were not intended to enhance your performance?”

Instruments (structure of the questions)

We used an online survey to investigate the prevalence of doping and sport induced self-medication. At the beginning of the survey, respondents were asked if they participate in sports and which sport disciplines they practice or have practiced. For these disciplines, we also asked if they participate/have participated in competitions. If the response was yes, we asked for their last competitive level, as well as for the highest competitive level they have ever reached. This was used to prioritise sport disciplines in our questionnaire (see below). These questions were followed by at least two and up to six questions concerning their sports-related substance use (see RRT procedure section). The questionnaire finished with socio-demographic information (i.e., age, gender, education level).

Specific instrument—The RRT questions

With RRT, we asked a maximum of two questions relating to doping (lifetime and last season) and two questions relating to medication (lifetime and last year). Because of that all participants got one question (sport-induced self-medication, lifetime) in minimum and could get up to four RRT questions. A summary of the questions surveyed using the RRT is shown in Fig 1.

The questions related to doping in the last season in one sport (basketball in our example, for the lifetime prevalence the question was modified accordingly) were asked in the following way: “Have you, as a basketball player, last season used prohibited substances or methods with the aim to enhance your performance?” This question was asked to those persons who participated in sport competitions over the previous season (or ever). The relevant sport was decided after the prioritisation (see instruments section).

In addition to the doping questions, we asked about sport-induced self-medication during the last year and over the respondent’s lifetime. For this question, we asked about use in the last year because non-competitive athletes were also questioned here and they might not find the term “season” helpful or intuitive, which may lead to respondents assuming different windows of time. These questions about sport-induced self-medication were not specific to any one type of sport, but according to the differentiation between doping and sport induced self-medication (see study’s definiton section), they explicitly excluded the intention to enhance sporting performance.

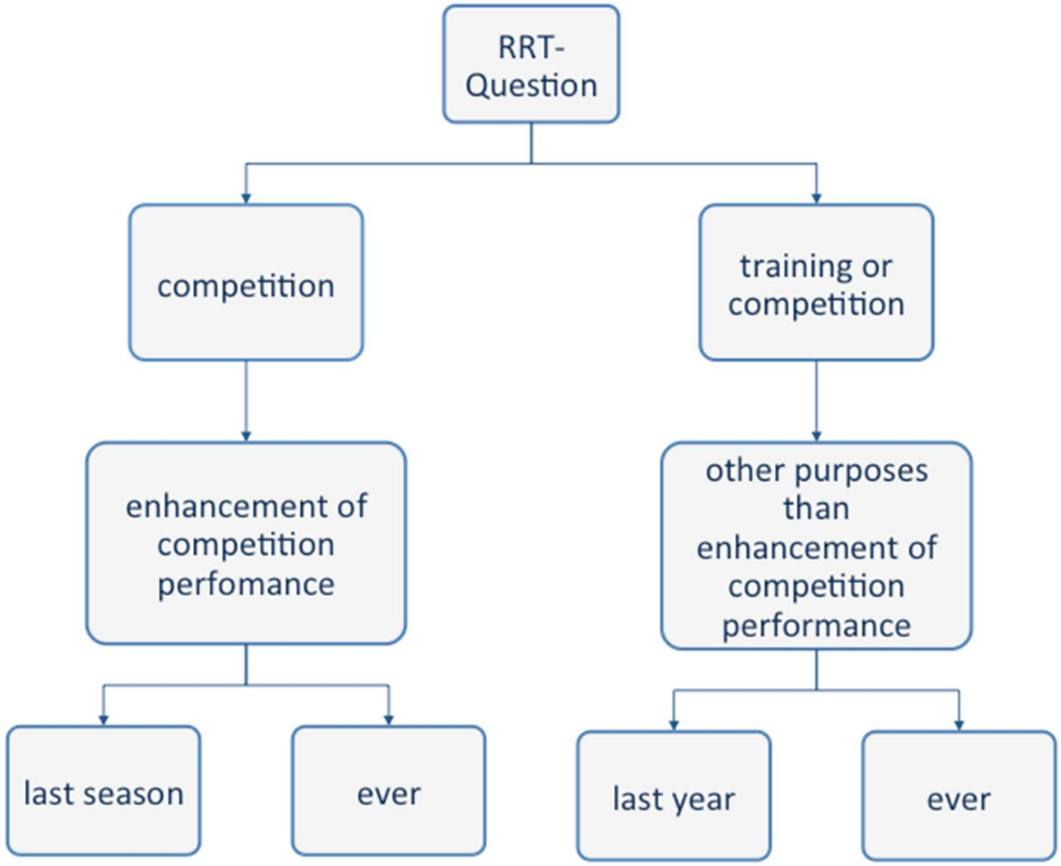


Fig 1. Overview of the questions about the embarrassing characteristic substance use that are asked using the RRT.

The RRT procedure

Because admitting to doping can be highly embarrassing, we used the randomized response technique (RRT) for questions on doping and medication. In addition to ensuring our results are comparable to other surveys [13, 19, 30–33], this approach leads to more reliable answers than those obtained by asking direct questions because RRT reduces distorting influences, such as those resulting from social desirability [34–35]. RRT cannot erase all social desirability bias, but in our opinion it guards against the worst fear, which is of exposure. We know from methodological studies on RRT [36–41] that results come closer to capturing true prevalence than direct questions. We only know two studies with lower prevalence estimations in the RRT questionnaire than when questioning directly [42–43]. That may be the result of reduced fear, since RRT questions can moderate the weight of expectation to answer "correctly" by means of social desirability. Additionally, recent publications have shown this technique produces reliable estimations of doping prevalence in sport (see [30–32] for elite sport as well as [13, 19, 33, 44], for sport below the professional realm).

This method protects against bias by giving an additional instruction when answering a question for an embarrassing property. Depending on which result is randomly generated from a known distribution, respondents either answer the embarrassing question or a corresponding innocuous question. In our case, this second question is structured in such a way that a cooperative person will always answer it with "yes".

The literature contains reports on the use of different random generators [45–47] or other randomly generated characteristics such as the final digit of telephone numbers [48–49]. However, for an Internet survey, it cannot be assumed that all respondents will have easy access to a random generator in the form of a coin or die precisely at the moment they are taking part in the survey. Other randomly generated characteristics, such as the final digit of a telephone number, assume the researchers know the distribution of telephone numbers at the moment the survey is administered. In this survey, we decided to use randomly generated numbers with equally distributed digits.

The process is illustrated by means of an example in Fig 2 that shows the forced answer model of the RRT for qualitative characteristics [50] (for other variants, such as the unrelated question model or for quantitative characteristics, see, for example, the summary in [45]).

Because the researcher does not know the random number generated for the respondent, he cannot conclude from a "yes" answer that the respondent has actually used prohibited substances or methods. However, because we do know the distribution from which the random number is generated, we can derive the probability that the respondent is instructed to answer the embarrassing question. From this, the proportion of people in the population exhibiting the characteristic (here, athletes who have taken substances) can be calculated.

Despite special instructions, cases continue to occur in which, for unknown reasons, the respondents do not comply with the RRT procedure [51–52]. These cases of so-called "cheating" can occur for a variety of reasons (for example, deliberately not complying with the instruction, not understanding the instruction, or similar errors) and thus reduce the precision of the estimate. To control for these biases, the "cheater detection model" [53] has been developed. The cheater detection is based on the assumption that RRT estimates of population shares are independent from the probability to answer one of the innocuous questions or the embarrassing question. To detect cheating, the sample is randomly split into (normally equally sized) subsamples with different probabilities. With these two groups, we can estimate three population proportions, namely the rate of honest yes responders, the rate of honest no responders and the rate of respondents who do not answer according to the instructions. In the literature on RRT, this third proportion is called "cheaters" although this does not imply that these respondents did not follow the instructions deliberately, nor that, in the present case, they cheated in the sense that they used prohibited substances. Nevertheless, we will use the term "cheater" for this population share according to the terminology in this special field of statistical methods. Further analysis of the RRT method with cheater detection is available elsewhere [54].

Data Analysis

The RRT can be used for exploratory prevalence estimation as well as for hypothesis testing, although the statistics needed for this second purpose are not straightforward. For the variants of RRT used here, error components are typically not distributed normally, but are heavily skewed [13]. Therefore, significance tests should be performed using bootstrap methods to calculate confidence intervals independently from any distribution assumption [55–56]. When using bootstrap statistics the meaning of a significant difference between two groups must be reinterpreted. For our purposes, the difference between the two groups under study and the (bootstrapped) confidence interval of the difference are used, and the reference bcα-(bias corrected and accelerated) value is reported in Tables A-D in S1 Appendix (5th column). For two groups to differ significantly, the value of the difference and the limit both must be positive or negative. Otherwise, the test fails and the hypothesis is rejected.

With this method, results of statistical tests are reported in a distinctive way because it lacks the value of a statistic (such as a t-value or an F-value from an ANOVA) or degrees of freedom.

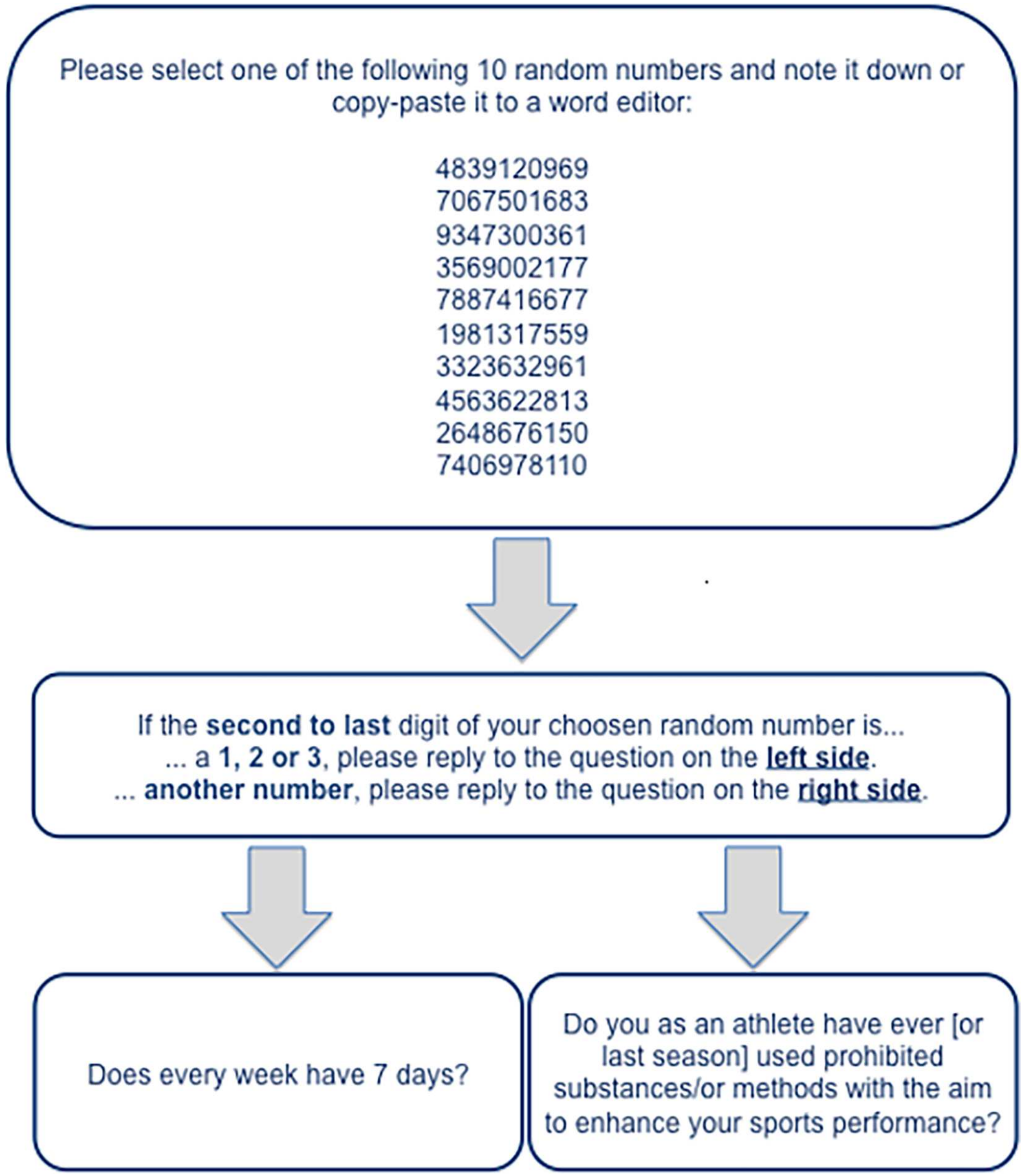


Fig 2. Example of an RRT question similar to those asked from respondents.

Additionally, the often-reported p-value, indicating the estimated significance level, becomes meaningless. A p-value could be reported [55], but would likely vary between different bootstrap-simulations even if the null hypothesis were consistently rejected at the selected level for significance (here: $p<0.05$). The results of the bootstrap analyses are shown in the additional tables (Tables A-D in S1 Appendix).

Ethical Issues

According to the guidelines of the German Research Association, no ethical approval was needed because the research did not pose any threats or risks to the respondents and the respondents were fully informed about the objectives of the study (http://www.dfg.de/

foerderung/faq/geistes_sozialwissenschaften/index.html). A member of the ethical committee of Saarland University confirmed that ethical approval was not needed for this study.

Conformity with the data privacy act of the European Union is confirmed by the data security officer at Sportbund Pfalz. This includes that neither participation nor non-participation could render negative consequences to the addressees, and that complete anonymity of the respondents was verified. As a result of this procedure, there is no possibility of de-identifying or de-anonymizing the records prior to analysis.

Implicit informed consent is given by participating after being fully informed of the objectives of the study. Written or verbal informed consent is not obtained during this study for two reasons. First, recording information that could be used to identify the participants (especially names and e-mail-addresses but also IP-addresses, time of participation) is explicitly prohibited by the data security officer. Therefore, written or verbal consent could not be obtained. Second, recording these data would have lowered the trust of the respondents in the anonymity of the study and would have foiled the logic of the RRT survey. The study was performed according to the principles of the Helsinki declaration.

Sample

The survey was carried out in the catchment area of the SBP, the umbrella organization for organised sport in the Palatinate. Access to recreational and amateur athletes was achieved with the cooperation of the SBP via different media (such as press releases and reports in regional newspapers) and using a snowball process. The sports-practicing public was notified about the survey via the Sports Association, with the organization and the group of researchers writing several times (in "special newsletters" to all subscribers, approximately 8,000 people, and in emails to the addresses held by the SBP for association officials, trainers, participants in events, etc., approximately 9,900 people). In all of the communications, the persons contacted were first asked to take part in the survey themselves. They were additionally asked to further distribute the information about the survey within their club, their training group, their circle of friends, etc.

Due to the snowball process, the transmission of information depended on a large number of factors that are almost impossible to influence, e.g., the commitment of those individuals initially contacted during the distribution, the (communication) networks in the clubs, the intensity and frequency of information exchange, etc. This gave rise to biased distributions with regard to the types of sport practiced (see Table 1). A further reason for this bias was the fact that, in the questionnaire, non-team sports were mentioned, while the distribution of athletes in the SBP was determined by the distribution of members across the constituent specialist clubs.

Two oddities in Table 1 (indicated by *) were influenced by the special structure of sports in Germany and do not reflect sampling errors. First, a rather small number of respondents stated "gymnastics" as their type of sport. The *Palatinate Gymnastics Association* covers different sport disciplines and sport clubs that are historically under the umbrella of German "Turnen". Additionally, this association covers all types of sport that lack a special regional umbrella organisation. Therefore, the number of members of the *Palatinate Gymnastics Association* far exceeded the share of respondents practicing gymnastics. The opposite phenomenon explained the difference between the number of athletes involved in "jogging/running" and in "athletics" and the number of members of the *Palatinate Athletics Association*. Running competitions are governed by the Athletics Association and there is no special runners' association. Nevertheless, it is possible for athletes who do not belong to any athletic club to take part in competitions, such as city marathons or fun runs, which can be organized to allow runners who do not

Table 1. Distribution of responses by types of sport practiced compared with the distribution of sports practice in the SB Pfalz. N = 1,620 data sets stating the type of sport practiced in the sports with the highest priority.

Sport	Number of respondents	Percentage of respondents	Percentage of athletes in SBP
Football	297	15,2	30,0
Running *	197	10,1	-
Athletics *	183	9,4	3,9
Tennis	137	7,0	6,3
Handball	100	5,1	3,5
Aquatics	91	4,7	1,7
Gymnastics *	83	4,3	18,9
Volleyball	69	3,5	1,0
Cycling	68	3,5	1,0
Triathlon	64	3,3	0,2
Badminton	60	3,1	0,9
Table tennis	58	3,0	2,9
Dancing	49	2,5	1,2
Judo	48	2,5	0,7
Shooting	37	1,9	3,8
Alpine Skiing	29	1,5	2,1
Equestrian	28	1,4	2,1
Bowling	22	1,1	0,5

* the asterisk denotes a sport with special characteristics that is explained in the text.

belong to a sports club to participate (that is, these competitions do not require athletes to have a starting license).

Our dataset comprised 1,930 responses containing individual data and 1,964 responses containing questions about types of sport (here, one person can contribute with two datasets about their sports behaviour). The first question asked for the participant's zip code to assure that the person lives in the region of Palatinate. Then we asked if the person is or had been formerly engaged in sports. For the RRT questions we prioritise one sport (see methods section) stated by the respondent.

It was not easy to quantify the response rate due to the snowball process. We only can estimate the rate from those persons who directly received the questionnaire. We scored all newsletter subscribers from SBP, a sample with more than 10,000 persons, with a special newsletter. Additionally, an email went out to all instructors and participants of an advanced training course (about 9,000). There was some overlap between these two groups, so that 10,000 direct mail recipients would be a conservative estimation. Using this number, we had a response rate of less than 20%. Because of other processes, like information sharing across sports clubs and the several newspaper articles about the survey, the response rate was clearly lower. 65.8% of the responses came from male athletes and 33.42% from female (the gap between male and female respondents comes from item non-responders). This distribution did not correspond to the population distribution in Rhineland-Palatinate, but it was very close to the distribution of sexes at the SBP and, thus, presumably to that of the population of the Palatinate who actively practice sport and who are the targets of the study.

Furthermore, the sample was characterized by high educational and professional status, which was evident from the high proportion of persons having completed secondary education and the that almost one third of respondents had a university degree.

Results

Results–prevalence of doping

When assessing the prevalence of substance use, we estimated the rate of honest yes respondents and of honest no respondents directly from the data. Typically, both shares did not sum up to 100%, as there remains a range of indifference because some participants (deliberately or by chance) did not obey the RRT instructions correctly ("cheaters" in the RRT-terminology). Therefore, we calculated an interval for the true prevalence: The lower boundary is the proportion of honest yes respondents (the red bars in Fig 3), while the upper boundary is defined by the proportion of honest no respondents (the green bars in Fig 3). The breadth of the interval between these two shares refers to the level of indifference. It ranges from the estimated proportion of honest yes respondents to 100% minus the share of honest no respondents.

Results–prevalence for sport-induced self-medication

Regarding whether athletes in the SBP deliberately used prohibited substances and/or methods with the aim of enhancing their competitive performance during the last season of their named sport, the proportion of honest "yes" respondents was 4.29%. With a proportion of honest no responders of 69.52%, the true prevalence falls between 4.29% and 30.48%. The percentage of respondents who doped at any point in their sports life to date falls between 3.35% and 10.55%. The proportion of honest no respondents was 89.45%. The much larger range of the interval due to non-compliance to the RRT instructions for the first question may be surprising at first, but it was similar to that seen in comparable studies of deviant behaviour [30–31]. A possible explanation for the fact that non-compliance to RRT-instructions on the question relating to

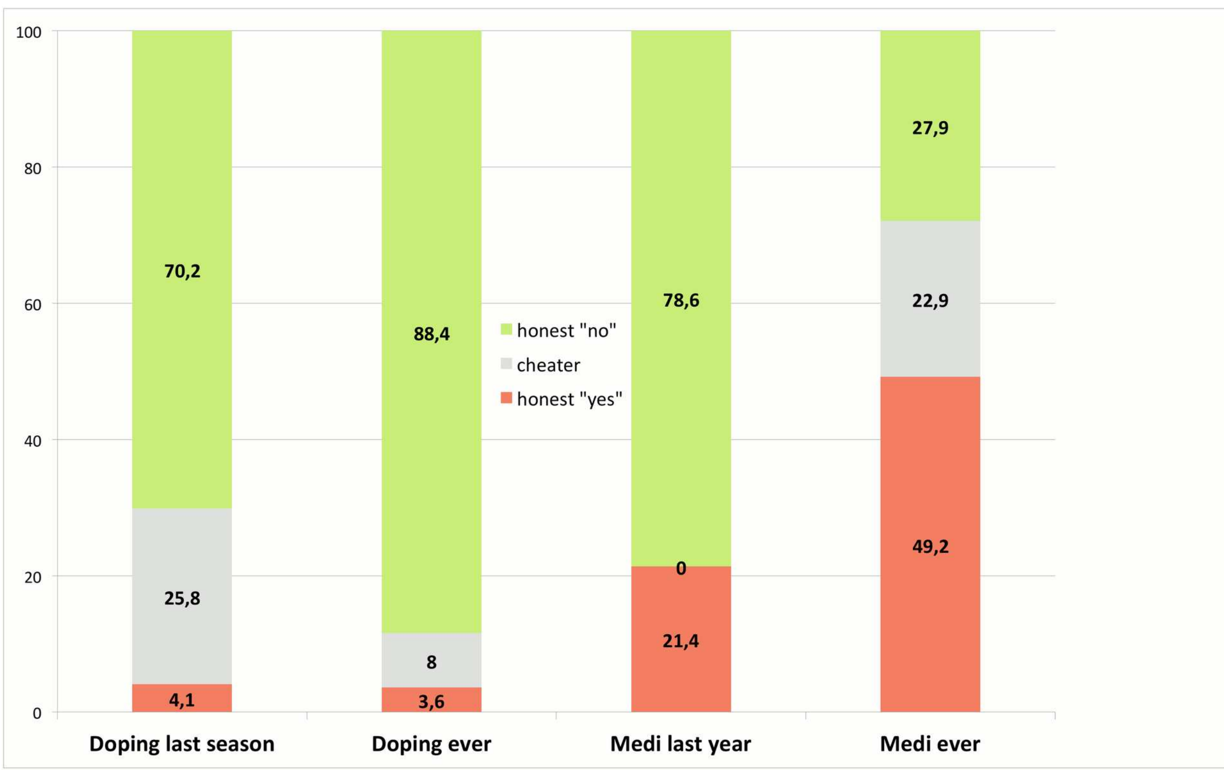


Fig 3. Results of the RRT questions for doping and sport-induced self-medication (medi).

the last season is seen more often than on the question relating to one's entire sporting life might lie in the proximity of the response to the embarrassing question to the deviant behaviour itself. The risk of being identified is higher for deviance that occurred recently than it is for deviance that occurred at some point during a long period of time.

Another question asked whether the athletes in the SBP used any substance for sport-induced self-medication without intending to enhance their competitive performance in the past year. With regard to substance use in the last year, the proportion of honest "yes" respondents in the last year was 21.40%. The proportion of honest no responders was estimated at 78.60%. Thus we did not have an interval for this question. With regard to the lifetime prevalence of substance use for self-medication in a sport-related context, we obtained an estimate of 49.20% honest yes respondents and 27.87% honest no respondents. The upper limit of the interval for lifetime use is 72.1%.

Hypotheses Tests

The results broken down by gender are shown in Tables A and B in S1 Appendix. The differences between male and female athletes are highest for the question regarding whether they have ever taken prohibited substances (doping), but the differences between genders did not reach significance for any question. The cell frequencies for the questions question were very low, so each estimate to compare sexes is relatively unreliable.

For the social status as well as the sport involvement hypotheses, the cell frequencies were even lower, which meant that this study did not provide any evidence either to support or to reject these hypotheses.

Without a pre-defined hypothesis, we additionally compared the substance use for doping purposes to the use of medication in the sport context for other purposes than to enhance performance. We have N = 616 respondents who answered both the doping question and the question for sport-induced self-medication during the last season or year and N = 786 answers for both questions about lifetime use. Although numerically different, the prevalence for doping and medication use did not differ significantly for the last season/last year question (see supporting information Table C in S1 Appendix). Likewise, there is a significant difference in these questions on the respondent's lifetime in both the number of honest "yes" and the number of honest "no" responses (see Table D in S1 Appendix).

Discussion

We found a last season prevalence of doping over 4% and a lower lifetime prevalence (3.6%). The common logic would say that lifetime prevalence can only be the same or higher than the last season's prevalence. So these results may puzzle at first and need more explanation. First, these results are estimations that contain a certain error term. Because of the result of the significance test we can say, that the difference isn't a relevant difference. Second, different arguments affect the RRT procedure. For each RRT question we have to use a randomization process to divide the sample into 2 groups. This means that between the two questions (last year and lifetime) the sample that has to answer honestly can be assembled in a different way. For example, in the first question respondents get the instruction to answer honestly and in the second to answer always yes (respectively to answer the right or left question). The information on the prevalence only can come from those who answer honestly, so that the randomization process can lead to a different estimation.

When interpreting our results in a public health context, we must conclude that in spite of the low prevalence rate of doping, from a population of (conservatively estimated) more than 20 million amateur and recreational athletes in Germany, nearly 900,000 are estimated to have used illegal substances in the last season to enhance their sporting performance. In addition, more than four million amateur and recreational athletes have used pharmaceuticals for purposes other than performance enhancement in the context of sport. Therefore, we must concede that substance use without the aim of performance enhancement in sports-related contexts is a larger public health problem than doping (substance use with the aim of performance enhancement), but that both problems indeed affect large numbers of individuals. These forms of sport-induced use of medication at least partly counteract health implications, which are often seen as a side effect of recreational sport. In contrast to the cited literature we could not find a gender difference in such a way that substance use is a male problem. Maybe the multidisciplinary sample and the recreational sports level can cause this. We found the reported differences especially at students and bodybuilding studies.

Regarding the RRT questions about prohibited substances that are used with the aim of enhancing performance, we highlight the tension between the fact that, at least in principle, the WADC applies to recreational and amateur athletes and the fact that we cannot assume that casual athletes are aware of the list of banned substances and its contents. Responses to our survey are thus biased by this tension in different ways. On the one hand, we had respondents who take substances that they regard as prohibited but that are not prohibited according to the WADC, leading to an overestimation of the prevalence of doping. On the other hand, we had respondents who take substances that they believe are permitted but are actually prohibited according to the WADC, leading to an underestimation of the prevalence of doping. Nevertheless, the blurred line that separates conforming behaviour from deviance in amateur and recreational sport is a strong argument for a tailored Anti-Doping regime in this realm to preserve fairness and the spirit of sport [57].

At this point, we lack information about how familiar recreational and amateur athletes are with the WADA list of banned substances, and thus which substances they consider to be prohibited or performance enhancing. Further studies are required on this topic so that this bias can be factored into future models. In this study we are measuring substance use, as we are focused on how many athletes reported taking substances they considered to be performance enhancing and prohibited.

When comparing the rates of substance use for sport-induced self-medication with rates of doping, we show that motives other than performance enhancement are the far more prevalent phenomenon. One argument for testing in or outside of competitions is that anti-doping measures protect athletes' health. It is therefore interesting that we find the more widespread behaviour is related to medication, which is not under the purview of competitions, federations, or sports organisations. In our opinion, there is no way to introduce a testing system similar to elite sports in this domain, and any type of organized regulation is destined to fail in this area. One reason is because there is no money to do it. Another reason is that recreational athletes use substances as they see fit if they are for medication purposes (but possibly prohibited in the sports environment). There are lots of areas in amateur sports that cannot be controlled by sports organizations (no license need, private sports participation, fitness studios etc.). If we look at the reported motivations for substance use, performance enhancement has less of an effect than do other motivations like pain reduction, recuperation, or mood improvement. Additionally, prohibited substances are less used than permitted substances.

Limitations of the Study

This study has limitations, so generalizations from the results should be made with caution.

First of all, the use of RRT aims to address the problem of social desirability bias, although this limits the reliability of any attempts to measure embarrassing properties by principle.

When estimating the proportions of the three response types (honest yes respondents, honest no respondents, and cheaters in RRT-terminology) it is apparent, as has already been shown in other studies, that the range of indifference due to respondents not obeying the RRT instructions correctly is lower when the embarrassing question (here, doping) asks about a broader and more distant period of time (here, any point in one's athletic life). When the behaviour takes place more recently, the proportion of indifference is higher. This finding is consistent with findings from work on elite German athletes [30–31] and on substance use by students [13]. It strongly supports the assumption that this range of indifference is largely influenced by deliberate cheating while random fluctuations in respondents' behaviour adds little to this proportion. The sensitivity of the proportion of cheaters to the temporal proximity of the doping behaviour, and the degree of threat from confessing to the embarrassing characteristic, is a strong argument for using RRT even in recreational and amateur sport where the social desirability bias is plausibly lower than in elite sport. This means that with the use of the RRT method we are sure to reduce this general limitation problem. Additionally, we can show that with the RRT the proportion of social desirable answers are lower and we can measure this proportion.

Second, the studied sample was not necessarily representative of all amateur athletes or of all sports represented by the SBP, due to the rather poor response rate in comparison to the athletes taking part in different sports (see Table 1). Nevertheless, it provided the first estimate of the prevalence of substance use in a population survey using a technique that elicits reliable results. These results are comparable to other studies in recreational sport, as well as to studies on doping in elite sports. Nevertheless, the results should be interpreted cautiously until they can be independently replicated.

Third, the athletes in our sample may not have known which products are banned and which are not in their particular sport. We made a distinction between doping and medication for other purposes than performance enhancement in the analysis. However, though the respondents were asked if they had "ever used prohibited substances or methods with the aim to enhance their performance" or "ever used substances with other aims than performance enhancement", we did not know if this distinction was clear for the individual respondents or if the substances they believed to be banned matched those actually banned by the WADC. Therefore, the results should be understood as reflecting the relative frequency of an openness to transgress between substance categories (i.e., from acceptable to prohibited) rather than as information on the frequency of genuine anti-doping rule infringements. Additionally, we cannot distinguish between substance use that is a kind of self-medication and those used under the guidance of a physician. We only asked about intent and not the kind of medication.

Finally, due to the challenges of implementing no-cheater detection for RRT [54], the setup may not have been not fully optimized to reduce error variance in the estimators. On the other hand, given the intrinsic difficulties in obtaining reliable prevalence data for sensitive behaviour and the paucity of doping prevalence data in the general population, our study had the distinction of being the first to report results in amateur athletes in a specific region across all sports using a method that allows us to reduce bias in surveys on sensitive issues.

Supporting Information

S1 Appendix. Explanation to the Bootstrap-tests for significance and Tables A-D with the different comparisons (with the independent variable "sex" and between the questions for doping and medi).
(DOCX)

Author Contributions

Conceptualization: EE WP MF.

Formal analysis: WP MF.

Investigation: WP MF.

Methodology: WP MF EE.

Project administration: EE MF.

Software: WP MF.

Visualization: MF.

Writing – original draft: EE MF WP.

Writing – review & editing: MF EE WP.

References

1. Boos C, Wulff P, Kujath P, Bruch HP. Medikamentenmißbrauch beim Freizeitsportler im Fitneßbereich. Deutsches Ärzteblatt. 1998; 95(16): A-953–A-957.
2. Kartakoullis NL, Phellas C, Pouloukas S, Petrou M, Loizou C. The Use of Anabolic Steroids and Other Prohibited Substances By Gym Enthusiasts in Cyprus. International Review for the Sociology of Sport. 2008; 43(3): 271–287.
3. Kersey RD. Anabolic-Androgenic Steroid Use Among California Community College Student-Athletes. Journal of Athletic Training. 1996; 31(3): 237–241. PMID: 16558405
4. Winsor R, Dimitru D. Prevalence of anabolic steroid use by male and female adolescents. Medicine and Science in Sports and Exercise. 1989; 21(5): 494–497. PMID: 2607943
5. Tanner SM, Miller DW, Alongi C. Anabolic Steroid Use by Adolescents: Prevalence, Motives, and Knowledge of Risks. Clinical Journal of Sport Medicine. 1995; 5: 108–115. PMID: 7882111
6. Nilsson S, Baigi A, Marklund B, Fridlund B. The prevalence of the use of androgenic anabolic steroids by adolescents in a county of Sweden. Eur J Public Health. 2001; 11(2): 195–197. PMID: 11420810
7. Thorlindsson T, Halldorsson V. Sport, and use of anabolic androgenic steroids among Icelandic high school students: a critical test of three perspectives. Substance Abuse Treatment, Prevention, and Policy. 2010; 5(1): 32.
8. Du Rant RH, Escobedo LG, Heath GW. Anabolic- Steroid Use, Strength Training, and Multiple Drug Use Among Adolescents in the United States. Pediatrics. 1995; 96: 23–28. PMID: 7596717
9. Pope HG, Katz DL, Champoux R. Anabolic-Androgenic Steroid Use Among 1,010 College Men. The Physician and Sportsmedicine. 1988; 16(7): 75–84. doi: 10.1080/00913847.1988.11709554 PMID: 27403827
10. Müller-Platz C, Boos C, Müller RK. Doping beim Freizeit- und Breitensport. Gesundheitsberichterstattung des Bundes: Vol. 34. Berlin: Robert Koch-Institut. 2006. Available: http://www.rki.de/cln_049/nn_199850/DE/Content/GBE/Gesundheitsberichterstattung/GBEDownloadsT/doping,templateId=raw,property=publicationFile.pdf/doping.pdf
11. Yesalis CE, Bahrke MS. Doping among adolescent athletes. Best Practice & Research Clinical Endocrinology & Metabolism. 2000; 14 (1): 25–35.
12. Tahtamouni LH, Mustafa NH, Alfaouri AA, Hassan IM, Abdalla MY, Yasin SR. Prevalence and risk factors for anabolic-androgenic steroid abuse among Jordanian collegiate students and athletes. The European Journal of Public Health. 2008; 18(6): 661–665. doi: 10.1093/eurpub/ckn062 PMID: 18603598
13. Pitsch W, Frenger M, Emrich E. Doping im Breiten- und Freizeitsport. Zur Überprüfung von Hypothesen mittels RRT-gewonnener Daten. In: Kempf H, Nagel S, Dietl H, editors. Im Schatten der Sportwirtschaft (Sportökonomie, 15). Schorndorf: Hofmann; 2013. pp. 111–125.
14. Thevis M, Sauer M, Geyer H, Sigmund G, Mareck U, Schänzer W. Determination of the prevalence of anabolic steroids, stimulants, and selected drugs subject to doping controls among elite sport students using analytical chemistry. Journal of sports sciences. 2008; 26(10): 1059–1065. doi: 10.1080/02640410801910293 PMID: 18608840

15. Buckman JF, Yusko DA, White HR, Pandina RJ. Risk Profile of Male College Athletes Who Use Performance-Enhancing Substances. Journal of Studies on Alcohol and Drugs, 2009; 919–923. PMID: 19895768

16. Kindlundh AMS, Isacson DGL, Berglund L, Nyberg F. Factors associated with adolescent use of doping agents: anabolic- androgenic steroids. Addiction. 1999; 94(4): 543–553. PMID: 10605850

17. Mallia L, Lucidi F, Zelli A, Violani C. Doping Attitudes and the Use of Legal and Illegal Performance-Enhancing Substances Among Italian Adolescents. Journal of Child and Adolescent Substance Abuse. 2013; 22(3): 179–190.

18. Bergsgard NA, Tangen JO., Barland B, Breivik G. Doping in Norwegian Gyms—A Big Problem? International Review for the Sociology of Sport. 1996; 31: 352–365.

19. Dietz P, Ulrich R, Dalaker R, Striegel H, Franke AG, Lieb K et al. Associations between physical and cognitive doping—A cross-sectional study in 2.997 triathletes. PLoS ONE. 2013; 8 (11), e78702. doi: 10.1371/journal.pone.0078702 PMID: 24236038

20. Wazaify M, Bdair A, Al-Hadidi K, Scott J. Doping in gymnasiums in amman: The other side of prescription and nonprescription drug abuse. Substance Use and Misuse. 2014; 49(10): 1296–1302. doi: 10.3109/10826084.2014.891625 PMID: 24611822

21. Razavi Z, Moeini B, Shafiei Y, Bazmamoun H. Prevalence of anabolic steroid use and associated factors among bodybuilders in Hamadan, western province of Iran. Journal of Research in Health Sciences. 2014; 14(2): 163–166. PMID: 24728754

22. Striegel H, Simon P, Frisch S, Roecker K, Dietz K, Dickhuth HH et al. Anabolic ergogenic substance users in fitness-sports: A distinct group supported by the health care system. Drug and Alcohol Dependence, 2006; 81(1): 11–19. doi: 10.1016/j.drugalcdep.2005.05.013 PMID: 16009506

23. Melia P, Pipe AL, Greenberg L. The Use of Anabolic-Androgenic Steroids by Canadian Students. Clinical Journal of Sport Medicine. 1996; 6: 9–14. PMID: 8925377

24. Nilsson S. Androgenic anabolic steroid use among male adolescents in Falkenberg. European Journal of Clinical Pharmacology. 1995; 48: 9–11. PMID: 7621856

25. Berendonk B. Doping-Dokumente. Von der Forschung zum Betrug. Berlin: Springer; 1991.

26. Thomas WI, Thomas DS. The Child in America: Behavior Problems and Programs. Knopf; 1928.

27. Partridge B, Lucke J, Hall W. "If you're healthy you don't need drugs": Public attitudes towards "brain doping" in the classroom and "legalised doping" in sport. Performance Enhancement & Health 2014; 3 (1): 20–25.

28. Schmidt S. Marathon-Dopingmittel—gefährlich und unfair. In: Laufplatz Rhein/Ruhr [Internet]. Düsseldorf. Available: http://www.laufplatz.de/lauftraining-marathon-das-wissens-lexikon-fuer-anfaenger/marathon-doping-kreatin-koffein-und-schmerzmittel-beim-lauftraining.html. Accessed 07 October 2015.

29. McGratt M. Is pain medication in sport a form of legal doping? BBC News. 5 June 2012. Available: http://www.bbc.com/news/science-environment-18282072. Accessed 07 October 2015.

30. Pitsch W, Emrich E, Klein M. Doping in elite sports in Germany: results of a www survey. European Journal of Sport and Society 2007; 4(2): 89–102.

31. Pitsch W, Emrich E. The frequency of doping in elite sport: Results of a replication study. International Review for the Sociology of Sport. 2012; 47: 559–580.

32. Striegel H, Ulrich R, Simon P. Randomized response estimates for doping and illicit drug use in elite athletes. Drug and Alcohol Dependence, 2010; 106 (2–3): 230–232. doi: 10.1016/j.drugalcdep.2009.07.026 PMID: 19740612

33. Schröter H, Studzinski B, Dietz P, Ulrich R, Striegel H, Simon P et al. A Comparison of the Cheater Detection and the Unrelated Question Models: A Randomized Response Survey on Physical and Cognitive Doping in Recreational Triathletes. PLoS ONE. 2016; 11: e0155765. doi: 10.1371/journal.pone.0155765 PMID: 27218830

34. Schänzer W. Comment of Laboratory B on the publication by Lundby et al. "Testing for recombinant human erythropoietin in urine: problems associated with current anti doping testing. J Appl Physiol 2008, doi: 10.1152/japplphysiol.90529.2008". Retrieved from http://www.dshs-koeln.de/biochemie/rubriken/07_Info/Comment_Lundby%20(2).pdf

35. Brown WM, College T. Ethics, Drugs, and Sport. Journal of the philosophy of sport. 1980; 7: 15–23.

36. Böckenholt U, Barlas S, van der Heijden PGM. Do randomized-response designs eliminate response bias? An empirical study of non-compliance behaviour. Journal of Applied Economics. 2009; 24: 377–392.

37. Coutts E, Jann B, Krumpal I, Näher AF. Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions. Journal of Official Statistics. 2011; 3, 439–448.

38. Jerke J, Krumpal I. Plagiate in studentischen Arbeiten. Eine empirische Untersuchung unter Anwendung des Triangular Models. Methoden, daten, analysen. 2013; 7: 347–368.
39. Kerkvliet J. Cheating by economics students: A comparison of survey results. Journal of Economic Education. 1994; 25: 121–133.
40. Scheers NJ, Dayton CM. Improved Estimation of Academic Cheating Behavior Using the Randomized Response Technique. Research in Higher Education. 1987; 26: 61–69.
41. Reckers PMJ, Wheeler SW, Wong-on-Wing B. A Comparative examination of audotir premature sign-off using the direct and the randomizied response methods. Auditing: A Journal of practice and Theory. 1997; 16: 69–78.
42. Duffy JC, Waterton JJ. Randomized Response versus direct questioning–Estimating the prevalence of alcohol-related in a field-survey. Australien Journal of statistic. 1988; 30: 1–14.
43. Krotki KJ, Fox B. The randomized response technique, the interview and the self-administered questionnaire–An empirical comparison of fertility report. Proceeding of the American statistical Association. 1974; 367–371.
44. Simon P, Striegel H, Aust F, Dietz K, Ulrich R. Doping in fitness sports: estimated number of unreported cases and individual probability of doping. Addiction. 2006; 101: 1640–1644. doi: 10.1111/j.1360-0443.2006.01568.x PMID: 17034444
45. Scheers NJ. Methods, Plainly Speaking. A Review of Randomized Response Techniques. Measurement and Evaluation in Counselling and Development. 1992; 24: 27–41.
46. Tracy PE, Fox JA. The Validity of Randomized Response for Sensitive Measurements. American Sociological Review. 1981; 46: 187–200.
47. Shimizu IM, Bonham GS. Randomized Response Techniques in a National Survey. Journal of the American Statistical Association. 1978; 73: 35–39. PMID: 12260868
48. Stem DE, Steinhorst K. Telephone Interview and Mail Questionnaire Applications of the Randomized Response Model. Journal of the American Statistical Association. 1984; 79: 555–564.
49. Goodstadt MS, Gruson V. The Randomized Response Technique: A Test on Drug Use. Journal of the American Statistical Association. 1975; 70: 814–818.
50. Boruch RF. Maintaining Confidentiality of Data in Educational Research: a Systemic Analysis. The American Psychologist. 1971; 26: 413–430.
51. Adolphsen J. Der Staat im Dopingkampf. Sportwissenschaft. 2008; 38: 82–88.
52. Wiseman F, Moriarty M, Schafer M. Estimating Public Opinion with the Randomized Response Model. Public Opinion Quarterly. 1975; 39: 507–513.
53. Clark SJ, Desharnais RA. Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model. Psychological Methods. 1998; 3: 160–168.
54. Feth S, Frenger M, Pitsch W, Schmelzeisen P. Cheater Detection bei Randomized Response-Technik. Herleitung, Analyse und Anwendung. Saarbrücken: universaar; 2014.
55. Efron B, Tibshirani R. An introduction to the bootstrap. New York, NY: Chapman, Hall; 1993.
56. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge university press; 1997.
57. Henning AD, Dimeo P. Questions of fairness and anti-doping in US cycling: The contrasting experiences of professionals and amateurs. Drugs Edu Prev Pol, 2015; 1–10.

5

The effect of contact sport expertise on postural control

Ying Liang [1]*, Michael Hiley [2], Kazuyuki Kanosue [3]

1 Department of Physical Education, Anhui Normal University, Wuhu, China, **2** School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, Leicestershire, United Kingdom, **3** Faculty of Sports Sciences, Waseda University, Tokorozawa, Saitama, Japan

* chloe_liangying@163.com

Abstract

It has been demonstrated that expertise in sport influences standing balance ability. However, little is known concerning how physical contact in sport affects balance ability. The aim of this study was to examine whether differences between contact and limited-contact sport experiences results in differences in postural control. Twenty male collegiate athletes (10 soccer/contact, 10 baseball/limited contact) and ten male untrained students stood quietly on a force plate under various bipedal and unipedal conditions, with and without vision. Significant differences for sway area and COP speed were found between the soccer players and the other two groups for unipedal stances without vision. Soccer players were found to have superior postural control compared with participants involved in limited contact sport or no sport at all. Contact sports may lead to increased postural control through enhanced use of proprioceptive and vestibular information.

Editor: Thomas A Stoffregen, University of Minnesota, UNITED STATES

Introduction

Expertise in sports which required good balance, e.g. gymnastics and dance, is of particular benefit to postural control [1,2,3,4,5,6]. Posture is controlled by integrating visual, proprioception and vestibular information [7]. These three types of information are obtained from the environment and the task [1,3,8,9,10,11]. In light of this fact, it should be expected that if non-gymnasts and non-dancers spend a great deal of time in environments with continual external disturbance, they may also develop greater adaptive ability in postural control. However, one limitation in postural control research is that previous studies may have been "contact sport biased" in non-gymnastics sports such as soccer, handball or American football [1,2,3]. Balance studies have been conducted on a variety of sports [12,13], however, the amount of physical contact involved has not been taken into account when attempting to clarify how expertise in sport contributes to postural control. Ideally, to obtain this information, a prospective study of contact experience is necessary. An extensive review of the literature on balance and different sports has been conducted [14], however, the consistency of measures (equipment and task difficulty) and sports of different amounts of contact has not been conducted. In the review of balance and various sports the consideration of the amount of contact involved in a sport is not addressed [14].

Competing interests: The authors have declared that no competing interests exist.

"Contact sport" is a term used in both competitive activity and in medical terminology to indicate a sport that emphasizes or requires physical contact between players [15]. Different classification has been used in different situations in relation to contact in sports. In order to categorize the degree of contact in different sports optimally, the system adopted by the United States for medical terminology has been used. This system uses the term "contact sport" to refer to sports such as soccer and basketball, in which athletes routinely make contact with each other or inanimate objects, but usually with less force than in "collision sports", such as rugby and American football. The term "limited-contact sport" denotes sports such as squash and baseball, in which contact with other athletes is infrequent or inadvertent [15]. The focus of the present study was to make comparisons between soccer players, baseball players, and controls (novices) to establish whether differences between contact and limited-contact sport experiences result in differences in postural control.

Postural responses induced by external perturbations have been thoroughly investigated in relation to standing positions such as bipedal and unipedal stances [1,2,5,16,17,18,19], which have aimed to differentiate the complexity of postural performance in line with decreases in the "supporting area". However, in many control studies, postural sway in bipedal stance showed no difference among athletes of different sports or compared with novices, while unipedal stance has been shown to be a less stressful task for gymnasts compared to non-gymnasts or for high-level soccer players compared to low-level soccer players [1,2,10]. Garcia et al. [6] reported that gymnastics training benefits postural control of bipedal standing only in younger children and suggested that more challenging stances should be investigated. Similarly, bipedal and unipedal tests may not be sufficiently challenging to compare the postural differences that may be present in the contact and limited-contact sports. With this in mind, the use of toe-stance [20] (i.e. standing on toes), which is more challenging than unipedal and bipedal stances, may be helpful to further determine to what extent expertise in sport contributes to postural regulation.

It is suggested that the cerebellar-cortical loop is responsible for adapting postural responses based on prior experiences [21]. The effect of sport experiences on postural control will relate to how the athlete more effectively uses sensory information. For example, the somatosensory inputs involved in the perception of the support conditions may play an important role in postural control in athletes participating in contact and limited-contact sports. Furthermore, since visual input is extremely important feedback information, postural control always deteriorates in eyes-closed conditions compared with eyes-open conditions [7,22,23,24]. Thus, the aim of this study was to investigate how postural performance differs from the amount of contact in sport (soccer and baseball) across bipedal, unipedal, unipedal on foam and toe stances with both eyes-open and eyes-closed.

It is hypothesized that (1) the soccer players (contact sport) will demonstrate greater postural stability compared to the baseball players (limited contact) and controls, especially when vision is removed, and that (2) this effect will become more pronounced as the difficulty of the task increases. More generally it is hypothesized that (3) less postural stability will demonstrated as the supporting area of the task decreases, especially in eyes-closed condition, however, as hypothesized above this effect will be less pronounced in the contact sport group.

Methods

Subjects

Thirty male college students, consisting of 10 collegiate soccer players (age = 21.5±1.9, height, 171.7±2.2cm; body mass, 64.3±4.8kg), 10 collegiate baseball players (age = 19.3±1.6, height, 174.3±4.0cm; body mass, 71.83±7.4kg), and 10 male students who had no special experience of any sport (age = 22.4±1.5, height, 173.3±3.6cm; body mass, 68.83±5.8kg) were recruited. The soccer and baseball players were selected based on a minimum of 8 years competitive training, playing only one sport, and having representation at primary, middle and high school, and university in Japan. The controls were selected based on not having taken part in any competitive sport or training. None of the participants had injuries inhibiting maximal exertion or conditions likely to be aggravated by maximal exertion. All participants agreed to the experimental procedure of the study that was specifically approved by the Human Research Ethics Committee in Faculty of Sport Sciences, Waseda University.

Data collection

The participants were initially asked to stand barefoot in front of the force plate. When recording was initiated, the participants were instructed to step onto the force plate and adopt a randomly assigned posture. Once quiet balance had been achieved a trigger (including a tone) was activated. For each trial 60 seconds of data were recorded from the force plate (AMTI model OR6-5-1) which were sampled at 100 Hz. The data from the first 10 seconds after the trigger were chosen for analysis. Four tasks/postures of increasing difficulty were tested. In the first posture (bipedal stance), participants stood comfortably on both feet, separated as they desired. In the second posture (unipedal stance), participants stood on their customary supporting foot (e.g. the supporting leg when kicking a ball) while the other foot was lifted with the big toe placed alongside the medial malleolus of the supporting leg. In the third posture (unipedal_foam stance), participants stood on a 9 cm thick foam mat (16g/cm^3) placed on top of the force plate in unipedal stance as described in the second posture. In the fourth posture (toe stance), the participants stood on their customary supporting foot and raised the heel, the other foot was lifted placing the big toe alongside the medial malleolus of the supporting leg.

The participants conducted each task with conditions of eyes-open and eyes-closed. When the participants had their eyes open, they were asked to fix their gaze on a letter 'E' (font size = 72) which was placed in front of them at eye level a distance of 5 m away. During the tests with eyes-closed, participants were asked to keep their "gaze" straight ahead [25] and maintain balance. In all trials, participants were instructed to keep their body straight with their hands on their hips. Participants performed three trials for each condition, so that 24 trials were completed for each participant. A one minute rest was taken between trials and the order of the 24 trials was randomized over the participants.

Data processing

The force plate data were low pass filtered with a second-order Butterworth filter (10 Hz). The displacement of the center of pressure (COP) in the anterio-posterior (AP) and medio-lateral (ML) directions was calculated from the vertical and horizontal reaction forces. Two dependent variables were used to investigate the participants' postural behavior. The mean speed of the COP displacement (mm/s) was calculated by the sum of the displacement scalars (i.e. the cumulated distance over the sampling period) divided by the sampling time [25] using the following equation:

$$COP\ Speed = \left(\frac{1}{T}\right)\sum_{i=1}^{N}|COP_i - COP_{i-1}|$$

where T is the time duration of the series and N is the total number of points in the series. The area of the stabilogram (AOS) was calculated by taking the ratio of the major and minor axes and then fitting an ellipse that included 85% of all the trajectory points [26] (Fig 1). COP speed

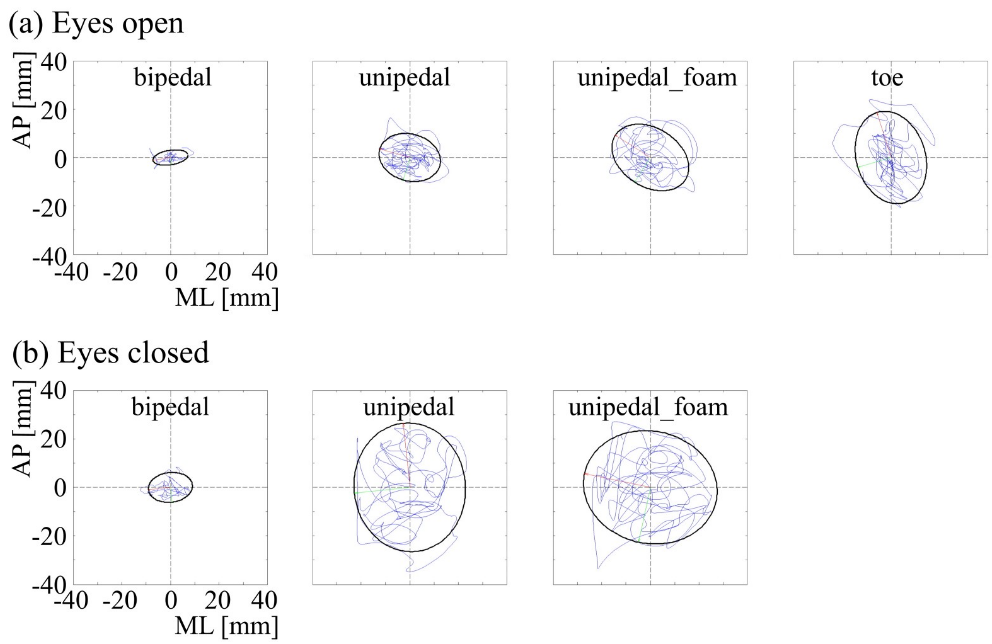


Fig 1. A typical example of the fluctuation of the COP in AP (anterior-posterior) and ML (medial-lateral) directions and the AOS as ellipse shown in a baseball subject with eyes-open and eyes-closed in the bipedal, unipedal, unipedal_foam, and toe stances.

and AOS were calculated using custom software written using MATLAB (The Mathworks Inc).

Statistical analysis

Mean values for each dependent variable were calculated across the three trials for each condition and posture. Since all participants were unable to perform toe stance with eyes-closed, and only a small number of participants in the novice group were capable of performing toe stance with eyes-open, the main analysis excluded toe stance.

The effects of the group, conditions and tasks were evaluated, 3 (groups: soccer, baseball, and novices) × 2 (conditions: eyes-open and eyes-closed) × 3 (postures: bipedal, unipedal and unipedal with foam), using three-factor analysis of variance (ANOVA) for each dependent variable. The interaction between two factors was evaluated in the simple main effects. Post hoc tests were made using t-Tests with a Bonferroni correction. In addition one-way ANOVA of expertise in toe stance with eyes-open between groups was conducted to compare the effect of group on AOS and COP speed of postural sway. The Shapiro-Wilk test was used for normality, and homogeneity of variances was investigated using Levene's test. Statistical significance was established a priori as p = 0.05 and partial eta squared (η^2) was used to calculate the effect size (small = 0.01, medium = 0.06 and large = 0.14) [27].

Results

The dependent variable of AOS revealed expertise ($F\,(2,162) = 4.460$, $p < 0.05$, $\eta^2 = 0.05$), vision ($F\,(1,162) = 174.458$, $p < 0.001$, $\eta^2 = 0.52$), and posture effects ($F\,(2,162) = 103.480$, $p < 0.001$, $\eta^2 = 0.56$) and also significant two-way interactions of expertise and vision ($F\,(2,162) = 3.714$, $p < 0.05$, $\eta^2 = 0.04$) (Fig 2), and of vision and posture ($F\,(2,162) = 41.846$,

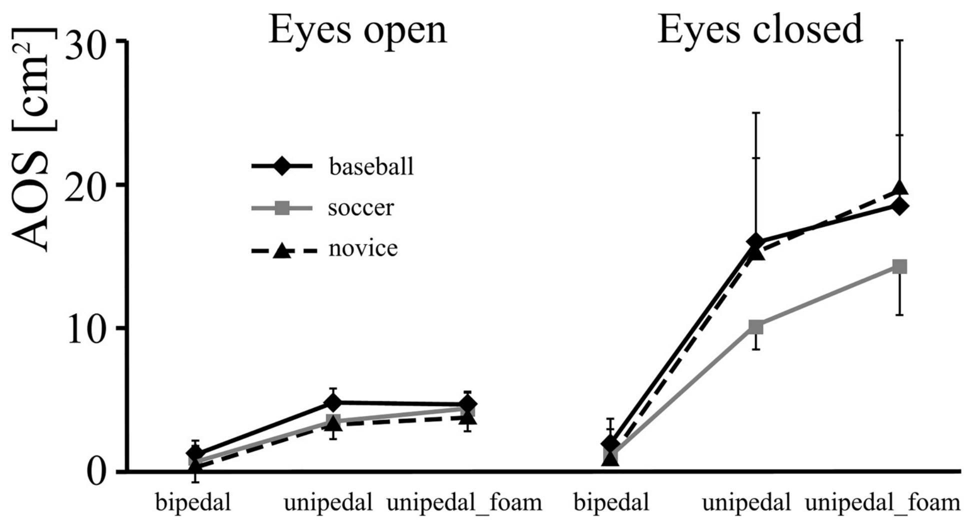


Fig 2. The AOS in the baseball, soccer and novice groups with eyes-open and eyes-closed in the bipedal stance, unipedal, and unipedal_foam stances.

$p < 0.001$, $\eta^2 = 0.34$). Soccer players had little sway in the eyes-closed condition in comparison to baseball players and novices for both unipedal and unipedal_foam. Postural sway in baseball players was comparable to that of the novices. Postural sway increased as the difficulty of posture increased only in the eyes-closed condition among the three groups.

The dependent variable of COP speed in the ML direction and AP direction revealed vision ((F (1,162) = 131.597, p<0.001, $\eta^2 = 0.55$) and (F (1,162) = 224.169, p<0.001, $\eta^2 = 0.58$), respectively), and posture effects ((F (2,162) = 142.448, p<0.001, $\eta^2 = 0.64$) and (F (2,162) = 308.479, p<0.001, $\eta^2 = 0.79$), respectively) and also significant two-way interactions of vision and posture ((F (2,162) = 29.641, p<0.001, $\eta^2 = 0.27$) and (F (2,162) = 53.235, p<0.001, $\eta^2 = 0.40$), respectively)) (Figs 3 and 4). That is, the eyes-closed condition lead to the COP

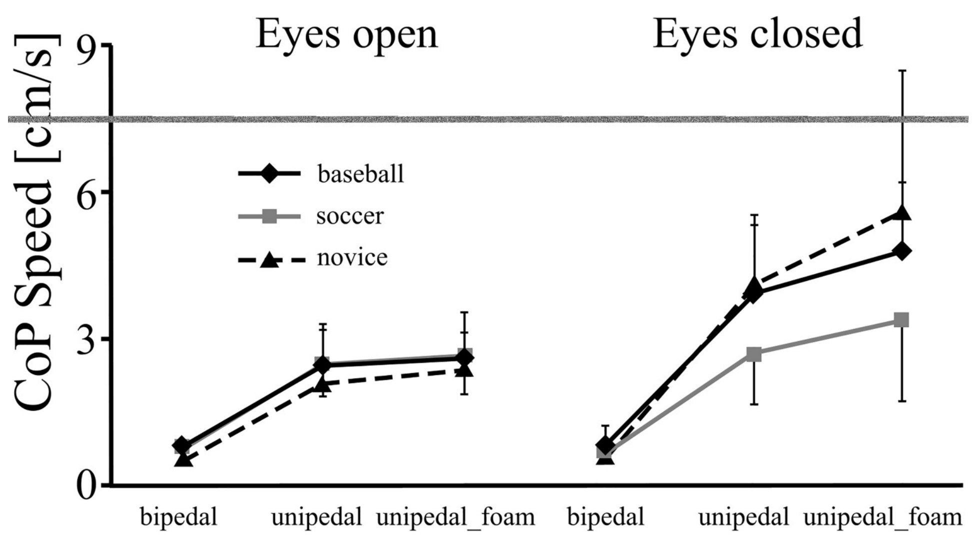


Fig 3. The COP speed in ML direction in the baseball, soccer and novice groups with eyes-open and eyes-closed in the bipedal, unipedal and unipedal_foam stances.

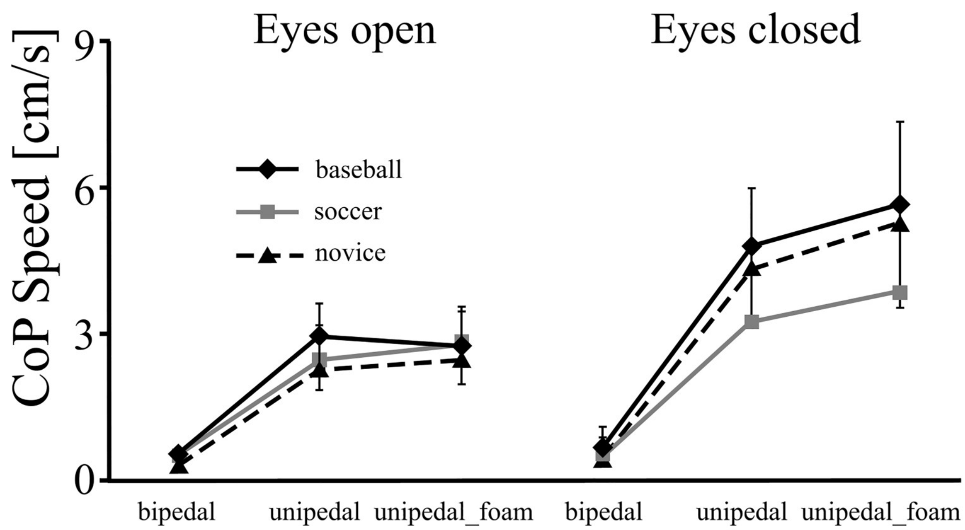


Fig 4. The COP speed in AP direction in the baseball, soccer and novice groups with eyes-open and eyes-closed in the bipedal, unipedal and unipedal_foam stances.

speed increasing more than in the eyes-opened condition only when the unipedal and unipedal_foam stances were performed.

The one-way ANOVA of expertise was made to clarify the effect of expertise with the amount of contact among three groups in the most challenging posture of toe stance. It showed there was no significant effect for expertise between the soccer, baseball and novice groups in regard to AOS of the toe stance in the eyes-open condition (F (2,24) = 0.451, $p > 0.05$, $\eta^2 = 0.04$). There was also no significant difference in COP speed in the AP direction (F (2,24) = 1.670, $p > 0.05$, $\eta^2 = 0.10$), and ML direction (F (2,24) = 1.255, $p > 0.05$, $\eta^2 = 0.12$). Only seven out of ten participants in the novice group were able to perform the posture of toe stance while all the participants from the sports groups could.

Discussion

Previous research has shown that expertise in sport results in superior postural control [14,18,19,28,29] although the effect of the amount of contact within those sports on postural control was still to be established. Additionally, across the range of available research the methods of assessing balance (field based and force plate) and the complexity of tasks (bipedal, unipedal, eyes open, eyes closed, on foam) has not been sufficiently consistent to make direct comparisons [14]. The purpose of the present study was to investigate whether an athlete's participation in a contact sport such as soccer resulted in better postural control than those who participated in limited-contact sports such as baseball or those who did not participate in any sport. A significantly lower postural sway area was found for the soccer players (contact sport) compared with baseball players (limited contact) and novices (no contact), during uni-pedal stance and uni-pedal_foam stance under the condition of no vision. This result supports the hypothesis that expertise in contact sport has a positive impact on postural control. That the baseball group were comparable with the control group is in contrast to Davlin [30] who found that expertise in sport resulted in better dynamic balance than controls. It also confirms that more challenging tests of balance than previously used [12,13] are required to determine differences between the various levels of contact in different sports. However, comparable postural performance was found amongst all three groups during toe stance. In particular none of

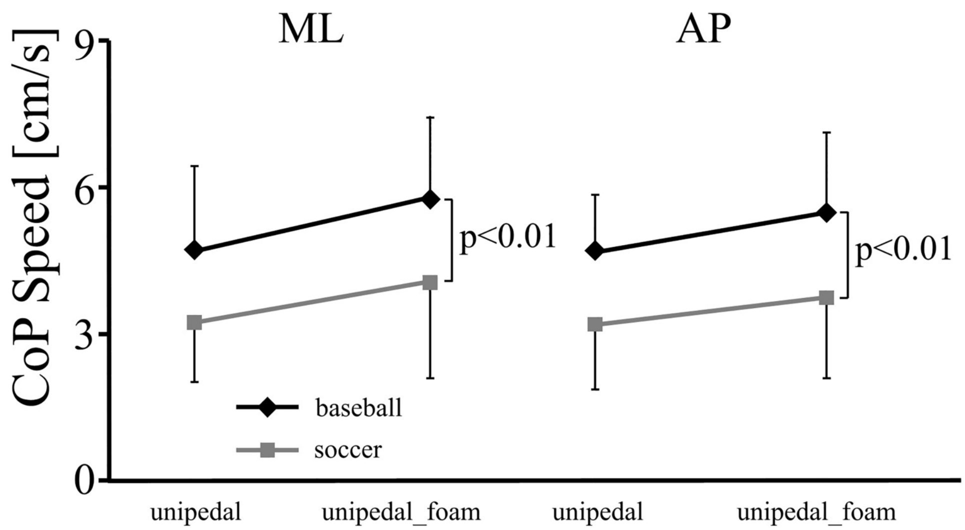


Fig 5. The COP speed in ML & AP directions in the soccer, baseball and novice groups with eyes-closed in the unipedal and unipedal_foam stances.

the participants was able to perform toe stance when vision was removed. This result only partially supports the hypothesis that as the task becomes more challenging, the benefits of expertise in contact sport become more pronounced in uni-pedal stances (Fig 3).

Increased postural stability may be developed with the diverse nature of expertise in sport, similar to being trained on an unstable compared with a stable surface [29] and requiring 'dynamic balance' as opposed to 'static balance' [19]. The aforementioned results can only show that contact experience benefits postural control in general. Whether the amount of contact experience would lead to differential effects on postural control, is still to be established. Thus, additional statistics (two-way ANOVA) were carried out in the distinguished stances of unipedal and unipedal_foam without vision, between just the soccer and baseball groups. Both AOS and COP speeds (ML and AP) in the soccer group were significantly lower than the baseball group (F (1,36) = 13.220, $p < 0.01$, $\eta^2 = 0.27$; F (1,36 = 8.915, $p < 0.01$, $\eta^2 = 0.20$; F (1,36) = 11.878, $p < 0.01$, $\eta^2 = 0.25$) (Fig 5). This, together with above results, confirmed the hypothesis that the contact group demonstrated greater postural stability than the limited-contact and non-sport groups, with special attention on the eyes-closed condition in unipedal stances. These findings have implications for the study of postural control in sport, as the level of expertise and contact experienced by the participants will have an effect on postural control. That is, care should be taken to avoid any "contact sport bias" when selecting participant groups.

It has been suggested that for healthy people the sensory contributions to quiet standing are 70% from somatosensory, 20% from vestibular and 10% from visual information [31,32]. In the present study, postural performance became worse when visual information was removed (Fig 2). It was also observed that the soccer players had significantly less dependence on vision compared to the other participants, which would suggest that they were better able to use somatosensory and vestibular information when vision was removed. This result is similar to those shown in a study on gymnasts and dancers by Vuillerme et al. [16], who suggested that gymnasts are capable of using the remaining sensory information to keep posture stable even with the loss of vision. Golomer et al. [33] and Paillard and Noe [10] presumed that soccer players and ballet dancers were able shift the sensorimotor dominance from vision to proprioception. It is interesting that specific balance skills are often practiced in gymnastics and dance, which might be expected to result in more efficient utilization of vestibular and proprioception sensory information. However, soccer is not supposed to highlight any particular balance training, but players seem more able to transfer to proprioception and vestibular systems when vision is not available, compared to baseball players and the untrained participants. Bressel et al. [34] also reported that soccer players and gymnasts did not differ in balance tests. Presumably, physical contact training works well for improving postural control. Perrin et al. [12] found that judo players, a sport which would be defined as a collision sport, performed better than dancers in bipedal stance with the eyes closed. It could be argued that judo, unlike soccer, involves specific balance training since one of the goals of the sport is to avoid being toppled by an opponent. Additionally, study [12] did not say how sports with more limited contact would compare.

Previous research has shown that soccer players demonstrate superior balance compared to basketball players and controls [11,34,35,36]. Basketball could be classed in the contact sport group, however, given that excessive contact is penalized by the referee, and the evidence from previous studies, it would fall into the limited contact group alongside such sports as baseball and squash. Again, it is difficult to directly compare results due to the limited number of conditions used and the predominantly field based testing used. Although, the present experimental designed aimed to resolve this issue, there are still limitations with the present study. The present cohort of participants was drawn from collegiate athletes and was relatively low in numbers. However, collegiate athletes have been used extensively in the literature [14,28,34,35,37] and all participants had undergone extensive training in their one sport for numerous years. While it would have been ideal to have a larger sample size, the effect sizes found in the present study were meaningful [26], and based on a review of the area [14] the majority of studies comparing balance in a variety of sports have also had comparatively small sample sizes [2,10,11,13,14,16,34,35,37,38].

What remains unclear is whether, by having a challenging stance, effects on postural control could arise from changes in the area of the base of support, support surfaces, or both. Introducing less supporting area might thus reveal a more complex phenomenon amongst subjects. However, the soccer players were not more stable in toe stance in the eyes-open condition. There are two possibilities; one is that the task of toe standing is overwhelming for all participants, which is supported by the completely failed trials in toe-standing with eyes-closed in the present study. The other is that the participants were using the control strategy (eg. ankle and hip strategies) for bipedal stance in the toe stance condition. Nolan & Kerrigan [20] concluded that despite more open loop corrections, there were no significant differences in the closed loop control between toe standing and bipedal stance. The present finding of comparable postural performance amongst the three groups in bipedal stance may indirectly support the latter possibility (Figs 2, 3 and 4). This finding is also consistent with results from previous studies where specific training experience has been shown to have a small effect on fine postural control in bipedal stance [1,2]. Bipedal stance was therefore found to be limited in revealing differences in postural stability due to transfer from particular training [39,40,41]. This is likely due to the somatosensory stimulation being below the physiological threshold leading to an intermittent process [42,43,44]. This implies that when attempting to establish difference in postural control between players of various sports it is necessary to design tests that are suitably challenging.

It may be argued that having a smaller AOS and lower COP speeds are indicative of participants who possess steady posture control within a changed environment. Results from the additional analysis support this view, with significantly less postural sway in the contact sport group compared with the limited contact group under the non-vision condition. Biec &

Kuczynski [17] proposed that soccer players exhibited different postural strategies from novices with a lower rate of postural corrections, more feedforward control and higher postural automaticity. As Deveau et al. [45] reported, specific training alters the brain so it is better able to respond to real life situations. More specifically, soccer players who were pushed off balance in such situations would react in a way that closely resembled balance training, particularly when the original balance was broken by an external disturbance. Thus, the proprioception and vestibular sensory systems are evoked and provide necessary input channels for sensory information when soccer players are working on ball control and combatting physical disturbance by an opponent. This would help explain why the soccer players are better able to cope with the loss of visual information. In that regard, contact-sport training such as in soccer may improve the proprioceptive and vestibular functions relevant for retaining balance, as sensory reweighting occurs when sensory systems change with environmental conditions [24,46].

It remains unclear whether the soccer players are better at detecting relevant sensory information or whether they are better able to respond to the information compared to the baseball players, due to having acquired different postural control strategies. According to Horak [32], there are two main types of movement strategies used to maintain balance during quiet stance, the ankle strategy and the hip strategy. It has been suggested that the former strategy is used for small perturbation in situations such as bipedal standing, and the latter strategy is used for larger perturbations, as in heel-toe standing [47]. In the same unperturbed stance while on a hard surface or with a foam support, Mesure et al. [48] confirmed that the experts with sport training preferred the ankle strategy, but the controls chose to use the hip strategy. Hence, the selection of postural strategy seems to be related to previous experience in sport, so that for a given situation the player is able to select the most appropriate strategy in order to respond to the perturbation [32,49,50].

Conclusion

Participants involved in sport with physical contact (soccer) were found to have superior postural control compared with participants involved in sport with limited contact (baseball). This was particularly evident during the more challenging unipedal stance. Routine participation in sport involving physical contact appears to be an effective method for training proprioceptive and vestibular plasticity to posture control, particularly when vision is lacking.

Supporting information

S1 File. Sway area and sway speed data for all participants.
(CSV)

Author Contributions

Conceptualization: Ying Liang, Michael Hiley, Kazuyuki Kanosue.

Data curation: Ying Liang.

Formal analysis: Ying Liang, Kazuyuki Kanosue.

Funding acquisition: Ying Liang, Kazuyuki Kanosue.

Investigation: Ying Liang, Kazuyuki Kanosue.

Methodology: Ying Liang, Michael Hiley, Kazuyuki Kanosue.

Project administration: Ying Liang.

Resources: Kazuyuki Kanosue.

Supervision: Michael Hiley, Kazuyuki Kanosue.

Writing – original draft: Ying Liang, Michael Hiley, Kazuyuki Kanosue.

Writing – review & editing: Ying Liang, Michael Hiley, Kazuyuki Kanosue.

References

1. Vuillerme N, Nougier V. Attentional demand for regulating postural sway: the effect of expertise in gymnastics. Brain Res Bull 2004; 63: 161–165. https://doi.org/10.1016/j.brainresbull.2004.02.006 PMID: 15130706
2. Asseman FB, Caron O, Crémieux J. Are there specific conditions for which expertise in gymnastics could have an effect on postural control and performance? Gait Posture 2008; 27: 76–81. https://doi.org/10.1016/j.gaitpost.2007.01.004 PMID: 17337190
3. Gautier G, Thouvarecq R, Larue J. Influence of experience on postural control: effect of expertise in gymnastics. J Motor Behav 2008; 40: 400–408.
4. Gautier G, Thouvarecq R, Vuillerme N. Postural control and perceptive configuration: influence of expertise in gymnastics. Gait Posture 2008; 28: 46–51. https://doi.org/10.1016/j.gaitpost.2007.09.007 PMID: 17976990
5. Bruyneel V, Mesure S, Paré JC, Bertrand M. Organization of postural equilibrium in several planes in ballet dancers. Neurosci Lett 2010; 485: 228–32. https://doi.org/10.1016/j.neulet.2010.09.017 PMID: 20849927
6. Garcia C, Barela JA, Viana AR, Barela AMF. Influence of gymnastics training on the development of postural control. Neurosci Lett 2011; 492: 29–32. https://doi.org/10.1016/j.neulet.2011.01.047 PMID: 21276829
7. Massion J. Postural control system. Curr Opin Neurobiol 1994; 4: 877–87. PMID: 7888772
8. Stoffregen TA, Riccio GE. An ecological theory of orientation and the vestibular system. Psychol Rev 1988; 95: 3–14. PMID: 3281178
9. Bardy BG, Marin L, Stoffregen TA, Bootsma RJ. Postural coordination modes considered as emergent phenomena. J Exp Psychol Human 1999; 25: 1284–301.
10. Paillard T, Noé F. Effect of expertise and visual contribution on postural control in soccer. Scand J Med Sci Sports 2006; 16: 345–348. https://doi.org/10.1111/j.1600-0838.2005.00502.x PMID: 16978254
11. Matsuda S, Demura S, Nagasawa Y. Static one-legged balance in soccer players during use of a lifted leg. Percept Motor Skill 2010; 111: 167–177.
12. Patti A, Messina G, Palma R, Barcellona M, Brusa J, Iovane A, et al. Comparison of posturographic parameters between young taekwondo and tennis athletes. J Phys Ther Sci 2018; 30: 1052–1055. https://doi.org/10.1589/jpts.30.1052 PMID: 30154599
13. Perrin P, Deviterne D, Hugel F, Perrot C. Judo, better than dance, develops sensorimotor adaptabilities involved in balance control. Gait Posture 2002; 15: 187–194. PMID: 11869913
14. Hrysomallis C. Balance ability and athletic performance. Sport Med 2011; 41: 221–231.
15. Rice SG. Medical Conditions Affecting Sports Participation. Pediatrics 2001; 107: 1205–1209. PMID: 11331710
16. Vuillerme N, Danion F, Marin L, Boyadjian A, Prieur JM, Weise I, et al. The effect of expertise in gymnastics on postural control. Neurosci Lett 2001; 303: 83–86. PMID: 11311498
17. Bieć E, Kuczyński M. Postural control in 13-year-old soccer players. Eur J Appl Physiol 2010; 110: 703–8. https://doi.org/10.1007/s00421-010-1551-2 PMID: 20582432
18. Marchetti PH, Hartigan EH, Duarte M. Comparison of the Postural Control Performance of Collegiate Basketball Players and Nonathletes. Athl Train Sports Heal Care 2012; 4: 251–256.
19. Negahban H, Aryan N, Mazaheri M, Norasteh AA, Sanjari MA. Effect of expertise in shooting and Taekwondo on bipedal and unipedal postural control isolated or concurrent with a reaction-time task. Gait Posture 2013; 38: 226–30. https://doi.org/10.1016/j.gaitpost.2012.11.016 PMID: 23245642
20. Nolan L, Kerrigan DC. Postural control: toe-standing versus heel—toe standing. Gait Posture 2004; 19: 11–15. PMID: 14741299
21. Jacobs JV, Horak FB. Cortical control of postural responses. J Neural Transm 2007; 114: 1339–48. https://doi.org/10.1007/s00702-007-0657-0 PMID: 17393068

22. Dichgans J. Moving visual scenes influence the apparent direction of gravity. Science 1972; 178: 1217–1219. PMID: 4637810

23. Lestienne F, Soechting J, Berthoz A. Postural readjustments induced by linear motion of visual scenes. Exp Brain Res 1977; 28: 363–384. PMID: 885185

24. Nashner L, Berthoz A. Visual contribution to rapid motor responses during postural control. Brain Res 1978; 150: 403–407. PMID: 678978

25. Ivanenko YP, Talis VL, Kazennikov OV. Support stability influences postural responses to muscle vibration in humans. Eur J Neurosci 1999; 11: 647–54. PMID: 10051765

26. Geurts C, Nienhuis B, Mulder TW. Intrasubject variability of selected force-platform parameters in the quantification of postural control. Arch Phys Med and Rehabil 1993; 74: 1144–1150.

27. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic; 1988.

28. Gerbino PG, Griffin ED, Zurakowski D. Comparison of standing balance between female collegiate dancers and soccer players. Gait Posture 2007; 26: 501–507. https://doi.org/10.1016/j.gaitpost.2006.11.205 PMID: 17197186

29. Williams DSB, Murray NG, Powell DW. Athletes who train on unstable compared to stable surfaces exhibit unique postural control strategies in response to balance perturbations. J Sport Heal Sci 2016; 5: 70–76.

30. Davlin CD. Dynamic balance in high level athletes. Percept Motor Skill 2004; 98: 1171–1176.

31. Peterka RJ. Sensorimotor Integration in Human Postural Control. J Neurophysiol 2002; 88: 1097–1118. https://doi.org/10.1152/jn.2002.88.3.1097 PMID: 12205132

32. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? Age Ageing 2006; 35: ii7–ii11. https://doi.org/10.1093/ageing/afl077 PMID: 16926210

33. Golomer E, Cre J, Dupui P, Isableu B. Visual contribution to self-induced body sway frequencies and visual perception of male professional dancers. Neurosci Lett 1999; 267: 189–192. PMID: 10381008

34. Bressel E, Yonker JC, Kras J, Heath EM. Comparison of static and dynamic balance in female collegiate soccer, basketball, and gymnastics athletes. J Athl Train 2007; 42: 42–6. PMID: 17597942

35. Thorpe JL, Ebersole KT. Unilateral balance performance in female collegiate soccer athletes. J Strength Cond Res 2008; 22: 1429–33. https://doi.org/10.1519/JSC.0b013e31818202db PMID: 18714247

36. Feizolahi F, Azarbayjani M. Comparison of static and dynamic balance in amateur male athletes. Sci J Rehab Med 2015; 3: 89–98.

37. Schmit JM, Regis DI, Riley MA. Dynamic patterns of postural sway in ballet dancers and track athletes. Exp Brain Res 2005; 163: 370–8. https://doi.org/10.1007/s00221-004-2185-6 PMID: 15655686

38. Paillard T, Noe F, Riviere T, Marion V, Montoya R, Dupui P. Postural performance and strategy in the unipedal stance of soccer players at different levels of competition. J Ath Train 2006; 41: 172–6.

39. Hugel F, Cadopi M, Kohler F, Perrin PH. Postural Control of Ballet Dancers. Physiol Biochem 1999; 20: 86–92.

40. Kiefer AW, Riley M, Shockley K, Sitton C, Hewett TE, Cummins-Sebree S, et al. Multi-segmental postural coordination in professional ballet dancers. Gait Posture 2011; 34: 76–80. https://doi.org/10.1016/j.gaitpost.2011.03.016 PMID: 21530267

41. Casabona A, Leonardi G, Aimola E, La Grua G, Polizzi CM, Cioni M, et al. Specificity of foot configuration during bipedal stance in ballet dancers. Gait Posture 2016; 46: 91–97. https://doi.org/10.1016/j.gaitpost.2016.02.019 PMID: 27131184

42. Bottaro A, Casadio M, Morasso PG, Sanguineti V. Body sway during quiet standing: Is it the residual chattering of an intermittent stabilization process? Hum Mov Sci 2005; 24: 588–615. https://doi.org/10.1016/j.humov.2005.07.006 PMID: 16143414

43. Bottaro A, Yasutake Y, Nomura T, Casadio M, Morasso P. Bounded stability of the quiet standing posture: An intermittent control model. Hum Mov Sci 2008; 27: 473–495. https://doi.org/10.1016/j.humov.2007.11.005 PMID: 18342382

44. Blenkinsop GM, Pain MTG, Hiley MJ. Evaluating feedback time delay during perturbed and unperturbed balance in handstand. Hum Mov Sci 2016; 48: 112–120. https://doi.org/10.1016/j.humov.2016.04.011 PMID: 27155963

45. Deveau J, Ozer DJ, Seitz AR. Improved vision and on-field performance in baseball through perceptual learning. Curr Biol 2014; 24: R146–7. https://doi.org/10.1016/j.cub.2014.01.004 PMID: 24556432

46. Assländer L, Peterka RJ. Sensory reweighting dynamics in human postural control. J Neurophysiol 2014; 111: 1852–1864. https://doi.org/10.1152/jn.00669.2013 PMID: 24501263

47. Horak FB. Clinical Measurement of Postural Control in Adults. Phys Ther 1987; 67: 1881–1885. PMID: 3685116

48. Mesure S, Amblard B, Crémieux J. Effect of physical training on head-hip coordinated movements during unperturbed stance. Neuroreport 1997; 8: 3507–3512. PMID: 9427316

49. Burleigh AL, Horak FB, Malouin F. Modification of postural responses and step initiation: Evidence for goal-directed postural interactions. J Neurophysiol 1994; 72: 2892–2902. https://doi.org/10.1152/jn.1994.72.6.2892 PMID: 7897497

50. Runge CF, Shupert CL, Horak FB, Zajac FE. Ankle and hip postural strategies defined by joint torques. Gait Posture 1999; 10: 161–170. PMID: 10502650

6

Biomechanics of fencing sport: A scoping review

Tony Lin-Wei Chen[1], Duo Wai-Chi Wong[1,2], Yan Wang[1,2], Sicong Ren[1], Fei Yan[1,3], Ming Zhang[1,2]*

1 Interdisciplinary Division of Biomedical Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China, **2** The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, Guangdong, China, **3** Department of Applied Mechanics, Sichuan University, Chengdu, Sichuan, China

* ming.zhang@polyu.edu.hk

Abstract

Objectives

The aim of our scoping review was to identify and summarize current evidence on the biomechanics of fencing to inform athlete development and injury prevention.

Design

Scoping review.

Method

Peer-reviewed research was identified from electronic databases using a structured keyword search. Details regarding experimental design, study group characteristics and measured outcomes were extracted from retrieved studies, summarized and information regrouped under themes for analysis. The methodological quality of the evidence was evaluated.

Results

Thirty-seven peer-reviewed studies were retrieved, the majority being observational studies conducted with experienced and elite athletes. The methodological quality of the evidence was “fair” due to the limited scope of research. Male fencers were the prevalent group studied, with the lunge and use of a foil weapon being the principal movement evaluated. Motion capture and pedabarography were the most frequently used data collection techniques.

Conclusions

Elite fencers exhibited sequential coordination of upper and lower limb movements with coherent patterns of muscle activation, compared to novice fencers. These elite features of neuromuscular coordination resulted in higher magnitudes of forward linear velocity of the body center of mass and weapon. Training should focus on explosive power. Sex- and equipment-specific effects could not be evaluated based on available research.

Editor: Tiago M Barbosa, Nanyang Technological University, SINGAPORE

Competing interests: The authors have declared that no competing interests exist.

1. Introduction

Modern fencing emerged as a competitive sport in Europe and is now a well-recognized Olympic sport, with over 150 member federations [1]. Both the sport and the culture of fencing have progressed significantly over the past decades, with an estimated 22,000 participants in the United States in 2006 [2] and 25,000 in Germany in 2008 [3]. The dressing culture and fighting traditions until the 19th century are likely to have contributed to the promotion of this combat sport [4].

Owing to its unique asymmetry in movement, fencing imposes high physiological demands in terms of neuromuscular coordination, strength and power, and the impact on the musculoskeletal system [5]. As an example, for the basic 'on-guard' stance, fencers align their leading foot with their opponent's position, with the trailing foot placed at an angle to the lead foot for stability [6]. To score against their opponent, fencers must thrust their weapon quickly toward their opponent, which requires an explosive extension of the trailing leg to perform a forceful forward lunge. These quick 'propulsion' and 'dodge' offense/defense movements further expose fencers to impacts, explosive forces, power absorption, and shear forces of varying magnitude, asymmetrically distributed across the body [7].

Resulted from this dynamic and repetitive movements in fencing matches, fencing injuries were quite prevalent among the athletes. In spite of the rare cases of severe trauma caused by penetration (puncture by broken blades, account for 2.7–3.2%) [2, 8], most of the fencing injuries arise from overuse. In a 5-year survey by the USFA [2], 184 cases of time-loss injuries were recorded for 610 exposures with an overall 30.0% of injury rate. Approximately 52% of all reportable injuries were first or second-degree strains and sprains. Lower limb is most susceptible to injuries. The injury rates were 19.6%, 15.2%, and 13.0% respectively for the knee, thigh, and ankle. These injuries also carry a high risk of chronic morbidity, predominantly achillodynia and patellofemoral pain [9]. Understanding the biomechanics and demands of a sport provides a pathway to injury prevention and safety promotion [10]. An analysis of the biomechanics of a sport can also improve athletes' skills, tactics and overall performance and competitiveness.

Currently for fencing, biomechanics of performance have been investigated for different movement components of the offensive and defensive manoeuvres and using varying methodologies, which has made interpretation of findings for practice difficult. Therefore, our aim was to perform a scoping review to identify, evaluate and summarize current evidence on the biomechanics of fencing to inform athlete development and injury prevention.

2. Methods

2.1 Search strategy and study selection

The research was approved by The Human Subject Ethics Sub-committee of The Hong Kong Polytechnic University. The reference number is HSEARS20150814001. As electronic search of five databases was conducted (PubMed, EBSCOhost, Wiley, Web of Science and Google Scholar), using a pre-defined keyword combination (fencing AND (biomechanics OR kinematics OR kinetics OR dynamics OR movements OR performance)) to identify relevant research published in English.

Publication time was not restricted. Nighty-seven articles were identified after duplication removal and screened for eligibility. Inclusion criteria were 1) studies that addressed fencers' neuromusculoskeletal features and the biomechanics of fencing movements; 2) studies that examined the performance of fencing-specific equipment and training strategy. Studies were excluded if they 1) did not involve human subjects; 2) did not provide numeric results; 3) recruited subjects for sports other than fencing. Literature search was performed on between March 3rd to March 11th, 2016.

During the article screening, titles and abstracts of identified studies were reviewed, independently, by the first two authors to ensure that studies were experimental in nature and addressed the biomechanics of fencing. Papers for retained titles were retrieved for full review to confirm relevance to the aim of our scoping review, as well as to extract required data for analysis: experimental setting and design, characteristics of the study group, sample size, and measured outcomes. Data extraction was done independently by two authors (DWW and YW) of this study. Any inconsistency in the results was solved by group discussion involving a third author (MZ). Based on these summaries of available research evidence, three emergent themes were identified and used to organize our data for analysis: (1) intrinsic, athlete-specific, factors; (2) extrinsic factors; and (3) basic biomechanics.

2.2 Quality assessment

Quality of the recruited studies was assessed by two authors (TLC and DWW) using the tool developed by the Effective Public Health Practice Project [11]. Each of following components was rated: selection bias (the likelihood that the selected subjects can represent the target population); study design (the bias resulted from allocation and the independence of exposure and outcomes); confounders (the inter-group imbalance associated with variables that influence intervention or exposure); blinding (concealment of subject allocation and outcome assessment); data collection method (the validity and reliability of outcome measurement); withdrawals and drop-outs; intervention integrity (the percentage of subjects received complete intervention and reports of unintended intervention); analysis appropriate to question (correct statistics and intention-to-treat analysis). A score of 'strong', 'moderate', and 'weak' was assigned to each study according to existing standard [11]. If consensus was not reached, a third author (MZ) made the final decision.

3. Results

3.1 Search results

The retrieve results are summarized in Fig 1. We identified 548 studies, with 37 retained for analysis. Among the retained studies, 24 examined the lunge manoeuvre, which was considered to be the core component of fencing (Fig 2). Nine studies did not specify the fencing manoeuvre (Table 1). The biomechanics of the lower limbs was evaluated in 27 studies, and the biomechanics of the upper limbs in 15. The majority of studies were conducted in Europe (70.3%), with three studies conducted in the United States. Expert/elite fencing athletes were the major components of research subjects, which increases the difficulty of enlarging sample size for the recruited studies because top athletes are always rare. All three types of fencing weapons were included in these studies-foils, épées, and sabers. Foils were addressed in 16 studies, épées in 10 and sabers in 6. All studies were lab-based experiments except one performed measurements during competitions (video footage) [12]. Measurements during competitions could provide valuable information of high standard game and athletes. However, the video-based analysis could not quantify fencing biomechanics as accurately as in-lab 3D motion capture technique.

3.2 Sample characteristics

Relevant characteristics of the study groups in the 37 retained studies are summarized in Table 1. Male fencers were included to a higher extent than female fencers, overall, and sex-

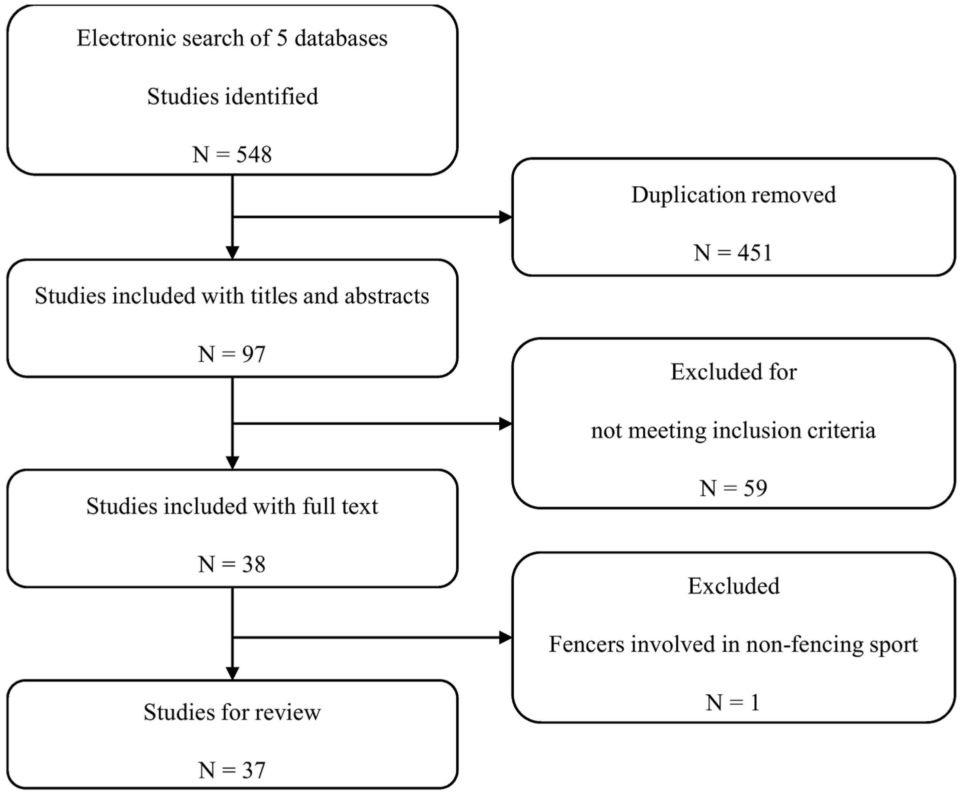


Fig 1. Flow diagram of the search strategy and screening of identified research for inclusion.

specific effects were not typically addressed. The Body Mass Index (BMI) of fencers was generally within normal limits or slightly lower (23.85kg/m^2 in fencers VS 24.95kg/m^2 in the untrained group) because fencers were commonly taller (1.83m VS 1.69m) than the general population [13]. Ethnicity, which would potentially influence anthropometry, was seldom reported, limiting the generalizability of findings. Moreover, two-thirds of the studies enrolled experienced or elite fencers, which further limits the application of findings in developing comprehensive programs for athlete development and injury prevention (in this study, fencing injury refers to those injury types associated with overuse in fencing sports).

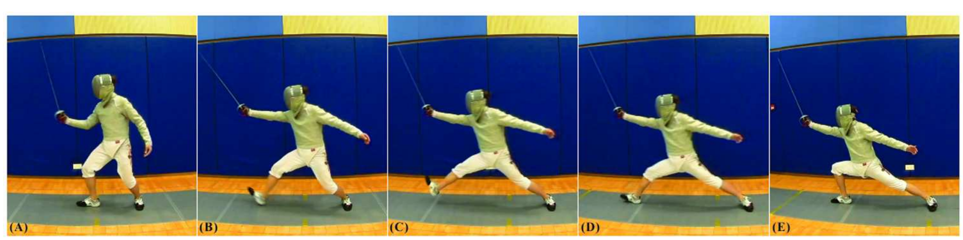


Fig 2. Sequence of movement for lunging using a saber fencer. (A) on-guard position; (B) lifting of the lead leg; (C) forward flying phase of the lead leg and push-off with the trail leg; (D) landing of the lead foot with the armed upper limb in full extension; and (E) final lunge position.

Table 1. Basic characteristics of the participants in studies included in our analysis.

Study	Reference number	Location	Subject	Handedness	Age (year)	Mass (kg)	Height (m)	BMI (kg/m^2)	Weapon	Experience and level
Akpinar et al., 2015	38	United States	4 males 4 females	Right	21.5	N/A	N/A	N/A	N/A	8.8 years on average
Aquili et al., 2013	14	Italy	32 males 25 females	N/A	25.7	71.2	1.79	22.2	Saber	World class
Bottoms et al., 2013	16	United Kingdom	9 males 5 females	Right	26.2	75.6	1.76	24.4	Epee	At least 3 years
Chang et al., 2009	22	Taiwan	8 males	N/A	19.0	65.7	1.73	22.0	Foil	At least 3 years
Cronin et al., 2003	31	New Zealand	21 males	N/A	23.1	76.3	1.76	24.6	N/A	Not fencer athletes
Do and Yiou, 1999	42	France	5 males	Right	25.0	66.0	1.72	22.3	Foil	Not fencer athletes
Frère et al., 2011	17	France	8 males	6 right 2 left	23.0	77.2	1.83	23.1	Epee	10.5 years on average
										National team
Geil, 2002	20	United States	9 males 4 females	N/A	N/A	N/A	N/A	N/A	N/A	3 to 27 years
Gholipour et al., 2008	25	Iran	8 males	Right	22.8	72.2	1.80	22.3	Foil	Novice fencer: college team
										Elite fencer: national team
Greenhalgh et al., 2013	27	United Kingdom	6 males 7 females	N/A	32.4	74.4	1.78	23.5	Foil	At least 3 years
Gresham-Fiegel et al., 2013	32	United States	12 males 13 females	N/A	20.3	N/A	N/A	N/A	N/A	National team
Guilhem et al., 2014	28	France	10 females	N/A	22.2	67.3	1.70	23.3	Saber	National team
Gutierrez-Davila et al., 2013	33	Spain	30 subjects	N/A	35.2	82	1.79	25.6	Epee	National team
Gutierrez-Davila et al., 2013	29	Spain	30 subjects	N/A	29.6	79.8	1.80	24.6	Foil	National team
Hassan and Klauck, 1998	18	Germany	4 females	N/A	16–17	N/A	N/A	N/A	Foil	At least 5 years
Irurtia et al., 2008	26	Spain	16 males 7 females	N/A	18.7	68.9	1.79	21.5	N/A	National team
Kim et al., 2015	36	Korea	9 males	Right	28.2	76.5	1.82	23.1	4 Epee	National team
									5 Saber	
Lin et al., 2010	21	Taiwan	12 males 3 females	Right	19.2	63.2	1.68	22.4	Foil	At least 3 years
Margonato et al, 1994	23	Italy	58 males	N/A	23.0	71.9	1.78	22.7	Foil	11.4 years on average
Morris et al., 2011	15	Canada	1 subject	Right	N/A	N/A	N/A	N/A	Foil	N/A
Mulloy et al., 2015	37	United Kingdom	10 subjects	N/A	22.8	73.6	1.79	23.0	N/A	Novice fencer: at least 1-year experience
										Elite fencer: regional team
Poulis et al., 2009	48	Greece	16 males 14 females	N/A	18.2	62.7	1.73	20.9	N/A	National team
Sinclair and Bottoms, 2015	45	United Kingdom	8 males 8 females	Right	26.1	69.5	1.73	23.2	Epee	Minimum of 3 training sessions per week
Sinclair et al., 2010	7	United Kingdom	19 males	17 right 2 left	25.6	76.8	1.78	24.2	2 Foil	At least 2 years
									16 Epee	
									1 Saber	

(*Continued*)

Table 1. (*Continued*)

Study	Reference number	Location	Subject	Handedness	Age (year)	Mass (kg)	Height (m)	BMI (kg/m^2)	Weapon	Experience and level
Sinclair and Bottoms, 2013	30	United Kingdom	8 males 8 females	Right	26.1	69.5	1.73	23.2	Epee	Minimum of 3 training sessions per week
Sterkowicz-Przybycień, 2009	13	Poland	30 subjects	N/A	23.3	79.1	1.83	23.6	10 Foil 10 Epee 10 Saber	World class
Steward and Kopetka, 2005	24	United Kingdom	15 subjects	Right	N/A	N/A	N/A	N/A	Epee	N/A
Trautmann et al., 2011	3	Germany	20 males 9 females	N/A	19.3	70.8	1.76	22.9	N/A	National team
Tsolakis et al., 2006	51	Greece	84 males 68 females	N/A	17.0	59.1	1.67	21.2	N/A	Age-based International ranking
Tsolakis et al., 2006	49	Greece	8 males	N/A	12.3	N/A	N/A	N/A	N/A	At least 1 year
Tsolakis et al., 2010	34	Greece	15 males 18 females	N/A	19.9	66.1	1.75	21.6	N/A	National team
Turner et al., 2016	47	United Kingdom	49 males 21 females	N/A	16.8	68.2	1.78	21.5	21 Foil 30 Epee 19 Saber	National team
Williams and Walmsley, 2000	67	New Zealand	5 males 1 female	N/A	19–26	N/A	N/A	N/A	Foil	Elite fencer: world class Novice fencer: local club
Williams and Walmsley, 2000	39	New Zealand	5 males 1 female	N/A	19–26	N/A	N/A	N/A	Foil	Elite fencer: world class Novice fencer: local club
Wylde, 2013	12	United Kingdom	9 males	Right	27.4	89.4	1.73	29.9	Foil	Elite fencer: world class Novice fencer: not previous fencing experience
Yiou and Do, 2000	40	France	11 males	Right	29.2	72.1	1.71	24.7	Foil	N/A
Yiou and Do, 2001	41	France	11 males	Right	31.0	73.0	1.70	N/A	Foil	N/A

Notes: N/A, not available.

3.3 Methodological quality

The EPHPP tool assesses many important aspects of the research quality that are critical for studies in public health and injury prevention. The tool is also commonly used by professionals of various topic areas to facilitate their decision-making based on high-quality evidence. In this study, the inter-rater reliability for EPHPP was 0.87 (Cohen's kappa coefficient), indicating a good agreement between the two reviewers. As showed in Tables 2 and 3, eleven studies (29.7%) were rated as high-quality studies (strong), eight (21.6%) were rated 'weak', and the remaining eighteen (48.7%) had 'moderate' quality. For most of the recruited studies, they were less scored due to the potential bias present in research design, confounders, and sample selection. Based on the descriptions in the paper, only five studies (13.5%) were randomized controlled trial. However, they were all classified as controlled clinical trial in EPHPP assessment because none of them clarifies the method of randomization in the text. The majority of the studies had case-control design (20 studies, 54.1%) while ten were cross-sectional studies

Table 2. Methodological characteristics of studies included in our analysis.

Study	Reference number	Objective	Moves	Measurement and Equipment	Variable of interest	Results
Akpinar et al., 2015	38	Examined the handedness and performance asymmetries in fencers	N/A	Motion capture (Flock of Birds)	Movement speed Movement time Movement accuracy (point error)	As compared to fencers, non-fencers showed greater inter-limb differences in error making and pointing path deviation under the non-choice condition (p<0.01).
					Movement quality (path deviation from linearity)	Fencers used less right arm to reach middle and left regions under the choice condition (17.0%-23.5% less).
Aquili et al., 2013	14	Time-motion characteristics of saber fencing	Complete bout	Motion capture (Casio & Dartfish)	Time-motion parameters (the type and quantity of actions during the bouts)	There were gender differences in saber data. Males were faster and more frequent in attacking (action/break ratio: 1:6.5 VS 1:5.1, lunge frequency: 23.9/s VS 20.0/s, time of direction change: 65.3s VS 59.7s).
						In both sexes, the percentage of offensive action (49% - 55%) was higher than for defensive action (26% - 31%). The number of lunges was high compared to the number of changes in direction.
Bottoms et al., 2013	16	Identified the kinematic determinants of lunge performance	Lunge	Motion capture (Qualisys)	Kinematics Regression analysis for lunge performance	The average sword velocity was 12.8±3.3m/s, Knee range of motion (30.7˚±10.7˚) and peak hip flexion of the trailing leg (9.7˚±10.9˚), and peak hip flexion of the leading leg (102.0˚±13.0˚) were significant predictors of sword velocity (R^2 = 0.14–0.36, p<0.01).
Chang et al., 2009	22	Determine appropriate foil handle shape which could reduce the load on grip force	Quarte sixte	EMG measurement (Biometrics)	Muscle activity	The defensive position has no significant effects on muscle activity (p = 0.39). Activity of the adductor pollicis and the extensor carpi radialis (average value = 0.16±0.02) was significantly lower when using the Poistol-Viscounti in comparison to other types of handles.
Cronin et al., 2003	31	Identified the strength qualities predictive of lunge performance	Lunge	Squat performance (Fitness Works)	Squat strength and velocity Explosive strength	The average lunge velocity was 1.62±0.21m/s (concentric velocity), The best three predictors of lunge velocity (R^2 = 0.85, p<0.05) were time to peak squat force (0.48±0.07s), leg length (83.9 ±5.2cm), and flexibility (171.0±12.5cm).
				Lunge test (Unimeasure)	Regression analysis for lunge performance	When lunge velocity was normalized by body mass, the three best predictors (R^2 = 0.87, p<0.05) were time to peak squat force, mean squat power (364.0±96.8W) and relative squat strength (1.65 ±0.32kg/body mass).
Do and Yiou, 1999	42	Examine the effects of anticipatory postural adjustments on fencing speed	Touche	Force plate (Unspecified) Accelerometer (Entran)	Displacement of the foot Acceleration of the foil	In lunge + touche condition, when touche was initiated (onset of anterior deltoid) before the postural adjustment of lunge (200ms prior to foot off), touche speed was comparable to that in the isolated touche condition.
			Lunge	EMG measurement (Unspecified)	Muscle activity	When touche was initiated during postural adjustment, touche speed dropped. The average touche speed was significantly lower when executed at the time of foot off (2.19±0.52m/s VS 2.54±0.44m/s, p<0.01) than that in the isolated touche condition.
Frère et al., 2011	17	Classify fencers based on kinematics and muscular activation pattern	Fleche	Motion capture (Vicon)	Kinematics	Experienced and elite fencers did not differ significantly in their anthropometries.
				EMG measurement (Noraxon)	Muscle activity	Fencers were firstly sorted into two groups based on the timing of maximal elbow extension (MEE, early group: 0.20±0.06s, late group: 0.47±0.03s).
						Further EMG-based classification was performed to the two groups. The results showed that early MEE group exhibited higher deltoid intensity (91 ±18%) than late MEE group (36±13%) in attacking (p<0.05). Spherical classification confirmed that muscular activity was different based on the strategies used in the two groups.

(*Continued*)

Table 2. (*Continued*)

Study	Reference number	Objective	Moves	Measurement and Equipment	Variable of interest	Results
Geil, 2002	20	Effects of different footwear on plantar pressure	Advance	Motion capture (Peak Performance)	Kinematics	The court shoes significantly reduced plantar pressure by 15.37–26.38% as compared to the fencing shoes in all fencing movements.
			Lunge	Plantar pressure (Pedar)	Plantar pressure	Pressures were consistently higher at the front foot. The major pressured regions are the front heel and back medial forefoot (average pressure normalized by body mass: 0.0611–0.0862N/kg·cm^2).
			Fleche			The court shoes altered fencers' kinematics by increasing the range of motion of the weapon hand. The consistency of repeated movement and lunge velocity also reduced in course shoes condition.
Gholipour et al., 2008	25	Compared the kinematics of fencers at different levels	Lunge	Motion capture (Kinemetrix)	Kinematics	Elite fencers had a higher mean lunge length (1.17 ±0.17m VS 1.02±0.10m), larger late-phase knee extension (51±9˚ VS 18±8˚), and shorter time gap in hand/foot motion (0.07±0.05s VS 0.13±0.15s) in comparison to the novice fencers.
Greenhalgh et al., 2013	27	Effects of sports surface on impact shock during a fencing movement	Lunge	Accelerometer (Biometrics)	Impact shock	Significantly larger impact shock magnitude (F = 17.07, p<0.001) was identified during a lunge on the concrete-based surface (14.9±8.5g) compared with the wooden-based surface (range: 11.1–12.0g). Use of a 'piste' had no significant effect on the overall impact shock magnitude (p = 0.38–0.69).
Gresham-Fiegel et al., 2013	32	Effects of trail leg displacement angle on lunge performance	Lunge	Power measurement (TENDO Weightlifting Analyzer)	Lunge power and velocity	For all fencers, their natural trail leg displacement angles were 68˚ to 100˚, 60% of them had a forward deviation, 12% had a perpendicular stance, and 28% had a backward deviation.
						A perpendicular placement (90˚) of the feet produced the greatest average power (411.1 ±97.8W) and velocity (0.59±0.11m/s) during lunging, while forward deviation (45˚) produced the lowest values (336.8±70.2W in power, 0.49 ±0.09m/s in velocity).
Guilhem et al., 2014	28	Investigated the coordination of the leg muscles in fencing execution	Advance lunge	Dynamometer (CMV)	Muscle strength and activity	Concentric contraction tests showed that peak torque produced by hip extensors (221.1 ±64.0N·m) and knee extensors (173.4±33.9 N·m) were significantly greater in the front leg than the rear leg. The front ankle dorsiflexor torque was 20% stronger than the rear leg on the whole range of motion.
				EMG measurement (Zerowire)	Lunge displacement	
				Force plate (Kistler)		E The fencers reached their peak velocity (2.6 ±0.9m/s) at the early phase 4 of lunge, while peak acceleration (6.5±0.9m/s), force (469.6±77.4N), and power (1051.8±231.5W) occurred in the middle of phase 3.
						Knee extensors of the trailing leg were mainly activated (25.3%-35.1% more than the front leg) during propulsive phases, and less activated than that of the front leg (10.4% more than the trailing leg) during the braking phase.
						Hip extensors of the leading leg were mainly activated (54.1% more than the trailing leg) during the final braking phase.
						Hip and knee extensors and ankle plantarflexors were earlier activated in the trailing leg, while ankle dorsiflexors were earlier activated in the front leg.
Gutierrez-Davila et al., 2013	33	Effects of target change on fencing performance	Lunge	Motion capture (Vicon)	Kinematics	A change in target location significantly increased the reaction time (by 28±32ms), movement time (by 69±50ms), the time used in acceleration (by 43 ±49ms), and errors made (by 18±19%) during lunge, while it also decreased the attacking velocity (by 0.33±0.35m/s) and action time in front foot (by 0.083±0.023s).
				Force plate (IBV)	Kinetics	
					Time-motion parameters	

(*Continued*)

Table 2. (*Continued*)

Study	Reference number	Objective	Moves	Measurement and Equipment	Variable of interest	Results
Gutierrez-Davila et al., 2013	29	Examined the differences between elite and medium fencers in response to changed lunge target	Lunge	Motion capture (Vicon)	Kinematics	Elite fencers generated higher flight time (36 ±37ms VS -2±12ms), late-phase horizontal foot velocity (4.56±0.75m/s VS 3.59±0.30m/s), sword velocity (2.55±0.42m/s VS 1.88±0.48m/s), and lunge length (1.40±0.15m VS 1.13±0.13m) as compared to the medium-level fencers.
				Force plate (IBV)	Kinetics	
					Time-motion parameters	Elite fencers made fewer errors (31±17% VS 43 ±12%) and maintained better arm-foot timing sequence in responses to target change.
Hassan and Klauck, 1998	18	Evaluate the fencing lunge movement based on quantitative analysis	Lunge	Motion capture (SELSPOT II)	Kinematics	The maximal horizontal foil velocity was 3.40–3.91m/s, horizontal foil velocity at hit time was 2.96–3.56m/s, and maximal horizontal hip velocity was 2.28–2.33m/s across four subjects.
Irurtia et al., 2008	26	Assessed the anthropometry and limb asymmetry in Spanish junior fencers	N/A	Anthropometric assessment	Anthropometries	Male fencers showed significantly (p = 0.01–0.05) larger forearm and thigh girths, as well as higher thigh muscle cross-sectional area (236±26 vs. 212 ±19cm^2) on the armed side than the Spanish reference population, while females fencers did not exhibit the advantages.
						No inter-limb significant differences were identified in both genders.
Kim et al., 2015	36	Effects of a specific training program in improving muscle imbalance	N/A	Motion capture (Motion Analysis Corporation)	Dispersion of center of mass and center of pressure	Fences showed significant improvement in mediolateral sway of the non-dominant leg during one-leg standing (8.55±4.46cm/fl VS 7.95±1.52cm/fl), mediolateral sway during deep squats (14.76 ±7.18cm/fl VS 9.95±2.54cm/fl) and the balance scale after (3.14±1.72 VS 1.81±0.92) training.
				Force plate (Kistler)	Balance score	
Lin et al., 2010	21	Evaluated the workload of the wrist muscles for different foil handle types	Quarte sixte	EMG measurement (Biometrics)	Muscle activity	The Viscounti-type handle elicited the most equal load distribution for all muscle groups in comparison to other handle types. However, Adductor pollicis and extensor carpi radialis were more activated in Viscounti-type handle condition (p = 0.011–0.017) and may be more vulnerable to fatigue.
						Handle angles for 21˚ and 24˚ increased risks of muscle fatigue, Grip strength was highest (8.66–9.52 (unknown unit)) at the two handle angles (p = 0.029).
Margonato et al, 1994	23	Investigate the bilateral differences in forearm muscle trophism and force	N/A	Anthropometric measurement	Dynamometer (Lafayette Instruments)	Fencers exhibited significant differences in cross-sectional area (51.7±8.2cm^2 VS 45.8±7.8cm^2) and isometric force (502±126N VS 449±115N) of the forearm between the dominate and non-dominate side (p<0.001).
						Fencers and the control groups were not significantly different on non-dominate side regarding muscular force and trophism.
						The absolute gains in muscular force and trophism of the dominate side were greater in fencers than the control group (5.9cm^2 VS 2.0cm^2, 53N VS 15N).
Morris et al., 2011	15	Investigated the characteristics of two fencing movements	Lunge	Motion capture (Vicon)	Kinematics	During the lunge, the ankle plantarflexors and knee extensors of the trailing leg contributed significantly to the attack. On the other hand, ankle plantarflexors and extensors of the hip and knee of both limbs contributed significantly to the progression of fleche. (Results are displayed by graphs only).
			Fleche	Force plate (Unspecified)	Kinetics	
Mulloy et al., 2015	37	Determined the kinematic chain in lunge	Lunge	Motion capture (Motion Analysis Corporation)	Kinematics	Expert fencers exhibited greater peak sword velocity (3.21±0.22m/s VS 2.63±0.29m/s), lunge distance (1.12±0.07 leg length VS 0.83±0.15 leg length), and peak ankle extension velocity (564 ±132˚/s VS 273±184˚/s).
						The sequential motion of the hip-knee-ankle sequential is more tightly coupled in elite than in non-elite fencers, allowing elite fencers to achieve greater ankle extension and forward sword velocity. (sequence identified by graphic comparison).

(*Continued*)

Table 2. (*Continued*)

Study	Reference number	Objective	Moves	Measurement and Equipment	Variable of interest	Results
Poulis et al., 2009	48	Examined the asymmetry of muscle strength in fencers	N/A	Anthropometric assessment	Anthropometries	Fencers had greater knee extension torque (112.3–221.4Nm VS 111.7–210.4Nm), flexion torque (66.9–119.7Nm VS 62.3–112.4Nm), and flexor/extensor peak torque ratio (F = 3.04–3.79, p = 0.01–0.03) compared to the non-fencer group, in regardless of the differences in angular velocity (30˚/s, 60˚/s, and 240˚/s).
				Strength test (Cybex)	Isokinetic strength	The differences in peak torque between the dominant and non-dominant legs were not significant in both groups.
Sinclair and Bottoms, 2015	45	Determined sex-differences in joint loading during lunge	Lunge	Motion capture (C-Motion)	Kinematics	Female fencers had significantly greater peak knee extension moment (2.05±0.22Nm·kg VS 1.72 ±0.25Nm·kg), patellofemoral joint contact force (2.90±0.58BW VS 2.18±0.43BW), and contact force loading rate (22.12±5.74BW/s VS 14.14 ±6.36BW/s) in comparison to male fencers.
				Force plate (Kistler)	Kinetics	
					Joint loading	
Sinclair et al., 2010	7	Effects of footwear on shock attenuating	Lunge	Accelerometer (Biometrics)	Impact shock	Traditional fencing shoes significantly increased the magnitude of peak impact shock in comparison to sports shoes that had shock absorbing qualities (p<0.01).
Sinclair and Bottoms, 2013	30	Investigated sex-differences in fencing kinematics and kinetics	Lunge	Motion capture (C-Motion)	Kinematics	FWhen variables were normalized by body weight., there were no significant inter-gender differences in both kinetics and kinematics except that, female fencers had significantly greater peak hip (42.79±12.42˚ VS 51.64±10.25˚) and knee abduction angles (1.91±6.44˚ VS -8.99±4.91˚) in comparison to male fencers.
				Force plate (Kistler)	Kinetics	
Sterkowicz-Przybycień, 2009	13	Body composition and somatotype of male fencers	N/A	Anthropometric assessment	Anthropometries	Fencers were characterized by higher mesomorphy and lower ectomorphy (p<0.05) compared to untrained males. Fencers' somatotypes differed from that of the untrained (3.3–4.8–2.3 vs. 3.7–4.3–3.1).
						Fencers using sabers were relatively heavier (84.4kg VS 74.9–77.9kg) and had higher mesomorphy (3.4–5.4–1.8 VS 3.6–4.9–2.5 and 2.9–4.2–2.8) than fencers using two other types of weapons.
Steward and Kopetka, 2005	24	Kinematic determinants of lunge speed	Lunge	Motion capture (Peak Motus)	Kinematics	Lunge velocity was significantly correlated to the time-to-peak angular velocity of the trailing knee (p = 0.022) and leading elbow (p = 0.047).
					Regression analysis for lunge performance	
Trautmann et al., 2011	3	Determined the foot loading characteristics of three fencing movements	Lunge	Plantar pressure (Pedar)	Plantar pressure	For the leading leg, the heel was predominately loaded during lunge (peak pressure: 551.8 ±113.9kPa, contact time: 705.4±166.9ms) while its hallux was more loaded during retreating movements (peak pressure: 341.0±122.4kPa, contact time: 205.0±43.0ms).
			Advance		Time-force parameters	For the trailing leg, the forefoot was generally loaded across the three different fencing movements (peak pressure: 170.2–352.7kPa, contact time: 191.8–682.2ms).
			Retreat			
Tsolakis et al., 2006	51	Investigated the anthropometric profile of young fencers	N/A	Anthropometric assessment	Anthropometries	There were generally no significant differences between male and female fencers in all age group in terms of anthropometric measurements.
						The mean somatotype of male fencers was 3.1–2.6–3.2 as compared to 3.8–14.8–3.3 in females. Female fencers were mainly situated in the ectomorph region.
						Cross-sectional area of the arms was higher in males compared to females and higher on the dominant side compared to the non-dominant side.
Tsolakis et al., 2006	49	Effects of a conditioning program to peripubertal fencers	N/A	Anthropometric assessment	Anthropometries	Increases in anthropometries, hormone level, and handgrip strength were detected in both groups. However, differences between fencers and the inactive children were not significant (p>0.05).
				Blood sampling	Hormones concentrations	
				Not fencing specific physical test (Psion XP)	Muscle strength	
					Physical performance	

(*Continued*)

Table 2. (*Continued*)

Study	Reference number	Objective	Moves	Measurement and Equipment	Variable of interest	Results
Tsolakis et al., 2010	34	Investigated selected correlates of fencing performance	Lunge	Anthropometric assessment	Anthropometries	Lunge time was best predicted (R^2 = 0.42, p = 0.001) by drop jump performance (30.5 ±8.58cm) and thigh cross-sectional area (205.3 ±38.50cm^2). When lunge time was normalized by body mass, only performance on the arm-driven counter-movement jump (37.7±9.26cm) was predictive of lunge time (R^2 = 0.71, p<0.001).
				Not fencing specific physical test (Psion XP)	Physical performance	
				Lunge test (Polifermo Radio Light)	Regression analysis for lunge performance	
Turner et al., 2016	47	Determined physical characteristics that underpinned lunge performance	Lunge	Anthropometric assessment	Anthropometries	Standing broad jump (177.7±0.32cm) was the strongest predictor of lunge velocity (3.35±0.70m/s, R^2 = 0.507, p<0.001) and change of direction speed (5.45±0.65m/s, R^2 = 0.425, p<0.001).
				Not fencing specific physical test (Optijump)	Physical performance	
				Lunge test (Casio)	Regression analysis for lunge performance	
Williams and Walmsley, 2000	67	Compare response profile between novice and elite fencers under several levels of target choice	Lunge	EMG measurement (Medicotest)	EMG signals timing	The elite fencers showed superiority over the novice fencers in reaction time (32-33ms less), total response time (19-23ms less), onset of activation in anterior deltoid (37-45ms less) and front rectus femoris (52-55ms less) in regardless of the changed target conditions (F = 10.29–34.46, p<0.05).
				Timing record	Timing parameters	Increased target number slightly elongated reaction time and delayed the onset of muscle activation for all fencers. However, the differences were not significant in both elite and novice groups.
						Elite fencers made fewer errors (11 VS 70) in hitting target than the novice fencers (X^2 = 12.18, p = 0.002).
						For both groups, the within-subject correlation was 75% (all variables of interest), indicating inter-trial consistency of movement pattern.
Williams and Walmsley, 2000	39	Compare response timing and muscle coordination between different fencers under target changing condition	Lunge	EMG measurement (Medicotest)	EMG signals timing	Elite fencers showed shorter reaction time (333 ±128ms, 40% of total response time VS 613 ±62ms, 66% of total response time) and total response time (808±53ms VS 934±34ms) in response to changed target.
				Timing record	Timing parameters	Elite fencers exhibited faster activation of selected muscle groups (178-378ms VS 301-617ms) in comparison to the novice fencers.
				Lunge distance measurement	Lunge distance	Elite fencers exhibited more coherent muscle synergy and consistent patterns of muscle coordination (rear knee extensor-front shoulder extensor-front knee extensor-front knee flexor).
Wylde, 2013	12	Time-motion analysis of foil fencing	Complete bout	Time-motion analysis (Sportstec)	Movement timing	High-intensity movements had a mean duration of 0.7s and accounted for 6.2% of the total bout time in elite women foil fencing.
					Movement duration	The work: recovery ratio of female's foil (15-touch) was 1:1.1, which was similar to that of men's epee (1:1), men's foil (1:3), and men's epee (8:10).
						For the 5-touch and team bouts the work: recovery ratio was 1:1, indicating an increased duration of moderate- and high-intensity movements.
Yiou and Do, 2000	40	Examine the differences between singular and combined training strategy of fencing performance	Touche	Force plate (Unspecified)	Foot pressure	There were no significant differences in body acceleration and peak velocity between the elite and novice fencers when touche and lunge were executed separately.
				Accelerometer (Entran)	Acceleration of body center	
			Lunge	EMG measurement (Unspecified)	Muscle activity	Elite fencers exhibited higher foil velocity (2.90 ±0.30m/s VS 2.66±0.29m/s) and postural velocity (0.41±0.20m/s VS 0.05±0.09m/s) in the sequential touché + lunge condition compared to the isolated touché condition, while no significant differences emerged for novice fencers.

(*Continued*)

Table 2. (*Continued*)

Study	Reference number	Objective	Moves	Measurement and Equipment	Variable of interest	Results
Yiou and Do, 2001	41	Examined the effects of "refractory period" on fencing movement	Touche	Force plate (Unspecified)	Displacement of the foot	There were no significant differences in postural velocity, speed performance, speed of focal movement, onset of anterior deltoid, and time of target hit between the elite and novice fencers in isolated touche condition.
				Accelerometer (Entran)	Acceleration of the foil	
			Lunge	EMG measurement (Unspecified)	Muscle activity	In lunge + touche condition, when the signal of touche was initiated more than 300ms prior to foot-off, there were no significant differences between groups. When touche was initiated within 200ms prior to foot-off or at the time of postural adjustment, speed performance and speed of focal movement were significantly higher in elite fencers.
						Maximum speed of touché was higher in elite fencers than in novice

Notes: N/A, not available.

(27.0%). Cross-sectional studies were assigned with 'weak' in EPHPP assessment. Some studies recruited less representative samples which had limited control on confounders regarding gender [14, 15], competitiveness [16–18], fencing events [12, 14, 19], and equipment features [20, 21]. Measurement method/collection (e.g. EMG signal processing and motion capture technique) was considered less reliable/not fully elaborated in four studies [13, 22–24]. Statistical analysis was not performed/not introduced in details in five studies [12, 15, 18, 25, 26]. Due to the nature of fencing sports and biomechanical study, outcome assessors could not be blinded to intervention/exposure. Since subjects were generally required to give their best performance in various fencing tasks, we assumed that they were not aware of the research questions for all recruited studies.

In addition to EPHPP score, the main limitation across the studies was the small sample size of study groups, with 14 studies including ≤10 participants. Sample size estimation, based on feasibility studies, was performed in only three studies [7, 27, 28]. Although significance level of 0.05, using two-tailed tests, was the most prevalent cut off for statistical analysis, the cut-off was not clarified in some studies [20, 22]. The small size of study groups increases the risk of violating the assumption of normal distribution required for parametric statistics. Yet, normality for analysis of variance and t-tests was verified in only six studies [7, 17, 27–30]. Therefore, the statistical methodology of studies included in our analysis was only "fair", with none of the studies reporting the statistical power of their results. Effect size, using Cohen's d-statistic (range in included studies, 0.29–2.26) and the eta-square statistic (range in included studies, 0.09–0.59) were reported in only six studies. The reliability of measured outcomes was evaluated by computing the interclass correlation coefficient (range in included studies, (0.570–0.988) in four studies [31–34].

3.4 Measurement methods

Techniques used for measurements are summarized in Table 2. Motion capture was used in 15 studies and force platform in 10 studies respectively, with electromyography measurement used in nine studies and accelerometers in five studies. However, there was noticeable variation in equipment and signal processing/analysis used across studies, even within a specific measurement technique. As an example, both infrared-based retroreflective marker [15, 17, 18, 29, 33, 35–37] and electromagnetic sensor tracking systems [38] were used in motion capture analysis. In terms of the processing of electromyography data, some studies used the root

Table 3. EPHPP score for the included studies.

Study	Research design	Number of subjects	EPHPP Score
Akpinar et al., 2015	Case-control	8	1
Aquili et al., 2013	Case-control	57	2
Bottoms et al., 2013	Cross-section	14	3
Chang et al., 2009	RCT	8	2
Cronin et al., 2003	Cross-section	21	2
Do and Yiou, 1999	Case-control	5	2
Frère et al., 2011	Case-control	8	2
Geil, 2002	RCT	13	2
Gholipour et al., 2008	Case-control	8	3
Greenhalgh et al., 2013	RCT	13	1
Gresham-Fiegel et al., 2013	Case-control	25	1
Guilhem et al., 2014	Case-control	10	1
Gutierrez-Davila et al., 2013	Cross-section	30	2
Gutierrez-Davila et al., 2013	Case-control	30	2
Hassan and Klauck, 1998	Cross-section	4	3
Irurtia et al., 2008	Cross-section	23	3
Kim et al., 2015	Cohort	9	2
Lin et al., 2010	RCT	15	3
Margonato et al, 1994	Case-control	58	2
Morris et al., 2011	Cross-section	1	3
Mulloy et al., 2015	Case-control	10	2
Poulis et al., 2009	Case-control	30	1
Sinclair and Bottoms, 2015	Case-control	16	1
Sinclair et al., 2010	RCT	19	2
Sinclair and Bottoms, 2013	Case-control	16	1
Sterkowicz-Przybycień, 2009	Case-control	30	2
Steward and Kopetka, 2005	Cross-section	15	3
Trautmann et al., 2011	Mixed design	29	1
Tsolakis et al., 2006	Case-control	152	2
Tsolakis et al., 2006	Controlled Clinical Trial	8	1
Tsolakis et al., 2010	Cross-section	33	2
Turner et al., 2016	Cross-section	70	2
Williams and Walmsley, 2000	Case-control	6	2
Williams and Walmsley, 2000	Case-control	6	2
Wylde, 2013	Cross-section	9	3
Yiou and Do, 2000	Case-control	11	1
Yiou and Do, 2001	Case-control	11	1

Notes: RCT, randomized controlled trial; N/A, not available.

mean square of the signal [21, 28] while others normalized the signal to the magnitude of a maximal voluntary contraction [17, 22, 39–42].

4. Discussion

4.1 Intrinsic, athlete-specific, factors

4.1.1 Sex-specific differences. As commonly reported for other sports, sex-specific differences in the kinematics of fencing were identified [35, 43, 44]. Specifically, females exhibited greater hip adduction, knee abduction/adduction, ankle eversion and patellofemoral contact forces of the leading leg during lunging movements [30, 45]. Some researchers have proposed that sex-specific differences in anthropometrics, neuromuscular functions and muscle strength may account for these differences in the kinematics of the leading leg [35, 43, 44]. Overall, the prevalence of injury was higher in female fencers (29%-44%, averaged: 36%) than in males (22%-32%, averaged: 27%) [2]. However, further evidence is required to characterize sex-specific differences based on high-quality evidence.

4.1.2 Anthropometry and muscle strength. The assessment of anthropometrics and strength is an important component for establishing functional profiles of fencers [46]. Performance of the lunge, considered to be the core component of fencing, was evaluated using stepping time or velocity, counter-movement jump strength and dynamometric tests [28, 47], with some studies evaluating the value of squat jump tests for predicting lunge performance [31, 34]. Despite differences in measurement, there is a consensus that greater lower limb strength and explosive power generated higher lunge velocity and quicker fencing moves [48, 49]. Ballistic training is therefore recommended to increase the rate of muscle force generation [50, 51], or a combination of resistance and ballistic training to optimize the explosive, power and endurance requirements of fencing [51, 52]. Turner et al. reported that most training programs were customized to a fencer's ability, sex and age by coaches, based on their experience or anecdotal knowledge rather than on evidence. Turner et al. [53] advocated the need to develop evidence-based training protocols which would consider a fencer's biomechanics, as well as physiological and psychological factors. Continued evaluation of the predictive value of somatotypes and measures of muscle strength and power to evaluate and improve fencing performance overall, and the lunge component more specifically, using multivariate analysis should be pursued in future studies.

4.1.3 Asymmetry. Fencing is clearly an asymmetric sports event based on its kinematic and kinetics [3, 15, 18]. Intuitively, the asymmetric sports contributed to asymmetry anthropometry due to the unilateral nature of the training. Irurtia et al. and Turner et al. [26, 53] showed that fencers had significantly higher handgrip strength and greater isokinetic leg strength on the dominant side than on the non-dominant side. Margonato and colleagues [23] conducted surveys on national-level fencers and discovered that they had higher muscle cross-sectional area in the dominant forearm. Arm and leg asymmetries were also observed in young fencers [49]

However, the effects of laterality on measures of muscle structure and performance have not been always consistently reported. Poulis et al. [48] failing to identify differences in peak knee and ankle isokinetic torques between the dominant and nondominant lower limb. It was argued that fencing movements do not necessarily induce unequal muscle growth between two sides of the lower limb. In fact, both trailing and leading legs contributed significantly to the progression of various fencing actions, especially for advance and fleche [15, 24]. Besides, Akpinar et al. [38] investigated the differences in movement accuracy, speed, multi-joint coordination, and handedness between fencers and non-fencers in a series of hand reaching tasks. They showed that fencers may have performance symmetry superior to non-fencers due to an underlying high skill in bilateral control, while Poulis also advocated that elite fencers have all-round development on their neuromuscular system [48]. In fact, the risk of injury associated with asymmetry is recognized, such that balance or weight training is often introduced to tackle with possible asymmetry-induced injuries [54–56] [36, 57]. The variation in outcome measures and training method may contribute to the inconsistency of results. Therefore, the development and role of asymmetry in fencing should be subject of future studies.

4.2 Extrinsic factors

4.2.1 Weapon. Foil, saber, and epee are the three major weapons used in fencing sports. Each weapon has its own rules and strategies. Though the basic offensive and defensive techniques are universally applicable in the three weapons, their biomechanics may differ due to the variances in blade type (e.g. length and weight), valid target (e.g. torso with and without the extremities) and scoring technique (e.g. thrusting and cutting). Both epee and foil are thrusting weapons which score only by landing their tips on the valid area, while epee is heavier (775g VS 350-500g, same in blade length: 90cm) than foil. In spite of the limited evidence from existing studies, there was a trend in the numbers present that foil fencers had slightly higher peak velocity in mass center (1.92m/s VS 1.72m/s), weapon (2.91m/s VS 2.49m/s), and the front foot (4.56m/s VS 4.10m/s) than epee fencers during the execution of simple lunge attack [16, 18, 24, 25, 29, 33, 40–42]. Though the comparison is based on different samples, there are also signs from some performance analyses that the length of action and break was shorter in foil than in epee (5.2s VS 12.7s, 15.6s VS 18.2s) [14]. Ratio of action to break was lower in foil (1:3) than in epee (1:1.0–1.4) [1, 58], indicating quicker-finished bouts in foil. Saber (weight: 500g, length: 88cm in the blade) is most distinctive from the other two weapons by its 'cutting' rule. Data from research on saber biomechanics was rare, but saber was thought to have a fast tempo and a quicker burst of speed than the other two weapons [59]. Length of action/break (2.5±0.6s, 16.5±2.7s) and ratio of action to break (1:6.5) was the lowest in saber [14]. Up to date, fencing weapons were usually studies in simple movement and controlled in-lab conditions, e.g. lunge and fleche. However, in-game combat is more complex involving many extensions of the fundamental actions. Saber fencing may be even faster paced than reports.

In addition to having to bear the weight of the long weapon itself, fencers' wrist further sustains abrupt and violent loading during the series of attack and defense fencing actions. Therefore, a well-designed weapon handle is important for reducing abnormal wrist joint motion and lowering the risk of wrist injury. Three types of handles are generally used in fencing: the French, the pistol, and the Italian handle. The French and Italian handles have evolved from the classical rapier handle from the Baroque period, with the French handle slightly contoured to the curve of the hand. The pistol handle is much like that of a pistol and commonly known as the "anatomical" or "orthopaedic" grip. Use of the pistol handle is advocated as it elicits lower activation in the adductor pollicis and extensor carpi radialis muscles compared to the French and Italian handles [22]. Moreover, the Visconti style pistol handle also promotes a more balanced activation of forearm muscles [21, 22] which would delay muscle fatigue and improve the capacity of the wrist to resists excessive motion when external forces are applied. This protective function of the pistol handle can be enhanced by placing the handle at an angle of 18°-21°, which also improves hit rate and accuracy while delaying the onset of early fatigue [21].

4.2.2 Footwear and the fencing piste. Fencers' feet are repeatedly exposed to large transient impact shock, especially during sudden forward thrusts, increasing the risk of lower limb injuries [60]. The metal carpet piste is the main source of these high impact forces. Although different overlay materials have been used for shock absorption, Greenhalgh et al. [27] could not confirm a significant attenuation of impact magnitude for different overlays.

Fencing shoes compound this problem by providing little intrinsic shock absorption, compared to standard court shoes. Geil [20] and Sinclair et al. [7] confirmed the lower shock absorption capacity of fencing shoes compared to squash and running shoes. Geil [20] did discuss the trade-off between increased shock absorption of shoes and a slower and less reliable performance, with shock absorbing materials reducing sensory information from the floor which provides important proprioceptive input for agility and balance [61]. Therefore, finding a balance between shock absorption and performance remains an issue to be resolved.

4.2.3 Training and conditioning. The success of fencing was largely determined by speed and explosive strength [53]. Therefore, ballistic training is recommended to increase the rate of muscle force generation [62], with most of the improvements occurring within the first 200 to 300 milliseconds of a single lunge movement [63, 64]. Research has shown that pure fencing training regime did not induce growth of muscle strength that overrode the progression of puberty for the adolescent fencers (compared to the inactive children: increases in arm cross sectional area: 17.1% VS 6.97%, increases in grip strength: 25.81% VS 18.07%) [51]. The increases of leg muscle cross sectional area (32±7%) and body mass (16±3%) in fencers was also insignificant when the effects of body height were ruled out. The author thus recommended strength training for young fencers to complement their training routine.

Training of balance and coordination is also a fundamental element [65], By taking specific balance training, fencers demonstrated better coordination (42.36% improvement in balance score) and less body sway (7.02% less in dispersion of the center of plantar pressure) in single-leg standing tasks. When coordinating touché and lunge movements, elite fencers produced higher sword velocity (2.90±0.30m/s VS 2.52±0.29m/s) and body center velocity (0.41±0.20m/s VS 0.04±0.10m/s), compared to novice fencers [36, 40]. In their review of training programs, Turner et al. [53] reported that most programs are customized to a fencer's ability, sex and age by coaches, based on their experience or anecdotal knowledge rather than on evidence. Turner et al. advocated the need to develop evidence-based training protocols which would consider a fencer's biomechanics, as well as physiological and psychological factors.

4.3 Biomechanics of fencing

Fencing is a highly asymmetric sport, with the armed side of the body leading movement over a substantial duration of a competitive bout, and during training. Moreover, the upper and lower extremities present distinctive motion patterns, which imposes a considerable burden on the neuromuscular system, including effects of dominance on kinematics and kinetics [66]. The advance, retreat, fleche, and lunge movements, commonly used in fencing, have been evaluated using motion capture, demonstrating the greater joint motion and force output required to perform the fleche and lunge movements [31, 67].

4.3.1 Posture and kinematics. In lunging, power during the propulsion phase is provided by the ankle plantarflexors and knee/hip extensors of the trailing leg, with additional contribution from the hip flexors and knee extensors of the leading leg during the subsequent flight phase [15]. Contrarily in fleche, both lower limbs provided power in a cyclical sprint-like manner. During the initial phase, the trailing leg provides the thrusting power as its ankle plantarflexors and knee extensors control the velocity of flexion at the ankle and knee joints of the leading leg. Once the trailing leg passes in front of the leading leg, the thrust-absorption cycle is repeated with a reversal of the power and absorption roles of the lower limbs [15]. Lower limb coordination and balance also significantly influence performance, with elite fencers generating greater hip flexion force of the leading leg at the end of a lunging movement and, hence, a higher sword velocity [16, 25]. Range of motion of the knee and peak hip flexion range of the trailing leg and hip flexion range of the leading leg were identified as significant predictors of lunge performance, allowing fencers to assume a low on-guard position and adjust movement of the leading leg in lunge to improve their performance [16]. Though these studies of fencing performance were based on small sample size, their subjects were mostly elite fencers who can represent the significance of fencing sports. Besides, the outcomes were quite consistent in terms of phase-contributors of lunge dynamics.

Study of in-shoe pressure revealed the asymmetric characteristics of weight bearing on the foot. Trautmann, Martinelli & Rosenbaum [3] and Geil [20] evaluated the distribution of plantar pressures during three fencing movements [3]. Load was predominantly placed on the heel of the leading foot and on the forefoot of the trailing foot during performance of a lunge and advance, with plantar pressure being a maximum under the hallux, bilaterally, during the retreat movement, regardless of the type of shoes worn. Steward and Kopetka [24] further reported time-to-peak angular velocity of the knee joint of the trailing leg and of the elbow of on the leading side to have a significant effect on overall lunge speed. The angle of the trailing foot relative to the lead foot also influenced lunge performance [68]. Gresham-Fieg, House & Zupan [32] evaluated the effects of three different angles of rear foot placement on lunge performance, confirming that placement of the rear foot perpendicular to the alignment of the forefoot produced higher magnitudes of peak and average power and average velocity of lunging.

4.3.2 Joint coordination and synergy. The proximal-to-distal coupling of upper and lower limb motion ensures an effective transfer of joint segmental angular velocity of the lower limb to the maximum linear velocity of the center of mass [69], a feature which differentiates elite from novice fencers [37]. Elite fencers extended the weapon arm prior to initiating front foot movement during lunge [18]. Focusing specifically on the coordination among lower limb joints, Mulloy et al. [37] identified that elite fencers exhibited greater peak horizontal sword velocity and lunge distance in comparison to novice fencers through a clearly sequential increase in angular velocity of joint extension from the hip, to the knee, to the ankle. This kinematic sequence was claimed to be the correct technique that increase fencing success in elite fencers [39]. Do and Yiou has identified a 'refractory period' existing between motor tasks that had negative effects on fencing performance [42]. Elite fencers were competent to inhibit the effects by closely linking 'touche' movement of the arm and 'lunge' movement of the legs in perfect timeline [41]. Technical training should take into account of the specific fencing movement pattern, with emphasis on practicing different movement components in combination rather than in separate form [40].

4.3.3 Muscle coordination and synergy. The proximal-to-distal sequence was also reported for muscle activation, with activation of the anterior deltoid of the armed upper limb, with extension of the armed hand, preceding the lifting of the lead foot at the initiation of lunging in expert fencers [40]. In contrast, novice fencers exhibited a delayed onset of upper limb muscle activity, associated with shortened propulsion phase by the trailing leg resulting in an earlier "kick off" of the trailing leg in novice fencers [18, 67]. Overall, elite fencers presented more coherent muscle synergies of the upper and lower limbs, compared to novice fencers, characterized by sequential activation of shoulder/elbow extensors of the armed upper limb and hip/knee extensors of the rear lower limb, followed by activation of the forelimb during lunging, with the ability to maintain this quasi-invariant pattern of activation despite changing task requirements during a fencing bout [39, 67]. In contrast, muscle activation patterns for novice fencers were more variable with changing task demands imposed by their opponents, often leading to interruptions in their movement [29, 33]. Therefore, novice fencers may not have consolidated neuromuscular strategies for complex, multi-segmental movements [70], while elite fencers are able to finely adjust muscle activation patterns to optimize attacking (lunge) efficiency without violating the "correct" kinematic sequence of upper and lower limb motions (as the sequence mentioned in section 4.3.2: rear knee extension-front shoulder extension-front knee extension-front knee flexion) [17].

5. Summary and remarks

Fencing is an idiosyncratic sport, with unique patterns of asymmetrical movements and biomechanics. Although athlete-specific (intrinsic) and external factors were identified as influencing performance and, probably the risk of injury, current evidence can be considered incomplete and of "fair" quality only. However, we did identify key points that can begin to inform practice, as well as providing a direction for future research. Foremost, fencing requires explosive force and as such, evidence regarding effects of sex, anthropometry, muscle structure and neuromuscular coordination is required. Although intuitively, effects of asymmetry have been discussed and evaluated, evidence of an effect of handedness on muscle strength should be more deeply studied. Moreover, elite fencers were found to have a higher capacity for bilateral performance of hand tasks than the novice or untrained individuals. However, neuromuscular control of multi-joint movements is essential to an elite fencing performance. A unique feature of fencing is the metal carpet piste and the poor shock absorbing characteristics of the fencing shoes which increase the magnitude of impact forces and the risk of foot/ankle and knee injuries. Strategies to mitigate impact forces while optimizing performance are required.

Current evidence is limited by the narrow scope of the research and "fair" methodological quality. Foremost, due to differences in the characteristics of different weapons and, therefore, movement requirements, findings are not transferable across weapon type [14]. Information available largely addresses the lunge component, which is understandable as it is the core movement component in fencing. Although a range of measures of lunge performance has been used, ranging from sword velocity to timing of target hit, the predictive value of these different measures to overall tactical performance and injury has yet to be determined. Moreover, fencing contest regularly lasts for more than an hour, during which fencers experience rapid alternation between resting and intensive activity [5]. Therefore, muscle fatigue and psychophysical exhaustion would influence performance measures and increase overall risk for injury and poor performance outcomes [1]. Future studies will need to evaluate effects of fatigue on fencing performance.

For future direction, current training programs mainly focus on improvement of muscle strength and power, with endurance training having received relatively less attention, despite its importance to injury prevention [71, 72]. Footwear design will also need to be addressed to reduce exposure to repetitive high magnitude impacts. Numerical modeling, in combination with neurophysiological measures of proprioception and muscle activation and performance-based outcomes, could assist in identifying optimal design criteria for fencing shoes by predicting internal loading across the geometrically complex anatomy of the foot and ankle [73, 74].

Supporting information

S1 Checklist. PRISMA checklist.
(PDF)

Acknowledgments

The authors would like to acknowledge the Hong Kong Research Institute of Textiles and Apparel and the Hong Kong Sports Institute for providing us with the figures of fencing weapons and movements.

Author Contributions

Conceptualization: TLC DWW MZ.

Formal analysis: TLC DWW YW.

Funding acquisition: MZ.

Investigation: SR FY.

Methodology: TLC DWW.

Project administration: MZ.

Resources: TLC SR FY.

Software: DWW MZ.

Supervision: MZ.

Visualization: TLC DWW.

Writing – original draft: TLC DWW.

Writing – review & editing: TLC DWW YW MZ.

References

1. Roi GS, Bianchedi D. The science of fencing: Implications for performance and injury prevention. Sports Med. 2008; 38: 465–481. PMID: 18489194
2. Harmer PA. Incidence and characteristics of time-loss injuries in competitive fencing: a prospective, 5-year study of national competitions. Clin J Sport Med. 2008; 18: 137–142. PMID: 18332688
3. Trautmann C, Martinelli N, Rosenbaum D. Foot loading characteristics during three fencing-specific movements. J Sports Sci. 2011; 29: 1585–1592. doi: 10.1080/02640414.2011.605458 PMID: 22077403
4. Angelo D, Kirby J. The School of Fencing: With a General Explanation of the Principal Attitudes and Positions Peculiar to the Art. London: Greenhill Books; 2005.
5. Murgu A-I. Fencing. Phys Med Rehabil Clin N Am. 2006; 17: 725–736. doi: 10.1016/j.pmr.2006.05.008 PMID: 16952760
6. Barth B. The Complete Guide to Fencing. Meyer & Meyer Verlag; 2006.
7. Sinclair J, Bottoms L, Taylor K, Greenhalgh A. Tibial shock measured during the fencing lunge: the influence of footwear. Sports Biomech. 2010; 9: 65–71. doi: 10.1080/14763141.2010.491161 PMID: 20806842
8. Wild A, Jaeger M, Poehl C, Werner A, Raab P, Krauspe R. Morbidity profile of high-performance fencers. Sportverletz Sportschaden Organ Ges Orthopadisch-Traumatol Sportmed. 2001; 15: 59–61.
9. Wild A, Jaeger M, Poehl C, Werner A, Raab P, Krauspe R. Morbidity profile of high-performance fencers. Sportverletz Sportschaden Organ Ges Für Orthop-Traumatol Sportmed. 2001; 15: 59–61.
10. Timpka T, Ekstrand J, Svanström L. From sports injury prevention to safety promotion in sports. Sports Med Auckl NZ. 2006; 36: 733–745.
11. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract. 2012; 18: 12–18. doi: 10.1111/j.1365-2753.2010.01516.x PMID: 20698919
12. Wylde MJ, Tan FHY, O'Donoghue PG. A time-motion analysis of elite women's foil fencing. Int J Perform Anal Sport. 2013; 13. Available: https://www.researchgate.net/publication/263213406_A_time-motion_analysis_of_elite_women's_foil_fencing
13. Sterkowicz-Przybycień K. Body composition and somatotype of the elite of Polish fencers. Coll Antropol. 2009; 33: 765–772. PMID: 19860102
14. Aquili A, Tancredi V, Triossi T, De Sanctis D, Padua E, D'Arcangelo G, et al. Performance analysis in saber. J Strength Cond Res. 2013; 27: 624–630. PMID: 23443217
15. Morris N, Farnsworth M, Robertson DGE. Kinetic analyses of two fencing attacks–lunge and fleche. Port J Sport Sci. 2011; 11: 343–346.
16. Bottoms L, Greenhalgh A, Sinclair J. Kinematic determinants of weapon velocity during the fencing lunge in experienced épée fencers. Acta Bioeng Biomech. 2013; 15: 109–113. PMID: 24479483

17. Frère J, Göpfert B, Nüesch C, Huber C, Fischer M, Wirz D, et al. Kinematical and EMG-Classifications of a Fencing Attack. Int J Sports Med. 2011; 32: 28–34. doi: 10.1055/s-0030-1267199 PMID: 21086241

18. Hassan SEA, Klauck J. Kinematics of lower and upper extremities motions during the fencing lunge: results and training implications. ISBS—Conf Proc Arch. 1998; 1: 170–173.

19. Turner A, James N, Dimitriou L, Greenhalgh A, Moody J, Fulcher D, et al. Determinants of Olympic Fencing Performance and Implications for Strength and Conditioning Training. J Strength Cond Res. 2014; 28: 3001–3011. PMID: 24714533

20. Geil MD. The role of footwear on kinematics and plantar foot pressure in fencing. J Appl Biomech. 2002; 18: 155–162.

21. Lin FL, Chang CL, Jou YT, Pan HC, Hsu TY. The study of influence of fencing handle type and handle angle on wrist for a fencing game. Industrial Engineering and Engineering Management (IE&EM), 2010 IEEE 17Th International Conference on. IEEE; 2010. pp. 1624–1627. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5646097

22. Chang C-L, Lin F-T, Li K-W, Jou Y-T, Huang C-D. A study of optimal handle shape and muscle strength distribution on lower arm when holding a foil. Percept Mot Skills. 2009; 108: 524–530. doi: 10.2466/PMS.108.2.524-530 PMID: 19544957

23. Margonato V, Roi GS, Cerizza C, Galdabino GL. Maximal isometric force and muscle cross-sectional area of the forearm in fencers. J Sports Sci. 1994; 12: 567–572. doi: 10.1080/02640419408732207 PMID: 7853453

24. Steward SL, Kopetka B. The kinematic determinants of speed in the fencing lunge. J Sports Sci. 2005; 23: 105.

25. Gholipour M, Tabrizi A, Farahmand F. Kinematics analysis of lunge fencing using stereophotogrametry. World J Sport Sci. 2008; 1: 32–37.

26. Irurtia A, Pons V, Carrasco M, Iglesias X, Porta J, Rodriguez FA. Anthropometric profile and limb asymmetries in Spanish junior elite male and female fencers. Fenc Sci Technol Barc General Catalunya-INEFC. 2008; 185–7.

27. Greenhalgh A, Bottoms L, Sinclair J. Influence of surface on impact shock experienced during a fencing lunge. J Appl Biomech. 2013; 29: 463–467. PMID: 22923353

28. Guilhem G, Giroux C, Couturier A, Chollet D, Rabita G. Mechanical and Muscular Coordination Patterns during a High-Level Fencing Assault: Med Sci Sports Exerc. 2014; 46: 341–350. PMID: 24441214

29. Gutierrez-Davila M, Rojas FJ, Antonio R, Navarro E. Response timing in the lunge and target change in elite versus medium-level fencers. Eur J Sport Sci. 2013; 13: 364–371. doi: 10.1080/17461391.2011.635704 PMID: 23834541

30. Sinclair J, Bottoms L. Gender differences in the kinetics and lower extremity kinematics of the fencing lunge. Int J Perform Anal Sport. 2013; 13: 440–451.

31. Cronin J, McNAIR P, Marshall R. Lunge performance and its determinants. J Sports Sci. 2003; 21: 49–57. doi: 10.1080/0264041031000070958 PMID: 12587891

32. Gresham-Fiegel CN, House PD, Zupan MF. The effect of nonleading foot placement on power and velocity in the fencing lunge. J Strength Cond Res. 2013; 27: 57–63. PMID: 22395272

33. Gutierrez-Davila M, Rojas FJ, Caletti M, Antonio R, Navarro E. Effect of target change during the simple attack in fencing. J Sports Sci. 2013; 31: 1100–1107. doi: 10.1080/02640414.2013.770908 PMID: 23421933

34. Tsolakis C, Kostaki E, Vagenas G. Anthropometric, flexibility, strength-power, and sport-specific correlates in elite fencing. Percept Mot Skills. 2010; 110: 1015–1028. PMID: 20865989

35. Sinclair J, Greenhalgh A, Edmundson CJ, Brooks D, Hobbs SJ. Gender differences in the kinetics and kinematics of distance running: implications for footwear design. Int J Sports Sci Eng. 2012; 6: 118–128.

36. Kim T, Kil S, Chung J, Moon J, Oh E. Effects of specific muscle imbalance improvement training on the balance ability in elite fencers. J Phys Ther Sci. 2015; 27: 1589–1592. doi: 10.1589/jpts.27.1589 PMID: 26157269

37. Mulloy F, Mullineaux D, Irwin G. Use of the kinematic chain in the fencing attacking lunge. Poitiers, France; 2015. http://eprints.lincoln.ac.uk/17817/

38. Akpinar S, Sainburg RL, Kirazci S, Przybyla A. Motor Asymmetry in Elite Fencers. J Mot Behav. 2015; 47: 302–311. doi: 10.1080/00222895.2014.981500 PMID: 25494618

39. Williams LR, Walmsley A. Response timing and muscular coordination in fencing: a comparison of elite and novice fencers. J Sci Med Sport. 2000; 3: 460–475. PMID: 11235010

40. Yiou E, Do MC. In fencing, does intensive practice equally improve the speed performance of the touche when it is performed alone and in combination with the lunge? Int J Sports Med. 2000; 21: 122–126. doi: 10.1055/s-2000-8864 PMID: 10727073

41. Yiou E, Do MC. In a complex sequential movement, what component of the motor program is improved with intensive practice, sequence timing or ensemble motor learning? Exp Brain Res. 2001; 137: 197–204. PMID: 11315548

42. Do MC, Yiou E. Do centrally programmed anticipatory postural adjustments in fast stepping affect performance of an associated" touche" movement? Exp Brain Res. 1999; 129: 462–466. PMID: 10591918

43. Katis A, Kellis E, Lees A. Age and gender differences in kinematics of powerful instep kicks in soccer. Sports Biomech Int Soc Biomech Sports. 2015; 14: 287–299.

44. Niu W, Wang Y, He Y, Fan Y, Zhao Q. Biomechanical gender differences of the ankle joint during simulated half-squat parachute landing. Aviat Space Environ Med. 2010; 81: 761–767. PMID: 20681236

45. Sinclair J, Bottoms L. Gender differences in patellofemoral load during the epee fencing lunge. Res Sports Med. 2015; 23: 51–58. doi: 10.1080/15438627.2014.975813 PMID: 25630246

46. Gonçalves C EB, Rama L ML, Figueiredo AB. Talent identification and specialization in sport: an overview of some unanswered questions. Int J Sports Physiol Perform. 2012; 7: 390–393. PMID: 22868280

47. Turner A. Physical Characteristics Underpinning Lunging and Change of Direction Speed in Fencing: J Strength Cond Res. 2016; 1.

48. Poulis I, Chatzis S, Christopoulou K, Tsolakis C. Isokinetic strength during knee flexion and extension in elite fencers. Percept Mot Skills. 2009; 108: 949–961. doi: 10.2466/PMS.108.3.949-961 PMID: 19725328

49. Tsolakis C, Bogdanis GC, Vagenas G. Anthropometric profile and limb asymmetries in young male and female fencers. J Hum Mov Stud. 2006; 50: 201–215.

50. Komi PV. Strength and Power in Sport. 2nd edition. Osney Mead, Oxford ; Malden, MA: Wiley-Blackwell; 2002.

51. Tsolakis C, Bogdanis G, Vagenas G, Dessypris A. Influence of a twelve-month conditioning program on physical growth, serum hormones, and neuromuscular performance of peripubertal male fencers. J Strength Cond Res. 2006; 20: 908–914. PMID: 17194232

52. Aşçi A, Açikada C. Power production among different sports with similar maximum strength. J Strength Cond Res Natl Strength Cond Assoc. 2007; 21: 10–16.

53. Turner A, Miller S, Stewart P, Cree J, Ingram R, Dimitriou L, et al. Strength and conditioning for fencing. Strength Cond J. 2013; 35: 1–9.

54. Gray J, Aginsky KD, Derman W, Vaughan CL, Hodges PW. Symmetry, not asymmetry, of abdominal muscle morphology is associated with low back pain in cricket fast bowlers. J Sci Med Sport. 2016; 19: 222–226. doi: 10.1016/j.jsams.2015.04.009 PMID: 26059231

55. Jayanthi N, Esser S. Racket sports. Curr Sports Med Rep. 2013; 12: 329–336. PMID: 24030308

56. Reeser JC, Joy EA, Porucznik CA, Berg RL, Colliver EB, Willick SE. Risk Factors for Volleyball-Related Shoulder Pain and Dysfunction. PM&R. 2010; 2: 27–36.

57. Hyun J, Hwangbo K, Lee C-W. The effects of pilates mat exercise on the balance ability of elderly females. J Phys Ther Sci. 2014; 26: 291–293. doi: 10.1589/jpts.26.291 PMID: 24648651

58. Bottoms L. Physiological responses and energy expenditure to simulated epee fencing in elite female fencers. Serbian J Sports Sci. 2011; 5: 17–20.

59. Hutton A, Martinez R. Cold Steel: The Art of Fencing with the Sabre. Courier Corporation; 2006.

60. Harmer PA. Getting to the point: injury patterns and medical care in competitive fencing. Curr Sports Med Rep. 2008; 7: 303–307. PMID: 18772692

61. Sekizawa K, Sandrey MA, Ingersoll CD, Cordova ML. Effects of shoe sole thickness on joint position sense. Gait Posture. 2001; 13: 221–228. http://dx.doi.org/10.1016/S0966-6362(01)00099-6 PMID: 11323228

62. Dahab KS, McCambridge TM. Strength Training in Children and Adolescents. Sports Health. 2009; 1: 223–226. doi: 10.1177/1941738109334215 PMID: 23015875

63. Newton RU, Kraemer WJ. Developing Explosive Muscular Power: Implications for a Mixed Methods Training Strategy. STRENGTH Cond J. 1994; 16: 20.

64. Häkkinen K, Pakarinen A, Alén M, Komi PV. Serum hormones during prolonged training of neuromuscular performance. Eur J Appl Physiol. 1985; 53: 287–293.

65. Gutiérrez-Cruz C, Rojas FJ, Gutiérrez-Davila M. Effect of defence response time during lunge in foil fencing. J Sports Sci. 2015; 1–7.

66. KawaŁek K, Ogurkowska MB. A comparison of selected biomechanical parameters in speed-endurance athletes. Trends Sport Sci. 2014; 21. Available: http://www.wbc.poznan.pl/Content/329892/5_TRENDS_Vol%2021_no.2_2014_85.pdf

67. Williams LR, Walmsley A. Response amendment in fencing: differences between elite and novice subjects. Percept Mot Skills. 2000; 91: 131–142. PMID: 11011884

68. Evangelista N. The Art and Science of Fencing. 1st edition. Indianapolis, IN: McGraw-Hill Education; 1999.

69. Bobbert MF, van Soest AJ. Why do people jump the way they do? Exerc Sport Sci Rev. 2001; 29: 95–102. PMID: 11474963

70. Zhang D, Ding H, Wang X, Qi C, Luo Y. Enhanced response inhibition in experienced fencers. Sci Rep. 2015; 5: 16282. doi: 10.1038/srep16282 PMID: 26541899

71. Hassanlouei H, Falla D, Arendt-Nielsen L, Kersting UG. The effect of six weeks endurance training on dynamic muscular control of the knee following fatiguing exercise. J Electromyogr Kinesiol. 2014; 24: 682–688. doi: 10.1016/j.jelekin.2014.06.004 PMID: 25112924

72. Weist R, Eils E, Rosenbaum D. The Influence of Muscle Fatigue on Electromyogram and Plantar Pressure Patterns as an Explanation for the Incidence of Metatarsal Stress Fractures. Am J Sports Med. 2004; 32: 1893–1898. PMID: 15572318

73. Chen TL, An WW, Chan ZYS, Au IPH, Zhang ZH, Cheung RTH. Immediate effects of modified landing pattern on a probabilistic tibial stress fracture model in runners. Clin Biomech. 2016; 33: 49–54.

74. Wong DW-C, Zhang M, Yu J, Leung AK-L. Biomechanics of first ray hypermobility: an investigation on joint force during walking using finite element analysis. Med Eng Phys. 2014; 36: 1388–1393. PMID: 24726375

7

General versus sports-specific injury prevention programs in athletes: A systematic review on the effects on performance

Ashley Plummer[1☯*], **Hendrik Mugele**[1,2☯], **Kathrin Steffen**[3], **Josefine Stoll**[4], **Frank Mayer**[4], **Juliane Müller**[4,5]

1 Clinical Exercise Science, University of Potsdam, Department of Sport and Health Sciences, Potsdam, Germany, **2** Department of Sport Science, University of Innsbruck, Innsbruck, Austria, **3** Oslo Sports Trauma Research Center, Oslo, Norway, **4** University Outpatient Clinic, Professorship of Sports Medicine and Orthopedics, University of Potsdam, Potsdam, Germany, **5** Professorship for Physiotherapy: Exercise Science and Applied Biomechanics, Trier University of Applied Science, Trier, Germany

☯ These authors contributed equally to this work.
* ashleyplummer@hotmail.com

Editor: Daniel Boullosa, University of Brasilia, BRAZIL

Competing interests: The authors have declared that no competing interests exist.

Abstract

Introduction

Injury prevention programs (IPPs) are an inherent part of training in recreational and professional sports. Providing performance-enhancing benefits in addition to injury prevention may help adjust coaches and athletes' attitudes towards implementation of injury prevention into daily routine. Conventional thinking by players and coaches alike seems to suggest that IPPs need to be specific to one's sport to allow for performance enhancement. The systematic literature review aims to firstly determine the IPPs nature of exercises and whether they are specific to the sport or based on general conditioning. Secondly, can they demonstrate whether general, sports-specific or even mixed IPPs improve key performance indicators with the aim to better facilitate long-term implementation of these programs?

Methods

PubMed and Web of Science were electronically searched throughout March 2018. The inclusion criteria were randomized control trials, publication dates between Jan 2006 and Feb 2018, athletes (11–45 years), injury prevention programs and included predefined performance measures that could be categorized into balance, power, strength, speed/agility and endurance. The methodological quality of included articles was assessed with the Cochrane Collaboration assessment tools.

Results

Of 6619 initial findings, 22 studies met the inclusion criteria. In addition, reference lists unearthed a further 6 studies, making a total of 28. Nine studies used sports specific IPPs, eleven general and eight mixed prevention strategies. Overall, general programs ranged from 29–57% in their effectiveness across performance outcomes. Mixed IPPs improved in 80% balance outcomes but only 20–44% in others. Sports-specific programs led to larger scale improvements in balance (66%), power (83%), strength (75%), and speed/agility (62%).

Conclusion

Sports-specific IPPs have the strongest influence on most performance indices based on the significant improvement versus control groups. Other factors such as intensity, technical execution and compliance should be accounted for in future investigations in addition to exercise modality.

Introduction

In the past decade, youth participation in organized sport has increased 21% resulting in approximately 30 million youth amateur athletes in the US alone [1]. This observation is likely driven by numerous health initiatives encouraging the many physical and psychosociological benefits of physical activity [2]. Aside from the clear benefits of participation, >30% of injuries among adolescents are sport-related making it the leading cause of injury for this population in western societies [3]. Risk factors such as increasing level of competition, training duration and intensity [4] were consistently identified across 15 different sporting disciplines [5] and deemed responsible for the high injury rates. However, despite increasing research and knowledge of injury mechanisms (e.g. anterior cruciate ligament injury [6]), risk factors and recovery methods, injury rates remain high in both the youth [5] and professional [7] ranks.

There is a general agreement among sport professionals that injury has an unfavorable impact on the sport they are involved in. In a survey that collaborated with 72 professionals within soccer, 77% believe that injury has a negative impact on their teams overall performance [8]. Additionally, 98% of soccer professionals share the belief that evidence based injury prevention exercises should be applied to their regular training routine [8]. Injury prevention programs (IPPs) such as the FIFA 11+ [9], [10], the PEP Program [11] and Sportsmetrics [12] have been previously designed and proven effective in preventing sports-related injuries. However, long-term implementation of IPPs is often problematic as proving effectiveness against injury is insufficient for coaches and athletes to convert from their conventional routine [13].

Additionally, modified versions of evidence based programs are partially implemented to add variation, progression and individualization as well as to align with specific training formats and goals [8]. In the survey, the FIFA 11+ was implemented in their original (12% of sessions) and modified version (28% of sessions) [8]. This could be attributed to the fact compliance to an IPP is highly dependent on the exercises being enjoyable, practicable, progressive and result in physical benefits for their athletes [10,14], which often manifests itself in the form of exercises that pertain directly to the given sport e.g. sports specific.

Specificity is recognized by coaches as a key principle of training that result in training adaptations and is therefore, frequently carried out in sessions [15]. If it can be revealed that specific IPPs lead to performance improvement in comparison to other types of IPPs, then coaches would be assured that specificity should remain an integral part of their training time with their athletes.

A previous systematic review conducted by the present authors aimed to determine whether specificity of an IPP affects the injury rates of young athletes [16]. One main finding was that there is currently a dearth of research using sports specific IPPs (n = 2) as per the authors´ definition, thus making practical recommendations challenging. Due to this, a new perspective that hasn't previously been addressed is: Does additional injury prevention training culminate in performance improvement in athletes? If so, then does the specificity of the program make a difference in enhancing crucial performance parameters? Or, could it be the case that programs with general training exercises attribute to this improvement and thus resulting in a one size fits all type approach. There are IPPs that integrate tasks that are more sports-specific e. g. using cutting maneuvers and jumping/landing exercises [17] [18]. Conversely, other IPPs aim to develop general fitness abilities indirectly related to the athlete's sport e.g. core stability exercises such as squats and side planks [19]. It remains unclear, which of these program types (sports specific or general IPPs) have the highest evidence-based justification for implementation regarding the positive effects on critical performance indices. Therefore, the aim of the present systematic review was to evaluate the effectiveness of evidence based IPPs on performance in young athletes.

Methods

Data sources

A systematic search of the literature, in compliance with the preferred reporting items for Systematic Reviews and Meta-Analyses [20] was conducted.

Relevant articles were identified via PubMed and Web of Science and were systematically searched between Jan 2006 until May 2019. The search terms used were "*athletes*" AND "*injury prevention*" *OR* "*exercise program*". Two authors (HM; AP) performed the literature search independently, with disagreements resolved by consensus and further consultation from a third author (JM) was sought if necessary. The search process entailed removing duplicates, screening titles, abstracts and eligible full texts. Additionally, reference lists of excluded systematic reviews, meta-analyses and reviews were manually searched to identify studies of relevance.

Eligibility criteria

Inclusion criteria: with the aim of evaluating the effectiveness on performance.

1. Individual and team sport athletes aged 11–45 years who had to have either participated in an IPP or have kept their usual training routine/standardized protocol.
2. The design of the study was restricted to randomized controlled trails (RCT) and non-randomized control trial (NRS).
3. Included at least one of the following outcome measures; a) dynamic balance (e.g. star excursion balance test), b) static balance (e.g. centre of pressure), c) power (e.g. jump height), d) strength (e.g. isokinetic strength at different angle velocities), e) endurance (e.g. repetitions per minute) or f) speed/agility (e.g. timed short distance tests) (S3 Table)
4. Had at least two measurement time points; before and after the IPP. In the event of multiple post intervention assessments, the measurement immediately post-intervention was used.

There were no limits placed on athletic level, sex or duration of the intervention. Passive IPPs, among others based on stretching, were considered irrelevant and excluded.

Data extraction

The following data was extracted from each eligible full text: the first author's last name, publication year, study design, country, follow-up period, study duration, type of sport, exposure data, subject information (sample size, dropout rate, sex and age) and intervention (name, description, type, dose, frequency, compliance, effects and categorization according to authors' definition).

Data analysis

The IPPs' were categorized into three sections: 1) general, 2) mixed and 3) sports-specific. Each exercise that comprises an IPP was determined by the authors and added to a category which most aptly describes them; **category a):** The exercise directly relates to the movement, task or skill performed in the athletes' respected sports (e.g. a jumping motion in a program for basketball players) or **category b)** The exercise focuses on developing general physical abilities that do not directly pertain to the movement, task or skill performed in their sport. Sports-specific IPPs were defined as IPPs which comprise primarily (i.e. >80%) of exercises from category a and general IPPs were defined as programs that consisted of exercises (i.e. >80%) from that described in category b. In cases where the components did not show a clear direction, i.e. either 21–79% not related or related to the sport, were categorized as a mixed IPP. For clarity, this is illustrated in the already published first part of this review assessing the effects of IPPs on injury rates [16]. This review focuses on the various IPPs effect on performance. Table 1 displays a summary of the main characteristics in each IPP type (e.g. intervention/control group, age range, sport type and target extremity). Table 2 identifies the tests used to assess the performance parameters in the included studies whereas Table 3 displays the overall effectiveness of the IPPs on various sports performance parameters. For clarity, the effectiveness of a given IPP is determined whether within each study the outcome measurement is significantly improved versus control e.g. In a study by Hermassi et al. [21], they used what would be defined as a sports specific IPP and assessed whether that increased throwing velocity [$m.s^{-1}$], which is a measurement of power. It could then be determined whether the intervention group improved this measure significantly against the control group.

Risk of Bias assessment

The internal validity of the RCTs was assessed using the Cochrane Collaborations' risk of bias assessment tool [22]. Independently, the two authors (HM; AP) examined the studies of interest for the following sources of bias: selection (sequence generation and allocation concealment), performance (blinding of participants/personnel), detection (blinding outcome

Table 1. Study characteristics of included studies.

IPP Type	General	Mixed	Sports Specific
Study No.	11	8	9
N Intervention	185	179	119
N Control	167	179	100
Age	12–25	10–21	14–27
Intervention type	9 Training programs [23] [24] [25] [26] [19] [27] [28] [29] [30] 2 Warm-up [31] [32]	3 Training programs [33] [34] [35] 5 Warm-up [36] [37] [14] [18] [38]	9 Training programs [39] [40] [41] [42] [43] [21] [44] [45] [46]
Target extremity	8 Lower extremity [23] [24] [25] [26] [19] [27] [28] [30] 1 Upper extremity [29]	8 Lower extremity [33] [34] [35] [36] [37] [38] [14] [18]	7 Lower extremity [39] [40] [42] [43] [21] [45] [46] 2 Upper extremity [41] [44]
Sport	6 Football [23] [24] [26] [32] [27] [30] 1 Handball [25]	7 Football [10] [14] [18] [47] [35] [17] [33] 1 Basketball [34]	3 Basketball [39] [40] [46] 1 Handball [25] [21] 1 Football [43] 1 Hockey [42] 1 Track Sprinting [45] 1 Tennis [44],

Table 2. Outcome measures used in the five performance categories.

Dynamic Balance	**SEBT (x3) [24] [32] [48]**	**SEBT [17]**	**SEBT [39]**
Static Balance	One legged stand time [19] Time to stabilization, BESS, COP_{sway} [31]	Stability [18] COP [34] Single leg balance [17]	COP, Centre of gravity control [49]
Power	CMJ (x8) [30] [25] [28] [29] [27] [26] [23] [32], Squat jump (x4) [23] [25] [30] [26], Vertical Jump [19] 5-jump test, Drop jump [30] 20 & 40cm drop jump [28], Kicking performance [27], Absolute power, Maximal force [23], Triple hop distance [32]	CMJ (x3) [14] [33] [36] Average kicking distance, Vertical jump height [35] 3 Step jump [14], Vertical drop jump, distance kick [36] Single legged triple hop [17]	Ball throwing velocity [21] Running based anaerobic sprint test- Max. and mean power [40] Standing throw, Jumping throw, Throw with run [25]
Strength/ Endurance	Push up, Sit up, Grip strength [19], Bench press, Full squat, [29] Trunk muscular strength [27] Hip extension/flexion torque [26] Maximal lower limb load [25]	Squatting, Stepping, Lunging, Leg raising, Push up [18], Concentric/eccentric knee flexors/extensors peak torque [47] Core stability [14] Isokinetic leg strength, Isometric leg strength [36]	Concentric and eccentric external/internal rotation [21] Concentric/Eccentric shoulder peak torque/Total work [44] Pull over, Bench press [25]
	Yo-Yo test, Maximal aerobic speed [30] 2.4km [28]	Bar jumps [17]	Running based anaerobic sprint test- fatigue [40]
Speed/ Agility	20m sprint (x4) [30] [26] [28] [32], 10m sprint (x3) [26] [30] [32], T-test (x2) [30] [27], Side Step [19] 10m, 30m sprint [30], 15m sprint [26] Sprint velocity [23] Illinois agility test [32]	Agility, 20m sprint [14] 40m sprint, Shuttle run [36] 9.1m,18.2m,27.3m,36.6m sprint, Illinois and pro agility test [33]	T-test [40] [46] Total/fastest/slowest/mean sprint time [42] 7.32m 10m Shuttle and Timed circuit test [43]

BESS, Balance Error Scoring System; **CMJ**, Countermovement jump; **m**, meters **min.**, minutes; **COP**, Centre of Pressure; **SEBT**, Star Excursion Balance Test; **s**, seconds

assessors), attrition (incomplete outcome data), reporting (selective reporting) and other potential bias.

Results

The initial search identified 6881 potentially relevant studies (Fig 1). Duplicates, reports, general review articles, current concepts, commentaries, systematic reviews and meta-analyses as well as studies that did not match the inclusion criteria were excluded following screening of titles and abstracts. Forty-four studies remained for full text screening and of which, 22 were found to be ineligible. Additional evidence was retrieved via reference lists of previously excluded reviews which unearthed a further 6 suitable articles. Consequently, 28 studies met the inclusion criteria for the investigation.

Table 3. A summary of the effectiveness of each outcome measure in each of the 3 IPP types.

IPP Type	**General**		**Mixed**		**Sports Specific**	
	Eff.	**NE**	**Eff.**	**NE**	**Eff.**	**NE**
Balance	*4*	*3*	*4*	*1*	*2*	*1*
Effectiveness	*57%*		*80%*		*66%*	
Power	7	14	4	5	5	1
Effectiveness	*33%*		*44%*		*83%*	
Strength	4	4	2	8	6	2
Effectiveness	*50%*		*20%*		*75%*	
Endurance	*3*	*0*	*1*	*0*	*1*	*0*
Effectiveness	*100%*		*100%*		*100%*	
Speed/Agility	*4*	*10*	*3*	*7*	*8*	*5*
Effectiveness	*29%*		*30%*		*62%*	

Eff.: effective; **NE:** not effective

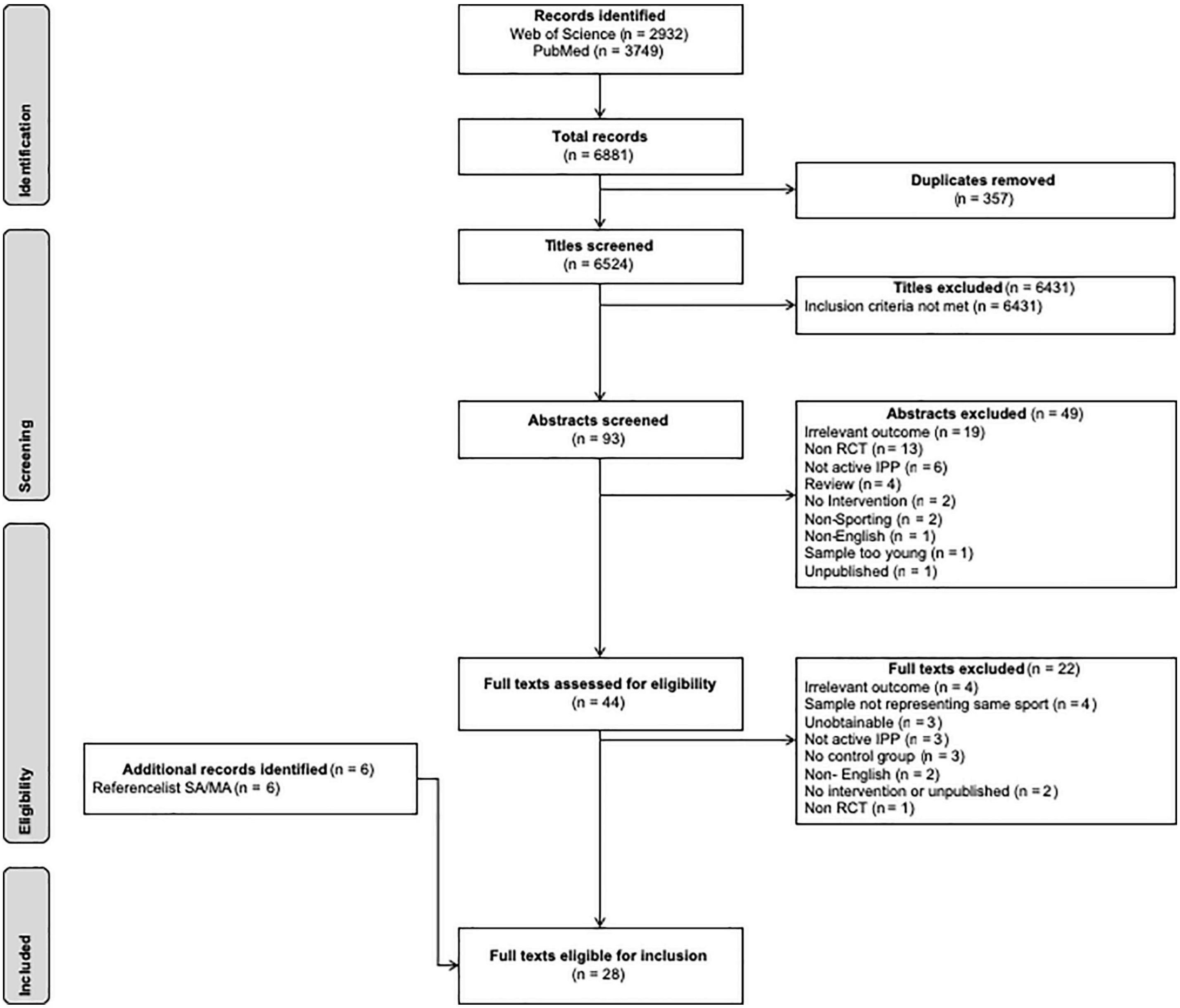


Fig 1. Flowchart for screening and selection of studies according to PRISMA.

Table 1 displays a summary of the study characteristics. Twenty one studies applied a training intervention in addition to regular training and 7 studies used a warm-up intervention in place of the usual warm-up routine [14] [18] [31] [32] [36] [37] [38]. Stated compliance in the 10 IPPs ranged from 24–100% [50] [26] [27] [48] [17] [18] [10] [14] [33] [21] whereas 18 were undefined [23] [24] [25] [19] [28] [29] [30] [47] [34] [35] [39] [40] [41] [42] [51] [49] [44] [46].

The S1–S3 Tables give a comprehensive overview of the interventions, study characteristics and results. In the interest of coherence and readability, the programs will be collectivized into first either a) General, b) Mixed or C) Sports-specific. No specific mention will be given to the type of intervention (e.g. strength program, plyometrics etc.) but references will be giving for every reporting of a result.

General IPPs

Of the 11 studies comprising of general IPPs, 3 assessed dynamic balance [32] [31] [24], 2 static balance [31] [19], 4 strength (10 overall measures) [19] [29] [27] [26], 8 power (20 overall measures) [30] [28] [29] [27] [26] [23] [32], 2 endurance [30] [28] and 7 speed/agility (14 overall measures) [19] [30] [28] [27] [26] [23] [32] as a function of performance. The IPPs ranged from 1–29 different exercises lasting between 10-90min. and were applied 2 to 5 times per week for the duration of 4 to 13 weeks.

Balance. Three studies measured dynamic balance as expressed by the Star Excursion Balance Test (SEBT). Two of which (warm up programs) show no improvement in SEBT score [32] [48]. The same general neuromuscular warm-up program assessed 3 measures of static balance and only the Balance Error Scoring System (BESS) improved whilst Time-to-Stabilization and Centre of Pressure (COP) sway were unaffected relative to control [31]. In active interventions, SEBT was improved in 3 out of 6 measures in one study [24] whilst another found that closed eyes one legged stand significantly improved compared to control [19].

Power. There were 17 measures of various jump types across 8 active IPPs [19] [30] [28] [29] [27], [26] [23] and a single warm up program [32]. Six jump measures were increased compared to control and 11 did not differ significantly. The other power measure that improved was maximal force [23]. Kicking performance [27] and absolute power [23] failed to show a significant increase versus control.

Strength. Bench press, full squat weight, hip extension torque (although not hip flexion torque) and maximal lower limb load were significantly improved following active IPPs [29] [26] [25]. Conversely, push-up/sit-up number and grip strength [19] remained unaffected. A general warm up IPP led to no significant changes in trunk muscular strength [27].

Endurance. Maximal aerobic speed / yo-yo test and a 2.4km time trial increased significant compared to control group following 2 active strength protocols [30], [28].

Speed/Agility. Five studies measured 9 types of sprint speed (e.g. 20m), only 3 improved following 2 different active strength IPPs [30] [28]. Highest sprint velocity was significantly improved following a warm up plyometric protocol [23].

For tests of agility, 4 measures in 4 different studies failed to yield significant improvement. Side Step [19], [30] [27] [32] failed to improve compared to control.

Mixed IPPs

Of the 8 studies comprising of mixed IPPs, a single study assessed dynamic balance [17] as an aspect of performance, 3 static balance [17] [34] [18], 5 power (9 overall measures) [35] [14] [36] [17] [33], 4 strength (10 overall measures) [18] [38] [14] [36], 3 speed/agility (10 overall measures) [14] [36] [33] and 1 endurance [17]. The IPPs comprised 12–41 different exercises lasting between 10-120min. and were applied 2–5 times per week for the duration of 4 weeks-20 months.

Balance. The FIFA 11+ is a neuromuscular warm-up program and is commonly applied in studies attempting to determine its efficacy, one study [17] found a significant improvement in 2 from 6 possible angles in the dynamic SEBT following FIFA 11+. As for static balance, the single legged eyes closed balance test in both legs [17] was significantly improved compared to the control group. Another study that applied the FIFA 11+ program failed to show improvement in functional stability amongst young male footballers [18]. A 20 week neuromuscular warm-up showed significant improvement COP, particularly in the dominant limb [34].

Power. CMJ was measured in 3 separate studies [33] [14] [36]. Following the PEP program and the F-MARC 11, this measure was not improved [33] [36] although a study implementing the same program (F-MARC 11) found significant improvement versus control [14]. Football kicking distance, vertical jump height and improved post plyometric training [35]. The F-MARC 11 program did not result in significant improvement in vertical drop jump [36] or kick distance [36] and the updated version of the program (FIFA 11+) did not significantly increase single legged triple hop distance [17].

Strength. A study that measured knee extensor torque found a concentric improvement without much effect on the knee flexors when working eccentrically following a specific knee IPP [47]. Amongst three studies, the F-MARC 11 or the FIFA 11+ program resulted in no difference in core stability [14] isokinetic/isometric leg strength [36], squatting, stepping, lunging, leg raising or push-ups [18].

Endurance. The only measure of endurance that was found for mixed programs was bar jumps which was assessed before and after the FIFA 11+ [17] and was significantly improved.

Speed/Agility. Sprint times for 20m [14] 27.3m and 36.6m [33] in 2 separate studies were significantly improved post IPP. Agility [14] 40m sprint, Shuttle run [36] 9.1m,18.2m sprint, Illinois and pro agility test [33] were all unaffected by the F-MARC 11 [14] [36] or the PEP program [33].

Sport-specific IPPs

Of the 9 studies comprising of sports specific IPPs, 1 assessed dynamic balance [39], 1 static balance (2 overall measures) [45], 3 power (7 overall measures) [21] [40] [25], 3 strength (8 overall measures) [21] [44] [25], 4 speed/agility (9 overall measures) [40] [42] [43] [46] and 1 endurance [40] as a function of performance. The IPPs ranged from 1–20 different exercises lasting between 15-60min. and were applied 2 to 5 times per week for the duration of 3 to 8 weeks.

Balance. Following plyometric training, the SEBT improved in 6 out of 8 directions [39], The progressive sprinter-specific proprioception training program found significantly better stability with eyes open in the medial-lateral plane in the experimental group. Furthermore, gravity control was measured and did not yield in statistical significant differences, although differences on the right side and the back were reported [45].

Power. Strength training led to no significant increase in ball throwing velocity compared to control [21]. Conversely, mid-season resistance training improved the standing, jumping and running throw in handball athletes [25]. Sports specific plyometric training led to significant mean and maximum (w) anaerobic sprinting power gains [40].

Strength. Eccentric and concentric shoulder peak torque improved during a strength training program albeit not compared to control. The total work of the internal/external rotators did significantly improve in the same study [44]. Another strength training IPP only found concentric and eccentric improvements of the external/internal rotators of the non-dominant arm [21]. Sports-specific resistance training resulted in increased strength (kg) in pull overs and bench press [25].

Endurance. Running based anaerobic sprint test- fatigue improved following plyometric training in a single study investigating endurance [40].

Speed/Agility. Two studies implementing plyometric training and dynamic balance training improved post-protocol t-test scores [40] [46]. A significant between group interaction effect was observed at total, fastest, slowest and mean linear sprint time as well as shuttle sprints with and without a ball after completing sprint training [42], although no difference was observed during the slalom runs. A group x time effect was observed following soccer specific training but no significant difference in the 7.32m, 10m Shuttle and Timed circuit test were apparent [43].

Risk of bias assessment

The results of the methodological quality assessment across all included RCTs is presented in Fig 2.

Discussion

The present systematic review presented the effectiveness of general versus sports-specific IPPs and their role on enhancing key indices of athletic performance. The current consensus amongst coaches is that the highest performance benefit can only be achieved through implementation of a sports-specific IPPs. Although not entirely in line of the presented evidence, there is no apparent necessity to deviate from this existing perception as the results are largely positive in favor of specificity to the sport concerned.

The first part of the current research attempted to define the definition of sports-specific. This has been discussed in a previous review [16]. The intention was to define a sports specific program as one that consisted of maneuvers whereby >80% are used in the sport itself. e.g. starting, stopping, twisting, turning, running, jumping, landing, shuffling, pushing, pulling, hitting, throwing, catching, hopping, accelerating, decelerating, sliding, blocking. Isolated actions such as planks, bench press and squats do not directly correspond to the movement pattern involved in any sport and would be considered as general movements. The research attempted to address this question and to identify possible avenues for future research.

As summarized in Table 3, general programs arguably show a modest effect (29–57%) on all performance outcomes except for endurance (100%). Mixed interventions resulted in a relatively low success rate in measurements of strength (20%), power (44%) and speed (30%) whereas the sports-specific programs shown significant improvement in every measurement of endurance, 75% of strength measurements and power, balance and speed/agility were largely positive in their effectiveness (63–66%).

The sports-specific IPPs generally show a more positive relationship in terms of effectiveness in power, strength and speed/agility when compared to the other types. In a sporting environment, being able to perform repeated explosive movements such as rapid acceleration/deceleration (i.e. power) and changing direction (i.e. agility) is critical to success and injury risk reduction. Power production is stated to be influenced by motor unit synchronization, neuronal adaptation [25] and increases in neural activation [30]. Whereas agility is linked to ATP production [40], ATP resynthesis [42] and thereby reducing the time required for voluntary muscle activation. This reduction would manifest itself in faster changes in movement and therefore more agility. In this regard, the speed of movement should provide the neuromuscular stimulation to mimic that performed in the explosive movements of match play. This stimulus often comes in the form of sports-specific exercises whilst a lot of mixed and general programs fail to replicate this [32] [19] [36] [14] [27] [29] [26] [17] [33] [30] perhaps

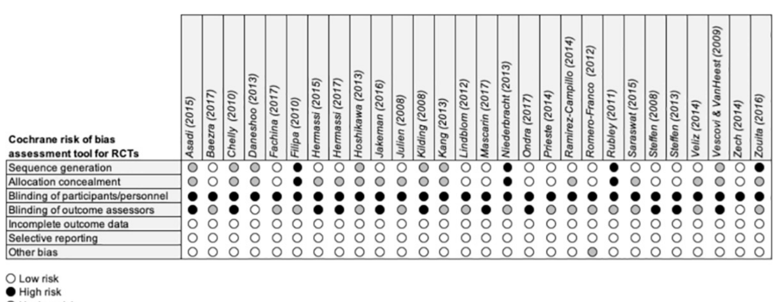


Fig 2. Risk of bias assessment of included RCTs.

thereby explaining the discrepancy in the results. A few notable exceptions is that a few general programs did see increases in at least one measure of power following strength training programs [30] [25] [28] so it would appear IPPs that progress their exercises (by either increasing the resistance or degree of difficulty) or have movements in the program that directly replicate the outcome of the assessed movement (e.g. in the case of Chelly et al. (2010), the IPP consisted of hurdle and drop jumps and used CMJ and squat jumps as the performance outcome) lead to some improvement in power. Additionally, there were incidences of improved sprint speeds [14] [33] and jumps [14] [35] following mixed IPPs and it was suggested in the literature that plyometric exercises could have prompted these enhancements [14].

If we examine strength acquisition, all three sports specific IPPs improved versus control [25] [21] [44]. The unifying factor of these IPPs is that resistance level was progressively increased during the time period to adapt to the athletes' current strength level. In one of these studies, sports specific strengthening resulted in a tangible performance improvement in handball players [25], it was observed that all throwing types were significantly faster ($m.s^{-1}$) than their control counterparts. This type of sports specific IPP may appeal to both coaches and players alike. If for example, a team has been suffering with multiple injuries of the same nature e.g. there is a high incidence of hamstring strains in football [52], general functional strengthening may be appropriate to either add to regular training or be implemented as an off season IPP as improvements are observed [27] [19], but in regards to warm-up programs of mixed exercises (e.g. the FIFA 11 and 11+), there was less significant improvements in strength [18] [47] [14] [36]. This effect is largely attributed to the program being short in its duration of 15-20mins and not providing the necessary stimulus for strength acquisition.

When examining the influence of IPP content on balance, certainly a necessary stimulus should replicate the unpredictable high-speed environment of sport, but simultaneously, having a certain level of general thigh, hip, ankle and core strength is also an important factor [17]. In this regard, a program of mixed exercises for balance improvement may be the most efficacious [17] [34]. Interestingly, in the mixed IPPs, there were two studies that investigated dynamic or static balance following FIFA 11+ (which is also considered a warm up program) [17] [18] with the only difference one was lasting longer (16 [17] weeks compared to 6 [18]) and was coach focused (i.e. supervised and corrected). The same authors uniquely tested for and indeed found a dose-response relationship between the volume of FIFA 11+ exercises performed and the magnitude of balance performance. On this basis, dosage and compliance should be factored in to any future investigations before assessing the true effectiveness of the program itself.

A prior review suggested that regardless of low volume, the FIFA 11+ warm-up is enough to improve balance [53]. Although the FIFA 11+ was originally intended for footballers, there is no reason why these abilities cannot be transferable to other sports that replicate some of the movement patterns e.g. hockey, basketball, rugby and handball. There is a general lack of studies that measured endurance, but the ones that do show significant improvement versus the control group regardless of IPP content [30] [28] [17] [40]. Future research should perhaps consider endurance more intently as an outcome measure after IPP intervention.

Recommendations

The authors can reasonably make some creditable recommendations based on the presented results. For future researchers, it is crucial that other factors such as compliance, age and standard of the athletes, time of season, facility access and execution of exercises are well considered to reveal the true effectiveness and therefore clinical relevance of IPPs. A dose-response relationship should be investigated which has previously been found over the course of 4 months for key performance indicators [17]. For coaches, particularly of young athletes, IPPs should be applied to prevent injuries [16]. However, do coaches need to consider the IPP type to gain an additional benefit to performance? It appears that IPPs that focus on general exercises do not provide an additional performance benefit and it is therefore recommended that sports specific IPPs are desirable.

In accordance with the TRIPP model, proving an injury reducing effect of IPPs (e.g. the FIFA 11+ [54]) is often insufficient for coaches and athletes alike to yield implementation [13]. Over the course of a season, providing tangible performance benefits (short and long-term) may increase the likelihood of real-world implementation. A further recommendation is that professionals with less time constraints or recreational athletes in the off season may benefit from longer duration (>30min) neuromuscular training that has been shown to improve strength [26] [29] [44] [21] [25], balance [19] [39], power [30] [28] [29] [21] [40] speed/agility [30] [28] [40] and endurance [30] [28] [40] in several studies regardless of IPP content.

Risk of bias assessment

The overall quality of the included RCTs was considered moderate. Sequence generation and allocation concealment in these sorts of studies are inherently difficult. Moreover, blinding of participants as well as personnel for exercise interventions is not feasible. Moreover compliance of such IPPs mostly rely on self-reports; therefore, all trials showed high risk of performance and detection bias. Nevertheless, the overall quality could have been underestimated by non-transparent reporting. Accordingly, consensus statements such as CONSORT [55] should be followed across studies.

Study limitations

It could be argued that the initial search strategy lacked robustness, with only two databases searched and six articles retrieved during the manual screening of reference lists from systematic reviews. The largest limitation is the considerable number of variables of an IPP and its implementation into a study design (e.g. the repetitions and execution of exercise, the compliance rate). The sizeable number of outcome measures that assess performance lead to further complications when attempting to determine the overall IPP effectiveness. Furthermore, no standard tests exist that truly correlates with the performance and the internal validity of performance studies may thus be constrained.

Future research

In order to make recommendations regarding the effectiveness of sports-specific IPPs on athletic performance (short or long-term), future investigation need to compare effects of sports-specific versus general IPPs. To allow further comparisons, outcome measures would ideally be similar across investigations. Moreover, to control for adequate compliance/application of the IPPs, future studies have to invest more resources in following up on athletes and coaches/training staff without rising concern about potential motivational bias.

Conclusion

There are contradictory findings regarding the effect that neuromuscular programmes may have in improving physical performance. A large majority of sports-specific IPPs did improve outome measure that the researchers found to be of use so in this regard, additional sports-specific training does indeed benefit performance. It cannot be unstated, however, that key performance improvements are potentially affected by a variety of other factors (e.g. compliance, frequency, intensity, technical execution) so it is difficult to assess the magnitude of influence that specificity to the sport actually has.

Supporting information

S1 Table. Summary of intervention programs used in the included studies (alphabetical order by program). Level**: E, professional/elite/highest level; AM, amateur; d, days; min., minutes; mo., months; N/D, not described; reps, repetitions; S, season; SE, session(s); VB, volleyball; wk., weeks; yrs., years.
Values presented as mean ± standard deviation if not otherwise stated.
(DOCX)

S2 Table. Exercise interventions applied by included studies (alphabetical order). ACL, anterior cruciate ligament; BB, basketball; CPL, compliance; EX, exercise(s); KLIP Program, Knee Ligament Injury Prevention Program; LE, lower extremity; min., minutes; N/D, not described; OSTRC, Oslo Sports Trauma Research Center; PEP Program, Prevent injury and Enhance Performance Program; Pt., part/phase; reps, repetitions; s, seconds; SE, session(s); UE, upper extremity; VB, volleyball; wk., week(s); yrs., years.
(DOCX)

S3 Table. Results and conclusion of included studies (alphabetical order). "Statistically significant difference within groups (pre- post-test). * Statistically significant difference between intervention and control group (pre- post-test).
5JT, 5 jump test; BAL, balance training group; BESS, Balance error scoring system; CG control, general gravity center control; CMJ, counter movement jump; COP, center of pressure; DEO, distance covered by the center of pressure with eyes open; DJ, drop jump; ECC, Eccentric strengthening exercises; ER, external rotation; IR, internal rotation; MAS, maximal aerobic speed; MSFT, multi-stage fitness test; PLYO, plyometric training group; RAST, running-based anaerobic sprint test; RJ; rebound jump,; RombergD, Romberg index about distance; RombergS, Romberg index about surface; RombergSp, Romberg index about speed; SE, session(s); SEBT, Star Excursion Balance Test; SEO, surface covered by the center of pressure with eyes open; SLH, single leg hop; SLTH, single leg triple crossover hop; SLS, single leg squat; SpEO, speed of center of pressure movement with eyes open; THDT, Triple hop distance; TTS, time to stabilization; UNS, unstable surface exercises; VDJ, vertical drop jump; VO2max, maximal oxygen uptake; XEO, mean position center of pressure in the medial-lateral plane with eyes open; YEO, mean position center of pressure in the anterior-posterior plane with eyes open; YYIRTL1, Yo-Yo intermittent recovery test level 1; Values presented as mean ± standard deviation or Δ% or if not otherwise stated.
(DOCX)

S1 File. PRISMA checklist.
(PDF)

S2 File. Search strategy.
(XLSX)

Acknowledgments

This work was completed in partial fulfillment of a Master of Science in Clinical Exercise Science at the University of Potsdam. The authors would like to thank the staff at the Universities' Outpatient Clinic who supported this study.

Author Contributions

Conceptualization: Ashley Plummer, Hendrik Mugele, Josefine Stoll, Frank Mayer, Juliane Müller.

Data curation: Ashley Plummer, Hendrik Mugele.

Formal analysis: Ashley Plummer.

Investigation: Ashley Plummer, Hendrik Mugele.

Methodology: Ashley Plummer, Hendrik Mugele, Josefine Stoll, Juliane Müller.

Project administration: Ashley Plummer, Hendrik Mugele.

Supervision: Juliane Müller.

Validation: Ashley Plummer, Hendrik Mugele.

Visualization: Ashley Plummer, Hendrik Mugele.

Writing – original draft: Ashley Plummer, Hendrik Mugele.

Writing – review & editing: Ashley Plummer, Kathrin Steffen, Juliane Müller.

References

1. Barber Foss KD, Myer GD, Hewett TE. Epidemiology of basketball, soccer, and volleyball injuries in middle-school female athletes. Phys Sportsmed. 2014;
2. Piercy KL, Dorn JM, Fulton JE, Janz KF, Lee SM, McKinnon RA, et al. Opportunities for public health to increase physical activity among youths. American Journal of Public Health. 2015.
3. Pickett W, Molcho M, Simpson K, Janssen I, Kuntsche E, Mazur J, et al. Cross national study of injury and social determinants in adolescents. Inj Prev. 2005;
4. Emery CA, Tyreman H. Sport participation, sport injury, risk factors and sport safety practices in Calgary and area junior high schools. Paediatr Child Health (Oxford). 2009;
5. Hootman JM, Dick R, Agel J. Epidemiology of Collegiate Injuries for 15 Sports: Summary and Recommendations for Injury Prevention Initiatives. J Athl Train. 2007; 42(2):311–9. PMID: 17710181
6. Olsen O, Myklebust G, Engebretsen L, Bahr R. Injury Mechanisms for Anterior Cruciate Ligament Injuries in Team Handball. Americen J Sport Med. 2004; 32(4):1002–12.
7. Ekstrand J, Hägglund M, Waldén M. Epidemiology of muscle injuries in professional football (soccer). Am J Sports Med. 2011;
8. O'Brien J, Finch CF. Injury Prevention Exercise Programs for Professional Soccer: Understanding the perceptions of the end-users. Clin J Sport Med. 2017;
9. Soligard T, Myklebust G, Steffen K, Holme I, Silvers H, Bizzini M, et al. Comprehensive warm-up programme to prevent injuries in young female footballers: Cluster randomised controlled trial. BMJ. 2009;
10. Steffen K, Bakka HM, Myklebust G, Bahr R. Performance aspects of an injury prevention program: A ten-week intervention in adolescent female football players. Scand J Med Sci Sport. 2008; 18(5):596–604.
11. Mandelbaum BR. Effectiveness of a Neuromuscular and Proprioceptive Training Program in Preventing Anterior Cruciate Ligament Injuries in Female Athletes: 2-Year Follow-up. Am J Sports Med. 2005; 33 (7):1003–10. https://doi.org/10.1177/0363546504272261 PMID: 15888716
12. Barber-Westin SD, Hermeto AA, Noyes FR. A six-week neuromuscular training program for competitive junior tennis players. J Strength Cond Res. 2010;
13. Finch C. A new framework for research leading to sports injury prevention. J Sci Med Sport. 2006; 9(1–2):3–9. https://doi.org/10.1016/j.jsams.2006.02.009 PMID: 16616614
14. Kilding AE, Tunstall H, Kuzmic D. Suitability of FIFA's "The 11" training programme for young football players—Impact on physical performance. J Sport Sci Med. 2008; 7(3):320–6.
15. Wilmore JH, Costill DL. Physiology of Sport and Exercise. J Athl Train. 2005;
16. Mugele H, Plummer A, Steffen K, Stoll J, Mayer F, Müller J. General versus sports-specific injury prevention programs in athletes: A systematic review on the effect on injury rates. PLoS One. 2018;
17. Steffen K, Emery CA, Romiti M, Kang J, Bizzini M, Dvorak J, et al. High adherence to a neuromuscular injury prevention programme (FIFA 11+) improves functional balance and reduces injury risk in Canadian youth female football players: a cluster randomised trial. Br J Sports Med. 2013; 47(12):794–802. https://doi.org/10.1136/bjsports-2012-091886 PMID: 23559666

18. Baeza G, Paredes G, Vega P, Monrroy M, Gajardo-Burgos R. Effect of "fifa 11+" on the pattern of fundamental movements in under-14 soccer players. Rev Bras Med do Esporte. 2017; 23(6).

19. Kang SH, Kim CW, Kim Y Il, Kim KB, Lee SS, Shin K ok. Alterations of Muscular Strength and Left and Right Limb Balance in Weightlifters after an 8-week Balance Training Program. J Phys Ther Sci [Internet]. 2013; 25(7):895–900. Available from: http://jlc.jst.go.jp/DN/JST.JSTAGE/jpts/25.895?lang=en&from=CrossRef&type=abstract https://doi.org/10.1589/jpts.25.895 PMID: 24259879

20. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Bmj. 2009; 89(9):873–80.

21. Mascarin NC, de Lira CAB, Vancini RL, da Silva AC, Andrade MS. The effects of preventive rubber band training on shoulder joint imbalance and throwing performance in handball players: A randomized and prospective study. J Bodyw Mov Ther. 2017; 21(4):1017–23. https://doi.org/10.1016/j.jbmt.2017.01.003 PMID: 29037617

22. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011; 343(oct18 2):d5928–d5928. https://doi.org/10.1136/bmj.d5928 PMID: 22008217

23. Chelly MS, Ghenem MA, Abid K, Hermassi S, Tabka Z, Shephard RJ. Effects of in-season short-term plyometric training program on leg power, jump-and sprint performance of soccer players. J Strength Cond Res. 2010; 24(10):2670–6. https://doi.org/10.1519/JSC.0b013e3181e2728f PMID: 20844458

24. Filipa A, Byrnes R, Paterno M V, Myer GD, Hewett TE. Neuromuscular training improves performance on the star excursion balance test in young female athletes. J Orthop Sports Phys Ther. 2010;

25. Hermassi S, Chelly MS, Fieseler G, Bartels T, Schulze S, Delank KS, et al. Effects of In-Season Explosive Strength Training on Maximal Leg Strength, Jumping, Sprinting, and Intermittent Aerobic Performance in Male Handball Athletes. Sportverletzung-Sportschaden. 2017;

26. Hoshikawa Y, Iida T, Muramatsu M, Ii N, Nakajima Y, Chumank K, et al. Effects of stabilization training on trunk muscularity and physical performances in youth soccer players. J Strength Cond Res. 2013;

27. Prieske O, Muehlbauer T, Borde R, Gube M, Bruhn S, Behm DG, et al. Neuromuscular and athletic performance following core strength training in elite youth soccer: Role of instability. Scand J Med Sci Sport. 2016;

28. Ramirez-Campillo R, Alvarez C, Henriquez-Olguinu C, Baez EB, Martinez C, Andrade DC, et al. Effects of Plyometric Training on Endurance and Explosive Strength Performance in Competitive Middle-and Long-Distance Runners. J Strength Cond Res. 2014;

29. Veliz RR, Requena B, Suarez-Arrones L, Newton RU, De Villarreal ES. Effects of 18-week in-season heavy-resistance and power training on throwing velocity, strength, jumping, and maximal sprint swim performance of elite male water polo players. J Strength Cond Res. 2014;

30. Zouita S, Zouita ABM, Kebsi W, Dupont G, Ben Abderrahman A, Ben Salah FZ, et al. Strength Training Reduces Injury Rate in Elite Young Soccer Players During One Season. Journal of strength and conditioning research / National Strength & Conditioning Association. 2016.

31. Zech A, Klahn P, Hoeft J, Zu Eulenburg C, Steib S. Time course and dimensions of postural control changes following neuromuscular training in youth field hockey athletes. Eur J Appl Physiol. 2014; 114(2):395–403. https://doi.org/10.1007/s00421-013-2786-5 PMID: 24318788

32. Lindblom H, Waldén M, Hägglund M. No effect on performance tests from a neuromuscular warm-up programme in youth female football: A randomised controlled trial. Knee Surgery, Sport Traumatol Arthrosc. 2012; 20(10):2116–23.

33. Vescovi JD, VanHeest JL. Effects of an anterior cruciate ligament injury prevention program on performance in adolescent female soccer players. Scand J Med Sci Sport. 2010; 20(3):394–402.

34. Lukas O, Natesta P, Bizovska L, Kubonova E, Svoboda Z. Effect of in-season neuromuscular and proprioceptive training on postural stability in male youth basketball players. Acta Gymnica. 2017;

35. Rubley MD, Haase AC, Holcomb WR, Girouard TJ, Tandy RD. The effect of plyometric training on power and kicking distance in female adolescent soccer players. J Strength Cond Res. 2011;

36. Steffen K, Myklebust G, Olsen OE, Holme I, Bahr R. Preventing injuries in female youth football—A cluster-randomized controlled trial. Scand J Med Sci Sport. 2008; 18(5):605–14.

37. Steffen K, Emery CA, Romiti M, Kang J, Bizzini M, Dvorak J, et al. High adherence to a neuromuscular injury prevention programme (FIFA 11+) improves functional balance and reduces injury risk in Canadian youth female football players: a cluster randomised trial. Br J Sports Med [Internet]. 2013; 47(12):794–802. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23559666 https://doi.org/10.1136/bjsports-2012-091886 PMID: 23559666

38. Daneshjoo A, Mokhtar AH, Rahnama N, Yusof A. The Effects of Comprehensive Warm-Up Programs on Proprioception, Static and Dynamic Balance on Male Soccer Players. PLoS One. 2012; 7(12):1–10.

39. Asadi A, Saez De Villarreal E, Arazi H. The Effects of Plyometric Type Neuromuscular Training on Postural Control Performance of Male Team Basketball Players. J Strength Cond Res. 2015;

40. De Freitas Guina Fachina RJ, Martins DS, Montagner PC, Borin JP, Vancini RL, Dos Santos Andrade M, et al. Combined plyometric and strength training improves repeated sprint ability and agility in young male basketball players. Gazz Medica Ital Arch per le Sci Mediche. 2017;

41. Hermassi S, Van Den Tillaar R, Khlifa R, Chelly MS, Chamari K. Comparison of In-Season-Specific Resistance vs. A Regular Throwing Training Program on Throwing Velocity, Anthropometry, and Power Performance in Elite Handball Players. J Strength Cond Res. 2015;

42. Jakeman JR, McMullan J, Babraj JA. Efficacy of a four-week uphill sprint training intervention in field hockey players. J Strength Cond Res. 2016;

43. Jullien H, Bisch C, Largouét N, Manouvrier C, Carlentg CJ, Amiard V. Does a short period of lower limb strength training improve performance in field-based tests of running and agility in young professional soccer players? J Strength Cond Res. 2008;

44. Niederbracht Y, Shim AL, Sloniger MA, Paternostro-Bayles M, Short TH. Effects of a shoulder injury prevention strength training program on eccentric external rotator muscle strength and glenohumeral joint imbalance in female overhead activity athletes. J Strength Cond Res. 2008; 22(1):140–5. https://doi.org/10.1519/JSC.0b013e31815f5634 PMID: 18296967

45. Romero-Franco N, Martínez-López E, Lomas-Vega R, Hita-Contreras F, Martínez-Amat A. Effects of proprioceptive training program on core stability and center of gravity control in sprinters. J Strength Cond Res. 2012; 26(8):2071–7. https://doi.org/10.1519/JSC.0b013e31823b06e6 PMID: 21997455

46. Saraswat A, Malhotra D, Sivaram C. EFFECT OF DYNAMIC BALANCE TRAINING ON AGILITY IN MALE BASKETBALL PLAYERS. Int J Physiother. 2015; 2(5):798–803.

47. Daneshjoo A, Mokhtar AH, Rahnama N, Yusof A. The effects of injury prevention warm-up programmes on knee strength in male soccer players. Biol Sport. 2013;

48. Zech A, Klahn P, Hoeft J, Zu Eulenburg C, Steib S. Time course and dimensions of postural control changes following neuromuscular training in youth field hockey athletes. Eur J Appl Physiol. 2014; 114 (2):395–403. https://doi.org/10.1007/s00421-013-2786-5 PMID: 24318788

49. Romero-Franco N., Martínez-López E., Lomas-Vega R., Hita-Contreras F., & Martínez-Amat A. Effects of proprioceptive training program on core stability and center of gravity control in sprinters. J Strength Cond Res. 2012; 26(8):2071–7. https://doi.org/10.1519/JSC.0b013e31823b06e6 PMID: 21997455

50. Lindblom H, Waldén M, Hägglund M. No effect on performance tests from a neuromuscular warm-up programme in youth female football: A randomised controlled trial. Knee Surgery, Sport Traumatol Arthrosc. 2012; 20(10):2116–23.

51. Aucouturier J, Boissière J, Pawlak-Chaouch M, Cuvelier G, Gamelin FX. Effect of dietary nitrate supplementation on tolerance to supramaximal intensity intermittent exercise. Nitric Oxide—Biol Chem. 2015; 49:16–25.

52. Croisier J-L, Ganteaume S, Binet J, Genty M, Ferret J-M. Strength imbalances and prevention of hamstring injury in professional soccer players: a prospective study. Am J Sports Med. 2008; 36(8):1469–75. https://doi.org/10.1177/0363546508316764 PMID: 18448578

53. Thorborg K, Krommes KK, Esteve E, Clausen MB, Bartels EM, Rathleff MS. Effect of specific exercise-based football injury prevention programmes on the overall injury rate in football: a systematic review and meta-analysis of the FIFA 11 and 11+ programmes. Br J Sports Med. 2017;bjsports-2016-097066

54. Al Attar WSA, Soomro N, Pappas E, Sinclair PJ, Sanders RH. How Effective are F-MARC Injury Prevention Programs for Soccer Players? A Systematic Review and Meta-Analysis. Sport Med. 2016; 46 (2):205–17.

55. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Bmj. 2010; 10(1):1–28.

8

The effects of short term detraining and retraining on physical fitness in elite soccer players

Chang Hwa Joo*

Department of Football Science, Honam University, Gwangsan-gu, Gwangju, South Korea

* footballer@honam.ac.kr

Editor: Alessandro Zagatto, Sao Paulo State University - UNESP, BRAZIL

Competing interests: The author has declared that no competing interests exist.

Abstract

Purpose

The aim of this study was to examine the effects of aerobic high-intensity training with reduced volume and training cessation on body composition and physical fitness after the end of season and the time required to recapture physical fitness with intensified retraining following two weeks of detraining in elite soccer players.

Method

Twenty male semi-professional soccer players participated in this study. The soccer players were assigned to either a group that completed high-intensity aerobic training (HAT, n = 10) or to a detraining and retraining group (DHAT, n = 10) for a 5-week period immediately after the end of the season. The first 2 weeks of the period, members of the HAT group performed high-intensity aerobic exercise (80–90% of HRmax, 12 min × 3, three times per week), whereas members of the DHAT group abstained from any physical activity. During the subsequent 3 weeks, members of both the HAT and DHAT groups completed high-intensity aerobic exercise. Exercise performance testing and body composition analysis were performed before; after 2 weeks of detraining; and at 1, 2 and 3 weeks of retraining.

Results

Intensified high-intensity training for 5 weeks maintained the performance in the Yo-Yo Intermittent Recovery level 2 test (Yo-Yo IR2) and repeated sprints at any time point ($P > 0.05$). However 2 weeks of detraining resulted in significant decreases in the performance on the Yo-Yo IR2 ($P < 0.01$) and repeated sprints test ($P < 0.05$). Performance on the Yo-Yo IR2 enhanced after 2 weeks of retraining and was maintained up to 3 weeks after retraining, with no significant differences between conditions ($P > 0.05$). In addition, repeated sprint performance markedly decreased after the detraining period ($P < 0.05$) and was continuously lower compared to the baseline at 2 weeks after retraining ($P < 0.05$). Furthermore, this value reached baseline level at the end of the experimental period ($P > 0.05$). There were no significant differences between conditions in body composition, performance of agility, or sprint ability throughout the 5-week experimental period ($P > 0.05$).

Conclusions

The present data suggest that short-term detraining after the competitive season can markedly decrease performances in the Yo-Yo IR2 test and repeated sprints. To return to a previous level of ability on the Yo-Yo IR2 and/or sprint test with retraining through high-intensity aerobic training after a period of detraining, a similar or longer period of retraining is required. However, the high-intensity training with reduced amount of training after competitive season can prevent reductions in physical fitness.

Introduction

Soccer is a high intensity intermittent exercise that requires a high level of physical fitness for players to successfully perform in the game. Elite soccer players perform 587±133 m of high-speed running (19.8–25.2 km/h) and 184 ± 87 m of sprinting (>25.2 km/h) during a typical game [1]. The total distance of high-intensity running depends on the position of the player and team success in a league [2]. The amount of high-intensity running performed during a game also depends on the competitive standards between leagues: top-class professional soccer player perform more high-intensity running compared with moderate professional soccer players [3]. Thus, high level of physical performance is an important factor in determining team success in soccer.

Due to the high intensity performance required in soccer, players should perform systematic and scientific physical fitness training. Several studies have shown that high-intensity training improves soccer players' fitness levels and skills, such as sprint, strength, and speed endurance [4, 5]. The organization of fitness training for soccer players varies according to the time frame of the periodization along with changes in training volume and intensity. These changes seek to the optimize player's physical condition and minimize injury [6]. For example, training is conducted to improve physical fitness during the preseason in preparation for the impending competitive season [7, 8].

Elite soccer players normally cease training or perform training with reduced volume and lower intensity for more than two weeks after the end of the competitive season for physical and mental recovery. A prolonged period of rest after the competitive season causes the partial or complete loss of training-induced physiological and performance adaptations, which is defined as detraining [9]. The magnitude of changes during training-induced adaptations after detraining is different depending on the fitness level and the duration of training cessation or insufficient training [9]. Three to six weeks of detraining did not result in changes in aerobic capacity and muscle strength in recreational players and untrained individuals [10–12]. However, decreases in physical fitness are inevitable after such a period of detraining in well-trained elite players who have a relatively higher level of fitness compared to recreational players [9, 13]. Unlike reduced physical fitness after a prolonged period of detraining in elite players, the effects of short-term detraining (~2 weeks) on fitness are controversial. Buchheit et al. [14] observed that short-term detraining after a competitive season improved levels of strength and cardiorespiratory fitness in Australian football players [14]. In contrast, several studies reported that physical fitness was reduced after a short-term detraining period in elite soccer players [5, 15]. The reasons for these contrasting results are not apparent, but may be due to differences in sports and testing methods.

During the preseason, the aim of training is mainly to improve physical fitness, while during the in-season period, it is performed to develop playing strategies and to enhance performance, while maintaining physical fitness. High-intensity training is a more efficient method of inducing skeletal muscle adaptation in comparison to moderate-intensity training [16]. High-intensity aerobic training has been widely used by athletes to improve physical fitness during the preseason. Indeed, high-intensity aerobic training consisting of 4 bouts of 4 min at 90–95% of the maximum heart rate during the preseason significantly improved aerobic fitness and match performance in soccer players [17]. Those results indicated that high-intensity aerobic training might be effective at improving the physical fitness of soccer players and inducing rapid training adaptation in skeletal muscle during the preseason.

In order to start the season without injury, athletes must gradually improve their post-season, resting period-induced reduction in physical fitness with an appropriate exercise intensity and volume. However, there is limited information available regarding the effects of retraining during pre-season training in well-trained elite soccer players. Therefore, the aim of the study was to investigate 1) the effects of aerobic high-intensity training with reduced volume and training cessation on body composition and physical fitness after the end of season and 2) the time required to return to the previous level of physical fitness with intensified retraining following two weeks of detraining in semi-professional soccer players.

Materials and methods

Participants

Twenty semi-professional male Korea soccer players (age: 22.1±1.8 years, height: 175.5±4.7 cm). The Korean professional soccer league is divided into K League Classic (first division) and K League Challenge (second division). The semi-professional league consists of the National League and K3 Leagues (K3 League Advanced [12 teams] and K3 League Basic [8 teams]). The soccer players participating in this study were members of K3-league teams. All participants had experience of elite soccer players for at least more than 7 years. All participants were non-smokers, no history of neurological disease or musculoskeletal abnormality and none were under any pharmacological treatment during the course of the study.

Ethics statement

Before testing, all participants gave written informed consent to participate after details and procedures of the study had been fully explained. All of the fitness testing and exercise were performed in the research institute for sport and exercise science at Honam Unviersity. All of the experimental protocols and related procedures were approved by the ethical committee of Honam University.

Intervention period and training

All players participating in the study trained for more than 2 hours per day for 4–5 times per week (excluding matches) during the previous season. An independent research assistant selected the 20 participants from among 35 players who were between 20 and 23 years of age by drawing a sealed envelope containing a player's name followed by drawing another sealed envelope containing the name of the group to which they were assigned (i.e., high-intensity aerobic training (HAT) or detraining and high-intensity aerobic training (DHAT) group). The two-week detraining period started immediately after the last match of the season. The fitness tests were conducted two days and one day before the last match as a pre-test; after two weeks of detraining; and at one, two, and three weeks of retraining.

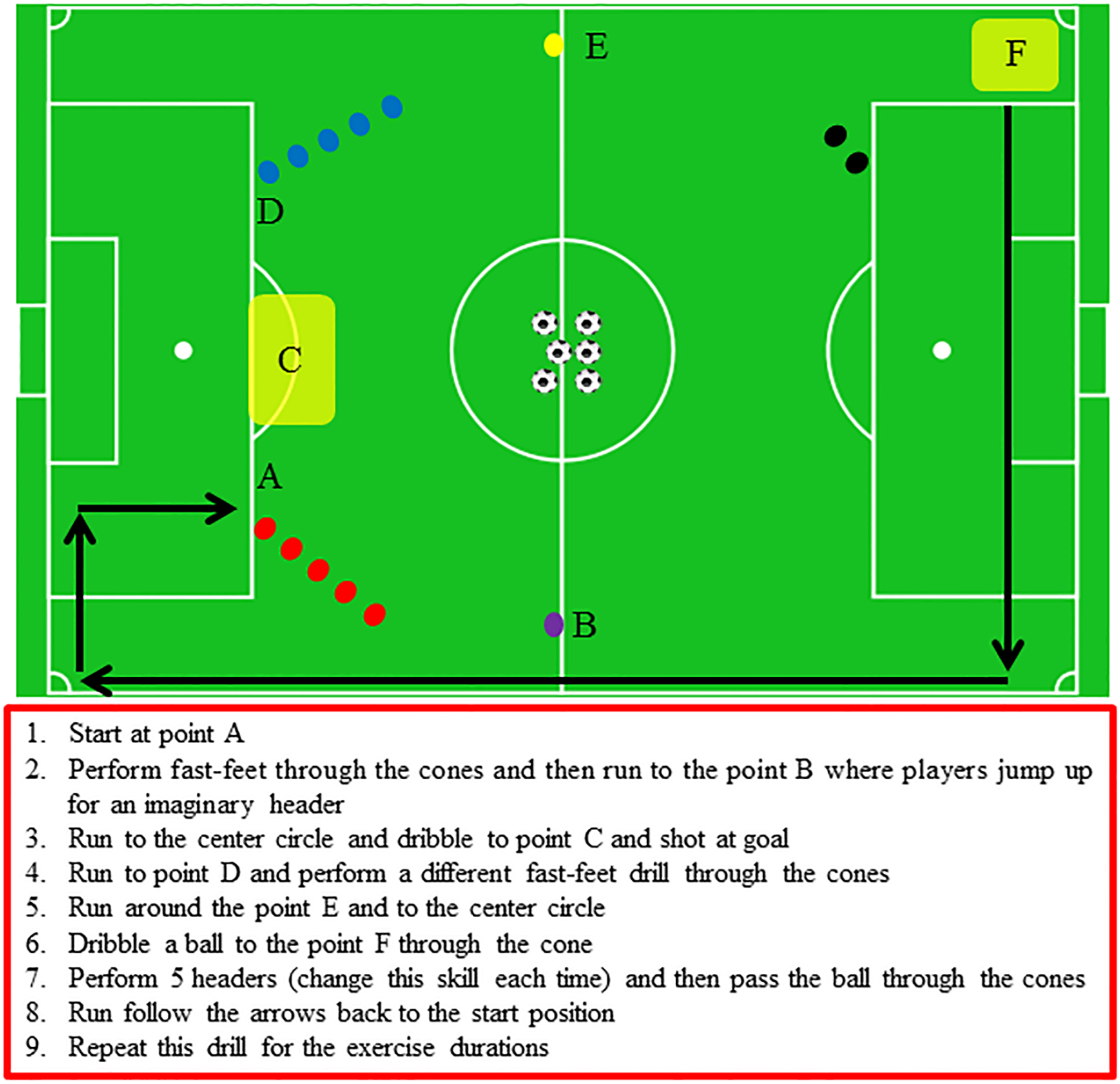


Fig 1. Diagram of high-intensity training.

During the detraining experimental period, high-intensity aerobic training was performed three times per week for two weeks in the HAT group. After approximately a 20-min warm-up period, the players performed a soccer drill (Fig 1) on an artificial grass surface and three repetitions of 12 min of exercise at 80–90% of the maximum heart rate (HRmax) measured during Yo-Yo IR2 test. These repetitions were interspersed by 3 min active recovery. The players controlled exercise intensity by watching their HR monitor that recorded at 5 s intervals (Polar Team System, Polar, Electro Oy, Kempele, Finland). These data were downloaded to a personal laptop for further analysis. The mean HR during the 12 min exercise sessions was 87.3±1.5% of HRmax. The DHAT group did not perform any exercise sessions during the two weeks of detraining and conducted normal daily activities.

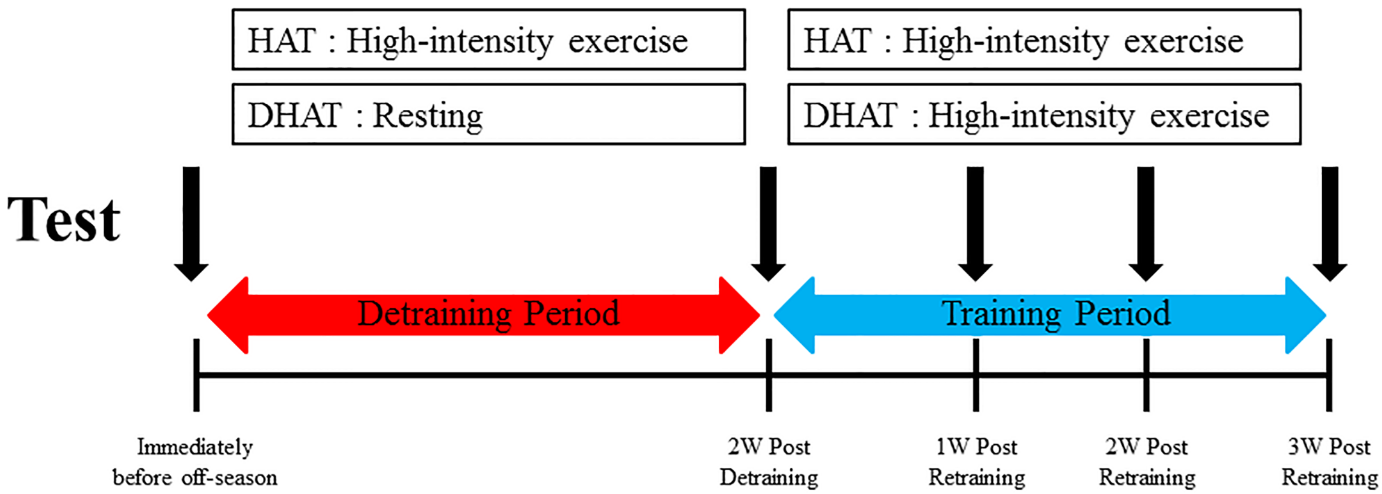


Fig 2. A schematic illustration of the experimental design.

During the retraining period, the HAT and DHAT groups performed high-intensity aerobic training (12 min × 3) four times per week for three weeks. The mean HR during the 12 min exercise sessions was 86.5±1.4% of HRmax in the HAT group. The DHAT group completed moderate intensity aerobic training (HRmax 70–80%; 76.5±3.2%) for two days before completing the high-intensity training (HRmax 80–90%; 87.7±1.3%).

Experimental protocol

A schematic illustration of the experimental design is shown in Fig 2. The subjects completed the 30 m sprint test, Yo-Yo intermittent recovery level 2 (Yo-Yo IR2) test, arrowhead agility test, repeated sprint test, and isokinetic strength test. The tests were conducted for two days. The participants refrained from alcohol and caffeine in the 24 h prior to the test. The participants arrived at the laboratory having completed the appropriate diet regime to monitor the diet level. The participants were instructed to ingest water 5 mL of water for every kilogram of their body mass 2 h before arriving at the laboratory. Upon the arrival at the laboratory, body composition (Inbody 520, Biospace, Seoul, Korea) and height (BSM, Seoul, Korea) were measured. Following the completion of the baseline assessments, the participants commenced the tests on an artificial grass surface. A 30-m sprint test, arrowhead agility test, and repeated sprints test were performed in the morning. The Yo-Yo IR2 test was conducted in the evening with 5 hours of recovery after lunch. Isokinetic strength tests were performed in the laboratory the next day. Body composition and exercise tests were completed immediately before the end of the season; after two weeks of detraining; and at one, two, and three weeks of retraining intervention.

30m sprint test

The sprint tests which consisted of 2 maximal sprints of 30 m with 2-minute rest between each sprint were conducted. The sprint times at 5, 10, 20 and 30 m were recorded using the photocell gates (Microgate, Bolzano, Itaia). The participants started to run 50 cm before the photocell gate recordings. The fastest times at the distances were recorded for data analysis.

Repeated sprint test

The repeated sprint test consisted of seven maximal 34.2 m sprints, interspersed by 25 s of active recovery (40 m jogging distance) [18]. Recovery was timed so that the subjects returned to the start line between the 23rd and 24th second. Additionally, verbal feedback was given at 5, 10, 15 and 20 s of the recovery. Performance was measured as the total sprint time in seconds.

Yo-Yo intermittent recovery test (level 2)

The Yo-Yo IR2 test was performed on an artificial turf. The Yo-Yo IR2 test consists of 2 × 20 m shuttle runs at increasing speeds, controlled by audio signals from a compact disk. Between each bout of running, the subjects completed 10 s of active recovery, consisting of 2 × 5 m jogging [19]. The test was terminated when the subjects failed twice to reach the start line on time and the distance (meters) covered at the end point was recorded [5].

Arrowhead agility test

The arrowhead agility tests consisted of 4 sprints (2 right, 2 left), with 2-minutes rest between each sprint [20]. Each subject started 50 cm behind the start line and sprinted 10 m forward to a cone. From the cone, the subjects turned at a right angle to a cone being apart from 5m before turning to a cone 15 m straight from the start line. They turned again from the cone to accelerate in a straight line for 15 m over the initial start line to complete the run. The fastest times were recorded for data analysis. Timing gates were used to accurately assess the time to completion.

Isokinetic strength

The subjects performed the Isokinetic dynamometry (Cybex MET-300, New York, USA) to evaluate the unilateral strength of the concentric contraction of the flexors and extensors of the knee [21]. The angular speed parameters of $60^\circ \times s^{-1}$, $180^\circ \times s^{-1}$, and $240^\circ \times s^{-1}$ were used for the measurements. The results of the measurements were expressed in absolute peak torque (Nm) for the purposes of the off-seasonal variation comparisons.

Statistical analysis

All data are presented as means ± SD. Two-way analysis of variance (ANOVA) with repeated measure was conducted to determine any treatment differences between the HAT and DHAT conditions. The assumption of sphericity (homogeneity of covariance) was assessed and corrected for using the Huynh-Feldt epsilon. Because there were only 2 levels in the main effect of condition, follow-up multiple comparisons were not necessary. A significant effect of time was followed up with planned multiple contrasts in line with the a priori hypotheses. Therefore, data at the specific time points were compared with the baseline (first) time point using Newman-Keuls multiple contrasts. Where a significant interaction between condition and time was observed, differences between conditions were examined at each time point using Newman-Keuls multiple contrasts. Baseline values were compared using an independent samples *t* test. The alpha level for evaluation of statistical significance was set at $P < 0.05$. Effect sizes were assessed by partial eta squared (η_p^2), which were defined as trivial (<0.1), small (0.1–0.3), moderate (0.3–0.5) and large (>0.5) [22].

Results

Body weight and body fat were similar between the HAT and DHAT groups throughout the experimental period ($P > 0.05$; Tables 1 and 2). There was no significant effect of condition

Table 1. Body composition of the subjects before, after two weeks of detraining and at one, two and three weeks of retraining (mean ± SD).

		Pre	2W DT	1W RT	2W RT	3W RT
Body weight (kg)	HAT	68.1±7.1	68.5±7.1	68.6±7.2	68.6±7.3	68.4±7.3
	DHAT	67.5±7.3	67.8±7.3	67.9±7.3	68.2±7.2	67.9±7.3
Body mass index (kg/m^2)	HAT	22.7±0.4	22.7±0.5	22.7±0.5	22.5±0.6	22.9±0.4
	DHAT	22.1±0.8	22.4±0.9	22.4±0.6	22.5±0.9	22.3±0.8
Skeletal muscle mass (kg)	HAT	32.6±3.5	32.5±3.3	32.3±3.2	32.6±3.6	32.4±3.3
	DHAT	33.0±3.5	33.2±2.3	33.1±3.2	33.4±3.3	33.3±3.3
Percent body fat (%)	HAT	9.6±0.7	9.7±0.7	9.8±1.3	9.5±0.9	9.9±1.3
	DHAT	9.3±1.2	9.8±1.3	9.8±1.1	9.7±1.2	9.5±1.4

Values are means ± standard deviation

nor was there an interaction of condition and time ($P > 0.05$) in the performance of players on sprint and agility tests (Tables 3 and 4). However, a significant effect of time was observed for sprint test at 5 m, 10 m, and 30 m as well as in the left direction of arrowhead agility ($P < 0.05$). Isokinetic strength at all angular speeds remained similar to baseline under both conditions throughout the experimental period, with no significant effects of time, condition, or an interaction between the two ($P > 0.05$; Tables 5 and 6). There was a significant interaction in the Yo-Yo IR2 test ($F = 3.273$; $P < 0.05$; $\eta_P^2 = 0.267$), while the measurement time ($F = 1.517$; $P > 0.05$; $\eta_P^2 = 0.144$) and condition were not significant ($F = 1.938$; $P > 0.05$; $\eta_P^2 = 0.177$). Compared to the pre-detraining performance, the Yo-Yo IR2 test performance decreased significantly after the two-week detraining period ($P < 0.01$) and the values reach before detraining level after two weeks of retraining in the DHAT group ($P > 0.05$). No differences were detected at three weeks post-retraining between conditions ($P > 0.05$), whilst values in the HAT group remained stable throughout the experimental period ($P > 0.05$; Fig 3). A main effect of time was found ($F = 3.539$; $P < 0.05$; $\eta_P^2 = 0.282$), along with a significant interaction between condition and time for repeated sprint performance ($F = 3.127$; $P < 0.05$; $\eta_P^2 = 0.258$). No changes in repeated sprint performance were observed at any time point under HAT conditions ($P > 0.05$), whereas repeated sprint performance declined after two weeks of detraining ($P < 0.05$) and remained lower than at baseline by two weeks post-

Table 2. Differences in the body composition of the subjects between conditions in each test (n = 20).

		F	*P*	η_P^2
Body weight	Condition	0.048	0.831	0.005
	Time	12.372	0.001	0.579
	Condition x Time	0.628	0.646	0.065
Body mass index	Condition	2.524	0.147	0.219
	Time	0.716	0.587	0.074
	Condition x Time	1.776	0.155	0.165
Skeletal muscle mass	Condition	0.228	0.644	0.025
	Time	0.178	0.948	0.019
	Condition x Time	0.117	0.976	0.013
Percent body fat	Condition	0.046	0.834	0.005
	Time	2.201	0.088	0.197
	Condition x Time	0.653	0.629	0.068

F; testing criteria level, *P*; level of statistical significance, η_P^2; partial eta squared

Table 3. Sprint and agility before, after two weeks of detraining and at one, two and three weeks of retraining (mean ± SD).

		Pre	2W DT	1W RT	2W RT	3W RT
5 m	HAT	1.04±0.04	1.05±0.03	1.06±0.04	1.02±0.05	1.02±0.05
	DHAT	1.05±0.04	1.04±0.03	1.05±0.03	1.01±0.04	1.01±0.04
10 m	HAT	1.75±0.06	1.74±0.10	1.73±0.05	1.71±0.04	1.72±0.06
	DHAT	1.78±0.05	1.73±0.05	1.73±0.05	1.71±0.06	1.72±0.07
20 m	HAT	3.00±0.09	3.01±0.13	3.02±0.08	2.99±0.06	2.99±0.09
	DHAT	3.05±0.05	3.07±0.08	3.03±0.06	2.99±0.09	2.99±0.09
30 m	HAT	4.13±0.11	4.22±0.17	4.25±0.12	4.21±0.11	4.23±0.13
	DHAT	4.23±0.07	4.30±0.12	4.29±0.09	4.23±0.12	4.25±0.10
Agility (R)	HAT	8.04±0.19	8.09±0.22	8.13±0.17	7.99±0.21	8.03±0.22
	DHAT	8.06±0.16	8.09±0.22	8.09±0.19	8.04±0.25	8.05±0.25
Agility (L)	HAT	7.99±0.17	8.12±0.20	8.10±0.20	7.98±0.23	8.00±0.18
	DHAT	8.04±0.18	8.14±0.24	8.14±0.18	8.08±0.15	8.08±0.20

Values are means ± standard deviation. R; right, L; left

retraining under DHAT conditions ($P < 0.05$). It reached baseline level at the end of the experimental period ($P > 0.05$; Fig 4).

Discussion

The major findings in the present study were that two weeks of detraining after competitive season decreased performance in the Yo-Yo IR2 test and repeated sprints. The detraining-induced reductions in the Yo-Yo IR2 test performance improved compared to baseline levels after two weeks of high-intensity aerobic training. Meanwhile, three weeks were required to

Table 4. Differences in sprint and agility between conditions in each test (n = 20).

		F	*P*	η_P^2
5 m	Condition	0.095	0.765	0.010
	Time	7.657	0.001	0.460
	Condition x Time	1.586	0.199	0.150
10 m	Condition	0.305	0.594	0.033
	Time	4.672	0.004	0.342
	Condition x Time	1.010	0.415	0.101
20 m	Condition	0.480	0.506	0.051
	Time	2.500	0.060	0.217
	Condition x Time	1.167	0.342	0.115
30 m	Condition	0.879	0.373	0.089
	Time	5.357	0.002	0.373
	Condition x Time	1.619	0.191	0.152
Agility (R)	Condition	0.013	0.912	0.001
	Time	2.516	0.058	0.218
	Condition x Time	0.357	0.838	0.038
Agility (L)	Condition	0.499	0.498	0.053
	Time	3.542	0.015	0.282
	Condition x Time	0.382	0.820	0.041

R; right, L; left, F; testing criteria level, *P*; level of statistical significance, η_P^2; partial eta squared

Table 5. Peak torques (Nm) during concentric knee flexion and extension before, after two weeks of detraining and at one, two and three weeks of retraining (mean ± SD).

		Pre	2W DT	1W RT	2W RT	3W RT
DL-PT-E-60	HAT	208.2±12.3	208.2±12.3	209.6±27.3	210.0±25.2	205.2±19.8
	DHAT	211.5±13.9	212.7±15.9	214.2±15.9	216.7±17.5	208.3±12.4
DL-PT-F-60	HAT	135.8±32.3	135.8±30.6	135.8±37.7	139.2±29.5	135.2±28.2
	DHAT	121.2±21.6	137.2±26.9	140.3±29.1	137.2±27.7	136.5±19.1
NL-PT-E-60	HAT	189.9±33.1	194.6±39.0	197.0±35.2	200.9±38.1	191.0±31.0
	DHAT	198.1±28.0	189.1±24.2	183.9±26.8	193.0±24.6	187.0±16.4
NL-PT-F-60	HAT	129.1±30.5	127.9±26.4	125.1±34.8	128.1±26.1	127.1±31.5
	DHAT	132.6±23.1	127.1±20.2	125.7±23.1	131.9±24.6	135.3±26.2
DL-PT-E-180	HAT	138.8±18.9	146.6±21.4	146.6±23.9	145.3±23.3	140.4±18.8
	DHAT	145.2±23.6	146.1±15.1	152.0±20.1	150.3±15.4	152.6±18.8
DL-PT-F-180	HAT	105.4±17.3	105.2±14.8	108.5±13.8	107.3±10.9	107.7±13.4
	DHAT	108.5±9.8	107.2±14.2	110.6±15.7	110.5±17.3	106.1±15.3
NL-PT-E-180	HAT	136.9±18.9	137.0±25.4	139.5±21.1	141.1±18.0	137.7±19.5
	DHAT	138.1±16.5	139.6±16.6	137.9±21.9	142.7±21.4	138.5±20.8
NL-PT-F-180	HAT	97.4±18.7	93.7±18.4	99.7±24.4	96.4±18.9	95.7±21.6
	DHAT	102.6±16.7	100.5±15.3	100.7±19.2	101.7±17.5	102.6±19.9
DL-PT-E-240	HAT	114.9±19.0	115.9±15.9	116.9±17.6	115.2±15.6	113.4±17.1
	DHAT	116.2±16.2	118.6±12.4	117.5±13.6	119.5±12.7	115.8±11.7
DL-PT-F-240	HAT	84.6±15.1	83.8±16.8	84.8±14.1	87.3±12.2	86.2±13.8
	DHAT	89.1±9.3	87.8±13.4	92.6±17.4	92.9±16.0	89.3±12.7
NL-PT-E-240	HAT	113.3±16.7	112.7±15.4	112.6±13.7	114.4±11.2	110.3±14.6
	DHAT	115.2±10.8	110.3±13.1	116.1±14.6	114.2±16.8	115.4±13.2
NL-PT-F-240	HAT	83.3±15.2	83.4±18.6	81.0±19.2	85.1±16.2	86.3±12.8
	DHAT	88.4±14.6	81.7±14.5	86.3±18.4	89.2±18.6	87.2±18.6

Values are means ± standard deviation. DL; dominant leg, NL; non-dominant leg, PT; peak torque, E; extensors, F; flexors, 60, 180, 240; angular velocities ($^{\circ}\cdot s^{-1}$)

return to the initial level of repeated sprint performance with retraining using high-intensity training. Ultimately, a reduced amount of high-intensity training after the competitive season facilitated the maintenance of physical fitness.

The HAT group that continued to perform high-intensity aerobic exercise after the competitive season maintained their performance level in the Yo-Yo IR2 test over the five week treatment period. These results are supported by previous studies, which indicate that, after the last match of the season, 10 training sessions, consisting of high-intensity training for two weeks, do not change performance in the Yo-Yo IR2 test in elite soccer players [15]. However, Nakamura et al. [23] observed that running and plyometric training for two days per week for three weeks after the completion of a competitive season did not prevent the decrease in performance in The Yo-Yo IR2 test in elite soccer players. The reason of these differences in results is unclear but it probably related to exercise intensity. Indeed, there was no significant decrease in performance in the Yo-Yo IR2 test during off-season in the present study and Christensen et al. [15]'s study applying high-intensity exercise despite reduced exercise time compared to that in-season. Furthermore, the exercise intensity was higher than that in the previous study conducted by Nakamura et al. [23], which modulated endurance training (70–80% of HRmax).

In the present study, we found that two weeks of detraining after the competitive season markedly decreased performance in the Yo-Yo IR2 test in elite soccer players. Accordingly,

Table 6. Differences in sprint and agility peak torques (Nm) during concentric knee flexion and extension between conditions in each test (n = 20).

		F	*P*	η_P^2
DL-PT-E-60	Condition	1.674	0.228	0.157
	Time	1.665	0.180	0.156
	Condition x Time	0.100	0.982	0.011
DL-PT-F-60	Condition	0.032	0.862	0.004
	Time	1.249	0.308	0.122
	Condition x Time	1.156	0.346	0.114
NL-PT-E-60	Condition	0.113	0.744	0.012
	Time	0.721	0.583	0.074
	Condition x Time	1.050	0.395	0.104
NL-PT-F-60	Condition	0.059	0.814	0.007
	Time	1.287	0.293	0.125
	Condition x Time	0.469	0.758	0.050
DL-PT-E-180	Condition	0.355	0.566	0.038
	Time	1.283	0.295	0.125
	Condition x Time	0.695	0.600	0.072
DL-PT-F-180	Condition	0.143	0.714	0.016
	Time	0.419	0.794	0.045
	Condition x Time	0.268	0.896	0.029
NL-PT-E-180	Condition	0.007	0.935	0.001
	Time	0.485	0.747	0.051
	Condition x Time	0.177	0.949	0.019
NL-PT-F-180	Condition	0.373	0.556	0.040
	Time	0.481	0.749	0.051
	Condition x Time	0.520	0.721	0.055
DL-PT-E-240	Condition	0.097	0.762	0.011
	Time	0.549	0.701	0.057
	Condition x Time	0.153	0.960	0.017
DL-PT-F-240	Condition	0.872	0.375	0.088
	Time	0.834	0.512	0.085
	Condition x Time	0.297	0.878	0.032
NL-PT-E-240	Condition	0.047	0.833	0.005
	Time	0.651	0.630	0.067
	Condition x Time	0.971	0.435	0.097
NL-PT-F-240	Condition	0.147	0.711	0.016
	Time	1.786	0.153	0.166
	Condition x Time	0.696	0.600	0.072

DL; dominant leg, NL; non-dominant leg, PT; peak torque, E; extensors, F; flexors, 60, 180, 240; angular velocities ($^\circ \cdot s^{-1}$), F; testing criteria level, *P*; level of statistical significance, η_P^2; partial eta squared

Thomassen et al. [5] and Christensen et al. [15] observed that the Yo-Yo IR2 test performance after detraining for two weeks decreased from 845 m to 654 m in elite soccer players. In addition, a study from another laboratory reported that a prolonged detraining period can induce an 8% decline in maximal oxygen consumption [24], which is strongly associated with distance on the Yo-Yo IR2 test [25]. The degree of deterioration of physical fitness over the course of the detraining period after the competitive season is closely related to the fitness level of

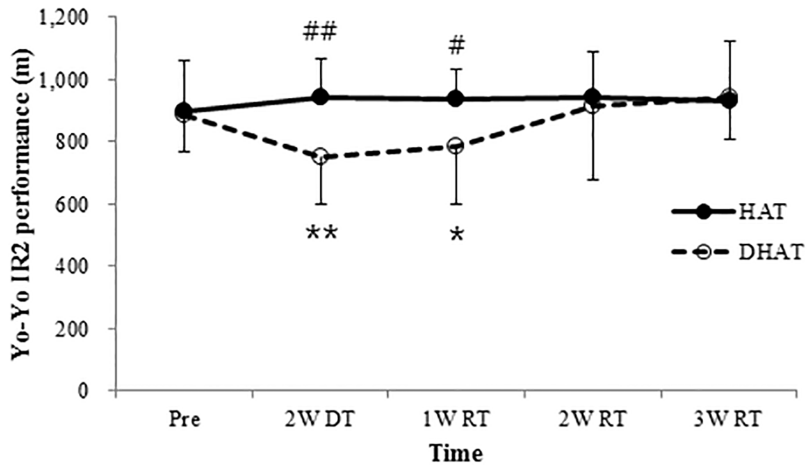


Fig 3. Yo-Yo IR2 performan for the high-intensity training (HAT; n = 10) and detraining + retraining (DHAT; n = 10) before, after two weeks detraining and at one, two and three weeks of retraining (n = 11, mean ± SD). **$P < 0.01$; significantly different from pre. *$P < 0.05$; significantly different from pre. ##$P < 0.01$; significantly between conditions. #$P < 0.05$; significantly between conditions.

athletes [23]. Therefore, these results can support the notion that performance in the Yo-Yo IR2 test can be reduced despite only a few days of detraining in elite soccer players with a high level of physical fitness. These decreases in performance in the Yo-Yo IR2 test can be explained at the muscle level. Several weeks of detraining lead to a return in muscle capillarization to

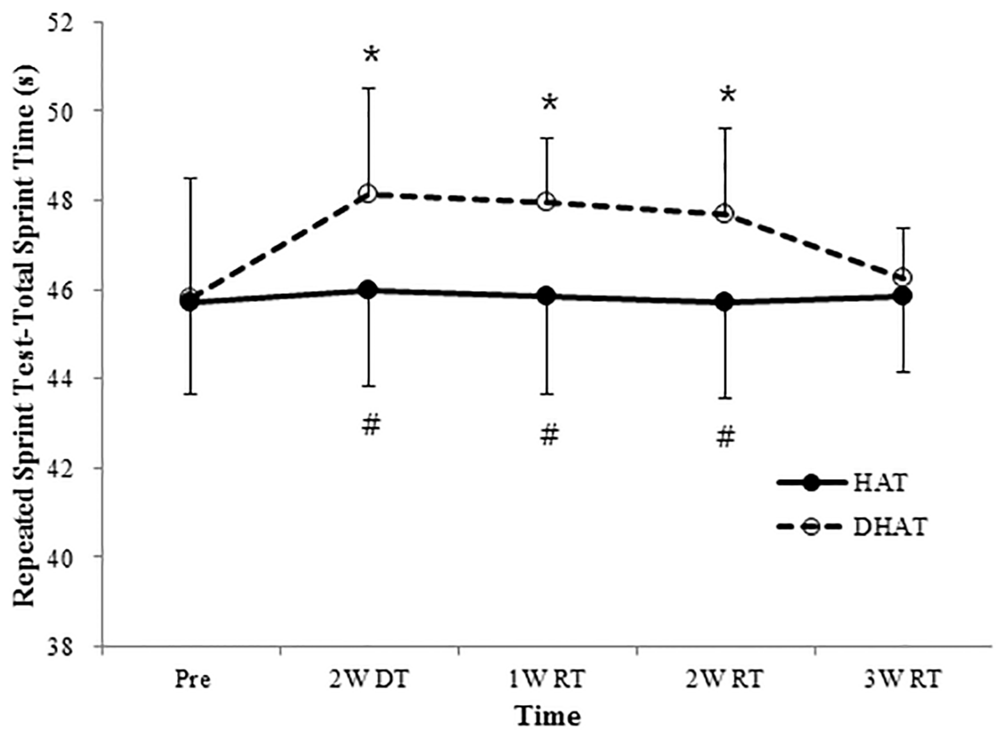


Fig 4. Repeated sprint test for the high-intensity training (HAT; n = 10) and detraining + retraining (DHAT; n = 10) before, after two weeks detraining and at one, two and three weeks of retraining (n = 11, mean ± SD). *$P < 0.05$; significantly different from pre. #$P < 0.05$; significantly between conditions.

baseline before detraining in athletes and a 25%-45% decline in oxidative enzyme activities, which result in reduced mitochondrial ATP production in skeletal muscle [9].

Several previous studies have reported that high intensity training improves the performance in the Yo-Yo IR2 test of elite soccer players [25, 26]. In line with these results, high-intensity aerobic training after two weeks of detraining was found to improve performance in the Yo-Yo IR2 test in the present study. Two weeks of retraining with high-intensity exercise is required to return close to the baseline level of performance. This result is inconsistent with a previous study that suggested that athletes with a high fitness level must perform exercise training for a period that is at least twice as long as the resting time period in order to improve their physical fitness to a level of before detraining [24]. The discrepancy in time periods required to return to the physical fitness level at baseline can be due to variations in the length of the detraining period (four weeks versus eight weeks) and the fitness level of the athletes (compared to end of season versus before the Olympic game). Indeed, the performance in the Yo-Yo IR2 test decreased by 11% at the end of the season compared to the start of the season and a 42% increase was observed during the eight weeks of pre-season training [25]. This phenomenon is likely due to accumulated fatigue experienced during the competitive season. This assumption is supported by the finding from the present study that the performance in the Yo-Yo IR2 test was higher at three weeks of post retraining compared to baseline. Furthermore, Noon et al. [20] and Oliver et al. [27] observed that perceptual well-being (e.g., motivation, sleep quality, recovery, appetite, fatigue, stress, muscle soreness, stiffness) deteriorated with an increase in training exposure and accumulated fatigue as the season progressed in elite athletes.

Repeated sprint performance did not change over five weeks of high-intensity training after competitive season in the present study. In contrast to the present study, previous studies reported that two weeks of high-intensity training immediately after the end of season enhanced repeated sprint performance in elite soccer players [5, 15]. These different results may be associated with the high-intensity training method used during the retraining period. Aguiar et al. [28] observed that intermittent training for 12 weeks consisting of 20 minutes per training session resulted in greater improvements in repeated sprint performance than did continuous training. Indeed, the training sessions in the present study largely comprised of high-intensity endurance exercise, whereas the training sessions used in previous studies consisted of five high-intensity aerobic training, including small-sided (4 vs. 4 and 3 vs. 3) soccer drills (8 × 2 min) and five speed endurance training (10–12 × 25–30 s sprints) over the course of two weeks. In other respects, since well-trained athletes are more sensitive to changes in physical fitness with inadequate training intensity and do not easily experience improvements following further training due to the ceiling effect [25], the capacity of repeated sprint performance of the players in the present study might be optimal by the end of the competitive season. This is supported by the observation that repeated sprint performance in players from the present study was similar to previous study conducted with professional soccer players during the competitive season [18].

It is well known that anaerobic exercise performance decreases in highly trained elite soccer players, despite a short period of detraining after the competitive season [9]. There was also a significant decrease in repeated sprint performance over two weeks of detraining after the end of a match in the present study. The detraining-induced decrease in performance gradually increased during the three weeks retraining period. The aerobic high-intensity training-induced increase in repeated sprint performance in the present study is likely to be the result of training-induced biochemical adaptation in skeletal muscles. Thomassen et al. [5] and Christensesn et al. [15] observed that two weeks of high-intensity exercise immediately after the last match of the season enhanced Na^+-K^+ pump α_2-isoform expression by 15%, increased the FXYD1ser68-to-FXYD1 ratio by 27%, increased the level of pyruvate dehydrogenase by 17%, and improved repeated sprint performance. In comparison, cessation of training for two weeks did not affect the expression of Na^+-K^+ pump isoform expression and resulted in a reduction of the AB_FXYD1ser68 signal by 18%; decreased pyruvate dehydrogenase level by 17%; a drop in citrate synthase and 3-hydroxyacyl-CoA activity to 12% and 18% of maximal, respectively; and a reduction in performance. However, repeated sprint performance at 3 weeks post-retraining was still lower than the performance recorded at baseline. As mentioned, aerobic high-intensity training is not optimal for improving repeated sprint performance, which represents the capacity for anaerobic exercise performance. On the contrary, improvements in repeated sprint performance through aerobic high-intensity training might be associated with the training period during the preseason. Recently, Teixeira et al. [29] reported that high-intensity aerobic training involving shuttle-run intervals (4 × 4 min) for five weeks during the preseason enhanced repeated sprint ability with increased aerobic performance in elite athletes. When considered, these findings suggest that more than three weeks of high-intensity aerobic training is required to develop repeated sprint performance during preseason in elite players.

The observed lack of changes in body composition and sprint performances (10 m, 20 m, 30 m) for five weeks during the study period in both groups disagrees with previous studies that engaged in more than two weeks of detraining [13, 30]. For example, Koundourakis et al. [13] examined the effect of detraining on exercise performance and body composition in professional soccer players. They observed that prolonged detraining period (six weeks) significantly increased body weight and body fat percentage and reduced maximal oxygen consumption and performances in squat-jump, countermovement-jump, and sprints (10 m, 20 m). These results suggest that a short period of detraining (approximately two weeks) may not lead to changes in body composition and explosive exercise performance in well-trained soccer players. This is supported by findings that there were changes in neither isokinetic strength at any angular speeds in the present study nor squat, vertical jump, or isometric and isokinetic knee force following two weeks detraining in high fitness athletes [31]. A possible explanation for the absence of changes in explosive exercise performance after a short period of detraining is the lack of changes in muscle fiber characteristics. Mujika and Padilla. [9] reported that two weeks of detraining did not alter muscle fiber distribution in well-trained athletes. However, three weeks of detraining after the first half of a competitive season in elite soccer players resulted in changes in skeletal muscle morphology, including a reduction in mean fast twitch (FT) fiber cross-sectional area and reduction in mitochondrial enzyme activities and exercise performance [32]. Taken together, these data suggest that more than two weeks of detraining in elite soccer players could have resulted in a decrease in explosive exercise performance by reduced ATP production in skeletal muscle.

Conclusions

In conclusion, the findings demonstrate that two weeks of detraining after the competitive season resulted in a marked decrease in performance in the Yo-Yo IR2 test and repeated sprints. To return to a previous level of physical fitness with retraining through high-intensity aerobic training after a period of detraining required a similar period of retraining for performance in Yo-Yo IR2 and/or more periods for repeated sprint performance. The off-season rest period did not result in changes in explosive exercise performances and body composition. Aerobic high-intensity training with reduced training volume after a competitive season can prevent reductions in performances in the Yo-Yo IR2 test and repeated sprints. On the contrary, the decrease in aerobic and anaerobic performance induced by two weeks of detraining was recovered within a few weeks of adequate training during the preseason. Therefore, these findings suggest that elite soccer players can be allowed to take short periods of rest (~2 weeks) without training during the off-season for the release of mental and physical stress that is accumulated throughout the competitive season.

Supporting information

S1 File. Raw data of Figs 3 and 4 and Tables 1, 2, 3, 4, 5 and 6.
(XLSX)

Author Contributions

Writing – original draft: Chang Hwa Joo.

References

1. Carling C, Bradley P, McCall A, Dupont G. Match-to-match variability in high-speed running activity in a professional soccer team. J Sports Sci. 2016; 34(24):2215–23. Epub 2016/05/05. https://doi.org/10.1080/02640414.2016.1176228 PMID: 27144879.
2. Di Salvo V, Gregson W, Atkinson G, Tordoff P, Drust B. Analysis of high intensity activity in Premier League soccer. Int J Sports Med. 2009; 30(3):205–12. Epub 2009/02/14. https://doi.org/10.1055/s-0028-1105950 PMID: 19214939.
3. Mohr M, Krustrup P, Bangsbo J. Match performance of high-standard soccer players with special reference to development of fatigue. J Sports Sci. 2003; 21(7):519–28. Epub 2003/07/10. https://doi.org/10.1080/0264041031000071182 PMID: 12848386.
4. Kotzamanidis C, Chatzopoulos D, Michailidis C, Papaiakovou G, Patikas D. The effect of a combined high-intensity strength and speed training program on the running and jumping ability of soccer players. J Strength Cond Res. 2005; 19(2):369–75. Epub 2005/05/21.
5. Thomassen M, Christensen PM, Gunnarsson TP, Nybo L, Bangsbo J. Effect of 2-wk intensified training and inactivity on muscle Na+-K+ pump expression, phospholemman (FXYD1) phosphorylation, and performance in soccer players. J Appl Physiol (1985). 2010; 108(4):898–905. Epub 2010/02/06. https://doi.org/10.1152/japplphysiol.01015.2009 PMID: 20133439.
6. Mara JK, Thompson KG, Pumpa KL, Ball NB. Periodization and physical performance in elite female soccer players. Int J Sports Physiol Perform. 2015; 10(5):664–9. Epub 2015/01/23. https://doi.org/10.1123/ijspp.2014-0345 PMID: 25611789.
7. Jeong TS, Reilly T, Morton J, Bae SW, Drust B. Quantification of the physiological loading of one week of "pre-season" and one week of "in-season" training in professional soccer players. J Sports Sci. 2011; 29(11):1161–6. Epub 2011/07/23. https://doi.org/10.1080/02640414.2011.583671 PMID: 21777053.
8. Buchheit M, Racinais S, Bilsborough JC, Bourdon PC, Voss SC, Hocking J, et al. Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. J Sci Med Sport. 2013; 16(6):550–5. Epub 2013/01/22. https://doi.org/10.1016/j.jsams.2012.12.003 PMID: 23332540.
9. Mujika I, Padilla S. Detraining: loss of training-induced physiological and performance adaptations. Part I: short term insufficient training stimulus. Sports Med. 2000; 30(2):79–87. Epub 2000/08/31. PMID: 10966148.
10. Moore RL, Thacker EM, Kelley GA, Musch TI, Sinoway LI, Foster VL, et al. Effect of training/detraining on submaximal exercise responses in humans. J Appl Physiol (1985). 1987; 63(5):1719–24. Epub 1987/11/01. https://doi.org/10.1152/jappl.1987.63.5.1719 PMID: 3693207.
11. Izquierdo M, Ibanez J, Gonzalez-Badillo JJ, Ratamess NA, Kraemer WJ, Hakkinen K, et al. Detraining and tapering effects on hormonal responses and strength performance. J Strength Cond Res. 2007; 21(3):768–75. Epub 2007/08/10.
12. Mujika I, Padilla S. Detraining: loss of training-induced physiological and performance adaptations. Part II: Long term insufficient training stimulus. Sports Med. 2000; 30(3):145–54. Epub 2000/09/22. PMID: 10999420.
13. Koundourakis NE, Androulakis NE, Malliaraki N, Tsatsanis C, Venihaki M, Margioris AN. Discrepancy between exercise performance, body composition, and sex steroid response after a six-week detraining period in professional soccer players. PLoS One. 2014; 9(2):e87803. Epub 2014/03/04. https://doi.org/10.1371/journal.pone.0087803 PMID: 24586293.

14. Buchheit M, Morgan W, Wallace J, Bode M, Poulos N. Physiological, psychometric, and performance effects of the Christmas break in Australian football. Int J Sports Physiol Perform. 2015; 10(1):120–3. Epub 2014/05/09. https://doi.org/10.1123/ijspp.2014-0082 PMID: 24806508.

15. Christensen PM, Krustrup P, Gunnarsson TP, Kiilerich K, Nybo L, Bangsbo J. VO2 kinetics and performance in soccer players after intense training and inactivity. Med Sci Sports Exerc. 2011; 43(9):1716–24. Epub 2011/02/12. https://doi.org/10.1249/MSS.0b013e318211c01a PMID: 21311360.

16. Bartlett JD, Close GL, MacLaren DP, Gregson W, Drust B, Morton JP. High-intensity interval running is perceived to be more enjoyable than moderate-intensity continuous exercise: implications for exercise adherence. J Sports Sci. 2011; 29(6):547–53. Epub 2011/03/02. https://doi.org/10.1080/02640414.2010.545427 PMID: 21360405.

17. Impellizzeri FM, Marcora SM, Castagna C, Reilly T, Sassi A, Iaia FM, et al. Physiological and performance effects of generic versus specific aerobic training in soccer players. Int J Sports Med. 2006; 27(6):483–92. Epub 2006/06/13. https://doi.org/10.1055/s-2005-865839 PMID: 16767613.

18. Abrantes C, Macas V, Sampaio J. Variation in football players' sprint test performance across different ages and levels of competition. J Sports Sci Med. 2004; 3(YISI 1):44–9. Epub 2004/11/01. PMID: 24778553.

19. Bangsbo J, Iaia FM, Krustrup P. The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. Sports Med. 2008; 38(1):37–51. Epub 2007/12/18. PMID: 18081366.

20. Noon MR, James RS, Clarke ND, Akubat I, Thake CD. Perceptions of well-being and physical performance in English elite youth footballers across a season. J Sports Sci. 2015; 33(20):2106–15. Epub 2015/09/19. https://doi.org/10.1080/02640414.2015.1081393 PMID: 26383605.

21. Kilinc BE, Kara A, Camur S, Oc Y, Celik H. Isokinetic dynamometer evaluation of the effects of early thigh diameter difference on thigh muscle strength in patients undergoing anterior cruciate ligament reconstruction with hamstring tendon graft. J Exerc Rehabil. 2015; 11(2):95–100. Epub 2015/05/12. https://doi.org/10.12965/jer.150100 PMID: 25960982.

22. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 2009; 41(1):3–13. Epub 2008/12/19. https://doi.org/10.1249/MSS.0b013e31818cb278 PMID: 19092709.

23. Nakamura D, Suzuki T, Yasumatsu M, Akimoto T. Moderate running and plyometric training during off-season did not show a significant difference on soccer-related high-intensity performances compared with no-training controls. J Strength Cond Res. 2012; 26(12):3392–7. Epub 2011/12/31. https://doi.org/10.1519/JSC.0b013e3182474356 PMID: 22207263.

24. Godfrey RJ, Ingham SA, Pedlar CR, Whyte GP. The detraining and retraining of an elite rower: a case study. J Sci Med Sport. 2005; 8(3):314–20. Epub 2005/10/27. PMID: 16248472.

25. Krustrup P, Mohr M, Nybo L, Jensen JM, Nielsen JJ, Bangsbo J. The Yo-Yo IR2 test: physiological response, reliability, and application to elite soccer. Med Sci Sports Exerc. 2006; 38(9):1666–73. Epub 2006/09/09. https://doi.org/10.1249/01.mss.0000227538.20799.08 PMID: 16960529.

26. Mohr M, Krustrup P. Comparison between two types of anaerobic speed endurance training in competitive soccer players. J Hum Kinet. 2016; 51:183–92. Epub 2017/02/06. https://doi.org/10.1515/hukin-2015-0181 PMID: 28149381.

27. Oliver JL, Lloyd RS, Whitney A. Monitoring of in-season neuromuscular and perceptual fatigue in youth rugby players. Eur J Sport Sci. 2015; 15(6):514–22. Epub 2015/09/15. https://doi.org/10.1080/17461391.2015.1063700 PMID: 26366619.

28. Aguiar M, Abrantes C, Maçãs V, Leite N, Sampaio J, Ibáñez S. Effects of intermittent or continuous training on speed, jump and repeated-sprint ability in semi-professional soccer players. The Open Sports Sciences Journal. 2008; 1:15–9.

29. Teixeira AS, Arins FB, De Lucas RD, Carminatti LJ, Dittrich N, Nakamura FY, et al. Comparative effects of two interval shuttle-run training modes on physiological and performance adaptations in female professional futsal players. J Strength Cond Res. 2017. Epub 2017/09/14. https://doi.org/10.1519/JSC.0000000000002186 PMID: 28902113.

30. Ostojic SM. Seasonal alterations in body composition and sprint performance of elite soccer players. Journal of Exercise physiology online. 2003; 6(3).

31. Hortobagyi T, Houmard JA, Stevenson JR, Fraser DD, Johns RA, Israel RG. The effects of detraining on power athletes. Med Sci Sports Exerc. 1993; 25(8):929–35. Epub 1993/08/01. PMID: 8371654.

32. Bangsbo J, Mizuno M. Morphological and metabolic alterations in soccer players with detraining and retraining and their relation to performance1988.

9

Leisure-time physical activity and sports in the Brazilian population: A social disparity analysis

Margaret Guimarães Lima[1☯*], Deborah Carvalho Malta[2☯], Camila Nascimento Monteiro[3,4‡], Neuciani Ferreira da Silva Sousa[5‡], Sheila Rizzato Stopa[6‡], Lhaís de Paula Barbosa Medina[1‡], Marilisa Berti de Azevedo Barros[1☯]

1 School of Medical Sciences, State University of Campinas, Campinas, SP, Brazil, **2** School of Nursing, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil, **3** Núcleo de Indicadores e Sistemas de Informação, Hospital Israelita Albert Einstein, São Paulo, SP, Brazil, **4** Department of Preventive Medicine, Medicine School, University of São Paulo, São Paulo, SP, Brazil, **5** Institute of Collective Health, Federal University of Mato Grosso, Cuiabá, MT, Brazil, **6** Department of Health Analysis and Noncommunicable Diseases Surveillance, Health Surveillance Secretariat, Ministry of Health, Brasília, DF, Brazil

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* mglima@unicamp.br

Editor: Brecht Devleesschauwer, Sciensano, BELGIUM

Abstract

Objectives

To estimate the prevalence of leisure-time physical activity (LTPA) or sports in the Brazilian population according to demographic and income variables.

Methods

Data from 60,202 Brazilian individuals (18 years and over) were analyzed, belonging to the National Health Survey 2013 sample. The prevalence of different modalities of LTPA and sports was estimated according to age, sex, skin color and income. The adjusted prevalence ratios were estimated by Poisson regression.

Results

Of every thousand Brazilians, 695 do not practice LTPA or sports. Walking is the most practiced LTPA (98/1000), followed by soccer (68/1000) and weight training (45/1000). For poor and black men, the most frequent LTPA was soccer, and, for women, gymnastics and walking. The prevalence of weight training and gymnastics was higher for white people compared with black people. All LTPA practices were more prevalent in individuals with higher income, except for soccer. Running on a treadmill and weight training had, respectively, 24.7 and 6.4 times higher prevalence in the richer quartile.

Conclusions

The study allowed identifying the type of LTPA and sport reported as the most frequent by the Brazilian population according to age, sex, skin color, and income, detecting strong social disparities in these practices.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Several studies [1–8] analyzed the levels of leisure-time physical activity (LTPA) based on information about duration and intensity of the practices, and some of them, providing results on the classification of the individuals in active, insufficiently active and inactive. However, fewer studies estimated the prevalence of the type of LTPA practices [9–13].

The guides and recommendations for physical exercise, aimed at promoting health, consider the type of physical activities in addition to levels, duration, frequency and intensity [14,15]. Knowledge on the type of LTPA practiced by the population is important, considering that different dimensions of physical capacity, such as the cardiorespiratory fitness, muscular strength and neuromotor function exert particular influences in the various aspects of health. The best cardiorespiratory fitness may be achieved through aerobic exercises, while increasing muscle strength is possible with weight training exercises and other activities using external weights or the body itself. Exercises as Yoga, Pilates and dance, in their turn, contribute to improve neuromotor functions [14–16].

Each type of LTPA may depend, in a specific way, on some factors for its accomplishment, such as location and income, considering that LTPA can be practiced indoors and paid, or outdoors, in squares, beaches, sidewalks, without necessarily demanding financial resources [17], and some activities require more expensive devices and accessories [18]. Some practices also are dependent on cultural factors [19–21].

Inequities in the health conditions of the population are widely known [22], and studies have shown that the disparities are also present in health behaviors and, particularly, in LTPA [1–5]. Although various studies have been conducted to monitor social disparities in LTPA levels, the analysis regarding type of physical activity are still scarce.

The Brazilian National Human Development Report, published by the United Nations Development Programme (UNDP) [17], showed, based on the National Household Sample Survey (PNAD) data, information on some of the main types of LTPA. Despite having descriptive analyses and not assessing associations regarding race/skin color, it points out to social disparities in the LTPA practice. In addition, the information available in the National Health Survey on the type of the practice of LTPA and sports was not analyzed, and only two population-based investigations assessed the socioeconomic disparities of these practices in the Brazilian population [10,17], but none evaluated them according to race or skin color. Other studies were performed with specific group [23] or in a particular region [24, 25].

The analysis of the disparities regarding type of physical activity may provide knowledge about associations in different directions regarding gender, age, race, income, in the several types of LTPA or sports, expanding and enriching our understanding about more and less widespread practices. These findings will be useful to guide more targeted and specific LTPA strategies.

This study aimed to estimate the prevalence of the type of LTPA or sport reported as the most frequent, and to verify the association with age, sex, skin color and income in the Brazilian population.

Methods

The study was carried out using data from the PNS, developed by the Brazilian Institute of Geography and Statistics (IBGE) in partnership with the Brazilian Ministry of Health and the Oswaldo Cruz Foundation, in 2013 and 2014. The PNS sampling was stratified and conducted in three stages: census, household, and resident tracts. In the first stage, the primary sampling units (PSU) were composed of one or more census tracts; in the second stage, 10 to 14 households of each PSU were selected randomly; and, in the third and final stage, an adult (18 years or older) in each household was selected with equiprobability among other residents for the individual interview. The number of households covered by the PNS was of 64,348, and 60,202 individuals were interviewed, with 86.1% response rate.

Data were collected using personal digital assistance (PDA), programmed for critic data entry processes at the time of data collection. A detailed description of the development of the PNS was published [26].

Information on LTPA was obtained from 2 questions. The first one was whether the interviewee practiced or not any type of LTPA in the last three months before the survey. If the answer was positive, they were questioned about the LTPA or sport more often practiced in the period and only one of the following was recorded: outdoor walking, treadmill walking, street running, treadmill running, biking (on the street or ergometric), aerobics, swimming, dancing (practicing PA purpose), muscle training, water aerobics, gymnastics/Pilates/stretching/Yoga, soccer, basketball, tennis, martial arts and wrestling, other practices. Each of these activities was analyzed as one variable, categorized in (0) do not practice and 1 (yes).

In order to describe the sample according to the main socio-demographic variables, we also analyzed sex, age in years (18 to 29; 30 to 59; 60 or more), self-reported skin color (black and brown; white), and monthly family income per capita in quartiles: (1) up to R$355.99 (= low income); (2) R$356 to R$677.99; (3) R$678 to 1,199.99; (4) R$1,200.00 or more (= high income).

The prevalence per thousand people of the reported practices was estimated according to age group, sex, skin color and income. Prevalence differences were tested using Pearson's Chi-square (*Rao-Scott correction*). Multiple Poisson regression models were used to estimate the prevalence ratios (PR) [27] and their respective confidence intervals (CI 95%). We made adjustments for sex and age, considering that these factors can be determinants of LTPA [4, 6, 17, 28]. To verify associations with skin color, we added income to the adjustment in order to know to what extent the relationship between race and type of LTPA can be explained by income, considering the strong income disparity by race [29]. Data analysis was performed using STATA software version 15.0, with *svy* commands, which consider the corrections for non-response and adjustments for the post-stratification [30].

The PNS was approved by the National Research Ethics Committee (CONEP) under process no. 328,159 of June 26, 2013.

Results

For every 1,000 inhabitants, 695 did not practice LTPA, and, in the elderly group, only 213 for every 1,000 people reported practicing it. The most frequent LTPA was walking (98/1,000), followed by soccer (68/1,000) and weight training (45/1,000). Soccer and muscle training were substantially higher among individuals aged from 18 to 29 years in relation to the elderly. In adults, the activities reported as more frequent were walking, followed by soccer and muscle training. The prevalence of outdoor walking (PR = 2.06) and gymnastics/stretching/Yoga (PR = 1.62) were higher in elderly people, in addition to water aerobics (PR = 26.2), which was rare among younger adults (Table 1).

Table 1. Prevalence and prevalence ratios of LTPA and sports reported as the most frequently practiced, according to age group, PNS 2013.

LTPA and sports reported as the most frequently practiced	Prevalence (per thousand inhabitants)					
	Total	Age			Prevalence ratios[a] adjusted for sex (CI 95%)	
		18 to 29 (1)	30 to 59 (2)	60 or more (3)	2/1	3/1
No LTPA	695.2	581.6	718.7	786.8	1.23 (1.20–1.27)	1.34 (1.30–1.38)
LTPA						
Outdoor walking	98.1	60.2	106.5	127.0	1.75 (1.54–1.99)	2.06 (1.77–2.39)
Soccer	67.7	149.4	50.0	4.2	0.35 (0.32–0.39)	0.03 (0.02–0.04)
Muscle training	45.1	91.1	35.2	9.2	0.39 (0.34–0.45)	0.10 (0.07–0.14)
Aerobics, spinning, step, jump	17.0	21.6	16.9	10.7	0.76 (0.59–0.99)	0.46 (0.31–0.68)
Street running	14.4	25.7	13.1	1.6	0.52 (0.40–0.68)	0.6 (0.41–0.11)
Gymnastics, stretching, Yoga	14.0	11.6	13.1	20.3	1.09 (0.81–1.48)	1.62 (1.15–2.27)
Biking (street/ergometric)	9.7	11.3	10.5	5.2	0.97 (0.70–1.35)	0.49 (0.32–0.76)
Treadmill walking	7.4	5.6	8.7	6.2	1.55 (1.03–2.32)	1.08 (0.61–1.94)
Martial arts and wrestling	4.5	9.2	3.3	1.3	0.37 (0.23–0.59)	0.15 (0.1–0.45)
Water aerobics	4.3	0.04	3.8	11.4	9.26 (3.05–28.0)	26.2 (8.57–80.2)
Swimming	3.2	4.3	3.0	2.1	0.72 (0.44–1.15)	0.51 (0.26–1.01)
Basketball, volleyball	2.9	6.7	1.9	0.8	0.28 (0.17–0.47)	0.12 (0.03–0.44)
Dancing (practicing PA purpose)	2.7	3.8	2.4	1.9	0.62 (0.37–1.06)	0.48 (0.24–0.98)
Treadmill running	2.6	2.3	3.2	1.9	1.26 (0.73–2.16)	0.79 (0.30–2.14)
Tennis	1.8	1.7	2.2	0.7	1.36 (0.39–4.72)	0.44 (0.1–2.1)
Others	9.0	13.3	7.2	8.4	0.55 (0.37–0.80)	0.64 (0.41–1.01)

[a] Reference category: 18 to 29 years old

The main LTPA practiced by men were soccer, walking, muscle training, street running and biking. Among women, the most frequent were walking, water aerobics, muscle training, gymnastics and treadmill walking. Regarding sex, the greater differences were in the following practices: soccer (PR = 0.03), martial arts (PR = 0.25) running on the street (PR = 0.31), biking (PR = 0.41), swimming (PR = 0.45) and muscle training (PR = 0.71), with lower frequencies of practice for women. On the other hand, practices markedly more frequent among the female population were water aerobics (PR = 6.23), aerobics (PR = 5.69) and other gymnastics (PR = 4.41), and dancing, which was almost three times higher in this population (Table 2).

All LTPA and sports were more prevalent among individuals who self-reported being white, except for dancing, tennis, martial arts and wrestling, for which no statistically significant difference was observed. Soccer was the only one whose prevalence was higher among black and brown people. After adjustment according to monthly income per capita, associations remained significant only for muscle training and gymnastics (exercises, stretching and Yoga), with a prevalence that was 17% and 39% higher in white people in relation to black, respectively. Soccer remained associated with higher prevalence in black people after adjustment (Table 3).

The prevalence of all LTPA practices, reported as the most frequent, was higher in individuals with higher income, comparing the first and last income quartiles, except for soccer. The greatest disparities were observed in treadmill running (PR = 24.7), treadmill walking (PR = 7.6), and muscle training (PR = 6.4). Gymnastics showed a growing gradient as income increased. Soccer was the only practice with higher prevalence in the poorest quartile of the population (Table 4).

Table 2. Prevalence (per 1,000) and prevalence ratios of LTPA and sports reported as the most frequently practiced, according to sex, PNS 2013.

LTPA and sports reported as the most frequently practiced	Sex		p	Prevalence ratios adjusted for age[a] (CI 95%)
	Male	Female		
No LTPA	644.9	740.0	<0.0001	1.14 (1.12–1.16)
LTPA				
Outdoor walking	76.6	117.2	<0.0001	1.50 (1,36–1.66)
Soccer	138.7	4.5	<0.0001	0.03 (0.02–0.04)
Aerobics, spinning, step, jump	5.0	27.8	<0.0001	5.69 (4.29–7.56)
Street running	22.8	6.9	<0.0001	0.31 (0.23–0.42)
Muscle training	54.4	36.8	<0.0001	0.71 (0.62–0.81)
Gymnastics, stretching, Yoga	5.0	22.1	<0.0001	4.41 (3.30–5.89)
Biking (street/ergometric)	14.2	5.8	<0.0001	0.41 (0.31–0.55)
Treadmill walking	4.7	10.0	0.0001	2.14 (1.47–3.11)
Martial arts and wrestling	7.5	1.8	<0.0001	0.25 (0.15–0.42)
Water aerobics	10.8	71.8	<0.0001	6.23 (3.22–12.04)
Swimming	4.5	2.0	0.0001	0.45 (0.29–0.69)
Basketball, volleyball	2.7	3.2	0.5285	1.24 (0.76–2.02)
Dancing (practicing PA purpose)	1.4	3.9	0.0001	2.77 (1.66–4.63)
Treadmill running	2.1	3.1	0.1196	1.45 (0.91–2.41)
Tennis	3.3	0.4	<0.0001	0.12 (0.04–0.32)
Others	10.9	7.4	0.0229	0.69 (0.49–0.97)

[a] Reference category: Male

Table 3. Prevalence (per 1,000) and prevalence ratios of LTPA and sports reported as the most frequently practiced, according to skin color, PNS 2013.

LTPA and sports reported as the most frequently practiced	Black and brown	White	p	PR[a] (CI 95%)	PR[b] (CI 95%)[c]
No LTPA	718.0	672.8	<0.0001	0.92 (0.91–0.94)	1.00 (0.98–1.02)
LTPA					
Outdoor walking	89.9	106.0	0.0005	1.13 (1.03–1.24)	0.99 (0.90–1.09)
Soccer	80.0	54.6	<0.0001	0.79 (0.71–0.88)	0.94 (0.75–0.94)
Muscle training	36.8	53.3	<0.0001	1.62 (1.41–1.86)	1.17 (1.02–1.34)
Street running	12.4	16.5	0.0168	1.48 (1.17–1.88)	1.05 (0.81–1.37)
Aerobics, spinning, step, jump	14.3	20.0	0.0013	1.44 (1.17–1.76)	1.05 (0.86–1.28)
Gymnastics, stretching, Yoga	9.6	18.8	<0.0001	1.89 (1.46–2.44)	1.39 (1.07–1.82)
Biking (street/ergometric)	8.6	10.9	0.0890	1.33 (1.00–1.77)	1.16 (0.86–1.56)
Treadmill walking	4.9	10.1	<0.0001	2.02 (1.43–2.83)	1.29 (0.90–1.87)
Martial arts and wrestling	4.8	3.9	0.4211	0.96 (0.61–1.51)	0.75 (0.47–1.20)
Water aerobics	30.7	57.3	0.0010	1.58 (1.09–2.30)	1.06 (0.72–1.57)
Swimming	2.3	4.2	0.0050	1.94 (1.25–3.01)	1.26 (0.84–1.89)
Basketball, volleyball	2.4	3.6	0.1247	1.67 (1.01–2.77)	1.39 (0.84–2.30)
Dancing (practicing PA purpose)	2.5	2.8	0.6595	1.15 (0.71–1.87)	0.87 (0.53–1.44)
Treadmill running	20.1	34.5	0.0329	1.75 (1.05–2.91)	1.12 (0.67–1.84)
Tennis	1.1	2.5	0.1692	2.30 (0.77–6.86)	1.27 (0.36–4.49)
Others	7.2	10.7	0.0216	1.56 (1.12–2.18)	1.18 (0.83–1.68)

[a] Prevalence ratio with adjustments by sex and age

[b] Prevalence ratio with adjustments by sex, age and income

[c] Reference category: black and brown.

Table 4. Prevalence (per 1,000) and prevalence ratios of LTPA and sports reported as the most frequently practiced, according to monthly family income per capita, PNS 2013.

LTPA and sports, reported as the most frequently practiced	Income in deciles of minimum wage				PR adjusted by sex and age (CI 95%)[a]		
	1st quartile (low income) (1)	2nd quartile (middle-low income) (2)	3rd quartile (middle-high income) (3)	4th quartile (high income) (4)			
	Prevalence per 1,000 inhabitants				2/1	3/1	4/1
No LTPA	790.6	746.2	709.1	555.8	0.93 (0.91–0.95)	0.86 (0.84–0.88)	0.68 (0.66–0.70)
LTPA							
Outdoor walking	67.9	86.5	96.9	135.7	1.25 (1.07–1.46)	1.33 (1.15–1.52)	1.87 (1.63–2.13)
Soccer	85.8	71.8	65.4	51.3	0.92 (0.80–1.06)	0.97 (0.83–1.12)	0.81 (0.69–0.95)
Muscle training	17.0	31.7	45.2	80.6	2.05 (1.52–2.77)	3.34 (2.48–4.50)	6.35 (4.84–8.32)
Street running	6.5	9.8	9.8	29.7	1.64 (0.93–2.88)	1.85 (1.08–3.16)	5.88 (3.54–9.78)
Aerobics, spinning, step, jump	7.6	12.2	17.0	29.3	1.68 (1.13–2.52)	2.50 (1.78–3.50)	4.35 (3.13–6.06)
Gymnastics, stretching, Yoga	5.9	10.0	11.0	27.4	1.68 (1.03–2.73)	1.78 (1.14–2.80)	4.44 (2.94–6.71)
Biking (street/ergometric)	5.8	9.1	1.0	1.3	1.7 (1.07–2.48)	1.93 (1.23–3.02)	2.49 (1.61–3.83)
Treadmill walking	2.3	2.4	7.0	1.7	1.07 (0.53–2.17)	3.13 (1.59–6.16)	7.58 (3.97–14.48)
Martial arts and wrestling	1.7	5.2	3.5	7.0	3.32 (1.62–6.80)	2.60 (1.36–4.96)	5.56 (3.16–9.80)
Water aerobics	1.2	1.6	3.8	9.9	1.12 (0.46–2.79)	2.08 (0.91–4.76)	5.79 (2.84–11.81)
Swimming	0.5	1.4	2.9	7.3	2.62 (0.89–7.70)	5.59 (1.96–15.95)	14.4 (5.40–38.6)
Basketball, volleyball	2.1	3.2	1.9	4.5	1.70 (0.84–3.43)	1.18 (0.63–2.20)	2.95 (1.69–5.14)
Dancing (practicing PA purpose)	1.2	1.5	3.8	4.0	1.29 (0.50–3.30)	3.49 (1.47–8.32)	3.76 (1.69–8.37)
Treadmill running	0.3	1.7	0.2	5.9	6.80 (2.54–18.19)	9.78 (3.81–25.07)	24.70 (11.04–55.27)
Tennis	0.6	0.03	0.5	5.7	0.55 (0.04–0.75)	0.81 (0.08–8.84)	10.19 (1.22–84.95)
Others	3.9	5.4	9.7	15.9	1.45 (0.88–2.39)	2.83 (1.71–4.69)	4.69 (3.02–7.27)

[a] Reference category: 1st income quartile.

Discussion

The prevalence of LTPA in this study was low, considering that only 305 out of 1,000 individuals reported practicing LTPA and sports and that this observation is even more alarming among the elderly. Women are 14% less active than men and studies in Brazil [6, 17] and in other countries [4,28] also pointed out to a higher prevalence of LTPA among men and younger individuals. In the Brazilian population, this evidence has been attributed to cultural aspects, considering that boys used to be more encouraged to participate in outdoor games. In the Brazilian culture, girls are most encouraged to engage in children games such as playing house and cooking, generally with dolls [31], stimulating the permanence of women at home in a sexist way.

Walking was the main practice reported, which was also verified by other authors [7, 17, 32], however, some groups of the population, such as men and younger individuals, did not report this activity as the most frequent. The prevalence of walking grew as age increased, similar to what was observed in a study using the PNAD data in 2015 [17]. Although it is a very common practice, national and abroad, the incentives to walking for leisure purposes should be valued and maintained since this exercise can be practiced by different population groups, with low cardiorespiratory and neuromotor demand and minimum cost because it does not require special equipment or accessories, and in democratic places [33]. In addition, special walking programs, such as trekking would be encouraged, which can also attract younger people and men. Brazil has a great potential to the development of these activities. Based on *Camino de Santiago de Compostela*, in Spain, there are three important routes in Brazil, in the states of Minas Gerais and São Paulo: *O Caminho do Sol*, *O Caminho da Fé*, *O Caminho da Luz* [33].

Soccer is the second most frequently practice sport nationwide, but the first among men (139 among 1,000). In Brazil, soccer has always been the favorite sport among men, with great cultural appeal and traditionally practiced in open and public spaces, in the cities, outskirts and neighborhoods, country towns and rural areas [10]. However, VIGITEL data showed a fall in the soccer practice after 2012 in Brazilian capitals and, according to Rodrigues (2008) [34], the decrease of the soccer game in major cities must be due to real estate speculation and the replacement of these spaces with buildings and roads.

Brazil has stood out in women's soccer, and a female player from the Brazilian team is the athlete who scored the most goals in the history of both women's and men's soccer cups. Nevertheless, the visibility of women in soccer in Brazil is still low, which is mainly due to the prejudice against women participating in this sport, considered for men [19,35]. Female soccer could be a great opportunity in LTPA practices and a good strategy to raise the women's level of physical activity if it was more encouraged among the female population.

Muscle training is the third LTPA reported and is more prevalent in young adults and men, noting that only 37 women and 9 elder people out of 1,000 individuals reported this practice as the most frequent. One emphasizes that muscle strength and resistance exercises are very important, especially for these population groups. Evidence shows that these activities can act positively on bone and muscular mass, being constituted of considerable exercises to prevent and control musculoskeletal diseases, such as osteoporosis [36, 37]. Muscle strength and resistance exercises, along with balance exercises, also assist in the prevention of falls and disabilities [15,38].

Water aerobics is a very common practice for the elderly in Brazil [21]. This phenomenon can lead to some generation conflict in LTPA, as young and adult people tend to avoid this practice because of prejudice and the elderly, on the other hand, lose the company of younger people.

In this study, the prevalence of the practice of almost all types of LTPA was greater, except for soccer, in individuals with white skin. However, after income adjustments, associations lost significance in most activities. In Brazil, the strong economic inequality in relation to race/skin color is known, disfavoring black people [29], and worse economic conditions tend to hinder the practice of some types of LTPA due to the need for expenses to practice them. In addition, people in worse economic condition usually also live in areas with less facilities and spaces for recreational physical activity. However, the associations between race/skin color and muscle training and gymnastics (localized exercises, stretching and Yoga) remain with higher prevalence among white people, even after adjustments by income, showing that, not just income, but other aspects related to segregation can contribute to the racial disparities in LTPA. Soccer also remains associated with black people and is more prevalent among them. As far as is known, no studies evaluated the association of race/skin color with the practice of several types of LTPA. Although scarce, investigations have only examined walking. Bates et al. (2005) [11], and Kruger et al. (2008) [33] verified a higher prevalence of walking in non-Hispanic white people when compared with black people.

The results of the analyses by income are coherent with the results of the PNAD in 2015, published in a UNDP report [17]. In Brazil, De Sá et al. (2014) [10] observed that the prevalence of muscle training and running are around 3 times higher for people with the better educational level, than for the less educated. It is noteworthy that this study detected a prevalence of treadmill running 24 times higher compared with the extreme income quartiles; treadmill walking was 7.6 times higher, and muscle training was 6.4 times higher in the segment of the highest income. Muscle training, as well as gymnastics, demand equipment and special accessories, which might hinder and decrease its practice among the poorest. Some municipalities have implemented outdoor academies, which include the fixing of exercise equipment for strength training and stretching in open and public places. This could contribute to decrease the disparities, however, the results of these strategies are still little studied [39].

In 2011, the Brazilian Ministry of Health created the Health Academy Program (PAS), which consist of activities integrated to the primary health care of the Brazilian Unified Health System (SUS), in order to promote physical activities, healthy eating and educational programs by territories [40]. Maintaining and analyzing this government strategy is important, since social disparities are still strong in several LTPA. The appropriation of the city, as well as of the public spaces, connects to aspects related to the promotion of health habits and healthy lifestyles and the regular practice of physical activities. Historically, cities have been built as spaces for production of disparities, since they have not been able to meet socially determined specificities [41].

Soccer is a strong social marker, being the only sport more present in the poor and black population and strategies to maintain and not to hinder this practice among this population is important to guarantee some activity for more socioeconomic vulnerable groups. According to the results of the 2015 PNAD, people with lower income reported practicing LTPA especially because they like to compete and to have fun [17], which is contemplated by a soccer match. On the other hand, individuals in the higher-income stratum mostly reported practicing LTPA or sports to improve their quality of life and well-being [17].

The results of this study bring important elements to understand existent disparities in prevalence of the type of LTPA or sport and to provide knowledge about more elite and more democratic practices, pointing out to the need for specific interventions to rise and maintain physical activity levels in the Brazilian population.

However, the study has some limitations. The question about physical activity was self-referred and tend to be overestimated [42, 43], the validity and reliability of the outcome questions are not known, however other studies assessed the LTPA using similar information [9,10,14]. In addition, self-reported data about the most frequent physical LTPA and sports practiced are self-referred and, possibly, the study could be influenced by some information bias. Also, only the most frequently practiced LTPA was analyzed instead of all the activities practiced by the individuals. Nonetheless, describing population subgroups that refer certain practice as the main one was possible. On the other hand, the study was elaborated with data from the PNS, a survey developed with representative sample of the Brazilian population. Information was discussed on the type of LTPA, which have been little analyzed in the national scope, especially focusing on social disparities, which are crucial when considering the population's health.

Conclusions

This study allowed identifying the types of LTPA and sports reported as the most frequent by the Brazilian population according to age, sex, race/skin color and income, detecting strong social and demographic disparities in these practices. These results reinforce the importance of encouraging the practice of LTPA, disseminating knowledge on the benefits of the different practices and facilitating these activities through appropriate spaces and means to perform them. These tasks should be maintained, assessed and regularly rethought considering especially the health of the less favored subgroups. In addition, thinking about strategies to try to avoid age and gender prejudice in the types of LTPA is important.

Acknowledgments

The authors thank the Espaço da Escrita–Pró-Reitoria de Pesquisa—UNICAMP—for the language services provided and the National Council for Scientific and Technological Development (CNPq) for the scholarship of productivity to D. C. Malta and M. B. A. Barros.

Author Contributions

Conceptualization: Margareth Guimarães Lima, Deborah Carvalho Malta, Marilisa Berti de Azevedo Barros.

Formal analysis: Margareth Guimarães Lima, Deborah Carvalho Malta, Marilisa Berti de Azevedo Barros.

Investigation: Deborah Carvalho Malta.

Methodology: Margareth Guimarães Lima, Deborah Carvalho Malta, Marilisa Berti de Azevedo Barros.

Resources: Margareth Guimarães Lima, Deborah Carvalho Malta, Camila Nascimento Monteiro, Neuciani Ferreira da Silva Sousa, Sheila Rizzato Stopa, Lhaís de Paula Barbosa Medina, Marilisa Berti de Azevedo Barros.

Validation: Margareth Guimarães Lima, Deborah Carvalho Malta, Camila Nascimento Monteiro, Neuciani Ferreira da Silva Sousa, Sheila Rizzato Stopa, Lhaís de Paula Barbosa Medina, Marilisa Berti de Azevedo Barros.

Visualization: Margareth Guimarães Lima, Deborah Carvalho Malta, Camila Nascimento Monteiro, Neuciani Ferreira da Silva Sousa, Marilisa Berti de Azevedo Barros.

Writing – original draft: Margareth Guimarães Lima, Deborah Carvalho Malta, Sheila Rizzato Stopa, Lhaís de Paula Barbosa Medina, Marilisa Berti de Azevedo Barros.

Writing – review & editing: Margareth Guimarães Lima, Deborah Carvalho Malta, Camila Nascimento Monteiro, Neuciani Ferreira da Silva Sousa, Sheila Rizzato Stopa, Lhaís de Paula Barbosa Medina, Marilisa Berti de Azevedo Barros.

References

1. Mielke GI, Malta DC, de Sá GB, Reis RS, Hallal PC. Regional differences and correlates of leisure time physical activity in Brazil: results from the Brazilian National Health Survey-2013. Rev Bras Epidemiol. 2015a; Suppl 2:158–169.
2. Barros MBA, Lima MG, Medina LP, Szwarcwald CL, Malta DC. Social inequalities in health behaviors among Brazilian adults: National Health Survey, 2013. Int J Equity in Health 2016. https://doi.org/10.1186/s12939-016-0439-0 PMID: 27852275

3. Oliveira GD, Oancea SC, Nucci LB, Vogeltanz-Holm N. The association between physical activity and depression among individuals residing in Brazil. Soc Psychiatry Psychiatr Epidemiol. 2018; 53(4):373–383 https://doi.org/10.1007/s00127-017-1441-6 PMID: 28889252

4. Bauman A, Ma G, Cuevas F, Omar Z, Waqanivalu T, Phongsavan P, et al. Cross-national comparisons of socioeconomic differences in the prevalence of leisure-time and occupational physical activity, and active commuting in six Asia-Pacific countries. J Epidemiol Community Health 2011; 65(1):35–43. https://doi.org/10.1136/jech.2008.086710 PMID: 20943821

5. Beenackers MA, Giskes CBMK, Brug J, Kusnt AE, Burdorf A, Lenthe FJV. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: A systematic review. Int J Behav Nutr Phys Act. 2012 Sep 19; 9:116. https://doi.org/10.1186/1479-5868-9-116 PMID: 22992350

6. Azevedo MR, Araújo CL, Reichert FF, Siqueira FV, da Silva MC, Hallal PC. Gender differences in leisure-time physical activity. Int J Public Health 2007; 52(1):8–15. https://doi.org/10.1007/s00038-006-5062-1 PMID: 17966815

7. Malta DC, Andrade SSA, Santos MAS, Rodrigues GBA, Mielke G. Tendências dos indicadores de atividade física em adultos: conjunto de capitais do Brasil 2006–2013. Rev Bras Ativ Fís Saúde 2015; 20 (2):141–151.

8. Moore LV, Harris CD, Carlson SA, Kruger J, Fulton JE. Trends in no leisure-time physical activity—United States, 1988–2010. Res Q Exerc Sport. 2012; 83(4):587–591. https://doi.org/10.1080/02701367.2012.10599884 PMID: 23367822

9. Malta DC, Moura EC, Castro AM, Cruz DKA, Morais Neto OL, Monteiro CA. Padrão de atividade física em adultos brasileiros: resultados de um inquérito por entrevistas telefônicas, 2006. Epidemiol. Serv. Saúde 2009; 18(1):7–16.

10. De Sa TH, Garcia LMT, Claro RM. Frequency, distribution and time trends of types of leisure-time physical activity in Brazil, 2006–2012. International Journal of Public Health 2014; 59(6):975–982. https://doi.org/10.1007/s00038-014-0590-6 PMID: 25047019

11. Bates JH; Serdula MK; Khan LK; Jones DA; Gillespie C; Ainsworth BE. Total and leisure-time walking among U.S. adults should every step count? Am J Prev Med 2005; 29(1):46–50. https://doi.org/10.1016/j.amepre.2005.03.011 PMID: 15958251

12. Stamatakis E, Chaudhury M. Temporal trends in adults' sports participation patterns in England between 1997 and 2006: the Health Survey for England. Br J Sports Med 2008; 42:901–908. https://doi.org/10.1136/bjsm.2008.048082 PMID: 18658250

13. Monteiro CA, Conde WM, Matsudo SM, Matsudo VR, Bonseñor, Lotufo PA. A descriptive epidemiology of leisure-time physical activity in Brazil, 1996–1997. Pan Am J Public Health 2003; 14(4):246–254.

14. World Health Organization. WHO | Global recommendations on physical activity for health. Geneva, 2010. Available from: https://www.who.int/dietphysicalactivity/global-PA-recs-2010.pdf

15. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc. 2011; 43(7):1334–1359. https://doi.org/10.1249/MSS.0b013e318213fefb PMID: 21694556

16. Prestes J; Assumpção CO. Ginástica em academias. In: Gaio T; Góis AAF; Batista JCF (Orgs.). A Ginástica em Questão–Corpo e Movimento: Phort Editora; 2010. pp. 137–158.

17. Brasil, 2017. Movimento é vida: Atividades físicas e esportivas para todas as pessoas. Relatório Nacional de Desenvolvimento Humano no Brasil. Programa das nações unidas para o desenvolvimento–PNUD. Brasília, DF, 2017. Avaliable from:http://movimentoevida.org/wpcontent/uploads/2017/09/PNUD_RNDH_completo.pdf

18. Liz CM, Crocetta TB, Viana MS, Brandt R, Andrade A. Aderência à prática de exercícios físicos em academias de ginástica. Motriz 2010; 16:181–188.

19. Fanzini F. Futebol é "coisa para macho"? Pequeno esboço para uma história de mulheres no país do futebol. Revista Brasileira de História 2005; 25(50):315–328.

20. Seabra AF, Mendonça DM, Thomis MA, Anjos LA, Maia JA. Biological and sócio-cultural determinants of physical activity in adolescents. Cad. Saúde Pública 2005; 24(4):721–736.

21. Locatelli PAPC, Cavedon NR. As gurias: exercício etnográfico realizado com mulheres idosas praticantes de hidroginástica. Ciências sociais em Perspectiva 2011; 10(18): 45–61.

22. Marmot M. WHO Commission on Social Determinants of Health. Closing the health gap in a generation: the work of the Commission on Social Determinants of Health and its recommendations. Glob Health Promot.2009; Suppl 1: 23–27. https://doi.org/10.1177/1757975909103742 PMID: 19477825

23. Silva KS, Lopes AS, Duca GFD, Garcia LMT, Nahas MV. Patterns of engagement in leisure-time physical activities of workers with different economic status: a descriptive analysis. Brazilian Journal of Kinanthropometry and Human Performance. 2013; 15(6):656–666.

24. Zanchetta LM. Barros MBA, César CLG, Carandina L, Goldbaum M, Alves MCGP. Inatividade física e fatores associados em adultos, São Paulo, Brasil. Rev Bras Epidemiol 2010; 13(3): 387–399. https://doi.org/10.1590/s1415-790x2010000300003 PMID: 20857026

25. Dumith SC, Domingues MR, Gigante DP. Epidemiologia das atividades físicas praticadas no tempo de lazer por adultos do Sul do Brasil. Rev Bras Epidemiol 646 2009; 12(4): 646–658.

26. Damacena GN, Szwarcwald CL, Malta DC, Souza-Júnior PRB, Vieira MLFP, Pereira CA, et al. O processo de desenvolvimento da Pesquisa Nacional de Saúde no Brasil, 2013. Epidemiologia e Serviços de Saúde 2015; 24:197–206.

27. Barros A, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Medical Research Methodology 2003; 3:21–34. https://doi.org/10.1186/1471-2288-3-21 PMID: 14567763

28. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, et al. Global physical activity levels: surveillance progress, pitfalls and prospects. Lancet 2012; 380(9839):20–30.

29. Brasil, OXFAM. A distância que nos une: Um retrato das desigualdades brasileiras. Set., 2017. Available from: https://oxfam.org.br/um-retrato-das-desigualdades-brasileiras/a-distancia-que-nos-une/

30. Souza-Júnior PRB, Freitas MPS, Antonaci GA, Szwarcwald CL. Sampling Design for the National Health Survey, Brazil 2013. Epidemiol. Serv. Saúde 2015 24(2). https://doi.org/10.5123/S1679-49742015000200003

31. Oliveira Pinto T, Lopes MF. Brincadeira no espaço da rua e a demarcação dos gêneros na infância. Revista Latinoamericana de Ciencias Sociales 2009; 7(2):861–885.

32. Kruger J, Ham SA, Berrigan D, Ballard-Barbash. Prevalence of transportation and leisure time walking among U.S. adults. Preventive Medicine 2008; 47:329–334. https://doi.org/10.1016/j.ypmed.2008.02.018 PMID: 18445507

33. Andrade COP, Godinho RCR, Magri RAF. Elaboração e aplicação de uma rota de trekking em uma área do arque Nacional da Serra da Canastra. Revista Brasileira de Ecoturismo 2016; 9(2):285–307.

34. Rodrigues CAFG. Olhares sobre a cidade: reflexos das alterações urbanas na formação do jogador de futebol ao longo do século XX. Monografy, Federal University of Uberlandia.2008. Available from: https://repositorio.ufu.br/bitstream/123456789/18870/1/OlharesCidadeFutebol.pdf

35. Teixeira FLS, Caminha IO. Preconceito no futebol feminino brasileiro: uma revisão sistemática. Movimento 2013; 19(1):265–287.

36. Suominen H. Muscle training for bone strength. Aging Clin Exp Res. 2006; 18(2):85–93. https://doi.org/10.1007/bf03327422 PMID: 16702776

37. Senderovich H, Tang H, Belmont. The role of exercises in osteoporotic fracture prevention and current care gaps. Where are we now? Recent updates. Rambam Maimonides Medical Journal 2017 Jul 8(3). https://doi.org/10.5041/RMMJ.10308 PMID: 28786812

38. Clemson L, Singh MAF, Bundy A, Cumming RG, Manollaras K, O'Loughlin P, et al. Integration of balance and strength training into daily life activity to reduce rate of falls in older people (the LIFE study): randomized parallel trial. BMJ 2012 Aug. https://doi.org/10.1136/bmj.e4547 PMID: 22872695

39. Lepsen AM, Silva MC. Perfil dos frequentadores das academias ao ar livre da cidade de Pelotas–RS. Rev Bras Ativ Fís Saúde 2015; 20(4):413–24.

40. Brasil, 2014a. Academia da Saúde. Ministério da Saúde, Brasília, DF, 2014. Avaliable from:http://bvsms.saude.gov.br/bvs/publicacoes/academia_saude_cartilha.pdf

41. Fernandes AP, Andrade ACS, Ramos CGC, Friche AAL, Dias MAS, Xavier CC et al. Leisure-time physical activity in the vicinity of Academias da Cidade Program in Belo Horizonte, Minas Gerais State, Brazil: the impact of a health promotion program on the community. Cad. Saúde Pública 2015. https://doi.org/10.1590/0102-311X00104514 PMID: 26648374

42. Boon RM, Hamlin MJ, Steel GD, Ross JJ. Validation of the New Zealand Physical Activity Questionnaire (NZPAQ-LF) and the International Physical Activity Questionnaire (IPAQ-LF) with accelerometry. Br J Sports Med 2010; 44:741–746. https://doi.org/10.1136/bjsm.2008.052167 PMID: 18981036

43. Biernat E, Stupnicki R, Lebiedziński B, Janczewska L. Assessment of physical activity by applying IPQ questionnaire. Physical Education and Sport 2008; 52:46–52.

Type D personality, stress, coping and performance on a novel sport task

Erika Borkoles[1☯*], Mariana Kaiseler[2‡], Andrew Evans[3‡], Chantal F. Ski[4‡], David R. Thompson[4‡], Remco C. J. Polman[5☯]

1 School of Public Health and Social Work, Queensland University of Technology, Brisbane, QLD, Australia, **2** Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, Yorkshire, United Kingdom, **3** Directorate of Sport, Exercise, & Physiotherapy, The University of Salford, Lancashire, United Kingdom, **4** School of Nursing and Midwifery, Queen's University, Belfast, Northern Ireland, **5** School of Exercise & Nutrition Sciences, Queensland University of Technology, Brisbane, QLD, Australia

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* erika.borkoles@qut.edu.au

Editor: Fulvio d'Acquisto, Queen Mary University of London, UNITED KINGDOM

Competing interests: The authors have declared that no competing interest exist.

Abstract

We investigated (1) the relationship between Type D personality, stress intensity appraisal of a self-selected stressor, coping, and perceived coping effectiveness and (2) the relationship between Type D personality and performance. In study one, 482 athletes completed the Type D personality questionnaire (DS14), stress thermometer and MCOPE in relation to a recently experienced sport stressor. Type D was associated with increased levels of perceived stress and selection of coping strategies (more emotion and avoidance coping) as well as perceptions of their effectiveness. In study two, 32 participants completed a rugby league circuit task and were assessed on pre-performance anxiety, post-performance affect and coping. Type D was associated with poorer performance (reduced distance; more errors), decreases in pre-performance self-confidence and more use of maladaptive resignation/withdrawal coping. Findings suggest that Type D is associated with maladaptive coping and reduced performance. Type D individuals would benefit from interventions related to mood modification or enhancing interpersonal functioning.

Introduction

Competitive sport is associated with stressful experiences [1]. For athletes to perform to the best of their ability and to feel satisfied with their performance, it is essential that they use adaptive coping strategies to deal with these stressors [2, 3]. A number of factors have been found to influence the stress and coping process including the personality of athletes [4, 5]. A personality construct which has been shown to have significant consequences for behavior, health, stress, and coping but which has not been examined in the domain of sport is the distressed or Type D personality [6]. To address this research gap, two empirical studies were conducted using a crossectional and quasi-experimental design respectively, to investigate how Type-D personality in a sport context might influence (i) stress, coping preferences and perception of coping effectiveness; (ii) performance on a novel sport task. Type D personality is likely to result in maladaptive coping with stress experienced in sport thereby influencing athletic performance adversely. However, this issue has not been examined to date. As such this research might have important implications for athletes and practitioners.

The cognitive-motivational-relational model of stress and coping describes the relation between stress and coping as a dynamic process between an individual's internal and situational environment [7, 8]. Through primary appraisal an individual will assess the significance of an event to their personal values, beliefs or intentions. Secondary appraisal involves the complex evaluative process in which a person analyses their available coping options in relation to the specific situation, maximizing gains or favorable outcomes and limiting harm [8]. Perceived stress intensity [9] as well as perceptions of control [10] have been found to be factors influencing the coping process.

Coping has been defined as "a constantly changing cognitive and behavioral effort to manage specific external and/or internal demands that are appraised as taxing or exceeding the resources of the person" (p.141) [8]. Coping responses have been classified into higher order dimensions. Nicholls and Polman [11] in their systematic review on coping in sport suggested three dimensions. First, problem-focused coping refers to cognitive and behavioral attempts to manage distress by reducing or eliminating the stressor. Second, emotion-focused coping is concerned with the regulation of emotional arousal and distress [7]. Finally, avoidance coping includes behavioral and psychological efforts to disengage from a stressor [12]. Endler and Parker [13], on the other hand, made a distinction between approach- (efforts to deal with a stressful situation) and avoidance-oriented (cognitive changes to avoid confrontation) coping, whereas Gaudreau and Blondin [14] identified three higher order dimensions (task, distraction and disengagement-oriented coping).

The use of any particular coping strategy does not guarantee its effectiveness; rather, its effectiveness depends on the context in which it is employed [7]. A number of approaches to coping effectiveness have been put forward [15]. The goodness-of-fit approach to coping effectiveness [16] has found that problem-focused coping strategies are more adaptive to athletic performance than either emotion-focused or avoidance coping strategies [3]. However, some avoidance coping strategies can be effective in dealing with acute stressors. Blocking (i.e., shut out thoughts, mentally withdraw from stressor) has been found to be an effective coping strategy in elite rugby union athletes [17]. The latter provides support for the choice of coping strategy [18] to coping effectiveness.

Personality has been found to influence the stress-coping process directly by restricting or assisting the use of specific coping strategies and indirectly, by influencing the type and intensity of the stressors experienced or coping effectiveness [19]. There is also evidence that personality has an influence on athletic performance. For example, Piedmont, Hill, and Blanco [20] showed that 23% of variance of coaches rating of female soccer players was predicted by their neuroticism and conscientiousness. Also, conscientiousness explained 8% of the variance in game statistics. Crust and Clough [21] showed that higher levels of mental toughness were associated with better performance on a physical endurance task. However, to our knowledge no study has examined how personality influences coping during actual sporting performance. A personality type which has received significant attention in the health domain is the distressed or Type D personality. Individuals with Type D personality experience a variety of negative emotions (NA; negative affectivity) and exhibit an inability to express emotions and/or behaviors in social interactions (SI; social inhibition) [6]. It is the interaction or synergetic effects of NA and SI which characterizes the Type D personality. Individuals high on NA are more likely to experience across time and situation negative emotions, depressed mood, anxiety, hostility, irritability, a negative self-view and an attentional bias towards adverse stimuli. Individuals high on SI feel inhibited, tense, have fewer personal ties, and feel uncomfortable and insecure in encounters with other people [22].

Type D personality represents a relatively homogeneous subgroup of personality traits grounded in psychological theory [23]. The subcomponents NA and SI are well represented within the five-factor framework of personality (supporting its construct validity) sharing less than 50% of the variance with neuroticism and extraversion [6]. Type D personality has incremental value above and beyond the five factor trait model [24] which suggests that Type D personality is a distinct construct. Individuals with Type D personality (those who score > 10 on both NA and SI on the DS14) appear to experience higher levels of chronic stress, emotional and social difficulties and adverse health events. Type D personality is closely associated with symptoms of depression and anxiety, chronic tension, pessimism, lack of perceived social support, lower subjective well-being and self-esteem, dissatisfaction with life, lower quality of life, and poor body image. It is also a predictor of adverse clinical events and poor coronary heart disease prognosis [22, 25–27].

In non-clinical populations, Type D personality is related to major psychosocial stressors [28], greater cardiovascular and neuroendocrine reactivity [29], increased cardiac output in response to acute laboratory stressors [30], and a hemodynamic maladaptation to an acute mental arithmetic stressor in women [31]. Type D individuals are also more likely to appraise stressful events as a threat and higher levels of perceived stress [32] and use more passive, maladaptive avoidance coping strategies, which are associated with burnout, decreased self-reported health and morale [33–35]. There is some evidence suggesting that when experiencing an acute laboratory stressor Type D individuals show greater stress reactivity [29]. However, studies have not yet examined the effects on actual (sport) performance.

Type D personality has been shown to be associated with detrimental health behaviors (e.g., more smoking) and avoidance of facilitative health behaviors (e.g., physical activity), adverse health outcomes as well as poor coping with stress [36–38]. However, to date no study has examined how Type D personality manifest itself in the realm of sport. This is an important research endeavor since sport participation has the potential to facilitate emotional regulation, which is an important need for individuals with Type D personality. Hence, being physically active has been associated with improved affective states [39]. In addition, understanding the relationship between Type D personality and stress and coping in sport can help inform effective applied stress management interventions. As such, the aim of study one was to examine the relationship between Type D personality, stress intensity appraisal of a self-selected stressor, coping, and perceived coping effectiveness in a sport context. We predicted that athletes classified as Type D would perceive the self-selected sport stressors with higher levels of intensity and would mainly use avoidance coping and emotion-focused strategies, and report lower levels of problem-focused coping. No a priori prediction was made with regard to coping effectiveness. The aim of study two was to examine the role of Type D on performance on a novel sport task as well as pre-performance anxiety levels, coping, and post-performance affect. For study two we predicted that Type D personality would be associated with increased levels of anxiety, lower levels of self-confidence prior to completing a rugby league novel task, poorer performance, and more negative emotions post-performance.

Study1

Materials and methods

Participants. Participants were 482 British, mainly Caucasian (93.8%) athletes (male *n* = 305; female *n* = 177) aged between 16 to 45 years (*M* age = 20.44 years, *SD* = 3.98), with experience in their sport from 1 to 35 years (*M* = 9.63, *SD* = 4.69). Athletes were recruited from different Universities (68%) and sports clubs in the region of Yorkshire, UK. See also Table 1 for additional demographic information. The study was approved by a University's

Research Ethics Committee and participants provided written informed consent prior to participating.

Instruments. First, the participants completed a demographic section. This included questions on their gender, age, sport (i.e., team/individual; contact/non-contact), years they participated in their sport, skill level (i.e., University/club, county, national or international).

The Type D scale 14 (DS14) provides a taxonomic and continuous assessments of distressed personality by measuring the traits of negative affectivity (NA, 7-items; e.g., "I often feel unhappy") and social inhibition (SI, 7-items; e.g., "I am a closed kind of person"). Participants respond using a 5-point Likert scale anchored at 0 = false to 4 = true. Based on cluster analysis and item response theory, a score of 10 or more on both subscales indicates the likelihood of a respondent fitting the Type D personality profile [40]. In addition to its taxonomic use the NA and SI subscales can be multi-plicatively combined to produce a continuous measure of distressed personality, with higher scores corresponding to a greater likelihood that the respondent behaves in a manner consistent with the distressed personality type. This dimensional interpretation of distressed personality is supported by psychometric evaluations of the DS14, which have indicated that item responses are additive at the subscale and total scale levels [41, 42]. The DS14 total score ($\alpha = .88$) and subscale scores (NA, $\alpha = .86$; SI, $\alpha = .80$) were internally consistent in the present sample, in accordance with previous studies [6].

The Modified Cope Inventory (MCOPE) [43] has 12 coping scales each consisting of four items which can be classified under 3 higher order dimensions (1) Problem focused coping: Active coping, seeking informational social support, planning, suppression of competing activities, increasing effort; (2) Emotion focused coping: Seeking emotional social support, humor, venting emotions, self-blame, wishful thinking; (3) Avoidance coping: Behavioral disengagement, denial. There is extensive evidence supporting the convergent validity of the MCOPE subscales and the MCOPE subscales have demonstrated acceptable levels of internal consistency in previous research [43, 44]. To measure coping effectiveness a 5-point Likert scale was added to the MCOPE which was anchored by 1 = extremely ineffective and 5 = extremely effective [17]. The three higher order dimensions achieved acceptable levels of reliability (Cronbach alpha between .76 and .89 for coping and coping effectiveness).

Prior to completing the MCOPE, participants reported the most intense sport stressor experienced in the last 14 days and appraised the stressor in terms of how much stress the event caused by dissecting a 10 cm bipolar line anchored by ‘not at all stressful’ vs. ‘extremely stressful’. The ‘stress thermometer’ has already demonstrated normal distribution properties and adequate variability for male and female athletes [45]. The MCOPE was completed in relation to the self-reported stressor.

Procedure. Athletes and coaches of sports teams received letters detailing the nature of the study and participant requirements. If the coaches granted permission for the data collection, an information letter and consent form was distributed to athletes. Research assistants, who had received training in quantitative data collection techniques, administered the paper based questionnaire pack in the same order prior to or following training sessions. All of the participants were actively involved in competitive sport and had participated competitively within 14 days of the questionnaires being administered.

Analysis strategy. The statistical analysis was conducted with SPSS version 22. After screening for outliers and normality, Cronbach alphas and descriptive statistics for all study variables were obtained. The influence of Type D personality on stress and coping was examined using both the taxonomic and continuous approach. For the taxonomic approach we first conducted a test to examine differences in stressor intensity as a function of Type D personality. Following this we conducted multivariate analysis of variance (MANOVA) to test differences in coping and coping effectiveness at the dimensional level. Type D was entered as a

Table 1. Sample characteristics of the athlete participants in study 1 and difference between Type D and non-Type D athletes.

Variable	N	Total Sample	N	Non-Type D	N	Type D	*p*
Age, years (M, SD)	482	20.4 (4.0)	342	20.4 (4.1)	140	20.5 (3.6)	.81
Males (N, %)		305 (63.3)		221 (64.6)		84 (60.0)	.35
Years in sport (M, SD)	482	9.6 (4.7)	342	9.9 (4.8)	140	8.9 (4.6)	.02
Skill level (N, %)	482		342		140		.007
No response		12 (2.5)		7 (2.0)		5 (3.6)	
University/club		175 (36.3)		110 (32.2)		65 (46.4)	
County		220 (45.6)		174 (50.9)		46 (32.9)	
National		60 (12.4)		40 (11.7)		20 (14.3)	
International		15 (3.1)		11 (3.2)		4 (2.9)	
Classification (N, %)	481		341		140		.19
Contact		267 (55.4)		196 (57.5)		71 (50.7)	
Non-contact		214 (44.5)		145 (42.5)		69 (49.3)	
Pursuit (N, %)	481		341		140		.20
Team		323 (67.0)		235 (68.9)		88 (62.9)	
Individuals		158 (33.0)		106 (31.1)		52 (37.1)	
Stress Intensity	342	6.2 (2.3)	342	6.0 (2.4)	140	6.7 (2.1)	.002

fixed factor for each analysis, and the coping dimensions were entered as the dependent variable in each analysis. MANOVA main effects were further analyzed with Bonferonni adjusted univariate analysis of variance (ANOVA) to test between group differences (Type D vs. non-Type D athletes).

We conducted regression analysis to examine the role of Type D personality as a continuous variable. Stress, and coping and coping effectiveness at the dimensional level were the dependent variables and NA, SI entered at step one and centered NA X SI at step two, the predictor variables.

Results study 1

The coping dimensions were multivariate normal when regressed on Type D. Box's M and Levene's test were not significant for the variables. Table 1 shows the distribution of Type D personality among the athlete participants in the present study based on demographic variables. Type D athletes reported fewer years of participation in sport than non Type D athletes (t (480) = -2.17, P = .03; *d* = -0.20) and were comparatively less likely to compete at a regional or county level and more likely to compete at a University or local level (χ^2 (4, 482) = 14.14, P = .02; Phi = .17). Type D groups were comparable with respect to age, gender, sport classification (contact vs. non-contact) and pursuit (team vs. individual sport).

Taxonomic analysis. The t-test for stressor intensity was significant (t = -3.07, P = .002; Cohen *d* = 0.32) with a small effect size and Type D individuals reporting higher levels of stress intensity.

Table 2 shows the mean standard deviations, and results, for coping and coping effectiveness at the dimensional level for the Type D and non-Type D athletes. The MANOVA for coping (Wilks' λ = 0.94; $p < .001$, η2p = .06) and coping effectiveness (Wilks' λ = 0.94; $p < .001$, η^2p = .06) were both significant with small effect sizes. As predicted, Type D athletes reported less frequent use of problem- and emotion-focused coping but more avoidance coping. Type D athletes also reported lower coping effectiveness for problem-focused coping but higher coping effectiveness for avoidance coping.

Table 2. Mean and standard deviations for each of the coping dimensions and coping effectiveness (CE) for the Type D and non-Type D participants and results of the analysis of variance (significant results only).

	Non-Type D		Type D		Results Coping			Results CE		
	Coping	CE	Coping	CE	F(1,480)	*p*	Eta	F(1,480)	*p*	Eta
Problem Focused Coping*$	3.30 (.50)	3.02 (.44)	3.16 (.55)	2.83 (.46)	7.38	.02	.02	16.40	< .001	.03
Emotion Focused Coping*	2.55 (.64)	2.41 (.51)	2.73 (.69)	2.43 (.55)	6.97	.009	.02		n.s	
Avoidance Coping*$	1.85 (.63)	2.13 (.79)	2.10 (.73)	2.28 (.74)	13.46	< .001	.03	4.00	.046	.01

* Denotes significant difference for coping between the Type D and non-Type D group.

$ Denotes significant difference for coping effectiveness between the Type D and non-Type D group.

Continuous analysis. The regression analysis for stress intensity was significant at step one ($F (2,479) = 11.51$; $P < .001$) but not step two with NA being a significant predictor (Beta = .199; $P < .001$). Table 3 provides the results of the regression analysis for Type D personality as a continuous variable. As predicted, for emotion-focused and avoidance coping the interaction between NA and SI added small but significant variance to the model. However, no such effect was found in problem-focused coping.

Discussion study 1

As predicted, the results of study one suggest that Type D personality influences the stress-coping process in sport. Athletes with Type D reported higher levels of stress independent of stressor type and made use of more emotion-focused and avoidance coping strategies and less use of problem-focused coping strategies. In addition, Type D athletes reported the emotion-focused strategies as more effective, and the problem-focused coping strategies as less effective.

As predicted, and supporting previous work in organizational settings [46] and with university students [33], athletes classified as Type D reported higher levels of stress intensity independent of stressor type. The continuous analysis suggests this was mainly due to higher levels of NA. These results are not dissimilar to a study by Schoormans, Husson, Denollet and Mols [47] in a sample of cancer survivors and provide some support for the notion that competing in sport is more stressful for Type D than non-Type D athletes. On the other hand, sport participation can facilitate emotional regulation of individuals with Type D. Regular and acute bouts of exercise have been shown to have positive effects on affective states [39]. It might be that sport participation moderates the experience of stress, making the differences between Type D and non-Type D individuals less pronounced in such environments. This issue would warrant further investigation because it provides a potential mechanism for intervention to help individuals with Type D to reduce their stress levels.

Table 3. Results of the regression analysis for Type D as a continuous variable (continuous analysis).

	Coping		Coping Effectiveness
	Step 1 R^2 and Beta predictors	Step 2 ΔR^2	Step 1 R^2 and Beta predictors
Problem Focused Coping	5.7%**; NA = -.141**; SI = -.131**	n.s.	5.7%**; NA = -.141**; SI = -.131*
Emotion Focused Coping	6.4%**; NA = .294**; SI = .122*	1.4%**; NAxSI = .389**	n.s.
Avoidance Coping	7.8%**; NA = .301**	1%*; NAxSI = -.328*	n.s.

*$P < .05$

**$P < .01$

NA = Negative affectivity; SI = Social inhibition

Both the taxonomic and continuous analysis supported the predictions that Type D was directly associated with significantly more use of emotion-focused and avoidance coping whereas the taxonomic analysis also indicated less use of problem-focused coping [33, 35]. The finding that Type D is associated with more use of emotion-focused coping might be due to the fact that participants experienced more intense emotions because of higher levels of stress. This is supported by recent findings that Type D is associated with lower HRV which in turn is associated with greater social threat and anxiety [48]. A priority for Type D athletes would be to down regulate such emotions. Lower resting HRV as found by Jandackova et al. [48] in Type D individuals is associated with reduced context appropriate emotional responses [49]. However, there was no difference in coping effectiveness for emotion-focused coping suggesting that Type D athletes might have developed strategies to deal with increased emotionality and stress. The physiological and psychological responses to being physically active might influence the experience of stress and emotions as well as the coping responses. However, this would require further study.

Avoidance coping is likely to be used in situations in which the athlete experiences limited power to change the outcome of a situation [50]. Such coping strategies can be adaptive if the athlete is unable to achieve their goal regardless of investment of effort [51, 52]. Avoidance coping is not an uncommon strategy used by athletes. Nicholls, Jones, Polman, and Borkoles [53] for example found that professional rugby union athletes used blocking as one of the most frequent coping strategies during training and competition. Although providing temporary relief from an acute stressor, in the long-term avoidance coping is maladaptive and can result in adverse psychological and physiological outcomes including higher stress levels [50, 54].

Most studies report that athletes make a greater use of problem-focused coping in comparison to emotion-focused or avoidance coping [1]. Problem-focused coping involves active problem solving and is regarded more adaptive in nature. For example, problem-focused coping is associated with positive affect [55] and subjective performance ratings [3]. The taxonomic analysis showed that individuals classified as Type D used less problem-focused coping and rated this form of coping as less effective.

The present findings support the cognitive-motivational-relational model of stress and coping [7], which predicts that variables, including personality, influence how frequently coping strategies are used and how effective they are perceived to be. The study's observational design precludes inferences about the causal links between personality, stress and coping. For example, it is possible that stressor intensity confounds the association between Type D personality and coping. However, the pattern of results may prove informative for future research.

The present results are also consistent with a functional account of behavior [56, 57]. In particular, the tendency of Type D individuals to perceive stressors as more intense in conjunction with their inclination to use avoidance and emotion-focused coping, and their disinclination to use problem-focused coping. A functional viewpoint conceives of coping as behavior that is based on predictions of one's own future behavior, and that seeks to maximize long-term rewards [56].

Distressed personality implies a susceptibility to acute stress, which may engender reliance on coping strategies that have utility in situations of acute stress. The utility of avoidance strategies becomes evident if one considers that the negative affect associated with avoidance coping may serve as a substitute for emotional states that are relatively more damaging to performance and self-esteem, such as shame associated with the prospect of failure. This interpretation implies that Type D personality, through the preferential use of avoidance and emotion-focused coping, are primarily motivated to manage aversive emotional states induced by stressors. They may correspondingly be less inclined (or less able) to cope by responding to aspects of the stressor that exist apart from its emotional impact, as is characteristic of problem-focused coping. The latter coping process implies a sensitivity to long-term contingencies that reach into the past and extend into one's future or imagined prospects. By contrast, so-called maladaptive coping strategies may predominantly arise from short-term contingencies whose immediate consequences are paramount [56]. This interpretation predicts that individuals responding to acute stress will be inclined to use coping strategies that have short-term utility because such strategies ostensibly forestall aversive emotional states in the short term. However, habitual use of avoidance coping, for instance, is likely to be a strategy of diminishing returns [56, 58]. Over the long run, avoidance coping may consequently prove detrimental to athletic performance. Study two investigates the effect of Type D personality and affective states on objective measures of athletic performance of athletes.

Study 2

Material and methods

Participants. Thirty-two healthy Caucasian British male university student athletes (M age = 20.5 years; SD = 1.0 years) took part in study two. Participants reported M = 11 (SD = 4.2) years of sporting experience, played M = 4.2 (SD = 3.1) competitive games per month, trained M = 2.38 (SD = 1.12) times per week with session of M = 78.9 (SD = 33.8) minutes' duration. The effort exerted was rated M = 7.7 (SD = 1.7) on the ten point Borg scale. The main sport of 27 participants was soccer, followed by cricket (n = 1), athletics (n = 1), fitness (n = 2) and tennis (n = 1). The study was approved by a University ethics committee. In addition, all participants provided written informed consent prior to their involvement in the study.

Instruments. Participants first completed a demographic section consisting of questions regarding age, ethnicity, sport played, years of participation, time spent training, duration and perceived effort in training and frequency of competition. Following this they completed the DS14.

Pre-performance state anxiety and self-confidence was assessed with the Competitive State-Anxiety Inventory-2 Revised [59]. The CSAI-2R consist of 7-items measuring somatic anxiety, 5-items cognitive anxiety and 5-items self-confidence and is scored on a 4-point Likert scale (1 = not at all to 4 = very much). The CSAI-2R has been shown to have a good factorial structure and internal consistency [59]. In the present study internal consistency was good for the somatic anxiety ($\alpha = .82$) and self-confidence ($\alpha = .80$) scales and adequate for the cognitive anxiety scale ($\alpha = .67$).

Post-performance affect was examined using the Positive and Negative Affect Scale (PANAS) [60]. The PANAS is scored on a 5-point Likert scale (1 = not at all to 5 = extremely) with 10 items measuring positive affect and 10 items measuring negative affect. Higher scores on each subscale indicate a greater presence of affect it measures. Both scales have been shown to be uncorrelated and stable over a 2 month period, with high internal consistency [60]. Internal consistency in the current study was acceptable (Cronbach's $\alpha = .87$ and .70 for PA, NA respectively). For both the CSAI-2 and PANAS the participants were asked to complete the instrument with how they felt 'right now'.

Because of coping being assessed immediately following performance and the inclusion of other constructs this study made use of the much shorter Brief Approach/ Avoidance Coping Questionnaire (BACQ) [61] to reduce the burden on participants. The BACQ is a 12-item construct, measuring three distinct coping dimensions. Six-items assess general approach coping, 3-items cognitive avoidance coping and 3-items measure resignation/ withdrawal. The scale is anchored by a 5-point Likert scale (1 = disagree completely to 5 = agree completely). Previous research has shown the BACQ to be internally consistent [61]. Weak internal consistency was shown for the cognitive avoidance (α = .55) and resignation/ withdrawal (α = .66) scales and good reliability for the approach coping scale (α = .83). Considering the low number of scale items [62], item total correlations [63] and arguments set out in study one all subscales were used for statistical analysis.

Procedure. Participants were recruited from one University in England through advertising and referrals. On the first visit to the laboratory participants were informed about the nature of the study before providing consent. Following this a questionnaire pack was completed. On the second visit the participants were made aware of previous performance charts regarding distance and success on the task to be completed. They were then provided with standard instructions regarding the aims and objectives of the task before completing the CSAI-2R.

The experimental task consisted of a rugby league circuit and was new to all participants. The task was extensively piloted to test its suitability and how effectively it induced stress. The task consisted of participants running as many circuits in 3 minutes in a sports hall. From a start line, 10 m were run before returning to the line. Once turned, a further 20 m were run, again returning to the start. The individual turned and ran 30 m, before progressing 5 m to a cone station located on the left. Side-stepping around 6 cones placed 1 m apart and 0.5 m away from the center, in an alternating fashion, followed by 5 m of running. On the first passing station, 4 rugby balls were passed from left to right, at a target 8 m away, 1.50 m parallel to the floor and 0.5 m in diameter. A 5 m length ladder was located 5 m from the first passing station, requiring one foot placed in each section of the ladder. Travelling a further 5 meters, a second passing station was positioned. Four rugby balls were passed from right to left, at a target 8 meters away, 1.50 m parallel to the floor and 0.5 m in diameter. The individual ran 5 meters back to the start and repeated until 3 minutes expired. Running one circuit lasted approximately 50 seconds, and was 130 meters in distance. The outcome measures for the performance task were distance travelled in 3 minutes and the number of errors made (targets missed).

To increase anxiety participants were provided with verbal prompts during the circuit: 'you are falling behind, what are you going to do about it (30 sec); 'you are 5 seconds adrift, keep working (90 sec), and 'last 30 seconds, put everything into it' (150 sec). Also, £50 was offered to the participant who performed the best on the circuit as assessed by a Great Britain rugby league player. To this end a video camera was used to film each performance. The researcher filmed performance, with two assistants as ball collectors. Shuttle run and ladder placement made achievement of success on the passing stations more difficult. Following this the participants completed the PANAS and BACQ which was followed by a debriefing.

Data analysis. Descriptive statistics for all dependent variables were calculated. Independent sample t-test were conducted to examine differences between those classified having Type D personality (NA and SI > 10) and non-Type D participants. Minimum group sizes of 12 participants were required to achieve an 80% chance of detecting a large effect (Cohen's $d > 1$) at a two-tailed α level of .05 [64]. We did not use regression analysis in this study because of lack of power due to its small sample size. Statistically significant contrasts are reported with bias-corrected (Hedges') Cohen's *d* effect sizes with 95% confidence intervals.

Results study 2

Ten participants were classified as Type D and 22 as non-Type D. The two groups did not differ on the demographic variables except duration of training. Type D participants (M = 57.58) trained for a considerably smaller length of time per session than non-Type D (M = 88.64) participants (z = 2.42, P = .01).

Table 4. Mean and standard deviations and statistical result for the dependent variables of study 2.

	Non-Type D (n = 22)		Type D (n = 10)				CIs difference	
	M (*SD*)		*M* (*SD*)		*t30*	*p*	Lower	Upper
CSAI						ns		
Somatic Anxiety	17.73	5.82	20.71	6.43		ns		
Cognitive Anxiety	17.45	5.42	20.40	4.97		ns		
Self-Confidence	29.45	5.79	22.00	6.60	-3.07	.003	-12.2	-2.7
PANAS								
Positive Affect	32.05	6.35	35.00	4.98		ns		
Negative Affect	12.86	2.23	14.50	4.14		ns		
BACQ								
Approach	21.18	4.01	20.50	5.48	-	ns		
Avoidance	9.27	2.14	9.10	2.02	-	ns		
Resignation/withdrawal	5.41	2.13	7.80	2.34	2.85	.008	0.68	4.10
Performance								
Distance (m)	372	19.9	320	25.7	-6.23	$< .001$	-69	-35
Targets missed	9.5	2.5	15.6	2.7	6.24	$< .001$	4.1	8.1

Table 4 provides the means and standard deviations for the two groups for the dependent variables as well as the results of the statistical analysis. Pre-performance the non- Type D athletes reported significantly higher scores for self-confidence than the Type D participants with a large effect size (t30 = -3.07; P = .003; *d* = 1.20; 95% CI 0.40 to 2.00). Although the Type D participants reported higher somatic and cognitive anxiety these were not significantly different from the non-Type D participants.

During performance of the 3 minute rugby league circuit task the Type D participants covered significantly less distance (t(30) = -6.23; P $<$ .001; *d* = -3.04; 95% CI -4.00 to -1.99) and made more errors (t(30) = 6.24; P $<$ .001; *d* = 2.38; 95% CI 1.38 to 3.26) than the non- Type D participants. The effect sizes were large for both distance travelled and errors made.

Post-performance the Type D and non-Type D participants did not report any differences in either positive or negative affect. However, Type D participants reported more use of resignation/withdrawal coping (t(30) = 2.85; P = .008; *d* = 1.06; 95% CI 0.27 to 1.85) but not approach or avoidance coping strategies.

Discussion study 2

Type D individuals in study two reported lower levels of self-confidence prior to completing the rugby league circuit task. In addition, they performed poorer as indexed by less distance completed and more mistakes made. This was associated with more frequent use of resignation/withdrawal coping. No differences in post-performance affect were observed.

Previous studies have indicated that Type D individuals report increased levels of stress and show higher levels of stress reactivity in laboratory tasks [30, 31]. This is the first study to indicate that Type D personality is also associated with decreased task performance. In addition, poorer performance was associated with lower self-confidence prior to the task and increased use of resignation/withdrawal coping after completion of the task. This is supported by findings that Type D individuals have a tendency to search the environment for trouble [65] and report lower levels of general self-efficacy [37]. The lower levels of self-efficacy in the study by Wiencierz and Williams [37] were associated with significant less total exercise participation and walking in Type D versus non Type D individuals. Our findings also support meta-analytic findings in sport [66] which have shown that self-confidence in particular (r = .24) has an influence on performance whereas cognitive anxiety (r = -.10) has a minimal influence.

Few studies have examined the association between coping and objective measures of performance. Our results suggest that the increased use of resignation/withdrawal coping by the Type D participants is associated with poorer performance [3]. We deliberately used a shorter coping questionnaire for this study to reduce demands on the participants. However, in line with study one it appears that Type D is associated with the use of more maladaptive coping strategies. More importantly, this is the first study to demonstrate that Type D personality negatively influences actual performance in sport rather than self-rated performance satisfaction [3].

No difference was found in post-performance affect between the Type D and non- Type D participants. This finding provides partial support for the observation from study one in that sport might provide an outlet for Type D individuals to regulate their emotions. This could be partly due to the fact that exercising, albeit at moderate intensity, is associated with positive affect [39]. An important limitation of the current study is the generalizability of the rugby task to real sporting competitions. Although the task selected has many similarities to real rugby competition behavior the environment in which it was conducted was very different. The task however allowed for the quantification of completion time and error. Future research could identify and examine behaviors which reflect this in real sporting competitions.

Overall discussion

This is the first study to examine the role of Type D personality, stress, coping, and performance in the context of sport. There is ample evidence that Type D personality has a negative impact on mortality and morbidity in CHD patients [26] and the general population [25]. The present study indicates that Type D personality has consequences for dealing with stress during competition and negatively affects performance in sport.

From a theoretical perspective, this study adds to our understanding of factors which influence the stress and coping process as well as athletic performance. The findings indicate that Type D personality influences actual sport performance and that the use of coping strategies that are maladaptive in the long run by Type D athletes may explain this finding.

In sport [67], health and medicine [68] individual or personalized approaches have been advocated. Although Type D is characterized by high levels of NA and SI the psychosocial, health and behavioral consequences are likely to vary across populations and settings. Based on our findings interventions for Type D athletes could focus on mood modification and enhancing interpersonal functioning [69]. With regard to mood modification it has been shown that a mindfulness-based stress reduction training program reduced NA and SI in the general population. Increased proficiency in mindfulness was found to be the main mechanism for this reduction [70]. Also, research in the sport context [71] suggests that higher levels of mindfulness are associated with reduced stress and improved coping effectiveness. Hence, future investigation are recommended to test the usefulness of mindfulness practice among Type D athletes. In addition, relaxation training (e.g., autogenic training, progressive muscle relaxation, relaxation imagery) or psychosocial coping interventions might be beneficial to Type D athletes. Psychosocial interventions could in particular target the appraisal process (perceive a stressful event as a challenge rather than a threat) through cognitive restructuring, development of emotion-focused skills to down regulate their emotional state (e.g., breathing), developing their problem-focused coping repertoire (e.g., planning, increasing effort) whilst reducing maladaptive avoidance coping strategies. Interpersonal functioning of Type D individuals could be improved through assertiveness training. This would allow Type D athletes to use more active and adaptive coping strategies in stressful encounters rather than using passive maladaptive coping strategies.

Our studies are not without limitations. Study one had a cross-sectional design which cannot support causal inferences. Also, athletes self-reported retrospectively on the stress experienced and coping used to deal with a stressful event. This can result in bias and forgetfulness. The sample used in study one was also heterogeneous in nature. In study two only males participated, as such it is unclear whether findings can be generalized to females. In addition, future research has to establish whether the task selected generalizes to behaviors in actual sporting competitions.

In conclusion, our novel findings suggest that Type D personality is associated with the use coping strategies which are considered to be maladaptive in the context of sport. In addition, Type D personality is associated with poorer performance on a sporting task which was associated with lower levels of self-confidence and more use of resignation/withdrawal coping.

Supporting information

S1 Data set. This is the dataset for study 1.
(SAV)

S2 Data set. This is the dataset for study 2.
(SAV)

Acknowledgments

We would like to thank Nick Reynolds for his assistance with this project.

Author Contributions

Conceptualization: Erika Borkoles, Remco C. J. Polman.

Data curation: Mariana Kaiseler, Andrew Evans, Remco C. J. Polman.

Formal analysis: Remco C. J. Polman.

Methodology: Erika Borkoles, Mariana Kaiseler, Andrew Evans, David R. Thompson, Remco C. J. Polman.

Project administration: Erika Borkoles.

Supervision: Erika Borkoles.

Writing – original draft: Erika Borkoles, Remco C. J. Polman.

Writing – review & editing: Erika Borkoles, Mariana Kaiseler, Andrew Evans, Chantal F. Ski, David R. Thompson, Remco C. J. Polman.

References

1. Polman RCJ. Elite athletes' experiences of coping with stress. In Thatcher J, Jones M, Lavallee D, eds. *Coping and emotion in sport*. London, UK: Routledge; 2012:284–301.
2. Doron J, Gaudreau P. A point-by-point analysis of performance in a fencing match: Psychological processes associated with winning and losing streaks. *J Sport Exerc Psychol*, 2014; 36:3–13. https://doi.org/10.1123/jsep.2013-0043 PMID: 24501140
3. Nicholls AR, Polman RCJ, Levy AR. (2012). A path analysis of stress appraisals, emotions, coping, and performance satisfaction among players. *Psychol Sport Exerc*. 2012; 13:263–270.
4. Kaiseler M, Polman RCJ, Nicholls AR. Effect of the Big Five personality dimensions on appraisal, coping, and coping effectiveness in sport. *Eur J Sport Sci*. 2012; 12: 62–72.

5. Kaiseler M, Levy A, Nicholls A. Madigan D. The independent and interactive effects of the Big Five personality dimensions upon dispositional coping and coping effectiveness in sport. *Int J Sport Exerc Psychol. in press* https://doi.org/10.1080/1612197X.2017.1362459
6. Denollet J. DS14: Standard assessment of negative affectivity, social inhibition, and Type D personality. *Psychosom Med*. 2005; 67:89–97. https://doi.org/10.1097/01.psy.0000149256.81953.49 PMID: 15673629
7. Lazarus RS. *Stress and emotion*: *A new synthesis*. New York, USA: Springer; 1999.
8. Lazarus RS, Folkman S. *Stress, appraisal and coping*. New York, USA: Springer; 1984.
9. Tamres LK, Janicki D, Helgeson VS. Sex differences in coping behaviour: A meta-analytic review and an examination of relative coping. *Pers Soc Psychol Rev*. 2002; 6:2–30.
10. Zakowski SG, Hall MH, Klein LC, Baum A. Appraised control, coping and stress in a community sample: A test of the goodness-of-fit hypothesis. *Ann Behav Med*. 2001; 23:158–165. https://doi.org/10.1207/S15324796ABM2303_3 PMID: 11495216
11. Nicholls AR, Polman RCJ. Coping in sport: A systematic review. *J Sports Sci*, 2007; 25:11–31. https://doi.org/10.1080/02640410600630654 PMID: 17127578
12. Krohne HW. Vigilance and cognitive avoidance as concepts in coping research. In Krohne HW, ed, *Attention and avoidance*: *Strategies in coping with averseness*. Seattle, WA: Hogrefe & Huber; 1993:19–50.
13. Endler NS, Parker JDA. Assessment of multidimensional coping: task, emotion, and avoidance strategies. *Psychol Assess*. 1994: 6;50–60.
14. Gaudreau P, Blondin J-P. Development of a questionnaire for the assessment of coping strategies employed by athletes in competitive sport settings. *Psychol Sport Exerc*. 2002; 3:1–34.
15. Nicholls AR. Effective versus ineffective coping. In Nicholls AR, ed, *Coping in sport: Theory, methods, and related constructs*. New York, USA: Nova Science; 2010:263–276.
16. Folkman S. Making the case for coping. In Carpenter BN, ed, *Personal coping*: *Theory*, *research and applications*. Westport, CT: Praeger; 1992:31–46.
17. Nicholls AR, Polman RCJ. Performance related stressors, coping, and coping effectiveness among international adolescent Rugby Union Football players: A 31-day diary study. *J Sport Behav*. 2007; 30:19–218.
18. Eubank M, Collins D. Coping with pre- and in-event fluctuations in competitive state anxiety: A longitudinal approach. *J Sports Sci*. 2000; 18:121–131. https://doi.org/10.1080/026404100365199 PMID: 10718568
19. Bolger N, Zuckerman A. A framework for studying personality in the stress process. *J Pers Soc Psychol*. 1995; 69:890–902. PMID: 7473036
20. Piedmont RL, Hill DC, Blanco S. Predicting athletic performance using the five- factor model of personality. *Pers Individ Differ*. 1999; 27:769–777.
21. Crust L, Clough PJ. Relationship between mental toughness and physical endurance. *Percept Motor Skills*. 2005; 100:192–194. https://doi.org/10.2466/pms.100.1.192-194 PMID: 15773710
22. Mols F, Denollet J. Type D personality in the general population: A systematic review of health status, mechanisms of disease, and work related problems. *Health Qual Life Outcomes*. 2010; 8:9. https://doi.org/10.1186/1477-7525-8-9 PMID: 20096129
23. Chapman BP, Duberstein PR, Lyness JM. The distressed personality type: Replicability and general health associations. *Eur J Pers*. 2007; 21:911–929.
24. Howard S, Hughes BM. Construct, concurrent, and discriminant validity of Type D personality in the general population: Associations with anxiety, depression, stress, and cardiac output. *Psychol Health*. 2012; 27:242–258. https://doi.org/10.1080/08870446.2011.603423 PMID: 21809947
25. Mols F, Denollet J. Type D personality among non-cardiovascular patient populations: A systematic review. *Gen Hosp Psychiatry*, 2010; 32:66–72. https://doi.org/10.1016/j.genhosppsych.2009.09.010 PMID: 20114130
26. O'Dell KR, Masters KS, Spielmans G, Maistro S. Does Type D personality predict outcomes among patients with cardiovascular disease? A meta-analytic review. *J Psychosom Res*. 2011; 71:199–206. https://doi.org/10.1016/j.jpsychores.2011.01.009 PMID: 21911096
27. Borkoles E, Polman RCJ, Levy A. Type D personality and body image in men: The role of exercise status. *Body Image*. 2010; 7:39–45. https://doi.org/10.1016/j.bodyim.2009.10.005 PMID: 19945926
28. Michal M, Wiltink J, Grande G, Beutel ME, Brahler E. Type D personality is independently associated with major psychosocial stressors and increased health care utilization in the general population. *J Affect Disord*. 2011; 134:396–403. https://doi.org/10.1016/j.jad.2011.05.033 PMID: 21663973

29. Habra ME, Linden W, Anderson JC, Weinberg J. Type D personality is related to cardiovascular and neuroendocrine reactivity to acute stress. *J Psychosom Res*. 2003; 55:235–245. PMID: 12932797

30. Williams L, O'Carroll RE, O'Connor RC. Type D personality and cardiac output in response to stress. *Psychol Health*. 2009; 24:489–500. https://doi.org/10.1080/08870440701885616 PMID: 20205007

31. Howard S, Hughes BM, James JE. Type D personality and hemodynamic reactivity to laboratory stress in women. *Int J Psychophysiol*. 2011; 80:96–102. https://doi.org/10.1016/j.ijpsycho.2011.02.006 PMID: 21333697

32. Grynberg D, Gidron Y, Denollet J, Luminet O. Evidence for a cognitive bias of interpretation toward threat in individuals with a Type D personality. *J Behav Med*. 2012; 35:95–102. https://doi.org/10.1007/s10865-011-9351-7 PMID: 21553240

33. Polman RCJ, Borkoles E, Nicholls A. Type D personality, stress and symptoms of burnout: The influence of avoidance coping and social support. *Br J Health Psychol*. 2010; 15:681–696. https://doi.org/10.1348/135910709X479069 PMID: 19930789

34. Yu X-N, Chen Z, Zhang J, Liu X. Coping mediates the association between Type D personality and perceived health in Chinese patients with coronary heart disease. *Int J Behav Med*. 2011; 18:277–284. https://doi.org/10.1007/s12529-010-9120-y PMID: 20941651

35. Williams L, Wingate A. Type D personality, physical symptoms and subjective stress: The mediating effects of coping and social support. *Psychol Health*. 2010; 27:1075–1085.

36. Pederson SS, Denollet J. Is Type D personality here to stay? Emerging evidence across cardiovascular patient groups. *Curr Cardiol Rev*. 2006; 2:205–213.

37. Wiencierz S. Williams L. Type D personality and physical inactivity: The mediating effect of low self-efficacy. *J Health Psychol*. 2017; 22:1025–1034. https://doi.org/10.1177/1359105315622557 PMID: 26837688

38. Williams L, Abbott C, Kerr R. Health behaviour mediates the relationship between Type D personality and subjective health in the general population. *J Health Psychol*. 2016; 21:2148–2155. https://doi.org/10.1177/1359105315571977 PMID: 25712490

39. Biddle SJH, Mutrie N, Gorely T. *Psychology of physical activity*: *Determinants*, *well-being and interventions*. 3rd ed. London, UK: Routledge; 2015.

40. Emons WHM, Meijer RR, Denollet J. Negative affectivity and social inhibition in cardiovascular disease: Evaluating Type D personality and its assessment using item response theory. *J Psychosom Res*. 2007; 63:27–39. https://doi.org/10.1016/j.jpsychores.2007.03.010 PMID: 17586335

41. Ferguson E, Williams L, O'Connor RC, Howard S, Hughes BM, Johnston DW et al. A taxometric analysis of type D personality. *Psychosom Med*. 2009; 71:981–986. https://doi.org/10.1097/PSY.0b013e3181bd888b PMID: 19834048

42. Straat JH, van der Ark LA, Sijtsma K. Multi-method analysis of the internal structure of the Type D scale-14 (DS14). *J Psychosom Res*. 2012; 72:285–265.

43. Crocker PRE, Graham TR. Coping with competitive athletes with performance stress: Gender differences and relationships with affect. *Sport Psychol*. 1995; 9:325–338.

44. Crocker PRE, Isaak K. Coping during competitions and training sessions: Are youth swimmers consistent? *Int J Sport Psychol*. 1997; 28:355–369.

45. Kowalski KC, Crocker PRE. Development and validation of the coping function questionnaire for adolescents in sport. *J Sport ExercPsychol*. 2002; 23:136–155.

46. Oginska-Bulik N. Occupational stress and its consequences in healthcare professionals: The role of Type D personality. *Int J Occup Med Environ Health*, 2006; 19:113–122. PMID: 17128809

47. Schoormans D, Husson O, Denollet J, Mols F. Is Type D personality a risk factor for all-cause mortality? A prospective population-based study among 2625 colorectal cancer survivors from the PROFILES registry. *J Psychosom Res*. 2017; 96:76–83. https://doi.org/10.1016/j.jpsychores.2017.03.004 PMID: 28545796

48. Jandackova VK, Koenig J, Jarczok MN, Fischer JE, Thayer J.F. Potential biological pathways linking Type D personality and poor health: A cross-sectional investigation. *PLoS ONE*, 2017; 12(4): e01766014.

49. Thayer JF, Brosschot JF. Psychosomatic and psychopathology: looking up and down from the brain. *Psychoneuroendocrinology*, 2005; 30:1050–1058. https://doi.org/10.1016/j.psyneuen.2005.04.014 PMID: 16005156

50. Carver CS, Scheier MF, Weintraub JK. Assessing coping strategies: A theoretically based approach. *J Pers Soc Psychol*. 1989; 56:267–283. PMID: 2926629

51. Wrosch C, Scheier MF, Carver CS, Schulz R. The importance of goal disengagement in adaptive self-regulation: When giving up is beneficial. *Self Identity*. 2003; 2:1–20.

52. Wrosch C, Scheier MF, Miller GE, Schulz R, Carver CS. Adaptive self-regulation of unattainable goals: Goal disengagement, goal reengagement, and subjective wellbeing. *Pers Soc Psychol Bull*. 2003; 29:1494–1508. https://doi.org/10.1177/0146167203256921 PMID: 15018681

53. Nicholls AR, Jones R, Polman RCJ, Borkoles E. Acute sport-related stressors, coping, and emotions among professional rugby union players during training and matches. *Scand J Med Sci Sports*. 2009; 19:113–120. https://doi.org/10.1111/j.1600-0838.2008.00772.x PMID: 18282223

54. Suls J, Fletcher B. The relative efficacy of avoidant and non-avoidant coping strategies: A meta-analysis. *Health Psychol*. 1985; 4:247–288.

55. Ntoumanis N, Biddle SJH. Relationship of intensity and direction of competitive anxiety with coping strategies. *Sport Psychol*. 2000; 14:369–371.

56. Rachlin H. In what sense are addicts irrational? *Drug Alcohol Depend*. 2007; 90:92–99.

57. Rachlin H. *Behavior and mind*: *The roots of modern psychology*. Oxford, UK: Oxford University Press; 1994.

58. Rachlin H. Rational thought and rational behavior: A review of bounded rationality: The adaptive toolbox. *J Exp Anal Behav*. 2003; 79:409–412.

59. Cox RH, Martens MP, & Russell WD. Measuring anxiety in athletics: The Revised Competitive State Anxiety Inventory-2. *J Sport Exerc Psychol*. 2003; 25:519–533.

60. Watson D, Clark LA, Tellegen A. Development and validation of a brief measure of positive and negative affect: The PANAS scales. *J Pers Soc Psychol*. 1988; 54:1063–1070. PMID: 3397865

61. Finset A, Steine S, Haugli L, Steen E, Laerum E. The brief approach/ avoidance coping questionnaire: development and validation. *Psychol Health Med*. 2002; 7:75–85.

62. Nunally JO. *Psychometric theory*. New York, USA: McGraw-Hill; 1978.

63. Briggs SR, Cheek JM. The role of factor analysis in the development and evaluation of personality scales. *J Pers*. 1986; 54:106–148.

64. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behav Res Methods*. 2007; 39:175–191. PMID: 17695343

65. Denollet J. Type D personality a potential risk factor refined. *J Psychosom Res*. 2000; 49:255–266. PMID: 11119782

66. Woodman T, Hardy L. The relative impact of cognitive anxiety and self-confidence upon sport performance: A meta-analysis. *J Sports Sci*. 2003; 21:443–457. https://doi.org/10.1080/0264041031000101809 PMID: 12846532

67. Hajkowicz SA, Cook H, Wilhelmseder L, Boughen N. *The future of Australian sport: Megatrends shaping the sport sector over coming decades*. A consultancy report for the Australian Sports Commission. CSIRO, Australia; 2013

68. Hamburg MA, Collins FS. The path to personalized medicine. *New Eng J Med*. 2010; 363:301–304. https://doi.org/10.1056/NEJMp1006304 PMID: 20551152

69. Pelle AJ, van den Broek KC, Denollet J. Interventions in the context of the distressed (Type D) personality. In Dornelas EA, ed, *Stress proof the heart*: *Behavioral interventions for cardiac patients*. London, UK: Springer; 2012:167–198.

70. Nyklicek I, van Beugen S, Denollet J. Effects of mindfulness-based stress reduction on distressed (Type D) personality traits: A randomized controlled trial. *J Behav Med*. 2013; 36:361–370. https://doi.org/10.1007/s10865-012-9431-3 PMID: 22585012

71. Kaiseler M, Poolton JM, Backhouse SH, Stanger N. The relationship between mindfulness and life stress in student-athletes: The mediating role of coping effectiveness and decision rumination. *Sport Psychol*. 2017; 31:288–298.

11

Acute effects of in-step and wrist weights on change of direction speed, accuracy and stroke velocity in junior tennis players

Joshua Colomar[1,2]*, Ernest Baiget[3], Francisco Corbi[4], Joshua Muñoz[1,2]

1 National Institute of Sport and Physical Education (INEFC), University of Barcelona, Barcelona, Spain, **2** Sports Science Department, Academia Sánchez-Casal, Barcelona, Spain, **3** Sport Performance Analysis Research Group (SPARG), University of Vic—Central University of Catalonia, Vic, Spain, **4** National Institute of Sport and Physical Education (INEFC), University of Lleida, Lleida, Spain

* joshuacolomar@gmail.com

Editor: Caroline Sunderland, Nottingham Trent University, UNITED KINGDOM

Abstract

The main aim of this study was to investigate the acute effects of the use of a weighting set (Powerinstep®) on measures of stroke velocity (StV), accuracy and change of direction speed (CODS) in junior tennis players. A within-subjects design was used to evaluate seventeen (6 female and 11 male) tennis players (mean ± SD; 16.5 ± 1.3 years old; 1.75 ± 8.4 m; 67.0 ± 8.1 kg; 22.04 ± 1.8 kg/m^2) on StV of three specific tennis actions (serve, forehand and backhand) and CODS for the following conditions: wearing a 50, 100, 150, 200 g weight or no weight at all (baseline). No significant differences were found between conditions for forehand (F = 0.412; p = 0.799), backhand (F = 0.269; p = 0.897) and serve (F = 0.541; p = 0.706) velocity and forehand (F = 1.688; p = 0.161), backhand (F = 0.567; p = 0.687) and serve (F = 2.382; p = 0.059) accuracy and CODS (F = 0.416; p = 0.797). Small-to-moderate effect sizes (ES) negatively affecting StV when using 200 g compared to the baseline (ES = 0.48, 0.35 and 0.45) could be observed. Moderate (ES = -0.49) and trivial (ES = -0.14 and -0.16) ES for a higher accuracy score were noticed in serve, forehand and backhand 100 g compared to the baseline. Moreover, small ES (ES = 0.41) for improvement in 200 g CODS comparing to baseline conditions were found. These results indicate that the use of a weighting set does not significantly affect StV or CODS respectively. Notwithstanding, small-to-moderate changes show impact in accuracy and no variance in velocity production when using 100 g alongside faster execution in CODS when using 200 g.

Introduction

The development of hitting or throwing velocity in overhead sports has often involved improving movement patterns, enhancement of conditioning or modifying implement such as racquets or baseballs [1]. As speed, power and stroke velocity (StV) have become determinant factors of tennis [2,3], it may become interesting to observe specific strategies to improve velocity production that practitioners can use to manage and plan new training methods. Concerning modification of implement, and focusing on tennis, customizing racquets in order to alter their weight, balance point and swing weight is an extended practice performed by players

Competing interests: The authors declare that there is no conflict of interest between the participants, the materials and equipment used, or any other procedure undertaken during the experiments and the researchers of this investigation.

and coaches [4]. This practice, in addition to other reasons, intends to use the transfer of momentum created by the mass of the racquet to hit the ball faster. In this line, heavier racquets will produce faster balls but consequently be swung slower than lighter versions [5]. Nevertheless, current literature is scarce about the effects of these variations and also offers doubts on how different customizing techniques (i.e., how the mass is distributed throughout the racquet) may affect speed [1,4] or accuracy. Moreover, intervention programs have suggested that the use of overweight implements or balls could be an effective way of improving throwing velocity in overhead sports [6,7] including tennis [8]. Although in this case tendencies have generally aimed to vary weight on the frame of the racquet, no investigations are available on how StV may be affected by the use of extra loading on extremities, raising uncertainty on how this may affect ball speed alongside kinetics and kinematics. Taking into account that the International Tennis Federation (ITF) does not prohibit the use of materials that modify the shape or physical properties of the racquet, the appearance of new equipment and training techniques may offer other ways of modifying momentum and consequently StV without modifying the racquet's features, giving insight on new ways of affecting velocity production. As a starting point, further knowledge on how StV and accuracy are affected when modifying swing weight could be interesting for developing specific intervention programs that seek to maximize the mechanical power output using light loads [9].

Added to this, around four changes of direction per point and as many as 1000 per match are produced during tennis match-play and cover on average a distance of 8–15 m per point [10,11], highlighting the importance of short distance sprinting, change of direction speed (CODS) and agility for competitive tennis players [3]. Following the aforementioned use of wearable resistance training systems in order to improve physical aspects in predominantly upper body actions, literature shows some interesting performance adaptations when using this kind of equipment. Aspects such as oxygen consumption or energy cost are increased when running using certain external light loads on the lower limbs [12]. Furthermore, the use of wearable devices on the trunk and limbs may also affect sporting aspects such as jumping and sprinting, decreasing or increasing performance [13,14]. The use of light loads that can easily be attached and don't interfere in the athlete's movement could enable higher execution velocities that performed in a sport-specific context may further optimize training adaptations [15]. However, literature seems to be limited when speaking of the effects of these wearables in over-the-ground sprinting or acyclic sporting actions such as agility or CODS [13], which would more appropriately fit those actions present in tennis match-play.

A mobile weighting set with the name of Powerinstep®, consisting of various weight capsules (50, 100, 150, 200 g) and a wristband or plastic pieces to place them on the player's wrist or instep could be one of the aforementioned systems that practitioners may be interested in using in order to develop velocity production on both, specific tennis strokes and change of direction performance. Therefore, the purpose of this study was to investigate the acute effects of the use of a weighting set (Powerinstep®) on the tennis player's wrist or shoe on measures of StV, accuracy and CODS in comparison with 5 different conditions (i.e., wearing 50, 100, 150, 200 g weights or no weights at all) in young competitive tennis players. It is hypothesized that the use of certain weights that increase the momentum of the swing without altering speed (i.e., 100 g and 150 g) and that do not exceed a certain weight and interfere in velocity production (i.e., 200 g) will improve StV without affecting accuracy. On the other hand, CODS will be negatively affected exponentially as weight increases.

Materials and methods

Subjects

Seventeen (6 female and 11 male) competitive tennis players (mean ± SD; age, 16.5 ± 1.3 years; height, 1.75 ± 8.4 m; weight, 67.0 ± 8.1 kg; BMI 22.04 ± 1.8 kg/m^2) with an International Tennis Number (ITN) ranging from 2 to 4 participated in this study. Based on the repeated-measures design and an anticipated statistical power of 0.80 with an effect size 1.2, it was determined that a minimal sample size of n = 15 subjects would be necessary (G-Power software version 3.1.9.5, University of Dusseldorf, Dusseldorf, Germany). The player's ITN was established by the consensus of three coaches accredited with RPT (Registry of Tennis Professionals) level 3, following the ITN Description of Standards. Out of the seventeen players, just one of them used a one-handed backhand style while the remaining subjects played two-handed. Participants had a weekly volume of training of 25h/week^{-1}, and were required to have a minimum of 1 year of experience in tennis and strength training. Also, they should not have experienced any pain in the trunk/upper body or other musculoskeletal discomfort in the six previous months.

Ethics statement

All subjects were informed in advance about the characteristics of the study and, before their participation, the participants and their legal tutors, in the case of being underage, voluntarily signed an informed consent. The study was conducted following the ethical principles for biomedical research with human beings, established in the Declaration of Helsinki of the AMM (2013) and approved by the Ethics Committee of the Catalan Sports Council (01/2019/CEICEGC).

Experimental design

A randomized, repeated measures within study design were assessed to compare the acute effects of wearing a set of weights (50, 100, 150, 200 g. Powerinstep®) with respect of not wearing them on StV, accuracy and CODS in young competitive tennis players. All weight sets were provided by Powerinstep® and consisted of one weight attached to a wristband for StV testing and two weights with instep plastic pieces for attachment to assess CODS (Fig 1). A familiarization session was carried out to inform on how to place the weights to avoid discomfort and possible inconveniences. Conditions were randomly distributed to avoid the influence of fatigue and test-learning effects. Subjects weren't familiarized with in-step or wrist weights. As dependent variables, StV (in km·h^1), accuracy points and CODS (in seconds) were recorded to compare between 4 different conditions (50, 100, 150, 200 g) and baseline conditions (0 g). The comparison between these situations aimed to investigate the effects of using light weight loads on StV, accuracy and CODS.

Measurements

The collection of data took place in March during a normal in-season training week in groups of 4 players and on 2 separate testing sessions, performed in the morning and executed at least 48h apart. Participants hadn't trained in the previous 24h to any of the testing sessions and received all information regarding the risks and benefits of the study to obtain the informed consent in advance. Players were allowed to consume water ad libitum. Isotonic, energetic and caffeinated drinks were not allowed before or during the testing sessions. The first session consisted of performing the CODS test while the second session was scheduled to obtain StV and accuracy parameters.

Fig 1. Powerinstep® wristband and in-step weight attachment.

Maximum stroke velocity and accuracy. Data collection was executed on a tennis hard court with stable wind conditions ($< 2\ m \cdot s^{1}$) using new tennis balls (Head ATP Pro, Spain). Before the test, subjects performed a standardized warm-up that included mobility exercises, 5 minutes of free rallies and 5 to 10 progressive serves. Each subject randomly executed 5 series of 8 serves (4 on each side of the court) with 2 minutes of rest between sets for each one of the analyzed conditions (i.e., wearing a 50, 100, 150, 200 g or no weight set on the dominant wrist as shown in Fig 1). Following the serves, and after a 5-minute rest, participants performed 5 random series of 8 forehands and 8 backhands (crossed-court) without alternating strokes following each testing condition and following the same resting periods, as explained in Fig 2. Participants wore one of the weight sets exclusively attached to the dominant extremity. Only the serves that were in the serve box and the groundstrokes that landed in the singles court were registered. Maximum StV was determined using a hand-held radar gun (Stalker ATS II, USA, frequency: 34.7 GHz [Ka-Band] ± 50 MHz). The radar was positioned in the center of the baseline, 2 m behind the line and at an approximate height of 2 m for the serves and behind the player following the trajectory of the ball. Hitting as hard and precise as possible was indicated and immediate feedback was provided to the subjects to encourage maximum effort. To avoid variability performing groundstrokes, balls were fed by a ball-throwing machine (Pop-

Set 1	*Rest*	*Set 2*	*Rest*	*Set 3*	*Rest*	*Set 4*	*Rest*	*Set 5*
x8 Serves	2min	x8 Serves	2min	x8 Serves	2min	x8 Serves	2min	x8 Serves
								Rest = 5min
x8 Forehands	2min	x8 Forehands	2min	x8 Forehands	2min	x8 Forehands	2min	x8 Forehands
								Rest = 5min
x8 Backhands	2min	x8 Backhands	2min	x8 Backhands	2min	x8 Backhands	2min	x8 Backhands

Fig 2. StV and accuracy experimental design.

Lob Airmatic 104, France) at a constant speed (68.6 ± 1.9 $km \cdot h^{-1}$). Also, accuracy of the strokes was registered for further analysis using a similar approach to Pialoux et al., 2015 [16] as explained in Fig 3. To assess serve accuracy, a ball that landed in the S1 area (1*1 m) accounted for 5 points; S2 (2*2 m), 3 points and S3 (remaining area of the serve box), 1 point. To assess groundstrokes, a ball that landed in the area FH1 or BH1 (2*2 m) accounted for 5 points; FH2 or BH2 (3*3 m), 3 points and FH3 or BH3 (rest of the tennis court besides doubles alleys), 1 point. All other ball placements resulted in zero points. Accuracy was defined by the sum of all points, with a higher score corresponding to a higher accuracy. StV assessment measurements showed good to excellent test-retest reliability (ICCs 0.73 to 0.96) with a coefficient of variation (CV) ranging from 4.6 to 5.9%. Accuracy showed poor to moderate test-restest reliability (ICCs <0.2 to 0.550), similar to previous investigations [17] but contrary to studies that found good reliability in similar assessments [18].

CODS assessment. To assess the ability to perform a single change of direction (CODS), the 505-agility test was performed on a tennis hard court [19]. Participants executed a standardized warm-up prior to the commencement of the test, consisting of a series of mobility exercises, a 5-minute jog and 3 progressive sprints. The 505-agility test consisted of sprinting from a standing position for 15 m (through the timing gates at 10 m) and executing a 180° change of direction on their preferred foot to further sprint through the timing gates [20]. Players assumed a preferred foot behind the starting position and started the test voluntarily. Results were registered using timing gates (Chronojump®, Barcelona, Spain), as they offer higher degrees of accuracy than stopwatch-recorded times [21]. All subjects executed the test two times with each one of the analyzed conditions (i.e., wearing a 50, 100, 150, 200 g on both feet (Fig 1) or no weight set in a randomized order. After every attempt, subjects were asked to rest for 1 minute prior to performing again. All measurements demonstrated a good to excellent test-retest reliability (ICCs 0.79 to 0.91) with CV ranging from 1.6 to 3.3%.

Statistical analyses

Descriptive data were expressed as mean ± standard deviation (SD). The normality of the distributions and homogeneity of variances were assessed with the Shapiro–Wilk and Levene tests, respectively. The reliabilities of test measurements were assessed using intraclass correlation coefficients (ICCs), all of agility, serve, forehand and backhand velocity measurements reached an acceptable level of reliability (ICC > 0.73). The typical error of measurement (TEM) was calculated for the intraindividual test–retest strokes (i.e., forehand, backhand and service) and CODS variables and expressed as a mean CV. Differences between the StV and accuracy and CODS 0 g (baseline) and the scores at 4 conditions (50, 100, 150 and 200 g) were evaluated using a one-way analysis of variance (ANOVA) with repeated-measures with Bonferroni-corrected post hoc analysis. Mean differences in absolute and percent values were also used. The magnitude of the differences in mean was quantified as effect size (ES) and interpreted according to the criteria used by Cohen [22] (<0.2 = trivial, 0.2–0.5 = small, 0.5–0.8 = moderate, >0.8 = large). Because forehand velocity 0 g and 150 g data were not normally distributed, Friedman's test was used to examine the differences between baseline and different weights in forehand velocity. The level of significance was set at $p \leq 0.05$. All statistical analyses were performed using SPSS 23.0 software (SPSS Inc., Chicago, IL, USA).

Results

No significant differences were found between conditions for forehand (F = 0.412; p = 0.799), backhand (F = 0.269; p = 0.897) and serve (F = 0.541; p = 0.706) velocity and forehand

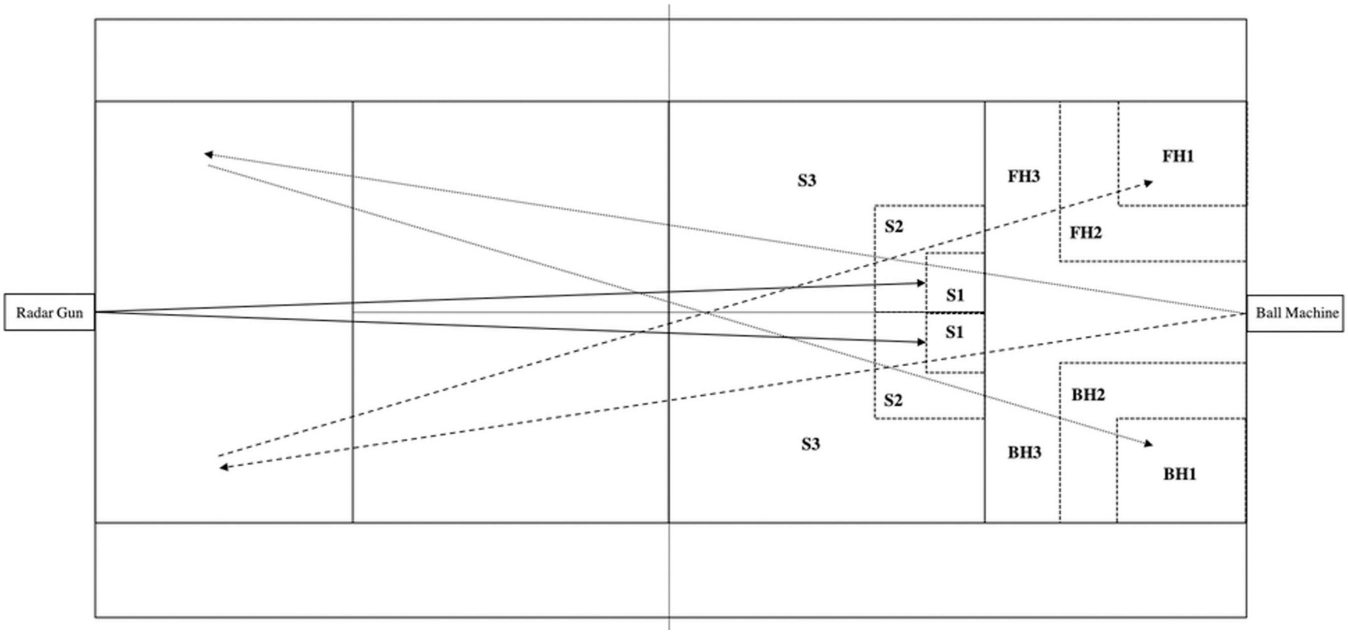


Fig 3. Tennis court layout for stroke velocity and accuracy assessment. S1, S2 and S3, the target areas for the serve; FH1, FH2 and FH3, the target areas for forehand drives; BH1, BH2 and BH3, the target areas for backhand drives. The full arrows indicate the ball trajectories for the serve, the dotted arrows indicate the ball trajectories for backhand drive, and the dash arrows the ball trajectories for forehand drive.

(F = 1.688; p = 0.161), backhand (F = 0.567; p = 0.687) and serve (F = 2.382; p = 0.059) accuracy and CODS (F = 0.416; p = 0.797).

There were no significant decreases and small-to-moderate effect sizes of StV in serve, forehand and backhand 200 g compared to the baseline (-4.5, -2.91 and -2.99%; ES = 0.48, 0.35 and 0.45) (Table 1). Moderate (23.04%; ES = -0.49) and trivial (6.06 and 7.33%; ES = -0.14 and -0.16) effect sizes for higher accuracy were found in serve, forehand and backhand 100 g compared to the baseline (Fig 4). A non-significant small effect size (-2.35%; ES = 0.41) for improvement in 200 g CODS comparing to the baseline conditions was observed (Fig 5).

Table 1. Magnitude and percentage changes from baseline (0 g) in serve, forehand and backhand velocity and accuracy and change of direction speed (CODS) between 4 conditions (50, 100, 150 and 200 g).

	50 g		100 g		150 g		200g	
	ES	%	ES	%	ES	%	ES	%
Serve								
Velocity (km·h^{-1})	0.06	-0.61	0.08	-0.76	0.29	-0.31	0.48	-4.50
Accuracy (points)	-0.08	4.30	-0.49	23.04	0.55	-29.58	0.11	-5.42
Forehand								
Velocity (km·h^{-1})	-0.06	0.52	0.06	-0.50	-0.01	0.10	0.35	-2.91
Accuracy (points)	0.48	-23.40	-0.14	6.06	0.25	-10.94	0.53	-21.47
Backhand								
Velocity (km·h^{-1})	-0.05	0.36	0.13	-0.96	0.02	-0.14	0.45	-2.99
Accuracy (points)	0.00	0.00	-0.16	7.33	-0.01	0.44	0.31	-12.09
CODS								
Time (s)	0.13	-0.60	0.13	-0.64	0.11	-0.48	0.41	-2.35

ES, Cohen's effect size; CODS, change of direction speed.

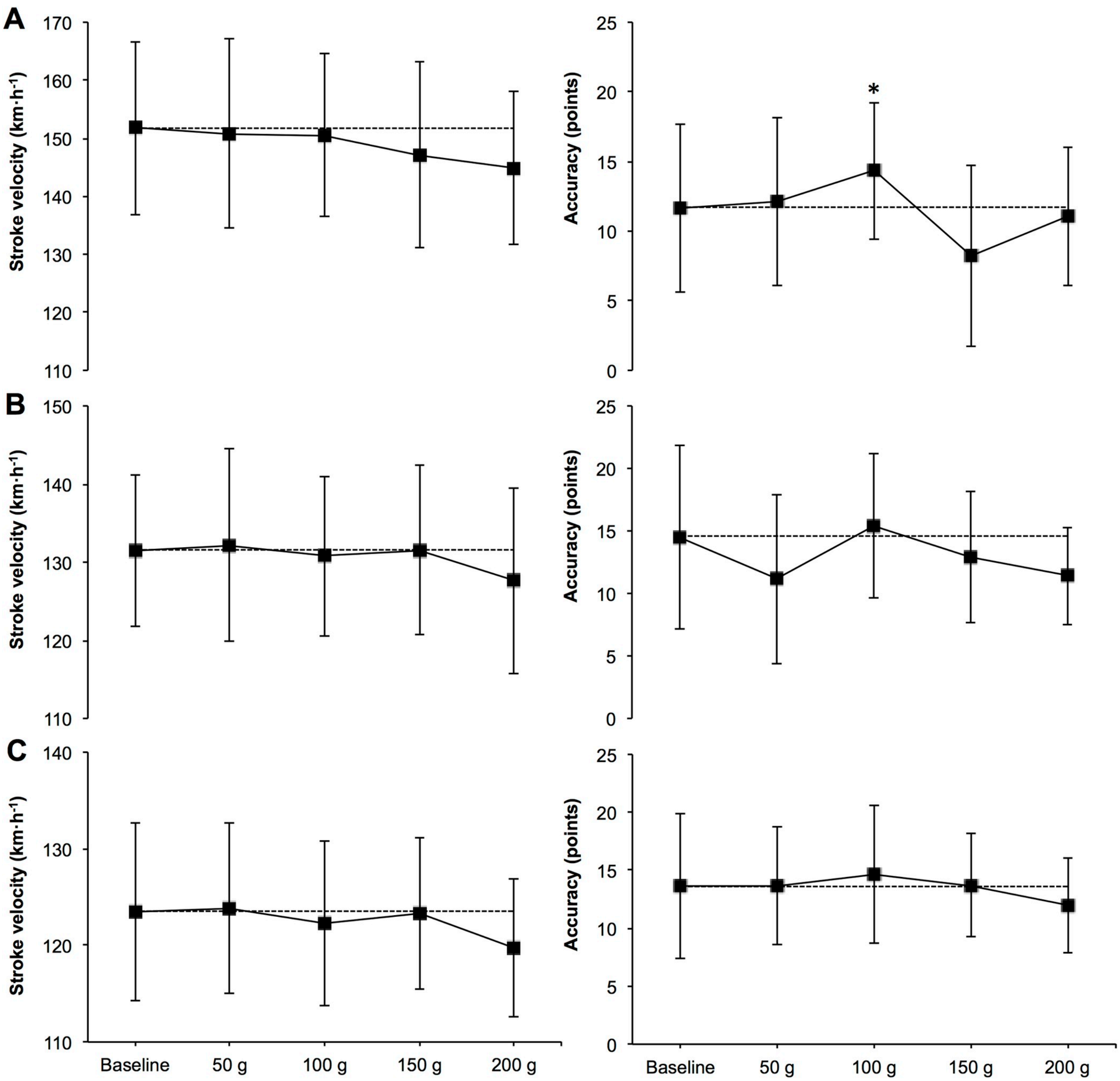


Fig 4. Comparisons of serve (A), forehand (B) and backhand (C) velocity and accuracy between 4 conditions (50, 100, 150 and 200 g). *Significant change from 150 g at $p \leq 0.05$.

Discussion

The main findings of this investigation were that the use of external light loads on upper and lower extremities do not seem to have significant effects on StV or CODS in junior tennis players. However, certain negative small-to-moderate changes were observed regarding StV when using heavier loads (200 g) and a higher accuracy without affecting velocity when using moderate loads (100 g). Regarding the use of weights on lower limbs, similar changes indicated that the use of heavier loads (200 g) could affect CODS in a positive way. Although no significant

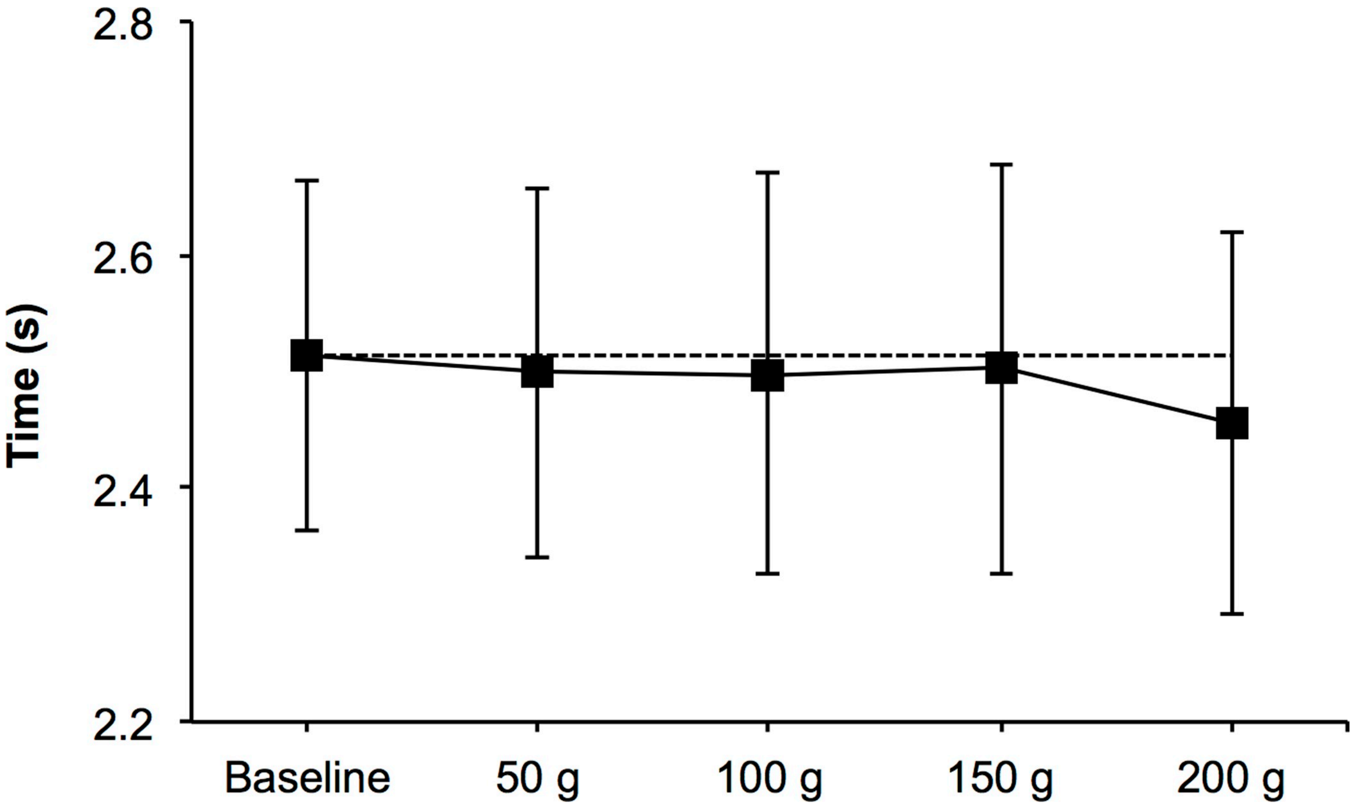


Fig 5. Comparisons of change of direction speed (CODS) between 5 conditions (0, 50, 100, 150 and 200 g).

increases in performance were observed by using a weighting set, no variables were diminished either.

More specifically, the lack of significant positive results regarding higher StV when using certain weights matches findings in other similar studies [1,4]. While literature has mainly focused on the acute effects of serve speed when adding weight to the racquet rather than the extremity as in this study, results did not find significant increases in velocity either. Even though a higher momentum caused by a heavier extremity could result in greater StV, the need of maintaining an optimal speed of the swing is necessary to benefit from this principle. As suggested by other authors, an increment in weight might cause deceleration in key determinant contributors to velocity production as internal rotation speed of the arm in the case of the serve [1,4,23]. Moreover, heavier loads placed on the extremity instead of the implement could reduce the speed of the racquet head due to a decreased linear and angular speed of the wrist, which is an important contributor to velocity production [23]. Precisely this issue may be the causative of no increases in speed in any of the weights used in this investigation and the greater loss of velocity that seems to happen when using 200 g weights (Fig 4). Interestingly, and focusing on groundstrokes, similar changes towards a decrease in StV occurred in players with a 2-handed backhand and the single subject that performed a 1-handed backhand with the weight and wrist band on his dominant extremity (2.99 and 1.12%; ES = 0.45 and ES = 0.49 respectively). Differences in both types of strokes rely on aspects such as a greater trunk rotation in the 2-handed backhand and a more rotated shoulder complex when playing with one hand [24]. In any case, as literature points out, players with either technique are able to produce similar horizontal racquet speed relying on a higher linear velocity in the 1-handed fashion or angular velocity in the 2-handed style [24]. The fact that two strokes that build speed around different kinematic aspects but obtain similar results when performing with extra light loading as in this study, may reinforce the idea that certain weights affect key factors that influence the players ability to provide speed to the stroke. Added to this, investigations have found important kinematic and physical differences between elite and competitive players, concluding that those of a greater level rely on certain variables to produce speed. Aspects such as a more efficient use of elastic energy in leg extensors [2] or horizontal shoulder and racquet velocities [25], among others, contribute to enhancing StV, highlighting the importance of specific strength and kinematic parameters. As stated previously, the use of weights on the player's extremity may affect some of the mentioned key factors. Moreover, only players of a certain age and level may be able to maintain arm and racquet swing speed invariable and benefit from a higher momentum at impact on both, groundstrokes and serves. As a limitation of this study and aspects further investigations could focus on, the analysis of kinematic differences between the use of different weights and maturity/age status differences of the players could be registered to offer a further approach to the results obtained. Regarding the differences observed in the use of moderate weights (100 g), results seem to indicate slight changes towards an increased accuracy with unaffected velocity. It may appear that this could be a suitable load to observe positive longitudinal effects on StV or accuracy. Unlike non-significant immediate results observed in investigations that focused on acute effects [1,4], longitudinal studies that proposed the use of extra light loading around the implement or mobile offered positive increases in other overhead sports [6,7] besides tennis [8]. As literature suggests, the use of these kinds of strength training programs seem to be a good way of enhancing velocity production [26], benefiting from the principle of overload. On the other hand, this approach could compromise other factors such as kinematics and kinetics of the sporting action or injury rates [7]. Following suggestions presented by other authors, these interventions could be a way of improving velocity production after achieving a certain strength level in previous programs to, after, transfer these gains into specific tennis actions such as the serve and groundstrokes [4]. Concerning accuracy, as results seem to show small-to-moderate differences for greater scores with velocity unaffected, the use of this approach to training may offer players and coaches some beneficial technical outcomes regarding skill acquisition based on variability during the training of the stroke itself, following modern coaching practices [27]. In any case, to our knowledge, this is the first study to examine the acute effects of increasing weight on extremities) on StV or accuracy, manifesting the need of further investigations to expose such statements. As a limitation, and regarding accuracy reliability, the test was probably limited by asking subjects to hit the ball at maximum speed, causing greater variability in accuracy and consequently decreasing it. This issue has previously been observed in tennis [17] and is frequent when testing accuracy.

Regarding the use of light weights on lower limbs, no studies, to our knowledge, have attempted to investigate the effects on agility aspects or, more specifically, on CODS. Linear sprinting has received attention from literature both on treadmill and over-the-ground conditions [13,28], showing no changes in running or sprinting technique but decreases in performance (maximum sprint running), especially in the acceleration phase due to a reduction in stride frequency [29]. Contrary to results noticed when analyzing StV, the differences observed in this study showed a small decrease in time when using the heavier load (200 g), unlike the mentioned researches. These contrary results could be due to the differences in the weight used in previous investigations. The loads presented ranged from 1–5% of bodyweight in the mentioned studies whereas the higher load in this investigation (i.e., 200 g) accounted for around 0.335% of bodyweight. Loads of a certain magnitude may interpose stride frequency and consequently sprinting velocity. Although little literature is available on this matter, presumably we will find differences when analyzing linear sprinting and change of direction or agility parameters such as the present here. In fact, some authors have analyzed kinematic factors affecting CODS and found better performances in those groups that had an increased stride frequency [30]. The use of wearable weights may cause greater stride rate triggered by the enhanced gravitational forces [31] and consequently result beneficial for agility-based tasks as the 505-agility test analyzed in this study. At any rate, further studies should focus on investigating longitudinally the effects of in-step weights on change of direction and agility and examine how loads may affect essential kinematic aspects such as stride length or frequency that are key determinants of CODS [13] performance before being able to state this.

In conclusion, the use of a weighting set on both wrists and in-steps does not significantly affect StV or CODS respectively. Although differences are not observed, the use of these light weights do not affect negatively velocity production or accuracy scores in junior tennis players either. Taking into account that further investigation is needed, small-to-moderate differences show an interesting improvement in accuracy and no variance in velocity production when using some of the weights tested (i.e., 100 g), suggesting that the use of this kind of apparel as a training tool could result in some way useful. This study also shows certain small changes for an increased performance in CODS when using 200 g in-step weights, suggesting that gear of these characteristics may affect change of direction or agility aspects to some extent. In any case, further investigations on the effects of the use of weighting sets on StV and CODS would be of great interest.

Practical applications

Taking into account that using certain external light loads on the upper limbs in the form of a weight set does not seem to affect negatively velocity production or accuracy scores in young competitive tennis players, the use of this kind of apparel as a training tool could result in improvements on StV in the mid-long term, as suggested in similar literature [8]. Most likely, it would be preferable that strength training preceded wearable weight interventions, being this type of protocols more adequate for in-season programs where the goal is to transfer strength gains into specific tennis actions. Furthermore, programs should be applied with caution and not be maintained during long periods of training or competition since some studies suggest compromised kinematics and kinetics of the sporting action or increases in injury rates when analyzing light-weight interventions [7]. Moreover, variability of practice may be induced by the use of this piece of equipment and offer coaches and players new insights in emergent methods of training [27]. Regarding the use of in-step weights and their effects on CODS, further studies are needed to examine how loads may affect essential kinematic aspects such as stride length or frequency that are key determinants of CODS performance.

Acknowledgments

The authors thank all the players and coaches for their enthusiastic participation. They would like to thank Academia Sánchez-Casal Barcelona. The research leading to these results has been conducted using funds from the agreement between the Universitat de Vic–Universitat Central de Catalunya and Powerinstep, SL. The authors declare that there is no conflict of interest between the participants, the materials and equipment used, or any other procedure undertaken during the experiments and the researchers of this investigation.

Author Contributions

Conceptualization: Joshua Colomar, Ernest Baiget, Francisco Corbi.

Data curation: Ernest Baiget.

Formal analysis: Ernest Baiget.

Funding acquisition: Ernest Baiget.

Investigation: Joshua Colomar, Joshua Muñoz.

Resources: Joshua Muñoz.

Supervision: Francisco Corbi.

Validation: Joshua Colomar, Ernest Baiget, Francisco Corbi.

Visualization: Joshua Colomar.

Writing – original draft: Joshua Colomar.

References

1. Whiteside D, Elliott B, Lay B, Reid M. The effect of racquet swing weight on serve kinematics in elite adolescent female tennis players. Journal of Science and Medicine in Sport. 2014; 17(1):124–8. https://doi.org/10.1016/j.jsams.2013.03.001 PMID: 23587536
2. Girard O, Micallef JP, Millet GP. Lower-limb activity during the power serve in tennis: Effects of performance level. Medicine and Science in Sports and Exercise. 2005; 37(6):1021–9. PMID: 15947729
3. Girard O. Physical determinants of tennis performance in competitive teenage players. Journal of Strength and Conditioning Research. 2009; 23(6):1867–72. https://doi.org/10.1519/JSC.0b013e3181b3df89 PMID: 19675471
4. Söğut M. Acute effects of customizing a tennis racket on serve speed. Baltic Journal of Sports Health and Science. 2017; 1(104):44–6.
5. Miller S. Modern tennis rackets, balls, and surfaces. British Journal of Sports Medicine. 2006; 40 (5):400–5.
6. DeRenne C, Szymanski DJ. Effects of Baseball Weighted Implement Training: A Brief Review. Strength and Conditioning Journal. 2009; 31(2):30–7.
7. Reinold MM, Macrina LC, Fleisig GS, Aune K, Andrews JR. Effect of a 6-week weighted baseball throwing program on pitch velocity, pitching arm biomechanics, passive range of motion, and injury rates. Sports Health. 2018; 10(4):327–33. https://doi.org/10.1177/1941738118779909 PMID: 29882722
8. Genevois C, Frican B, Creveaux T, Hautier C, Rogowski I. Effects of two training protocols on the forehand drive performance in tennis. Journal of Strength and Conditioning Research. 2013; 27(8):677–82.
9. Wilson GJ, Newton RU, Murphy AJ et al. The optimal training load for the development of dynamic athletic performance. Med Sci Sports Exerc 1993; 25:1279–86. PMID: 8289617
10. Kovacs MS. Applied physiology of tennis performance. British Journal of Sports Medicine. 2006; 40:381–386. https://doi.org/10.1136/bjsm.2005.023309 PMID: 16632565
11. Fernandez J, Sanz D, Mendez-Villanueva A. A review of the activity profile and physiological demands of tennis match play. Strength and Conditioning Journal. 2009; 31(4):15–26.
12. Martin PE. Mechanical and physiological responses to lower extremity loading during running. Medicine and Science in Sports and Exercise. 1985; 18(4):415–9.
13. Macadam P, Cronin JB, Simperingham KD. The effects of wearable resistance Training on metabolic, kinematic and kinetic variables during walking, running, sprint running and jumping: A Systematic Review. Sports Medicine. 2017; 47(5):887–906. https://doi.org/10.1007/s40279-016-0622-x PMID: 27638041
14. Macadam P, Simperingham KD, Cronin JB. Acute kinematic and kinetic adaptations to wearable resistance during sprint acceleration. Journal of Strength and Conditioning Research. 2017; 31(5):1297–304. https://doi.org/10.1519/JSC.0000000000001596 PMID: 27548784
15. Hrysomallis C. The effectiveness of resisted movement training on sprinting and jumping performance. Journal of Strength and Conditioning Research. 2012; 26(1):299–306. https://doi.org/10.1519/JSC.0b013e3182185186 PMID: 22158137
16. Pialoux V, Genevois C, Capoen A, Forbes SC, Thomas J, Rogowski I. Playing vs. nonplaying aerobic training in tennis: Physiological and performance outcomes. PLoS One. 2015; 10(3):1–10.

17. Menayo R, Moreno FJ, Fuentes JP, Reina R, Damas J. Relationship between motor variability, accuracy, and ball speed in the tennis serve. Journal of Human Kinetics, 2012; 33(1):45–53.

18. Guillot A, Di Renzo F, Pialoux V, Simon G, Skinner S, Rogowski I. Implementation of motor imagery during specific aerobic training session in young tennis players. PLoS One. 2015; 10(11):e014331.

19. Stewart PF, Turner AN, Miller SC. Reliability, factorial validity, and interrelationships of five commonly used change of direction speed tests. Scandinavian Journal of Medicine and Science in Sport. 2014; 24 (3):500–6.

20. Draper JA, Lancaster MG. The 505 test: A test for agility in the horizontal plane. Australian Journal of Science and Medicine in Sport. 1985; 17(1):15–18.

21. Hetzler RK, Stickley CD, Lundquist KM, Kimura IF. Reliability and accuracy of handheld stopwatches compared with electronic timing in measuring sprint performance. Journal of Strength and Conditioning Research. 2008; 22(6):1969–76. https://doi.org/10.1519/JSC.0b013e318185f36c PMID: 18978613

22. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 1988. United States of America: Lawrence Erlbaum Associates.

23. Elliot BC, Marshall RM, Noffal GJ. Contributors of upper limb segment rotations during the power serve in tennis. Journal of Applied Biomechanics. 1995; 11:433–442.

24. Reid M, Elliot BC. The one- and two-handed backhands in tennis. Sports Biomechanics. 2002; 1(1):47–68. https://doi.org/10.1080/14763140208522786 PMID: 14658135

25. Landlinger J, Lindlinger S, Stöggl T, Wagner H, Müller E. Kinematic differences of elite and high-performance tennis players in the cross court and down the line forehand. Sport Biomechanics. 2010; 9 (4):280–95.

26. Van den Tillaar R. Effect of different training programs on the velocity of overarm throwing: a brief review. Journal of Strength and Conditioning Research. 2004; 18(2):388–96. https://doi.org/10.1519/R-12792.1 PMID: 15142008

27. Reid M, Crespo M, Lay B, Berry J. Skill acquisition in tennis: Research and current practice. Journal of Science and Medicine in Sport. 2007; 10(1): 1–10. https://doi.org/10.1016/j.jsams.2006.05.011 PMID: 16809063

28. Simperingham KD, Cronin JB. Changes in sprint kinematics and kinetics with upper body loading and lower body loading using exogen Exoskeletons: a pilot study. Journal of Australian Strength and Conditioning. 2014; 22(5):69–72.

29. Ropret R, Kukolj M, Ugarkovic D, Matavulj D, Jaric S. Effects of arm and leg loading on sprint performance. European Journal of Applied Physiology and Occupational Physiology. 1998; 77(6):547–50, https://doi.org/10.1007/s004210050374 PMID: 9650741

30. Hewit JK, Cronin JB, Hume PA. Kinematic factors affecting fast and slow straight and change-of-direction acceleration times. Journal of Strength and Conditioning Research. 2013; 27(1):69–75. https://doi.org/10.1519/JSC.0b013e31824f202d PMID: 22362087

31. Rusko H, Bosco C. Metabolic response of endurance athletes to training with added load. European Journal of Applied Physiology and Occupational Physiology. 1987; 56(4):412–8. https://doi.org/10.1007/bf00417768 PMID: 3622484

12

An author keyword analysis for mapping Sport Sciences

Luis-Millán González[1], Xavier García-Massó[2], Alberto Pardo-Ibañez[1], Fernanda Peset[3], José Devís-Devís[1]*

1 Departamento de Educación Física y Deportiva, Universidad de Valencia, Valencia, Spain, **2** Departamento de Didáctica de la Expresión Musical, Plástica y Corporal, Universidad de Valencia, Valencia, Spain, **3** Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, Valencia, Spain

* jose.devis@uv.es

Abstract

Scientific production has increased exponentially in recent years. It is necessary to find methodological strategies for understanding holistic or macro views of the major research trends developed in specific fields. Data mining is a useful technique to address this task. In particular, our study presents a global analysis of the information generated during last decades in the Sport Sciences Category (SSC) included in the Web of Science database. An analysis of the frequency of appearance and the dynamics of the Author Keywords (AKs) has been made for the last thirty years. Likewise, the network of co-occurrences established between words and the survival time of new words that have appeared since 2001 has also been analysed. One of the main findings of our research is the identification of six large thematic clusters in the SSC. There are also two major terms that coexist ('REHABILITATION' and 'EXERCISE') and show a high frequency of appearance, as well as a key behaviour in the calculated co-occurrence networks. Another significant finding is that AKs are mostly accepted in the SSC since there has been high percentage of new terms during 2001–2006, although they have a low survival period. These results support a multidisciplinary perspective within the Sport Sciences field of study and a colonization of the field by rehabilitation according to our AK analysis.

Editor: Sergi Lozano, Institut Català de Paleoecologia Humana i Evolució Social (IPHES), SPAIN

Competing interests: The authors have declared that no competing interests exist.

Introduction

Sport Sciences is a field of study that embraces knowledge produced by research around those social practices such as sport, physical activity, play, game and exercise which is developed from different epistemological and methodological research perspectives. In order to understand the nature of this field, the present paper undertakes a grand overview by analyzing the articles published within the research journals that circulate through Web of Science (WOS), one of the main international data bases. This task is particularly difficult since the scientific production in this and other fields has grown exponentially over the years [1–3], and this is closely linked to the growth of scientific journals. These journals increase in diversity and specialization and multiply the number of articles per year. This inordinate growth has evolved with the advent of the digital age, which has facilitated the reduction of publishing costs associated with the old paper-based system. Researchers have also seen how the amount of information contained in an article has spread with the appearance of supplementary materials. In short, this boost in scientific information is causing researchers to have great problems in knowing the state of the art of a particular topic of study.

Paradoxically, this augment of information is not always an advantage for research groups, since much of the findings are diluted in the large number of bytes that are published [4]. For this reason, research teams are including specialists in the recovery, classification and analysis of information, which improves the efficiency of the scientific documentation search [5]. However, despite the fact that information retrieval systems have become more sophisticated over the last few years, the analysis of scientific literature still requires crafting processes that involve expert reading and interpretation.

Researchers write reviews, systematic reviews, meta-analyses or letters to the editor in order to make the main findings accessible in a given field [6]. These works have been gathered in specialized products in the form of journals or databases (e.g., Cochrane Library). The most successful authors concisely summarize the knowledge they have acquired in their laboratories or the interpretation of the results of the best published articles. With enormous effort and remarkable cognitive abilities, researchers use their experience in this type of article to highlight the strengths and weaknesses in a particular subject of study. It seems that the ability of the human brain to include or discard relevant information is one of the most effective ways to address this type of task [7]. Undoubtedly, this huge effort to successfully manage the information overload deserves the production of papers that usually receive more citations and academic recognition.

However, the highly specialized nature of these review works should be balanced with other works of a more general focus in a certain field of study [8]. Science is increasingly universal and interdisciplinary [9], and yet the specialized literature suffers from the lack of an overview. Cross-sectional knowledge of what is being done in a field might feed the imagination of future researchers, thus leading them to pose hypotheses that could break the barriers imposed by super-specialized knowledge. These grand overviews are even more necessary in fields, such as Sports Sciences, which gathers biological, social and humanities-based studies accompanied by other works of inter and cross-disciplinary character. Therefore, the challenge relies on how we address the analysis of an entire field of study with hundreds of thousands published works.

The increase in data and academic documents has boosted data-mining and text-mining disciplines to manage the huge amount of data. Both disciplines automatically look for patterns that underlie the large masses of information. Although data-mining is a promising alternative, intellectual property protection and the lack of open source data are, to date, two obstacles that seem insurmountable in the short term. Nonetheless, text-mining presents some interesting advantages since what the authors write in their articles is usually expressed in natural language, and most of their information is available in open access (e.g., titles, abstracts, keywords, etc.) [10].

Text-mining refers to the process of extracting useful and non-trivial knowledge from different textual databases using various techniques that are automatically applied to digital environments [11]. Although the text-mining term mainly refers to the analysis of an unstructured text, many of its techniques (e.g. link analysis, reduction dimensionality or clusterization) are also used for structured texts. These techniques are mainly used to analyse content published on the Internet [12–14], although they are also used on scientific documents published in repositories or databases. Bibliometrics, as a discipline that studies the behaviour of scientific publications, also applies these analyses to full texts or parts of the text, such as the title, abstract or keywords [15–17].

Among the different sections stored from a document in scientific databases, Author Keywords (AKs) play a prominent role. The investigations around the AKs are numerous and developed from different approaches, although most of them use the counting terms and co-occurrence networks as the main tools [18–24]. Only a few papers over the years have focused their research interests on the dynamics of keywords [25–27]. Several studies also analysed AKs in the whole field of Sport Sciences [28,29] through the subject matter contained in the documents. However, to the best of our knowledge, no work to date has conducted a global analysis of the issues addressed in the Sport Sciences field.

Against this backdrop, the purpose of this study is threefold. First, to detect the most relevant AKs of the Web of Science (WOS) Sport Sciences Category (SSC) as a representation of the field of study that bears the same name. Second, to discover the dynamics of these words over the years and how these AKs are related to each other, thus giving rise to major research topics. Third, to quantify the most innovative words and how they survive throughout the subsequent years.

Materials and methods

The authors voluntarily choose AKs, thus performing a succinct exercise of representing the whole text of a document [30]. According to Jones and Jackson, "Keywords are a list of words or phrases that are provided by the author and signify the meaning or main ideas presented in the paper" [31].

AKs have advantages when compared to other sections that are stored in databases (e.g., title, abstract). For instance, their volume of information is easier to manage in storing and analysing than other sections. In fact, AKs consume few bytes of information because there are no connectors between words, which consequently facilitates their storage and handling through computer systems. Obviously, other similar sections like the title, abstract or even the full text contain more information, but the analysis of hundreds of thousands of works supposes a very large volume of text that requires very exclusive computer systems that are accessible to few researchers. The AKs sections do not contain irrelevant information, and everything is 'edible'. In this sense, AKs do not allow for the manipulation of information by the researchers, thus creating an unbiased position compared with the management of sections with more words. Since there is no possibility of selecting or transforming words, AKs allow to accomplish the positivist science postulate that the observer should not influence the phenomenon under study.

Author keywords selection

AKs selection criteria include belonging to SSC journals and the Science Citation Index (SCI) in which AKs were published. Eighty-one journals were indexed in this category in 2016. All the AKs published in journals of the SSC were the universe of our study.

The SSC ranks 86 out of 234 categories based on the number of journals published. The median Impact Factor of this category is 1.681, which ranks it as the 104th out of 234 categories. Fig 1 shows the WOS categories from which Sport Sciences documents are retrieved. As can be observed, the three categories with the highest representation beyond the specific SSC are (in this order) Physiology, Orthopaedics and Rehabilitation.

The Journal Citation Reports from WOS was used as the tool for searching the journals and the AKs. The ISSNs of each of the journals included in the SSC were downloaded and combined into a single equation. The file used for searching is included in Supplementary Material (S1 File).

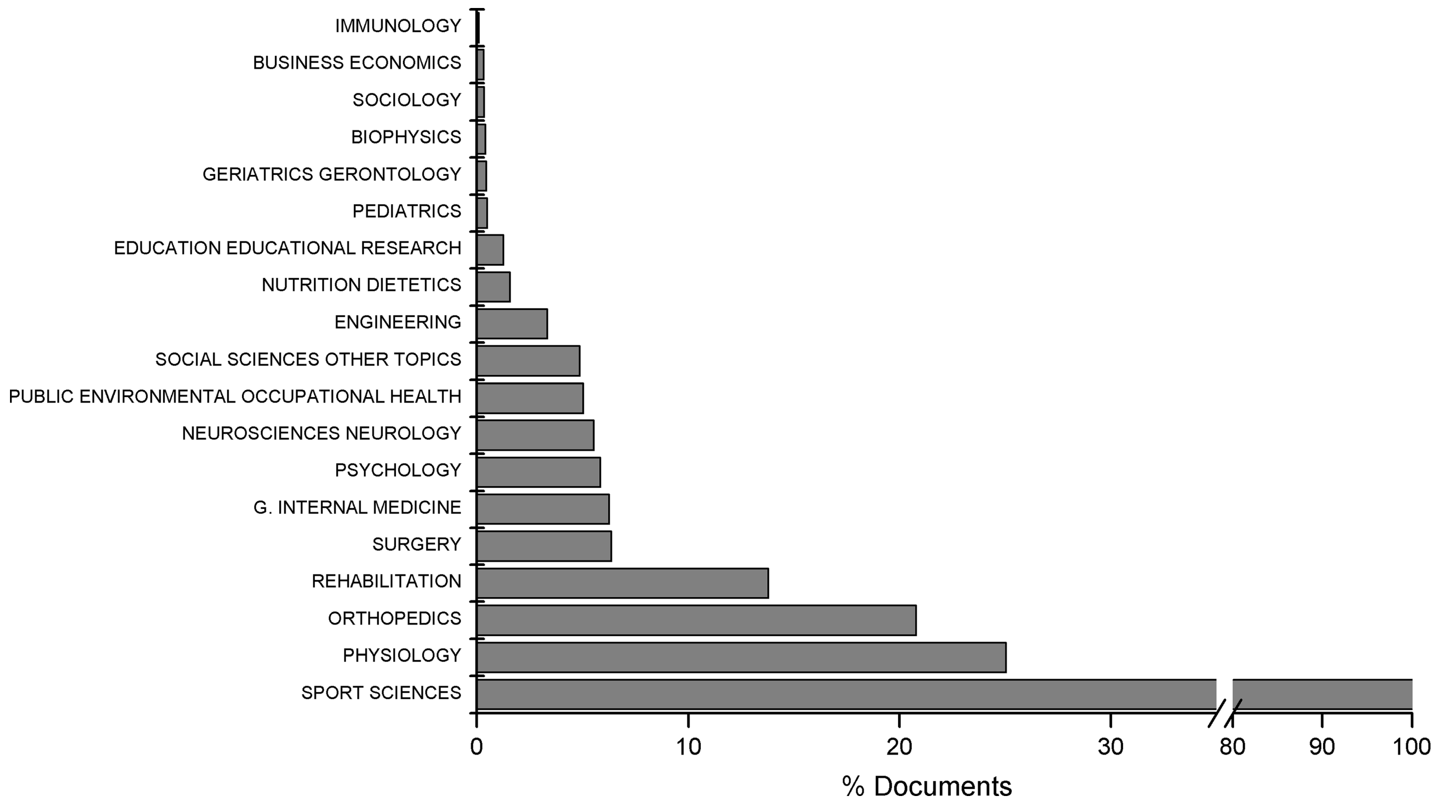


Fig 1. WOS categories from which documents were retrieved.

The equation was introduced 2017-09-19 in the WOS Core Collection database and it was established the publications before 2017 as the time limit. Only articles, reviews and letters were considered for this analysis. There was no restriction to the language because English is compulsory for title, abstract and keywords. The aforementioned search yielded 168,299 documents. The records were downloaded in batches of 500 documents in plain text format with the following fields: i. Author keywords, ii. Year Published, iii. Subject Category, iv. Publication Name, v. ISSN, vi. Times Cited and vii. Unique Article Identifier. The downloaded files were stored on a hard drive for further analysis.

To obtain and store the AKs in a single document, the software bibexcel (version 2011-02-03, Olle Persson, Umeå University, Umeå, SWE) was used. Since there was no declared AK field in many journals or it was recently declared, the resulting file contained fewer references than the original search. The final document was stored in plain text, and the fields associated with each keyword were separated through a tabulator.

Author keywords description and co-occurrence analysis

The retrieved AKs were described through frequency tables. The number of times that AKs appeared and their dynamics throughout the period were studied. Those keywords with high frequency appear in the results section, and all of them can be observed in S1 Table. The cleaning process of the 100 most frequent words was done manually, mainly unifying the terms that were in singular and plural and acronyms. For instance, SPORT and SPORTS or ELECTROMYOGRAPHY and EMG.

A citation analysis of the articles in which the AKs had been published was also carried out. With the total number of appointments, the following parameters were calculated: i. Number of citations per article that contained the AK, ii. Percentage of articles that contained the AK that had never been cited, iii. Number of citations obtained by the most cited article that contained the AK, and iv. Hirsch index [32]. Additionally, the three most frequent words that appeared in the journals belonging to the Sport Sciences category were calculated.

A co-occurrence analysis was also performed with the AKs in order to show the number of times that two words appeared simultaneously in a published article. This relationship is established with greater or lesser strength depending on the repetition of this pair of words in a published paper. The co-occurrences of AKs form a graph in which the nodes are the AKs and the edges are the co-occurrence relationships between them. As in any graph, the importance of the nodes can be measured through different parameters of centrality.

We used the bibexcel software to create the AKs' co-occurrence network, and only the co-occurrences that appeared at least two times or more were taken into account. Pajek software (version 5.01, Batagelj and Mrvar, University of Ljubljana, Ljubljana, Slovenia) was used to visualize and perform the centrality calculations of the network. To facilitate the visual interpretation, several reductions of the original network were made, such as eliminating those edges that had smaller values. The following centrality parameters were calculated: i. All Degree, ii. All Proximity Prestige, iii. Betweenness centrality and iv. Average Distance from All Domain. These are usual parameters to describe the importance each node has within the network. A more detailed description of the equations used for its calculation can be found in previously published works [18,33,34]. In order to locate the nodes in a two-dimensional space a Kamada-Kawai algorithm was used [35]. Once nodes were located, small manual changes were developed to improve the visibility of labels.

The clusters originated from the most important co-occurrence relationships in the last network were also calculated. We used VOSviewer software (version 1.6.4, Nees Jan van Eck and Ludo Waltman, Leiden University, Leiden and Erasmus University, Rotterdam, Netherlands) to establish the clusters, especially the equation that produces the density maps and calculates the resulting clusters [36–38]. This software uses a new mapping technique called VOS, which stands for visualization of similarities. In a simplified way, the equation proposed by the authors calculates the forces of repulsion and attraction of the different nodes as a function of the distance and the strength that joins them. Finally, the number of clusters depends on the resolution that is applied. In our case, this parameter took the value of $\gamma = 1$.

An analysis of the clusters dynamic through time was performed using a heat map, in a similar way as it has been done with AKs frequencies. In this analysis the results were expressed as parts per unit regarding to the highest count of a specific AK within a cluster during a year (highest value = 1).

New author keywords search from 2001–2006

We choose a six-year period (2001–2006) as a first step to locate new AKs, since there were no previous studies that use this type of methodology. This period is in between of many previous years and sufficient subsequent years with AKs to conduct the analysis of new words with guarantees.

In order to locate the new words and to track their frequencies of appearance during the 10 subsequent years, custom-written software routines were established (MATLAB R2013a, MathWorks Inc., Natick, MA, USA). Any word variation from previous ones was considered as a 'new' word and included those that were written with some orthographic or typographical mistakes. However, the 'mistaken' words will likely disappear over the years, thus becoming anomalies without impacts. If anomalies still survive, then it would be a case of new use accepted by the scientific community, and consequently, it would have an impact on the scientific writings.

We only made two exceptions to this rule: i) differences between uppercase and lowercase were not considered, and ii) Hyphens were removed. For instance, Meta-analysis → Meta analysis.

As a necessary step prior to the analysis, the records were divided into three periods based on the years in which the AKs were published: 1) the historical period [1889–2001], 2) the onset period [2001–2006], and c) the survival period [10 subsequent years 2002–2011, 2003–2012 (. . .) 2007–2016]. Once the periods were established, a retrospective search was started with the words published in the period of appearance looking back into the historical period. Those words that had never been published in the historical period were selected as new words in the SSC.

Later, a search was made in the survival period with the new words selected. The search was conducted year after year, thus resulting in a vector of 10 columns for each word in which 1 indicated the appearance of the term and 0 indicated no appearance. In addition, the number of times the word appeared every year was counted. Frequency tables were calculated for the entire generated file.

New author keywords analysis during the survival period

With the new words detected, a first analysis was carried out. It consisted of calculating the probability that each new word had to survive or disappear. An analysis of the survival through Kaplan-Meier curves was then proposed. This type of analysis estimates the time that passes until a certain event occurs. This analysis can be used beyond the estimation time until death since the survival analysis can be applied to all those events that occur over time and have been previously defined. In fact, this type of analysis has been applied to a large number of fields of study, such as medicine, economics, production engineering and social sciences [39].

Before conducting this analysis, it is necessary to establish three fundamental aspects: the time of observation, the moment in which the event of interest occurs and when a subject is censored. In our analysis, the research subjects are the AKs, and they were followed up over the survival period. That is, we are going to test over 10 years if these words appear in an article published in the SSC.

Regarding the definition of the event, it is necessary to consider that a word can appear and disappear discontinuously over a period of 10 years. Therefore, it could be argued that the data is interval-censored [40]. However, we think that as long as there are subsequent records in which the word appears, a particular AK is alive as researchers use it while preparing their articles. As a general criterion, it was established that a word 'died' in the following year of its last appearance in which there were records. When this circumstance occurred, this year was indicated as the moment in which the event occurred and then the analysis stopped for this word [39].

The last step in our analysis was to establish which words were censored. In our work, only those words that appeared throughout the observation period were considered censored (right-censoring).

Once the data matrix was prepared, the analysis was performed using the SPSS 20 software (IBM, Armonk, USA). We calculated the Kaplan-Meier curves for the total set of new words, the average survival times, and the 95% confidence intervals.

Results

Descriptive statistics

In all, 111,606 documents that contained AKs were obtained from a total of 168,299 records. The first AK of the SSC appeared in 1983. As expected, the amount of AKs has been growing over the years due to the increase in scientific production and the popularization of AKs usage in the field for indexing papers. AKs usage was not generalized until 1991, as can be observed in Fig 2. A total quantity of 504,479 AKs were recorded between 1983 and 2016. This amount

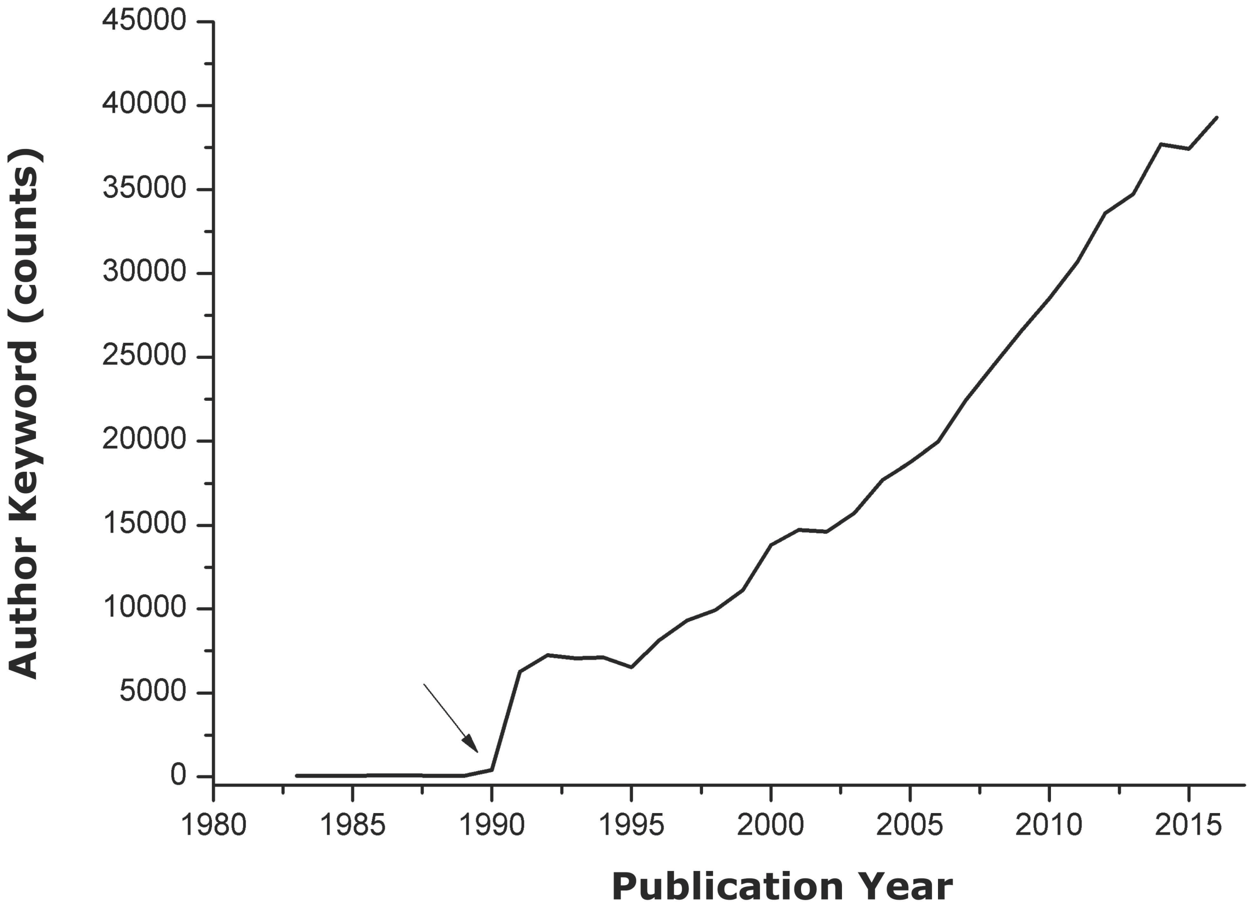


Fig 2. Frequency of AK appearance in sport sciences category throughout the 1983–2016 period. The arrow indicates that 1991 was when AK usage began to be generalized.

is reduced to 101,824 AKs when duplicate words are discarded. Although there is an average of four AKs per article, 27 AKs were observed in a particular paper [41].

Twenty-six words appeared more than 1000 times over the tested years (Table 1). The most repeated word was 'REHABILITATION' followed by 'EXERCISE'. Both terms represent approximately 1.5% of the total. Words with a lower frequency of appearance can be found in S1 Table. Regarding the evolution of these words over the years analysed, Fig 3 shows how 'EXERCISE' was the most used AK until 1999, and 'REHABILITATION' became the predominant AK in the following years.

The AKs had an average number of 19 characters in length. Particularly, 0.4% of the AKs had a greater length than 50 characters, and only 5 AKs had a length of 1 character.

Regarding citations received by articles which contain the main AKs, Table 2 shows how the words 'AGING' and 'PHYSICAL ACTIVITY' were more efficient. On the contrary, 'SPORT' and 'ATHLETES' were the less efficient words.

It can be observed in Table 3 the AKs most used within the journals with top ten impact factor in 2016 in SCC.

Author keyword co-occurrence and clusters

A co-occurrence network was built with a total of 504,479 AKs. Only those words that at least appeared accompanied by others a minimum of 2 times were considered for elaborating the

Table 1. Author keywords that appeared more than 1000 times during the 1983–2016 period in the sport sciences category.

Author Keyword	Counts
REHABILITATION	7647
EXERCISE	5975
ELECTROMYOGRAPHY	3136
BIOMECHANICS	2961
PHYSICAL ACTIVITY	2510
KNEE	2273
ANTERIOR CRUCIATE LIGAMENT	2199
GAIT	1996
FATIGUE	1700
AGING	1671
PERFORMANCE	1622
KINEMATICS	1610
SPORT	1519
SHOULDER	1394
OXYGEN CONSUMPTION	1353
STROKE	1316
ATHLETES	1313
STRENGTH	1253
SPINAL CORD INJURIES	1235
HEART RATE	1223
TRAINING	1214
INJURY	1186
MAGNETIC RESONANCE IMAGING	1180
CHILDREN	1119
SOCCER	1050
RUNNING	1007

co-occurrence matrix. The resulting network had 101,757 nodes and 729,800 edges. To visually represent it with a manageable number of words (nodes), several reductions were made. The process of reducing the general network is graphically represented in Fig 4.

Fig 4A shows the first reduction (threshold$\geq$ 5). It is a network with 3994 nodes and a total of 15037 edges. The average degree of the network is 7.52. The largest distance between nodes was 10 jumps between the words 'BREATH HOLD' and 'SENSORIMOTOR SYNCHRONIZATION'. In Fig 4B the edges with values less than 25 (threshold$\geq$ 25) have been eliminated and a network with 469 nodes was presented. Finally, a more drastic reduction is presented in Fig 4C. In this graph, only those edges with values above 50 co-occurrences (threshold$\geq$ 50) appear. It is then possible to visualize a network with 185 nodes and 302 lines. This network visually facilitates its interpretation, but the number of nodes is still very large. Therefore, a network only with co-occurrences greater than 100 was used for centrality calculations and cluster extraction. All networks centrality parameters can be observed in S2 Table.

Table 4 shows the centrality values of the 20 most prestigious AKs within the network. The AK that reached the greatest number of co-occurrences was 'REHABILITATION', with it concurrently being the word with the highest degree and betweenness. This AK was also the one that had the lowest average distance (1.44 jumps) from the other AKs. The remaining values for the set of nodes can be observed in S2 Table.

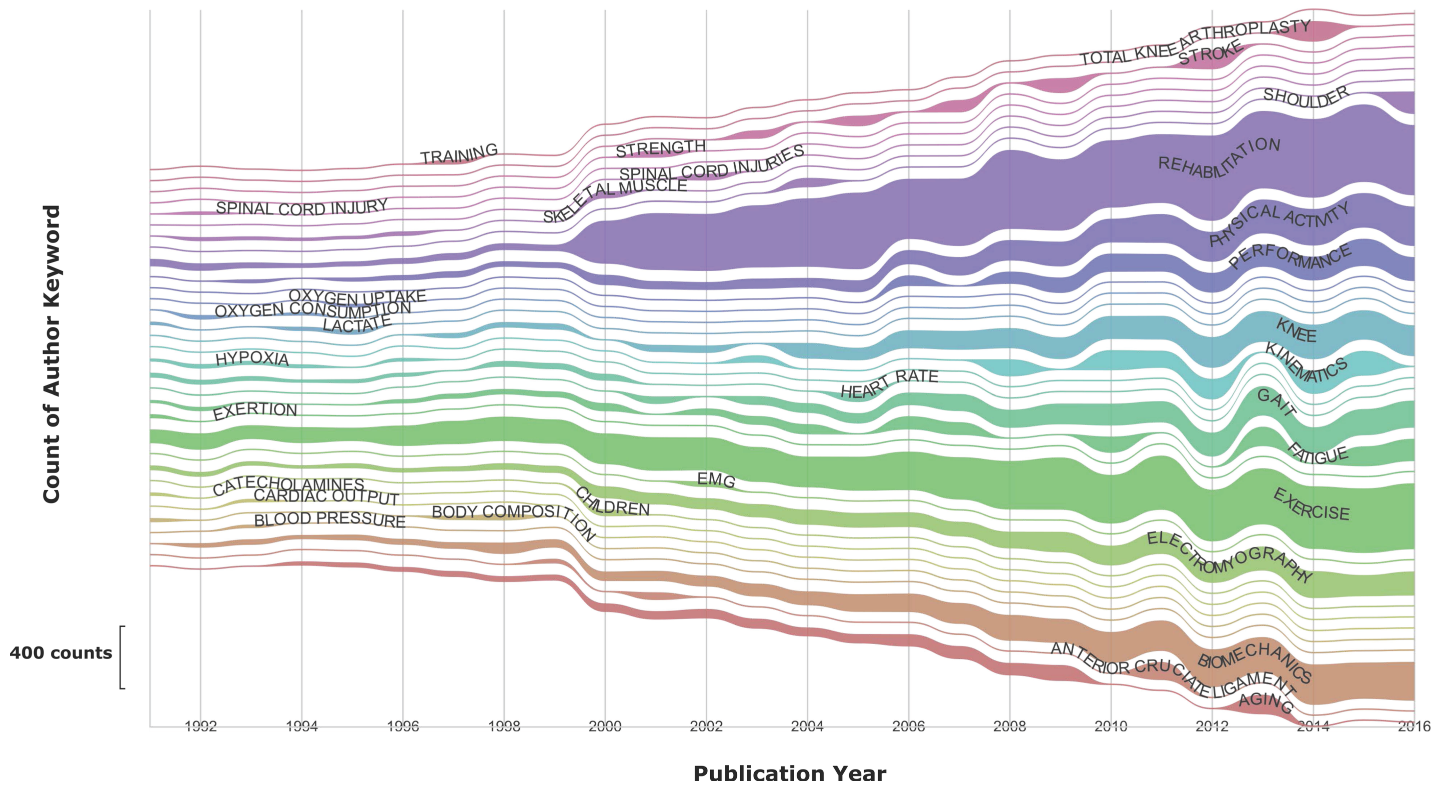


Fig 3. Changes in the frequency of the 10 most frequent AKs over the years from 1991–2016. No AKs before 1991 appear in the figure because the total frequencies are under 30 AKs. A total of 31 AKs among the top frequent ten were identified during the analysed years. The size of the lines is proportional to the count of the number of times they appear in a given year. In the first year (1991), the total frequency was 363 AKs, and in the last year the total frequency of the 10 most used words was 3294 AKs.

In a second level of analysis, we grouped the AKs into clusters in order to indicate which topics are the most common in the SSC. The clusters are represented with different colours in Fig 5. In our analysis, there are 6 major themes that are related to each other.

The most central cluster of all is commanded by the word 'REHABILITATION' and appears in red. Several medical terms about musculoskeletal and nervous system pathology share a space in this cluster with different physiotherapy techniques.

The cluster commanded by the word 'EXERCISE' appears in yellow, and it exerts absolute control over other terms. It is a cluster in which training and physiology have an important position.

In pink and blue are two clusters close to each other. The first refers to the 'BIOMECHANICS' and the second to postural control. In between biomechanics and postural control, a cluster whose main theme is traumatology is coloured in garnet.

A sixth cluster is formed by the words 'PHYSICAL ACTIVITY' and is related to different moments of the life cycle.

It can be seen in Fig 6 the dynamics of each cluster along the years. The 'EXERCISE AND TRAINING' cluster 'was the predominant at the end of the past century, given this place to 'REHABILITACION' later on.

New author keywords appeared in the period 2001–2006 and survival in the following years

During the years from 2001–2006, a total of 21,662 new AKs were published, which corresponds to 42.31% of the total AKs published in this period. Among the new ones that appeared

Table 2. Citation parameters for articles which contain the most popular AKs in the sport sciences category.

Author Keyword	citations / artícle	% artícles without citations	h-index	Maximum number of citation of an article	Unique Article Identifier*
AGING	31.55	5.45	101	1111	WOS:000227665400002
PHYSICAL ACTIVITY	29.01	9.04	115	2417	WOS:A1993KG53700011
EXERCISE	27.56	8	146	3185	WOS:000089257400009
CHILDREN	25.79	7.77	75	805	WOS:000207606400001
GAIT	25.72	5.91	96	979	WOS:000177649600002
ANTERIOR CRUCIATE LIGAMENT	24.96	6.91	89	362	WOS:000169666000011
STROKE	24.47	4.94	80	567	WOS:000231077900028
FATIGUE	24.32	6.88	86	455	WOS:A1996VA58600029
REHABILITATION	24.06	5.57	133	921	WOS:000170275000007
OXYGEN CONSUMPTION	23.66	6.95	75	1185	WOS:000292773000025
STRENGTH	23.56	8.06	79	589	WOS:000185850300013
MAGNETIC RESONANCE IMAGING	23.31	5.34	64	703	WOS:000088230400011
INJURY	22.89	9.19	74	369	WOS:000072475800012
RUNNING	22.75	9.04	71	649	WOS:000084834900012
SPINAL CORD INJURIES	22.66	3.97	60	517	WOS:A1992HX73900002
HEART RATE	21.95	9.4	72	565	WOS:000167227800019
ELECTROMYOGRAPHY	21.83	7.08	87	603	WOS:000178034600020
KINEMATICS	21.5	8.7	83	369	WOS:000072475800012
BIOMECHANICS	21.18	6.35	98	555	WOS:A1992JX14100009
SOCCER	21.11	12.76	72	446	WOS:000230141300007
KNEE	20.94	7.22	92	362	WOS:000169666000011
TRAINING	20.57	8.98	70	423	WOS:A1995RV19200001
PERFORMANCE	18.91	8.2	76	431	WOS:000263752200026
SHOULDER	18.39	8.32	70	335	WOS:000083353800002
SPORT	15.41	15.14	66	391	WOS:A1997XJ67700017
ATHLETES	14.42	14.85	58	289	WOS:000230326800005

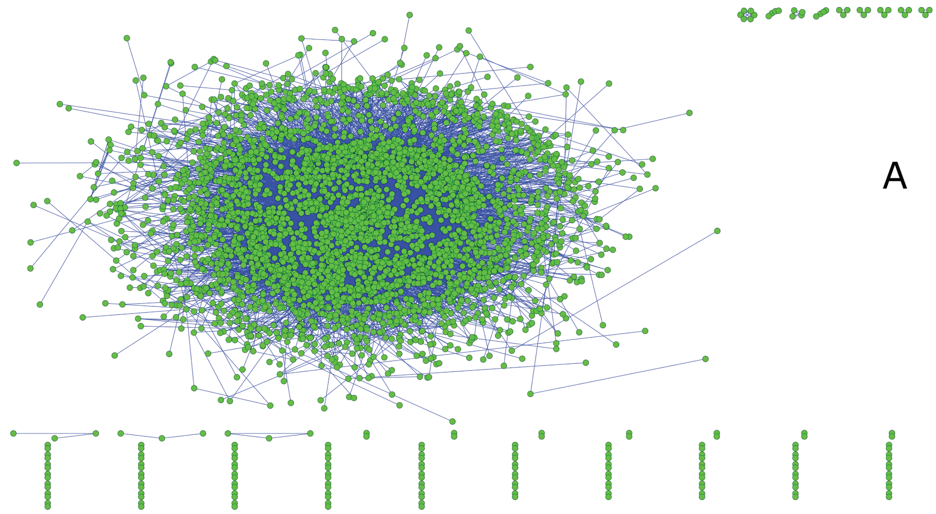
Data are ordered according to the number of citations by article which contains the AK. h-index = Hirsch index.

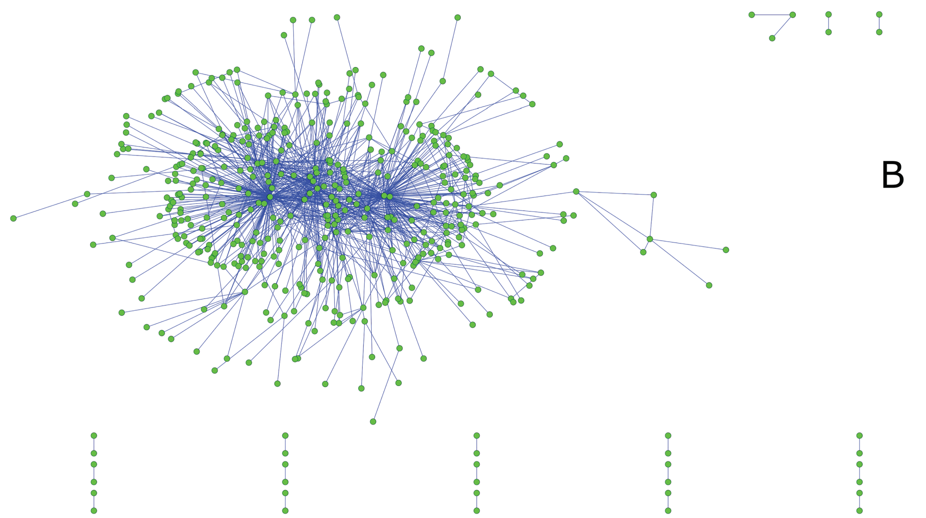
* Indicates the WOS identifier of the most cited paper (for searching purposes, include this identifier in the Advanced Search tag field UT).

Table 3. Most used author keywords within the top ten journals of the sport sciences category.

Journal*	AK 1	AK 2	AK 3
Exercise Immunology Review	EXERCISE	INFLAMMATION	CYTOKINES
British Journal of Sports Medicine	EXERCISE	INJURY PREVENTION	EPIDEMIOLOGY
American Journal of Sports Medicine	KNEE	ANTERIOR CRUCIATE LIGAMENT	SHOULDER
Exercise and Sport Sciences Reviews	EXERCISE	AGING	SKELETAL MUSCLE
Medicine and Science in Sports and Exercise	EXERCISE	PHYSICAL ACTIVITY	AGING
Journal of Science and Medicine in Sport	PHYSICAL ACTIVITY	EXERCISE	CHILDREN
Journal of Applied Physiology	EXERCISE	SKELETAL MUSCLE	HYPOXIA
Scandinavian Journal of Medicine & Science in Sports	EXERCISE	PHYSICAL ACTIVITY	SOCCER
Archives of Physical Medicine and Rehabilitation	REHABILITATION	SPINAL CORD INJURIES	STROKE
Knee Surgery Sports Traumatology Arthroscopy	KNEE	ANTERIOR CRUCIATE LIGAMENT	TOTAL KNEE ARTHROPLASTY

*Sports Medicine is not included in the table because this journal does not accept AKs. AK1, AK2, AK3 the three more frequent words in each journal, ordered from major to minor.


A


B

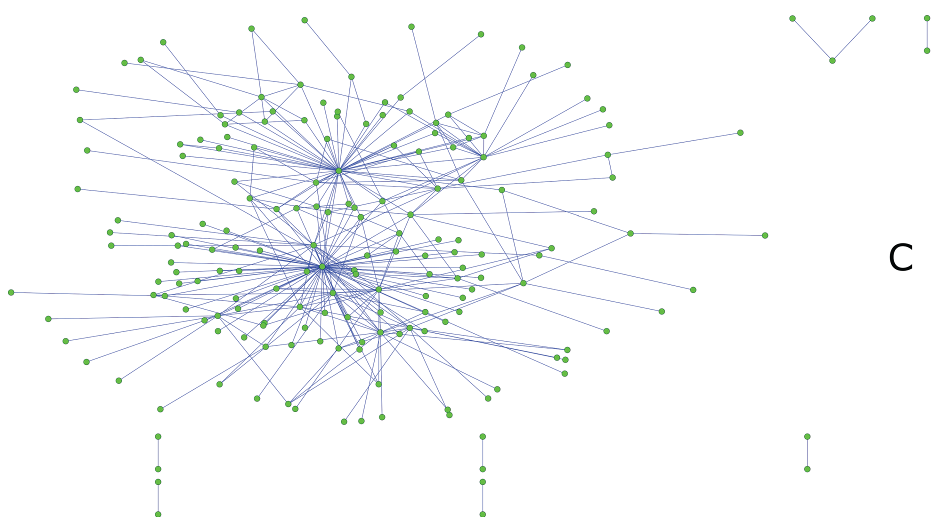

C

Fig 4. Author keyword networks co-occurrence. (A) shows the network of co-occurrences with 3994 nodes (threshold≥ 5). (B) shows the reduction of the network (threshold≥25) with a total of 469 nodes. (C) shows 185 nodes with the highest co-occurrence values (threshold≥50).

in this period, those that had greater success during the following years were 'POSTURAL BALANCE', 'DOUBLE BUNDLE', 'PERFORMANCE ANALYSIS' and 'COMBAT SPORTS', which had total frequencies above 100 (Table 5). The word that had a stronger debut was 'PHYSICAL THERAPY TECHNIQUES', which was used for the first time 11 times in 2003.

Only 61 journals accepted new AKs during the 2001–2006 period out of the total number of journals in the SSC. The Journal of Applied Physiology, Medicine and Science in Sports and Exercise and Aviation Space and Environmental Medicine were the ones that published a greater number of new AKs among all the analysed ones.

Fig 7 shows the survival curves of AKs over the 10 years after their appearance. It can be observed that during the first year, more than half of the words disappeared and were not used again until the end of the analysed period. Since then, a soft fall is observed that is accentuated slightly towards the end. Only 9.4% of the words arrived at the end of the period without any event being observed. Moreover, only 2027 words were used during the ten years after its appearance. The average time (95% CI) of survival for the series was 2.93 (2.88 to 2.98) years.

Discussion

This is the first article in performing an empirical global analysis of the Sport Sciences research, by using AKs from articles contained in WOS, to identify major research trends in

Table 4. Centrality parameters of author keywords co-occurrence network.

Author Keyword	All Degree	All Proximity Prestige	Betweenness centrality	Average Distance from All Domain
REHABILITATION	39	0.616	0.662	1.446
EXERCISE	15	0.459	0.283	1.938
BIOMECHANICS	8	0.416	0.027	2.138
AGING	3	0.416	0.024	2.138
MUSCLE	3	0.416	0.005	2.138
KNEE	9	0.405	0.089	2.200
GAIT	10	0.402	0.044	2.215
ELECTROMYOGRAPHY	7	0.394	0.042	2.262
ANTERIOR CRUCIATE LIGAMENT	4	0.391	0.024	2.277
SHOULDER	5	0.388	0.058	2.292
OSTEOARTHRITIS	2	0.381	0.000	2.338
BALANCE	3	0.378	0.001	2.354
STROKE	2	0.376	0.000	2.369
WALKING	2	0.376	0.000	2.369
CEREBRAL PALSY	2	0.376	0.000	2.369
POSTURE	2	0.369	0.000	2.415
SPINAL CORD INJURIES	1	0.366	0.000	2.431
BRAIN INJURIES	1	0.366	0.000	2.431
OUTCOME ASSESSMENT (HEALTH CARE)	1	0.366	0.000	2.431
QUALITY OF LIFE	1	0.366	0.000	2.431

The table shows the 20 words with the highest Proximity Prestige value. The parameters have been calculated over a network with 74 nodes (threshold≥100)

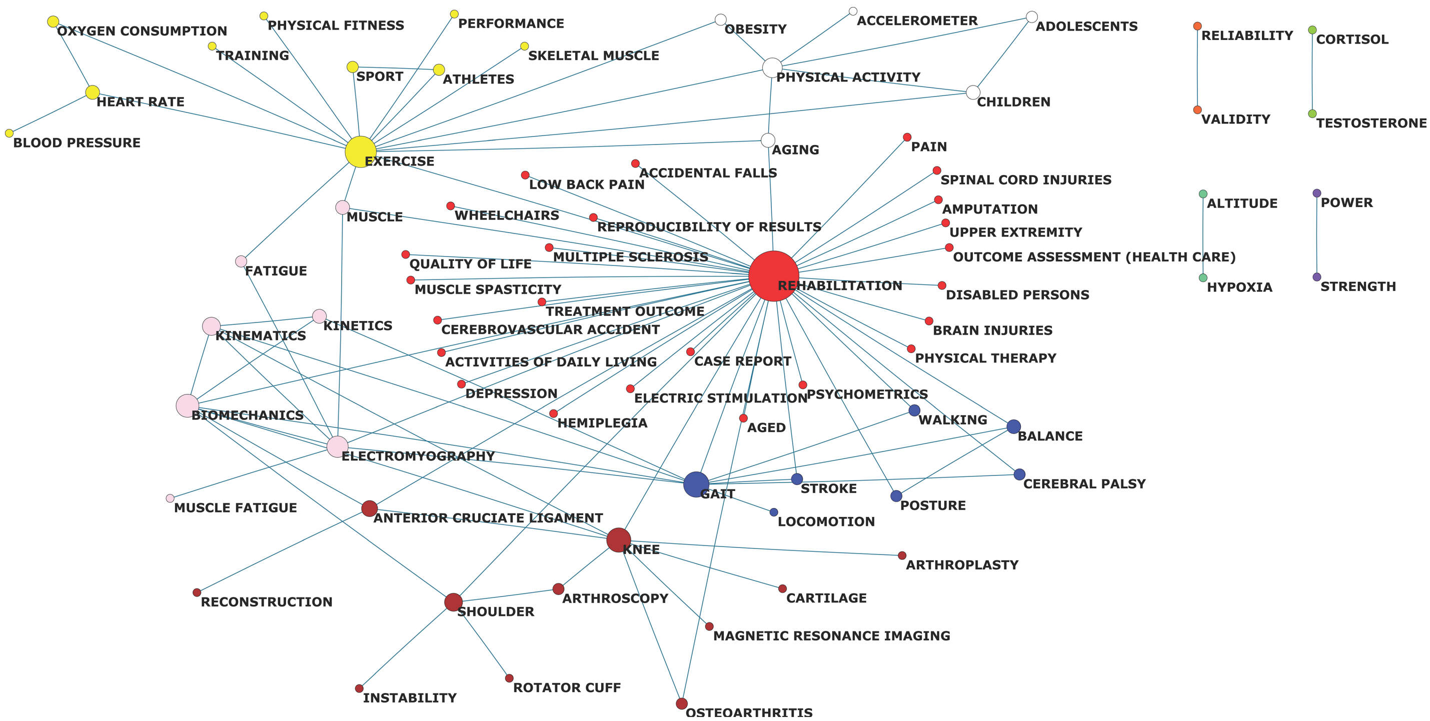


Fig 5. Main research topics in the sport sciences category according to the co-occurrence of author keywords. 6 large clusters appear with different colours as follows: Red = Rehabilitation, Yellow = Exercise and training, Pink = Biomechanics, Garnet = traumatology, Blue = Gait and balance, and White = Physical activity. The size of the nodes indicates co-occurrence between related terms, and a larger size means a greater co-occurrence.

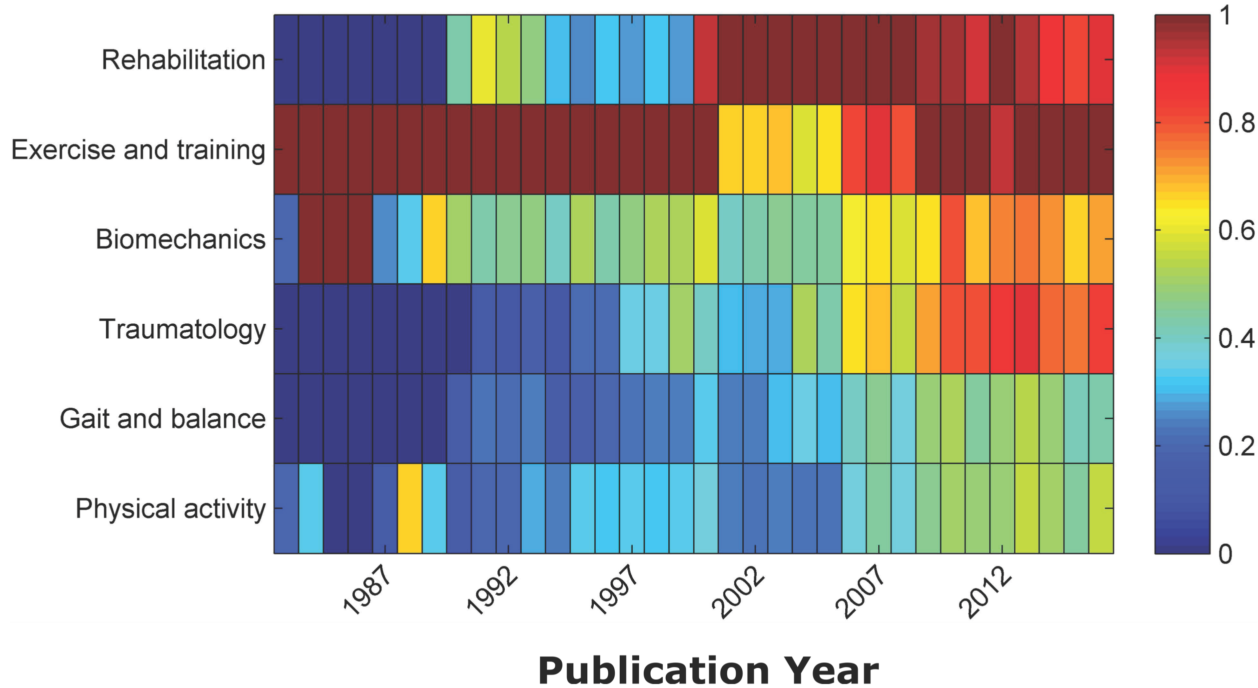


Fig 6. Heat map on the AKs dynamic contained in each cluster. Data were expressed as parts per unit regarding to the highest count of a specific AK within a cluster during a year (highest value = 1).

Table 5. New author keywords in 2001–2006 period and frequency of use during the following 10 years.

New Keyword 2001	Freq. (2002–2011)	New Keyword 2002	Freq. (2003–2012)	New Keyword 2003	Freq. (2004–2013)	New Keyword 2004	Freq. (2005–2014)	New Keyword 2005	Freq. (2006–2015)	New Keyword 2006	Freq. (2007–2016)
PARKINSON DISEASE	46	ICF	89	TRAINING LOAD	46	POSTURAL BALANCE	158	DOUBLE BUNDLE	126	OSTEOCHONDRAL LESION	46
FEMOROACETABULAR IMPINGEMENT	34	AUTOLOGOUS CHONDROCYTE IMPLANTATION	57	CAREGIVERS	31	LOCKING PLATE	57	PERFORMANCE ANALYSIS	123	AUTONOMY SUPPORT	45
EVIDENCE BASED MEDICINE	32	TYPE 2 DIABETES	57	REVIEW (PUBLICATION TYPE)	31	REPEATED SPRINT ABILITY	49	COMBAT SPORTS	108	SEDENTARY LIFESTYLE	44
MUSCLE, SKELETAL, PHYSIOLOGY	32	POSTACTIVATION POTENTIATION	42	MEDIATION	30	ANTERIOR CRUCIATE LIGAMENT (ACL) RECONSTRUCTION	46	TALENT DEVELOPMENT	50	TROCHLEAR DYSPLASIA	41
RANDOMIZED CONTROLLED TRIALS	31	NIRS	33	PHYSICAL THERAPY TECHNIQUES	30	CORE STABILITY	44	TIBIAL SLOPE	42	AMPK	39
SPORTS NUTRITION	28	AKT	28	HANDWRITING	29	CONCUSSIONS	36	GLOBAL POSITIONING SYSTEM	36	RELATIVE AGE EFFECT	39
ACHILLES	27	GLUTEUS MEDIUS	28	MOSAICPLASTY	28	ADIPONECTIN	33	SMALL SIDED GAMES	34	PGC 1 ALPHA	38
SHOULDER ARTHROPLASTY	27	SUBSCAPULARIS	24	ACTIGRAPH	24	POLICY	33	MYOSTATIN	31	ORTHOPAEDIC TRAUMA	37
QUALITATIVE	26	OVERHEAD ATHLETE	21	BIOMARKERS	23	SLAP	30	VALIDATION STUDIES	29	INERTIAL SENSORS	36
AUSTRALIAN FOOTBALL	24	MINIMALLY INVASIVE SURGERY	20	DOUBLE BUNDLE RECONSTRUCTION	23	HIGH INTENSITY INTERVAL TRAINING	29	COMBAT SPORT	23	ANATOMIC RECONSTRUCTION	34
HAMSTRING TENDON	24	WHIPLASH INJURIES	20	FALL PREVENTION	21	PACING STRATEGY	29	INCLUSION	22	MESENCHYMAL STEM CELLS	31
STEADINESS	24	COMPUTER NAVIGATION	19	NEURONAL PLASTICITY	21	INTRA ARTICULAR	28	SURFING	20	MTOR	30
MALUNION	23	BALL SPEED	18	MULTIFIDUS	20	GAIT VARIABILITY	27	AUTOLOGOUS CHONDROCYTE IMPLANTATION (ACI)	19	SPORTS INJURY PREVENTION	30
PROSTHESES AND IMPLANTS	23	SUTURE	18	NEUROMUSCULAR TRAINING	20	POST ACTIVATION POTENTIATION	27	TROCHLEOPLASTY	19	TIME MOTION	30
SICK LEAVE	23	ARTHROSCOPIC SURGERY	17	SEMG	20	MICROARRAY	25	BUILT ENVIRONMENT	18	SUBCHONDRAL BONE	27
CENTRAL ACTIVATION	19	MMG	17	GHRELIN	18	COLLISION SPORT	24	MPFL RECONSTRUCTION	18	OSTEOCHONDRAL ALLOGRAFT	26
CHONDROCYTES	19	MUSCULOSKELETAL DISEASES	17	PATELLOFEMORAL INSTABILITY	18	IDENTITY	24	MESENCHYMAL STEM CELL	17	LOCKED PLATING	24
ENDOTHELIAL DYSFUNCTION	19	PERFORMANCE INDICATORS	17	TRAINING STATUS	18	CORE	23	REVISION TOTAL KNEE ARTHROPLASTY	17	VIBRATION TRAINING	24
AUGMENTATION	18	OLDER PEOPLE	16	VIBRATION EXERCISE	18	GAME ANALYSIS	22	SELF CONTROL	17	POST EXERCISE HYPOTENSION	23
HEART RATE RECOVERY	18	ELITE SPORT	15	PHYSICAL EDUCATION TEACHER EDUCATION	17	SKELETALLY IMMATURE	22	SPORT PHYSIOLOGY	17	POSTEROLATERAL BUNDLE	23

The Author Keywords are ordered from the highest to the lowest frequency. The table only shows the 20 most used Author Keywords for each respective year.

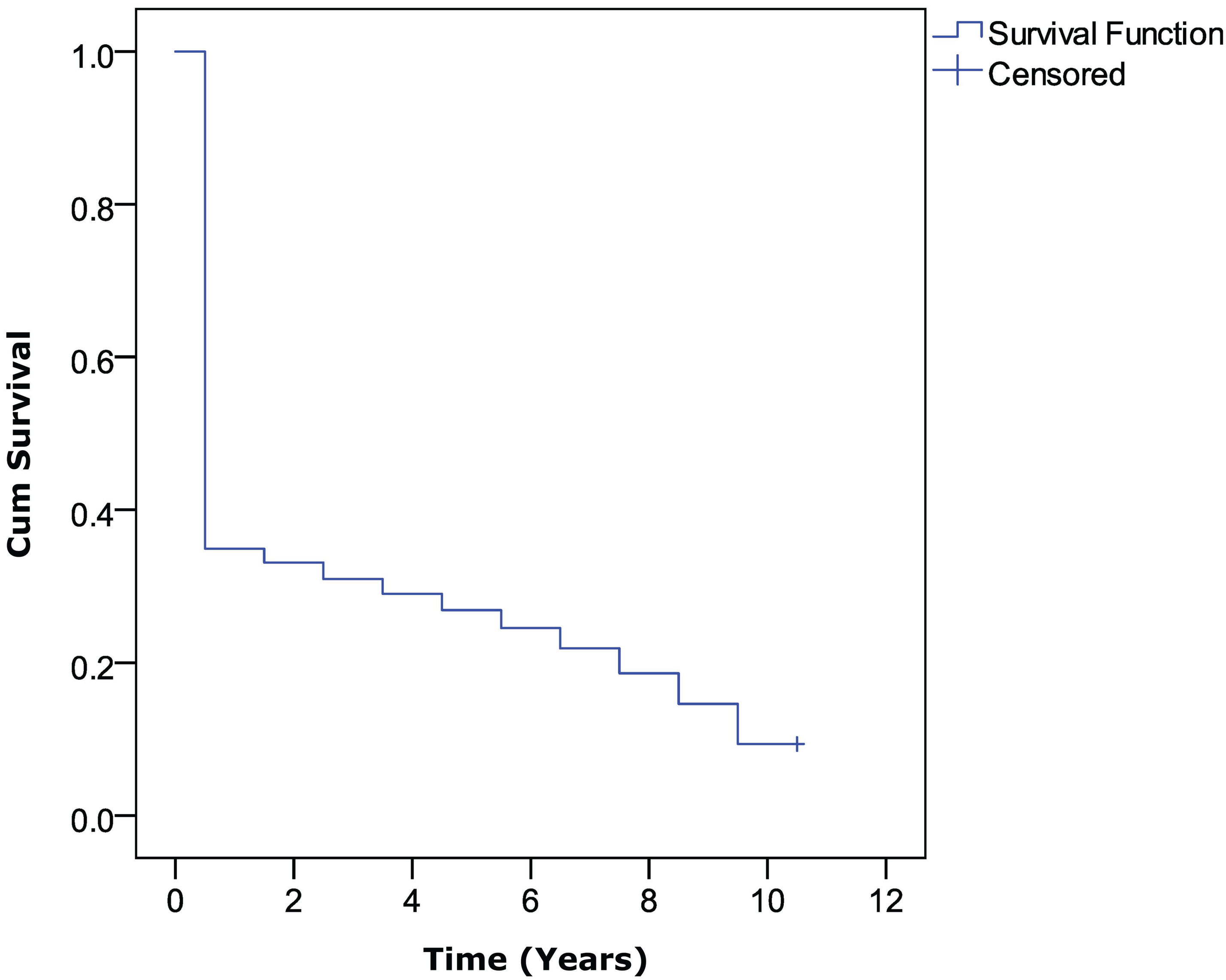


Fig 7. Kaplan–Meier curves of the ten years following of a new AK.

this field of study. Previously, only partial and regional analyses had been developed, [28,29, 42] but global analyses are necessary for understanding holistically this complex field.

A particular surprising result is the emergence of 'REHABILITATION' as the most frequent AK in the SSC. It is paradoxical that the most common term in this subject category is the one that gives name to another subject category precisely called *Rehabilitation*. According to WOS, the SSC would encompass the following topics:

> "Sport Sciences covers resources on the applied physiology of human performance, physical conditioning for sports participation, optimal nutrition for sports performance, and the prevention and treatment of sports-related injuries and diseases. This category also includes resources on sport psychology and sociology".

These topics make it possible that rehabilitation has a place in this category as part of the treatment of injuries produced during sport practice. However, such a large number of words in SSC requires a more detailed reflection.

Since journals may be assigned to more than one WOS category, 13% of our analysed documents have been published in journals that belong to Rehabilitation and SSC. This percentage is not enough to explain the predominant position of Rehabilitation AK. Since the second category of WOS that shares the most number of documents related to Sports Sciences is Physiology, it was more likely that any word related to this category would have obtained the first positions. Moreover, in one of the few studies analysing the AKs that appeared in one of the most important congresses on sports sciences held in Europe [29], the authors observed a relatively low weight of rehabilitation compared to physiology (2% vs 22%, respectively). This dissonance between our results and those described in the cited study may be due to the European College of Sport Science congress organization and its journal that consciously controls the main themes under which authors present their communications. However, the WOS category is made up of different journals, which have different editors that make decisions in competition with other journals, and consequently, their management is characterized by the decentralization of decisions. Although our data does not allow us to infer why rehabilitation has been imposed in SSC, it seems that journals editors themselves have prioritized the contents related to the secondary prevention of sports injuries.

The second most frequent AK is 'EXERCISE', a key term that undoubtedly is a fundamental part of the SSC. In essence, the word expresses the orderly development of physical activity for the purpose of maintaining or improving physical fitness [43]. Although this word is usually associated with training (as will be seen later in this discussion), its use permeates much of what is published in the area [28]. However, the absence of some terms that should theoretically emerge among the most cited in the SSC is astonishing. For instance, the term 'SPORT' is not among the most repeated words, as would be expected since it names the category. However, the term could be obliterated because the authors may assume it is not necessary to explicitly indicate the word of the category in their publications. Despite this explanation, it is still surprising that among the most cited words there is only one sport, 'SOCCER'.

The academic sport disinterest shown by our results has already been previously proven by Stone et al. in 2004 [44]. According to these authors, sport as a subject matter is being displaced by others more focused on biomedical aspects related to the practice of exercise. They argue for the methodological difficulties (i.e., internal validity vs. external validity), little training of the coaches who are consumers of the final product, deficient training of university students about the sport and scarce employability of sport scientists. Although our results do not allow us to know why researchers choose a certain subject of study, we think that the system of academic rewards are influencing the decision-making of researchers about their topics of study. This idea is reinforced by the statistics obtained regarding the citation parameters, where the word 'SPORT' gets the worst results.

Researchers who publish in journals with a high impact factor are more likely to obtain relevant positions within academic institutions and more funding for their projects [45,46]. Researchers choose topics that are of interest to the journals with the greatest impact, and they are inclined towards topics that are more likely to be cited, thus entering a vicious circle from which it is difficult to leave. The journals are also part of this 'game' and are not exempt from pressure. Since prestige is associated with journal citations, it is possible that the best placed journals tend not to accept breakthrough ideas because they increase the quotations with mainstream themes that allow them to preserve their prestige. It is possible that, since the biomedical sciences have a long research tradition and well-structured methods, the topics associated with the word 'SPORT' may not be too attractive for journals. However, our experimental data does not allow us to conclude in this direction. We believe that future work should address this problem by looking at other factors or co-variables that explain the phenomenon.

Regarding the dynamics of the AKs, the last 20 years are characterized by showing few changes with respect to the analysis of the total frequency. It indicates that the SSC is quite stable over time. Both words, 'REHABILITATION' and 'EXERCISE', have a hegemonic position since AKs were introduced in the normalization of SSC journals. There are no themes that appear and then disappear. Perhaps the only exceptions to this rule are the 'AGING', 'STROKE' and 'TOTAL KNEE ARTHROPLASTIA' AKs that during the years 2012–2014 show a slight increase in their frequency of appearance. These words are directly or indirectly associated with the ageing process. Therefore, their appearance among the most cited words in the last decade may be strengthened, due to the growing concern with the progressive ageing of the population in developed countries.

An analysis of AKs alone does not offer an accurate view of what occurs in a field of study, since AKs only express a part of the articles content and are usually used as a claim for readers. Therefore, a more complex analysis focused on the connections among words (such as their co-occurrence) may enrich the view of the field. In fact, our analysis of co-occurrences among AKs has yielded interesting results. When authors match two terms in a single article, they are indicating the use of different topics to solve a particular problem. In our co-occurrence networks, it can be seen that the words 'REHABILITATION', 'EXERCISE' and 'BIOMECHANICS' are the ones that obtain the highest values in the centrality parameters that were analysed. It is especially revealing that the betweenness value obtained reflects the mediating role that 'KNEE' plays as a gateway for traumatology in our graph. Beyond the individual values that each AK obtains, the co-occurrence analysis ultimate aim is to obtain clusters that trace the predominant themes.

The clusters obtained in the results show that the rehabilitation and biomechanics AKs (two names of recognized disciplines) come together with physiology and traumatology within the SSC. It is an amalgam of multiple AKs that are fundamentally related with biomedical disciplines, while words from other social and human sciences are absent. If the SSC is intended to represent the Sport Sciences field of study in the WOS, it should expand the articles from the social and human sciences. The editors of the journals and WOS managers should assume this task, since the editors are members of the Sport Sciences community with responsibilities within the field of study [47,48], and the WOS managers are accountable for a balanced selection of journals from the field of study as a whole.

However, authors of the scientific community make the decision to write the AKs in their articles, and thus they are also responsible for the disciplinary and thematic mixtures reflected in the clusters. These clusters reflect a field of study that has not achieved an international consensus to define itself as a scientific discipline with a clear subject matter, as various authors have noted [49–51]. Since Henry [52] opened up this issue in the mid-1960s, various disciplinary proposals [53–55] and several contributions around its subject matter [56–59] have been made. These proposals have been unevenly followed, and a multidisciplinary perspective seems to have been imposed. The interdisciplinary and cross-disciplinary proposals that involve a greater integration of knowledge from biomedical, human and social disciplines present conceptual and practical problems that make them difficult to materialize [49,55,60–62].

Our analysis of AKs supports this multidisciplinary perspective that the Sport Sciences field of study has experienced since the middle of the last century. As Henry stated [63] more than fifty years ago, this field still displays a few common interests, key issues and conceptual systems, and it is not characterized by a single body of knowledge. The Sports Sciences field has been developed as an amalgam of isolated sub-disciplines that are derived from the mother ones and seek rapid academic respectability [64]. Therefore, there are few or no connections among sub-disciplines to build common aspects from all of them. This is indicated by the empirical analysis of the AKs conducted in this work and which is closer to the reflections of authors that refer to chaos [50], fragmentation and over-specialization [49,65] or the lack of integration [63] in this field of knowledge, despite some recent integrative advances [66,67]. In addition, AK analysis reflects a lack of agreement, even in the subject matter that revolves around historically consolidated concepts in the field of study, since terms such as sport, game, body and movement are missing in our analysis. However, the situation of colonization is even worse when there is a predominance of 'REHABILITATION' over other concepts, thus endangering the identity of the field of study by dissolving and absorbing the key concepts of another field of close study that has attracted the scientific interest of our community of researchers. Only 'EXERCISE' and 'PHYSICAL ACTIVITY' have a presence in the cluster analysis. This makes us think that, in addition to epistemological, methodological and conceptual problems, there are practical problems linked to prestige, recognition, employability and ultimately personal survival within academia and science in general.

Finally, the survival analysis of AKs during the years 2002–2007 allows us to establish the quantity of new terms that appear in the category and the average survival time that they have in the following years. In particular, the analysis shows that 40% of new terms are basically small variations of terms already known. Only two of them stand out for their great acceptance in the subsequent years, 'POSTURAL BALANCE' and 'DOUBLE BUNDLE'. Although there are no similar studies with which we can compare our data, it seems that the area has a good attitude towards the new words. However, they have scarce relevance since more than half of the AKs fail to pass the first year. According to our survival analysis, the average lifetime for a new word is 3 years. Only 9% of the words were used throughout the survival period. The works of Santos and Irizo [68,69] employ a model of analysis closer to ours, using the citations received by the articles. Obviously, the behaviour of citations and keywords does not have to be similar; however, we have found some similarities. Although the results section has simplified the analyses carried out to improve reading fluency, like Santos and Irizo we have tested our empirical model with different theoretical models. As with their findings, the distribution that best fits here is the Weibull distribution ($k = 0.69$, $SE = 0.01$, where k is the shape parameter), which indicates that the failure rate decreases over time. In other words, despite the sharp decline of the first year, data indicate that after a while, words begin to gain strength in the SSC.

The main limitations of our work are related to the methodology used. In our research, we choose AKs as an indicator of the contents that appear in the articles, but authors may not properly select them. Moreover, AKs from the SSC present some inaccuracies in the way that they are written, since some of them show an excessive extension. However, the impact of these singularities on our results is diluted since we have analysed the entire universe of AKs that have appeared within the WOS category. Works that choose smaller samples will have to take this limitation into account.

A second limitation refers to the period of AK analysis of 40 years and not since the beginning of the SSC. The studies whose objectives include the historical evolution of concepts should necessarily opt for the use of other fields of search in the database. The third is about the survival analysis that is restricted to the appearance of AKs in the SSC, although the words could appear in a different WOS category or other parts of the paper (title, abstract or main text).

Finally, the way in which the results are shown in this article (ordered by their frequency of appearance) may highlight the most common topics but not the most important ones. To save space, the tables and figures of our article only contain those terms that reached a high number of repetitions and excluded those terms that are more residual or less frequent from a quantitative point of view. This method limits our results, probably because the front of the SSC knowledge is made of AKs with low or medium frequency. For this reason, we choose to increase the information of this paper and make it available to the scientific community in the supplementary materials for future interpretations and analyses. Despite the effort made by the research team to maintain a neutral tone in the discourse of this article, our own background as researchers may have influenced the way of ordering and discussing the results. This is especially relevant in the case of a global analysis such as the one presented in this paper. Future studies should discuss our results from the point of view of researchers from other disciplines or people who, because of their professional work (e.g., journal editors), have a global but different view of the Sport Sciences field of study.

Conclusions

One of the main findings of our research is the identification of 6 large thematic clusters in the SSC. There are also two major terms that coexist ('REHABILITATION' and 'EXERCISE') and show high frequencies of appearance, as well as a key behaviour in the calculated co-occurrence networks. Another significant finding is that new AKs are mostly accepted in the SSC since a high percentage of new terms during 2001–2006 were observed, although with a low survival period. These results support a multidisciplinary perspective within the Sport Sciences field of study and a colonization of the field by rehabilitation according to our AK analysis. This global view of the SSC has been possible through the methodology used, which includes data mining methods for the analysis of a large amount of data. Of special interest is the survival analysis developed because it represents a new methodology in the AKs analysis. This type of analysis opens new possibilities in different areas of research to study trends and introduction of new words, not only in the academic world but also in the information and communication professional contexts.

Supporting information

S1 File. ISSN.
(PDF)

S1 Table. KW frequency.
(XLSX)

S2 Table. Centrality parameters.
(XLSX)

Author Contributions

Conceptualization: Luis-Millán González, Xavier García-Massó, Alberto Pardo-Ibañez, Fernanda Peset, José Devís-Devís.

Data curation: Luis-Millán González.

Formal analysis: Luis-Millán González, Xavier García-Massó, Alberto Pardo-Ibañez.

Methodology: Luis-Millán González, Xavier García-Massó, Alberto Pardo-Ibañez, Fernanda Peset.

Software: Alberto Pardo-Ibañez, Fernanda Peset.

Supervision: Xavier García-Massó, Fernanda Peset.

Visualization: Luis-Millán González, Xavier García-Massó.

Writing – original draft: Luis-Millán González, José Devís-Devís.

Writing – review & editing: Luis-Millán González, Xavier García-Massó, Alberto Pardo-Ibañez, Fernanda Peset, José Devís-Devís.

References

1. Tabah AN. Literature dynamics: Studies on growth, diffusion, and epidemics. Annu Rev Inf Sci Technol ARIST. 1999; 34: 249–86.
2. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. J Assoc Inf Sci Technol. 2015; 66: 2215–2222.
3. de Solla Price DJ. Science Since Babylon. Am J Phys. 1961; 29: 863–864. https://doi.org/10.1119/1.1937650
4. Ke Q, Ferrara E, Radicchi F, Flammini A. Defining and identifying Sleeping Beauties in science. Proc Natl Acad Sci. 2015; 112: 7426–7431. https://doi.org/10.1073/pnas.1424329112 PMID: 26015563
5. Cooper ID, Crum JA. New activities and changing roles of health sciences librarians: a systematic review, 1990–2012. J Med Libr Assoc JMLA. 2013; 101: 268. https://doi.org/10.3163/1536-5050.101.4.008 PMID: 24163598
6. Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. The Lancet. 2011; 377: 108–109. https://doi.org/10.1016/S0140-6736(10)60903-8
7. Norman D, Dunaeff T. Things That Make Us Smart: Defending Human Attributes In The Age Of The Machine. Edición: Revised. Boston, MA: Addison-Wesley Publishing Company; 1993.
8. Morillo F, Bordons M, Gómez I. Interdisciplinarity in science: A tentative typology of disciplines and research areas. J Am Soc Inf Sci Technol. 2003; 54: 1237–1249. https://doi.org/10.1002/asi.10326
9. Lungeanu A, Huang Y, Contractor NS. Understanding the assembly of interdisciplinary teams and its impact on performance. J Informetr. 2014; 8: 59–70. https://doi.org/10.1016/j.joi.2013.10.006 PMID: 24470806
10. Rajman M, Besançon R. Text mining: natural language techniques and text mining applications. Data mining and reverse engineering. Springer; 1998. pp. 50–64.
11. Feldman R, Dagan I. Knowledge discovery in Textual Databases (KDT). Proceedings of the First International Conference on Knowledge Discovery and Data Mining. AAAI Press; 1995. pp. 112–117.
12. He W, Zha S, Li L. Social media competitive analysis and text mining: A case study in the pizza industry. Int J Inf Manag. 2013; 33: 464–472.
13. Tseng Y-H, Lin C-J, Lin Y-I. Text mining techniques for patent analysis. Inf Process Manag. 2007; 43: 1216–1247.
14. Spiliopoulou M. Web usage mining for web site evaluation. Commun ACM. 2000; 43: 127–134.
15. Hung J. Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. Br J Educ Technol. 2012; 43: 5–16.
16. Liu X, Yu S, Janssens F, Glänzel W, Moreau Y, De Moor B. Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database. J Assoc Inf Sci Technol. 2010; 61: 1105–1119.
17. Glenisson P, Glänzel W, Janssens F, De Moor B. Combining full text and bibliometric information in mapping scientific disciplines. Inf Process Manag. 2005; 41: 1548–1572.
18. Radhakrishnan S, Erbis S, Isaacs JA, Kamarthi S. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. PloS One. 2017; 12: e0172778. https://doi.org/10.1371/journal.pone.0172778 PMID: 28328983
19. Névéol A, Doğan RI, Lu Z. Author keywords in biomedical journal articles. AMIA Annual Symposium Proceedings. American Medical Informatics Association; 2010. p. 537.
20. Yang S, Han R, Wolfram D, Zhao Y. Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. J Informetr. 2016; 10: 132–150.
21. Uddin S, Khan A. The impact of author-selected keywords on citation counts. J Informetr. 2016; 10: 1166–1177.
22. Su H-N, Lee P-C. Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. Scientometrics. 2010; 85: 65–79.

23. Aizawa A, Kageura K. Calculating association between technical terms based on co-occurrences in keyword lists of academic papers. Syst Comput Jpn. 2003; 34: 85–95.

24. Dotsika F, Watkins A. Identifying potentially disruptive trends by means of keyword network analysis. Technol Forecast Soc Change. 2017; 119: 114–127.

25. Mela CF, Roos J, Deng Y. A keyword history of marketing science. Mark Sci. 2013; 32: 8–18.

26. Gil-Leiva I, Alonso-Arroyo A. Keywords given by authors of scientific articles in database descriptors. J Assoc Inf Sci Technol. 2007; 58: 1175–1187.

27. Lee S. A Study on Research Trends in Public Library Research in Korea Using Keyword Networks. Libri. 2016; 66: 263–274.

28. Hristovski R, Aceski A, Balague N, Seifert L, Tufekcievski A, Cecilia A. Structure and dynamics of European sports science textual contents: Analysis of ECSS abstracts (1996–2014). Eur J Sport Sci. 2017; 17: 19–29. https://doi.org/10.1080/17461391.2016.1207709 PMID: 27460778

29. Champely S, Fargier P, Camy J. Disciplinarity and sport science in Europe: A statistical and sociological study of ECSS conference abstracts. Eur J Sport Sci. 2017; 17: 5–18. https://doi.org/10.1080/17461391.2016.1197318 PMID: 27344922

30. Kevork EK, Vrechopoulos AP. CRM literature: conceptual and functional insights by keyword analysis. Mark Intell Plan. 2009; 27: 48–85.

31. Jones KS, Jackson DM. The use of automatically-obtained keyword classifications for information retrieval. Inf Storage Retr. 1970; 5: 175–201.

32. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 2005; 102: 16569–16572. https://doi.org/10.1073/pnas.0507655102 PMID: 16275915

33. Nooy W de, Mrvar A, Batagelj V. Exploratory Social Network Analysis with Pajek. Edición: Expanded edition. England; New York: Cambridge University Press; 2012.

34. Peset F, Ferrer-Sapena A, Villamón-Herrera M, González LM, Toca-Herrera JL, Aleixandre-Benavent R. Scientific literature analysis of Judo in Web of Science®. Arch Budo 2013 Vol 9 Num 2 P 81–91. 2013;

35. Kamada T, Kawai S. An algorithm for drawing general undirected graphs. Inf Process Lett. 1989; 31: 7–15. https://doi.org/10.1016/0020-0190(89)90102-6

36. Waltman L, Van Eck NJ. A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys J B. 2013; 86: 471.

37. van Eck NJ, Waltman L, Dekker R, van den Berg J. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. J Assoc Inf Sci Technol. 2010; 61: 2405–2416.

38. Waltman L, van Eck NJ, Noyons EC. A unified approach to mapping and clustering of bibliometric networks. J Informetr. 2010; 4: 629–635.

39. Singer JD, Willett JB. It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. J Educ Stat. 1993; 18: 155–195. https://doi.org/10.3102/10769986018002155

40. Box-Steffensmeier JM, Cunha RC, Varbanov RA, Hoh YS, Knisley ML, Holmes MA. Survival analysis of faculty retention and promotion in the social sciences by gender. PloS One. 2015; 10: e0143093. https://doi.org/10.1371/journal.pone.0143093 PMID: 26580565

41. Schwartz MD. Fever in the returning traveler, part one: a methodological approach to initial evaluation. Wilderness Environ Med. 2003; 14: 24–32. PMID: 12659246

42. Pérez-Gutiérrez M, Lagos-Hernández RI, Izquierdo-Macón E. Sport sciences' scientific production published in Chile (1912–2014): A bibliometric approach. Movimento. 2016; 22: 1121–1136.

43. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985; 100: 126. PMID: 3920711

44. Stone MH, Sands WA, Stone ME. The Downfall of Sports Science in the United States. Strength Cond J. 2004; 26: 72–75.

45. Hall CM. Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. Tour Manag. 2011; 32: 16–27.

46. Nosek BA, Spies JR, Motyl M. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. Perspect Psychol Sci. 2012; 7: 615–631. https://doi.org/10.1177/1745691612459058 PMID: 26168121

47. Aguinis H, Gottfredson RK, Culpepper SA, Dalton DR, de Bruin GP. Doing good and doing well: On the multiple contributions of journal editors. Acad Manag Learn Educ. 2013; 12: 564–578.

48. McGinty S. Gatekeepers of Knowledge: Journal Editors in the Sciences and the Social Sciences. Warrendale, Pennsylvania: Bergin & Garvey; 1999.

49. Greendorfer SL. Specialization, fragmentation, integration, discipline, profession: What is the real issue? Quest. 1987; 39: 56–64.

50. Newell KM. Physical education in higher education: Chaos out of order. Quest. 1990; 42: 227–242.

51. Freeman WH. Physical education, exercise and sport science in a changing society. Burlington, MA: Jones & Bartlett Publishers; 2013.

52. Henry FM. Physical Education. J Health Phys Educ Recreat. 1964; 35: 32–69. https://doi.org/10.1080/00221473.1964.10621849

53. Bouchard C. Les sciences de lactivité physique: un concept fondamentale dans notre organisation disciplinare et professionnell. Mouvemment. 1974; 9: 117–129.

54. Lawson HA, Morford WR. The crossdisciplinary structure of kinesiology and sports studies: Distinctions, implications, and advantages. Quest. 1979; 31: 222–230.

55. Renson R. From physical education to kinanthropology: a quest for academic and professional identity. Quest. 1989; 41: 235–256.

56. Arnold PJ. Meaning in Movement, Sport and Physical Education. London: Heinemann Educational Publishers; 1979.

57. Cagigal JM. ¿ La educación física, ciencia? Citius Altius Fortius. 1968; X: 5–26.

58. Harris JC. Social contexts, scholarly inquiry, and physical education. Quest. 1987; 39: 282–294.

59. Renshaw P. The nature of human movement studies and its relationship with physical education. Quest. 1973; 20: 79–86.

60. Rikli RE. Kinesiology—A "homeless" field: Addressing organization and leadership needs. Quest. 2006; 58: 287–309.

61. Thomas JR. Are we already in pieces, or just falling apart? Quest. 1987; 39: 114–121.

62. Thomas JR. The public face of kinesiology in the 21st century. Quest. 2014; 66: 313–321.

63. Henry FM. The academic discipline of physical education. Quest. 1978; 29: 13–29.

64. Bressan ES. An academic discipline for physical education: What a fine mess! Proceedings National Association for Physical Education in Higher Education Annual Conference. 1982. pp. 26–27.

65. Hoffman SJ. Specialization+ fragmentation = extermination: A formula for the demise of graduate education. J Phys Educ Recreat Dance. 1985; 56: 19–22.

66. Freedson P. Back to the Future: Reflecting on the Past and Envisioning the Future for Kinesiology Research. Kinesiol Rev. 2014; 3: 1–3. https://doi.org/10.1123/kr.2014-0044

67. Reeve TG. Kinesiology: Defining the Academic Core of Our Discipline: Introduction. Quest. 2007; 59: 1–4. https://doi.org/10.1080/00336297.2007.10483531

68. Santos J, Irizo F. Modelling citation age data with right censoring. Scientometrics. 2005; 62: 329–342.

69. Santos J, Irizo F. Modelización de la antigüedad de las citas en la literatura científica con datos censurados a la derecha. Rev Esp Doc Científica. 2002; 25: 141–150.

13

Prediction equation for estimating cognitive function using physical fitness parameters in older adults

Gian Pietro Emerenziani[1]*, Maria Grazia Vaccaro[2], Giulia Izzo[1], Francesca Greco[1], Luca Rotundo[1], Roberto Lacava[3], Sandro La Vignera[4], Aldo Eugenio Calogero[4], Andrea Lenzi[5], Antonio Aversa[1]

1 Department of Experimental and Clinical Medicine, University of Catanzaro "Magna Græcia", Catanzaro, Italy, **2** Neuroscience Centre, Magna Graecia University, Catanzaro, Italy, **3** PST Umberto Primo, DSS CZ Ambulatorio Geriatria Disturbi Cognitive e Demenze, Catanzaro, Italy, **4** Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy, **5** Department of Experimental Medicine, Policlinico Umberto I, Sapienza University of Rome, Rome, Italy

* emerenziani@unicz.it

Abstract

Ageing is associated with declines in cognitive functions and physical fitness (PF). Physical exercise training and physical activity (PA) have been shown to have positive effects on cognitive functions and brain plasticity. This study aims to establish a practical equation for evaluating cognitive functions using PF parameters in healthy older adults. One-hundred and two older subjects were physically and clinically evaluated. Participants performed the Short Physical Performance Battery (SPPB) and handgrip test (HG); general cognitive functions were examined using the Mini Mental State Examination (MMSE). For all of them, a multiple regression analysis was used to predict MMSE from age, SPPB and HG variables. The new equation was cross validated to determine its prediction accuracy. Considering that SPPB and MMSE reference score are not different between genders, only one equation was developed for females and males. Age, SPPB and HG correlated significantly ($p<0.01$) with the MMSE score. The developed equation was MMSE = 19.479 + (1.548 x SPPB)–(0.130 x age) (R^2 = 0.72 and root mean square errors of 3.6). The results of PF are useful for exercise specialists to achieve the best physical exercise training and PA in older adults. In conclusion, this study showed for the first time that our new equation can be used to predict subjects' cognitive functions based on SPPB results and subject age. We suggest its use when patients' cognitive functions or more appropriate clinical tests cannot be pursued.

Editor: Giuseppe Curcio, University of L'Aquila, ITALY

Competing interests: The authors have declared that no competing interests exist.

Introduction

Between 2015 and 2050, the number of people over 60 will almost double from 12% to 22% [1]. As a result, the population structure will be changing in developed countries, with fewer children and more elderly people. Because of this change, the pace of population aging around the world is also increasing [1]. A longer life might bring new opportunities or disabilities depending on whether people can experience these additional years of life in a healthy or non-healthy condition. Indeed, if these added years are dominated by declines in physical and mental capacity the prevalence of many chronic conditions is expected to increase. Healthy ageing is defined as the process of developing and maintaining the functional ability that enables well-being in old age [1]. With increased longevity it is very important to prevent the age-related impairment of cognitive functions such as mild cognitive impairment (MCI) and dementia. Dementia, the end stage of brain diseases and Alzheimer disease (AD), is the most common one, while MCI is a heterogeneous state between normal ageing and early dementia. There are different methods to assess subjects' cognition state, however, the most common one, employed in 80% of studies, is the Mini-Mental State Examination (MMSE) or its modified version [2]. A healthy lifestyle, including correct diet, abstinence from cigarette smoking and regular physical activity (PA) has a pivot role for healthy aging. Through the manuscript we use the term "physical activity" to indicate any bodily movements, "exercise" to indicate a subset of physical activity characterized by a planned and purposeful training, and "physical fitness" (PF) to indicate a set of attributes that are health related [3]. Many studies support the idea that PA might be considered as non-invasive therapy for physical and mental health improvements [4–6]. For instance, Blair et al. (1989) showed that high level of PF appears to delay all-cause mortality decreasing the rates of cardiovascular diseases and cancer [7]. It has been recognized that healthy lifestyle might counteract physical and cognitive decline in subjects affected by illness or impairment. PA has been recognized as the stronger factor to counteract the development of AD [8]. PA might maintain or improve cognitive functions in ageing and reduce the risk of AD in subjects older than 35 years old [9]. Moreover, physical active behaviors and PE might produce benefit in executive functions, and memory counteracting cognitive aging [10,11]. However, to reach these positive health effects, volume, intensity, frequency, and the type of exercise should be planned to achieve the best clinical results. When, these exercise parameters are not planned correctly, it could be possible that exercise might induce health complications or the training goals will not be achieved. Recent data suggest that functional mobility impairment in cognitive dual task is correlated to cognitive decline in patients with AD [12]. Dual-task actions require simultaneous motor and cognitive tasks and they are frequently used during daily living activity. Scientific research has been focused on the effects of dual-task training in older adults with [13] and without [14,15] mild-to moderate dementia and with Parkinson's disease [16] showing an enhanced cognitive and physical functions after training. Moreover, Vaccaro et al. (2019) showed that dancing practice might increase fitness performance, memory functions and anxiety in older adults [17]. As previously reported, a decline in physical functions has been associated with cognitive decline [18]. Indeed, slow gait speed and weaker grip strength are strongly associated with worse cognitive performance [18]. Given that the evaluation of subjects' physical functions is usually a non-invasive and well tolerated procedure, it could be useful to consider it as an additional marker for the assessment of MCI to validated expensive instrumental tests, i.e. positron emission tomography and functional magnetic resonance imaging. Moreover, whether it could be possible to estimate the global cognitive functions from physical fitness tests, sport science specialist could optimize the training program (e.g. choosing the most appropriate type of exercise) in order to counteract the subjects' cognitive decline, using as example a dual-task training program. Therefore, the aim of this study was to establish a practical equation for evaluating cognitive functions using PF parameters in healthy older adults.

Materials and methods

Participants

One hundred and two older adults (65 females; 37 males) (age = 74.3±6.7 years, BMI = 28.3 ±4.0 kg/m^2) were recruited in this study from patients admitted to Geriatric Evaluation Unit for Cognitive Disorders—Azienda Sanitaria Provinciale Catanzaro and Endocrinology Unit—Department of Experimental and Clinical Medicine, University Magna Graecia, Catanzaro. Inclusion criteria consisted of older age (> 65 years). The exclusion criteria were: physical impairment, severe psycho-cognitive diseases (major depressive disorder or psychosis), any neuropathy or autonomic dysfunction, significant renal or liver disease, uncontrolled cardiovascular disease, i.e., myocardial infarction/myocardial ischemia or ventricular tachycardia/obstructive valvular heart disease during the previous 6 months, uncontrolled hypertension (blood pressure values exceeding 140 mm Hg systolic or 90 mm Hg diastolic), uncontrolled hyperglycaemia, thyroid disease including autoimmunity, or any treatment with thyroid hormone preparations or amiodarone, methimazole or propylthiouracil in the prior 3 months. All participants underwent clinical examination to exclude any contraindications to PA and were recruited according to their willingness to participate to the study protocol and signature of informed consent. Moreover, independent samples of forty-five subjects (26 females; 19 males) (age = 78.4±6.4 years, BMI = 28.1±4.7 kg/m^2) were selected for cross-validation analysis. These subjects were recruited using the same inclusion/exclusion criteria and from the same Centers. All tests were performed in the morning from 9:00 AM to 2:00 PM and MMSE was assessed with face-to-face interview by a trained physician. After that, subjects performed the Short Physical Performance Battery and Handgrip test in order to evaluate subjects' physical fitness. Each participant provided a written informed consent before entering the study. This study was conducted according to the Declaration of Helsinki and was approved by the Ethical Committee of the Magna Graecia University, (approval number 149, 2017) as an amendment to baseline screening evaluation included in Eudract protocol n. 2016-005198-11.

General cognitive functions, anthropometric and physical fitness assessment

Subjects' general cognitive impairment were assessed by using the standardized neuropsychological Mini Mental State Examination (MMSE) test [19]. Height (to nearest 0.01 cm) and weight (to nearest 0.1 kg) were measured using a stadiometer with weighting station. Body mass index (BMI) was calculated dividing body weight in kilograms by height in meters squared (kg/m^2). After a familiarization session, subjects performed the Short Physical Performance Battery (SPPB) [20] and Hangrip test (HG) [21]. The individual score of SPPB was derived from three functional tests that evaluate balance (Bal), lower body strength (CST) and gait speed (GS). The procedure is described in detail elsewhere [20]. Grip strength was measured using a JAMAR handheld dynamometer (BK-7498, Fred Sammons, Inc.) with participants seated, with their elbow by their side and flexed to right angles. The participants' hand grip strength data were evaluated as left or right according to the dominant hand (the hand used in performing heavy tasks or using heavy tools). Subjects performed three trials and the average of the three attempts was used for data analysis. To minimize the effects of fatigue 45 seconds of recovery time was allowed between each trial.

Statistical analysis

The Kolmogorov-Smirnov test was used to ensure normally distributed data. All data are presented as mean values ± standard deviation (SD). Differences between males and females were evaluated with an unpaired *t*-test. Correlation analysis was used to explore the relationships between MMSE and the physical fitness variables. Stepwise regression analysis was performed to identify which combination of significantly related variables would best predict MMSE measured by the interview. The coefficient of determination (R^2) and the SEE were estimated. The criterion for inclusion (addition and retention) of predictors was the highest R^2 model and the lowest SEE. Statistical significance was assumed at the conventional level of $p \leq 0.05$. In the current study, cross-validation of predicted equations was performed by using the root mean squared error (RMSE) methods [22] to an independent sample. RMSE is a measure of the performance of prediction equation when applied to an independent sample. It is calculated as the square root of the sum of squared differences between the observed and the predicted values divided by the number of subjects in the cross-validation sample. All statistical analyses were performed with the SPSS statistical package (Version 24.0 for Windows; SPSS Inc., Chicago, IL, USA).

Results

This section was divided by subheadings to provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

Cognitive functions, anthropometric and physical fitness results

Subjects' cognitive functions, anthropometric characteristics and physical fitness results are shown in Table 1. As expected, height, weight, and HG were significantly higher ($P < 0.01$) in male than in female subjects. No differences were observed for age, body mass index (BMI), SPPB and MMSE variables between males and females (Table 1).

Stepwise and multiple regression analyses between MMSE and independent variables

Considering that SPPB and MMSE reference score are not different between genders, we decided to develop only one equation for both female and male subjects. MMSE showed significant ($p<0.001$) negative correlation with age ($R = -0.532$) (Fig 1a) and significant ($p<0.001$) positive correlation with SPPB ($R = 0.841$) (Fig 1b) and HG ($R = 0.558$) (Fig 1c).

The results of the stepwise multiple regressions showed that age, and SPPB data can give the best predictive model ($R = 0.85$, $R^2 = 0.72$) as shown in Table 2.

Table 1. Subjects' anthropometrics characteristics, physical fitness and cognitive functions results. Data are presented as mean ± SD.

Parameters	Female (N = 65)	Male (N = 37)	Pooled (N = 102)	P value
Age (years)	73.4 ± 6.9	75.7 ± 6.1	74.3 ± 6.7	P = 0.10
Height (m)	1.52 ± 0.08	1.66 ± 0.07**	1.57 ± 0.09	P<0.01
Weight (kg)	66.5 ± 11.5	76.5 ± 11.1**	68.9 ± 12.7	P<0.01
BMI (kg/m)	28.6 ± 4.4	27.7 ± 3.2	28.3 ± 4.0	P = 0.27
CST (score)	2.9 ± 1.5	3.2 ± 1.1	3.0 ± 1.4	P = 0.28
GS (score)	2.8 ± 1.1	3.2 ± 1.0	3.0 ± 1.1	P = 0.10
Balance (score)	3.5 ± 0.9	3.5 ± 0.8	3.5 ± 0.9	P = 0.86
SPPB (score)	9.2 ± 3.1	9.9 ± 2.5	9.4 ± 2.9	P = 0.24
HG (score)	19.5 ± 6.8	26.4 ± 9.3**	22.0 ± 8.4	P<0.01
MMSE (score)	24.1 ± 6.4	25.1 ± 4.6	24.5 ± 5.8	P = 0.44

BMI = body mass index; CST = lower body strength; GS = gait speed; SPPB = Short Physical Performance Battery; HG = handgrip test; MMSE = Mini Mental State Examination;

** Statistically significant vs female ($P < 0.01$)

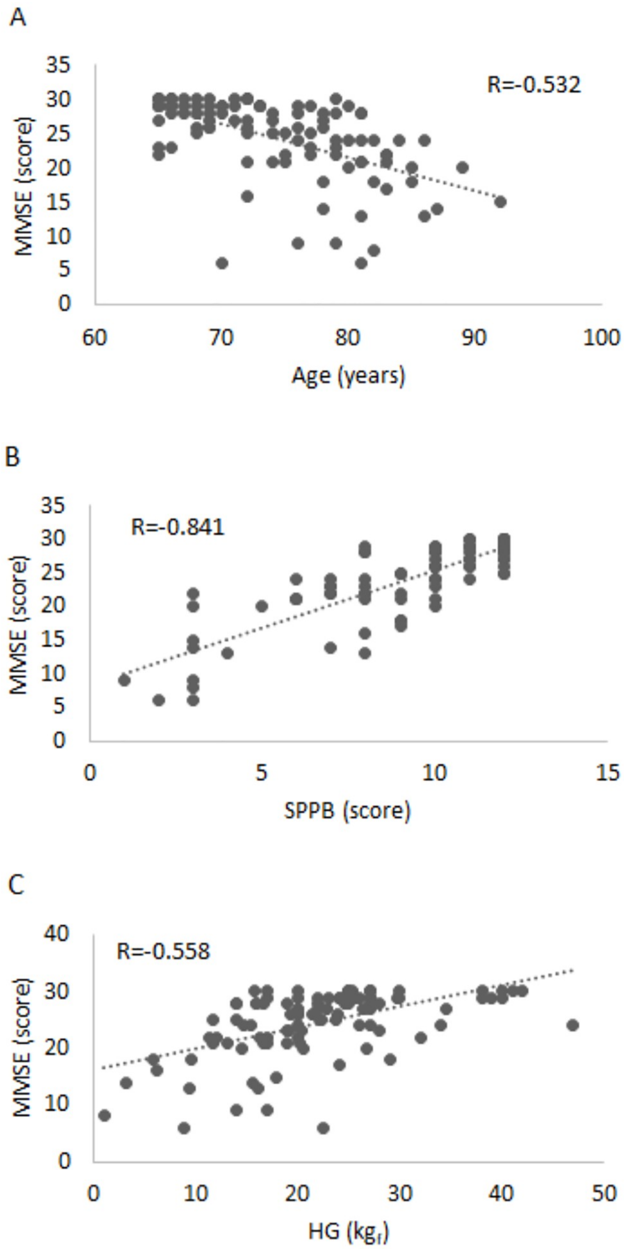


Fig 1. Correlation between Mini Mental State Examination (MMSE) score and age (a), Short Physical Performance Battery (SPPB) score (b) and Handgrip (HG) (c) in one hundred and two older adults (age range: 65 to 92 years).

Table 2. Stepwise regression analysis results.

Title 1	Coefficient	SE	R	SEE	P
Constant	19.479	4.812			
SPPB	1.548	0.127			
Age	-0.130	0.055			
Total model			0.850	3.1	<0.01

[1] SPPB = Short Physical Performance Battery

From the result of multiple regression analysis, the prediction equations to estimate MMSE is:

$$MMSE = 19.479 + (1.548 \times SPPB) - (0.130 \times age)$$

For cross-validation analysis, prediction equations were used on forty-five subjects (age = 78.4 ± 6.4 years, SPPB = 9.0 ± 2.9 score. Subjects' MMSE was 24.2 ± 5.6 score and predicted mean MMSE was 23.3 ± 4.9 score. The MMSE values for the RMSE were 3.0 score, therefore, RMSE was 13% of the range of target property value.

Discussion

In this study we evaluated the correlation between PF parameters and MMSE score to establish a simple and practical equation that may help to predict cognitive functions in older adults. The results showed for the first time that, age and SPPB could be used as predictive variables of MMSE in older adults of both genders. It is worthy to mention that the subjects enrolled in our study were physically healthy, without any severe physical acute problem and severe psychological diseases that could negatively influence the results of PF tests. Moreover, we should point out the attention to the claim of our study that was not to establish a new method for MCI diagnosis; rather, we aimed to estimate subjects' cognitive functions to prescribe the best PE protocols in older adults. In fact, more appropriate and validated methods are available for the evaluation of subjects' cognitive functions, i.e. positron emission tomography and functional magnetic resonance imaging. With the increasing of population aging, it is important to apply any tool that could lead to a healthy aging such as be involved in PE practice. However, different PE parameters such as intensity, frequency and the type of exercise might influence the training effects. Therefore, the use of our practical equation might give to the sport science specialist more details in order to choose the most appropriate type of physical exercise (e.g. dual task exercise instead of strength exercise). Physical inactivity is associated with increased risk of cardiovascular and metabolic diseases that in turn are associated with increased risk of dementia [23]. Increased level of PA and PF result in a 20% lower mortality rate [24]. Moreover, Hu and colleagues (2004) showed that physically inactive middle-aged women have a 52% increase of all cause of mortality when compared with physically active subjects [25]. Korpelainen and co-workers (2016) reported that exercise capacity is the strongest predictor of cardiovascular diseases and all-causes of mortality in both genders especially for cardiovascular deaths in women [26]. Moreover, Myers et al. (2002) showed that each one metabolic equivalent (1 MET = 3.5 ml/kg/min) increase in exercise capacity conferred a 12% survival improvement in men [27]. Lower extremity muscle efficiency is also important in delaying the onset of disability since it correlates with gait and balance [28]. Falls are one of the causes of morbidity and mortality in older adults and gait and balance are also strongly associated with the risk of falls [28]. Once again, it has been shown that PE might prevent falls in community-dwelling older people [28]. However, PA and exercise do not have only positive effects on physical health but also on psychological well-being and cognitive functions, decreasing symptoms of anxiety and depression [5,29]. Indeed, it has well known that PA and PE may have a positive effect on cognition in multiple sclerosis [30], depression [31], stress disorders [32], and Parkinson's disease [33]. The PE-related improvement in cognitive functions and psychological state seem to be associated with an increased expression of brain-derived neurotrophic factor (BDNF), glial cell-derived neurotrophic factor (GDNF) in some brain areas and insulin growth factor-1 (IGF-1) [34; 4]. BDNF is a growth factor expressed in the brain and throughout the rest of the central nervous system [35] and enhances the survival and differentiation of neurons, even dendritic arborization and synaptic plasticity [36]. Moreover, like BDNF, IGF-1 plays a fundamental role in many exercise-induced adaptations in the brain. The positive effects of PA on psychological state is also due to the increase of β-endorphin in peripheral blood resulted after exercise and it depends on the exercise intensity performed [37].

Low level of PF is linked to low cognitive performance and this relation could be explained by changes in the neurotrophic factors in the brain. Scientific literature showed that decreased physical performance is associated with poor cognitive functions [38]. Veronese and colleagues (2016) showed that slow walking speed precedes the onset of poor cognitive functions and that poor SPPB score is significantly associated with the onset of cognitive impairment in both genders. Moreover, chair standing time predicts the onset of cognitive impairment in females [38]. The authors [38], elucidated that one reason of the relation between gait speed and cognitive impairment should be that gait speed is closely associated with an impaired balance and fear of falling which has been associated with grey matter volume loss. Our results are in agreement with those reported by Veronese and co-workers; indeed, muscle strength and SPPB are positively correlated to MMSE showing that muscle strength and PF are correlated to subjects' cognitive functions. However, to reach these positive health and physiological effects, volume, intensity, frequency and the type of exercise should be planned to achieve the best clinical results.

Scientific evidence showed that both endurance and resistance training may lead to positive results on subjects' physical health by decreasing the risk of fall and by increasing the cardiovascular capacity and cognitive functions in older people [13,28,39]. Indeed, it is known that resistance-exercise training improves cognitive functions in healthy older adults [40] and that the types of PA might influence differently the structural and functional brain [41]. Recently, the number of studies on the effects of physical-cognitive dual task training on cognitive functions has increased [14,42]. Dual-task exercise requires a multitasking ability since subjects must simultaneously perform two tasks (physical and cognitive). For instance, subjects might walk while processing a cognitive task (e.g. counting backwards) simultaneously. As previously reported by Falbo and colleagues (2016) the addition of dual-task exercise to physical training enhance gait performance in general, suggesting to include dual-task exercise into physical training. To date, no equation allowing estimation of MMSE from SPPB and age in older adults has been published. The possibility to estimate the subjects' cognitive level in older adults might lead the physical exerciser specialists to choose the best training protocol to reach the greatest clinical positive effects. Our regression model might be useful and suitable to all professionals that work in interdisciplinary teams to realize and optimize PE intervention. Our results have shown that SPPB was the highest predictor of MMSE whit a correlation coefficient equal to 0.841. The second predictor was people age with a correlation coefficient of 0.532. As expected, PF (SPPB) and age variables are strongly correlated with MMSE. In fact, a high level of PF resulted in a positive while age in a negative relationship with MMSE, respectively. A lower SPPB score and higher age will result in a low MMSE, on the contrary, higher SPPB score and younger age will result in a high MMSE score. The standard error of our predicted equation was 3.1. As described by Alexander and co-workers (2015) [22], the coefficient of determination (R^2), the value of the root mean squared error (RMSE), and the use of an independent test set are recommended to characterize the external predictively of the model. In detail, values of $R^2 > 0,6$ and MRSE $< 10\%$ are suggested [22]. In our equation, R^2 value was 0.72 and RMSE value from the cross-validation results was 13% of the range of target property values. Although the well-known and more invasive (positron emission tomography) and expensive test (functional magnetic resonance imaging) remain the gold standard method for the assessment of cognitive impairments, this study suggests a valid alternative and an easier method to estimate MMSE when these methods are not available.

Nevertheless, when using this equation researchers and exercise professionals should be cautious to exclude older adult with physical and psychological diseases that might influence the SPPB results. We are aware of some important study limitations. For instance, different physical tests including other variables may be used for future studies to establish a new equation to estimate MMSE with RMSE lower than 10%. In addition, in our study the evaluation of subjects' cognitive level was not supported with diagnostic imaging tests. Moreover, our cohort was made up of physical healthy subjects without severe psychological disease that could influence the PF tests and that could be able to attend an exercise training program. For all these reasons, future studies may implement the current equation with new parameters and different healthy subjects' variables to achieve the highest correlation.

Conclusions

This is the first study aimed to establish a practical reference equation to estimate the MMSE in healthy older adults. SPPB and age might be used to predict MMSE in both genders. This practical reference equation may a valid and alternative method to evaluate the cognitive functions in elderly when gold standard methods are not applicable or available in clinical practice and it could be useful to the sport science specialist in order to choose the most appropriate type of exercise training.

Finally, this design suggests several clinical research perspectives. In the next studies it will be interesting to evaluate if there are different levels of oxidative stress [43] capable of interfering on the validation of this equation and contextually evaluate the correlation with the quality of sexual activity [44], as well as compare aerobic exercise patterns vs other types of exercise.

Supporting information

S1 File.
(PDF)

Author Contributions

Conceptualization: Gian Pietro Emerenziani.

Data curation: Gian Pietro Emerenziani, Maria Grazia Vaccaro, Giulia Izzo, Francesca Greco, Luca Rotundo, Roberto Lacava.

Formal analysis: Francesca Greco, Luca Rotundo.

Funding acquisition: Gian Pietro Emerenziani, Antonio Aversa.

Methodology: Sandro La Vignera, Aldo Eugenio Calogero, Andrea Lenzi.

Supervision: Roberto Lacava, Sandro La Vignera, Andrea Lenzi, Antonio Aversa.

Writing – original draft: Gian Pietro Emerenziani, Antonio Aversa.

Writing – review & editing: Maria Grazia Vaccaro, Giulia Izzo, Francesca Greco, Luca Rotundo, Roberto Lacava, Sandro La Vignera, Aldo Eugenio Calogero, Andrea Lenzi.

References

1. WHO 2- World Health Organization. 1st World Report on Ageing and Health, WHO. Geneva: WHO, 2015. https://www.who.int/ageing/events/world-report-2015-launch/en/
2. Demnitz N, Esser P, Dawes H, Valkanova V, Johansen-Berg H, Ebmeier KP, et al. A systematic review and meta-analysis of cross-sectional studies examining the relationship between mobility and cognition

in healthy older adults. Gait Posture 2016, 50, 164–174, https://doi.org/10.1016/j.gaitpost.2016.08.028 PMID: 27621086

3. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985, 100(2), 126–131. PMID: 3920711

4. Vina J, Sanchis-Gomar F, Martinez-Bello V, Gomez-Cabrera MC. Exercise acts as a drug; the pharmacological benefits of exercise. Br J Pharmacol 2012, 167(1), 1–12, https://doi.org/10.1111/j.1476-5381.2012.01970.x PMID: 22486393

5. Ruegsegger GN, Booth FW. Health Benefits of Exercise. Cold Spring Harb Perspect Med 2018, 8(7), pii: a029694, https://doi.org/10.1101/cshperspect.a029694 PMID: 28507196

6. Macera CA, Hootman JM, Sniezek JE. Major public health benefits of physical activity. Arthritis Rheum 2003, 49(1), 122–128, https://doi.org/10.1002/art.10907 PMID: 12579603

7. Blair SN, Kohl HW 3rd, Paffenbarger RS Jr, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. JAMA 1989, 262(17), 2395–2401, https://doi.org/10.1001/jama.262.17.2395 PMID: 2795824

8. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol 2011, 10(9), 819–828, https://doi.org/10.1016/S1474-4422(11)70072-2 PMID: 21775213

9. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015, 385 (9984), 2255–2263, https://doi.org/10.1016/S0140-6736(15)60461-5 PMID: 25771249

10. Young J, Angevaren M, Rusted J, Tabet N. Aerobic exercise to improve cognitive functions in older people without known cognitive impairment. Cochrane Database Syst Rev 2015, (4):CD005381, https://doi.org/10.1002/14651858.CD005381.pub4 PMID: 25900537

11. Hillman CH, Erickson KI, Kramer AF. Be smart, exercise your heart: exercise effects on brain and cognition. Nat Rev Neurosci 2008, 9(1), 58–65, https://doi.org/10.1038/nrn2298 PMID: 18094706

12. Borges SM, Radanovic M, Forlenza OV. Correlation between functional mobility and cognitive performance in older adults with cognitive impairment. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2018, 25(1), 23–32, https://doi.org/10.1080/13825585.2016.1258035 PMID: 27934540

13. Chen YL, Pei YC. Musical dual-task training in patients with mild-to-moderate dementia: a randomized controlled trial. Neuropsychiatr Dis Treat 2018, 14, 1381–1393, https://doi.org/10.2147/NDT.S159174 PMID: 29881275

14. Falbo S, Condello G, Capranica L, Forte R, Pesce C. Effects of Physical-Cognitive Dual Task Training on Executive Function and Gait Performance in Older Adults: A Randomized Controlled Trial. Biomed Res Int 2016, 2016:5812092, https://doi.org/10.1155/2016/5812092 PMID: 28053985

15. Norouzi E, Vaezmosavi M, Gerber M, Pühse U, Brand S. Dual-task training on cognition and resistance training improved both balance and working memory in older people. Phys Sportsmed 2019, 47(4), 471–478, https://doi.org/10.1080/00913847.2019.1623996 PMID: 31155997

16. Yogev-Seligmann G, Giladi N, Brozgol M, Hausdorff JM. A training program to improve gait while dual tasking in patients with Parkinson's disease: a pilot study. Arch Phys Med Rehabil 2012, 93(1), 176–181, https://doi.org/10.1016/j.apmr.2011.06.005 PMID: 21849167

17. Vaccaro MG, Izzo G, Ilacqua A, Migliaccio S, Baldari C, Guidetti L, et al. Characterization of the Effects of a Six-Month Dancing as Approach for Successful Aging. Int J Endocrinol 2019, 2019:2048391, https://doi.org/10.1155/2019/2048391 PMID: 31316562

18. Hooghiemstra AM, Ramakers IHGB, Sistermans N, Pijnenburg YAL, Aalten P, Hamel REG, et al. Gait Speed and Grip Strength Reflect Cognitive Impairment and Are Modestly Related to Incident Cognitive Decline in Memory Clinic Patients With Subjective Cognitive Decline and Mild Cognitive Impairment: Findings From the 4C Study. J Gerontol A Biol Sci Med Sci 2017, 72(6), 846–854, https://doi.org/10.1093/gerona/glx003 PMID: 28177065

19. Measso G, Cavarzeran F, Zappalà G, Lebowitz BD, Crook TH, Pirozzolo FJ, et al. The mini-mental state examination: Normative study of an Italian random sample. Developmental Neuropsychology, 1993, 9:2, 77–85, https://doi.org/10.1080/87565649109540545

20. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG et al. A short physical performance battery assessing lower extremity functions: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994, 49, 85–94, https://doi.org/10.1093/geronj/49.2.m85 PMID: 8126356

21. Wang CY, Chen LY. Grip strength in older adults: test-retest reliability and cut off for subjective weakness of using the hands in heavy tasks. Arch Phys Med Rehabil 2010, 91(11), 1747–1751, https://doi.org/10.1016/j.apmr.2010.07.225 PMID: 21044721

22. Alexander DL, Tropsha A, Winkler DA. Beware of R(2): Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. J J Chem Inf Model 2015, 55(7), 1316–1322, https://doi.org/10.1021/acs.jcim.5b00206 PMID: 26099013

23. Profenno LA, Porsteinsson AP, Faraone SV. Meta-analysis of Alzheimer's disease risk with obesity, diabetes, and related disorders. Biol Psychiatry 2010, 67(6), 505–512, https://doi.org/10.1016/j.biopsych.2009.02.013 PMID: 19358976

24. Myers J, Kaykha A, George S, Abella J, Zaheer N, Lear S, et al. Fitness versus physical activity patterns in predicting mortality in men. Am J Med 2004, 117(12), 912–918, https://doi.org/10.1016/j.amjmed.2004.06.047 PMID: 15629729

25. Hu FB, Willett WC, Li T, Stampfer MJ, Colditz GA, Manson JE. Adiposity as compared with physical activity in predicting mortality among women. N Engl J Med 2004, 351(26), 2694–2703, https://doi.org/10.1056/NEJMoa042135 PMID: 15616204

26. Korpelainen R, Lämsä J, Kaikkonen KM, Korpelainen J, Laukkanen J, Palatsi I, et al. Exercise capacity and mortality—a follow-up study of 3033 subjects referred to clinical exercise testing. Ann Med 2016, 48(5), 359–366, https://doi.org/10.1080/07853890.2016.1178856 PMID: 27146022

27. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE. Exercise capacity and mortality among men referred for exercise testing. N Engl J Med 2002, 346(11), 793–801, https://doi.org/10.1056/NEJMoa011858 PMID: 11893790

28. Sherrington C, Michaleff ZA, Fairhall N, Paul SS, Tiedemann A, Whitney J, et al. Exercise to prevent falls in older adults: an updated systematic review and meta-analysis. Br J Sports Med 2017, 51(24), 1750–1758, https://doi.org/10.1136/bjsports-2016-096547 PMID: 27707740

29. Kujala UM. Born to be rich, physically active, fit and healthy? Scand J Med Sci Sports 2010, 20(3):367, https://doi.org/10.1111/j.1600-0838.2010.01137.x PMID: 20598095

30. Beier M, Bombardier CH, Hartoonian N, Motl RW, Kraft GH. Improved physical fitness correlates with improved cognition in multiple sclerosis. Arch Phys Med Rehabil 2014, 95(7), 1328–1334, https://doi.org/10.1016/j.apmr.2014.02.017 PMID: 24607835

31. Mura G, Moro MF, Patten SB, Carta MG. Exercise as an add-on strategy for the treatment of major depressive disorder: A systematic review. CNS Spectr 2014, 19(6), 496–508, https://doi.org/10.1017/S1092852913000953 PMID: 24589012

32. Schoenfeld TJ, Rada P, Pieruzzini PR, Hsueh B, Gould E. Physical exercise prevents stress-induced activation of granule neurons and enhances local inhibitory mechanisms in the dentate gyrus. J Neurosci 2013, 33(18), 7770–7777, https://doi.org/10.1523/JNEUROSCI.5352-12.2013 PMID: 23637169

33. Mattson MP. Interventions that improve body and brain bioenergetics for Parkinson's disease risk reduction and therapy. J Parkinsons Dis 2014, 4(1), 1–13, https://doi.org/10.3233/JPD-130335 PMID: 24473219

34. Liu PZ, Nusslock R. Exercise-Mediated Neurogenesis in the Hippocampus via BDNF. Front Neurosci 2018, 12:52, https://doi.org/10.3389/fnins.2018.00052 PMID: 29467613

35. Salehi A, Delcroix JD, Mobley WC. Traffic at the intersection of neurotrophic factor signaling and neurodegeneration. Trends Neurosci 2003, 26(2), 73–80, https://doi.org/10.1016/S0166-2236(02)00038-3 PMID: 12536130

36. Park H, Poo MM. Neurotrophin regulation of neural circuit development and functions. Nat Rev Neurosci 2013, 14(1), 7–23, https://doi.org/10.1038/nrn3379 PMID: 23254191

37. Schwarz L, Kindermann W. Changes in beta-endorphin levels in response to aerobic and anaerobic exercise. Sports Med 1992, 13(1), 25–36, https://doi.org/10.2165/00007256-199213010-00003 PMID: 1553453

38. Veronese N, Stubbs B, Trevisan C, Bolzetta F, De Rui M, Solmi M, et al. What physical performance measures predict incident cognitive decline among intact older adults? Exp Gerontol 2016, 81, 110–118, https://doi.org/10.1016/j.exger.2016.05.008 PMID: 27235850

39. Angevaren M, Aufdemkampe G, Verhaar HJ, Aleman A, Vanhees L. Physical activity and enhanced fitness to improve cognitive functions in older people without known cognitive impairment. Cochrane Database Syst Rev 2008, (3):CD005381, https://doi.org/10.1002/14651858.CD005381.pub3 PMID: 18646126

40. Chang YK, Pan CY, Chen FT, Tsai CL, Huang CC. Effect of resistance-exercise training on cognitive functions in healthy older adults: a review. J Aging Phys Act 2012, 20(4), 497–517, https://doi.org/10.1123/japa.20.4.497 PMID: 22186664

41. Voelcker-Rehage C, Niemann C. Structural and functionsal brain changes related to different types of physical activity across the life span. Neurosci Biobehav Rev 2013, 37(9 Pt B), 2268–2295, https://doi.org/10.1016/j.neubiorev.2013.01.028 PMID: 23399048

42. Wollesen B, Voelcker-Rehage C. Training effects on motor-cognitive dual-task performance in older adults: a systematic review. Eur Rev Aging Phys Act 2014, 11, 5–24.

43. Sessa F, Messina G, Russo R, Salerno M, Castruccio Castracani C, Distefano A, et al. Consequences on aging process and human wellness of generation of nitrogen and oxygen species during strenuous exercise [published online ahead of print, 2018 Jun 27]. Aging Male. 2018;1–9. https://doi.org/10.1080/13685538.2018.1482866 PMID: 29950140

44. Duca Y, Calogero AE, Cannarella R, Giacone F, Mongioi LM, Condorelli RA, et al. Erectile dysfunction, physical activity and physical exercise: Recommendations for clinical practice. Andrologia. 2019; 51(5): e13264. https://doi.org/10.1111/and.13264 PMID: 30873650

14

Sport motivation and doping in adolescent athletes

Jiri Mudrak[1☯*], Pavel Slepicka[1☯], Irena Slepickova[2☯]

1 Department of Pedagogy, Psychology and Didactics, Faculty of Physical Education and Sport, Charles University, Prague, Czech Republic, **2** Department of Kinanthropology and Humanities, Faculty of Physical Education and Sport, Charles University, Prague, Czech Republic

☯ These authors contributed equally to this work.
* mudrak@ftvs.cuni.cz

Abstract

Background

Although performance-enhancing drugs appear to be prevalent in adolescent sports, relatively little attention has been paid to why adolescent athletes decide to use these drugs. In this study, we examine doping among adolescents from a motivational perspective and explore how motivational variables, such as achievement goal orientations and the perceived self-determination of sports activities, may be related to moral attitudes, doping intentions and doping behavior in adolescents who participate in competitive sports.

Methodology

The study included 1035 adolescents participating in competitive sports from all regions of the Czech Republic (mean age = 16.3 years). The respondents completed a battery of questionnaires assessing their achievement goal orientations (task, ego), sports motivation at various levels of self-determination (intrinsic motivation, external regulation, amotivation), moral attitudes toward sport competition (acceptance of cheating, keeping winning in proportion, attitudes toward doping), doping intentions and doping behavior. A structural equation model was used to test the relations among motivational variables, attitudes, intentions and doping behavior.

Principal results

Our analyses indicated a good fit with the proposed model, which explained 59% of the variance in doping intentions and 17.6% of the variance in doping behavior. Within the model, task orientation was positively associated with intrinsic motivation and lower amotivation, whereas ego orientation was positively associated with extrinsic regulation and amotivation. Furthermore, intrinsic motivation was positively associated with keeping winning in proportion and negatively associated with acceptance of cheating and attitudes toward doping; the less self-determined forms of motivation showed opposite relationships. However, only the acceptance of cheating and attitudes toward doping were related to doping intention, which subsequently predicted doping behavior.

Editor: Amir H. Pakpour, Qazvin University of Medical Sciences, ISLAMIC REPUBLIC OF IRAN

Competing interests: The authors have declared that no competing interests exist.

Conclusions/Significance

The results provide further evidence that sports motivation represents a psychological variable that should be considered in anti-doping policies, programs, and interventions aimed at the adolescent population because motivation was linked to the doping-related attitudinal variables and also partially mediated the effect of achievement goal orientations in this regard. On the basis of these results, we may argue that the focus on intrinsic enjoyment, self-referenced criteria of success and self-improvement may be related to more negative attitudes toward doping and cheating, lower doping intentions and less frequent doping behavior, whereas the emphasis on competition, comparison with others and external motivation appear to be related to the opposite outcomes.

Introduction

The abuse of performance-enhancing drugs (PEDs) represents a significant problem in both competitive and leisure sports. The use of PEDs violates the spirit of fair play [1] and represents a significant health concern because doping has been linked to a number of health issues, including cardiovascular, neurological, and psychiatric disorders [2–3]. The World Anti-Doping Agency [4] reported that approximately 1% of the tested samples from Olympic sports athletes and approximately 3% of the tested samples from non-Olympic sports athletes showed positive results for doping. However, these relatively low numbers contrast with the results of questionnaire surveys that suggest a much higher prevalence of doping: approximately 10%-15% of competitive and recreational athletes report past or current use of doping, with some studies suggesting even higher percentages [5]. (Doping represents an umbrella term encompassing PED use, blood doping, gene doping, etc. However, we use the term doping in the article to refer only to the use of PEDs.)

Adolescent athletes may be considered particularly vulnerable to the abuse of PEDs. From a health perspective, adolescent users are at high risk of the side effects of PEDs such as anabolic steroids [3, 6]. From a psychological perspective, adolescents are especially susceptible to social pressures and expectations regarding sports competition and physical appearance [7] and tend to participate in risky behavior with possible harmful long-term effects [8]. A large-scale international meta-analytic study [9] observed that approximately 3%-6.5% of boys and 1%- 2% of girls reported current or past use of anabolic steroids. Other national surveys have found that, depending on the methodology used, 2.1%-11% of adolescents reported past or current use of PEDs [10–14].

A number of behavioral and psychological factors have been related to PED abuse in adolescents. Adolescent users of PEDs report more positive attitudes toward doping, show higher levels of moral disengagement toward doping and perceive higher approval of doping abuse by other people [12, 15]. Adolescent users of PEDs also report lower self-confidence and lower status in their peer group [7] and experience higher levels of anxiety [16], more frequent depression [17], lower self-regulation [12, 15], and more frequent use of other addictive substances, such as alcohol, tobacco, and hard drugs [11, 13]. Adolescent users of PEDs also experience more frequent eating disorders [17] and engage in other types of risk behavior, ranging from school absences [11] to membership in violent groups [18].

Two major motivations for adolescents' use of PEDs have been discussed in the literature. First, adolescents use PEDs because they strive for physical attractiveness [7], which appears to be an especially dominant motive among adolescent athletes not engaged in competitive sports

[2, 14]. For example, Sas-Novosielski [14] observed that a majority of adolescent PED users predominately strove for a "better body" with the primary goal of gaining muscle and losing body fat. Although more than half of the participants reported side effects from the substances (such as acne, hair loss, depression, and sexual disorders), they insisted that they would continue to use PEDs to improve their physical appearance. Second, adolescent athletes use PEDs to obtain a competitive advantage and succeed in sports competition. It appears that a focus on victory and success in competition has become a dominant discourse even in youth sports, which has increased the incidence of problematic behavior such as cheating and doping [19]. Motivational orientations that emphasize competitive performance and "winning at all costs" have been related to positive attitudes toward doping as well as toward doping behavior [20–21]. Although adolescent athletes generally report negative attitudes toward doping, they sometimes admit that they would be willing to use PEDs to develop their professional athletic careers [22].

Numerous research studies have suggested that attitudes toward doping, intentions to dope and actual doping abuse are significantly influenced by sports motivation; i.e., the subjective reasons underlying why athletes participate in sports affect the decision to use PEDs [23–30]. However, there are some limits to current research on the relationship between sports motivation and doping. Despite a recent surge of interest in this research topic [23–28], studies focusing on the relationships between sports motivation, doping-related attitudes, intentions and behavior in adolescents remain limited. In addition, some of the studies have investigated relatively small samples in the context of selected sports [27, 31], further limiting the generalizability of the current findings. Finally, there have been calls for more thorough implementation of coherent theoretical frameworks in doping research that would enable a better understanding of the psychological processes underlying doping behavior in adolescence [25, 31]. To address these limitations, this study's primary goal was to further explore the motivational perspective on doping in a large sample of Czech adolescent athletes participating in competitive sports. In this manner, we integrate some key ideas from the achievement goal theory, self-determination theory and integrative model of behavioral prediction [32–33], and we postulate a series of relationships among achievement goal orientations, sports motivation, sports-related moral attitudes, doping intentions and actual doping behavior. On the basis of these hypotheses, we formulate a structural model, which we test within the structural equation-modeling framework.

Theoretical framework

To understand the complexity of the psychosocial influences that determine an intentional behavior, Fishbein and Capella [33] proposed hierarchical relationships among behavior, an intention to carry out the behavior and behavior-related attitudes. Those authors suggested that "any given behavior is most likely to occur if one has a strong intention to perform the behavior, has the necessary skills and abilities required to perform the behavior, and there are no environmental or other constraints preventing behavioral performance" [33]. From this perspective, behavior-related intention is the direct determinant of a behavior, and we should strive to understand the factors that influence how an individual's intentions to carry out a behavior are formed. Numerous studies have shown that doping-related attitudes represent a significant predictor of doping intentions [24, 27, 29]. However, we may hypothesize that the key attitudes related to doping intentions may be broader in scope and include more general attitudes toward cheating and winning in sports competition [19, 27] because the doping represents an instance of a cheating behavior [27] and has been related to increased emphasis on competition and winning in youth sports [19]. The attitudinal variables may then be considered proximal predictors of doping intentions that mediate the effects of other distant variables, including motivational beliefs [21, 24–25, 27].

To further explore the mediating role of attitudes in the relationship between sports motivation and doping, we adopted two well-established theories of sports motivation in our research. These theories include the achievement goal theory [14, 34–37] and the self-determination theory [25, 27, 38]. Specifically, based on these sports motivation theories, we consider task and ego achievement goal orientations and different positions on the self-determination continuum, including intrinsic motivation, external regulation, and amotivation, as possible predictors of doping attitudes and sports-related moral attitudes that further predict doping intentions and behavior. Below, we provide an introduction to these sports-motivation theories and their possible associations with these attitudes, intentions and behavior.

Based on self-determination theory [38], we expected that motivational states characterized by different levels of self-regulation (i.e., intrinsic motivation, external regulation and amotivation) may have different effects on moral and doping-related attitudes and behavior [23–25, 27]. Self-determination theory suggests that people strive to fulfill several basic psychological needs, such as the needs for autonomy, inner organization and better relationships with others. These basic needs are predominantly manifested by "intrinsically motivated behavior" or behavior that people engage in for its own sake, such as for the enjoyment stemming from the activity itself. The other end of the self-determination spectrum is represented by "externally regulated behavior," which people engage in for external reasons, such as obtaining a reward or avoiding punishment. The least self-regulated motivational state is "amotivation," in which people perceive a lack of self-regulation and personal agency toward the behavior. On the basis of extensive research, Deci and Ryan [38] asserted that engagement in intrinsically motivated behavior (as opposed to extrinsically regulated or amotivated behavior) is related to a range of positive outcomes, such as better performance, better relationships, and a higher level of well-being. With regard to doping, researchers have found that motivational states with higher self-determination were negatively related to attitudes toward doping [25], doping intentions [24], and past doping use [23], whereas external regulation was associated with moral disengagement in sports situations [27] or positive attitudes toward doping [28]. Furthermore, several other studies found that less self-determined forms of motivation predicted antisocial moral values and a lack of sportspersonship in athletes, including acceptance of cheating and winning-at-all-costs attitudes [39, 40]. On the basis of this research, we may argue that athletes who experience low levels of self-determination in their sport participation are lacking in some of their basic psychological needs and may compensate for this deficiency with more positive doping-related attitudes, intentions and behavior.

By contrast to self-determination theory, which focuses on why people engage in an activity, another group of motivational theories explores how different people subjectively prefer different achievement outcomes. Achievement goal theory [34, 41–43] conceptualizes these achievement outcomes through the dichotomy of "success in comparison with past performance" and "success in comparison with others." In this framework, a subjective preference for one of these two dimensions has been determined to have different effects on achievement-related beliefs, choices, intentions and behavior. Various authors proposed different terms for these two dimensions, such as task-ego [34] and mastery-performance [42] orientations. It is important to note that in the context of doping research, the 2x2 model of achievement goal orientation has been used [23–24] to distinguish between two dimensions: mastery x performance and approach x avoidance (i.e., striving for success versus avoiding failure). However, effects on doping intentions and behavior have been observed in the mastery/performance rather than the approach/avoidance dimension, with mastery-oriented athletes showing lower doping intentions and behavior [23]. In our research, we applied the more traditional distinction of task-ego, which has also been used in the doping research [14, 36]. These two dimensions appear to be relevant in the context of doping: a negative relationship has been identified between the orientation toward improving past performance (task, mastery) and doping-related intentions, attitudes and behaviors, whereas the orientation toward comparison with other people (ego, performance) generally showed the opposite relationships [14, 24, 35–36, 44]. It seems that “task”-related goals do not provide an incentive for doping because these types of goals may be achieved solely through practice, and task-oriented athletes may be not motivated to expose themselves to the health and social risks related to doping. Conversely, “ego”-related goals appear to be supportive of doping because the standards of performance based on other athletes are much more difficult (or even impossible) to achieve, and athletes are more motivated to use immoral or even illegal practices to reach these goals [35–36, 44].

We may argue further that a link exists between the achievement goal orientations and self-determination [38, 45–47]. Task and ego goal orientations may be seen as different interpretative frameworks that influence the ways in which athletes perceive their autonomy, competence and relatedness to others. In this way, these achievement goal orientations affect the degree to which the athletes perceive themselves as self-determined [45–47]. For example, Duda and her colleagues [46] provided evidence that task orientation (as opposed to ego orientation) predicted sports-related intrinsic motivation in youth athletes. These authors argued that task orientation reduced the probability that athletes would perceive themselves to be incompetent because they compared themselves with self-referenced standards of achievement rather than standards set by other athletes, some of whom inevitably performed at a higher level. Similarly, other authors argued that task-oriented athletes experience fewer social constraints on their autonomy because they do not feel obliged to meet the performance standards set by other people and also experience better relationships because they do not perceive themselves to be in competition with others [45–47].

Aim of the study

Based on the theoretical framework introduced above, we formulated a set of hypotheses regarding the relationships among the constructs of achievement goal orientation, self-determined sport motivation, attitudes, intentions, and doping behavior. We empirically tested these hypotheses within the structural equation modeling framework on a large sample of Czech adolescents involved in competitive sports. The implemented structural model was based on the following hypotheses:

H1: Task orientation is positively related to intrinsic sports motivation and negatively related to less self-determined forms of sports motivation (amotivation, extrinsic motivation).

H2: Ego orientation is positively related to less self-determined sports motivation (amotivation, extrinsic motivation) and negatively related to intrinsic motivation.

H3: Intrinsic motivation is negatively related to attitudes toward doping and acceptance of cheating and positively related to attitudes toward keeping winning in proportion.

H4: External regulation and amotivation are positively related to attitudes toward doping and acceptance of cheating and negatively related to attitudes toward keeping winning in proportion.

H5: Attitudes toward doping, acceptance of cheating, and keeping winning in proportion are directly related to doping intentions.

H6: Doping intentions are directly related to doping behavior.

Methods

Design of the study

The present paper is a component of the research project "Doping in Czech adolescents: Prevalence, correlates and experiences," which was conducted with the support of the World Anti-Doping Agency. The data were collected from November 2014 –May 2015. The main part of the data collection occurred at high schools and elementary schools throughout the Czech Republic. The data collection at schools was facilitated by the Czech Association of School Sport Clubs, a nationwide educational organization that works with children and adolescents who are engaged in sports. Additionally, competitive and elite adolescent athletes were contacted through various Czech sports associations. In total, 60 schools and 7 sports associations participated in the research. Based on the preferences of the schools and sports associations, the questionnaires were administered either in paper form or by identical electronic questionnaires. The questionnaires were administered at schools or at training camps of the sports associations by the research team members and research assistants, who were PhD students of the Faculty of Physical Education and Sport, Charles University. Before the beginning of the data collection, the research was approved by the ethics committee of the Faculty of Physical Education and Sport, Charles University. The data collection was voluntary and anonymous, and the questionnaire was constructed in a way that prevented the identification of individual schools or respondents. Because the questionnaires were collected at the schools during school hours, the response rate was high (95%). Prior to the data collection, the children's parents/guardians were informed of the research by the assisting teachers/coaches and requested to provide written consent. Approximately 5% of the contacted students did not participate in the data collection either because the parents/guardians did not provide consent or because the students were not willing to participate; these students were provided with an alternative activity under the supervision of the assisting teachers/coaches during the data collection.

Sample

In total, we collected fully completed questionnaires from 2851 respondents (mean age 16.2 years, SD = 1.84). However, in the present article, we based our findings only on participants involved in competitive sports (n = 1035). The description of the sample is provided in Table 1.

Measures

In the first section of the questionnaire, the respondents were asked about demographic variables such as gender, age, and type of school and about their participation in sports (see

Table 1. Demographic description of the respondents.

Demographic variable		
Age (years)	M (SD)	16.3 (1.92)
Gender	Male	64.4%
	Female	35.5%
Type of school	Elementary	32.6%
	Vocational	2.4%
	Secondary	42.1%
	Grammar	22%
	Other	.9%
Attending a sports school	Yes	25.1%
	No	74.9%

Table 1). In the following section, the respondents were asked about their experiences with doping. The World Anti-Doping Agency defined doping as "breaking one or more anti-doping rules," meaning that athletes who were found to be "positive" either used substances or methods present on the list of banned substances or were not compliant with doping control regulations [48]. On the basis of the WADA definition, some studies have examined the prevalence of doping by inquiring about the substances respondents used in the past that were subsequently classified according to the WADA list [15]. However, we deemed this approach appropriate for the population of adult athletes but not for adolescents. For the purpose of our study, we defined doping in the questionnaire as the "use of any substance which aims to enhance sport performance artificially and unfairly." Therefore, we explored subjective evaluations of the respondents' experiences with doping; i.e., we measured the extent to which the participants *believed* they doped to gain an unfair competitive advantage rather than measuring actual doping behavior. The respondents evaluated the frequency of their perceived experiences with doping on a six-point scale ranging from 1 (no) to 6 (yes, regularly). Similar research methods for doping prevalence have been implemented by other studies [13].

To assess the respondents' attitudes toward doping, we used the Performance Enhancement Attitude Scale (PEAS) [49]. The PEAS is a one-dimensional 17-item scale measuring general attitudes toward doping in sports (unrelated to personal intentions to use doping). In the PEAS, respondents indicate on a 6-point Likert scale ranging from 1 ("completely disagree") to 6 ("completely agree") their agreement with statements evaluating various aspects of doping, such as "Doping is not cheating since everyone does it," "Athletes are pressured to take performance-enhancing drugs," or "The risks related to doping are exaggerated." In scoring the PEAS, the overall score is obtained as the mean of all items. Overall, the PEAS shows good psychometric properties [47] and has been used in studies focusing on adolescent populations [15]. In our study, the PEAS showed acceptable reliability (Cronbach's alpha = .755), although it did not originally fit well with our model. Confirmatory factor analysis of the one-dimensional scale showed poor fit indices (RMSEA = 0.075; 90% CI [0.069 to 0.081]; CFI = 0.826). We performed exploratory factor analysis of the scale that suggested 4 factors, which we labeled "Necessity of doping," "Harmlessness of doping," "Recreational drugs as doping," and "Doping in media." The four-factor model was further supported by confirmatory factor analysis that showed good fit (RMSEA = 0.054; 90% CI [0.047 to 0.060]; CFI = 0.926). In the structural equation model, we used a 7-item shortened version of the PEAS that included only items with factor loadings in the first factor higher than .5 (i.e., items 1–2, 6, 8, and 13–15 from the original scale).

To measure doping intentions, we utilized four items from an older Czech study focusing on the doping intentions of Czech adolescents [50]. The respondents answered on a scale ranging from 1 ("definitely not") to 6 ("definitely yes") regarding whether they would use doping in four hypothetical situations: 1) "Would you use doping if you strove for an important victory and were absolutely certain that nobody would find out?" 2) "Would you take a performance-enhancing substance that is not illegal but could have undesirable health effects?" 3) "Would you use doping if you were certain that it would help you succeed and would not have undesirable health effects?" and 4) "Would you use doping to enhance your performance if you knew that it would help you to achieve the highest level of sports success, such as winning the Olympic games?" On this basis, we computed doping intention as the mean of these four items. This scale showed good reliability (Cronbach's alpha = .873). Furthermore, we used the Acceptance of Cheating and Keeping Winning in Proportion scales from the Attitudes to Moral Decisions in Sport Questionnaire [51] to measure general moral attitudes in sports situations. On these scales, respondents are asked to indicate on a 5-point Likert scale (from 1 – strongly agree to 5 –strongly disagree) how much they agreed with statements presenting them with sports situations, including a moral dilemma such as "It is OK to cheat if nobody knows?" or "Winning and losing are a part of life." The Acceptance of Cheating scale appeared to have good reliability (Cronbach's alpha = .895); the Keeping Winning in Proportion scale also had acceptable reliability (Cronbach's alpha = .746).

To assess motivation-related constructs, we used selected scales from two questionnaires: the Perception of Success Questionnaire (PSQ) [52] and the Sport Motivation Scale-6 (SMS-6) [53]. The PSQ measures achievement goal orientations on a 5-point Likert scale, i.e., which types of sport-related outcomes are perceived as "success" by the respondents. The PSQ stems from the two-dimensional approach to achievement goal orientations and distinguishes between success stemming from mastering a task ("task") and success in comparison with others ("ego"). Respondents indicate on a scale ranging from 1 ("strongly disagree") to 5 ("strongly agree") the sports situation in which they perceive themselves to be most successful. An example of the "task" item is "When playing sport, I feel most successful when I really improve," and an example of an "ego" item is "When playing sport, I feel most successful when I am the best." The "task" and "ego" dimensions are computed as the means of all corresponding items. Additionally, the PSQ showed good reliability in our study (Cronbach's alpha = .857).

To measure the reasons why respondents participate in sports, we implemented several constructs based on self-determination theory [54]. Specifically, we used the dimensions of intrinsic motivation, external regulation, and amotivation from the SMS-6 [53], which represents a revised version of the Sport Motivation scale [54]. Each of these dimensions was measured by four items on a 5-point Likert scale ranging from 1 ("Does not correspond at all") to 5 ("Corresponds completely"). Respondents were prompted by the statement, "Using the scale below, please indicate to what extent each of the following items corresponds to one of the reasons for which you are presently practicing your sport", based on which the respondents indicated their reasons for their participation in sports. The items used to measure each dimension included "For the excitement I feel when I am really involved in the activity" (intrinsic motivation), "Because it allows me to be well regarded by people I know" (external regulation), and "I don't know anymore; I have the impression of being incapable of succeeding in this sport" (amotivation). The SMS-6 questionnaire showed good psychometric properties (Cronbach's alpha = .888 in our sample) and has been widely used in sports psychology research [53]. All scales and items used are provided in S1 Appendix.

Analysis

We tested the hypothesized relationships within a structural equation modeling (SEM) framework using the statistical open source software R [55] and Lavaan, an R structural equation modeling package [56]. Only data from complete questionnaires were included in the analysis; therefore, there were no missing values. No outliers were identified in the data, and all the reported coefficients from our analyses were standardized. The model fit was assessed using standard measures of model fit: the chi-square statistic and corresponding p-value; the standardized root mean square residual (SRMR, which should approximate or be less than .08 for a good-fitting model) [57]; the root mean square error of approximation (RMSEA, with values approximately .05 or less being indicative of a close fit and values of .08 or less being indicative of a good fit) [58]; and the comparative fit index (CFI, in which values should be higher than 0.90 for adequately fitting solutions) [59].

Results

Descriptive statistics and correlations of all variables included in the analysis are reported in Table 2. We observed significant but rather weak correlations among the majority of the

Table 2. Descriptive statistics and correlations.

	1. Task orientation	2. Ego orientation	3. Intrinsic motivation	4. External regulation	5. Amotivation	6. Winning in proportion	7. Acceptance of cheating	8. Attitudes toward doping	9. Doping intention	10. Doping behavior
1.	-									
2.	.341**	-								
3.	.396**	.237**	-							
4.	.110**	.315**	.460**	-						
5.	-.103**	ns	.ns	.259**	-					
6.	.259**	ns	.190**	-.091**	-.165**	-				
7.	-.125**	.243**	ns	.230**	.346**	-.197**	-			
8.	ns	.132**	ns	.135**	.204**	-.107**	.467**	-		
9.	ns	.250**	ns	.123**	.187**	-.132**	.663**	.513**	-	
10.	ns	.100**	.084**	.210**	.207**	-.107**	.346**	.265**	.391**	-
M	4.33	3.66	3.45	2.65	1.73	3.88	2.01	2.40	2.43	1.18
SD	.60	.80	.85	.94	.81	.83	.88	.78	1.15	.639

** Correlation significant at $p < .001$ level.

ns—correlation not significant

variables included in the analyses. There were moderate to strong correlations among the motivational variables, such as task orientation-intrinsic motivation (r = .396) and ego orientation-external regulation (r = .315). Additionally, we observed strong correlations between doping intention and attitudes toward doping (r = .513) and between doping intention and acceptance of cheating (r = .663).

Based on our hypotheses, we formulated a structural model in which achievement goal orientations (task and ego) predicted sport motivation at different levels of self-determination (intrinsic motivation, extrinsic regulation and amotivation). The sports motivation was then associated with moral- and doping-related attitudes (attitudes toward doping, acceptance of cheating, and keeping winning in proportion) that were further related to doping intentions, which subsequently predicted doping behavior. Overall, we determined that the SEM model fit well with our data ($\chi^2 = 19789.3$; df = 946; $p < 0.001$; RMSEA = 0.042; 90% CI [0.040 to 0.044]; SRMR = 0.055; CFI = 0.913) and explained 59% of doping intentions and 17.6% of doping behavior. The measurement loadings for all latent variables were moderately high to high (range, 0.419–0.895) and highly significant ($p < .001$). The SEM model is presented in Fig 1. All measurement loadings are provided in S1 Appendix.

In accordance with our hypotheses, we observed that task and ego goal orientations were inversely related to sports motivation variables. Task orientation was linked to more self-determined sports motivation because it was moderately to strongly positively related to intrinsic motivation (β = .474) and negatively related to amotivation (β = -.185). At the same time, ego orientation was moderately to strongly positively related to extrinsic regulation (β = .395) and to amotivation (β = .103). Within the SEM model, the sports motivation was further associated with the attitudinal variables. Specifically, intrinsic motivation was negatively related to attitudes toward doping (β = -.180) and acceptance of cheating (β = -.260) and strongly and positively related to keeping winning in proportion (β = .607). Extrinsic regulation was positively associated with acceptance of cheating (β = .388), attitudes toward doping (β = .251), and negatively associated with keeping winning in proportion (β = -.510). Amotivation was positively associated with acceptance of cheating (β = .249) and attitudes toward doping (β = .161). The sports motivation variables also appeared to moderate the indirect effects of achievement goal

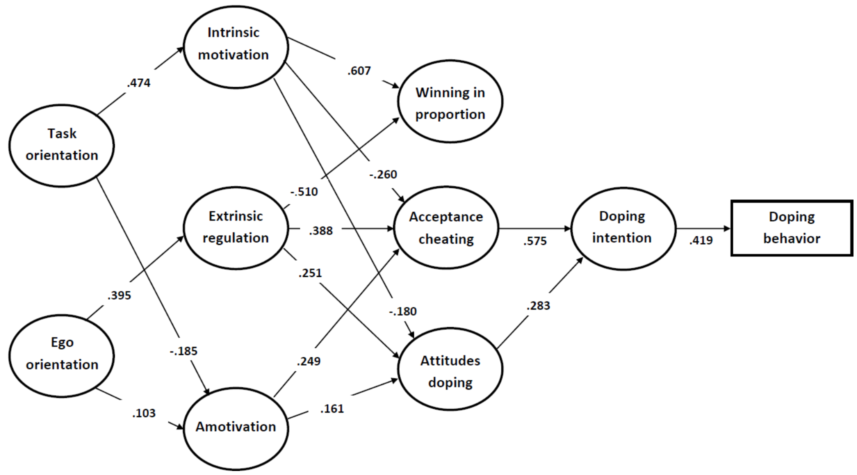


Fig 1. SEM model.

orientation on the attitudinal variables. Specifically, we observed a negative indirect effect of task orientation on the acceptance of cheating (β = -.169), attitudes toward doping (β = -.115), and positive indirect effects on keeping winning in proportion (β = .288). In ego orientation, we observed a reversed direction of relationships because ego orientation was indirectly and positively related to the acceptance of cheating (β = .179), attitudes toward doping (β = .116), and indirectly and negatively related to keeping winning in proportion (β = -.202). Furthermore, the acceptance of cheating and attitudes toward doping were moderately to strongly related to doping intention, with acceptance of cheating having approximately twice the effect (β = .575) of attitudes toward doping (β = .283). Notably, keeping winning in proportion was not significantly related to doping intention. This suggests that excessive focus on winning may not be as important for the occurrence of doping behavior as athletes' positive attitudes toward doping and cheating in sports. Finally, doping intention was moderately to strongly related to doping behavior (β = .419) within the SEM model.

Discussion

As we hypothesized, the proposed model showed a good fit, and the observed relationships largely confirmed our expectations regarding the possible effects of sports motivation variables on doping attitudes, intentions and behavior in adolescent athletes. From the model, we may infer several main findings. First, doping intention may be perceived as an important predictor of doping behavior in competitive adolescent athletes, although not a perfect predictor. Overall, our model explained nearly the same amount of variance in doping behavior as an aggregate model based on meta-analysis of studies stemming from the theory of planned behavior, in which doping intentions were also used as the proximal predictor of doping behavior [5]. As argued by Fishbein [33], behavior-related intentions are particularly predictive of behavior when people have an opportunity to engage in the behavior and no environmental constraints are present. This is, of course, not a case of doping in adolescent athletes because this group can be expected to have limited access to doping and also perceive severe penalties related to doping. Therefore, we may argue that other variables also should be considered to explain doping behavior; simultaneously, however, doping intentions represent an important factor that should be targeted in doping prevention [24, 27, 29].

Second, a large portion of the variance in doping intentions was explained by the attitudinal variables included in the model. By contrast to the studies focusing solely on attitudes toward doping [12, 15, 24–25, 28–29, 31, 60], we broadened our scope and also included attitudes toward winning and cheating in sports as complementary attitudinal variables. These additions appeared to be particularly productive in the case of the acceptance of cheating. Hodge et al. [27] suggested that doping should be approached as an instance of cheating behavior; our findings corroborate this idea because the association between the acceptance of cheating and doping intentions was more than twice the size of the attitudes toward doping–the doping intention relationship. In any case, we may see both of these attitudinal variables as related, which has been supported by our results as well as by other studies [60]. Preventive anti-doping programs frequently target attitudes toward doping [61–63], and our findings suggest that focusing on broader moral values rather than on doping-specific attitudes may represent a more effective manner of understanding and preventing doping behavior. However, contrary to some authors, who suggested that the growing focus on winning in contemporary youth sports leads to more frequent occurrences of doping [19], the attitudes toward winning (i.e., keeping winning in proportion) did not show a significant relationship with doping intentions in the model. On this basis, we may argue that the doping may stem not from the focus on winning itself but rather from the growing acceptance of cheating and more positive attitudes toward doping that may be related to contemporary trends in (youth) sports [63].

Third, the intrinsic motivation showed negative relationships between attitudes toward doping and acceptance of cheating whereas the less self-determined forms of motivation showed relationships moving in the opposite direction. On the basis of the self-determination theory, we may argue that athletes who engage in competitive sports for enjoyment have satisfied through sports their "basic needs" of autonomy, competence and relatedness [38]. Therefore these athletes may place less value on the behavior that would provide them with further unfair advantages, such as doping or cheating. Conversely, athletes at a lower level of self-determination who experience a lack in some of these basic needs could be expected to have less restraint and demonstrate more positive attitudes toward these undesirable practices [39]. Similar results were reported by other authors: Zucchetti et al. [28] found that extrinsic motivation was related to more positive attitudes toward doping, and Chan et al. [25] observed that autonomous motivation predicted doping avoidance-related attitudes and, indirectly, the intention to avoid doping. Barkoukis et al. [23] determined that athletes high in amotivation reported higher doping intentions and higher past PED use whereas athletes high in external regulation reported higher past use of PEDs compared with other athletes. Therefore, we may argue that the positive effects of self-determined motivation [38] apply also to doping-related attitudes, intentions and behavior and that sports environments supporting such a positive motivational climate should be endorsed as a component of anti-doping efforts.

Fourth, sports motivation also mediated the effect of achievement goal orientations on the attitudinal variables within the model. Significant effects of achievement goal orientations on doping-related variables have been observed in numerous other studies: Barkoukis et al. [23] observed that athletes who emphasized mastery goals and de-emphasized performance goals also reported the lowest levels of past doping use and the lowest doping intentions. Sas-Nowosielski and Swiatkowska [36] determined that athletes with high task and low ego orientation reported the most negative attitudes toward doping whereas athletes with low task/high ego goals reported the most positive attitudes toward doping. These contradictory effects of mastery and performance orientations were also observed with regard to cheating and cheating intention in sports situations [64]. Our results suggest that these effects may be partially mediated by the relation between achievement goal orientations and sports motivation. Consistent with other researchers [45–47], we argue that a subjective preference of task or ego-related goals structures the experiences of athletes in a manner that affects the degree to which they perceive their sports participation as self-determined. Being "task oriented," i.e., focusing on self-improvement and self-referenced standards of achievement, allows for disregarding societal constraints, comparison and competition with other athletes, which may be beneficial for the fulfillment of the basic needs of autonomy, competence and relatedness. However, "ego-oriented" athletes who set their standards of achievement based on the results of others may more easily question their competence, feel controlled in their sporting activity or experience worse relationships with other athletes, which affect the level of their self-determination. Because the sports motivation variables appear to be directly related to attitudes toward doping and cheating, we argue that the achievement goal orientations are related to doping intentions and behavior by this path.

Our study has some limitations that should be considered. Most importantly, the study employed a cross-sectional design that limits causal interpretations of the proposed relationships. We based our hypotheses on a review that suggested that the proposed direction of relationships would be at least partially valid; however, it is necessary to acknowledge that these relationships may be bi-directional, and we must interpret our results with caution. In addition, we included variables in the study that we hypothesized were important to doping in adolescents; however, a number of other variables that were not included may have similar or even greater effects. Additionally, although we recruited a large number of respondents from all regions of the Czech Republic and the response rate was high, our sample differed in some attributes from the general population of Czech adolescents. However, we believe that the sample showed sufficient diversity for the performed SEM analyses. We should also emphasize that we did not use objective methods to evaluate the prevalence of doping; instead, we relied on participants' self-reports. Although self-reports of doping prevalence have been commonly used in studies of doping in adolescents [13], these methods may have significant limitations [64]. For example, respondents may perceive substances that are not on the list of banned substances to be doping, or they may conceal doping because it is generally a condemned behavior that may even lead to potential penalties. It is also important to note that the relationships between the sport motivation and doping-related variables were significant but weak-to-moderate in magnitude, which suggests that although motivational variables appear to play a role in doping among adolescents, this role should not be exaggerated. Finally, because our model explained a relatively low portion of variance in doping behavior, we may argue that the effect of sport motivation and attitudes toward doping/cheating is much more noticeable with regard to doping intentions than actual doping behavior. Other variables not included in our study, such as the availability or affordability of doping [65], may moderate the relations among motivation, attitudes, and doping behavior.

Conclusion

The present research makes theoretical as well as practical contributions. Theoretically, we used well-established constructs of sports motivation and tested their hypothesized relations with doping-related variables in a complex model, largely confirming our hypotheses regarding the possible effects of achievement goal orientations and self-determined sports motivation. The tested model suggests a series of relationships between sports motivation and doping-related variables that are partially modifiable. Our findings thus suggest practical implications that may be used in doping-prevention efforts. First, it seems that it would be useful to target both doping-specific attitudes and general moral attitudes to decrease doping intentions and perhaps doping behavior. Second, sports motivation appears to play a significant role in attitudes toward doping and cheating and consequently toward doping intentions and actual doping behavior. The dimensions of sports motivation related to intrinsic immersion in the activity appear to have beneficial effects whereas the less self-determined forms of sports motivation may have some undesirable effects with regard to doping. Third, our results also suggest that achievement goal orientations are related to different levels of self-determination in sports activities and through this path, also to moral attitudes. In this manner, self-referenced task-goal orientations focusing on self-improvement appear to be beneficial whereas ego-goal orientations toward competition and comparison with others seem to have some detrimental effects. Therefore, our results further support the suggestions of numerous authors [9, 19–21, 63, 66] that the values present in contemporary youth sports that emphasize high-level performance, success in competition and victory at all costs may have negative consequences, including a greater susceptibility to doping. Positive change could come from parents, coaches, and teachers as well as sports and educational organizations, which all co-create a motivational climate and provide feedback that shapes individual motivational orientations [39, 67]. On this basis, we should once again endorse the classic Coubertin motto that "the important thing is not winning but taking part; the essential thing is not conquering but fighting well".

Supporting information

S1 Appendix. Questionnaire items and measurement loadings.
(DOCX)

Author Contributions

Conceptualization: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Data curation: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Formal analysis: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Funding acquisition: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Investigation: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Methodology: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Project administration: Pavel Slepicka, Irena Slepickova.

Supervision: Pavel Slepicka.

Visualization: Jiri Mudrak.

Writing – original draft: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

Writing – review & editing: Jiri Mudrak, Pavel Slepicka, Irena Slepickova.

References

1. Ehrnborg C, Rosén T. The psychology behind doping in sport. Growth hormone & IGF research. 2009 Aug 31; 19(4):285–7. http://dx.doi.org/10.1016/j.ghir.2009.04.003
2. Kanayama G, Hudson JI, Pope HG. Long-term psychiatric and medical consequences of anabolic–androgenic steroid abuse: A looming public health concern?. Drug and alcohol dependence. 2008 Nov 1; 98(1):1–2. http://dx.doi.org/10.1016/j.drugalcdep.2008.05.004
3. Maravelias C, Dona A, Stefanidou M, Spiliopoulou C. Adverse effects of anabolic steroids in athletes: a constant threat. Toxicology letters. 2005 Sep 15; 158(3):167–75. https://doi.org/10.1016/j.toxlet.2005.06.005 PMID: 16005168
4. World Anti-Doping Agency (2014): 2014 Anti-Doping Testing Figures Report. [staženo 30.7. 2015 z https://wada-main-prod.s3.amazonaws.com/wada_2014_anti-doping-testing-figures_full-report_en.pdf]

5. Ntoumanis N, Ng JY, Barkoukis V, Backhouse S. Personal and psychosocial predictors of doping use in physical activity settings: a meta-analysis. Sports Medicine. 2014 Nov 1; 44(11):1603–24. https://doi.org/10.1007/s40279-014-0240-4 PMID: 25138312

6. Anderson S. J., Bolduc S. P., Coryllos E., Griesemer B., McLain L., Rowland T. W, Risser W. L. (1997): Adolescents and anabolic steroids: a subject review. Pediatrics 99, 904–908. PMID: 9190555

7. Kindlundh AM, Hagekull B, Isacson DG, Nyberg F. Adolescent use of anabolic—androgenic steroids and relations to self-reports of social, personality and health aspects. The European Journal of Public Health. 2001 Sep 1; 11(3):322–8. http://dx.doi.org/10.1093/eurpub/11.3.322 PMID: 11582614

8. Blatný M, Hrdlicka M, Sobotková V, Jelínek M. Prevalence antisocialního chovani ceskych adolescentu z mestskych oblasti. Ceskoslovenská psychologie. 2006 Jul 1; 50(4):297.

9. Yesalis CE, Bahrke MS. Doping among adolescent athletes. Best Practice & Research Clinical Endocrinology & Metabolism. 2000 Mar 31; 14(1):25–35. https://doi.org/10.1053/beem.2000.0051

10. Johnson MD, Jay MS, Shoup B, Rickert VI. Anabolic steroid use by male adolescents. Pediatrics. 1989 Jun 1; 83(6):921–4. PubMed PMID: 2726346

11. Kindlundh AM, Isacson DG, Berglund L, Nyberg F. Doping among high school students in Uppsala, Sweden: A presentation of the attitudes, distribution, side effects, and extent of use. Scandinavian Journal of Public Health. 1998 Jan 1; 26(1):71–4. https://doi.org/10.1177/14034948980260010201

12. Lucidi F, Zelli A, Mallia L, Grano C, Russo PM, Violani C. The social-cognitive mechanisms regulating adolescents' use of doping substances. Journal of sports sciences. 2008 Mar 1; 26(5):447–56. https://doi.org/10.1080/02640410701579370 PMID: 18274942

13. Pedersen W, Wichstrøm L. Adolescents, doping agents, and drug use: A community study. Journal of Drug Issues. 2001 Apr 1; 31(2):517–41. https://doi.org/10.1177/002204260103100208

14. Sas-Nowosielski K. The abuse of anabolic-androgenic steroids by Polish school-aged adolescents. Biology of Sport. 2006 Jan 1; 23(3):225.

15. Zelli A, Mallia L, Lucidi F. The contribution of interpersonal appraisals to a social-cognitive analysis of adolescents' doping use. Psychology of sport and exercise. 2010 Jul 31; 11(4):304–11. http://dx.doi.org/10.1016/j.psychsport.2010.02.008

16. Laure P, Binsinger C. Doping prevalence among preadolescent athletes: a 4-year follow-up. British journal of sports medicine. 2007 Oct 1; 41(10):660–3. https://doi.org/10.1136/bjsm.2007.035733 PMID: 17473000

17. Irving LM, Wall M, Neumark-Sztainer D, Story M. Steroid use among adolescents: findings from Project EAT. Journal of Adolescent Health. 2002 Apr 30; 30(4):243–52. http://dx.doi.org/10.1016/S1054-139X(01)00414-1 PMID: 11927236

18. Pedersen W, Wichstrøm L, Blekesaune M. Violent Behaviors, Violent Victimization, and Doping Agents A Normal Population Study of Adolescents. Journal of interpersonal violence. 2001 Aug 1; 16(8):808–32. https://doi.org/10.1177/088626001016008005

19. Knop PD. European trends in youth sport: a report from 11 European countries. European Journal of Physical Education. 1996 Jan 1; 1(1–2):36–45. http://dx.doi.org/10.1080/1740898960010104

20. Ehrnborg C, Rosén T. The psychology behind doping in sport. Growth hormone & IGF research. 2009 Aug 31; 19(4):285–7. http://dx.doi.org/10.1016/j.ghir.2009.04.003

21. Petróczi A. Attitudes and doping: a structural equation analysis of the relationship between athletes' attitudes, sport orientation and doping behaviour. Substance abuse treatment, prevention, and policy. 2007 Nov 9; 2(1):1. https://doi.org/10.1186/1747-597X-2-34 PMID: 17996097

22. Lentillon-Kaestner V, Carstairs C. Doping use among young elite cyclists: a qualitative psychosociological approach. Scandinavian journal of medicine & science in sports. 2010 Apr 1; 20(2):336–45. https://doi.org/10.1111/j.1600-0838.2009.00885.x PMID: 19486486

23. Barkoukis V, Lazuras L, Tsorbatzoudis H, Rodafinos A. Motivational and sportspersonship profiles of elite athletes in relation to doping behavior. Psychology of sport and exercise. 2011 Jun 30; 12(3):205–12. http://dx.doi.org/10.1016/j.psychsport.2010.10.003

24. Barkoukis V, Lazuras L, Tsorbatzoudis H, Rodafinos A. Motivational and social cognitive predictors of doping intentions in elite sports: An integrated approach. Scandinavian journal of medicine & science in sports. 2013 Oct 1; 23(5):e330–40

25. Chan DK, Dimmock JA, Donovan RJ, Hardcastle SA, Lentillon-Kaestner V, Hagger MS. Self-determined motivation in sport predicts anti-doping motivation and intention: A perspective from the trans-contextual model. Journal of Science and Medicine in Sport. 2015 May 31; 18(3):315–22. https://doi.org/10.1016/j.jsams.2014.04.001 PMID: 24793786

26. Chan DK, Lentillon-Kaestner V, Dimmock JA, Donovan RJ, Keatley DA, Hardcastle SJ, et al. Self-control, self-regulation, and doping in sport: a test of the strength-energy model. Journal of sport & exercise psychology. 2015 Apr; 37(2):199–206. https://doi.org/10.1123/jsep.2014-0250 PMID: 25996110

27. Hodge K, Hargreaves EA, Gerrard D, Lonsdale C. Psychological mechanisms underlying doping attitudes in sport: Motivation and moral disengagement. Journal of sport & exercise psychology. 2013 Aug 1; 35(4):419–32. https://doi.org/10.1123/jsep.35.4.419

28. Zucchetti G, Candela F, Villosio C. Psychological and social correlates of doping attitudes among Italian athletes. International Journal of Drug Policy. 2015 Feb 28; 26(2):162–8. https://doi.org/10.1016/j.drugpo.2014.07.021 PMID: 25168080

29. Lazuras L, Barkoukis V, Tsorbatzoudis H. Toward an integrative model of doping use: an empirical study with adolescent athletes. Journal of Sport and Exercise Psychology. 2015 Feb; 37(1): 37–50. https://doi.org/10.1123/jsep.2013-0232 PMID: 25730890

30. Gucciardi DF, Jalleh G, Donovan RJ. An examination of the Sport Drug Control Model with elite Australian athletes. Journal of Science and Medicine in Sport. 2011 Nov; 14(6): 469–476. https://doi.org/10.1016/j.jsams.2011.03.009 PMID: 21514883

31. Morente-Sánchez J, Zabala M. Doping in sport: a review of elite athletes' attitudes, beliefs, and knowledge. Sports Medicine. 2013 Jun; 43(6): 395–411. https://doi.org/10.1007/s40279-013-0037-x PMID: 23532595

32. Fishbein M, Cappella JN. The role of theory in developing effective health communications. Journal of communication. 2006 Aug 4; 56(s1): S1–S17. https://doi.org/10.1111/j.1460-2466.2006.00280.x

33. Fishbein M. A reasoned action approach to health promotion. Medical Decision Making. 2008 Nov 17; 28(6), 834–844. https://doi.org/10.1177/0272989X08326092 PMID: 19015289

34. Nicholls JG. Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. Psychological review. 1984 Jul; 91(3):328. http://dx.doi.org/10.1037/0033-295X.91.3.328

35. Allen J, Taylor J, Dimeo P, Dixon S, Robinson L. Predicting elite Scottish athletes' attitudes toward doping: examining the contribution of achievement goals and motivational climate. Journal of sports sciences. 2015 May 28; 33(9):899–906. https://doi.org/10.1080/02640414.2014.976588 PMID: 25537139

36. Sas-Nowosielski K, Swiatkowska L. Goal orientations and attitudes toward doping. International journal of sports medicine. 2008 Jul; 29(07):607–12. https://doi.org/10.1055/s-2007-965817 PMID: 18214812

37. Vealey RS, Hayashi SW, Garner-Holman M, Giacobbi P. Sources of sport-confidence: Conceptualization and instrument development. Journal of Sport and Exercise Psychology. 1998 Mar 1; 20:54–80.

38. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American psychologist. 2000 Jan; 55(1):68. http://dx.doi.org/10.1037/0003-066X.55.1.68 PMID: 11392867

39. Ntoumanis N, Standage M. Morality in sport: A self-determination theory perspective. Journal of Applied Sport Psychology. 2009; 21(4): 365–380. http://dx.doi.org/10.1080/10413200903036040

40. Vallerand RJ, Losier GF. An integrative analysis of intrinsic and extrinsic motivation in sport. Journal of applied sport psychology. 1999; 11(1): 142–169.

41. Dweck C. Mindset: The new psychology of success. Random House; 2006 Feb 28.

42. Pintrich PR. Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. Journal of educational psychology. 2000 Sep; 92(3):544. http://dx.doi.org/10.1037/0022-0663.92.3.544

43. Kavussanu M, Ntoumanis N. Participation in sport and moral functioning: Does ego orientation mediate their relationship?. Journal of Sport and Exercise Psychology. 2003; 25:501–18.

44. Duda JL, Olson LK, Templin TJ. The relationship of task and ego orientation to sportsmanship attitudes and the perceived legitimacy of injurious acts. Research quarterly for exercise and sport. 1991 Mar 1; 62(1):79–87. https://doi.org/10.1080/02701367.1991.10607522 PMID: 2028097

45. Ntoumanis N. Empirical links between achievement goal theory and self-determination theory in sport. Journal of Sports Sciences. 2001; 19(6): 397–409. https://doi.org/10.1080/026404101300149357 PMID: 11411776

46. Duda JL, Chi L, Newton ML, Walling MD, Catley D. Task and ego orientation and intrinsic motivation in sport. International journal of sport psychology. 1995; 26(1): 40–63.

47. Standage M, Duda JL, Ntoumanis N. A model of contextual motivation in physical education: Using constructs from self-determination and achievement goal theories to predict physical activity intentions. Journal of educationa psychology. 2003; 95(1): 97.

48. World Anti-Doping Agency. The 2015 Prohibiter List. World Anti-Doping Agency; 2015. Downloaded from https://wada-main-prod.s3.amazonaws.com/resources/files/wada-2015-prohibited-list-en.pdf.

49. Petróczi A, Aidman E. Measuring explicit attitude toward doping: Review of the psychometric properties of the Performance Enhancement Attitude Scale. Psychology of Sport and Exercise. 2009 May 31; 10(3):390–6. http://dx.doi.org/10.1016/j.psychsport.2008.11.001

50. Slepička P, Slepičková I. Social Aspects of Doping and Antidoping Prevention Possibilities in Children and Youth. Acta Universitatis Carolinae Kinanthropologica. 1996; 32: 23–34.

51. Lee MJ, Whitehead J, Ntoumanis N. Development of the attitudes to moral decision-making in youth sport questionnaire (AMDYSQ). Psychology of Sport and Exercise. 2007 May 31; 8(3):369–92. http://dx.doi.org/10.1016/j.psychsport.2006.12.002

52. Roberts GC, Treasure DC, Balague G. Achievement goals in sport: The development and validation of the Perception of Success Questionnaire. Journal of Sports Sciences. 1998 Mar 8; 16(4), 337–347. https://doi.org/10.1080/02640419808559362 PMID: 9663958

53. Mallett C, Kawabata M, Newcombe P, Otero-Forero A, Jackson S. Sport motivation scale-6 (SMS-6): A revised six-factor sport motivation scale. Psychology of Sport and Exercise. 2007 Sep 30; 8(5):600–14. http://dx.doi.org/10.1016/j.psychsport.2006.12.005

54. Pelletier LG, Fortier MS, Vallerand RJ, Tuson KM, Briere NM, Blais MR. Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: The Sport Motivation Scale (SMS). Journal of sport and Exercise Psychology. 1995 Mar 1; 17:35–53.

55. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2014.

56. Rosseel Y. lavaan: An R package for structural equation modeling. Journal of Statistical Software. 2012 May; 48(2): 1–36.

57. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: a Multidisciplinary Journal. 1999; 6(1):1–55. http://dx.doi.org/10.1080/10705519909540118

58. MacCallum RC, Browne MW, Sugawara HM (1996). Power analysis and determination of sample size for covariance structure modeling. Psychological methods. 1996; 1(2): 130–149. http://dx.doi.org/10.1037/1082-989X.1.2.130

59. Marsh HW, Hau KT, Wen Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Structural equation modeling. 2004; 11(3): 320–341. http://dx.doi.org/10.1207/s15328007sem1103_2

60. Jalleh G, Donovan RJ, Jobling I. Predicting attitude toward performance enhancing substance use: A comprehensive test of the Sport Drug Control Model with elite Australian athletes. Journal of Science and Medicine in Sport. 2014; 17(6): 574–579. https://doi.org/10.1016/j.jsams.2013.10.249 PMID: 24268440

61. World Anti-Doping Agency. Dangers of doping: Get the facts. World Anti-doping Agency; 2017. Downloaded from https://www.wada-ama.org/en/resources/education-and-prevention/dangers-of-doping-get-the-facts.

62. Czech Anti-doping Comittee. Education. Czech Antidoping Comittee; 2017. Downloaded from http://www.antidoping.cz/vzdelavani.php.

63. Coakley JJ. Sport in society: issues & controversies. McGraw-Hill; 1997.

64. Petróczi A, Aidman EV, Hussain I, Deshmukh N, Nepusz T, Uvacsek M, et al. Virtue or pretense? Looking behind self-declared innocence in doping. PloS one. 2010; 5(5): e10457. https://doi.org/10.1371/journal.pone.0010457 PMID: 20463978

65. Pluim B. A doping sinner is not always a cheat. British journal of sports medicine. 2008; 42(7): 549–550. PMID: 18606830

66. Derenne JL, Beresin EV. Body image, media, and eating disorders. Academic psychiatry. 2006 May 1; 30(3):257–61. https://doi.org/10.1176/appi.ap.30.3.257 PMID: 16728774

67. Alderman MK. Motivation for achievement: Possibilities for teaching and learning. Routledge; 2013.

15

Responsiveness of the Calf-Raise Senior test in community-dwelling older adults undergoing an exercise intervention program

Helô-Isa André[1☯¤a*], Vera Moniz-Pereira[1¤a‡], Fátima Ramalho[2¤b‡], Rita Santos-Rocha[2¤b], António Veloso[1¤a], Filomena Carnide[1☯¤a]

1 CIPER, Sports & Health Department, Biomechanics and Functional Morphology Laboratory, Faculty of Human Kinetics, University of Lisbon, Cruz Quebrada-Dafundo, Lisbon, Portugal, **2** Sport Sciences School of Rio Maior, Polytechnic Institute of Santarém, Rio Maior, Portugal

☯ These authors contributed equally to this work.
¤a Current address: Estrada da Costa, Cruz Quebrada-Dafundo, Lisbon, Portugal
¤b Current address: Avenida Dr. Mário Soares, Rio Maior, Portugal
‡ These authors also contributed equally to this work.
* heloandre@fmh.ulisboa.pt

Editor: Gianluigi Forloni, Istituto Di Ricerche Farmacologiche Mario Negri, ITALY

Abstract

Introduction

Mobility significantly depends on the ankle muscles' strength which is particularly relevant for the performance of daily activities. Few tools are available, to assess ankle strength with all of the measurement properties tested. The purpose of this study is to test the responsiveness of Calf-Raise Senior Test (CRS) in a sample of elderly participants undergoing a 24-week community exercise program.

Methods

82 older adults participated in an exercise program and were assessed with CRS Test and 30-second chair stand test (CS) at baseline and at follow-up. Effect size (ES), standardized response mean (SRM) and minimal detectable change (MDC) measures were calculated for the CRS and CS tests scores. ROC curves analysis was used to define a cut-off representing the minimally important difference of Calf-Raise Senior test.

Results

Results revealed a small (ES = 0.42) to moderate (SRM = 0.51) responsiveness in plantarflexion strength and power across time, which was lower than that of CS test (ES = 0.64, SRM = 0.67). The responsiveness of CRS test was more evident in groups of subjects with lower initial scores. A minimal important difference (MID) of 3.5 repetitions and a minimal detectable change (MDC) of 4.6 was found for the CRS.

Conclusion

Calf-Raise Senior Test is a useful field test to assess elderly ankle function, with moderate responsiveness properties. The cutoff scores of MDC and MID presented in this study can

be useful in determining the success of interventions aiming at improving mobility in senior participants.

Competing interests: NO authors have competing interests

Introduction

Population aging is a major global demographic trend. The increase in life expectancy raises the related concern of increasing morbidity, prolonged disability and dependency, with a reduction in quality of life [1]. This demographic transition is influencing the economy, care, social development, welfare and well-being. Thus, it is crucial to adapt policy, health services and intervention programs to an aging and frailer population.

The quality of life of the elderly depends on their health and ability to perform activities of daily living (ADLs) [2]. Mobility is a prerequisite for the performance of most common ADLs and its maintenance is a major goal of geriatric health professionals [3]. Mobility greatly depends on the strength in the lower limbs, especially in the ankle muscles, which are particularly relevant in gait function [2, 4, 5] climbing stairs and rising from a chair [5–7]. Plantar-flexors (PF) muscles act to support and to propel the body forward in late stance and their strength is positively related to gait velocity and step length [8]. Lower strength levels are very common in older adults [9, 10] and are associated with poor gait and balance [11, 12]. PF muscles, in particular, reveal large decline in strength with aging, presenting a decrease of 2.3% in very old adults (>85 years-old) per year [13], and differences of about 40% when comparing elderly with young men [14, 15].

PF assessment is especially important to design programs or implement strategies, aiming the preservation of the mobility. This issue is even more relevant in the design of exercise programs tailored to functional status of older adults [16–18]. In order to ensure meaningful and quality data related to the functional loss and changes in mobility parameters with age, the use of measurement tools with relevant psychometric properties is essential. Studies reporting validity and responsiveness of strength and mobility assessment tools are relatively scarce [19, 20]. Additionally, few of the aforementioned studies report sensibility and sensitivity data, allowing the establishment of the minimal importance difference. To our knowledge, only the "30 second chair stand test" (CS) [19] and the "Timed Up and Go" (TUG) [20] have been tested for responsiveness in community-setting interventions with healthy and independent older adults. Although both tests are used to assess lower limbs muscle function, none of them provides specific information about the PF strength, which has already been mentioned as being relevant for preserving the quality of gait in the elderly [2, 4–5]

The Calf-Raise Senior (CRS) test is the only field assessment tool developed to evaluate ankle muscle function in the elderly and has shown to have an excellent test-retest reliability (ICC = 0.90), inter-rater reliability (ICC = 0.93–0.96) and a good intra-rater reliability (ICC = 0.79–0.84) [21,22]. The test also presented a significant association between its scores and laboratory strength assessments (isometric, $r = 0.87$, $r^2 = 0.75$; isokinetic, $r = 0.86$, $r^2 = 0.74$; and rate of force development, $r = 0.77$, $r^2 = 0.59$), demonstrating to be an excellent indicator of ankle strength in older adults [21]. Despite good results regarding the reliability and validity of the CRS test [21,22], its responsiveness has not yet been established. Therefore, the purpose of this study is to test the responsiveness of the CRS test in a sample of elderly participants undergoing a 24 weeks' community exercise program.

Materials and methods

Study design and subjects

A prospective multi-site cohort study was conducted, with a 24-week follow-up period. The follow up interval was defined by having into account that although major adaptations in strength and power in the elderly occur after 12 weeks of training [23–25] higher effect has been showed after longer training periods (≥24 weeks) [26]. All participants underwent a multicomponent community-based exercise intervention included in the "More Active Aging Project" (MAAP), which was implemented in 5 municipalities in the West and Ribatejo Regions of Portugal. The study was coordinated by the Faculty of Human Kinetics, University of Lisbon and Sport Sciences School of Rio Maior, between September 2014 and December 2015. All detailed information about the protocol and methods used in the MAAP intervention can be found in Ramalho et al. [27].

A sample of 82 older adults from the abovementioned cohort was recruited through advertising in local centers and sports community services by a multi-stage sample method. Using the results from our previous study [21], the minimum sample size of 61 participants was determined, considering and effect size of $d = 0.80$, with 80% power and alpha at 0.

To be considered eligible for this study, participants should be 65 years or more, live independently in the community, be autonomous and correctly understand the Portuguese language. Exclusion criteria were considered: self-reported cognitive, neurological, bone and joints, or other impairments that could inhibit the performance of exercises in the standing position autonomously; inability to walk independently and/or without assistance of a walking aid and not having a hip or knee prosthesis.

A written informed consent was obtained from all participants at the beginning of the intervention. The Faculty of Human Kinetics Ethics Committee (Lisbon University) approved the study protocol.

Exercise program

The MAAP 24-weeks exercise program comprised group-based multicomponent 50-min exercise sessions twice a week. Graduated exercise specialists received 20 hours of training regarding the program methodology and follow-up guidelines during the intervention period. The structure of the exercise program is fully explained elsewhere [27]. In brief, the program aims to develop postural control, balance (static and dynamic), endurance, mobility, walking pattern, and to improve strength and muscle resistance. In order to provide continuous and progressive stimulus to the participants' functional capacities, weekly progressive changes in intensity and complexity, and monthly variation in exercise mode, were incorporated into the program. The progression of the exercise program was controlled through periodic and unscheduled visits by the research team, in order to verify the compliance of the guidelines. Additionally, the instructors recorded a monthly qualitative classification of the participants' performance.

Assessments

All assessments were conducted at baseline (BL) and after 24 weeks (follow-up, FU) and were administered by examiners who received specific training in applying the test protocols.

To evaluate the eligibility of participants, a health and falls assessment questionnaire designed and validated by the Portuguese Language and Culture [28] was administered through face-to-face interviews. The questionnaire included questions about demographics, health, self-perception status, medication intake, medical history, and falls history. It was used to verify the eligibility of participants in the study, as well as, identifying health conditions that could prevent participation in the exercise program sessions and interfere with performance in the assessment tests.

All participants were evaluated using the CRS and CS tests on the same day in each phase (BL and FU) by the same examiners. The CS test was chosen as an external reference measure (anchor) in this study, as it measures the same attribute of the CRS (lower limbs strength) and presents results that can be partially explained by the PF strength ($\beta = 0.358$, $P = 0.074$) [29]. The CS test protocol consists of the performance of the maximum possible repetitions of the stand/sit down movements in 30 seconds [19]. The test was administered using an armless chair (height: 43.2 cm), which was supported against a wall to ensure stability. The final score corresponded to the total number of complete cycles performed, i.e., the participants should extend their knees and sit fully on the chair, not lift the feet off the floor and keep the arms folded over the chest. The CRS test protocol is fully described elsewhere [21]. Briefly, the protocol includes the performance of a maximum number of heel lifting / lowering repetitions in the standing position, in 30 seconds, with the knees extended, at maximum possible range and velocity, without transferring the body weight to the hands. The test score corresponded to the number of cycles correctly executed at the end of 30 seconds.

Data analyses

Descriptive statistical analyses were performed to characterize the sample. Central tendency parameters were determined for continuous variables (mean, standard deviation and median) and relative frequency was calculated for categorical and ordinal variables. The normal distribution of continuous variables was checked with the Kolmogorov-Smirnov Test.

The responsiveness of the CRS test was determined using two different methods: a distribution-based approach and an anchor-based approach.

For the distribution-based approach the results of the two assessment phases were used to calculate the change in scores (FU score—BL score) of the CRS and CS tests. The following statistical parameters, commonly used to assess the responsiveness of instruments, were computed: 1) Effect size (ES)—provides information about the magnitude of change over time by dividing the Mean Change Score of a variable by the SD of its BL Scores[30,31]. To interpret the effect-size data, the cutoff points proposed by Hopkins [32] ($ES < 0.20$ = trivial effect; $0.20 \geq ES < 0.60$ = small effect; $0.60 \geq ES < 1.20$ = moderate effect; $1.20 \geq ES < 2.0$ = large effect; $2.0 \geq ES < 4.0$ = very large; and $ES \geq 4.0$ = nearly perfect); 2) Standardized response mean (SRM)—parameter that indicates if the change of the results over time were large relative to the variability in the measurements. The SRM can be calculated as the Mean Change Score of the variable divided by the SD of the same Change Score. SRM values of 0.20, 0.50, and 0.80 are considered as small, moderate and large change, respectively [33]; 3) Minimal Detectable Change (MDC)—measure that reflects the smallest change in score that can be interpreted as a 'true' change, i.e. beyond measurement error [34]. The formula for the calculation of MDC can be expressed as: 1.96 ($\sqrt{2}$ x Standard Error of Measurement). The calculation of the Standard Error of Measurement was based on the results of our previous study using the following equation: SEM = SD of BL scores ($\sqrt{1\text{-ICC}}$). The proportion of participants achieving a degree of improvement beyond the MDC was then determined [35].

The anchor-based approach was performed using a Receiver Operating Characteristic (ROC) curve analysis in order to verify whether the CRS test could discriminate between participants with positive change (improved) versus no change (stable) [36,37]. The cut-off of 3.01 was considered to dichotomize sample in accordance with the minimal detectable change (MDC) determined in a previous test-retest reliability study (CS change score < 3.01 = stable group; CS change score ≥ 3.01 = improved group). An area under the curve (AUC) was used to determine specificity and sensitivity [38], and the cut-off corresponding to the point closer to the upper left corner was defined as the score that best classifies participants who had improved or maintained their state. This cut-off represents the "minimally important difference" (MID) of this test [34], that is, the smallest change in the CRS scores that is considered clinically relevant, or worthwhile to the participants [35], also frequently referred to in the literature as the "minimal clinically important difference" (MCID) [36, 39]. Paired t-tests (or non-parametric Wilcoxon tests) were used to compare data at BL and FU within groups of change.

Results

Eighty two healthy older adults were eligible to participate in the responsiveness study of the CRS test. All participants were present in both assessment periods and met a minimum attendance threshold in the training sessions. Furthermore, none of the participants showed signs of overexertion, pain in the lower limbs or other signs of discomfort that prevented them from complying with the requirements for a satisfactory assessment.

Participants were mainly women (87.8%) with good general health perception ($\geq$3 points-scale) and a mean BMI of 29.9±5.1 kg/m^2 (Table 1), indicating that this group is overweight ($\geq$ 25,0 kg/m^2) [40], although out of the range for increased mortality risk (BMI $<$23.0 or $>$ 33.0) [41].

In general, participants underwent statistically significant improvements in their lower limb strength (CRS and CS, $P < 0.001$). Regarding the distribution-based approach, the effect size was low to moderate, ranging from 0.4 to 0.6 (CRS and CS, respectively) and SRM values were moderate, ranging from 0.5 to 0.7 (Table 2).

The change detected using the CRS was higher, by a statistically significant difference, for the group of participants who improved (CRS change score = 5.8±5.4) when compared with the stable group (CRS change score = 3.0±6.5) (Table 3). Accordingly, it is also possible to verify that the proportion of changes related to initial values in CRS was higher in the improved group (CRS relative change = 37.9 ± 54.9%) than in the stable group (CRS relative change = 15.4 ± 31.5%). The results of the comparison between BL and FU scores validate the

Table 1. Demographics and functional fitness measures in baseline from the total group of participants and subgroups of CRS scores.

Demographic and health parameters	
N = 82	Mean ± SD (median)/ %
Age (years)	72.3 ± 5.0 (72,0)
Gender, female (%)	87.8
BMI (kg/m^2)	29.9 ± 5.2 (29.4)
HPS (1–4 scale)	3,3 ± 0,8 (3,0)
Functional fitness parameters	
N = 82	Mean ± SD (median)/ %
CRS (x/30s)	25.0 ± 8,8 (24,0)
CS (x/30s)	16,1 ± 4,6 (15,5)

Data are presented as mean ± standard deviation (median) for continuous variables, and percentage for categorical variables on Baseline.

Abbreviations: BMI = Body mass index; HPS = Health Perception Status; CRS = Calf-raise Senior test; CS = 30 s chair stand test.

1

Table 2. Responsiveness of FF measures for the Total Sample and by Subgroups of Lower and Higher CRS Scores.

	Change Score (N = 82)			
Parameters	**mean**	**± SD (median)**	**ES**	**SRM**
CRS score (x/30s)	3.4	± 6.6 (3.0)	0.4	**0.5***
CS score (x/30s)	3.6	± 5.3 (3.0)	**0.6†**	**0.7***

Data are presented as mean ± standard deviation (median). Cohen's Effect Size (ES) and Standardized Response Mean (SRM) from the comparison between baseline and follow up scores. Dagger (†) indicates ES|d| > 0.6 and asterisk (*) indicates SRM > 0.5 (medium to high Effect Size or response mean, here considered as important differences between group means).

Abbreviations: CRS = Calf-raise Senior test; CS = 30 s chair stand test; ES: Effect Size; SRM = standardized response mean.

lack of significant changes in the stable group, as expected. Subgroups of change did not reveal equivalence on BL (stables = 27.9± 9.0 vs improved 2.00±6.7; $P = 0.006$) but the differences were non-significant on FU ($P = 0.61$).

A ROC curve analysis shows a change score in the CRS test greater than or equal to 3.5 reps has a sensitivity of 0.68 and specificity of 0.33. The area under the curve found in this test was 0.67 ($P < 0.01$) indicating moderate discriminative ability.

Fig 1 shows the proportion of participants in the exercise program who had greater improvements in plantar-flexors (PF) strength and power than the MDC values (42.7%) and beyond MID (47.6%). Thirty five participants reached or overcame the MDC cut-off point of 4.6 and presented a mean CRS change score of 9.60 (± 4.33), while 37 elderly subjects improved PF function (≥ MID of 3.5) and revealed a CRS change score of 9.03 (± 4.45).

Discussion

This study aimed to test the responsiveness of CRS test, in a sample of older-adult participants undergoing a 24-weeks' community exercise program.

The CRS test showed a small to moderate responsiveness. A higher responsiveness was found for the CS test both in this study, as well as in other studies [42].

The responsiveness of the CRS test was also performed using an external measure to compare changes and establish cut-off values associated with meaningful improvements in physical function as a result of the intervention. The comparative analysis of the CRS scores obtained in the two time points (BL and FU) between the groups of positive change (improved group) and without change (stable group) in the reference test (CS) revealed significant differences in absolute and relative changes. Considering that the attribute assessed in both tests is the same —strength in the lower limbs; and that the two tests are evaluated in a similar way—the number of movements performed in 30s (revealing the same limitations in cognitive and sensory terms); it seems reasonable to state that the CRS test is responsive to discriminate elderly subjects with relevant changes after an exercise intervention program.

The ROC curve analysis revealed an optimal cut-off point of 3.5 repetitions allowing us to establish the minimal important difference for the CRS test. With this analysis it was possible to identify about 70% of the participants who underwent a truly important change after the intervention, while recognizing approximately 30% of the elderly who did not show a real change in their strength in any of the methods. The MDC value revealed that a change score of 4.6 would be required for the resulting change in participant status to be outside the test error range, which is higher than the MID estimate. This is in accordance with other responsiveness studies, in which anchor-based approaches outweigh the values found in the distribution

Table 3. Responsiveness of CRS test to the 24 weeks-exercise program, considering groups of change in the CS test (stable and improved).

	CS stable group (N = 34)		CS improved group (N = 48)			
Parameters	**mean**	**± SD (median)**	**mean**	**± SD (median)**	***P value***	**ES (d)**
CRS Baseline (x/30s)	27.8	± 9.0 (27,0)	22.0	± 6.7 (23.0)	**0,00****	**1.2†**
CRS Follow up (x/30s)	30.8	± 9.1 (30,0)	27.8	± 6.5 (28.0)	NS	0.4
CRS Change score (x/30s)	3.0	± 6,5 (2.0) NS	5.8	± 5.4 (5.0)§	**0.03***	**0.6†**
CRS relative change (%)	15.4	± 31.,5 (15.0)	37.9	± 54.,9 (32.0)	**0.02***	**0.5†**

Means ± standard deviations (median) of CRS test results on baseline and follow up; change scores and relative change (%) in groups of participants classified as stable or improved (CS test change scores, MDC = 3.01).

*$P<0.05$,

**$P<0.001$, comparison between groups (T-Student or Mann Whitney tests);

§$P<0.001$, comparison between baseline and follow-up scores (Wilcoxon test);

†ES |d |>0.5 (medium to high Effect Size, here considered as clinically relevant differences between group means); NS indicates non-significant differences.

Abbreviations: CRS = Calf-raise Senior test; CS = 30 s chair stand test.

approaches [40,43], highlighting the difficulties in using the former method due to the lack of an optimal threshold which can accurately determine the cut-offs to set the MID. Therefore, if only the value of MID is considered to evaluate the effects of an intervention, the results may indicate that improvements can be attributed to test error and not necessarily to a true change. Moreover, if the MDC is used as the only reference of change, then a small but significant effect can be neglected [40]. Several authors state that the value of the MID may be related to the SEM, depending on the degree of improvement defined by the anchor [44,45]. It has been shown that a cut-off point for MID between "slightly improved" and "moderately improved" may be similar to 2.0–2.3 * SEM [34], while in studies requiring "moderate" or "much improvement", MID corresponds to about 2.5–2.6 times the SEM value [34]. In the present study, the SEM value calculated for the CRS test was 2.8, indicating that the closest cut-off point of this relationship would be the MDC = 2.3 * SEM. In this case, the MID would be on the order of 1.25 * SEM, which is more in line with other studies [44] in which similar values using clinical parameters as the anchor were observed. Therefore, it is suggested that both values should be used in assessing ankle strength improvements resulting from an exercise program. In practice, we can establish that changes in CRS scores below 3.5 must be considered insufficient, values between 3.5 and 4.6 may be viewed as acceptable for slight to moderate changes (but with a chance of being inside the range of measurement errors), while scores above 4.6 can be considered as a true change.

The anchor-based approach used in the present study was based on an ecological perspective, considering that the effects of community programs are usually assessed through field tests, which are easy to administer, have good acceptability and motivation, and reach a large number of participants in screenings [17]. Therefore, taking into account that the CS test is widely used in this context, and that it has been showing very positive indicators of validation, reliability and sensitivity to change in previous studies [19,46,29], its use seems to us acceptable for a preliminary approach in the scope of CRS test responsiveness assessment. Nonetheless, to indicate a true change in the plantar-flexors strength, a comparison with a quantitative direct measures as gold-standard measures would be more accurate, such as a dynamometer strength test, force platform, a biomechanical gait analysis, or the use of other specific clinical test for ankle strength assessment (e.g. manual muscle testing) [47].

The lack of other studies assessing CRS responsiveness prevents comparison of the results found in the present study. As suggested by Revicki [48], the estimate of MID should be

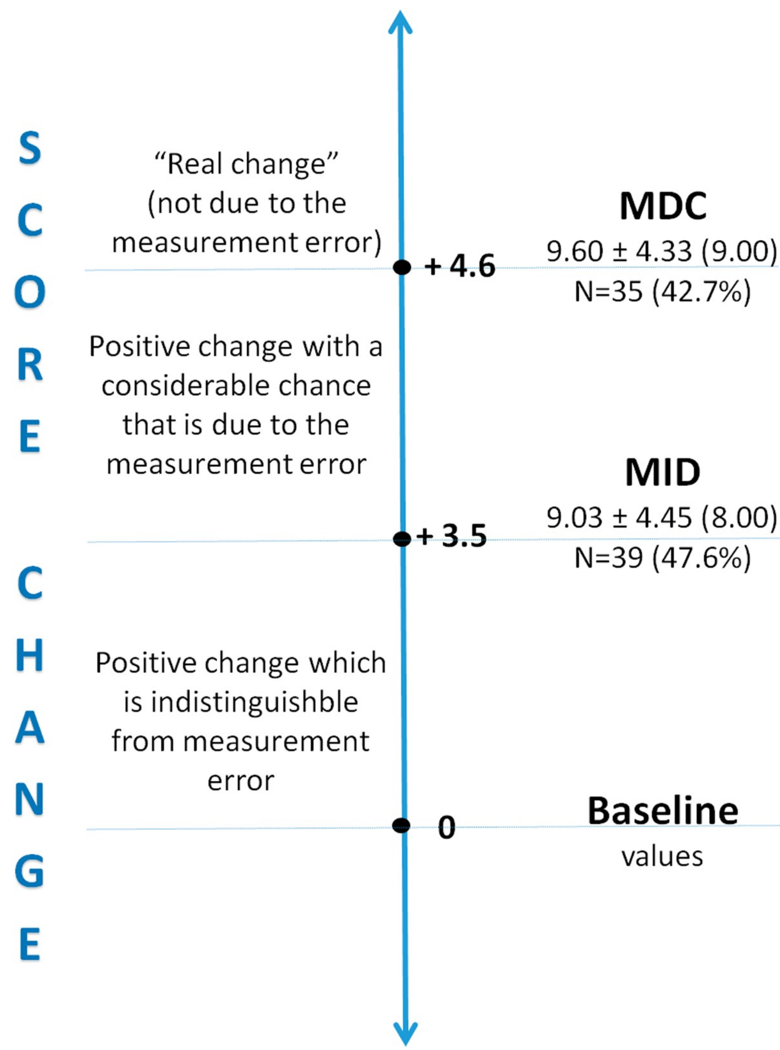


Fig 1. Means and standard deviations of change scores, and proportion of participants who reached or exceeded the MDC and MID cut-off points.

confirmed through on the accumulation of evidence from several studies, in order to build greater confidence in the defined cut-offs values. Therefore, it is necessary to develop further studies using a larger sample composed of participants with higher / lower functional-fitness levels, as well as performing interventions that address longer-term changes in physical function. We also suggest carrying out additional research to evaluate whether the test can detect relevant changes in people with a higher baseline physical condition since the sensitivity to change demonstrated by the CRS test was more evident in groups of subjects with lower initial scores. Despite the weaknesses identified, this is the first article that defines the responsiveness of the CRS test, identifying cut-off values of MDC and MID that may help to establish a basis for future studies focused on plantar flexion strength and power interventions in the elderly.

Conclusions

This study aimed to examine the responsiveness of the Calf-Raise Senior (CRS) test through a 24-week exercise intervention designed to improve muscle strength, endurance, flexibility and balance, as key factors affecting physical function.

The results strengthen the psychometric properties of CRS, revealing its ability to detect change after a 24 week community exercise program focused on improving functional mobility in the elderly. In addition to its excellent validity, reliability, and acceptability by participants and professionals, the CRS test revealed good responsiveness in detecting changes in plantar flexion function over time. The present study also provides data relevant to the field application of these measures, reporting cutoffs of 3.5 and 4.6 for the MDC and MID estimates, respectively.

Supporting information

S1 File.
(ZIP)

Acknowledgments

The authors are grateful to all of the seniors who voluntarily participated in this study, and acknowledge the valuable assistance of Sara Gabriel and all the other MAAP project technicians who gave their support in data collection.

Author Contributions

Conceptualization: Helô-Isa André, Vera Moniz-Pereira, Fátima Ramalho, Filomena Carnide.

Formal analysis: Helô-Isa André, António Veloso, Filomena Carnide.

Funding acquisition: Rita Santos-Rocha.

Investigation: Helô-Isa André.

Methodology: Helô-Isa André, Fátima Ramalho, Filomena Carnide.

Project administration: António Veloso.

Resources: Rita Santos-Rocha.

Supervision: António Veloso, Filomena Carnide.

Validation: Filomena Carnide.

Writing – original draft: Helô-Isa André, Filomena Carnide.

Writing – review & editing: Helô-Isa André, Vera Moniz-Pereira, Fátima Ramalho, Rita Santos-Rocha, António Veloso, Filomena Carnide.

References

1. World Health Organization. World report on ageing and health: World Health Organization; 2015.
2. Moreland JD, Richardson JA, Goldsmith CH, Clase CM. Muscle weakness and falls in older adults: a systematic review and meta-analysis. Journal of the American Geriatrics Society. 2004; 52(7):1121–9. https://doi.org/10.1111/j.1532-5415.2004.52310.x PMID: 15209650
3. King MB, Judge JO, Whipple R, Wolfson L. Reliability and responsiveness of two physical performance measures examined in the context of a functional training intervention. Physical therapy. 2000; 80(1):8–16. PMID: 10623956
4. Landi F, Liperoti R, Russo A, Giovannini S, Tosato M, Capoluongo E, et al. Sarcopenia as a risk factor for falls in elderly individuals: results from the ilSIRENTE study. Clinical nutrition. 2012; 31(5):652–8. https://doi.org/10.1016/j.clnu.2012.02.007 PMID: 22414775
5. Wang C, Olson S, Protas E. Lower extremity muscle performance associated with community ambulation in elderly fallers. Asian J Gerontol Geriatr. 2009; 4(2):52–7.

6. Hashish R, Samarawickrame SD, Wang M-Y, Yu SS-Y, Salem GJ. The association between unilateral heel-rise performance with static and dynamic balance in community dwelling older adults. Geriatric Nursing. 2015; 36(1):30–4. https://doi.org/10.1016/j.gerinurse.2014.09.003 PMID: 25457285

7. Pijnappels M, Reeves ND, van Dieën JH. Identification of elderly fallers by muscle strength measures. European journal of applied physiology. 2008; 102(5):585–92. https://doi.org/10.1007/s00421-007-0613-6 PMID: 18071745

8. McGowan CP, Neptune RR, Kram R. Independent effects of weight and mass on plantar flexor activity during walking: implications for their contributions to body support and forward propulsion. Journal of applied physiology. 2008; 105(2):486–94. https://doi.org/10.1152/japplphysiol.90448.2008 PMID: 18556431

9. Fielding RA, Vellas B, Evans WJ, Bhasin S, Morley JE, Newman AB, et al. Sarcopenia: an undiagnosed condition in older adults. Current consensus definition: prevalence, etiology, and consequences. International working group on sarcopenia. Journal of the American Medical Directors Association. 2011; 12 (4):249–56. https://doi.org/10.1016/j.jamda.2011.01.003 PMID: 21527165

10. Trudelle-Jackson EJ, Jackson AW, Morrow J. Muscle strength and postural stability in healthy, older women: implications for fall prevention. Journal of Physical Activity & Health. 2006; 3(3):292.

11. Lauretani F, Russo CR, Bandinelli S, Bartali B, Cavazzini C, Di Iorio A, et al. Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. Journal of applied physiology. 2003; 95(5):1851–60. https://doi.org/10.1152/japplphysiol.00246.2003 PMID: 14555665

12. Katula JA, Rejeski WJ, Marsh AP. Enhancing quality of life in older adults: a comparison of muscular strength and power training. Health and quality of life outcomes. 2008; 6(1):45.

13. Winegard KJ, Hicks AL, Sale DG, Vandervoort AA. A 12-year follow-up study of ankle muscle function in older adults. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 1996; 51(3):B202–B7. https://doi.org/10.1093/gerona/51a.3.b202 PMID: 8630696

14. Davies C, Thomas D, White M. Mechanical properties of young and elderly human muscle. Acta Medica Scandinavica. 1986; 220(S711):219–26.

15. Simoneau E, Martin A, Van Hoecke J. Muscular performances at the ankle joint in young and elderly men. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2005; 60 (4):439–47. https://doi.org/10.1093/gerona/60.4.439 PMID: 15933381

16. Runge M, Hunter G. Determinants of musculoskeletal frailty and the risk of falls in old age. Journal of Musculoskeletal and Neuronal Interactions. 2006; 6(2):167. PMID: 16849828

17. Rikli RE, Jones CJ. Assessing physical performance in independent older adults: Issues and guidelines. Journal of aging and physical activity. 1997; 5(3):244–61.

18. Rikli R, Jones C. Development and validation of a functional fitness test for community-residing older adults. Journal of aging and physical activity. 1999; 7(2):129–61.

19. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Research quarterly for exercise and sport. 1999; 70(2):113–9. https://doi.org/10.1080/02701367.1999.10608028 PMID: 10380242

20. Lin MR, Hwang HF, Hu MH, Wu HDI, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. Journal of the American Geriatrics Society. 2004; 52(8):1343–8. https://doi.org/10.1111/j.1532-5415.2004.52366.x PMID: 15271124

21. André H-I, Carnide F, Moço A, Valamatos M-J, Ramalho F, Santos-Rocha R, et al. Can the calf-raise senior test predict functional fitness in elderly people? A validation study using electromyography, kinematics and strength tests. Physical Therapy in Sport. 2018; 32:252–9. https://doi.org/10.1016/j.ptsp.2018.05.012 PMID: 29883924

22. André H-I, Carnide F, Borja E, Ramalho F, Santos-Rocha R, Veloso AP. Calf-raise senior: a new test for assessment of plantar flexor muscle strength in older adults: protocol, validity, and reliability. Clinical interventions in aging. 2016; 11:1661. https://doi.org/10.2147/CIA.S115304 PMID: 27895473

23. Latham NK, Bennett DA, Stretton CM, Anderson CS. Systematic review of progressive resistance strength training in older adults. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2004; 59(1):M48–M61.

24. De Vos NJ, Singh NA, Ross DA, Stavrinos TM, Orr R, Fiatarone Singh MA. Optimal load for increasing muscle power during explosive resistance training in older adults. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2005; 60(5):638–47.

25. Granacher U, Gruber M, Gollhofer A. Resistance training and neuromuscular performance in seniors. International journal of sports medicine. 2009; 30(09):652–7.

26. Silva NL, Oliveira RB, Fleck SJ, Leon AC, Farinatti P. Influence of strength training variables on strength gains in adults over 55 years-old: a meta-analysis of dose–response relationships. Journal of Science

and Medicine in Sport. 2014; 17(3):337–44. https://doi.org/10.1016/j.jsams.2013.05.009 PMID: 23806877

27. Ramalho F, Carnide F, Santos-Rocha R, André H-I, Moniz-Pereira V, Machado ML, et al. Community-based exercise intervention for gait and functional fitness improvement in an older population: study protocol. Journal of aging and physical activity. 2017; 25(1):84–93. https://doi.org/10.1123/japa.2015-0290 PMID: 27623478
28. Valente S. Validação de um Questionário de Saúde e Identificação de Factores de Risco de Quedas para a População Idosa Portuguesa. MS Thesis, Faculty of Human Kinetics—University of Lisbon; 2013. Available from: http://hdl.handle.net/10400.5/5202
29. McCarthy EK, Horvat MA, Holtsberg PA, Wisenbaker JM. Repeated chair stands as a measure of lower limb strength in sexagenarian women. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2004; 59(11):1207–12.
30. van Iersel MB, Munneke M, Esselink RA, Benraad CE, Rikkert MGO. Gait velocity and the Timed-Up-and-Go test were sensitive to changes in mobility in frail elderly patients. Journal of clinical epidemiology. 2008; 61(2):186–91. https://doi.org/10.1016/j.jclinepi.2007.04.016 PMID: 18177792
31. Cohen J. Statistical power analysis for the behavioral sciences: Academic press; 2013.
32. Hopkins WG. Measures of reliability in sports medicine and science. Sports medicine. 2000; 30(1):1–15. https://doi.org/10.2165/00007256-200030010-00001 PMID: 10907753
33. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. Journal of clinical epidemiology. 2003; 56(5):395–407. https://doi.org/10.1016/s0895-4356(03)00044-1 PMID: 12812812
34. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health and quality of life outcomes. 2006; 4(1):54.
35. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. Physical therapy. 2006; 86(5):735–43. PMID: 16649896
36. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. BMC musculoskeletal disorders. 2006; 7(1):82.
37. de Yébenes Prous MJG, Salvanés FR, Ortells LC. Responsiveness of outcome measures. Reumatología Clínica (English Edition). 2008; 4(6):240–7.
38. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures statistics and strategies for evaluation. Controlled clinical trials. 1991; 12(4):S142–S58.
39. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires. Sports Medicine. 2010; 40(7):525–37. https://doi.org/10.2165/11531370-000000000-00000 PMID: 20545379
40. Vieira AC, Moniz S, Fernandes R, Carnide F, Cruz EB. Responsiveness and interpretability of the Portuguese version of the Quebec Back Pain Disability Scale in patients with chronic low back pain. Spine. 2014; 39(5):E346–E52. https://doi.org/10.1097/BRS.0000000000000159 PMID: 24573078
41. Winter JE, MacInnis RJ, Wattanapenpaiboon N, Nowson CA. BMI and all-cause mortality in older adults: a meta-analysis. The American journal of clinical nutrition. 2014; 99(4):875–90. https://doi.org/10.3945/ajcn.113.068122 PMID: 24452240
42. Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Ni P, Slavin MD, et al. Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. Physical therapy. 2012; 92(3):388–97. https://doi.org/10.2522/ptj.20100398 PMID: 22114200
43. Pereira M, Cruz EB, Domingues L, Duarte S, Carnide F, Fernandes R. Responsiveness and interpretability of the Portuguese version of the Neck Disability Index in patients with chronic neck pain undergoing physiotherapy. Spine. 2015; 40(22):E1180–E6. https://doi.org/10.1097/BRS.0000000000001034 PMID: 26110663
44. Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy–Lung (FACT-L) questionnaire?: results from Eastern Cooperative Oncology Group (ECOG) Study 5592. Journal of clinical epidemiology. 2002; 55(3):285–95. https://doi.org/10.1016/s0895-4356(01)00477-2 PMID: 11864800
45. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. Journal of clinical epidemiology. 1999; 52(9):861–73. https://doi.org/10.1016/s0895-4356(99)00071-2 PMID: 10529027
46. Applebaum EV, Breton D, Feng ZW, Ta AT, Walsh K, Chassé K, et al. Modified 30-second Sit to Stand test predicts falls in a cohort of institutionalized older veterans. PLoS One. 2017; 12(5):e0176946. https://doi.org/10.1371/journal.pone.0176946 PMID: 28464024

47. Hislop H, Avers D, Brown M. Daniels and Worthingham's muscle Testing-E-Book: Techniques of manual examination and performance testing: Elsevier Health Sciences; 2013

48. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of clinical epidemiology. 2008; 61(2):102–9. https://doi.org/10.1016/j.jclinepi.2007.03.012 PMID: 18177782