Quantifying throwing load in handball: a method for measuring the number of throws

Skejø, Sebastian Deisting; Liaghat, Behnam; Jakobsen, Claes Christian; Møller, Merete; Bencke, Jesper; Papi, Giovanni; Kunwald, Nikolaj Pelle; Sørensen, Henrik

Published in:
Sports Biomechanics

DOI:
10.1080/14763141.2021.1951345

Publication date:
2021

Document version:
Submitted manuscript

Received: 26 February 2021
For correspondence:
sdsk@ph.au.dk

Quantifying throwing load in handball: a method for measuring the number of throws

Sebastian Deisting Skejø[a], Behnam Liaghat[b], Claes Christian Jakobsen[c], Merete Møller[b], Jesper Bencke[d], Giovanni Papi[a], Nikolaj Pelle Kunwald[a], Henrik Sørensen[a]

[a]Department of Public Health, Aarhus University, Aarhus, Denmark, [b]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark,[c]CC Athletics ApS, Odense, Denmark, [d]Human Movement Analysis Laboratory, Dep. Orthopaedic Surgery, Copenhagen University Hospital at Amager-Hvidovre, Denmark

Please cite as: Skejø, S., Liaghat, B. Jakobsen, C., Møller, M., Bencke, J., Papi, G., Kunwald, N & Sørensen, H. (2021). Quantifying throwing load in handball: a method for measuring the number of throws. *SportRχiv*. https://doi.org/10.31236/osf.io/87wsx

All authors have read and approved this version of the manuscript. This article was last modified on February 26, 2021.

Authors Sebastian Deisting Skejø @sebastianskejoe, Behnam Liaghat @behnam_liaghat, Claes Christian Jakobsen @Makstrero, Merete Møller @Merete_Moller and Jesper Bencke @jesperbencke can be reached on Twitter.

ABSTRACT

Shoulder injuries are a common problem in handball. One likely cause of such injuries is excessive throwing. However, it is difficult to measure the number of player throws in large cohort studies using existing methods accurately. Therefore, the purpose of this study is to develop and validate a method for identifying overhead throws using a low-cost inertial measurement unit (IMU) worn on the wrist. In a two-stage approach, we developed a threshold-based automatic identification method of overhead throws in a laboratory study using the IMU. Subsequently, we validated the suggested thresholds in a field setting by comparing throws identified by the threshold-method to throws identified by video recordings of handball practices. The best set of threshold values resulted in a per-player median sensitivity of 100% (range: 84-100%) and a median positive predictive value (PPV) of 96% (range: 86-100%) in the development study. In the validation study, the per-player median sensitivity dropped to 78% sensitivity (range: 52-91%), while the per-player median PPV dropped to 79% (range: 47-90%). The proposed method is a promising method for automatically identifying handball throws in a cheap and feasible way.

Keywords: inertial measurement unit, throwing load, handball exposure

INTRODUCTION

Previous handball studies have reported lifetime shoulder pain prevalence ranging from 41 to 75% and seasonal shoulder pain prevalence between 17 and 52%.[1–5] The incidence of shoulder injuries rank amongst the highest in handball[6] with reported incidence rates ranging from 0.2 to 1.4 per 1000 playing hours.[6–9] Therefore, it is essential to develop ways to prevent such injuries, which requires an in-depth understanding of the causes of shoulder injuries.[10]

When trying to understand why shoulder injuries occur in handball, it is imperative to accurately measure the exposure that puts players at risk for such injuries.[11] Many movements can place load on the shoulder structures, but the most common seems to be throwing. For instance, one study by Prestkvern et al. reported that Norwegian handball players make more than 100 throws per hour during practices.[12] Their method consisted of reviewing video footage to determine the number of throws, which seems to be an infeasible approach in studies with large sample sizes. An alternative approach, which is both relatively cheap and easy to scale, could be using wearable devices to identify throws. In baseball[13,14] and cricket[15], wearable devices have shown excellent ability to identify throws, but these devices do not seem immediately transferable to handball. For instance, the baseball studies[13,14] placed the device on the upper arm, which may be problematic in handball due to large amounts of physical contact[16], and the cricket study[15] relied on GPS measurements, which are difficult to obtain indoor. Additionally, handball players throw with more varied arm positions than in baseball and cricket. Apart from the usual overhead arm position, another relatively common throw is the sidearm throw.[17] Other variations include “flicking” the ball to another player using only the wrist or making a “push pass” to a pivot through the legs of a defender. Of these variations, the overhead throw appears to be most important for injury research for two reasons: 1) they seem to be the most common variant, although no previous research have established this, and 2) most of the high-intensity throws are likely overhead throws, while the non-overhead throws have a larger range of intensity.

To the best of our knowledge, there have been no attempts to develop a method for automatically identifying handball throws.[18] Therefore, the primary aim of the present study is to develop and validate a method to identify overhead throws in handball players based on measurements from a wearable inertial measurement unit (IMU) placed on the forearm. Our secondary aim is to investigate whether the proposed method can also identify non-overhead throws in handball.

MATERIALS AND METHODS

Study approach and design

We developed and validated the proposed method in two separate studies. In both studies, we used an inertial measurement unit (IMU) to measure linear acceleration and angular velocity of the forearm. These measurements formed the basis of identifying throws from non-throws.

The first study was a development study where participants completed a protocol simulating common handball movements including different types of overhead throws. The aim of this study was to establish threshold values for acceleration and angular velocity that could distinguish throws from non-throws. As the golden standard to which we compared automatically identified throws, we used the timestamps of the actual throws that were given by the protocol.

The second study was a validation study aiming to validate the threshold values established in the development study. We video recorded handball practices and manually identified throws in this footage. We then used the manually identified throws as a golden standard to which we compared the automatically identified throws in the IMU-signal.

Both studies were assessed by the Central Denmark Region Committees on Health Research Ethics, who waived the need for formal approval (journal number 1-10-72-1-20) and approved according to the Data Protection Act by the local Data Protection unit at Aarhus University (journal number 2016-051-000001).

Development study

Participants

We recruited ten active or former handball players (five males and five females) aged 18 years or older. The only exclusion criterion was any shoulder pain at the time of recording. Players gave informed consent before participating.

Data collection

To collect data, we used a custom measuring device containing an ADXL377 accelerometer (range ± 200g) and an ICM20600 IMU, from which we sampled the gyroscope (range ± 2000 °/s). Both accelerometer and gyroscope were sampled at 200 Hz to a built-in microSD card.

Following a standardized warm-up consisting of jogging, skipping, sidestepping, arm swings, push-ups, throws with increasing intensity and one round of the protocol (not recorded), the measurement device was strapped to the wrist of the player (Figure 1). Players then completed ten rounds of a two-minute protocol (Table 1) aiming to simulate common handball movements. During each round, the player threw five overhead throws, resulting in 50 throws over the entire recording.

Table 1. The protocol used in the development study. Participants completed 10 rounds of the protocol back-to-back, and all throws were overhead throws.

Action	Time period	Instruction to participant
Change of direction	0:00-0:05	Run without the ball and change direction once in each direction
Standing throw without run-up and whip-like wind-up	0:15	Throw the ball as hard as possible into the net.
Clap	0:25	Clap your hands three times
Jump throw	0:35	Throw the ball as hard as possible into a net with a jump throw.
Simulated fall	0:50	Sit on your knees and fall to the ground while bracing with your hands.
Standing throw without run-up and circular wind-up	1:05	Throw the ball as hard as possible into the net.
Run and dribble	1:15-1:20	Run around while dribbling.
Low intensity standing throw and catch without run-up	1:30	Make a short pass towards a wall and catch the ball as it comes back

High intensity standing throw without run-up and catch	1:40	Make a long pass towards a wall and catch the ball as it comes back
10m sprint	1:50	Sprint as fast as you can towards a cone.

Figure 1. Placement of the measurement device.

Data analysis

For each recording, we analyzed each sample from the gyroscope to see if the absolute value of any of the three axes surpassed a threshold, t_{gyro}, which we defined as a gyroscope hit and checked if the total acceleration surpassed another threshold, t_{acc}, within a time window of ±250 ms of the gyroscope hit. If both t_{gyro} and t_{acc} were surpassed, a throw was registered at the time of the gyroscope hit and the next full second of recording was skipped to avoid duplicate hits.

If the registered throw was within ±3 seconds of an actual throw (as per the protocol), we recorded a true positive. If not, we recorded a false positive. After checking the entire recording, we recorded the actual throws that were not associated with a true positive as false negatives.

To establish the best set of threshold values, we pooled all throws and analyzed the dataset multiple times, but with different threshold values. All combinations of $t_{gyro} = 0, 100, 200, \dots, 2000$ °/s and $t_{acc} = 0, 10, 20, \dots, 200$ g were checked. We then computed the sensitivity, i.e. the proportion of actual throws that was correctly identified, and the positive predictive value (PPV), i.e. the proportion of identified throws that was actual throws, for all sets of threshold values. The set that produced the highest F_1 score as given by

$$F_1 = 2 * \frac{Sensitivity * PPV}{Sensitivity + PPV}$$

was selected as the set of threshold values used in the validation study.

Validation study

Sample size calculation and participants

Conservatively, we assumed that each player would make 100 throws during a practice[12], and that the player-specific sensitivity would follow a β(20,6)-distribution. Simulation indicated that we would need to include 13 players to have a 95% chance of seeing 95% confidence interval widths for the overall sensitivity of 5% or less (see Appendix 1). Since we had two devices, we included 14 players (six males and eight females) aged 18 years or older from teams in the Central Jutland region of Denmark. All players gave informed consent beforehand.

Data collection

Before commencing a regular handball practice, we strapped a measurement device similar to the one used the development study to the wrist of the player (Figure 1), and the player participated in the regular practice. Throughout the entire practice, the measurement device recorded linear acceleration and angular velocity.

Simultaneously, we video recorded the entire training using a single, wide-angle camera (Hero 4, GoPro, Inc., USA). Subsequently, one author (SDS) reviewed all video recordings, to identify the timestamp of all throws manually. In this regard, we registered any throw from one player towards either another player or the goal. In addition to registering the time stamp, we also assessed the arm position of the throw. A priori, we defined the possible arm positions as overhead, double overhead (a variation sometimes used during the throwing warm-up) or non-overhead.

Time synchronization

As the IMU and video recordings did not start at the exact same time, we synchronized the recordings by manually locating an identified throw in the IMU recording that corresponded to an actual throw in the video recording. Then we realigned the timescales to align the identified throw with the actual throw.

Data analysis

Following synchronization, we identified throws in the IMU-signal using the same threshold-based approach as in the development study, but considering only the set of threshold values that produced the best performance in the development study. We then compared the timestamps of the automatically identified throws to the timestamps of the actual throws, which were manually identified throws in the corresponding video recording. As in the development study, we recorded a true positive if an automatically identified throw was within ±3 seconds of an actual throw. Subsequently, we removed the associated actual throw from further consideration in the analysis to avoid the same actual throw being associated with different identified throws. If an identified throw was not within ±3 seconds of an actual throw, we recorded a false positive. We recorded all actual throws that were not associated to an identified throw as false negatives.

For the primary outcome, we only considered the subset of actual throws classified as overhead throws, while all throws, irrespective of arm position, were included in the secondary analysis.

Finally, we computed the median and range of per-player sensitivities and PPVs.

RESULTS

Development study

Considering the per-player best performance as judged by the F_1 score, the median sensitivity was 100% (range: 84-100%) and the median PPV was 96% (range: 86-100%). When pooling all throws, the best set of threshold values was $(t_{gyro}, t_{acc}) = (1700\ °/s, 10\ g)$ which yielded an F_1 score of 93% (sensitivity: 90%; PPV: 96%). Figure 3 shows how the sensitivity, PPV and F_1 score changes as t_{gyro} and t_{acc} are varied.

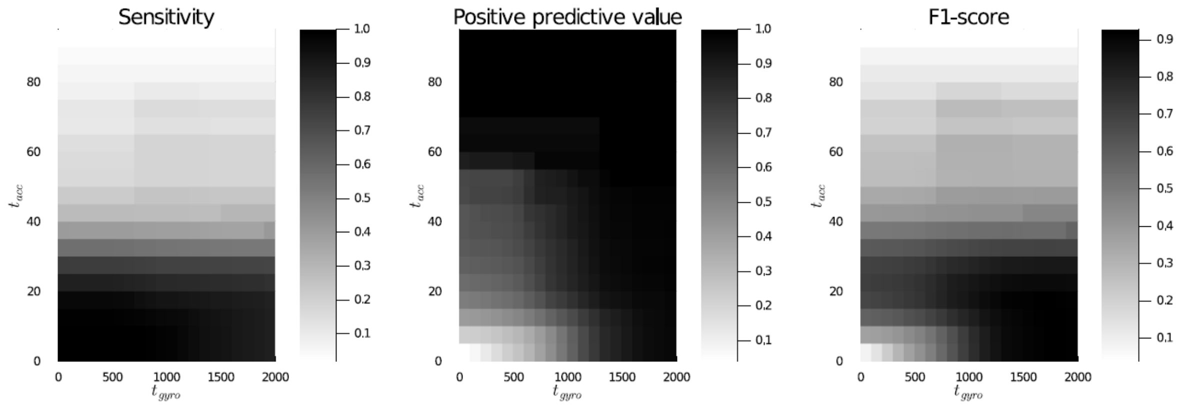

Figure 2. Changes in sensitivity, positive predictive value and F_1 score as the threshold values, t_{gyro} and t_{acc}, are varied.

Validation study

We recruited eight females (seven backcourt players and one wing player) and six males (three backcourt players, two pivots and one wing player). All played at the highest national level. The median practice time was 85.5 minutes (range: 70.4-117.6), and players threw a median of 146.5 throws (range: 51-226) of which a median of 121.5 were classified as overhead throws (range: 34-186).

Considering only overhead throws, the median sensitivity was 78% (range: 52-91%), and the median PPV was 79% (range: 47-90%). Considering all types of throws, the median sensitivity was 69% (range: 35-82%), and the median PPV was 83% (range: 47-92%). There was a large variation in per-player results (see Table 2). Figure 3 compares the peak total accelerations measured by the IMU at the time of each actual throw ± 500 ms for all true positives and all false negatives.

Table 2. Per-player performance in the validation study.

Player	Overhead throws			All types of throws		
	Sensitivity	PPV	Number of actual throws	Sensitivity	PPV	Number of actual throws
1	91%	74%	67	80%	78%	80
2	89%	81%	186	82%	90%	226
3	88%	77%	117	74%	84%	152
4	88%	79%	129	82%	81%	142
5	86%	82%	88	70%	83%	110
6	86%	79%	166	75%	83%	201
7	80%	69%	64	77%	73%	70
8	77%	78%	113	67%	78%	129
9	76%	82%	149	66%	88%	186
10	73%	54%	112	67%	62%	138
11	73%	90%	142	69%	91%	151
12	69%	66%	126	65%	84%	169
13	53%	47%	34	35%	47%	51
14	52%	89%	150	49%	92%	167

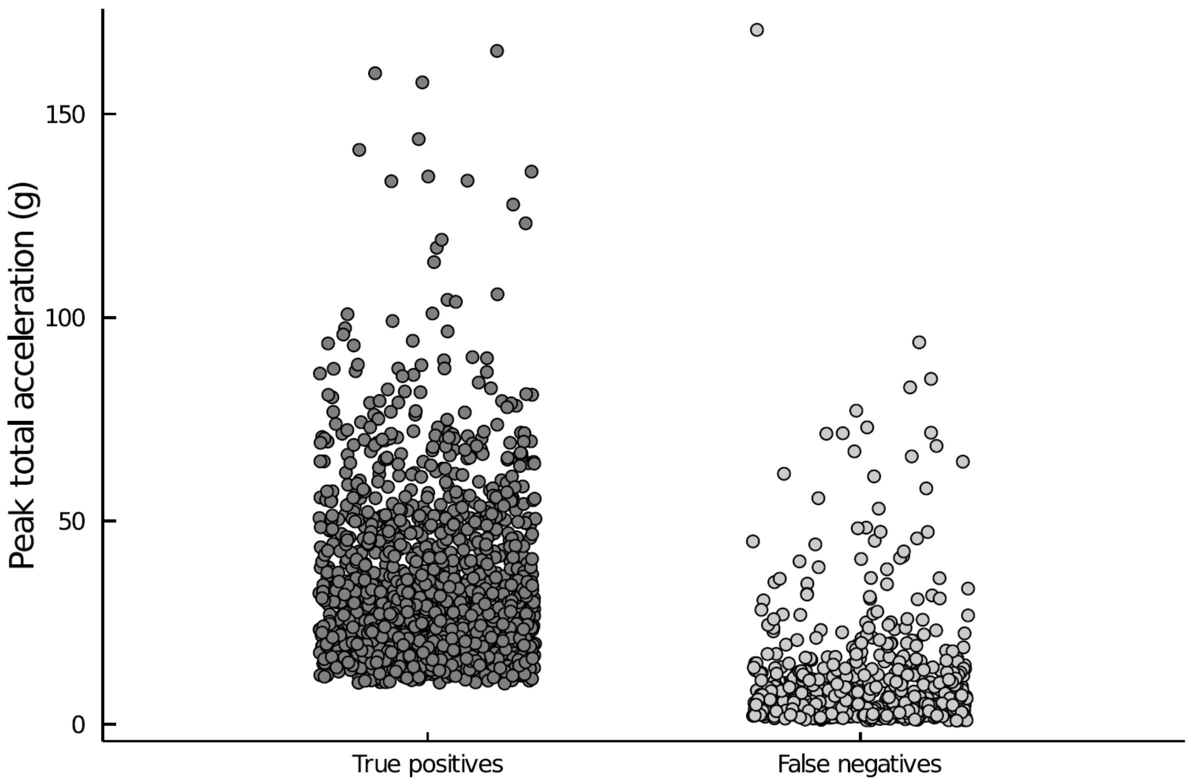

Figure 3. Comparison of the peak total acceleration measured by the IMU at time of an actual throw ± 500 ms.

DISCUSSION

This is the first study to develop and validate a method for counting handball throws using an IMU. In the development study, we found a median best per-player sensitivity of 100% and a median best per-subject PPV of 96%. This is higher or equal to what other methods have found in studies of baseball and cricket[13–15] where the sensitivity was between 94 and 98%, and the PPV was between 81 and 95%. However, in the validation study the sensitivity was lower with a median sensitivity of 79% for overhead throws. Similarly, the PPV dropped to 79%.

Despite the notable decrease in sensitivity and PPV in the validation study, our proposed method represents a significant step forward in the quantification of throwing load volume in handball. We will use the following example to demonstrate this. Consider, for instance, two players, A and B, who participated in the same practice in the validation study. The usual approach to estimating exposure time would consider only the playing time,[7,9] therefore estimating the same exposure for both players. Compare this to our method which estimated that player A made 93 throws and that player B threw 38 times, indicating 2.4 times more throwing-related exposure for player A than for player B. The video analysis revealed that player A made 88 overhead throws – 2.6 times more throws than player B who made 34 overhead throws.

Development vs validation performance

A likely explanation for the performance drop in sensitivity is that players may have more low-intensity throws in the validation study than in the development study. Since peak total acceleration predicts throwing speed,[16] the accelerometer threshold of 10 g used in the validation study excludes throws with a speed of approximately 9.8 m/s or lower. Indeed, as shown in figure 3, the false negatives were, on average, thrown with a lower intensity than the true positives. Since shoulder load likely increases with throwing intensity, low-intensity throws are less likely to have an important impact on shoulder injury risk. Therefore, the false negatives are less of a problem if the device is used to understand throwing load's impact on shoulder injury risk.

Another likely explanation for the difference between the development and the validation study may be the chaotic nature of handball play. While we attempted to simulate a wide range of common handball movements in the development study, a lab protocol is unlikely to capture the complexities of real handball play perfectly. For instance, the development study only contained overhead throws, while we observed a wide range of different arm positions and throwing types in the validation study. While some of the non-overhead throws certainly could represent a significant shoulder load, many of the recorded non-overhead throws, e.g. wrist flicks and push passes, did not. Therefore, we argue that the results concerning only the overhead throws are most relevant when it comes to estimating shoulder load using our proposed method.

A third potential explanation for the performance drop is the difference between players in the development study and players in the validation study. While some players in the development study were no longer active and/or did not have experience at the highest national level, all players in the validation study were active at highest national level. . Since elite players throw with a higher speed[19], there is a risk that the threshold values selected for the validation study are lower because of the potential of lower throwing speeds in the development study. Lower threshold values likely lead to more false positives, and for this reason, player differences might explain some of the drop in positive predictive value, but not in sensitivity.

Finally, there is a risk that some overhead throws were classified as non-overhead throws and vice versa. Likewise, other errors might have affected the performance in the validation study such as not noting the timestamp of an actual throw. However, these errors could only serve to decrease the performance of our proposed method. Only if we incorrectly identified non-throws as throws at a time where the IMU detected a false positive, would errors falsely improve performance. As such, it seems likely that the sensitivities and positive predictive values presented here are estimates of the lower bounds on the true median performance.

Per-player variability

We noted a relatively large range of per-player sensitivity and positive predictive value. For instance, player 14 had the lowest sensitivity of all players when considering only overhead throws, while maintaining a high positive predictive value. This could indicate that the threshold values were too high for this particular player, resulting in many false negatives, but few false positives (see Figure 2). It is possible that a simple calibration procedure, such as the one used in the development study, could be used to find individualized threshold values. However, since the overall performance seemed good (11/14 players had a sensitivity of 70% or higher), we speculate that the potential performance increase would not be worth the time spent by individually calibrating the method. Further studies could investigate whether this speculation holds true.

On the other hand, player 13 showed both low sensitivity and low positive predictive value. This might indicate that lowering the thresholds would not solve the performance problems. However, it is worth noting that player 13 also had the fewest throws of all players (only 34 overhead throws), which could be explained by the fact that player 13 mostly played defense in the practice we recorded. As such, it seems plausible that making few high-intensity throws could explain the low sensitivity as well as the low positive predictive value, since the positive predictive value depends on the prevalence,[20] i.e. the number of throws. Despite the low positive predictive value, the method still estimated only 38 throws for player 13, which makes it more unlikely that the error would have an impact in the decision making of whether that player is at risk of shoulder injury or not. As such, the low positive predictive value may not be clinically impactful for players with few throws such as defenders.

In some cases, a researcher may want to only estimate the number of throws, but not engage in further processing of the identified throws, e.g. estimate speed of the throws.[16] In this case, the relationship between sensitivity and PPV determines whether the number of throws is over-, under- or correctly estimated as according to the mathematical definitions of sensitivity and PPV.[21] If sensitivity is higher than PPV, the method overestimates the number of throws; if PPV is higher than sensitivity, the method underestimates the number of throws; if sensitivity and PPV are equal, the method correctly estimates the number of throws. In our sample, the median difference between sensitivity and PPV was 6.5%-point with sensitivity being higher. These findings would suggest that the method slightly overestimates the number of throws, but future studies are needed to firmly establish whether this is the case.

In the present article, we have focused solely on identifying throws. Consequently, the present results only relate to what has been termed training amount.[22] However, the IMU may be able to provide additional information related to other aspects of training load. For instance, training load magnitude may be estimated either by using the raw acceleration and angular velocities or by computing the throwing speed from the peak total acceleration.[16] Similarly, it may be possible to estimate distribution-related variables such as approach type[23] using only an IMU. In this light, a primary task for future studies is to develop ways to incorporate all these different aspects into a single exposure that is suitable for use in epidemiological studies.

In conclusion, while we did observe a performance drop in the validation studies, our method is accurate and provides a better estimate of throwing load volume than playing time. Our method is relatively cheap and easy to use in large cohorts, making it a prime candidate to measure exposure in large epidemiological studies of shoulder injuries in handball.

CONTRIBUTIONS

Contributed to the conception and design: All
Acquisition of data: SDS, GP, NPK
Analysis and interpretation of the data: SDS
Drafted and/or revised the article: All
Approved the submitted version for publication: All

DATA AND SUPPLEMENTARY MATERIAL ACCESSIBILITY

Due to the use of video recordings, the data is on available upon request and in concordance with the regulations laid out by the General Data Protection Regulation.

REFERENCES

1. Andersson SH, Bahr R, Clarsen B, Myklebust G. Risk factors for overuse shoulder injuries in a mixed-sex cohort of 329 elite handball players: previous findings could not be confirmed. *British Journal of Sports Medicine*. 2018;52(18):1191-1198. doi:10.1136/bjsports-2017-097648

2. Asker M, Holm LW, Källberg H, Waldén M, Skillgate E. Female adolescent elite handball players are more susceptible to shoulder problems than their male counterparts. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2018;26(7):1-9. doi:10.1007/s00167-018-4857-y

3. Clarsen B, Bahr R, Andersson SH, Munk R, Myklebust G. Reduced glenohumeral rotation, external rotation weakness and scapular dyskinesis are risk factors for shoulder injuries among elite male handball players: a prospective cohort study. *British Journal of Sports Medicine*. 2014;48(17):1327-1333. doi:10.1136/bjsports-2014-093702

4. Forthomme B, Croisier J-L, Delvaux F, Kaux J-F, Crielaard J-M, Gleizes-Cervera S. Preseason Strength Assessment of the Rotator Muscles and Shoulder Injury in Handball Players. *Journal of Athletic Training*. 2018;53(2):174-180. doi:10.4085/1062-6050-216-16

5. Myklebust G, Hasslan L, Bahr R, Steffen K. High prevalence of shoulder pain among elite Norwegian female handball players. *Scandinavian Journal of Medicine & Science in Sports*. 2013;23(3):288-294. doi:10.1111/j.1600-0838.2011.01398.x

6. Møller M, Attermann J, Myklebust G, Wedderkopp N. Injury risk in Danish youth and senior elite handball using a new SMS text messages approach. *British Journal of Sports Medicine*. 2012;46(7):531-537. doi:10.1136/bjsports-2012-091022

7. Asker M, Waldén M, Källberg H, Holm LW, Skillgate E. Preseason Clinical Shoulder Test Results and Shoulder Injury Rate in Adolescent Handball: A Prospective Study on Female and Male Elite Players. *Journal of Orthopaedic & Sports Physical Therapy*. 2020;50(2):67-74. doi:10.2519/jospt.2020.9044

8. Edouard P, Degache F, Oullion R, Plessis J-Y, Gleizes-Cervera S, Calmels P. Shoulder Strength Imbalances as Injury Risk in Handball. *International Journal of Sports Medicine*. 2013;34(07):654-660. doi:10.1055/s-0032-1312587

9. Møller M, Nielsen RO, Attermann J, et al. Handball load and shoulder injury rate: a 31-week cohort study of 679 elite youth handball players. *British Journal of Sports Medicine*. 2017;51(4):231-237. doi:10.1136/bjsports-2016-096927

10. Finch C. A new framework for research leading to sports injury prevention. *Journal of Science and Medicine in Sport*. 2006;9(1-2):3-9. doi:10.1016/j.jsams.2006.02.009

11. Stovitz SD, Shrier I. Injury rates in team sport events: tackling challenges in assessing exposure time. *Br J Sports Med*. 2012;46(14):960-963. doi:10.1136/bjsports-2011-090693

12. Prestkvern SR, Clarsen B, Myklebust G. *Er det en sammenheng mellom pasnings- og skuddeksponering og skulderproblemer i norsk herrehåndball?* Norges idrettshøgskole; 2013. Accessed December 18, 2020. https://nih.brage.unit.no/nih-xmlui/handle/11250/171860

13. Rawashdeh SA, Rafeldt DA, Uhl TL. Wearable IMU for Shoulder Injury Prevention in Overhead Sports. *Sensors*. 2016;16(11):1847. doi:10.3390/s16111847

14. Rose MJ, McCollum KA, Freehill MT, Cain SM. Quantifying Throw Counts and Intensities Throughout a Season in Youth Baseball Players: A Pilot Study. *Journal of Biomechanical Engineering*. 2021;143(3):034502. doi:10.1115/1.4049025

15. McNamara DJ, Gabbett TJ, Chapman P, Naughton G, Farhart P. The Validity of Microsensors to Automatically Detect Bowling Events and Counts in Cricket Fast Bowlers. *International Journal of Sports Physiology and Performance*. 2015;10:71-75.

16. Skejø SD, Bencke J, Møller M, Sørensen H. Estimating Throwing Speed in Handball Using a Wearable Device. *Sensors*. 2020;20(17):4925. doi:10.3390/s20174925

17. Wagner H, Buchecker M, von Duvillard SP, Müller E. Kinematic Comparison of Team Handball Throwing With Two Different Arm Positions. *International Journal of Sports Physiology & Performance*. 2010;5(4):469-483.

18. Black GM, Gabbett TJ, Cole MH, Naughton G. Monitoring Workload in Throwing-Dominant Sports: A Systematic Review. *Sports Medicine*. Published online April 12, 2016. doi:10.1007/s40279-016-0529-6

19. Wagner H, Buchecker M, von Duvillard SP, Müller E. Kinematic description of elite vs. low level players in team-handball jump throw. *Journal of Sports Science & Medicine*. 2010;9(1):15-23.

20. Mischlinger J, Schernhammer E. A common trap of diagnostic tests: Disease prevalence and positive predictive value. *Wien Klin Wochenschr*. 2017;129(15-16):583-584. doi:10.1007/s00508-017-1204-0

21. Eusebi P. Diagnostic Accuracy Measures. *Cerebrovasc Dis*. 2013;36(4):267-272. doi:10.1159/000353863

22. Nielsen RO, Bertelsen ML, Møller M, et al. Methods matter: exploring the 'too much, too soon' theory, part 1: causal questions in sports injury research. *Br J Sports Med*. 2020;54(18):1119-1122. doi:10.1136/bjsports-2018-100245

23. van den Tillaar R, Bhandurge S, Stewart T. Detection of Different Throw Types and Ball Velocity with IMUs and Machine Learning in Team Handball. *ISBS Proceedings Archive*. 2020;38(1). https://commons.nmu.edu/isbs/vol38/iss1/48

APPENDIX 1: SAMPLE SIZE SIMULATION

We simulated the required sample size using the following R script, R version 4.0.3 and the R packages GenBinomApps (version 1.1) and ggplot2 (version 3.3.2).

```
require(GenBinomApps)

# The distribution from which we pull sensitivities
alpha = 20
beta = 6
plot(dbeta(seq(0,1,length=100),alpha, beta))

m = 100 # Number of throws per session

results = data.frame(N=numeric(0), CIwidth=numeric(0))
# Run 1000 simulations
for (simno in 1:1000) {

  # Check between 2 and 20 players
  for (N in 2:20) {

    df = data.frame(ID=numeric(0), TPs=numeric(0))

    # Sample sensitivity for each player
    for (id in 1:N) {
      # Sample sensitivity
      Sensitivity = rbeta(1, alpha, beta)

      TPs = round(m*Sensitivity) # Number of true positives

      df <- rbind(df, data.frame(ID=id, TPs=TPs))
    }

    # Compute CI width
    total = m*N
    ci = clopper.pearson.ci(sum(df$TPs), total, alpha=0.05, CI="two.sided")
    results <- rbind(results, data.frame(N=N, CIwidth=(ci$Upper.limit-
ci$Lower.limit)))
  }
}

table(subset(results, CIwidth < 0.05, select=c("N")))

require(ggplot2)
ggplot(results, aes(x=N, y=CIwidth)) + geom_point() +
geom_abline(intercept=0.05, slope=0) +
  theme_minimal()
```