



OPEN

Implications of sample size and acquired number of steps to investigate running biomechanics

Anderson Souza Oliveira[✉] & Cristina Ioana Pircscoeanu

Low reproducibility and non-optimal sample sizes are current concerns in scientific research, especially within human movement studies. Therefore, this study aimed to examine the implications of different sample sizes and number of steps on data variability and statistical outcomes from kinematic and kinetics running biomechanical variables. Forty-four participants ran overground using their preferred technique (normal) and minimizing the contact sound volume (silent). Running speed, peak vertical, braking forces, and vertical average loading rate were extracted from > 40 steps/runner. Data stability was computed using a sequential estimation technique. Statistical outcomes (p values and effect sizes) from the comparison normal vs silent running were extracted from 100,000 random samples, using various combinations of sample size (from 10 to 40 runners) and number of steps (from 5 to 40 steps). The results showed that only 35% of the study sample could reach average stability using up to 10 steps across all biomechanical variables. The loading rate was consistently significantly lower during silent running compared to normal running, with large effect sizes across all combinations. However, variables presenting small or medium effect sizes (running speed and peak braking force), required > 20 runners to reach significant differences. Therefore, varying sample sizes and number of steps are shown to influence the normal vs silent running statistical outcomes in a variable-dependent manner. Based on our results, we recommend that studies involving analysis of traditional running biomechanical variables use a minimum of 25 participants and 25 steps from each participant to provide appropriate data stability and statistical power.

There is a growing concern regarding the actual quality and relevance of published studies in all research fields. Factors such as low reproducibility and non-optimal sample sizes have been at the forefront of discussions in different research areas^{1,2}. However, increasing concerns have been recently reported for human movement biomechanics, where sample sizes and the number of events recorded (trials or steps) might be insufficient to support the tested hypotheses²⁻⁵. On the core of this issue is the tradeoff between experiment duration and ecological validity, as the scientific community has no standardized experimental protocols to acquire “natural” human movement. Cyclical activities such as walking and running consist of repeating a defined movement pattern^{1,6}. However, identical kinematics/kinetics performance in cyclical movements, like walking and running, are unlikely for humans due to intrinsic and extrinsic factors, such as equipment, surface, and/or metabolism^{1,6}. Therefore, acquiring a number of repetitions (steps) to represent the variability of such a movement pattern is crucial to assure ecological validity. Calculating the peak of averaged data segments representing trials/steps is therefore an “acceptable” approximation of the recorded natural pattern⁷. However, calculating the average across the peaks identified from individual trials/steps provides better data representations^{8,9}.

Studies regarding running biomechanics in the past 20 years have been defining the number of running steps based on previous literature. For instance, a guideline to achieve 90% statistical power in sports biomechanics proposed that by evaluating 5, 10, or 20 participants would require 10, 5, and 3 steps respectively³. This early recommendation has been extensively used in running biomechanics for various types of biomechanical variables^{2,10-12}. Recommendations for an optimal number of recorded steps have been proposed using sequential estimation technique (SET)^{2,7}, from which the ideal number of steps range from 13 to 20^{2,11}. However, these variables may not represent the variability contained in running kinematics/kinetics. Inter-stride running variability

Department of Materials and Production, Aalborg University, Fibigerstræde 16, Building 4, 9220 Aalborg E, Denmark.
✉ email: oliveira@mp.aau.dk

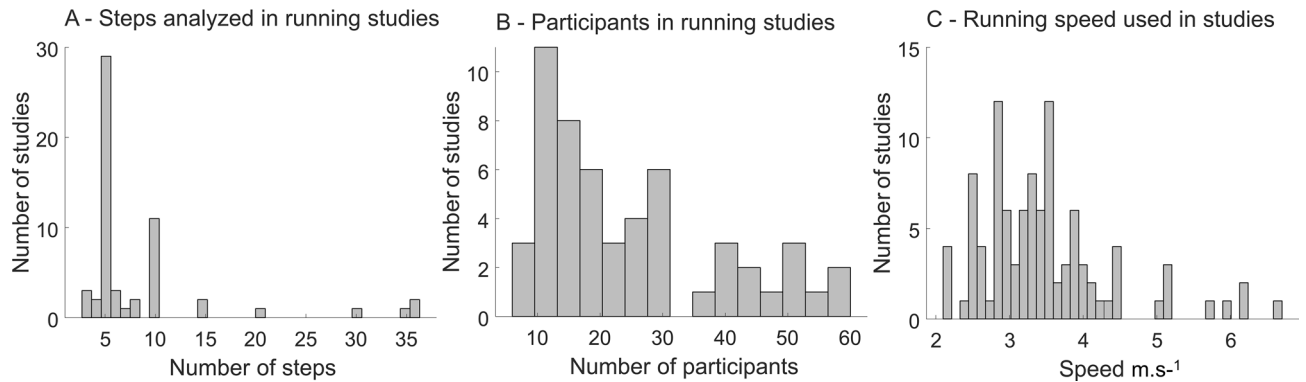


Figure 1. Distribution of running biomechanics studies describing the number of steps (A), participants (B), and running speed (C) from 2000 until present days. In total 60 studies were included in a short literature review (see Supplementary File 1). Some studies were not included in the plots to maximize visualization quality: In (A) Wouda et al. (2018)⁴³ reported around 200 steps in their analysis. In (B) Bredeweg et al. (2013)⁴⁴ and Stiffler-Joachim et al. (2019)⁴⁵ reported 170 and 210 participants, respectively. In (C) Stiles et al. (2013)⁴⁶ reported a running cadence of 130–140 steps/min, respectively 145–160 steps/min, as well as Morin et al. (2011)⁴⁷, Rowlands and Stiles⁴⁸ (2012) and Pollard et al. (2018)⁴⁹ used the preferred running speed of the participants.

is related to fluctuations in the control of motor function¹³ and the inherent complexity of the motor system^{10,13}. As human locomotion can present considerable variability, it is highly relevant to take such factors into account when defining the amount of data to be acquired for running biomechanical studies.

In sports biomechanics, there are two types of parameters: global parameters, defined as the output of the human system as an entity (i.e., speed, contact time, step length, center of mass); and localized parameters, defined as output related to movement performance (i.e., joint angles, moments, power, ground reaction force)¹. Furthermore, a decreased variability of a global parameter does not directly imply reduced variability in movement performance parameters¹. Therefore, it has been suggested that special attention needs to be devoted to the variability of both global and localized parameters in studies related to sports biomechanics^{1,6}. The coefficient of variation (CV) of global parameters such as running speed can vary between 2¹⁴ and 760%¹⁵, whereas foot contact time varies between 3.8¹⁶ and 10%¹⁷ (see Supplementary Table 1 for details on each variable). However, localized parameters may present greater variability. The CV from the vertical loading rate ranges from 12¹⁸ to 109%¹⁷, whereas the CV from foot angle at initial contact ranges from 24¹⁹ up to 3288%¹⁷. These statistics suggest that different variables in running biomechanics present specific variability patterns. Therefore, defining the number of steps and/or sample size for a running experiment based on only one variable may affect the results of all other recorded variables.

In a short literature review, we included 56 manuscripts evaluating uninjured runners from the year 2000 to 2019 (Supplementary Table 2). In general, running biomechanics studies reported data with an average of 12 running steps per runner (median: 5; range: 3–200, Fig. 1A), and 51 out of 56 studies (91%) reported results with up to 10 steps. Only 5 out of 56 studies (9%) in our review used more than 10 steps to describe running performance. Interestingly, no clear arguments were found in the literature to justify the need for a large number of steps in running biomechanics. Moreover, studies comprising of multiple variables of different nature (kinetics, kinematics, neurophysiological) defined the number of trials or steps based on one of the variables investigated. Therefore, although establishing adequate sample sizes and number of steps analyzed is considered a priority in biomechanical studies²⁰, the differences across variables might have been overlooked until present days.

Another relevant factor that assures an adequate statistical power of biomechanics research is establishing an adequate sample size⁴. Unfortunately, studies with a small sample might lead to a decreased statistical power, poor reliability, and inflated effect size^{4,10}, whereas studies with large sample sizes may be more expensive and may subject their participants to unnecessary interventions²¹. The running biomechanics studies we reviewed reported results using 31 runners on average (median: 21; range 8–210 runners, Fig. 1B), with 8 out of 56 studies (14.3%) that reported data from only 10 runners. Previous review papers in biomechanics have shown a median sample size of 14.5¹⁰ or ranging from 12 to 18 participants⁴. It seems that researchers are increasing the number of runners included in their studies within the last two decades. However, more research is necessary to further establish normative guidelines regarding data acquisition factors such as sample size and number of steps for research in running biomechanics. Similarly, the running speed is widely variable in running biomechanics studies, from which the mean value is 3.46 m s⁻¹ (median: 3.33 m s⁻¹; range: 2.13–6.7 m s⁻¹, Fig. 1C). Running speed can dramatically influence running kinematics and kinetics²², which requires caution when comparing results from studies with non-comparable running speed. Therefore, understanding the implications that may arise when selecting varying sample sizes, number of events (trials or steps), and locomotion speed is highly relevant in designing experimental biomechanics studies.

Therefore, the purpose of this study was to answer two main questions: (1) how to select an appropriate number of steps and sample size to conduct a study regarding running biomechanics involving multiple variables? (2) What are the implications of the selected number of steps and sample sizes on data variability and statistical outcomes? To answer these questions, a simple intervention consisting of reducing the sound volume produced

from running footfalls was applied in this study. The silent running performance was compared to normal running at a similar running speed. This intervention was implemented to induce expected changes in running kinematics/kinetics. For instance, it is known that silent running significantly reduces loading rate and peak vertical forces^{23,24}, however, no changes are found for running active peak and foot contact time²⁴. A sample of 44 runners performing 40 steps in each condition was used to investigate the implications of using several combinations of reduced sample sizes and steps on data variability [sequential estimation technique (SET) and CV], as well as statistical outcomes (*p* values and effect size from the comparison between normal vs silent running).

Methods

Participants. Forty-four healthy recreational runners (8 females and 36 males, age: 26.81 ± 4.04 years, weight: 77.29 ± 11.76 kg, height 178.41 ± 9.34 cm, BMI 24.20 ± 2.74) agreed to participate in our study. Runners reported a moderate to a high level of physical activity according to the International Physical Activity Questionnaire (3269.36 ± 1542.92 MET) or with a minimum level of running experience of 3 years and weekly mileage of 13.9 ± 9.1 km. None of the participants had any history of lower extremity musculoskeletal injury or running-related injuries during the testing or in the previous six months. All the participants provided verbal and written informed consent before participating. All experimental methods were carried out in accordance with relevant guidelines and regulations, and the experimental procedures of the present study were in accordance and approved by the Ethical Committee of North Jutland (Region Nordjylland).

Experimental procedure. The experiment consisted of a single session where participants performed two different running tasks: normal running, as well as running while minimizing the sound volume of their footfalls (silent running). The running tasks were performed overground, allowing for the recording of ground reaction forces and motion capture. Familiarization procedures were conducted prior to data recordings, as well as during the session for the silent running condition. The order for performing normal and silent running was randomized across all participants.

Running tasks. Participants performed a 5-min warm-up consisting of walking lunges, skipping, leg swings, and run-throughs (Phan et al. 2017). Following the warm-up, participants spent 5–10 min familiarizing themselves with the running track and established their preferred running speed. A 20-m indoor overground obround running track (2-m running turns and 8-m sprinting lines) was used. A force platform and four microphones surrounding it were located in the middle of one sprint line. The microphones did not interfere with the runner's sprint-line trajectory. All runners were provided with a standard running shoe (Nike Air Pegasus) to perform the familiarization and testing procedures.

For the normal running, participants were instructed to run as naturally as possible at a self-selected speed, aiming to hit the force plate with the right foot. For the silent running, participants were instructed to actively minimize the volume of sound produced during their running footfalls throughout the track, especially at the sprint line containing the force platform and microphones. Runners were not instructed regarding how they should change their running technique to achieve silent running. Participants performed both normal and silent running conditions in three sets of 3-min for each condition. Participants were instructed to run continuously around the track at their preferred running speed while stepping with the right foot on the force plate at every lap. As participants ran at their preferred speed, the number of laps in the track varied across participants (between 14 and 19 laps per set). In order to attenuate the effects of fatigue in the experiment, a 3-min rest period was offered to runners between each recording. The running speed was monitored on a trial-by-trial basis by a researcher recording the time (and subsequently the speed) runners took to pass through a 3-m sector of the track that contained the force plate in the middle. This speed monitoring allowed instructing runners regarding changes in their running speed throughout a set, as well as between sets and between conditions. In case the running speed changed 10% above or below the pre-established value, runners were instructed to increase or reduce their pace in the next lap.

Data collection. Data were recorded on an indoor motion capture laboratory, on a 20-m obround running track. The track had an 8-m sprinting line with a mid-data collection landmark consisting of a floor embedded force plate (AMTI Optima Gen 5, Watertown, MA, USA), sampled at 1000 Hz that provided three-dimensional forces and moments. Moreover, an 8-camera motion capture system (Qualisys 1.0, Qualisys Oqus, Sweden) sampled at 200 Hz was used to record position data from 8 retro-reflective markers (16 mm diameter). Four of these markers were fitted equidistantly onto an elastic strap that was positioned on the participant's head, and the remaining four markers were positioned on the right shoe at the foot tip, heel, medial, and lateral metatarsal heads of the running shoe following previous literature²⁵. Both running kinetic and kinematic data were synchronously captured through the motion capture software.

Data analysis—variables extraction. Customized scripts on Matlab (2019b, The Mathworks, Natick, USA) were used to analyze kinetics and kinematics data. Individual running steps were defined using the raw vertical ground reaction force data, defining foot contact to the force plate when the vertical force exceeded 25 N. Subsequently, the force data were filtered using a fourth-order Butterworth low pass filter (60 Hz cut-off frequency) and normalized to the participant's body weight (%BW). Foot contact time (FCT), peak vertical force (PKF), anterior–posterior peak braking force (BRK), and vertical loading rate, which was defined as the slope of the line between 20 and 80% of the impact peak^{26–28}, were extracted from the force data. If the impact peak was not found, a pre-set point in time, 13% of the total contact time was used instead^{26,27,29–31}.

The participants running speed was calculated within a range of one meter before and after stepping onto the force plate, by computing the average displacement across all head markers for each successful running cycle. The foot contact angle (FCA) was defined as the angle between the vector from the forefoot marker to the heel marker with respect to a vector representing the anterior–posterior direction in the laboratory coordinate system. The FCA was offset by the angle between these two vectors at 30% of the stance phase when the foot made full contact with the force plate. Positive FCA values were achieved when the heel marker was higher than the toe marker and negative values when the toe marker was higher.

Data analysis—sequential estimation technique (SET). SET determines the point of mean stability of a variable recorded continuously and was implemented in the present study following previous literature¹¹. Then, the sequential data points from all 40 steps were averaged to create the target stability, from which a bandwidth of $\pm 25\%$ was used to determine whether the individual steps fell within this bandwidth. Subsequently, a moving point mean was calculated, starting from the first two points and increasing after each interaction by one point until the total amount of steps was reached. The point of stability is the first point that is within the bandwidth along with all subsequent points recorded. SET was applied in all variables to define their point of stability. Therefore, data average stability was defined as the average stability of the individual participant for each parameter investigated where the number of steps analyzed should not influence the statistical outcome.

Data permutation (number of steps vs number of runners). Although each participant performed a different number of valid steps on the force plate, they all successfully performed at least 40 valid steps. The data from these 40 steps across all 44 participants was defined as the full running dataset, which would then be compared to sub-samples of steps (sequences of 5, 10, 15, 20, 25, 30, 35, and 40 steps), combined with sub-samples of participants (random samples of 10, 15, 20, 25, 30, 35 and 40 runners). As an example, the loading rate was computed for 10 runners performing 5 steps. There are $\sim 2.5 \cdot 10^{10}$ possible combinations of 10 runners from a total of 44 runners. Moreover, there are 658,008 possible combinations of 5 steps from a total of 40 steps for each runner. Therefore, for our example there are $\sim 1.6 \cdot 10^{16}$ possible combinations to represent 10 runners performing 5 steps each. From this total, 100,000 combinations were randomly selected from both normal and silent running (with the same runners in both conditions). Moreover, the number of steps was averaged for each runner and subsequent statistical comparison between normal and silent running was conducted for each combination. The colormaps in the results section show the average across the 100,000 statistical comparisons. Forrester et al. (2015)² used 10,000 different samples to run simulations and reach relevant conclusions. Preliminary calculations demonstrated that the use of 100,000 permutations was enough to represent similar variability as ≥ 10 million number of permutations. Therefore, data from the six dependent variables (foot contact time, loading rate, peak vertical force, peak braking force, running speed, and foot contact angle) were permuted for both normal and silent running.

Statistical analysis. Differences between normal and silent running using the standard dataset (44 runners, 40 steps per runner) for all dependent variables were assessed using two-tailed paired t-Student tests. Cohen's D effect size was computed for all comparisons. Similarly, the 100,000 paired samples were compared using similar t-Student tests, and the respective Cohen's D effect size ("small" values around 0.2, "medium" for 0.5, and "large" above 0.8³⁶) were computed for each pair. In addition, interquartile ranges (IQR, the difference between the 25th and 75th percentile from a given sample) and coefficient of variation (CV) were computed from all 100,000 samples of both conditions. Furthermore, a sample size estimation was conducted following the formula:

$$n = \left[\frac{z^2 * p * (1 - p)}{e^2} \right] / \left[1 + \frac{z^2 * p * (1 - p)}{e^2 * N} \right] \quad (1)$$

where $z = 1.96$ for a confidence level of 95%, $p = 0.7$ proportion of the population, $N = 44$ population size and $e = 0.05$ the margin of error³². A recommended sample size of 39 participants was found.

Results

Full running dataset analysis (44 runners, 40 steps/runner). There were significant reductions in loading rate ($p < 0.0001$, effect size: 1.41, Fig. 2), peak braking force ($p < 0.0001$, effect size: 0.66), running speed ($p < 0.05$, effect size: 0.167) and foot contact angle ($p < 0.0001$, effect size: 1.03) during silent running when compared to normal running. No changes were found for foot contact time and peak vertical force ($p > 0.05$).

The interquartile range across different number of steps. An illustration of such distribution using the loading rate (Fig. 3A) and running speed (Fig. 3B) depicts no substantial changes in data distribution when comparing data from 5 to 40 steps using all 44 runners. This pattern was maintained across all variables. Interestingly, the IQR may show differences both across the number of steps as well as between conditions. In particular, running speed presented a reduction in IQR that resulted in significant differences when using 20 steps or more. All other variables presented consistent IQR across the number of steps.

Sequential estimation technique. SET was used to determine the minimum number of trials required to reach a stable data average within 40 steps. It was evident that different variables may present different distribution over time (Fig. 4A), therefore influencing the minimum number of steps required to reach stability. Furthermore, Fig. 4A—right panels, illustrate the within parameter variability of one illustrative runner, where

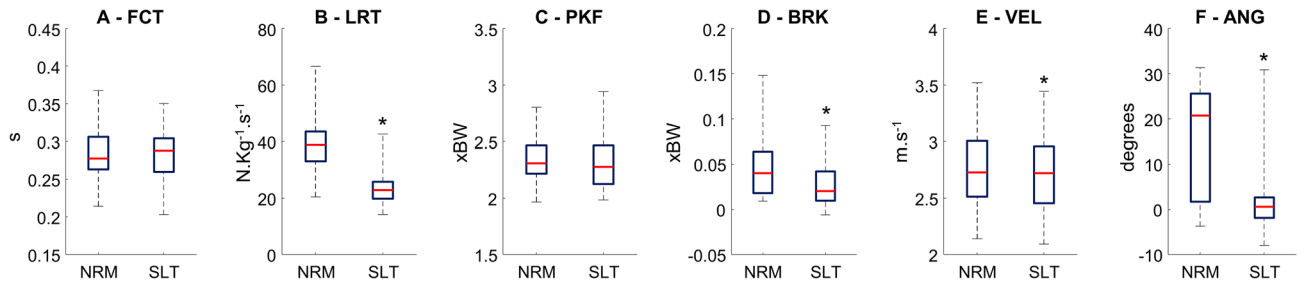


Figure 2. Comparison normal (NRM) vs silent running (SLT) for foot contact time (FCT, **A**), vertical loading rate (LRT, **B**), peak vertical force (PKF, **C**), peak braking force (BRK, **D**), running speed (VEL, **E**) and foot contact angle (ANG, **F**). Box plots represent 25th and 25th percentiles, and data range (*dashed vertical lines*) during normal (NRM) and silent running (SLT). * denotes a significant difference in relation to normal running ($p < 0.05$).

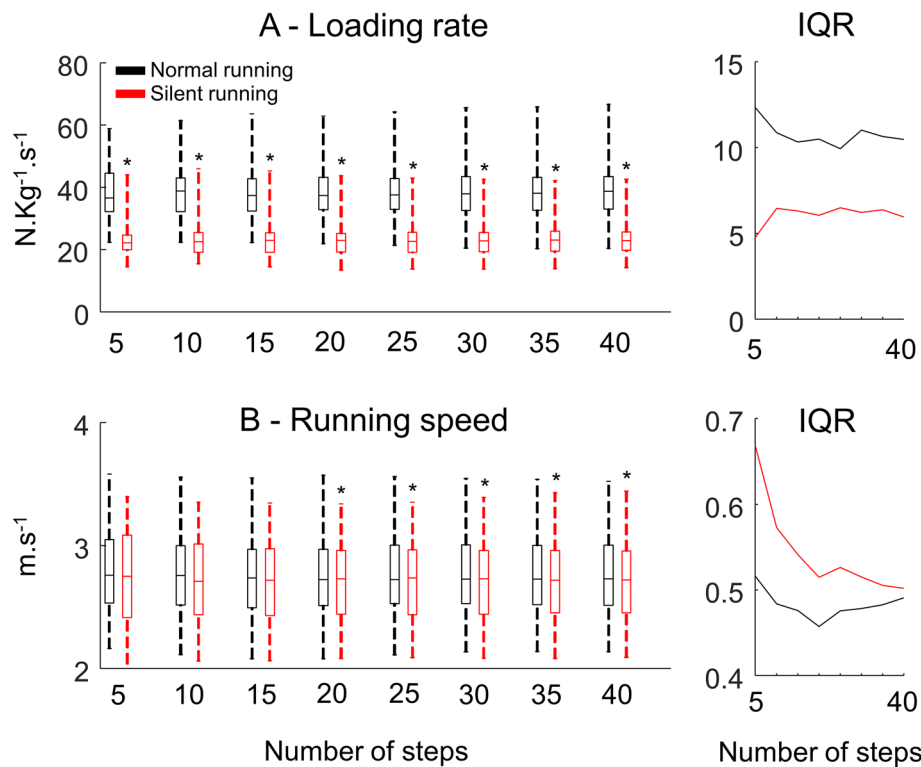


Figure 3. Comparison normal (*black*) versus silent running (*red*) for vertical loading rate (**A**), and running speed (**B**). Box plots represent 25th and 25th percentiles and data range (*dashed vertical lines*) across different numbers of steps. On the right, interquartile ranges (IQR) across the different number of steps. * denotes a significant difference in relation to normal running ($p < 0.05$).

sequential estimation shows how increasing the total number of steps analyzed decreased the within step variability and increases parameter stability. An evaluation across all 44 runners demonstrated that the average number of steps to reach stability ranges from 15 ± 8 (peak braking force) to 20 ± 7 (foot contact time). Moreover, the inter-subject variability is high across all variables and conditions (Fig. 4B), in which the 75th percentile is above 20 trials across all variables. The percentage of runners requiring 10 or fewer steps to reach stability is below 30% of the overall sample of runners (Fig. 4C), and this number is variable dependent on the variable. Most of the running variables require 15 to 20 steps to reach stability. It was determined that at least 20 steps are needed to reach more than 50% of the sample of runners with a stable average.

Data distribution—10 vs 40 runners. Figure 5 demonstrates that using 10 runners can drastically influence the representation of the data for both conditions, compared to 40 runners. Box plots, median values, and range across random sub-samples vary substantially when only using 10 runners, and this effect is substantially reduced using 40 runners.

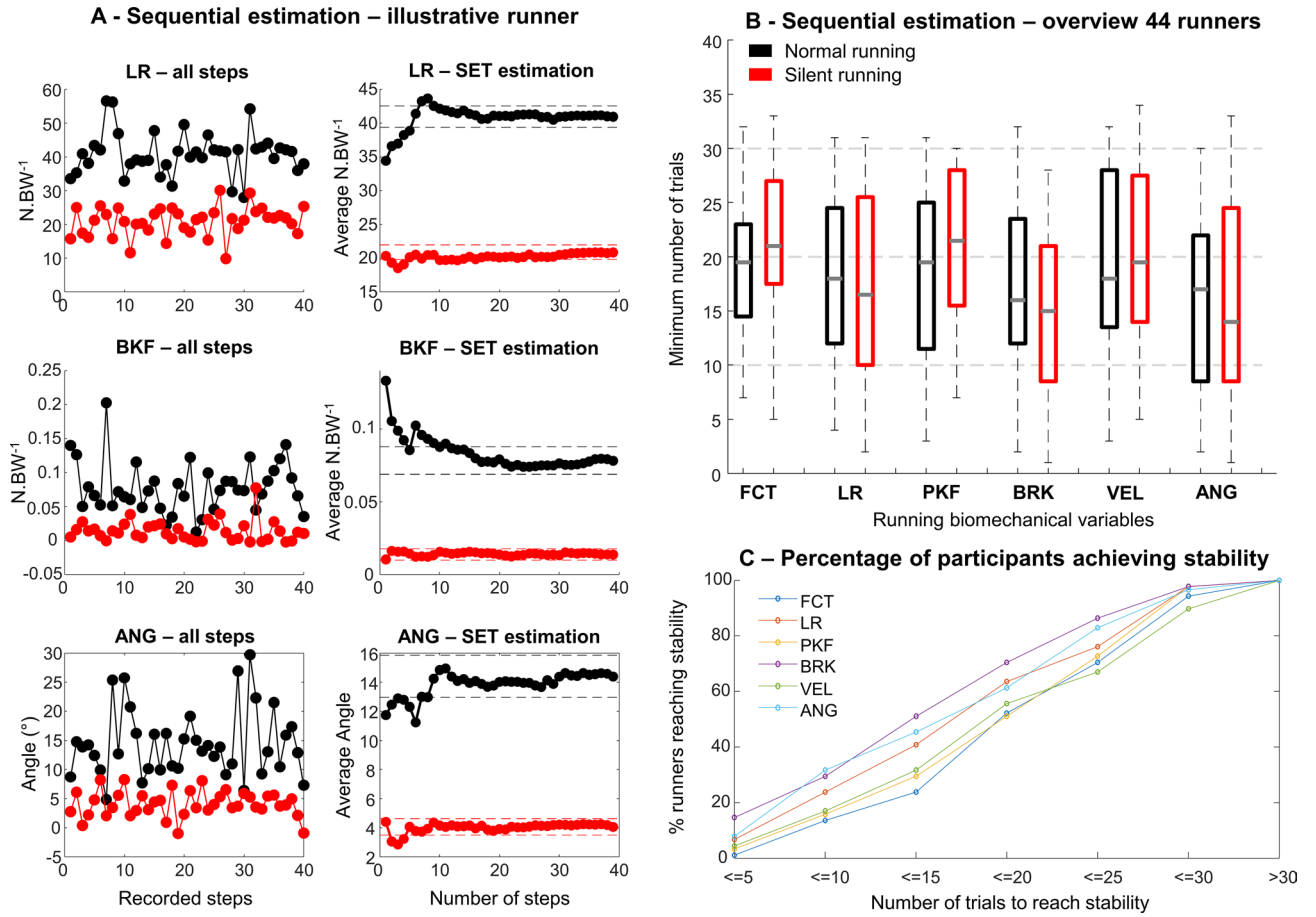


Figure 4. Illustration of sequential estimation to determine the number of steps needed to achieve stability from 10 (left) and 40 runners (right) during both normal (black) and silent running (red). **(A)** Within parameter variability of a representative runner (left) and sequential estimations (right) for loading rate (LR), peak braking force (BKF), and foot contact angle (ANG). **(B)** Box plots represent 25th and 75th percentiles, and data range (dashed vertical lines) for multiple parameters (foot contact time—FCT, loading rate—LR, peak vertical force—PKF, peak braking force—BRK, running speed—VEL, foot contact angle—ANG). **(C)** The linear plots illustrate parameter variability and the percentage of participants as well as the number of steps required to achieve stability across both normal and silent running.

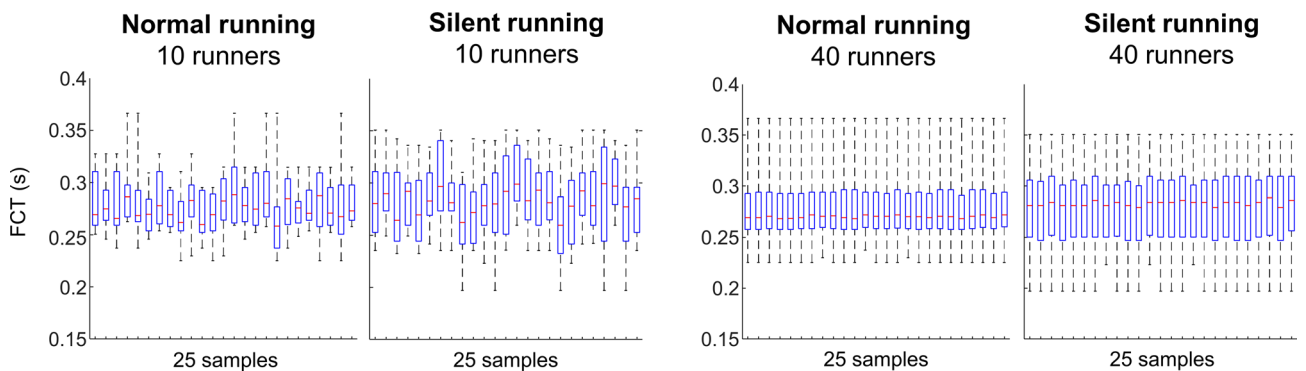


Figure 5. Illustration of 25 subsamples containing foot contact time (FCT) from 10 (left) and 40 runners (right) during both normal and silent running. Box plots represent 25th and 75th percentiles, and data range (dashed vertical lines). The 25 subsamples were randomly selected from 100,000 paired samples generated for further analysis.

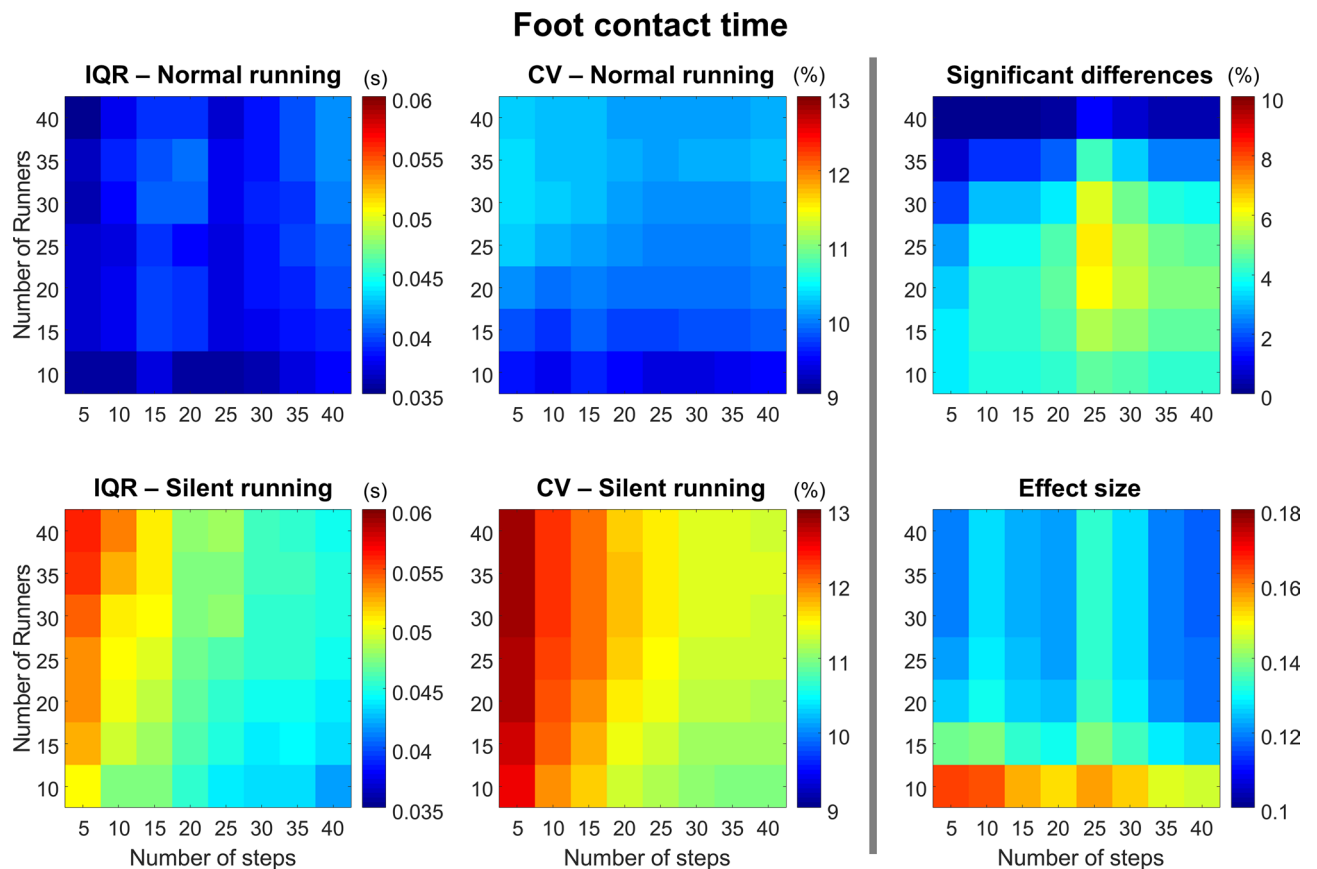


Figure 6. Colormaps demonstrating foot contact time interquartile range (IQR, *left column*) and coefficient of variation (CV, *central column*) from the combination of 10–40 runners versus 5–40 running steps. Data is shown from normal running (*top panels*) and silent running (*bottom panels*). Each value is an average from 100,000 paired permutations for each combination. The number of significant differences ($p < 0.05$, expressed as a percentage of 100,000 comparisons, *top right panel*) and their respective effect size (*bottom right panel*) illustrate the changes in statistical outcomes depending on the combination of number or runners versus number of steps.

Effect of number of runners and steps on normal vs silent comparison. *Foot contact time.* The foot contact time presented greater IQR (Fig. 6, *left column*) and CV (Fig. 6, *central column*) for the silent condition when compared to normal. Overall, the CV is low ($< 15\%$); however, for the silent condition, the CV is reduced when a greater number of steps are used. The number of runners does not influence the CV. However, the lower the number of runners, the higher the probability of finding significant differences (Fig. 6, *top right panel*). The effect size (Fig. 6, *bottom right panel*) is generally low across all combinations, but it tends to be greater when using a lower number of participants (below or equal to 15).

Peak vertical force. The peak vertical force resulted in greater IQR and CV for the silent condition when compared to normal (Fig. 7, *left and central columns*). Overall, the CV is low ($< 15\%$); however, for the silent condition the CV is reduced with fewer runners (< 20) and with a greater number of steps. There was a greater probability to find significant differences when using 15–30 runners with more than 30 steps. However, the effect size was usually small (< 0.15).

Peak braking force. The peak braking force presented a greater IQR for the normal condition, which generally increased with a lower number of steps (Fig. 8, *left column*). Overall, the CV is high ($> 60\%$), increasing with a lower number of steps (< 10 steps) when compared to the other step categories. No influence of the number of participants is seen. There was a greater probability to find significant differences when using > 20 runners, with no relevant effect of the number of steps on statistical outcomes. The effect size was moderate across all combinations (from 0.6 to 0.7).

Vertical loading rate. The loading rate presented greater IQR for the normal running, with no influence of steps or sample size (Fig. 9, *left column*). The CV was influenced differently depending on the condition. The normal condition presented a greater CV when analyzing > 15 steps regardless of the number of runners used, whereas during silent running the CV increased when using less than 25 steps and more than 20 participants steps (Fig. 9, *central column*). Significant differences between conditions were found for any combination of steps and runners

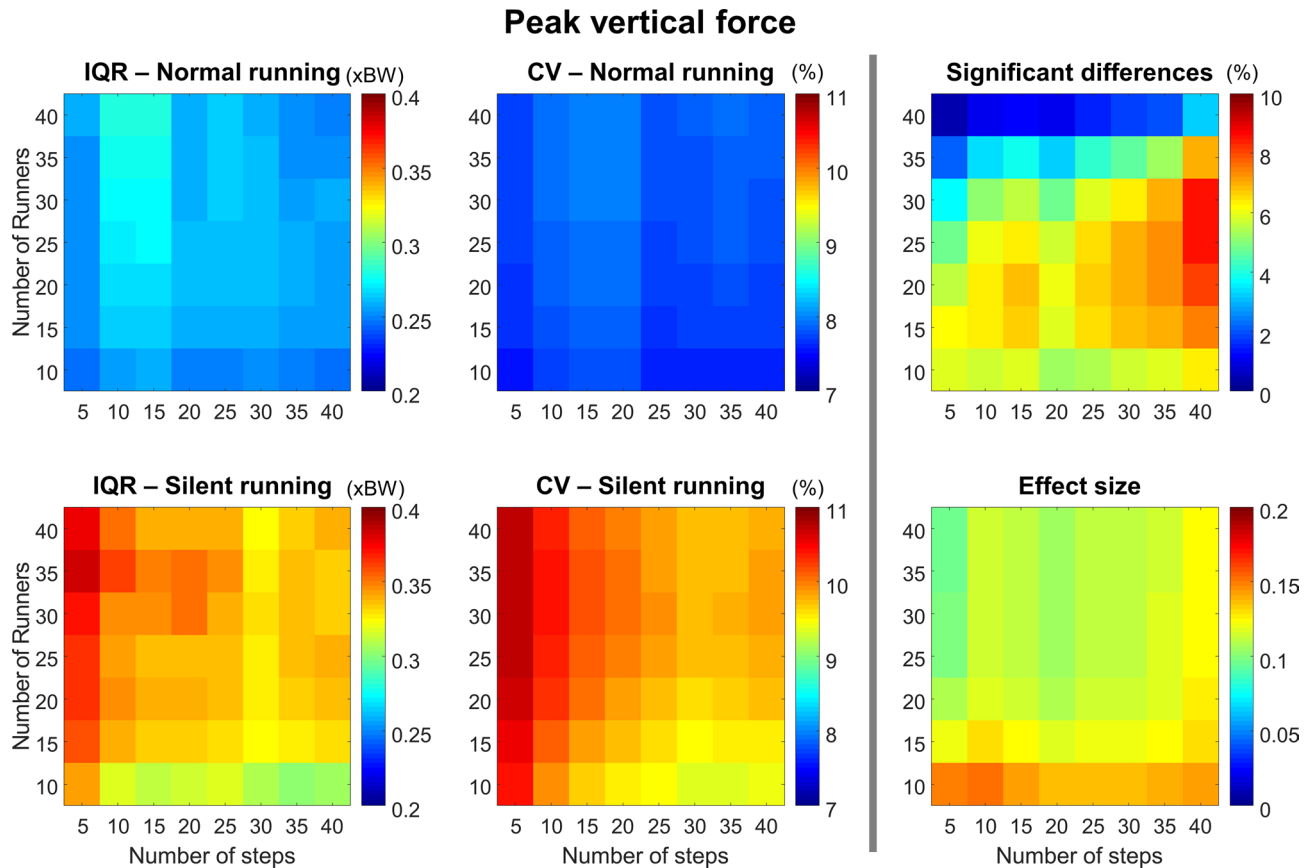


Figure 7. Colormaps demonstrating peak vertical force interquartile range (IQR, *left column*) and coefficient of variation (CV, *central column*) from the combination of 10–40 runners versus 5–40 running steps. Data is shown from normal running (*top panels*) and silent running (*bottom panels*). Each value is an average from 100,000 paired permutations for each combination. The number of significant differences ($p < 0.05$, expressed as a percentage of 100,000 comparisons, *top right panel*) and their respective effect size (*bottom right panel*) illustrate the changes in statistical outcomes depending on the combination of number or runners versus number of steps.

with an increased effect size (> 1.4). However, the effect size was lower when using > 30 steps and > 30 runners (Fig. 9, *right column*).

Running speed. The running speed presented a greater IQR for the silent running, resulting in greater IQR with fewer number of steps (Fig. 10, *left column*). The CV was generally low for both running conditions (Fig. 10, *left column*), with a tendency to increase when using a lower number of steps during silent running. The probability of finding significance was greater when analyzing > 30 runners with > 25 steps. However, using less than 15 steps dramatically reduced the probability of finding differences, regardless of the number of steps the runner used (Fig. 10, *right column*). Despite these differences, the effect size was, in general, small (< 0.2), increasing with the number of steps.

Foot contact angle. The foot contact angle presented greater IQR in the normal condition when compared to silent (Fig. 11, *left column*). Moreover, the IQR was greater when analyzing > 20 participants with 15 steps or less. The CV in both normal and silent conditions is greater when analyzing only 5 steps, with a low influence of the number of participants. The substantially greater CV during silent running is also worth noting. The probability of finding significant differences is assured by using more than 10 runners, regardless of the number of steps. The effect size is generally high (> 1.0) but tends to be lower when using more than 20 runners and 25 steps.

Specific recommendations regarding sample size and number of steps. In Fig. 12, we summarized specific recommendations to the number of runners and number of steps to be recorded with respect to different running biomechanical variables. These recommendations were based on the results provided by the SET analysis and the permutation analysis across each variable.

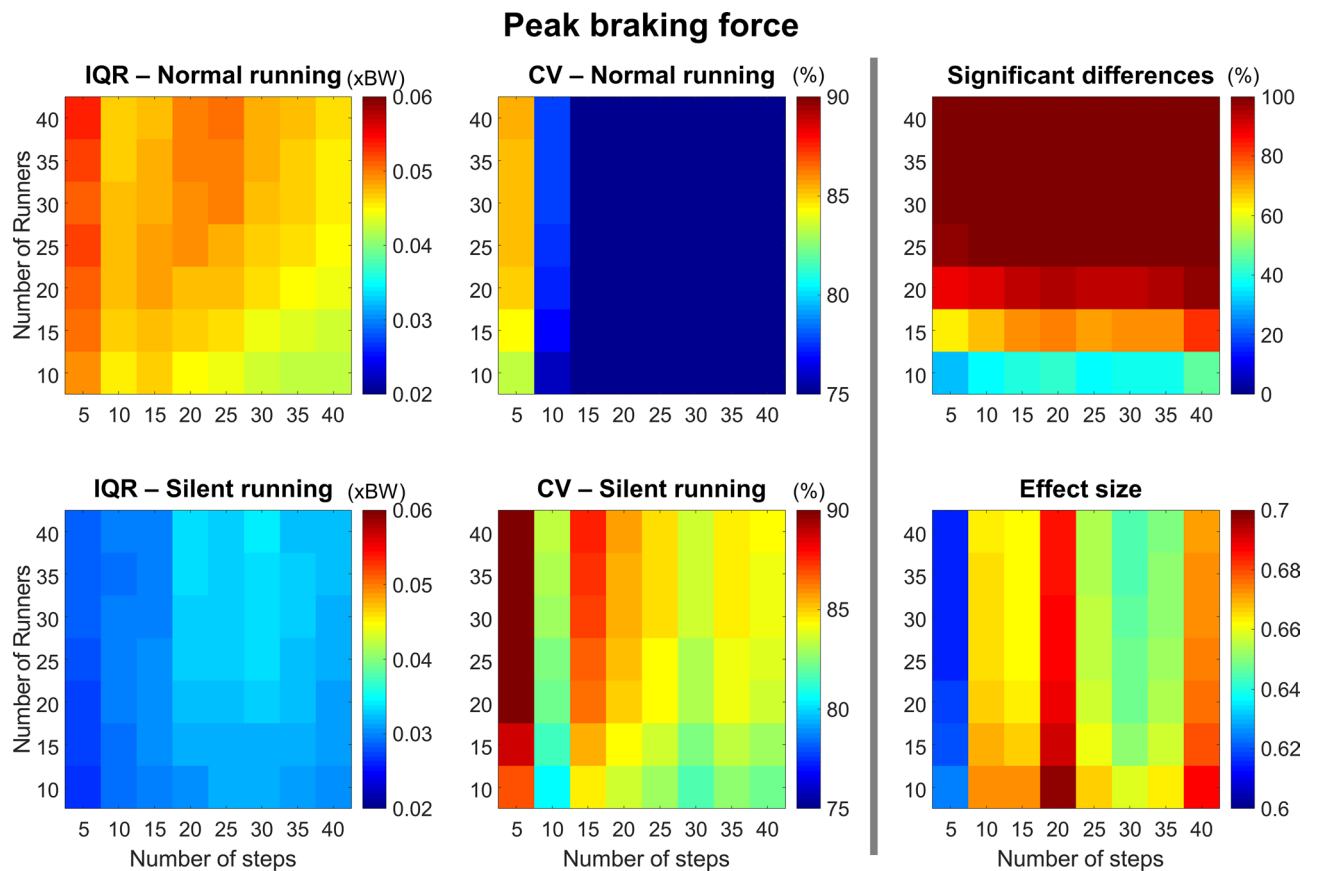


Figure 8. Colormaps demonstrating peak braking force interquartile range (IQR, *left column*) and coefficient of variation (CV, *central column*) from the combination of 10–40 runners versus 5–40 running steps. Data is shown from normal running (*top panels*) and silent running (*bottom panels*). Each value is an average from 100,000 paired permutations for each combination. The number of significant differences ($p < 0.05$, expressed as a percentage of 100,000 comparisons, *top right panel*) and their respective effect size (*bottom right panel*) illustrate the changes in statistical outcomes depending on the combination of number or runners versus number of steps.

Discussion

The purpose of this study was to provide information on (1) how to select sufficient steps and sample sizes and (2) the implication of the selected sample sizes and steps on the variability and statistical outcomes of various running biomechanical variables. A total of 44 runners performing 40 steps in two running conditions (normal and silent) were used to explore the implications of using several combinations of sample sizes and steps on the variability of running biomechanical parameters. The main findings of our study were that: (1) stable averages across the investigated biomechanical variables require more than 10 steps for the majority of runners, as less than 35% of runners reached a stable average within 10 steps or less. (2) Variables expected not to be influenced by the silent running intervention, such as foot contact time and peak vertical force, presented increased likelihood of reporting significant differences when the sample size was ≤ 25 runners. (3) Running kinetic variables (vertical loading rate and peak braking force), which were expected to demonstrate significant differences, presented greater IQR for normal running. Nonetheless, vertical loading rate consistently presented significance and large effect sizes, whereas peak braking force reached an expressive probability of significance only when using more than 20 runners. Finally, (4) foot contact angle demonstrated substantial variability in the silent condition, especially when using less than 10 steps. Moreover, the foot contact angle presented an optimal probability of significance when more than 10 runners were analyzed. These results highlight the importance of defining the adequate number of steps and sample size with respect to a particular running biomechanics study, due to the fact that different variables may require different amounts of steps and/or runners. Therefore, it is recommended that studies involving analysis of traditional running biomechanical variables use a minimum of 25 participants to increase the likelihood of appropriate statistical power and large effect sizes. Moreover, acquiring data from at least 25 steps from each participant may increase the likelihood of achieving data average stability across different variables for the majority of the study sample. Our results may serve as guidelines to optimize experimental protocols in running biomechanics, as well as standardize minimum requirements for appropriate comparison between results achieved from different research groups worldwide.

An important outcome from our short literature review (see Supplementary Table 2) was that 91% of biomechanical studies (51 out of 56) draw scientific conclusions based on 10 or fewer steps. According to our findings, the stability of the mean is achieved for only 10–30% of runners when using up to 10 steps across the

Vertical loading rate

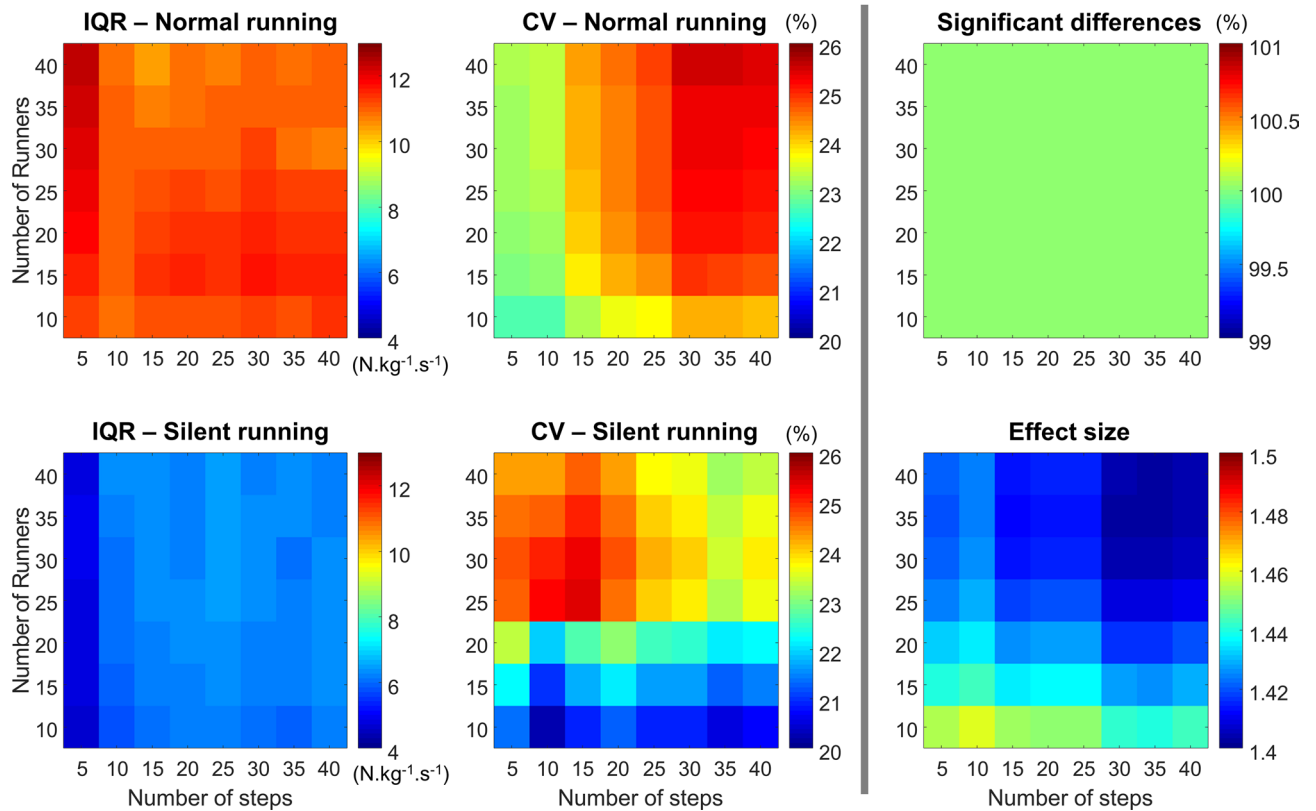


Figure 9. Colormaps demonstrating vertical loading rate interquartile range (IQR, *left column*) and coefficient of variation (CV, *central column*) from the combination of 10–40 runners versus 5–40 running steps. Data is shown from normal running (*top panels*) and silent running (*bottom panels*). Each value is an average from 100,000 paired permutations for each combination. The number of significant differences ($p < 0.05$, expressed as a percentage of 100,000 comparisons, *top right panel*) and their respective effect size (*bottom right panel*) illustrate the changes in statistical outcomes depending on the combination of number of runners vs number of steps.

investigated biomechanical variables (see Fig. 4C). Moreover, the number of steps required to reach stability is dependent on the type of variable targeted. Some previous recommendations regarding the number of steps are based on simulations and not experimental data^{2,3}. Our results suggest that a lower number of steps (< 10) and low sample size (< 20) may prevent finding significant differences for most of the variables. These results corroborate studies from Bates et al. (1992)³ and Forrester et al. (2015)² that recommended evaluating 20 participants³ and approximately 20 steps². However, Bates et al. (1992)³ compiled the recommendations using 1000 possible simulations based only on vertical ground reaction force data, disregarding differences across biomechanical variables. Furthermore, Forrester et al. (2015)² generated recommendations solely based on breast displacement during running, which is a highly specific variable that might not fully represent the variability from classic running biomechanical variables. Salo et al. (1997)³³ found that kinematic variables (e.g. hip, knee, and ankle angle of the lead foot) from sprint hurdle can achieve 90% statistical power using on average 6 steps for female sprinters (median: 3, range 1–61 steps) and 16 steps for male sprinters (median: 7, range 1–78). On the other hand, the same study³³ demonstrated that the time of maximum angular velocity requires 61 steps to achieve 90% statistical power for females, and horizontal velocity lost requires 78 steps to achieve 90% statistical power for males. Thus, it has been a general consensus that movement variability is variable-dependent^{2,33}, and the present study provides estimations of number of steps and sample size to reach data average stability and relevant effect sizes in different biomechanical variables. Moreover, our results further support the statement that variability is variable-dependent, but also that the outcomes of objective comparison (such as normal vs silent running) may be influenced by both sample size and number of steps. Therefore, our results represent an important advance towards standardizing experimental procedures to investigate different aspects of running biomechanics.

A positive aspect of movement variability is the possibility to achieve success on a motor action through several solutions¹. However, either high or low variability may reveal the presence of a pathology, pain, overuse injuries, as well as demonstrate learning effects or environmental adaptations^{1,6}. Early³ and more recent² methodological research on this topic has shown that a reduced number of steps can increase the mean difference between conditions due to a greater variability while reducing the effect size and power². Our results corroborate such a statement. For normal running, the IQR of running speed remains constant as a function of the number

Running speed

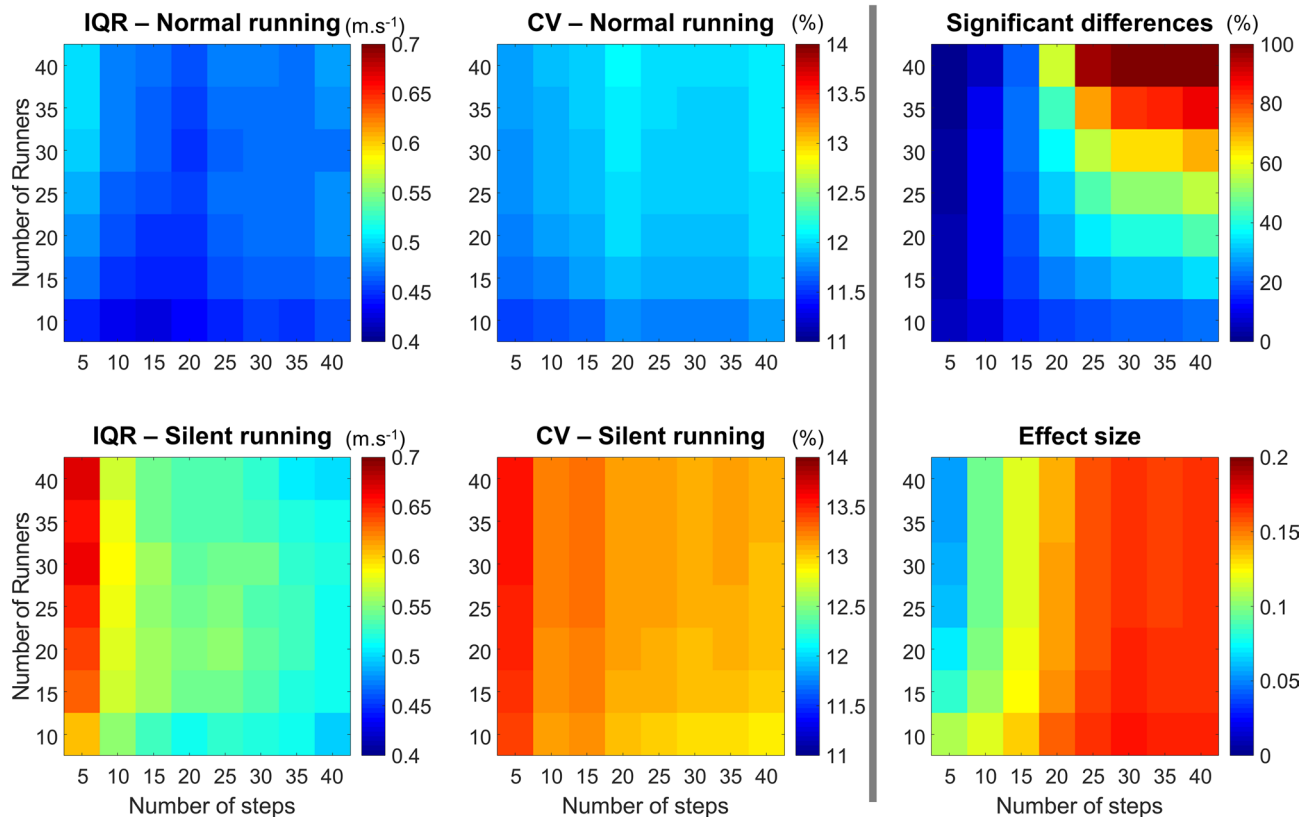


Figure 10. Colormaps demonstrating running speed interquartile range (IQR, left column) and coefficient of variation (CV, central column) from the combination of 10–40 runners versus 5–40 running steps. Data is shown from normal running (top panels) and silent running (bottom panels). Each value is an average from 100,000 paired permutations for each combination. The number of significant differences ($p < 0.05$, expressed as a percentage of 100,000 comparisons, top right panel) and their respective effect size (bottom right panel) illustrate the changes in statistical outcomes depending on the combination of number or runners versus number of steps.

of steps while the IQR was inversely related to the number of steps during silent running ($n = 44$, Fig. 3B, right panel). Statistical differences were only achieved when analyzing at least 20 steps. However, 96.5% (54/56) of the studies from our review used less than 20 steps to describe running speed. Although the difference seems marginal between conditions, by providing enough data to minimize variability, it is possible to reach significance. Therefore, studies comparing conditions may have different outcomes depending on the number of steps used for such statistical analysis.

The silent running condition proposed in this study aimed to change a runner's approach to performing the stance phase, eliciting major changes in movement kinematics to minimize the sound output from each foot strike. The IQR demonstrated distinct differences between normal and silent running depending on the variable investigated in this study. Moreover, the manipulation of sample size and number of steps can influence IQR considerably. Variables such as foot contact time and peak vertical force presented greater IQR for silent running, mostly due to the different motor strategies participants used to perform the task. Interestingly, significant differences in running speed between normal and silent running were found when using more than 20 steps and 20 runners. Runners were expected to maintain their running speed across both conditions. In fact, the difference when using all 44 runners was on average 0.04 m.s^{-1} ($\sim 3\%$ difference across all different numbers of steps). Moreover, the effect size across all runners versus steps combinations was low (maximum 0.2), suggesting that the differences might not be relevant despite being statistically significant. Our data revealed that the IQR for silent running speed gets lower as we increase the number of steps when using all 44 runners (see Fig. 3B), which may be the driver of statistical differences for such a large sample size. When evaluating the combinations of sample size versus number of steps, there is a shift from having no significant differences when using less than 20 steps to at least a 70% chance of finding significant results when using more than 20 runners and more than 20 steps. In addition, the number of runners highly influences the outcomes when using greater amounts of steps, and the number of steps is the major determinant of changes in effect size for the running speed.

A previous study comparing running patterns between treadmill and overground running using different shoe drops reported significant changes between running conditions, which were based on a different number of steps (20 treadmill steps and 7 overground steps)³⁴. These authors reported a 51% increase in loading rate and greater foot contact angle during overground running when compared to treadmill running³⁴. These authors stated

Foot contact angle

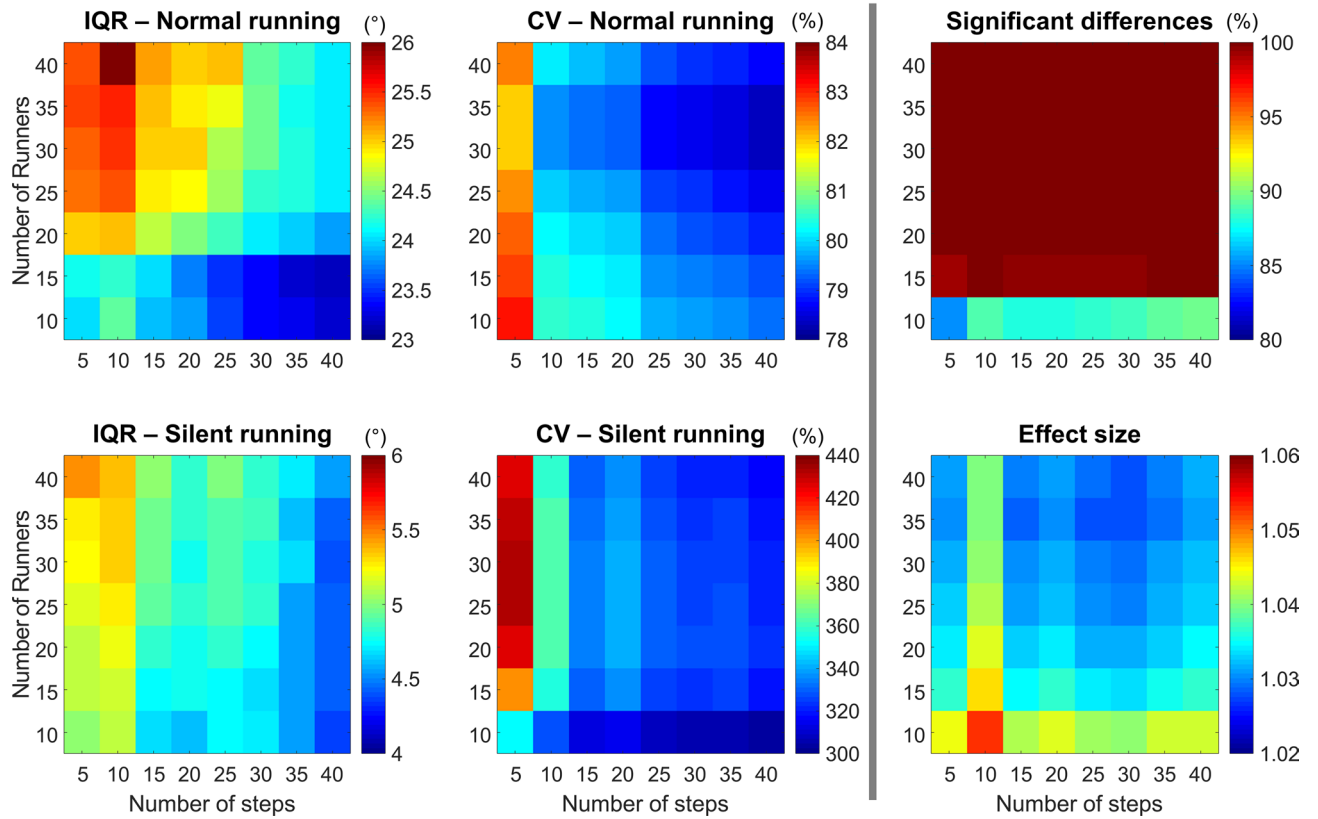


Figure 11. Colormaps demonstrating foot contact angle interquartile range (IQR, left column) and coefficient of variation (CV, central column) from the combination of 10–40 runners versus 5–40 running steps. Data is shown from normal running (top panels) and silent running (bottom panels). Each value is an average from 100,000 paired permutations for each combination. The number of significant differences ($p < 0.05$, expressed as a percentage of 100,000 comparisons, top right panel) and their respective effect size (bottom right panel) illustrate the changes in statistical outcomes depending on the combination of number of runners versus number of steps.

Recommendation for selection of sample size and number of steps

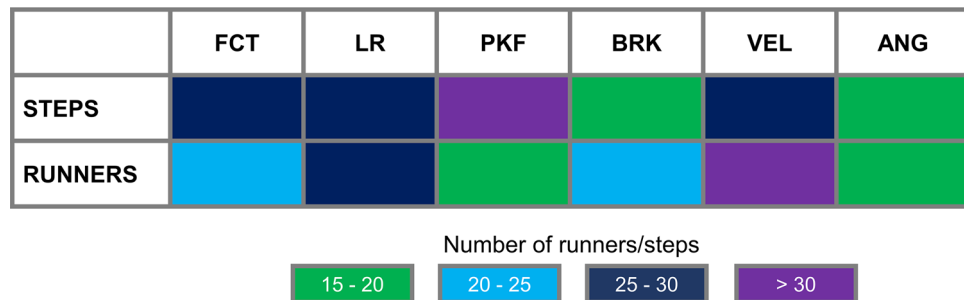


Figure 12. Colormap representing the recommended number of steps and sample size required to achieve individual stable average, as well as reduce variability and maximize effect size. *FCT* foot contact time, *LR* loading rate, *PKF* peak vertical force, *BRK* peak braking force, *VEL* running speed, *ANG* foot contact angle.

that treadmill running is easier to accomplish whereas overground running presents greater intra-individual variability due to variable running speed ($\pm 5\%$) and foot placement (within the force plate limits). Interestingly, other studies found no differences between treadmill and overground running variables analyzing a similar number of steps. Oliveira et al. (2016)³⁵ have reported that the peak vertical tibial acceleration is similar between these conditions when using 20 running steps. However, these authors found greater variability in the vertical peak tibial acceleration during overground running. Moreover, Fellin et al. (2010)³⁶ have reported no significant

differences in lower limb kinematics (hip, knee, and ankle angle at foot strike and midstance) between the two conditions, reported from the same number of steps (5 running steps) in both conditions. Therefore, our results contribute to the current body of literature by establishing benchmarks for data variability and expected outcomes regarding sample size and number of steps for widely used running biomechanical variables.

Studies using a low number of steps to represent a runner's performance may present the limitation of not achieving the required average stability. In this study, we found that a median of at least 15 steps is required to reach stability across all variables, for both normal and silent running (Fig. 4B). Moreover, the use of up to 10 steps dramatically reduced the number of runners reaching stable mean (< 35%), regardless of the biomechanical variable. This result is highly relevant for future studies, as the number of steps can influence the data representation of most of the runners, generating misleading results and interpretations. Our results demonstrated that a stable average for the majority of runners (> 50%) across all variables is achieved by using > 20 running steps, whereas each variable may require a different number of steps to reach stability (see Fig. 4C). A previous study comparing ground reaction force variables across different running shoes recorded 15–35 steps per runner³⁷. However, only two steps per shoe were analyzed. The study reported between-gender differences, as well as increased injury risks depending on footwear³⁷ for a sample size of (10 males and 10 females). Around 6 studies (11%) from our review have used less than 5 steps to describe running mechanics^{19,28,38–41}. Based on our results, the use of fewer than 4 steps to report data from ground reaction force parameters might not provide ideal and/or stable intra-subject representation to describe the running biomechanical parameter. Therefore, the interpretation of results generated from low sample sizes and number of steps must be carried out with caution.

The reductions in variability for foot contact time, peak vertical force, and running speed (CV < 15%) in silent running may be related to substantial movement constraints in order to achieve reduced sound outputs. However, the vertical loading rate presented conflicting trends and increased variability during normal running. There was a greater coefficient of variation during normal running when evaluating more than 15 runners and steps, whereas for silent running there was a greater variability when evaluating a lower number of steps (< 25 steps) and runners (< 20 runners). These results are crucial to demonstrate that different running biomechanics variables may present unique trends within the same experimental design. Therefore, studies using several variables must account for the variability pattern of all of them when defining experimental procedures. It is noteworthy that variables expected to show differences between conditions (e.g., peak braking force and foot contact angle) presented substantial variability, especially during silent running (see Figs. 8 and 11, *central columns*). Again, the motor control demands to successfully achieve silent running may explain the larger variability for these variables⁶. Furthermore, runners were not instructed on how to achieve and/or perform silent running. It is suggested that each runner may have adopted a unique running style that was effective to reduce sound outputs, contributing to increases in inter-subject variability for some biomechanical variables. Therefore, studies investigating changes in running technique and/or motor control strategies should consider recording a sufficient number of steps in order to generate stable averages for appropriate representation of all described biomechanical variables. Furthermore, pilot studies involving several participants should be used to determine the preliminary minimum number of steps required to reach average stability across all targeted biomechanical variables. In this way, experimental designs will account for reaching stability across all variables.

This study has some limitations. Firstly, participants performed overground running in an indoor environment that may not be ideal for describing running kinematics/kinetics. The running track layout favored the repeated acquisition of running steps, but runners did not maintain constant speed over the course of the entire track. This was caused due to turns and changes in approaching speed to the force platform. In our study, the average running speed across all participants and conditions was approximately 2.7 m s^{-1} . The vast majority of studies investigating novice runners report a running speed within 2.5 and 3.5 m s^{-1} (see P3, L79–94, and Fig. 1C), which contains the average running speed from this experiment. Since running biomechanical variables are speed-dependent, extrapolations of our results to running conditions where the running speed is outside this range should be done with caution. Secondly, another limitation of the study was the use of unfamiliar shoes by the runners. Despite the extensive procedures provided to runners to familiarize themselves with the shoes, the lack of long-term familiarity might affect the variability of the investigated biomechanical variables^{38,42}. Future studies comparing the stability of biomechanical variables between short-term and long-term familiarization to running shoes are needed to address this limitation. Lastly, there was a large number of statistical comparisons in this study (> 600,000 paired tests) and another limitation of the study is the lack of correction for multiple comparisons. Due to the large scale of comparisons performed in this study, the possibility of type I errors cannot be overruled. Therefore, extrapolating the conclusions from our study to different types of shoes (e.g., motion-control shoes), running surfaces (outdoors or treadmill) and other populations (elite athletes) must be done with caution. Furthermore, the interpretation of the results should take into consideration that the task is performed indoors, therefore translating these results to outdoor conditions with changing terrains and/or environmental conditions might not be ideal.

In summary, our results suggest that the selection of both sample size and number of steps can influence the statistical outcomes for the comparison of normal vs silent running. Moreover, the influence of sample size and number of steps is variable-dependent. Therefore, our results provided experimental evidence to recommend that studies involving analysis of traditional running biomechanical variables use a minimum of 25 participants and 25 steps from each participant to increase the likelihood of describing accurate results with large effect sizes when comparing running in different conditions. These guidelines may assist researchers to design optimal experimental protocols in running biomechanics, as well as provide standard requirements to assess the quality of experimental outcomes in this research field.

Received: 3 February 2020; Accepted: 20 January 2021

Published online: 04 February 2021

References

1. Preatoni, E. *et al.* Movement variability and skills monitoring in sports. *Sports Biomech.* **12**, 69–92 (2013).
2. Forrester, S. E. Selecting the number of trials in experimental biomechanics studies. *Int. Biomech.* **2**, 62–72 (2015).
3. Bates, B., Dufek, J. & Davis, H. The effect of trial size on statistical power. *Off. J. Am. Coll. Sports Med.* **0195–9131**, 1059–1065 (1992).
4. Knudson, D. Confidence crisis of results in biomechanics research. *Sports Biomech.* **16**, 425–433 (2017).
5. Oliveira, A. S., Gizzi, L., Farina, D. & Kersting, U. G. Motor modules of human locomotion: influence of EMG averaging, concatenation, and number of step cycles. *Front. Hum. Neurosci.* **8**, 1–9 (2014).
6. Bartlett, R., Wheat, J. & Robins, M. Is movement variability important for sports biomechanists?. *Sports Biomech.* **6**, 224–243 (2007).
7. Bates, B., Osternig, L. R., Sawhill, J. A. & James, S. L. An assessment of subject variability, subject-shoe interaction, and the evaluation of running shoes using ground reaction force data. *J. Biomech.* **16**, 181–191 (1983).
8. Dos' Santos, T., Comfort, P. & Jones, P. A. Average of trial peaks versus peak of average profile: impact on change of direction biomechanics. *Sports Biomech.* **19**, 483–492 (2020).
9. Dames, K. D., Smith, J. D. & Heise, G. D. Averaging trials versus averaging trial peaks: impact on study outcomes. *J. Appl. Biomech.* **33**, 233–236 (2017).
10. Mullineaux, D. R., Bartlett, R. M. & Bennett, S. Research design and statistics in biomechanics and motor control. *J. Sports Sci.* **19**, 739–760 (2001).
11. Taylor, P. G., Lee, K. Y., Landeo, R., O'Meara, D. M. & Millett, E. Determining optimal trial size using sequential analysis. *J. Sports Sci.* **33**, 300–308 (2015).
12. Vagenas, G., Palaiothodorou, D. & Knudson, D. Thirty-year trends of study design and statistics in applied sports and exercise biomechanics research. *Int. J. Exerc. Sci.* **11**, 239–259 (2018).
13. Stergiou, N., Harbourne, R. T. & Cavanaugh, J. T. Optimal movement variability: a new theoretical perspective for neurologic physical therapy. *J. Neurol. Phys. Ther.* **30**, 120–129 (2006).
14. Tam, N., Tucker, R. & Astephen Wilson, J. L. Individual responses to a barefoot running program. *Am. J. Sports Med.* **44**, 777–784 (2016).
15. Messier, S. P. *et al.* A 2-year prospective cohort study of overuse running injuries: the runners and injury longitudinal study (TRAILS). *Am. J. Sports Med.* **46**, 2211–2221 (2018).
16. Rice, D., Jamison, S. & Davis, I. Footwear matters: influence of footwear and foot strike on load rates during running. *Med. Sci. Sports Exerc.* **48**, 2462–2468 (2016).
17. Phan, X. *et al.* Running quietly reduces ground reaction force and vertical loading rate and alters foot strike technique. *J. Sports Sci.* **35**, 1636–1642 (2017).
18. Willson, J. D. *et al.* Short-term changes in running mechanics and foot strike pattern after introduction to minimalistic footwear. *PM and R* **6**, 34–43 (2014).
19. Breine, B. *et al.* Initial foot contact and related kinematics affect impact loading rate in running. *J. Sports Sci.* **35**, 1556–1564 (2017).
20. Knudson, D. Significant and meaningful effects in sports biomechanics research. *Sports Biomech.* **8**, 96–104 (2009).
21. Batterham, A. M. & Atkinson, G. How big does my sample need to be? A primer on the murky world of sample size estimation. *Phys. Ther. Sport* **6**, 153–163 (2005).
22. Brughelli, M., Cronin, J. & Chaouachi, A. Effects of running velocity on running kinetics and kinematics. *J. Strength Cond. Res.* **25**, 933–939 (2011).
23. Pircoveanu, C.-I. & Oliveira, A. S. The use of multi-directional footfall sound recordings to describe running vertical impact properties. *J. Sports Sci.* **39**, 267–274 (2020).
24. Pircoveanu, C.-I., Dam, P., Brandi, A., Bilgram, M. & Oliveira, A. S. Fatigue-related changes in vertical impact properties during normal and silent running. *J. Sports Sci.* 1–9 (2020). <https://doi.org/10.1080/02640414.2020.1824340>.
25. Shih, H. T. *et al.* Four weeks of training with simple postural instructions changes trunk posture and foot strike pattern in recreational runners. *Phys. Ther. Sport* **35**, 89–96 (2019).
26. Ching, E. *et al.* Impact loading during distracted running before and after auditory gait retraining. *Int. J. Sports Med.* **39**, 1075–1080 (2018).
27. Tate, J. J. & Milner, C. E. Sound-intensity feedback during running reduces loading rates and impact peak. *J. Orthop. Sports Phys. Ther.* **47**, 565–569 (2017).
28. Mullineaux, D., Milner, C., Davis, I. & Hamill, J. Normalization of ground reaction forces. *J. Appl. Biomech.* **22**, 230–233 (2006).
29. Bowser, B. J., Fellin, R., Milner, C. E., Pohl, M. B. & Davis, I. S. Reducing impact loading in runners: a one-year follow-up. *Med. Sci. Sports Exerc.* **50**, 2500 (2018).
30. Willy, R. W., Pohl, M. B. & Davis, I. S. Calculation of vertical load rates in the absence of vertical impact peaks. *American Society of Biomechanics Meeting* (2008).
31. Blackmore, T., Willy, R. W. & Creaby, M. W. The high frequency component of the vertical ground reaction force is a valid surrogate measure of the impact peak. *J. Biomech.* **49**, 479–483 (2016).
32. Daniel, W. *Biostatistics: A Foundation for Analysis in the Health Sciences* (Wiley, New York, 1999).
33. Salo, A., Grimshaw, P. N. & Viitasalo, J. T. Reliability of variables in the kinematic analysis of sprint hurdles. *Med. Sci. Sports Exerc.* **29**, 383–389 (1997).
34. Chambon, N., Delattre, N., Guéguen, N., Berton, E. & Rao, G. Shoe drop has opposite influence on running pattern when running overground or on a treadmill. *Eur. J. Appl. Physiol.* **115**, 911–918 (2015).
35. Oliveira, A. S., Gizzi, L., Ketabi, S., Farina, D. & Kersting, U. G. Modular control of treadmill vs overground running. *PLoS ONE* **11**, 4 (2016).
36. Fellin, R. E., Davis, I. S. & Manal, K. Comparison of lower extremity kinematic curves during overground and treadmill running. *J. Appl. Biomech.* **26**, 407–414 (2010).
37. Logan, S., Hunter, I., Hopkins, J. T., Feland, J. B. & Parcell, A. C. Ground reaction force differences between running shoes, racing flats, and distance spikes in runners. *J. Sports Sci. Med.* **9**, 147–153 (2010).
38. Breine, B., Malcolm, P., Frederick, E. C. & De Clercq, D. Relationship between running speed and initial foot contact patterns. *Med. Sci. Sports Exerc.* **46**, 1595–1603 (2013).
39. Breine, B. *et al.* Running speed-induced changes in foot contact pattern influence impact loading rate. *Eur. J. Sport Sci.* **19**, 774–783 (2019).
40. Van den Berghe, P., Six, J., Gerlo, J., Leman, M. & De Clercq, D. Validity and reliability of peak tibial accelerations as real-time measure of impact loading during over-ground rearfoot running at different speeds. *J. Biomech.* **86**, 238–242 (2019).
41. Khassestarash, A., Hassannejad, R., Etefagh, M. M. & Oskouei, A. E. Vibration settling time of the gastrocnemius remains constant during an exhaustive run in rear foot strike runners. *J. Biomech.* **93**, 140–146 (2019).

42. Hunter, J. G. *et al.* Standardized lab shoes do not decrease loading rate variability in recreational runners. *J. Appl. Biomech.* **36**, 340–344 (2020).
43. Wouda, E. J. *et al.* Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors. *Front. Physiol.* **9**, 1–14 (2018).
44. Bredeweg, S. W., Kluitenberg, B., Bessem, B. & Buist, I. Differences in kinetic variables between injured and noninjured novice runners: A prospective cohort study. *J. Sci. Med. Sport* **16**, 205–210 (2013).
45. Stiffler-Joachim, M. R., Wille, C. M., Kliethermes, S. A., Johnston, W. & Heiderscheidt, B. C. Foot angle and loading rate during running demonstrate a nonlinear relationship. *Med. Sci. Sports Exerc.* **51**, 2067–2072 (2019).
46. Stiles, V. H., Griew, P. J. & Rowlands, A. V. Use of accelerometry to classify activity beneficial to bone in premenopausal women. *Med. Sci. Sports Exerc.* **45**, 2353–2361 (2013).
47. Morin, J. B., Samozino, P. & Millet, G. Y. Changes in running kinematics, kinetics, and spring-mass behavior over a 24-h run. *Med. Sci. Sports Exerc.* **43**, 829–836 (2011).
48. Rowlands, A. V. & Stiles, V. H. Accelerometer counts and raw acceleration output in relation to mechanical loading. *J. Biomech.* **45**, 448–454 (2012).
49. Pollard, C. D., Ter Har, J. A., Hannigan, J. J. & Norcross, M. F. Influence of maximal running shoes on biomechanics before and after a 5K run. *Orthop. J. Sports Med.* **6**, 1–5 (2018).

Acknowledgements

The authors would like to thank Juliana Excel Santana for providing a scientific review of the final version of this manuscript and Christine Dzialo for proofreading the manuscript. This project was fully funded by Kulturministeriets Forskningsudvalg (FPK.2018-0048).

Author contributions

Design of experiment: A.O. and C.I.P.; Data collection: C.I.P.; Data processing and analysis: A.O. and C.I.P.; Data interpretation: A.O. and C.I.P.; Results/figures preparation: A.O. and C.I.P.; Drafting of the manuscript: C.I.P.; Final approval of manuscript: A.O. and C.I.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82876-z>.

Correspondence and requests for materials should be addressed to A.S.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021