

Original article

Classification of higher- and lower-mileage runners based on running kinematics

Christian A. Clermont^a, Angkoon Phinyomark^{a,b}, Sean T. Osis^{a,c}, Reed Ferber^{a,c,d,*}

^a Faculty of Kinesiology, University of Calgary, Calgary, AB T2N 1N4, Canada

^b Institute for Scientific Interchange (ISI) Foundation, Turin 10126, Italy

^c Running Injury Clinic, Calgary, AB T2N 1N4, Canada

^d Faculty of Nursing, University of Calgary, Calgary, AB T2N 1N4, Canada

Received 21 January 2017; revised 27 April 2017; accepted 5 June 2017

Available online 18 August 2017

Abstract

Background: Running-related overuse injuries can result from the combination of extrinsic (e.g., running mileage) and intrinsic risk factors (e.g., biomechanics and gender), but the relationship between these factors is not fully understood. Therefore, the first purpose of this study was to determine whether we could classify higher- and lower-mileage runners according to differences in lower extremity kinematics during the stance and swing phases of running gait. The second purpose was to subgroup the runners by gender and determine whether we could classify higher- and lower-mileage runners in male and female subgroups.

Methods: Participants were allocated to the “higher-mileage” group (≥ 32 km/week; $n = 41$ (30 females)) or to the “lower-mileage” group (≤ 25 km; $n = 40$ (29 females)). Three-dimensional kinematic data were collected during 60 s of treadmill running at a self-selected speed (2.61 ± 0.23 m/s). A support vector machine classifier identified kinematic differences between higher- and lower-mileage groups based on principal component scores.

Results: Higher- and lower-mileage runners (both genders) could be separated with 92.59% classification accuracy. When subgrouping by gender, higher- and lower-mileage female runners could be separated with 89.83% classification accuracy, and higher- and lower-mileage male runners could be separated with 100% classification accuracy.

Conclusion: These results demonstrate there are distinct kinematic differences between subgroups related to both mileage and gender, and that these factors need to be considered in future research.

2095-2546/© 2019 Published by Elsevier B.V. on behalf of Shanghai University of Sport. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Biomechanics; Clinical biomechanics; Gait analysis; Kinematics; Motion analysis; Running mileage; Running subgroups

1. Introduction

Running is one of the most common sports and recreational activities around the world. However, many runners will experience a running-related injury each year, with studies reporting incidence rates ranging between 18.2% and 92.4%.^{1–3} The etiology of overuse running injuries is multifactorial and can result from the interaction of many extrinsic (e.g., weekly training days, footwear, and weekly training mileage) and intrinsic risk factors (e.g., age, anatomic factors, foot strike pattern, biomechanics, and gender).⁴ Some research suggests weekly

running mileage to be a leading extrinsic risk factor associated with running-related injuries,^{2,5} while other studies have shown no significant differences in injury proportions between higher- and lower-mileage runners.^{6,7}

Recreational runners with a long history of running have been shown to demonstrate less frequent knee pain, as compared to non-runners, which may be the result of differences in running gait patterns, among other intrinsic and extrinsic factors.⁸ On the other hand, a study by Silvernail et al.⁹ demonstrated that older adults exhibited similar movement patterns and coordination variability to younger runners when the groups were matched for weekly running mileage. These similarities may indicate that maintaining a running program with higher weekly mileage may be protective against an increased risk of certain types of injuries associated with the

Peer review under responsibility of Shanghai University of Sport.

* Corresponding author.

E-mail address: rferber@ucalgary.ca (R. Ferber)

neuromuscular degenerative effects of aging. Other studies have also shown runners who engage in high weekly running mileage tend to sustain more hip and hamstring injuries,¹⁰ whereas runners with low weekly running mileage sustain more injuries around the knee, such as patellofemoral pain.⁶ To better understand the underlying interaction between running mileage and specific types of injuries, a biomechanical analysis of high- and low-mileage runners may be useful. However, there has been limited research into the interaction between running kinematics and weekly training mileage.

To our knowledge, only one study has explored the differences in running gait kinematics between higher- and lower-mileage runners to gain insight into potential atypical or injurious pathomechanics. Boyer et al.¹¹ reported differences between higher- and lower-mileage runners with respect to transverse plane pelvic and hip; frontal plane hip and knee; and sagittal and frontal plane foot kinematics. However, only stance phase kinematics were analyzed, and Watari et al.¹² reported that lower limb kinematics during swing phase are also important to consider when discriminating between running subgroups related to injury. Additionally, Schmitz et al.¹³ reported certain kinematic variables during swing (e.g., foot and thigh positions) to be significant predictors for impact peak and loading rate during running, which are considered important metrics related to overuse running-related injuries and associated with running injuries such as tibial stress fractures.¹⁴ Therefore, it is important to consider gait kinematics throughout both the stance and swing phases.

Boyer et al.¹¹ also grouped males and females together in each mileage group. It is well documented that distinct kinematic differences exist between male and female runners, which are primarily related to frontal and transverse motion of the knee and hip.^{15,16} Therefore, it is important to further subgroup runners with different weekly running mileage in order to improve homogeneity of these running subgroups and accurately quantify their running kinematics. In support of this statement, Gehring et al.¹⁷ reported gender-related differences in frontal plane hip and knee kinematics in high mileage runners, but runners with low mileage were not included in this investigation. However, these authors¹⁷ did not explore any sagittal or transverse plane kinematic differences. Thus, a more thorough analysis of three-dimensional (3D) kinematic gait differences related in male and female subgroups between higher- and lower-mileage runners is needed to provide greater insight into the relationship between weekly mileage, gender, and running biomechanics.

Despite the potential for many features needed to classify male and female, higher- and lower-mileage runners, too many discrete kinematic variables in the analysis can complicate the clinical interpretation of the gait patterns for these homogeneous running subgroups. Furthermore, analyzing the discrete kinematic variables with traditional inferential statistics ignores important information about the entire lower limb's coordination of segment motions during running. To overcome these shortcomings, recent studies have used a principal component analysis (PCA) in order to analyze all kinematic variables and time points and allow for a better

understanding of the segment couplings and joint motion in all 3 planes of motion.^{12,16,18-21} A PCA approach, combined with machine learning methods (e.g., support vector machine (SVM)), have also been applied successfully in running biomechanics research to separate and classify running subgroups according to age, gender, performance, and running-related injuries with greater than 80% accuracy.^{16,18-20,22} Thus, a similar approach may be used to separate and classify runners according to gender and mileage.

The first purpose of this study was to determine whether we could classify higher- and lower-mileage runners according to differences in lower extremity kinematics during the stance and swing phases of running gait. The second purpose was to subgroup the runners by gender and determine whether we could classify higher- and lower-mileage runners in male and female subgroups. We hypothesized that we could classify higher- and lower-mileage runners in all groups (i.e., (1) both genders, (2) females, and (3) males) with greater than 80% accuracy. We further hypothesized, based on the aforementioned research, that differences between mileage groups would be primarily related to frontal and transverse plane kinematics.

2. Methods

2.1. Subjects

Kinematic data during treadmill running were queried from an existing database²¹ and 81 healthy runners were included in this study. To build upon previous research, we chose to use separate lower- and higher-mileage runners by using the same dichotomy as Boyer et al.¹¹ As a result, 40 runners (29 females and 11 males) who ran a weekly self-reported average of ≤ 25 km (20.06 ± 4.22 km/week, mean \pm SD) were labelled as "lower-mileage" runners, and 41 runners (30 females and 11 males) who ran a weekly self-reported average of ≥ 32 km (49.65 ± 16.31 km/week) were labelled as "higher-mileage" runners. A recall bias may be associated with self-reported weekly running mileage, but Diderisken et al.²³ compared self-reported subjective evaluation of running mileage with Global Position System-measured distance, and found no significant difference between the 2 methods. Based on this evidence, we are confident that the self-reported weekly mileage from runners in our study accurately reflects whether they meet the higher- or lower-mileage running groups. Data collection was approved by the University of Calgary's Conjoint Health Research Ethics Board (CHREB: REB15-0557), and all participants provided written informed consent prior to participating.

2.2. Data collection

An 8-camera VICON motion capture system (MX3+; Vicon Motion Systems, Oxford, UK) was used to collect 3D kinematic data at 200 Hz during running on an instrumented-treadmill (Bertec Corp., Columbus, OH, USA). Twenty-two spherical retro-reflective markers (9-mm diameter; Mocap Solutions, Huntington Beach, CA, USA) were placed over the following anatomic landmarks bilaterally (Fig. 1): 1st and 5th metatarsal heads; medial and lateral malleoli; tibial tuberosity; head of the

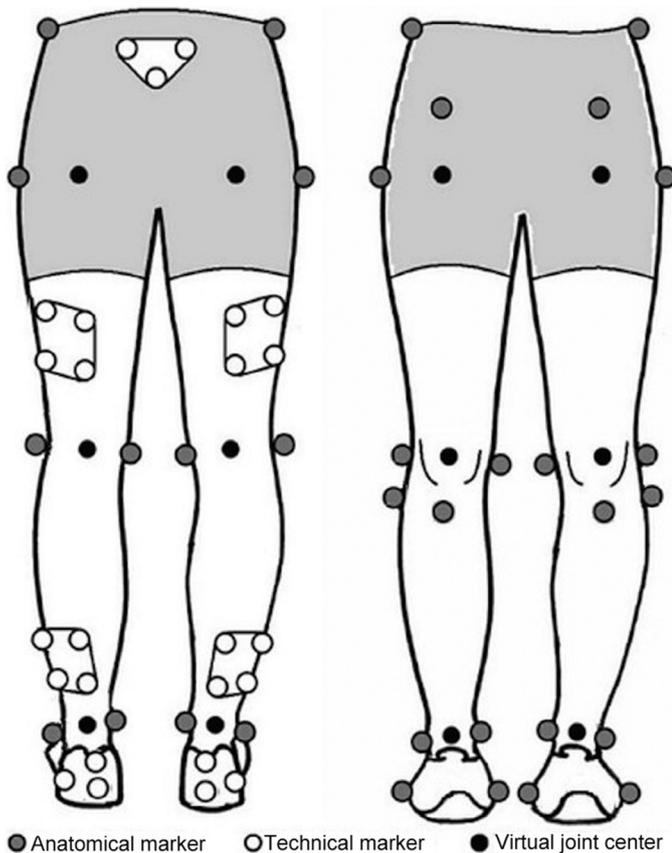


Fig. 1. Placement of the reflective markers on the lower extremities. Anatomic markers (gray circles) were used to create segmental anatomic coordinate systems while the technical clusters (open circles) were used for tracking purposes during the running trials, and virtual joint centers (black circles) were defined relative to the technical coordinate system using the technical marker clusters during the standing calibration trial, and were created at the hip, knee, and ankle.

fibula; medial and lateral femoral condyles; greater trochanter; anterior superior iliac spine; and iliac crest. Technical marker clusters, comprised of rigid, plastic shells, were positioned on the pelvis (3 markers) and bilateral thigh and shank (4 markers each) with self-adhering straps. Three markers were taped to the heel contour of each of the test shoes. These 25 technical markers represented 7 rigid segments.

Following placement of all the anatomic and technical markers, the participant was asked to stand on the treadmill for a static trial. Standing position was controlled using a graphic template placed on the treadmill with their feet positioned 0.3 m apart and pointing straight ahead. Once the participant was in the standardized position, he or she was asked to cross his or her arms over the chest and stand still while 1 s of marker location data were recorded. Upon completion of the static trial, the anatomic markers were removed to allow the participant to move with less restriction. Following approximately 5 min of acclimation, running kinematic data were collected for 60 s while each participant ran on the treadmill at a self-selected speed (2.61 ± 0.23 m/s, mean \pm SD) while wearing standardized footwear (Air Pegasus, Nike, Beaverton, OR, USA).

2.3. Data analysis

2.3.1. Initial input data

Joint angles were calculated using 3D GAIT custom software (Running Injury Clinic Inc., Calgary, Canada) according to previously published methods^{24,25} and normalized to 35 data points for stance and 65 data points for swing. To detect the timing of foot strike and toe-off events, methodology using a PCA combined with a machine learning approach was employed, and these methods have been previously detailed in Osis et al.²⁶ Kinematic data were averaged from 10 consecutive strides of data to produce mean angles for all 3 planes of motion for each of the 3 lower extremity joints (ankle, knee, and hip) in the local coordinate system as well as the pelvis segment in the global coordinate system. Kinematic angles were also calculated for the foot segment in the global coordinate system, but only for the transverse and sagittal planes. The gait data points (14 waveforms \times 100 time points) were combined into one 1400-dimensional row vector for each participant.

2.3.2. Feature extraction

Given the large number of data points and potential redundancy of data, a PCA was used to reduce the data. PCA is a linear transformation technique used to convert a set of possibly correlated variables into a set of linearly uncorrelated variables by determining new bases (principal components (PCs)) that maximize variance sequentially in each PC.²⁷ Prior to the PCA, all variables were standardized to a mean of 0 and standard deviation of 1 by calculating the z-score for each variable. As a result, the input vector for the PCA was an 81-by-1400 matrix, which generated an 81-by-80 PC scores matrix (i.e., a total of 80 PCs).

2.3.3. Machine learning

To examine the utility of PCA features in identifying and discriminating the differences between higher- and lower-mileage runners in the 3 groups of interest (i.e., (1) both genders, (2) female runners, and (3) male runners) an SVM approach was used and evaluated based on classification accuracy. The models were created using the PC scores as input for the linear SVM^{16,28} with a soft margin parameter c set at 1 based on the methods reported by Fukuchi et al.¹⁹ and Phinyomark et al.¹⁶ to perform the classifications for each group comparison. To compare higher- vs. lower-mileage groups, irrespective of gender, an 81-by-80 matrix was used as baseline feature vector for the SVM classifier. To separate and classify higher- and lower-mileage runners in the female subgroup, a 59-by-80 initial input vector was used as initial input for the SVM classifier. To separate and classify higher- and lower-mileage runners in the male subgroup, a 22-by-80 initial input vector was used as initial input for the SVM classifier.

For the feature selection procedure, the number of PC scores used as features in the SVM classifiers were increased in descending order of Cohen's d effect size, and once the maximum classification rate of the SVM was reached according to the 10-fold cross validation methods, these features were considered the optimal number and used as the output for the SVM.^{19,28} For the 10-fold cross validation, PC data were

randomly partitioned into 10 equally sized sub-datasets. Then, a single sub-dataset was retained as testing data while the remaining 9 sub-datasets were used as training data for the classification model. The cross-validation process was then repeated 10 times, and a single classification rate was computed by averaging from all 10 results.¹⁶

2.4. Evaluating functions

All analyses were performed using customized MATLAB 9.0 software (The Mathworks Inc., Natick, MA, USA). Three separate independent *t* tests ($p < 0.05$) compared anthropometric (i.e., height and mass), speed, and demographic (i.e., age, running experience, and weekly running mileage) measures between (1) higher- and lower-mileage runners, (2) higher- and lower-mileage female runners, and (3) higher- and lower-mileage male runners.

To visualize differences between groups, all PC scores that were used to yield the maximum accuracy for discriminating between groups were used to reconstruct the original joint and segment angle waveforms. Specifically, all selected PC scores were multiplied by the transpose of the PC coefficient matrix. Then, to reconstruct gait waveforms in the original coordinate space, each subject's sample was multiplied by the sample's standard deviation vector with addition of the mean vector. To interpret the biomechanical meaning of the waveforms, the standardized effect size, Cohen's *d* was used to identify meaningful differences.²⁹ Based on Cohen's conventional criteria,³⁰ a meaningful difference between groups was defined as a large effect size ($d \geq 0.8$).

3. Results

3.1. Higher- vs. lower-mileage runners (both genders)

Anthropometric and demographic comparisons between higher- and lower-mileage running groups can be seen in Table 1. Overall, there were no significant differences between these groups in terms of age, height, mass, or self-reported years of running experience. The higher-mileage runners exhibited a higher self-reported weekly mileage and a significantly faster self-selected speed ($p < 0.01$) during the data collection.

Higher- and lower-mileage runners could be separated with 92.59% classification accuracy using a linear SVM 10-fold cross-validation method with the top 57 ranked PCs explaining 52.71% of the total variance. Meaningful differences ($d > 0.8$) in kinematic variables for higher- and lower-mileage runners

Table 1 Participant characteristics (both genders) (mean ± SD).

| Parameter | Lower-mileage group (n = 40) | Higher-mileage group (n = 41) | <i>p</i> |
|---------------------------|------------------------------|-------------------------------|----------|
| Age (year) | 41.38 ± 11.39 | 42.00 ± 10.23 | 0.80 |
| Height (cm) | 168.80 ± 9.34 | 171.13 ± 7.43 | 0.22 |
| Mass (kg) | 70.77 ± 14.61 | 66.77 ± 11.46 | 0.18 |
| Running experience (year) | 8.49 ± 8.43 | 10.83 ± 10.59 | 0.28 |
| Weekly mileage (km) | 20.06 ± 4.22 | 49.65 ± 16.31 | <0.01 |
| Speed (m/s) | 2.53 ± 0.26 | 2.68 ± 0.16 | <0.01 |

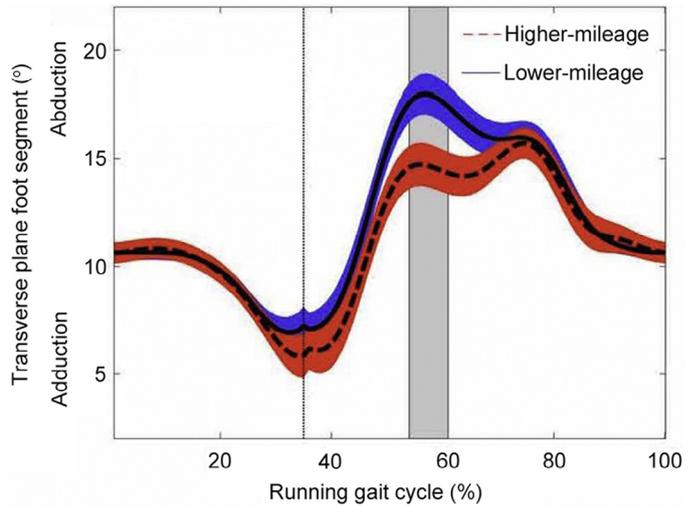


Fig. 2. Both genders: the mean of individual time-normalized transverse plane foot angles from kinematic analyses for higher- and lower-mileage runners (both genders) during stance phase (1%–35%) and swing phase (36%–100%) of running. The shaded area indicates meaningful differences ($d > 0.8$) between groups.

were found in the transverse plane foot angles during mid-swing (54%–61% of the gait cycle, Fig. 2).

3.2. Higher- vs. lower-mileage female runners

For female runners, anthropometric and demographic comparisons between higher- and lower-mileage groups can be seen in Table 2. There were no significant differences between these groups in terms of age, mass, or self-reported years of running experience. The higher-mileage female runners were significantly taller ($p = 0.04$), exhibited a higher self-reported weekly mileage, and a significantly faster self-selected running speed ($p < 0.01$) during the data collection.

Higher- and lower-mileage female runners could be separated with 89.83% classification accuracy using the top 47 PCs explaining 45.17% of the total variance. There were meaningful differences between higher- and lower-mileage female runners in terms of sagittal plane knee motion at different periods of the swing phase (49%–57% and 73%–88% of the gait cycle, Fig. 3A), sagittal foot motion during late swing (77%–82% of the gait cycle, Fig. 3B), as well as transverse knee motion during mid-stance (12%–19% of the gait cycle, Fig. 3C).

Table 2 Female participant characteristics (mean ± SD).

| Parameter | Lower-mileage group (n = 29) | Higher-mileage group (n = 30) | <i>p</i> |
|---------------------------|------------------------------|-------------------------------|----------|
| Age (year) | 41.48 ± 11.35 | 40.30 ± 9.79 | 0.67 |
| Height (cm) | 164.76 ± 6.36 | 168.32 ± 6.28 | 0.04 |
| Mass (kg) | 66.30 ± 12.49 | 61.78 ± 8.05 | 0.10 |
| Running experience (year) | 6.62 ± 5.86 | 10.10 ± 9.68 | 0.10 |
| Weekly mileage (km) | 19.93 ± 4.31 | 47.77 ± 14.01 | <0.01 |
| Speed (m/s) | 2.45 ± 0.26 | 2.69 ± 0.08 | <0.01 |

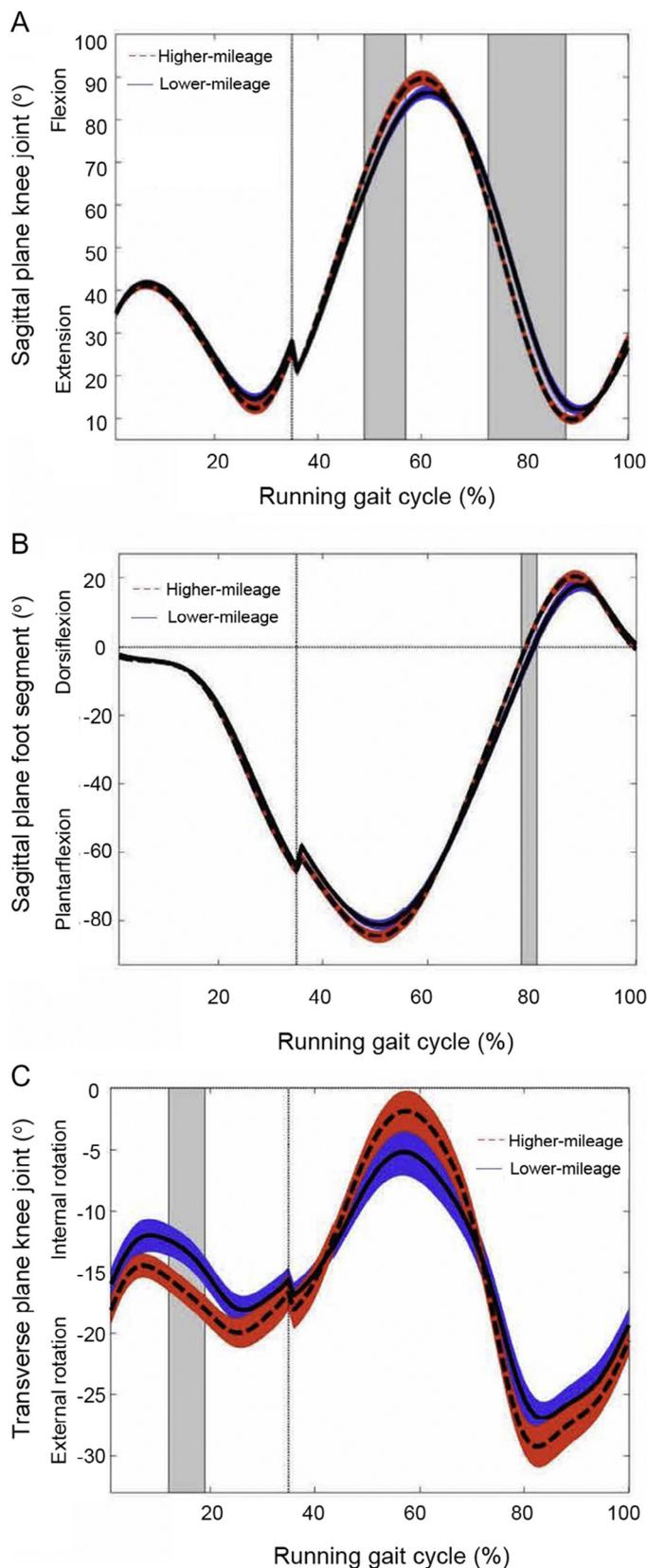


Fig. 3. Female: the mean of individual time-normalized joint and segment angles from kinematic analyses for higher- and lower-mileage female runners during stance phase (1%–35%) and swing phase (36%–100%) of running. (A) sagittal plane knee joint; (B) sagittal plane foot segment; and (C) transverse plane knee joint. The shaded area indicates meaningful differences ($d > 0.8$) between groups.

Table 3
Male participant characteristics (mean \pm SD).

| Parameter | Lower-mileage group (n = 11) | Higher-mileage group (n = 11) | p |
|---------------------------|------------------------------|-------------------------------|-------|
| Age (year) | 41.09 \pm 12.03 | 46.63 \pm 10.39 | 0.26 |
| Height (cm) | 179.45 \pm 7.40 | 178.82 \pm 4.20 | 0.81 |
| Mass (kg) | 82.53 \pm 13.62 | 80.39 \pm 7.66 | 0.66 |
| Running experience (year) | 13.41 \pm 12.0 | 12.68 \pm 12.95 | 0.89 |
| Weekly mileage (km) | 20.41 \pm 4.16 | 54.77 \pm 21.35 | <0.01 |
| Speed (m/s) | 2.73 \pm 0.14 | 2.68 \pm 0.29 | 0.65 |

3.3. Higher- vs. lower-mileage male runners

For male runners, anthropometric and demographic comparisons between higher- and lower-mileage groups can be seen in Table 3. There was only 1 significant difference between groups, which was the higher-mileage group exhibited a higher self-reported weekly running mileage ($p < 0.01$).

Higher- and lower-mileage male runners could be separated with 100% classification accuracy by using the top 12 ranked PCs which explain 6.13% of the total variance. There were meaningful differences between male higher- and lower-mileage runners at different points in the gait cycle in all joints and segments for various plane of motion. In brief, lower-mileage male runners demonstrated greater anterior pelvic tilt throughout 92% of the gait cycle compared to their higher-mileage counterparts (Fig. 4A). Higher-mileage male runners displayed greater transverse pelvic rotation toward the ipsilateral leg throughout most of the gait cycle (5%–78% and 85%–93%, Fig. 4B). Greater hip adduction was observed in the higher-mileage runners during most of stance and early swing (9%–55%). However, during mid- to late-swing (59%–90%), the opposite effect occurred, in which lower-mileage runners displayed greater hip adduction than higher-mileage male runners (Fig. 4C). With respect to knee motion, higher-mileage runners exhibited a more flexed knee throughout stance (6%–35%) and during the later stages of swing (69%–92%) (Fig. 4D). Finally, lower-mileage male runners demonstrated greater foot abduction (i.e., toe-out) throughout 95% of the gait cycle (Fig. 4E).

4. Discussion

4.1. Higher- vs. lower-mileage runners (both genders)

The first purpose of this study was to determine whether we could separate and classify higher- and lower-mileage runners, both independently and according to gender, based on differences in running kinematics during stance and swing phases. In support of our hypothesis, higher- and lower-mileage runners could be separated with 92.59% classification accuracy with the top 57 ranked PCs based on lower extremity gait kinematics and using a multivariate analysis and a machine learning approach.

The findings of the current study support those of Boyer et al.¹¹ and indicate there are systematic differences in running kinematics for higher- and lower-mileage runners. These findings are also supported by previous research showing that

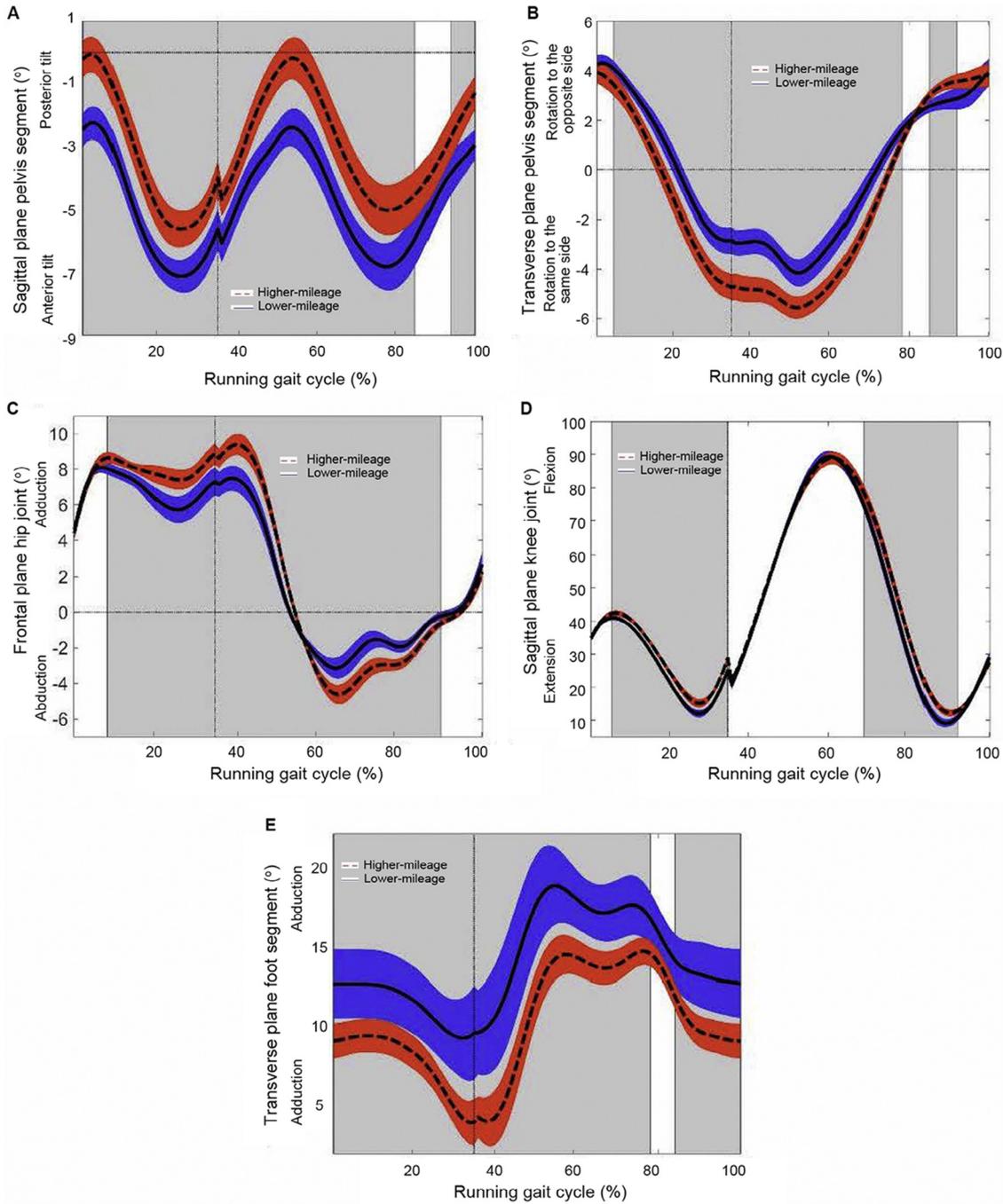


Fig. 4. Male: the mean of individual time-normalized joint and segment angles from kinematic analyses for higher- and lower-mileage male runners during stance phase (1%–35%) and swing phase (36%–100%) of running. (A) sagittal plane pelvis segment; (B) transverse plane pelvis segment; (C) frontal plane hip joint; (D) sagittal plane knee joint; and (E) transverse plane foot segment. The shaded area indicates meaningful differences ($d > 0.8$) between groups.

intermediate and higher-order PCs are associated with the subtle movement patterns of running gait³¹ and contain valuable information about between-group variation to help classify different types of runners. Our results provide further evidence that a machine learning approach using low-, intermediate-, and higher-order PC scores can provide insight into the complex relationships of biomechanical gait variables for dominant and subtle movements throughout the entire gait cycle for higher- and lower-mileage runners.

It was hypothesized that the meaningful differences between higher- and lower-mileage running groups would be primarily related to transverse and frontal motion. In partial support of this hypothesis, we found meaningful ($d > 0.8$) differences for the foot in the transverse plane, where higher-mileage runners demonstrated greater internal rotation of the foot during mid-swing. Boyer et al.¹¹ reported that higher-mileage runners exhibited greater internal rotation of the foot during stance phase, and the current study significantly builds on their

research, wherein the focus was limited to stance phase running kinematics. Swing phase lower-limb kinematics have been shown to be important when classifying runners and potentially studying the relationship between running mechanics and injury.^{12,13} In particular, Schmitz et al.¹³ indicated that the foot position during mid-swing is predictive of the magnitude of impact peak at foot contact. Impact peak is inherently related to loading rate, which has been associated with lower leg overuse running-related injuries, such as tibial stress fractures.¹⁴ Although Schmitz et al.¹³ did not explore the relationship between joint or segment angles with impact peak and loading rate, the differences in transverse plane foot angles between higher- and lower-mileage runners in our study may be indicative of different levels of impact peaks and loading rates. This result may help explain why more highly trained, competitive runners experience more lower-leg injuries than recreational runners³² but further research is necessary to better understand how these variables are related to injury.

4.2. Higher- vs. lower-mileage female runners

The second purpose of our study was to determine whether we could classify higher- and lower-mileage runners in male and female subgroups separately. In support of our hypothesis, higher- and lower-mileage female runners could be separated with 89.83%, and the differences were primarily related to transverse motion of the knee. Specifically, during mid-stance, lower-mileage female runners demonstrated greater internal rotation of the knee in the transverse plane of motion as compared to their higher-mileage female counterparts. Increased knee internal rotation can result in greater torsional strain to tissues at the knee joint, such as the iliotibial band.³³ This postulation is supported by previous research that has shown that females with iliotibial band syndrome exhibit increased knee internal rotation angles during stance in comparison to female controls.^{18,33} Moreover, knee injuries have been shown to be more common in lower-mileage runners;^{6,8,10} therefore, the increased knee internal rotation observed in our lower-mileage female runners may provide some biomechanical insight as to the etiology of this injury pattern.

4.3. Higher- vs. lower-mileage male runners

Higher- and lower-mileage male runners could be separated with 100% classification accuracy using the top 12 ranked PCs, which explains 6.13% of the total variance, and there were meaningful differences between higher- and lower-mileage runners in several joints and segments. At the pelvis, lower-mileage male runners demonstrated greater anterior tilt throughout 92% of the gait cycle and these results complement the work of Preece et al.,³⁴ who also reported that recreational male runners exhibit a more pronounced thoracic and pelvic inclination than highly trained, elite males. Although Preece et al.³⁴ grouped their runners according to race performance, it can be speculated that their elite runners had higher weekly training volume than recreational runners. Similar to Preece et al.,³⁴ our findings challenge some training protocols that

instruct recreational runners to increase trunk lean or anterior pelvic tilt to improve performance or reduce injury.^{35,36} Nevertheless, due to the cross-sectional nature of our study, it is difficult to speculate on the long-term effects of pelvic tilt on performance and injury.

Higher-mileage male runners exhibited greater transverse plane pelvic rotation toward the ipsilateral leg throughout stance and most of swing compared to lower-mileage male runners. Boyer et al.¹¹ also found similar results during the stance phase, in which their higher-mileage group also demonstrated greater transverse plane pelvic rotation toward the ipsilateral leg. Furthermore, we found that higher-mileage runners displayed greater hip adduction during stance, which is also consistent with Boyer et al.¹¹ Given the potential relationship between increased hip adduction and knee injury,³⁷ these results may be at odds with literature indicating that knee injuries are less frequent in higher-mileage runners. However, it is possible that higher-mileage runners also exhibit improved strength and coordination, which mitigates this risk.

Higher-mileage male runners exhibited a more flexed knee throughout stance, and during the later stages of swing in preparation for heel strike as compared to lower-mileage male runners. These results are in agreement with previous research that has shown that with training, runners adapt their gait patterns toward a more flexed knee during the stance phase of running gait and in preparation for toe off.³⁸ Greater knee flexion during stance may increase tibial shock and result in a higher risk for tibial stress injuries.^{39,40} Therefore, this running adaptation may increase the risk of experiencing lower leg injuries, which is a more common injury site for trained, competitive, long-distance runners compared to recreational runners.³²

Finally, higher- and lower-mileage male runners exhibited meaningful differences in transverse plane foot motion. Specifically, lower-mileage male runners demonstrated greater foot abduction (i.e., toe-out) throughout 95% of the gait cycle as compared to their higher-mileage counterparts. Increased foot abduction or greater toe-out during the stance phase of running has been associated with the presence of patellofemoral pain syndrome in runners.⁴¹ A greater toe out can increase the tibial external rotation, which may then lead to greater knee external rotation and result in patellofemoral pain due to repetitive loading.

4.4. Limitations

Limitations to the current study are acknowledged. First, the runners self-selected their preferred running speed, and there were significant differences in running speed in 2 of the subgroup comparisons. The differences in self-selected speed were likely due to differences in performance level and training habits between higher- and lower-mileage runners. Differences in speed have potential confounding effects on running biomechanics;⁴² however, due to the different training habits between these 2 groups, we chose a self-selected running speed, rather than a pre-defined speed, to better represent each runner's "typical" running patterns at their usual performance and

training level. Nevertheless, speed may be a confounding factor with regard to the kinematics of the female runners.

Secondly, a relatively small sample size was used in this study, especially when using machine learning methods to subgroup by gender (30 females and 11 males). When employing a machine learning approach and the ratio of the number of subjects to the number of samples is low, there is a risk of overfitting the data,²¹ which may reduce the generalizability of the results. To build more accurate classification models, a larger number of samples are needed in future studies.⁴³ Nevertheless, we are confident in our findings because similar differences were found with Cohen's effect size (d) and the features used in the SVM underwent a 10-fold cross-validated classification procedure. Although a larger sample size would have allowed for a more robust classification model, we believe our findings would be similar.

5. Conclusion

Using kinematic waveforms, PCA, and an SVM classifier, the current study was able to separate and classify lower- and higher-mileage runners based on differences in lower-limb kinematic gait patterns throughout stance and swing during treadmill running. The current study also demonstrated that there are distinct lower-limb kinematic differences between higher- and lower-mileage runners when further subgrouping by gender. A unique finding was that sagittal plane kinematics were an important consideration when subgrouping higher- and lower-mileage runners, especially in males. Our results suggest that gender needs to be considered to better understand the specific kinematic differences between higher- and lower-mileage runners. These findings may be used to monitor biomechanical adaptations associated with training that may help runners, coaches, and clinical professionals to assess performance and assess injury risk. For example, if a "lower-mileage" runner's gait pattern begins to resemble that of a "higher mileage" runner, then it may help predict improved running performance or protective mechanisms against running-related knee injuries. With the change in running biomechanics, however, this runner should then be aware of an increased risk of other types of running-related injuries (e.g., tibial stress fractures). On the other hand, if the "higher-mileage" runner's gait kinematics start to resemble that of a "lower-mileage" runner, then this may be predictive of an increased risk of running-related knee injuries. Finally, the findings of the current study also suggest that runners should employ individualized preventative strategies to reduce the risk of certain types of running-related injuries that are associated with the combination of running biomechanics, gender, and mileage.

Acknowledgments

Funding for this research was partially provided by a Discovery Grant (No. 1028495) and Accelerator Award (No. 1030390) through the Natural Sciences and Engineering Research Council of Canada (NSERC) along with the Faculty of Kinesiology Dean's Doctoral Studentship Program at the University of Calgary.

Authors' contributions

CAC drafted the manuscript and was involved in the study design and data analysis; AP helped draft the manuscript and was involved in the study design and data analysis; STO performed the data processing, helped with data analysis and interpretation, and was involved in critical review of the manuscript; RF was involved in the conception of the study, participated in its design and coordination, participated in data interpretation, and was involved in critical review of the manuscript. All authors have read and approved of the final version of the manuscript, and agree with the order of presentation of the authors.

Competing interests

The authors declare that they have no competing interests.

References

1. Saragiotto BT, Yamato TP, Lopes AD. What do recreational runners think about risk factors for running injuries? A descriptive study of their beliefs and opinions. *J Orthop Sports Phys Ther* 2014;**44**:733–8.
2. van Gent BR, Siem DD, van Middelkoop M, van Os TA, Bierma-zeinstra SSMA, Koes BBW. Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. *Br J Sports Med* 2007;**41**:469–80.
3. Taunton JE, Ryan MB, Clement DB, McKenzie DC, Lloyd-Smith DR, Zumbo BD. A retrospective case-control analysis of 2002 running injuries. *Br J Sports Med* 2002;**36**:95–101.
4. van der Worp MP, Ten Haaf DSM, van Cingel R, De Wijer A, Nijhuis-van Der Sanden MWG, Bart Staal J. Injuries in runners; a systematic review on risk factors and sex differences. *PLoS One* 2015;**10**:e0114937. doi:10.1371/journal.pone.0114937.
5. Hootman JM, Macera CA, Ainsworth BE, Martin M, Addy CL, Blair SN. Predictors of lower extremity injury among recreationally active adults. *Clin J Sport Med* 2002;**12**:99–106.
6. Messier S, Davis S, Curl W, Lowery R, Pack R. Etiologic factors associated with patellofemoral pain in runners. *Med Sci Sports Exerc* 1991;**23**:1008–15.
7. van Middelkoop M, Kolkman J, van Ochten J, Bierma-Zeinstra SMA, Koes BW. Risk factors for lower extremity injuries among male marathon runners. *Scand J Med Sci Sport* 2008;**18**:691–7.
8. Lo GH, Driban JB, Kriska AM, McAlindon TE, Souza RB, Petersen NJ, et al. History of running is not associated with higher risk of symptomatic knee osteoarthritis: a cross-sectional study from the osteoarthritis initiative. *Arthritis Care Res* 2016;**69**:183. doi:10.1002/acr.22939.
9. Silvernail JF, Boyer K, Rohr E, Bruggemann GP, Hamill J. Running mechanics and variability with aging. *Med Sci Sport Exerc* 2015;**47**:2175–80.
10. Wen D, Puffer J, Schmalzried T. Lower extremity alignment and risk of overuse injuries in runners. *Med Sci Sport Exerc* 1997;**29**:1291–8.
11. Boyer KA, Silvernail JF, Hamill J. The role of running mileage on coordination patterns in running. *J Appl Biomech* 2014;**30**:649–54.
12. Watari R, Kobsar D, Phinyomark A, Osis S, Ferber R. Determination of patellofemoral pain sub-groups and development of a method for predicting treatment outcome using running gait kinematics. *Clin Biomech (Bristol, Avon)* 2016;**38**:13–21.
13. Schmitz A, Pohl MB, Woods K, Noehren B. Variables during swing associated with decreased impact peak and loading rate in running. *J Biomech* 2014;**47**:32–8.
14. Zadpoor AA, Nikooyan AA. The relationship between lower-extremity stress fractures and the ground reaction force: a systematic review. *Clin Biomech (Bristol, Avon)* 2011;**26**:23–8.
15. Ferber R, Davis IM, Williams DS. Gender differences in lower extremity mechanics during running. *Clin Biomech (Bristol, Avon)* 2003;**18**:350–7.

16. Phinyomark A, Hettinga BA, Osis ST, Ferber R. Gender and age-related differences in bilateral lower extremity mechanics during treadmill running. *PLoS One* 2014;**9**:e105246. doi:10.1371/journal.pone.0105246
17. Gehring D, Mornieux G, Fleischmann J, Gollhofer A. Knee and hip joint biomechanics are gender-specific in runners with high mileage. *Int J Sports Med* 2014;**35**:153–8.
18. Phinyomark A, Osis S, Hettinga BA, Leigh R, Ferber R. Gender differences in gait kinematics in runners with iliotibial band syndrome. *Scand J Med Sci Sport* 2015;**25**:744–53.
19. Fukuchi RK, Eskofier BM, Duarte M, Ferber R. Support vector machines for detecting age-related changes in running kinematics. *J Biomech* 2011;**44**:540–2.
20. Clermont CA, Osis ST, Phinyomark A, Ferber R. Kinematic gait patterns in competitive and recreational runners. *J Appl Biomech* 2017;**2**:1–26.
21. Ferber R, Osis ST, Hicks JL, Delp SL. Gait biomechanics in the era of data science. *J Biomech* 2016;**49**:3759–61.
22. Nigg BM, Baltich J, Maurer C, Federolf P. Shoe midsole hardness, sex and age effects on lower extremity kinematics during running. *J Biomech* 2012;**45**:1692–7.
23. Diderisken M, Soegaard C, Nielsen RO. Validity of self-reported running distance. *J Strength Cond Res* 2016;**30**:1592–6.
24. Pohl MB, Lloyd C, Ferber R. Can the reliability of three-dimensional running kinematics be improved using functional joint methodology? *Gait Posture* 2010;**32**:559–63.
25. Osis ST, Hettinga BA, Macdonald S, Ferber R. Effects of simulated marker placement deviations on running kinematics and evaluation of a morphometric-based placement feedback method. *PLoS One* 2016;**11**:e0147111. doi:10.1371/journal.pone.0147111 e0147111.
26. Osis ST, Hettinga BA, Leitch J, Ferber R. Predicting timing of foot strike during running, independent of striking technique, using principal component analysis of joint angles. *J Biomech* 2014;**47**:2786–9.
27. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2010;**2**:433–59.
28. Phinyomark A, Osis ST, Hettinga BA, Kobsar D, Ferber R. Gender differences in gait kinematics for patients with knee osteoarthritis. *BMC Musculoskelet Disord* 2016;**17**:157. doi:10.1186/s12891-016-1013-z.
29. Knudson D. Significant and meaningful effects in sports biomechanics research. *Sport Biomech* 2009;**8**:96–104.
30. Cohen J. A power primer. *Psychol Bull* 1992;**112**:155–9.
31. Phinyomark A, Hettinga BA, Osis S, Ferber R. Do intermediate- and higher-order principal components contain useful information to detect subtle changes in lower extremity biomechanics during running? *Hum Mov Sci* 2015;**44**:91–101.
32. Kluitenberg B, van Middelkoop M, Diercks R, van der Worp H. What are the differences in injury proportions between different populations of runners? A systematic review and meta-analysis. *Sport Med* 2015;**45**:1143–61.
33. Ferber R, Cat C, Noehren ATCB, Hamill J, Davis PI. Competitive female runners with a history of iliotibial band syndrome demonstrate atypical hip and knee kinematics. *J Orthop Sports Phys Ther* 2010;**40**:52–8.
34. Preece SJ, Mason D, Bramah C. How do elite endurance runners alter movements of the spine and pelvis as running speed increases? *Gait Posture* 2016;**46**:132–4.
35. Dallam GM, Wilber RL, Jadelis K, Fletcher G, Romanov N. Effect of a global alteration of running technique on kinematics and economy. *J Sports Sci* 2005;**23**:757–64.
36. Dreyer D, Dreyer K. *ChiRunning: a revolutionary approach to effortless, injury-free running*. New York, NY: Simon & Schuster; 2008.
37. Powers CM. The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. *J Orthop Sports Phys Ther* 2010;**40**:42–51.
38. Moore IS. Is there an economical running technique? A review of modifiable biomechanical factors affecting running economy. *Sport Med* 2016;**46**:793–807.
39. Hamill J, Derrick TR, Holt KG. Shock attenuation and stride frequency during running. *Hum Mov Sci* 1995;**14**:45–60.
40. Greenberg ET, Greenberg S, Brown-Budde K. Biomechanics and gait analysis for stress fractures. In: Miller T, Kaeding CC, editors. *Stress fractures in athletes: diagnosis and management*. New York, NY: Springer; 2015.p.33–50.
41. Willson JD, Davis IS. Lower extremity mechanics of females with and without patellofemoral pain across activities with progressively greater task demands. *Clin Biomech (Bristol, Avon)* 2008;**23**:203–11.
42. Padulo J, Annino G, Migliaccio GM, D'ottavio S, Tihanyi J. Kinematics of running at different slopes and speeds. *J Strength Cond Res* 2012;**26**:1331–9.
43. Phinyomark A, Petri G, Ibáñez-Marcelo M, Osis ST, Ferber R. Analysis of big data in gait biomechanics: current trends and future directions. *J Med Biol Eng* 2017. doi:10.1007/s40846-017-0297-2.