Zeng *et al. Sports Medicine - Open* (2022) 8:86
https://doi.org/10.1186/s40798-022-00477-0

Sports Medicine - Open

SYSTEMATIC REVIEW

Open Access

Validity and Reliability of Inertial Measurement Units on Lower Extremity Kinematics During Running: A Systematic Review and Meta-Analysis

Ziwei Zeng, Yue Liu, Xiaoyue Hu, Meihua Tang and Lin Wang*

Abstract

Background: Inertial measurement units (IMUs) are useful in monitoring running and alerting running-related injuries in various sports settings. However, the quantitative summaries of the validity and reliability of the measurements from IMUs during running are still lacking. The purpose of this review was to investigate the concurrent validity and test–retest reliability of IMUs for measuring gait spatiotemporal outcomes and lower extremity kinematics of health adults during running.

Methods: PubMed, CINAHL, Embase, Scopus and Web of Science electronic databases were searched from inception until September 2021. The inclusion criteria were as follows: (1) evaluated the validity or reliability of measurements from IMUs, (2) measured specific kinematic outcomes, (3) compared measurements using IMUs with those obtained using reference systems, (4) collected data during running, (5) assessed human beings and (6) were published in English. Eligible articles were reviewed using a modified quality assessment. A meta-analysis was performed to assess the pooled correlation coefficients of validity and reliability.

Results: Twenty-five articles were included in the systematic review, and data from 12 were pooled for meta-analysis. The methodological quality of studies ranged from low to moderate. Concurrent validity is excellent for stride length (intraclass correlation coefficient (ICC) (95% confidence interval (CI)) = 0.937 (0.859, 0.972), $p < 0.001$), step frequency (ICC (95% CI) = 0.926 (0.896, 0.948), r (95% CI) = 0.989 (0.957, 0.997), $p < 0.001$) and ankle angle in the sagittal plane (r (95% CI) = 0.939 (0.544, 0.993), $p = 0.002$), moderate to excellent for stance time (ICC (95% CI) = 0.664 (0.354, 0.845), r (95% CI) = 0.811 (0.701, 0.881), $p < 0.001$) and good for running speed (ICC (95% CI) = 0.848 (0.523, 0.958), $p = 0.0003$). The summary Fisher's Z value of flight time was not statistically significant ($p = 0.13$). Similarly, the stance time showed excellent test–retest reliability (ICC (95% CI) = 0.954 (0.903, 0.978), $p < 0.001$) and step frequency showed good test–retest reliability (ICC (95% CI) = 0.896 (0.837, 0.933), $p < 0.001$).

Conclusions: Findings in the current review support IMUs measurement of running gait spatiotemporal parameters, but IMUs measurement of running kinematics on lower extremity joints needs to be reported with caution in healthy adults.

Trial Registration: PROSPERO Registration Number: CRD42021279395.

Keywords: Inertial measurement unit, Kinematics, Running, Validity, Reliability

---

*Correspondence: wanglin@sus.edu.cn

School of Kinesiology, Shanghai University of Sport, Shanghai, China

Key Points

- IMUs are reliable tools for measuring gait spatiotemporal parameters during running in healthy adults but should be reported with caution for lower extremity joint kinematics.
- Future studies need to include more subjects and use more rigorous protocols to provide evidence that supports the use of IMUs in the prevention of running-related injuries.
- Guidelines for applying IMUs for running kinematic measurement need to be established.

Introduction

Running is one of the popular physical activities around the world and has positive effects on both physical and mental health [1, 2]. Unfortunately, overuse due to the increased frequency and volume of running is the main mechanism for the occurrence of running-related injuries (RRIs), particularly in the lower limbs [3–7]. Therefore, a thorough understanding of biomechanical changes in the lower limbs during running is of great importance to the prevention of RRIs.

As a portable alternative to optical motion capture systems, inertial sensors are becoming increasingly popular in many fields, including sports science, owing to their low cost, portability, lightness and unlimited research environment [8–11]. Inertial sensors usually include accelerometers, gyroscopes or magnetometers, also known as inertial measurement units (IMUs), which measure gravitational acceleration, angular velocity and heading in the Earth's magnetic field, respectively [12, 13].

Along with the increasing popularity of IMUs, the number of studies examining their validity and reliability for a variety of populations (e.g., healthy people, multiple sclerosis and stroke patients) during different movements (e.g., walking, running and jumping) has increased [14–16]. Recent systematic reviews have examined the validity and reliability of measurements from IMUs of walking in healthy adults [17] and the impact of IMUs' position on the validity and reliability of stride variables during running [18]. However, three-dimensions kinematics data for the validity and reliability of measurements from IMUs during running have not been synthesized and quantified. Meanwhile, limiting the study population to healthy adults may render the findings more homogeneous. Therefore, the aim of the current systematic review and meta-analysis was to determine the concurrent validity and test–retest reliability of IMUs for measuring gait spatiotemporal and lower-extremity kinematics outcomes during running in healthy adults.

Methods

The protocol was registered on the International Prospective Register of Systematic Reviews (PROSPERO) (Registration number: CRD42021279395) and followed the Preferred Reporting Items for Systemic Reviews and Meta-Analysis (PRISMA) guidelines [19].

Search Strategy

PubMed, CINAHL, Embase, Scopus and Web of Science electronic databases were searched from inception until 27 September 2021. The search terms and strategies included: (wearable sensor* OR inertial sensor* OR inertial motion capture OR "Wearable Electronic Devices"[Mesh] OR inertial measurement unit* OR IMU OR "Micro-Electrical–Mechanical Systems"[Mesh] OR MEMS OR acceleromet* OR gyroscop* OR magnetomet* OR smart phone OR "Smartphone"[Mesh]) AND (running speed OR cadence OR (step frequency) OR (stride frequency) OR (step time) OR (stride time) OR (cycle time) OR (contact time) OR (swing time) OR (flight time) OR (step length) OR (stride length) OR spatiotemporal OR "Spatio-Temporal Analysis"[Mesh] OR kinematic* OR biomechanic* OR (joint angle) OR hip OR knee OR ankle OR range of motion OR "Range of Motion, Articular"[Mesh]) AND (running OR jogging OR sprinting) AND (validity OR reliability OR feasibility OR repeatability OR consistency OR "Reproducibility of Results"[Mesh] OR "Data Accuracy"[Mesh]). Minor adjustments were made for different databases. Full search strategies for each database can be found in Additional File 1.

Inclusion and Exclusion Criteria

Articles that met the following criteria were included in this systematic review: (a) evaluated the validity or reliability of IMUs, (b) measured specific gait spatiotemporal and lower extremity kinematics parameters, (c) compared the measurements captured by IMUs with those obtained using reference systems, (d) collected data during running, jogging or sprinting, (e) assessed human beings and (f) were published in English. Any studies that only measured activity/movement identification or energy expenditure were excluded from this review. Additional details on the inclusion and exclusion criteria and definitions for the spatiotemporal parameters can be found in Additional File 2.

Study Selection

After duplicate articles were removed, two independent reviewers (Zeng and Liu) screened the titles and abstracts according to the eligibility criteria. The full-text screening of the potentially eligible articles was examined by one author (Zeng) and rechecked by a second author (Hu). All reference lists and bibliographies of the retrieved studies were reviewed in case relevant studies were missed by the electronic search. Disagreements were discussed and resolved by a third reviewer (Wang).

Assessment of Risk of Bias

Assessment of risk of bias was assessed using a modified version of the Critical Appraisal of Study Design for Psychometric Articles [20], which was adjusted by Kobsar et al. [17] to specifically evaluate the psychometric properties of studies about inertial sensors. This checklist contains 12 items, which assess the methodological quality of five domains, namely, study question, study design, measurements, analyses and recommendations [17]. Each item comprises three descriptors. The maximum score is 24 and the final total score and percentage will be presented. Initially, two assessors (Tang and Liu) reviewed two articles at the same time, and then a consensus on the scoring and interpretation of each item was performed before the remaining articles were evaluated separately. The process described above in case of disagreement was used. Assessors were blinded to any identifiable information related to the studies to avoid bias in quality assessment. Furthermore, agreement between the two assessors was calculated using the Cohen's kappa coefficient with a 95% confidence interval (95% CI) [21]. Cohen's kappa coefficient of <0.40, 0.40–0.75 or >0.75 were regard as poor, fair to good or excellent, respectively [22].

To grade the quality of the study, a previously described classification scheme was applied (Table 1) [17]. Quality assessment scoring was then used in determining the strength of recommendations [23].

Table 1 Study Methodological quality grading scheme [17]

Level	Score
High quality (HQ)	85–100%
Moderate quality (MQ)	70–85%
Low quality (LQ)	50–70%
Very low quality (VLQ)	< 50%

Data Extraction

Data extraction was completed by two authors (Zeng and Tang) using a pre-defined form. The data consisted of (1) study identification information; (2) participant characteristics: sample size, sex, age, height, weight and recruited population; (3) IMUs' specifications: name, manufacturer, composition, used number, placement and sample frequency; (4) reference systems used; (5) study design: running speed/running distance and research field; (6) specific parameters; and (7) reported statistical outcomes.

For validity, statistical outcomes extracted were Pearson correlation coefficient (*r*), coefficient of determination (r^2), coefficient of multiple correlation (CMC), concordance correlation coefficient (CCC) and intraclass correlation coefficient (ICC) with 95% CI, root mean square error (RMSE; measurement error between the IMUs and reference systems), bias (mean difference between the IMUs and reference systems) and limits of agreement (LoA; 1.96*standard deviation of the difference between the IMUs and reference systems). For reliability, the statistical outcomes extracted were ICC (95% CI), RMSE, coefficient of variation (CV; the ratio of the standard deviation to the mean), and standard error of the mean. It should be mentioned that ICCs were not included in this review if they were only shown graphically without specific values and all differences were unified as the IMUs' measurement minus the reference systems' measurement if mentioned in the articles. While these statistical outcomes were extracted for the qualitative assessments, data pooling was a priori set to assess only the ICCs, *r* and sample size for validity and ICCs and sample size for reliability.

Statistical Analysis

In data pooling, validity and reliability were first dichotomised. Then, a further division was made among specific parameters. Given that reported ICC and *r* values need to be classified, a single study may contribute to multiple independent data pooling based on validity, reported statistical outcomes and parameter measured. The agreement metrics of ICCs were interpreted as poor (<0.500), moderate (0.500–0.749), good (0.750–0.899) or excellent (≥ 0.900) [24], and *r* was interpreted as no correlation (<0.250), fair relationship (0.250–0.500), moderate to good relationship (0.500–0.750) or good to excellent relationship (≥ 0.750) [25].

Review Manager (RevMan 5.3) was used in the meta-analysis. Heterogeneity was examined using Tau^2, Chi^2 and I^2 statistics where $\text{Tau}^2 = 0$ suggests no heterogeneity; I^2 values of $<25\%$, 26%–50% and $>75\%$ suggest low, moderate and high heterogeneity, respectively, and a significant Chi^2 indicates heterogeneity [26]. The level of significance was $P < 0.05$. Due to the heterogeneity of the experimental conditions and population, a random effects model was used with 95% CI [27]. When the number of studies is sufficient ($n \geq 3$), subgroup analyses were conducted to explore the potential sources of heterogeneity. Subgroup were defined a priori and included running speed, IMUs' position and running surface. The running speed was set to two levels: low (speed ≤ 15 km/h) and fast (speed > 15 km/h), and the running surface was divided into treadmill and ground. Sensitivity analyses were performed by deleting one study at a time to evaluate the stability of the results [28].

Weighting of individual point estimates was based on sample size. Given the non-normality of ICCs and *r*, point estimates were variance-stabilised using Fisher's z-transformation as follows [29]:

$$\text{Fisher}'\text{s}\ \text{Z}_{\text{ICC}} = 0.5 \times \ln\frac{1+\text{ICC}}{1-\text{ICC}} \tag{1}$$

$$\text{Fisher}'\text{s}\ \text{Z}_{r} = 0.5 \times \ln\frac{1+r}{1-r} \tag{2}$$

$$v_z = \frac{1}{n-3} \tag{3}$$

$$\text{SE}_{\text{ICC}} = \frac{1}{\sqrt{n-3/2}} \tag{4}$$

$$\text{SE}_r = \sqrt{v_z} \tag{5}$$

$$\text{Summary ICC}/r = \frac{\text{e}^{2\text{Z}}-1}{\text{e}^{2\text{Z}}+1} \tag{6}$$

where n represents sample sizes, SE depicts standard error and Z is Summary Fisher's Z value [30]. Data were then transformed back to ICCs or *r* for reporting. The results of the meta-analysis were interpreted using the same agreement metrics outlined above.

Statistical results that were not included in the quantitative analysis were included in the qualitative analysis to support the interpretation. An adapted rating system from the Cochrane collaboration back review group [23] was used in determining the level of evidence for each parameter (Table 2) [17, 31].

Table 2 Definitions of levels of evidence [17]

Level of evidence	Criteria
Strong evidence	Consistent results in HQ studies ($n \geq 2$)
Moderate evidence	Consistent results among multiple MQ studies ($n \geq 2$)
Limited evidence	Consistent results among multiple LQ studies ($n \geq 2$)
Conflicting evidence	Inconsistent results among multiple studies
Very limited evidence	Only one LQ or MQ study or multiple VLQ studies

HQ high-quality, *MQ* moderate-quality, *LQ* low-quality, *VLQ* very low quality

Results

Characteristics of the Included Studies

A total of 2316 articles were identified through database screening and cross-referencing. After the removal of duplicates, screening of titles and abstracts, and full-text screening, 25 studies met the eligibility criteria and were included in this systematic review [12, 32–55]. An outline of the screening process using the PRISMA flow diagram is presented in Fig. 1.

A summary of the characteristics of the included 25 studies is presented in Table 3. A total of 541 healthy adults (mean [sd] sample size: 22 [12] participants; range: 7–51 participants; 151 females and 354 males; sex was not described in Fox et al. [41] and Li et al. [48]) were included in this review. In terms of the population, it has been divided it into four categories, namely competitive runners ($n = 111$) [32, 33, 36, 37, 51, 52], experienced runners ($n = 28$) [43, 53], amateur runners ($n = 200$) [33, 41, 42, 45, 47, 54, 55] and non-runners ($n = 202$) [12, 34, 35, 38–40, 44, 46, 48–50]. The most common IMU systems used were the Xsens system ($n = 3$) [12, 34, 53] and RunScribe™ system ($n = 3$) [38, 42, 45]. Using two ($n = 9$) [32, 35, 40–42, 44, 45, 51, 54] or one ($n = 8$) [33, 34, 36, 37, 43, 46, 47, 52] IMU was the most preferable, and some studies used five ($n = 1$) [49], seven ($n = 3$) [39, 48, 50], eight ($n = 1$) [50] or seventeen ($n = 2$) [12, 53] IMUs. In addition, studies installed IMUs in diverse sites, including dorsum of the foot [12, 32, 34, 38–40, 42, 44, 45, 49, 50, 53, 55], ankle [38, 48, 51, 55], heel [38, 47, 55], shank [12, 35, 38, 39, 44, 49, 50, 53], knee [48], thigh [12, 35, 39, 50, 53], hip [46, 48], waist [36, 37, 43, 53], sacrum [12, 49, 50], chest [38, 41], sternum [12, 52, 53], back [33, 39, 41], upper arm [12, 53], lower arm [12, 53], hand [12, 53], shoulder [12, 53], head [12, 53] and shoes midsole [54, 55]. The most common sampling frequencies used in assessing running were 200 Hz ($n = 6$) [33, 41, 45, 49, 54, 55] and 500 Hz [36, 37, 40, 42, 43] ($n = 5$; range: 50–1000 Hz). For the sports settings, the present study included running on an indoor track or walkway, running on a treadmill and running outside, and running speed ranged from 7.2 km/h to 21 km/h.

Risk of Bias of the Included Studies

No articles were rated as HQ or VLQ, 11 as MQ and 14 as LQ (Table 4). Agreement between both assessors was good (Cohen's kappa = 0.75; 95% CI = 0.68–0.82). The items for which articles generally scored higher were '1- Background and research question', '4- Study design', and '12- Conclusion'. By contrast, only two studies (8%) provided justification about their sample sizes.

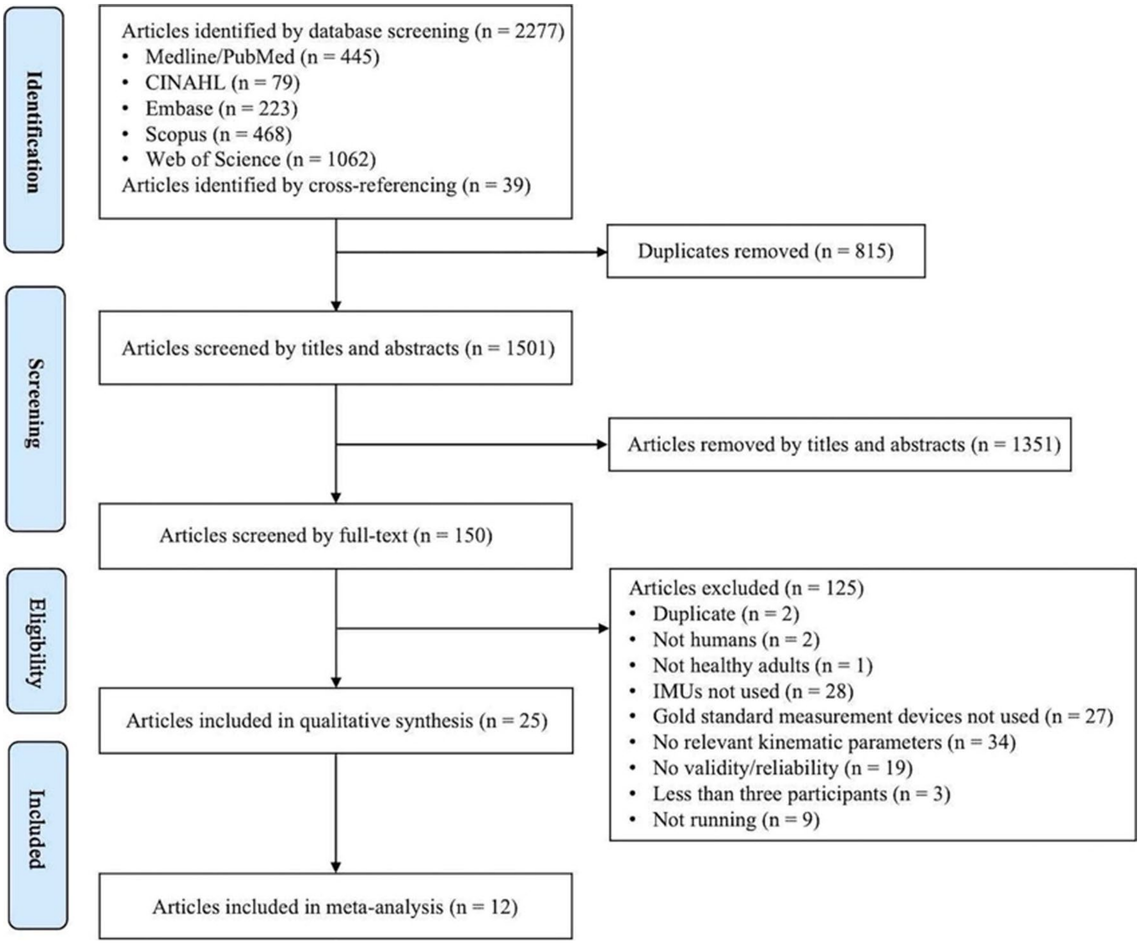


Fig. 1 Flowchart of the systematic review selection process.

Synthesis of Results

Validity

Validity was assessed using optical motion capture system ($n = 18$), instrumented treadmill ($n = 7$), force plate ($n = 3$), timing light system ($n = 1$) and photocell system ($n = 3$) as criteria. Overall, nine gait spatiotemporal and 31 lower extremity joint kinematics parameters were assessed across the 25 studies that examined the validity of IMUs. From these outcomes, one joint kinematics and five gait spatiotemporal parameters presented sufficient study quality and statistical outcomes for data pooling (Figs. 2, 3, 4, 5, 6, 7, 8). Meta-analysis was not possible on other outcomes because of the limited number of studies or the lack of consistency in data reporting, as many studies reported only RMSE or bias. Studies that were not included in the meta-analysis were qualitatively summarised according to outcomes in Additional File 3.

Quantitative Pooling for Validity *Stance time* Data from four MQ and one LQ studies suggested that the validity for stance time derived from IMUs, as reported by ICCs, was moderate (ICC (95% CI) = 0.664 (0.354, 0.845), $I^2 = 95\%$, $p = 0.0003$) (Fig. 2) [32, 37, 42, 43, 45], but pooled r values from one MQ and two LQ studies indicated validity for stance time from IMUs was excellent (r (95% CI) = 0.811 (0.701, 0.881), $I^2 = 99\%$, $p < 0.001$) (Fig. 3) [36, 42, 49]. The validity of stance time reported by r values can only be analysed for subgroups based on running surface due to variable running speed and the involvement of multiple attachment locations of IMUs. Subgroup analysis showed no significant effect of running speed on the validity for stance time derived from IMUs ($p = 0.54$), while IMUs at the shoe (ICC (95% CI) = 0.929 (0.869, 0.961), $I^2 = 71\%$) showed higher agreement compared to at the waist (ICC (95% CI) = 0.226 (−0.282, 0.641), $I^2 = 94\%$) ($p < 0.001$) (Fig. 2). The validity reported via ICC and r values did not differ significantly between the two running surfaces ($p \geq 0.05$) (Figs. 2 and 3). Sensitivity analysis showed that the results were stable.

Table 3 Study characteristics

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
Ammann et al., 2016 [32]	12 subjects (5 F, 7 M; age: 25.3 ± 3.2 years; height: 174.4 ± 7.9 cm; weight: 64.8 ± 10.2 kg) High-level running athletes	PARTwear (PW, HuCE-microLab, University of Applied Sciences, Biel, Switzerland)	3D accelerometer (± 16 g); 3D gyroscope; 3D magnetometer	2	Lace of the shoe	1000 Hz	OMC (Camera Marathon Ultra CL600, Videal AG, Niederonz, Switzerland)	Maximal sprinting speed (8.0 ± 0.5 m/s); intense training speed (6.2 ± 0.7 m/s); normal training speed (4.3 ± 0.7 m/s); all speeds (6.2 ± 1.6 m/s) (40 m)	Indoor track	Stance time
Bergamini et al., 2012 [33]	Group A: 6 amateur athletes (2F, 4 M; height: 172 ± 12 cm; weight: 63.50 ± 10.84 kg) Group E: 5 elite athletes (2F, 3 M; height: 177 ± 76 cm; weight: 65.00 ± 8.25 kg) Total: 11 participants (4F, 7 M; height: 174 ± 10 cm; weight: 64.18 ± 9.78 kg)	IMU (FreeSense, Sensorize, Italy)	3D accelerometer (± 6 g); 3D gyroscope (± 500°/s)	1	Lower back trunk (L1 level)	200 HZ	OMC (Casio Exilim EX-F1, Japan); 9 force platforms (Z20740AA, Kistler, Switzerland)	Three sprint runs of 60 m	Indoor track; outdoor training track	Stance time, stride time
Brahms et al., 2018 [34]	11 healthy young adults (4F, 7 M; age: 22.3 ± 1.5 years; height: 175.2 ± 23.1 cm; weight: 76.04 ± 3.19 kg)	Xsens (MTw)	–	1	Right foot	100 Hz	OMC (Vicon)	A range of typical distance running speeds/10 m trial runs (3.55 ± 0.34 m/s, range: 2.71–4.36 m/s)	Indoor	Stride length

Table 3 (continued)

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
Cooper et al., 2009 [35]	7 subjects (2F, 5 M; age: 30 ± 6 years; height: 170 ± 20 cm; weight: 70 ± 11 kg)	IMU (ETB Ltd, Codicote, UK)	3D accelerometer (± 5 g); 3D gyroscopes (± 1200°/s)	2	Thigh and shank	100 Hz	OMC (Qualysis)	5 mile/hour	Treadmill	Knee joint flexion/extension angles
Day et al., 2021 [36]	30 subjects (21F; weight: 54.0 ± 5.3 kg; 9 M; weight: 63.6 ± 6.7 kg) National Collegiate Athletic Association Division 1 cross country runners	IMU (IMeasureU, Auckland, New Zealand)	–	1	Over waistband	500 Hz	Instrumented treadmill (Treadmetrix, Park City, UT/Bertec, Columbus, OH)	M: 3.8, 4.1 and 5.4 m/s; F: 3.8 and 4.9 m/s	Treadmill	Stance time
Deflandre et al., 2018 [37]	Reliability: 10 young male athletes (age: 14 ± 0.5 years; height: 168 ± 7 cm; weight: 56 ± 9 kg) Validity: 20 male athletes (age: 32 ± 14 years; height: 181 ± 16 cm; weight: 71 ± 7 kg)	Myotest (Myotest SA, Sion, Switzerland)	3D accelerometer	1	The iliac crests mark, in the alignment with the umbilicus	Hz	The Optogait (Microgate, Bolzano, Italy); 3D optoelectronic CX1 units (Codamotion, Charnwood Dynamics Ltd, UK)	8 km/h, 12 km/h and 16 km/h (80 m); 8 km/h and 16 km/h	Outside: artificial turf field; indoor: treadmill	Stance time, step length, stride length, step frequency

Table 3 (continued)

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
De Fontenay et al., 2020 [38]	32 healthy participants (13F, 19 M; age: 27.0 ± 5.5 years; height: 174.4 ± 8.5 cm; weight: 69.1 ± 11.4 kg) 32 participants were analyzed for RunScribeTM, 31 participants were analyzed for MilestonePod and TgForce, 30 for Zoi, and 25 for Moov NowTM	Moov Now™ (Moov, San Mateo, California, USA); MilestonePod (Milestone Sports, Long beach, California, USA); RunScribe™ (Montara, California, USA); Zoi (Runteq, Tampere, Finland); TgForce (Kelsec Systems Inc.,Montréal, Canada)	–	–	Moov Now™: outside of the ankle and the loop end of the band forward; MilestonePod: shoelaces; RunScribe™: heel mount; Zoi: chest strap and shoelaces; TgForce: medial end of tibia	–	OMC (Vicon); Instrumented treadmill with force plates (Bertec, Columbus, OH, USA)	-	Treadmill	Step frequency
Dorschky et al., 2019 [39]	10 healthy male subjects (age: 27.1 ± 2.6 years; height: 182 ± 5 cm; weight: 76.9 ± 8.6 kg)	Custom-built IMUs (Portabiles GmbH, Erlangen, DE)	3D accelerometers (± 16 g); Gyroscopes (± 2000°/s)	7	Lower back, right and left lateral thigh, lateral shank, and upper midfoot	1000 Hz	OMC (Vicon MX, Oxford, UK); Force plate (Kistler Instruments Corp, Winterhur, CH)	3.0–4.9 m/s	Indoor	Hip, knee, and ankle angles in the sagittal plane
Falbriard et al., 2018 [40]	41 healthy adults (13F, 28 M; age: 29 ± 6 years; height 174 ± 8 cm; weight 70 ± 10 kg)	IMU (Physilog 4, Gait Up, Switzerland)	3D accelerometer (± 16 g); 3D gyroscope (± 2000 o/s)	2	Dorsum of each foot	500 Hz	Instrumented treadmill (T-170-FMT, Arsalis, Belgium)	Starting at 8 km/h and increasing by 2 km/h up to maximum speed	Treadmill	Stance time, flight time, swing time, step time
Fox et al., 2019 [41]	26 recreationally active participants (age: 32.2 ± 11.0 years; height: 173.3 ± 9.9 cm; weight: 74.2 ± 16.2 kg)	Polar Team Pro Sensor (Polar Electro, Kempele, Finland)	GPS; accelerometer; gyroscope; digital compass	2	Upper-torso between the scapulae and the centre of the chest at the level of the xiphoid process	200 Hz	Electronic timing light (Fusion Sport, Coopers Plains, QLD, Australia)	Medium (moderate, jogging) speed; high (maximal, sprint) speed	Indoor (sprung hardwood floor)	Running speed

Table 3 (continued)

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
García-Pinillos et al., 2019 [42]	49 amateur endurance runners (5F, 44 M; age: 26 ± 8 years; height: 174 ± 7 cm; weight: 71 ± 10 kg)	Stryd™ (Stryd Powermeter, Stryd Inc. Boulder CO, USA); RunScribe™ (Scribe Lab. Inc. San Francisco CA, USA)	Stryd™: 3D gyroscope; 3D accelerometer RunScribe™: 3D gyroscope; 3D accelerometer; 3D magnetometer	2	Lace shoe of the right leg	RunScribe™: 500 Hz	OMC (Imaging Source DFK 33UX174, The Imaging Source Europe GmbH; Germany)	Self-selected comfortable running velocity: 3.25 ± 0.36 m/s	Treadmill	Stance time, flight time, step length, step frequency
Gindre et al., 2016 [43]	20 habitual male runners (age: 31.6 ± 9.2 years, height: 178 ± 5.4 cm, weight: 72.5 ± 9.8 kg)	Myotest®	–	1	Around the waist of participants	500 Hz	The Optojump Next®; OMC (Casio High Speed EXILIM EX-FH25®, CASIO Europe GmbH, Norderstedt, Germany)	12 m/h; 15 km/h; 18 km/h; 21 km/h (60 m)	Indoor	Stance time, aerial time, step frequency
Kim et al., 2021 [44]	10 healthy male participants (age: 30.2 ± 5.3 years, height: 171 ± 15.3 cm, weight: 73.6 ± 12.4 kg	Adafruit BNO055 IMU sensors (Adafruit, New York, NY, USA)	–	2	Top of the instep of the right foot, and the right shin	100 Hz	OMC (VICON, Oxford, UK)	2.68 m/s	Indoor	Ankle dorsiflexion/plantarflexion and eversion/inversion angle
Koldenhoven and Hertel, 2018 [45]	12 recreational runners (6F, 6 M; age: 23.1 ± 5.5 years)	RunScribe™ (Scribe Labs, Inc., Half Moon Bay, CA, USA)	3D accelerometer; gyroscope	2	The back of each shoe	Hz	OMC (Vicon Motion Systems, Inc., Lake Forest, CA, USA); Instrumented treadmill (Bertec, Columbus, OH, USA)	Preferred speed: 2.7 ± 0.1 m/s; 1.5 miles	Treadmill	Stance time, cycle time, stride length, running speed

Table 3 (continued)

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
Konharn et al., 2016 [46]	30 normal-weight participants (15F, 15 M; age: 21.7 ± 1.0 years; height: 163.3 ± 19.5 cm; weight: 59.4 ± 8.5 kg)	Apps (Runtastic pedometer, Footsteps pedometer, and Walker pedometer) were downloaded into iPhone5 (iOS 7.0.3, Apple, Inc, CA, USA)	–	1	Right hip at the midline	–	The OZ1 Marathon treadmill (Marathon (Thailand) Co., Ltd., Bangkok, Thailand); The HJ-203 Omron pedometer (Omron Healthcare, Co., Ltd., Kyoto, Japan)	Moderate: 6.4 km/h; vigorous: 8 km/h	Treadmill	Running speed
Koska et al., 2018 [47]	51 recreational runners (15F, 36 M; 33.9 ± 8.2 years, height: 177.9 ± 7.6 cm; weight: 70.9 ± 10.1 kg)	IMU (aims®, Xybermind, Tübingen, Germany)	3D accelerometer (± 16 g); gyroscope (± 2000°/s)	1	The heel cup of the right shoe	400 Hz	OMC (Qualisys, Gothenburg, Sweden)	10 m/h; 12 km/h; 15 km/h	Treadmill	Rearfoot sagittal/eversion ROM
Li et al., 2020 [48]	10 healthy subjects (age > 18 years; height: 170 ± 10 cm; weight: 75 ± 10 kg)	Microelectromechanical system (MEMS) IMUs	-	7	The left side of the waist, bilateral knees above and below, bilateral feet	50 Hz	NDI system (NDI, Ontario, Canada)	5.1 km/h	Indoor	Hip, knee, and ankle angles in the sagittal plane
Mavor et al., 2020 [12]	20 civilian participants (10F, 10 M; age: 23.7 ± 3.44 years; height: 175 ± 7.93 cm; 71.9 ± 13.2 kg)	IMU (MVN BIOMECH, Xsens, Enschede, the Netherlands)	–	17	The back of the head, sacrum, sternum, and bilaterally on the upper arms, forearms, hands, shoulders, thighs, shanks, and feet	240 Hz	OMC (Vantage 5, Vicon, Oxford, UK)	-	Indoor	Hip, knee, and ankle flexion–extension, ab/adduction and axial rotation angle

Table 3 (continued)

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
Mo and Chow, 2018 [49]	11 healthy volunteers (4F, 7 M; age: 25.5 ± 4.2 years; height: 168.3 ± 9.1 cm; weight: 58.8 ± 5.3 kg)	IMU system (MyoMOTION MR3, Noraxon, USA)	3D accelerometer (± 16 g)	5	Sacrum (L5-S1), shanks (anteromedial distal aspect of the tibia) and feet (the dorsal surface of the shoe)	200 Hz	Force-platforms (Bertec, FP4060-07, USA)	Jog (3.1 ± 0.1 m/s); run (4.1 ± 1.2 m/s)	Indoor: 10 m walkway	Stance time
Nüesch et al., 2017 [50]	20 healthy subjects (12F, 8 M; age: 27.4 ± 8.3 years; height: 175 ± 8 cm; weight: 66.5 ± 12.5 kg)	IMU (RehaGait®, Hasomed, Magdeburg, Germany)	3D accelerometer (± 16 g); 3D gyroscope (± 2000°/s); 3D magnetometer (± 1.3 Gs)	7	The sacrum and bilaterally on the lateral thigh (middle), lateral shank (lower third), and lateral foot (on the shoe, below lateral malleolus)	400 Hz	OMC (Vicon MX, Vicon Motion Systems Ltd., Oxford, UK)	Self-selected running speed (2.93 ± 0.35 m/s)	Treadmill	Hip, knee, and ankle angles in the sagittal plane/ at initial contact; maximal/minimal ankle angle; ankle dorsiflexion/ plantarflexion ROM; hip and knee ROM (first and second half stride)
Schmidt et al., 2016 [51]	12 track and field athletes (2F, 10 M)	IMU (MPU-9150) from InvenSense	3D accelerometer (16 bit and ± 16 g range); 3D gyroscope (16 bit and ± 1000°/s); 3D magnet field sensor	2	Ankles	1000 Hz	OptojumpNext photocell system (Microgate, Bolzano, Italy / OJ)	Maximal sprints on a 60 m track	Track	Stance time
Watari et al., 2016 [52]	22 semi-elite runners (8F, 14 M; age: 28.2 ± 10.1 years; height: 173 ± 75 cm; weight: 65.4 ± 8.1 kg)	Built-in accelerometer (Forerunner 620, Garmin International Inc., Olathe, KS)	–	1	Torso of the runner, near the xiphoid process of the sternum	–	Instrumented treadmill (Bertec, Columbus, OH); OMC (Vicon Motion System, Vicon MX3, Oxford, UK) motion capture system	2.7 m/s; 3.0 m/s; 3.3 m/s; 3.6 m/s; 3.9 m/s	Treadmill	Stance time

Table 3 (continued)

Author(s), Year [Reference No.]	Participant (size, age, height, weight, population)	IMUs					Reference system	Running speed/running distance	Research field	Parameters
		Name (manufacturer)	Composition	Number	Placement	Sample frequency				
Wouda et al., 2018 [53]	8 healthy experienced male runners (age: 25.1 ± 5.2 years; height: 183.7 ± 4.5 cm; weight: 77.7 ± 9.4 kg)	Xsens MVN Link inertial motion capture system (Xsens, Enschede, the Netherlands)	–	17	Both shoulders, upper arms, lower arms, hands, upper legs, lower legs, feet, head, sternum, and pelvis	240 Hz	OMC (Nexus 1.8.5, Vicon, Oxford, UK); S-Mill instrumented treadmill (ForceLink, Culemborg, the Netherlands)	10 m/h; 12 km/h; 14 km/h	Treadmill	Maximum knee flexion/extension angle during Stance
Zrenner et al., 2018 [54]	27 amateur runners (6F, 21 M; age: 24.9 ± 2.4 years; height: 178.6 ± 8.0 cm)	miPod IMU sensor	Accelerometer (± 16 g); gyroscope (± 2000 o/s)	2	A cavity in the right and left shoes midsole	200 Hz	OMC (Vicon Motion Systems Inc., Oxford, UK)	2–6 m/s	Indoor	Stride length, running speed
Zrenner et al., 2020 [55]	29 amateur runners (6F, 23 M; age: 24.9 ± 2.4 years)	miPod IMU sensors	Accelerometer (± 16 g); gyroscope (± 2000 o/s)	8	A cavity in the sole of the running shoe, laterally under the ankle, at the heel, and on the instep	200 Hz	OMC (Vicon Motion Systems Inc., Oxford, UK)	2–6 m/s	Indoor	Stance time, stride time, stride length, running speed, rearfoot ROM in the frontal plane

F female, *M* male; *IMU* inertial measurement unit, *OMC* optical motion capture system, *ROM* range of motion

Table 4 Quality assessment scoring of 25 included studies

Author(s), Year [Reference No.]	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Total	%	Quality
Ammann et al., 2016 [32]	2	2	1	2	0	N	1	2	1	2	2	2	17/24	70.8%	MQ
Bergamini et al., 2012 [33]	2	1	0	2	1	N	1	2	1	1	1	1	13/24	54.2%	LQ
Brahms et al., 2018 [34]	2	2	2	2	1	N	1	2	1	2	2	2	19/24	79.2%	MQ
Cooper et al., 2009 [35]	2	1	1	2	0	N	2	2	1	0	1	1	13/24	54.2%	LQ
Day et al., 2021 [36]	1	2	0	2	1	N	1	2	1	1	1	2	14/24	58.3%	LQ
Deflandre et al., 2018 [37]	2	2	2	2	1	0	1	1	2	1	2	2	18/24	75.0%	MQ
De Fontenay et al., 2020 [38]	2	2	2	2	1	2	1	0	2	2	1	2	19/24	79.2%	MQ
Dorschky et al., 2019 [39]	2	1	0	2	1	N	2	2	1	1	1	2	15/24	62.5%	LQ
Falbriard et al., 2018 [40]	2	2	0	2	1	1	2	2	2	1	1	2	18/24	75.0%	MQ
Fox et al., 2019 [41]	2	1	0	2	2	N	1	1	1	2	2	2	16/24	66.7%	LQ
García-Pinillos et al., 2019 [42]	2	2	1	2	1	N	1	1	1	2	2	2	17/24	70.8%	MQ
Gindre et al., 2016 [43]	2	2	2	2	1	N	1	2	2	1	1	2	18/24	75.0%	MQ
Kim et al., 2021 [44]	2	2	0	2	1	N	2	2	2	1	1	2	17/24	70.8%	MQ
Koldenhoven and Hertel, 2018 [45]	1	1	2	2	1	N	1	2	1	2	1	2	16/24	66.7%	LQ
Konharn et al., 2016 [46]	2	2	0	2	1	N	1	1	1	1	2	2	15/24	62.5%	LQ
Koska et al., 2018 [47]	2	1	0	2	1	N	1	2	1	2	1	2	15/24	62.5%	LQ
Li et al., 2020 [48]	2	1	0	2	1	N	2	1	2	2	1	2	16/24	66.7%	LQ
Mavor et al., 2020 [12]	2	1	0	2	1	2	2	2	2	0	1	2	17/24	70.8%	MQ
Mo and Chow, 2018 [49]	2	0	0	2	0	N	2	2	1	1	1	2	13/24	54.2%	LQ
Nüesch et al., 2017 [50]	2	2	2	2	1	N	2	1	2	2	2	2	20/24	83.3%	MQ
Schmidt et al., 2016 [51]	1	0	0	2	1	N	1	1	1	2	1	2	12/24	50.0%	LQ
Watari et al., 2016 [52]	1	2	0	2	2	N	1	1	1	2	1	2	15/24	62.5%	LQ
Wouda et al., 2018 [53]	2	2	0	2	0	N	1	2	2	2	1	2	16/24	66.7%	LQ
Zrenner et al., 2018 [54]	2	1	0	2	1	N	1	2	2	2	2	2	17/24	70.8%	MQ
Zrenner et al., 2020 [55]	2	0	0	2	1	N	1	2	2	1	1	2	14/24	58.3%	LQ

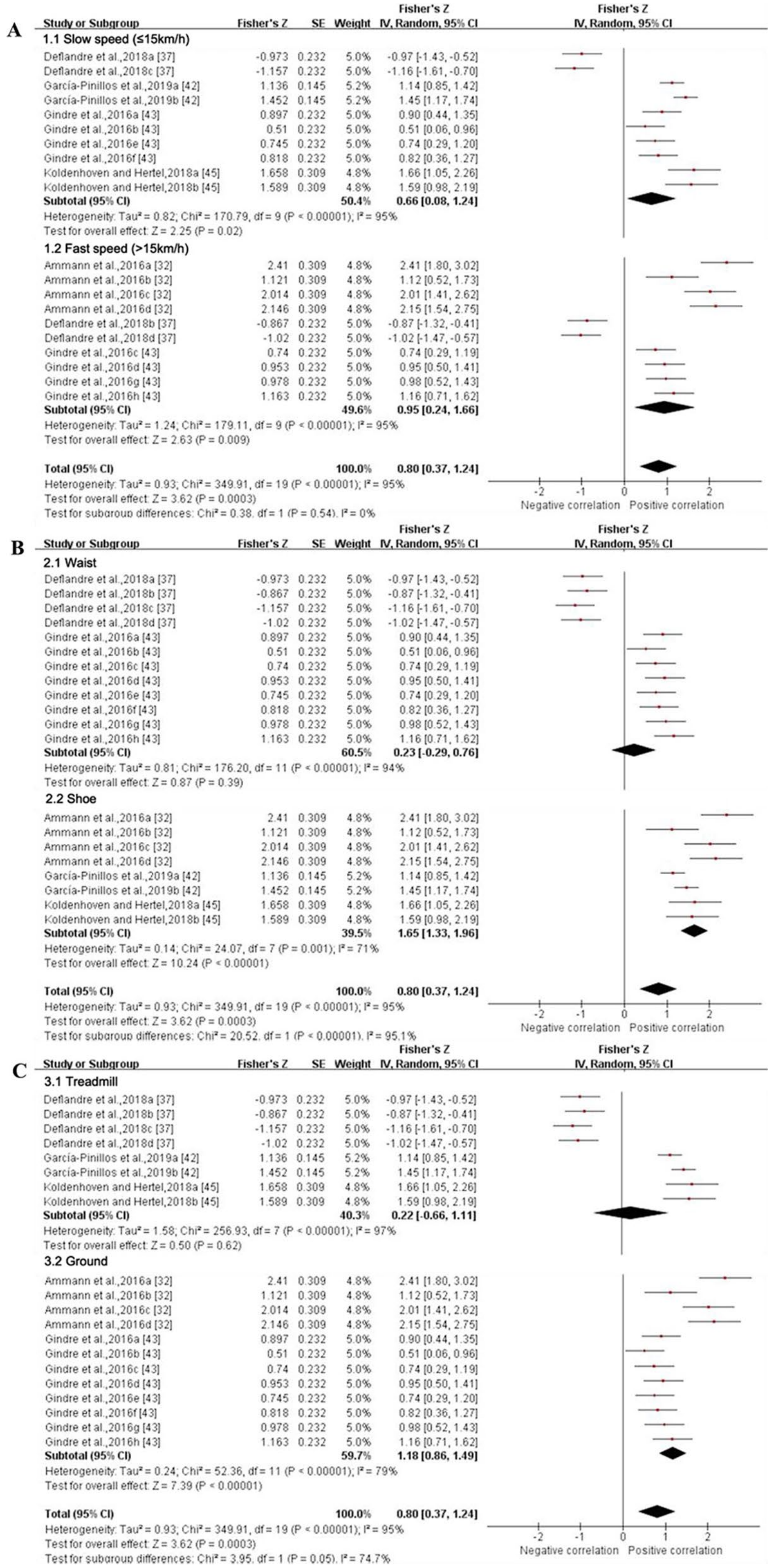
N not mentioned, *MQ* moderate quality, *LQ* low quality

Flight time Data from three MQ studies suggested that the validity for flight time measured by IMUs was poor with no statistical significance (ICC (95% CI) = 0.371 (−0.110, 0.711), $I^2 = 95\%$, $p = 0.13$) (Fig. 4) [37, 42, 43]. Subgroup analysis was not conducted as the results were not statistically significant. The sensitivity analysis showed that after excluding the study of Deflandre et al. [37], the I^2 reduced ($I^2 = 0\%$), summary ICC value increased (ICC (95% CI) = 0.774 (0.716, 0.818), $p < 0.001$). Sensitivity analysis showed that the results were unstable.

Stride length Data from two MQ and one LQ study suggested that the validity for stride length derived from IMUs was excellent (ICC (95% CI) = 0.937 (0.859, 0.972), $I^2 = 79\%$, $p < 0.001$) (Fig. 5) [34, 37, 45]. The results of the subgroup analysis based on running speed, IMUs' position and running surface were not statistically significant ($p \geq 0.2$) (Fig. 5). Sensitivity analysis showed that after excluding the study of Deflandre et al. [37] the I^2 reduced ($I^2 = 40\%$), and the agreement was good (ICC (95% CI) = 0.890 (0.744,

(See figure on next page.)

Fig. 2 Subgroup analysis describing the validity of stance time measured using IMU (Intraclass correlation coefficient, ICC). A based on running speed, B based on location and C based on running surface. Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Ammann et al. 2016a (combined speeds), 2016b (maximal sprinting speed), 2016c (intense training speed), 2016d (normal training speed) [32]; Deflandre et al. 2018a (8 km/h, IMUs vs optical motion capture system), 2018b (16 km/h, IMUs vs optical motion capture system), 2018c (8 km/h, IMUs vs Optogait), 2018d (16 km/h, IMUs vs Optogait) [37]; García-Pinillos et al.,2019a (IMUs vs optical motion capture system, IMUs: Stryd™), 2019b (IMUs vs optical motion capture system, IMUs: RunScribe™) [42]; Gindre et al., 2016a (12 km/h, IMUs vs optical motion capture system), 2016b (15 km/h, IMUs vs optical motion capture system), 2016c (18 km/h, IMUs vs optical motion capture system), 2016d (21 km/h, IMUs vs optical motion capture system), 2016e (12 km/h, IMUs vs Optojump), 2016f (15 km/h, IMUs vs Optojump), 2016 g (18 km/h, IMUs vs Optojump), 2016 h (21 km/h, IMUs vs Optojump) [43]; Koldenhoven and Hertel, 2018a (left limb), 2018b (right limb) [45]. *SE* standard error, *IV* inverse variance, *CI* confidence interval



A

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
1.1 Slow speed (≤15km/h)				
Deflandre et al.,2018a [37]	-0.973	0.232	5.0%	-0.97 [-1.43, -0.52]
Deflandre et al.,2018c [37]	-1.157	0.232	5.0%	-1.16 [-1.61, -0.70]
García-Pinillos et al.,2019a [42]	1.136	0.145	5.2%	1.14 [0.85, 1.42]
García-Pinillos et al.,2019b [42]	1.452	0.145	5.2%	1.45 [1.17, 1.74]
Gindre et al.,2016a [43]	0.897	0.232	5.0%	0.90 [0.44, 1.35]
Gindre et al.,2016b [43]	0.51	0.232	5.0%	0.51 [0.06, 0.96]
Gindre et al.,2016e [43]	0.745	0.232	5.0%	0.74 [0.29, 1.20]
Gindre et al.,2016f [43]	0.818	0.232	5.0%	0.82 [0.36, 1.27]
Koldenhoven and Hertel,2018a [45]	1.658	0.309	4.8%	1.66 [1.05, 2.26]
Koldenhoven and Hertel,2018b [45]	1.589	0.309	4.8%	1.59 [0.98, 2.19]
Subtotal (95% CI)			**50.4%**	**0.66 [0.08, 1.24]**
Heterogeneity: Tau² = 0.82; Chi² = 170.79, df = 9 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 2.25 (P = 0.02)				
1.2 Fast speed (>15km/h)				
Ammann et al.,2016a [32]	2.41	0.309	4.8%	2.41 [1.80, 3.02]
Ammann et al.,2016b [32]	1.121	0.309	4.8%	1.12 [0.52, 1.73]
Ammann et al.,2016c [32]	2.014	0.309	4.8%	2.01 [1.41, 2.62]
Ammann et al.,2016d [32]	2.146	0.309	4.8%	2.15 [1.54, 2.75]
Deflandre et al.,2018b [37]	-0.867	0.232	5.0%	-0.87 [-1.32, -0.41]
Deflandre et al.,2018d [37]	-1.02	0.232	5.0%	-1.02 [-1.47, -0.57]
Gindre et al.,2016c [43]	0.74	0.232	5.0%	0.74 [0.29, 1.19]
Gindre et al.,2016d [43]	0.953	0.232	5.0%	0.95 [0.50, 1.41]
Gindre et al.,2016g [43]	0.978	0.232	5.0%	0.98 [0.52, 1.43]
Gindre et al.,2016h [43]	1.163	0.232	5.0%	1.16 [0.71, 1.62]
Subtotal (95% CI)			**49.6%**	**0.95 [0.24, 1.66]**
Heterogeneity: Tau² = 1.24; Chi² = 179.11, df = 9 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 2.63 (P = 0.009)				
Total (95% CI)			**100.0%**	**0.80 [0.37, 1.24]**
Heterogeneity: Tau² = 0.93; Chi² = 349.91, df = 19 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 3.62 (P = 0.0003)				
Test for subgroup differences: Chi² = 0.38, df = 1 (P = 0.54), I² = 0%				

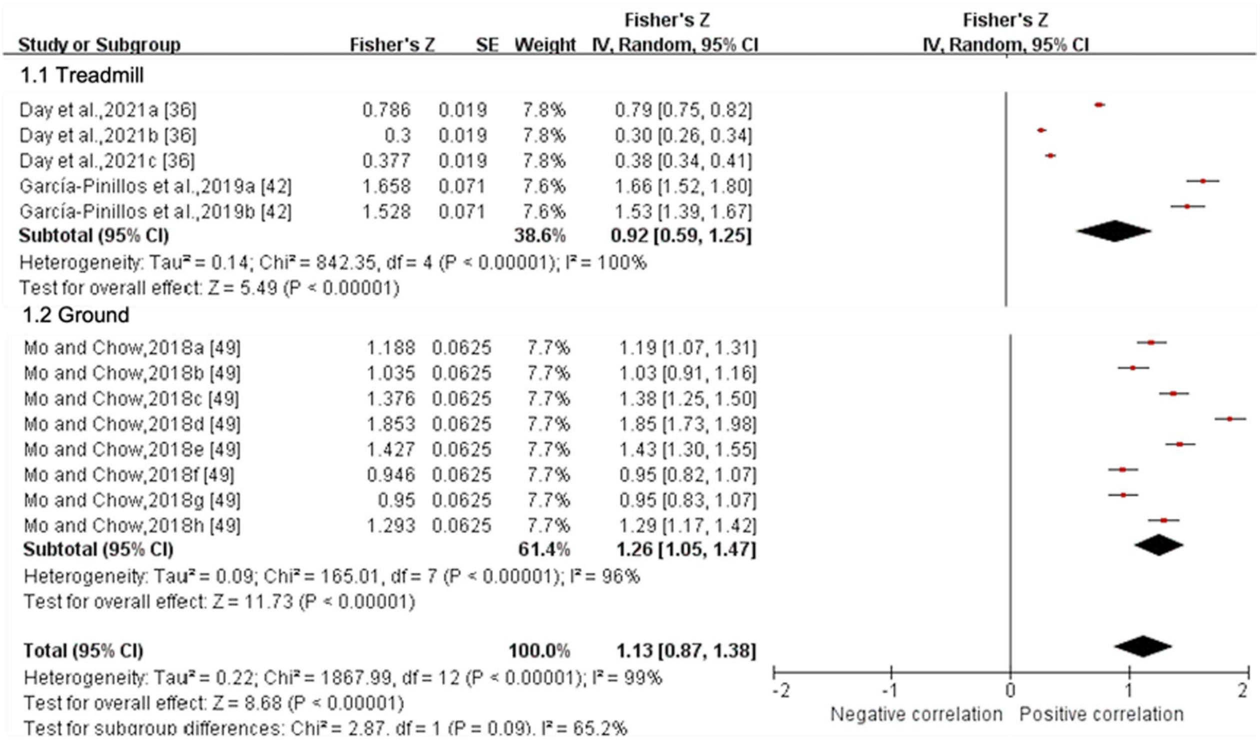
B

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
2.1 Waist				
Deflandre et al.,2018a [37]	-0.973	0.232	5.0%	-0.97 [-1.43, -0.52]
Deflandre et al.,2018b [37]	-0.867	0.232	5.0%	-0.87 [-1.32, -0.41]
Deflandre et al.,2018c [37]	-1.157	0.232	5.0%	-1.16 [-1.61, -0.70]
Deflandre et al.,2018d [37]	-1.02	0.232	5.0%	-1.02 [-1.47, -0.57]
Gindre et al.,2016a [43]	0.897	0.232	5.0%	0.90 [0.44, 1.35]
Gindre et al.,2016b [43]	0.51	0.232	5.0%	0.51 [0.06, 0.96]
Gindre et al.,2016c [43]	0.74	0.232	5.0%	0.74 [0.29, 1.19]
Gindre et al.,2016d [43]	0.953	0.232	5.0%	0.95 [0.50, 1.41]
Gindre et al.,2016e [43]	0.745	0.232	5.0%	0.74 [0.29, 1.20]
Gindre et al.,2016f [43]	0.818	0.232	5.0%	0.82 [0.36, 1.27]
Gindre et al.,2016g [43]	0.978	0.232	5.0%	0.98 [0.52, 1.43]
Gindre et al.,2016h [43]	1.163	0.232	5.0%	1.16 [0.71, 1.62]
Subtotal (95% CI)			**60.5%**	**0.23 [-0.29, 0.76]**
Heterogeneity: Tau² = 0.81; Chi² = 176.20, df = 11 (P < 0.00001); I² = 94%				
Test for overall effect: Z = 0.87 (P = 0.39)				
2.2 Shoe				
Ammann et al.,2016a [32]	2.41	0.309	4.8%	2.41 [1.80, 3.02]
Ammann et al.,2016b [32]	1.121	0.309	4.8%	1.12 [0.52, 1.73]
Ammann et al.,2016c [32]	2.014	0.309	4.8%	2.01 [1.41, 2.62]
Ammann et al.,2016d [32]	2.146	0.309	4.8%	2.15 [1.54, 2.75]
García-Pinillos et al.,2019a [42]	1.136	0.145	5.2%	1.14 [0.85, 1.42]
García-Pinillos et al.,2019b [42]	1.452	0.145	5.2%	1.45 [1.17, 1.74]
Koldenhoven and Hertel,2018a [45]	1.658	0.309	4.8%	1.66 [1.05, 2.26]
Koldenhoven and Hertel,2018b [45]	1.589	0.309	4.8%	1.59 [0.98, 2.19]
Subtotal (95% CI)			**39.5%**	**1.65 [1.33, 1.96]**
Heterogeneity: Tau² = 0.14; Chi² = 24.07, df = 7 (P = 0.001); I² = 71%				
Test for overall effect: Z = 10.24 (P < 0.00001)				
Total (95% CI)			**100.0%**	**0.80 [0.37, 1.24]**
Heterogeneity: Tau² = 0.93; Chi² = 349.91, df = 19 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 3.62 (P = 0.0003)				
Test for subgroup differences: Chi² = 20.52, df = 1 (P < 0.00001), I² = 95.1%				

C

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
3.1 Treadmill				
Deflandre et al.,2018a [37]	-0.973	0.232	5.0%	-0.97 [-1.43, -0.52]
Deflandre et al.,2018b [37]	-0.867	0.232	5.0%	-0.87 [-1.32, -0.41]
Deflandre et al.,2018c [37]	-1.157	0.232	5.0%	-1.16 [-1.61, -0.70]
Deflandre et al.,2018d [37]	-1.02	0.232	5.0%	-1.02 [-1.47, -0.57]
García-Pinillos et al.,2019a [42]	1.136	0.145	5.2%	1.14 [0.85, 1.42]
García-Pinillos et al.,2019b [42]	1.452	0.145	5.2%	1.45 [1.17, 1.74]
Koldenhoven and Hertel,2018a [45]	1.658	0.309	4.8%	1.66 [1.05, 2.26]
Koldenhoven and Hertel,2018b [45]	1.589	0.309	4.8%	1.59 [0.98, 2.19]
Subtotal (95% CI)			**40.3%**	**0.22 [-0.66, 1.11]**
Heterogeneity: Tau² = 1.58; Chi² = 256.93, df = 7 (P < 0.00001); I² = 97%				
Test for overall effect: Z = 0.50 (P = 0.62)				
3.2 Ground				
Ammann et al.,2016a [32]	2.41	0.309	4.8%	2.41 [1.80, 3.02]
Ammann et al.,2016b [32]	1.121	0.309	4.8%	1.12 [0.52, 1.73]
Ammann et al.,2016c [32]	2.014	0.309	4.8%	2.01 [1.41, 2.62]
Ammann et al.,2016d [32]	2.146	0.309	4.8%	2.15 [1.54, 2.75]
Gindre et al.,2016a [43]	0.897	0.232	5.0%	0.90 [0.44, 1.35]
Gindre et al.,2016b [43]	0.51	0.232	5.0%	0.51 [0.06, 0.96]
Gindre et al.,2016c [43]	0.74	0.232	5.0%	0.74 [0.29, 1.19]
Gindre et al.,2016d [43]	0.953	0.232	5.0%	0.95 [0.50, 1.41]
Gindre et al.,2016e [43]	0.745	0.232	5.0%	0.74 [0.29, 1.20]
Gindre et al.,2016f [43]	0.818	0.232	5.0%	0.82 [0.36, 1.27]
Gindre et al.,2016g [43]	0.978	0.232	5.0%	0.98 [0.52, 1.43]
Gindre et al.,2016h [43]	1.163	0.232	5.0%	1.16 [0.71, 1.62]
Subtotal (95% CI)			**59.7%**	**1.18 [0.86, 1.49]**
Heterogeneity: Tau² = 0.24; Chi² = 52.36, df = 11 (P < 0.00001); I² = 79%				
Test for overall effect: Z = 7.39 (P < 0.00001)				
Total (95% CI)			**100.0%**	**0.80 [0.37, 1.24]**
Heterogeneity: Tau² = 0.93; Chi² = 349.91, df = 19 (P < 0.00001); I² = 95%				
Test for overall effect: Z = 3.62 (P = 0.0003)				
Test for subgroup differences: Chi² = 3.95, df = 1 (P = 0.05), I² = 74.7%				

Fig. 2 (See legend on previous page.)



Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
1.1 Treadmill				
Day et al.,2021a [36]	0.786	0.019	7.8%	0.79 [0.75, 0.82]
Day et al.,2021b [36]	0.3	0.019	7.8%	0.30 [0.26, 0.34]
Day et al.,2021c [36]	0.377	0.019	7.8%	0.38 [0.34, 0.41]
García-Pinillos et al.,2019a [42]	1.658	0.071	7.6%	1.66 [1.52, 1.80]
García-Pinillos et al.,2019b [42]	1.528	0.071	7.6%	1.53 [1.39, 1.67]
Subtotal (95% CI)			**38.6%**	**0.92 [0.59, 1.25]**
Heterogeneity: Tau² = 0.14; Chi² = 842.35, df = 4 (P < 0.00001); I² = 100%				
Test for overall effect: Z = 5.49 (P < 0.00001)				
1.2 Ground				
Mo and Chow,2018a [49]	1.188	0.0625	7.7%	1.19 [1.07, 1.31]
Mo and Chow,2018b [49]	1.035	0.0625	7.7%	1.03 [0.91, 1.16]
Mo and Chow,2018c [49]	1.376	0.0625	7.7%	1.38 [1.25, 1.50]
Mo and Chow,2018d [49]	1.853	0.0625	7.7%	1.85 [1.73, 1.98]
Mo and Chow,2018e [49]	1.427	0.0625	7.7%	1.43 [1.30, 1.55]
Mo and Chow,2018f [49]	0.946	0.0625	7.7%	0.95 [0.82, 1.07]
Mo and Chow,2018g [49]	0.95	0.0625	7.7%	0.95 [0.83, 1.07]
Mo and Chow,2018h [49]	1.293	0.0625	7.7%	1.29 [1.17, 1.42]
Subtotal (95% CI)			**61.4%**	**1.26 [1.05, 1.47]**
Heterogeneity: Tau² = 0.09; Chi² = 165.01, df = 7 (P < 0.00001); I² = 96%				
Test for overall effect: Z = 11.73 (P < 0.00001)				
Total (95% CI)			**100.0%**	**1.13 [0.87, 1.38]**
Heterogeneity: Tau² = 0.22; Chi² = 1867.99, df = 12 (P < 0.00001); I² = 99%				
Test for overall effect: Z = 8.68 (P < 0.00001)				
Test for subgroup differences: Chi² = 2.87, df = 1 (P = 0.09), I² = 65.2%				

Fig. 3 Subgroup analysis describing the validity of stance time measured using IMU (Pearson correlation coefficient, *r*). Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Day et al.,2021a (5 Hz cutoff), 2021b (10 Hz cutoff), 2021c (30 Hz cutoff) [36]; García-Pinillos et al.,2019a (IMUs vs optical motion capture system, IMUs: Stryd™), 2019b (IMUs vs optical motion capture system, IMUs: RunScribe™) [42]; Mo and Chow, 2018a (jogging, L-method), 2018b (jogging, M-method), 2018c (jogging, S-method), 2018d (jogging, MS-method), 2018e (running, L-method), 2018f (running, M-method), 2018 g (running, S-method), 2018 h (running, MS-method) [49]. *SE* standard error, *IV* inverse variance, *CI* confidence interval.

0.954), $p < 0.001$). Sensitivity analysis showed that the results were stable.

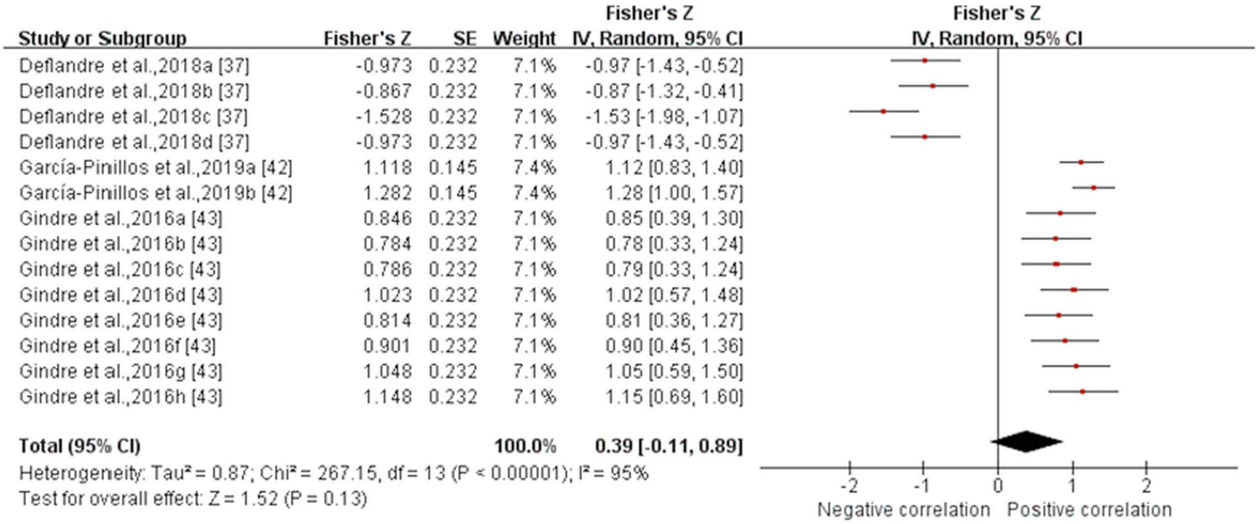
Step frequency Data from four MQ studies suggested that the validity for step frequency derived from IMUs was excellent [(ICC (95% CI) = 0.926 (0.896, 0.948), $I^2 = 61\%$, $p < 0.001$) (Fig. 6) [37, 42, 43] and (r (95% CI) = 0.989 (0.957, 0.997), $I^2 = 100\%$, $p < 0.001$) (Fig. 7) [38, 42]]. The results of subgroup analysis based on running speed showed that the summary ICC value at fast speed (ICC (95% CI) = 0.890 (0.827, 0.932), $I^2 = 49\%$) was lower than that at slow speed (ICC (95% CI) = 0.945 (0.919, 0.962), $I^2 = 42\%$) (Fig. 6). The IMUs at the waist showed good to excellent agreement (ICC (95% CI) = 0.912 (0.879, 0.937), $I^2 = 43\%$), and the shoelace showed excellent agreement (ICC (95% CI) = 0.965 (0.948, 0.976), $I^2 = 0\%$) (Fig. 6). Running on the treadmill (ICC (95% CI) = 0.949 (0.914, 0.969), $I^2 = 66\%$) showed higher consistency compared to the ground (ICC (95% CI) = 0.900 (0.862, 0.926), $I^2 = 0\%$) (Fig. 6). Due to the limited amount of literature, no subgroup analysis was performed on the validity of the step frequency measured by IMUs as reported through the r values. Sensitivity analysis showed that the results were stable.

Running speed Data from two LQ studies suggested that the validity for running speed measured by IMUs was good (ICC (95% CI) = 0.848 (0.523, 0.958), $I^2 = 88\%$, $p = 0.0003$) (Fig. 8) [45, 46]. Subgroup analyses could not be performed due to the insufficient number of studies. Sensitivity analysis showed that the results were stable.

Ankle angle in the sagittal plane Data from one MQ and one LQ study suggested that the validity for ankle angle in the sagittal plane measured by IMUs was excellent (r (95% CI) = 0.939 (0.544, 0.993), $I^2 = 99\%$, $p = 0.002$) (Fig. 9) [39, 44]. Subgroup and sensitivity analyses could not be performed due to the insufficient number of studies.

Reliability

Six gait spatiotemporal outcomes and 22 lower extremity joint kinematics outcomes were assessed across the six studies that examined reliability for IMUs. From this group, only two gait spatiotemporal outcomes presented sufficient study quality and statistical outcomes



Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
Deflandre et al.,2018a [37]	-0.973	0.232	7.1%	-0.97 [-1.43, -0.52]
Deflandre et al.,2018b [37]	-0.867	0.232	7.1%	-0.87 [-1.32, -0.41]
Deflandre et al.,2018c [37]	-1.528	0.232	7.1%	-1.53 [-1.98, -1.07]
Deflandre et al.,2018d [37]	-0.973	0.232	7.1%	-0.97 [-1.43, -0.52]
García-Pinillos et al.,2019a [42]	1.118	0.145	7.4%	1.12 [0.83, 1.40]
García-Pinillos et al.,2019b [42]	1.282	0.145	7.4%	1.28 [1.00, 1.57]
Gindre et al.,2016a [43]	0.846	0.232	7.1%	0.85 [0.39, 1.30]
Gindre et al.,2016b [43]	0.784	0.232	7.1%	0.78 [0.33, 1.24]
Gindre et al.,2016c [43]	0.786	0.232	7.1%	0.79 [0.33, 1.24]
Gindre et al.,2016d [43]	1.023	0.232	7.1%	1.02 [0.57, 1.48]
Gindre et al.,2016e [43]	0.814	0.232	7.1%	0.81 [0.36, 1.27]
Gindre et al.,2016f [43]	0.901	0.232	7.1%	0.90 [0.45, 1.36]
Gindre et al.,2016g [43]	1.048	0.232	7.1%	1.05 [0.59, 1.50]
Gindre et al.,2016h [43]	1.148	0.232	7.1%	1.15 [0.69, 1.60]
Total (95% CI)			**100.0%**	**0.39 [-0.11, 0.89]**

Heterogeneity: Tau² = 0.87; Chi² = 267.15, df = 13 (P < 0.00001); I² = 95%
Test for overall effect: Z = 1.52 (P = 0.13)

Fig. 4 Forest plot describing the validity of flight time measured using IMU (Intraclass correlation coefficient, ICC). Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Deflandre et al.,2018a (8 km/h, IMUs vs optical motion capture system), 2018b (16 km/h, IMUs vs optical motion capture system), 2018c (8 km/h, IMUs vs Optogait), 2018d (16 km/h, IMUs vs Optogait) [37]; García-Pinillos et al.,2019a (IMUs vs optical motion capture system, IMUs: Stryd™), 2019b (IMUs vs optical motion capture system, IMUs: RunScribe™) [42]; Gindre et al.,2016a (12 km/h, IMUs vs optical motion capture system), 2016b (15 km/h, IMUs vs optical motion capture system), 2016c (18 km/h, IMUs vs optical motion capture system), 2016d (21 km/h, IMUs vs optical motion capture system), 2016e (12 km/h, IMUs vs Optojump), 2016f (15 km/h, IMUs vs Optojump), 2016 g (18 km/h, IMUs vs Optojump), 2016 h (21 km/h, IMUs vs Optojump) [43]. *SE* standard erroSr, *IV* inverse variance, *CI* confidence interval.

for meta-analysis (Figs. 10 and 11). Similar to validity, the inability to pool outcomes were due to either a limited number of studies or a lack of consistency in data reporting. Studies that were unable to be pooled were qualitatively summarised according to outcomes in Additional File 3.

Quantitative Pooling for Reliability *Stance time* Data from three MQ studies suggested that the reliability for stance time measured by IMUs was excellent (ICC (95% CI) = 0.954 (0.903, 0.978), $I^2 = 74\%$, $p < 0.001$) (Fig. 10) [32, 37, 43]. Subgroup analysis showed no significant effect of running speed, IMU's position and running surface on the reliability for stance time derived from IMUs ($p \geq 0.37$). Sensitivity analysis showed that the results were stable.

Step frequency Data from two MQ studies suggested that the reliability for flight time measured by IMUs was good (ICC (95% CI) = 0.896 (0.837, 0.933), $I^2 = 29\%$, $p < 0.001$) (Fig. 11) [37, 43]. Subgroup analyses could not be performed due to the insufficient number of studies. Sensitivity analysis showed that the results were stable.

Discussion

The aim of this review was to determine the concurrent validity and test–retest reliability of biomechanical outcomes derived from IMUs during running in healthy adults. The main findings of this review were as follows: (1) among the studies examining the validity or reliability of measurements from IMUs during running, there have been noticeably more studies involving gait spatiotemporal outcomes than those involving lower limb joint kinematics. (2) Regarding validity: the stride length, step frequency and ankle angle in the sagittal plane showed excellent agreement, the stance time depicted moderate to excellent agreement and running speed was good, with statistical significance ($p < 0.01$), whereas the summary Fisher's Z value of flight time was not statistically

(See figure on next page.)

Fig. 5 Subgroup analysis describing the validity of stride length measured using IMU (Intraclass correlation coefficient, ICC). A based on running speed, B based on location and C based on running surface. Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Deflandre et al. 2018a (8 km/h, IMUs vs optical motion capture system), 2018b (16 km/h, IMUs vs optical motion capture system), 2018c (8 km/h, IMUs vs Optogait), 2018d (16 km/h, IMUs vs Optogait) [37]; Koldenhoven and Hertel, 2018a (left limb), 2018b (right limb) [45]. *SE* standard error, *IV* inverse variance, *CI* confidence interval.

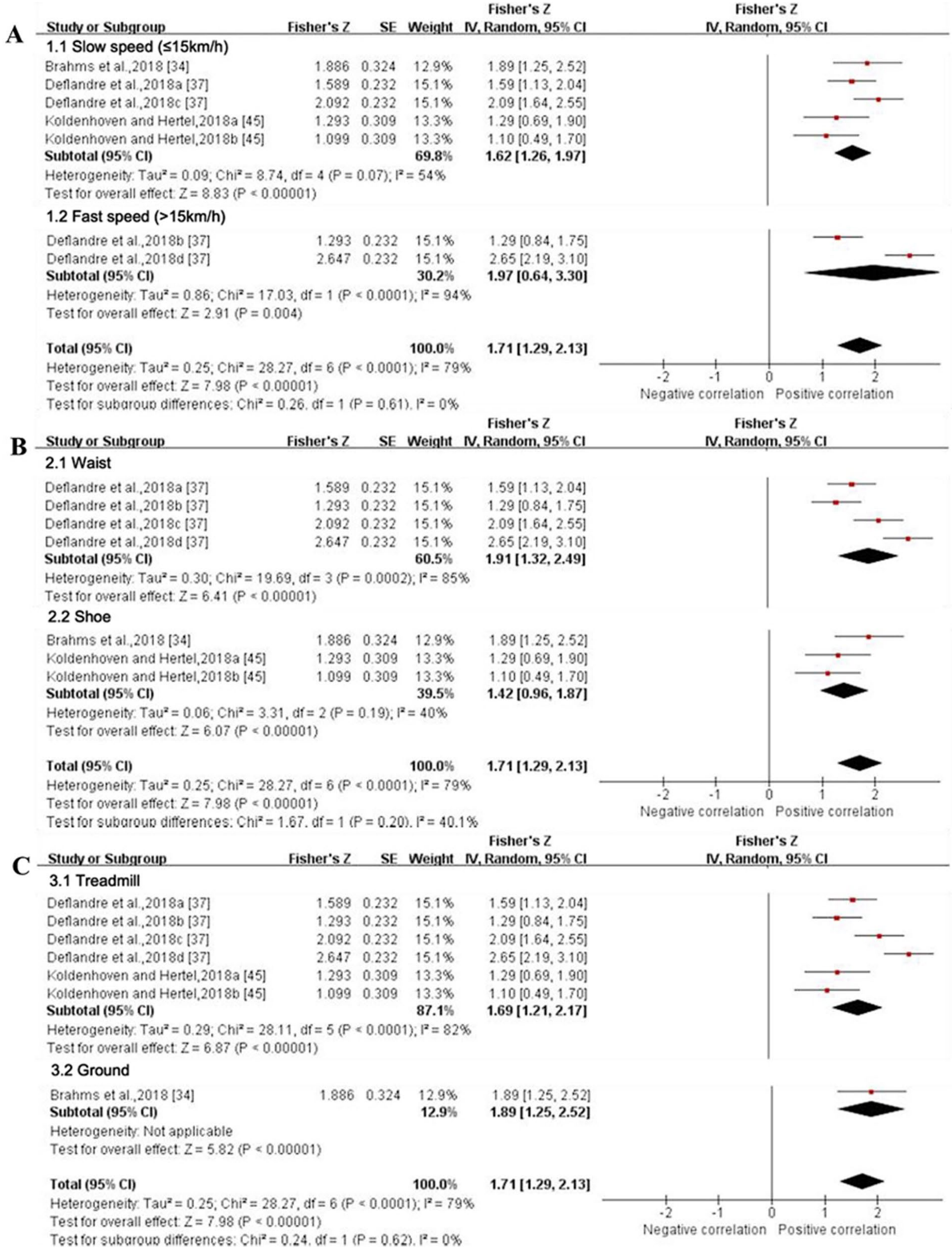


Fig. 5 (See legend on previous page.)

significant ($p = 0.13$). (3) For reliability: stance time showed excellent test–retest reliability and step frequency showed good test–retest reliability, and summary Fisher's Z values were statistically significant ($p < 0.001$).

This systematic review used a similar review process to the previous study, which evaluated the validity and reliability of measurements from IMUs during walking [17]. As far as the authors know, this is the first meta-analysis involving the assessment of validity and reliability of lower limb joint kinematics measured by IMUs during running. It has been reported that running speed, IMUs' position and running surface are the main factors related to the validity and reliability of measurements from IMUs [41, 56]. Therefore, to explore the specific effects of these factors on the validity and reliability of gait spatiotemporal outcomes and lower extremity joint kinematics derived from IMUs, the subgroup analyses based on running speed, IMUs' position and running surface were conducted for the parameters that could be pooled.

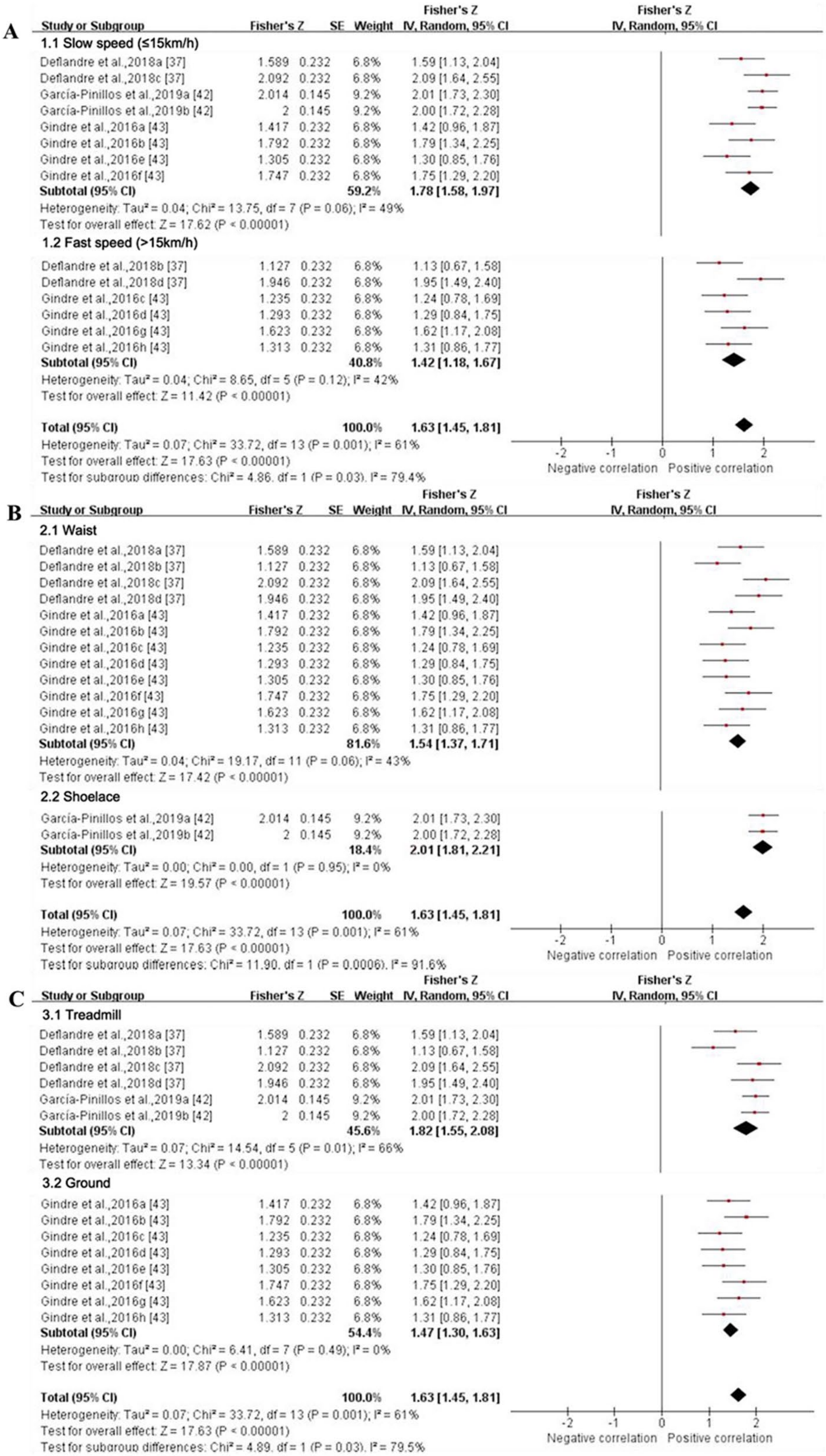
In this systematic review and meta-analysis, the included studies have measured the validity and reliability of IMUs' measurements at a variety of running speeds (7.2–21 km/h). Specifically, two studies measured gait spatiotemporal parameters and sagittal joint kinematics at preferred running speed [42, 50]. Although difference in preferred running speed was found between them (2.93 ± 0.35 m/s and 3.25 ± 0.36 m/s), both studies suggested that the measurements from IMUs and optical motion capture systems had an almost perfect association (ICC > 0.81 and CMC > 0.950) [42, 50]. In addition, five studies assessed running at maximum speed, and a general conclusion was that speed had an impact on the validity of the measurements from IMUs [32, 33, 40, 41, 51]. Although only the validity of step frequency derived from IMUs was statistically different in the subgroup analysis based on running speed ($p = 0.03$), a summary of other statistical outcomes in Additional File 3 also showed a general decrease in the validity and reliability of gait spatiotemporal parameters and lower limb joint kinematics as running speed increased [32, 37, 41, 43, 49, 52, 56]. This can be explained by the fact that increase in running speed increases peak vertical acceleration at impact, resulting in soft tissue artifacts [57, 58]. However, the results of Watari et al. [52] showed that the validity is the lowest at the lowest speed (2.7 m/s), similar to the results of Gindre et al. [43]. Difference in IMUs' placement may be one of the main reasons for the discrepancies in the conclusion [58, 59].

Collectively, foot and shank were the most common IMUs attachment locations when assessing the validity or reliability of IMUs measurement of running kinematics, which is consistent with previous studies [40, 49, 60]. Subgroup analysis based on IMU's position depicted that for stance time and step frequency, placing the IMUs on the shoe yielded more accurate measurements than placing the IMUs around the waist ($p < 0.001$). Previous studies generally suggested that IMUs' position closer to the foot can more accurately capture acceleration signals and thus recognise gait events [40, 49, 60]. However, a recent review showed that placing the IMUs on the foot, tibia and lumbar spine yielded valid and reliable stride data, suggesting measurement position may not be a critical factor [18]. Since only subgroup analyses on the waist and shoe of two gait parameters were performed, it is therefore unclear whether other IMUs' placements affect the gait spatiotemporal and lower extremity kinematics outcomes.

In this review, running on an indoor track or walkway, running on a treadmill and running outside were included. Previous evidence suggests that running on a treadmill and running on the ground/track are associated with different biomechanical performance [56, 61–63], so we divided the running surfaces into treadmill and ground before carrying out subgroup analyses. The results of the subgroup analysis based on running surface showed that the validity of running on a treadmill was better than running on the ground for step frequency. Interestingly, close to half of the studies in our review were conducted on a treadmill [35–38, 40, 42, 45–47, 50, 52, 53]. The effective control of running speed on treadmills is the main reason. On the other hand, it also provides support for researchers to assess the effect of specific running speed on the validity and reliability of the IMUs' measurements. Moreover, to improve ecological validity, IMUs are necessary for outdoor measurement.

(See figure on next page.)

Fig. 6 Subgroup analysis describing the validity of step frequency measured using IMU (Intraclass correlation coefficient, ICC). A based on running speed, B based on location and C based on running surface. Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Deflandre et al. 2018a (8 km/h, IMUs vs optical motion capture system), 2018b (16 km/h, IMUs vs optical motion capture system), 2018c (8 km/h, IMUs vs Optogait), 2018d (16 km/h, IMUs vs Optogait) [37]; García-Pinillos et al.,2019a (IMUs vs optical motion capture system, IMUs: Stryd™), 2019b (IMUs vs optical motion capture system, IMUs: RunScribe™) [42]; Gindre et al.,2016a (12 km/h, IMUs vs optical motion capture system), 2016b (15 km/h, IMUs vs optical motion capture system), 2016c (18 km/h, IMUs vs optical motion capture system), 2016d (21 km/h, IMUs vs optical motion capture system), 2016e (12 km/h, IMUs vs Optojump), 2016f (15 km/h, IMUs vs Optojump), 2016 g (18 km/h, IMUs vs Optojump), 2016 h (21 km/h, IMUs vs Optojump) [43]. *SE* standard error, *IV* inverse variance, *CI* confidence interval



A

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
1.1 Slow speed (≤15km/h)				
Deflandre et al.,2018a [37]	1.589	0.232	6.8%	1.59 [1.13, 2.04]
Deflandre et al.,2018c [37]	2.092	0.232	6.8%	2.09 [1.64, 2.55]
García-Pinillos et al.,2019a [42]	2.014	0.145	9.2%	2.01 [1.73, 2.30]
García-Pinillos et al.,2019b [42]	2	0.145	9.2%	2.00 [1.72, 2.28]
Gindre et al.,2016a [43]	1.417	0.232	6.8%	1.42 [0.96, 1.87]
Gindre et al.,2016b [43]	1.792	0.232	6.8%	1.79 [1.34, 2.25]
Gindre et al.,2016e [43]	1.305	0.232	6.8%	1.30 [0.85, 1.76]
Gindre et al.,2016f [43]	1.747	0.232	6.8%	1.75 [1.29, 2.20]
Subtotal (95% CI)			**59.2%**	**1.78 [1.58, 1.97]**

Heterogeneity: Tau² = 0.04; Chi² = 13.75, df = 7 (P = 0.06); I² = 49%
Test for overall effect: Z = 17.62 (P < 0.00001)

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
1.2 Fast speed (>15km/h)				
Deflandre et al.,2018b [37]	1.127	0.232	6.8%	1.13 [0.67, 1.58]
Deflandre et al.,2018d [37]	1.946	0.232	6.8%	1.95 [1.49, 2.40]
Gindre et al.,2016c [43]	1.235	0.232	6.8%	1.24 [0.78, 1.69]
Gindre et al.,2016d [43]	1.293	0.232	6.8%	1.29 [0.84, 1.75]
Gindre et al.,2016g [43]	1.623	0.232	6.8%	1.62 [1.17, 2.08]
Gindre et al.,2016h [43]	1.313	0.232	6.8%	1.31 [0.86, 1.77]
Subtotal (95% CI)			**40.8%**	**1.42 [1.18, 1.67]**

Heterogeneity: Tau² = 0.04; Chi² = 8.65, df = 5 (P = 0.12); I² = 42%
Test for overall effect: Z = 11.42 (P < 0.00001)

Total (95% CI) **100.0%** **1.63 [1.45, 1.81]**
Heterogeneity: Tau² = 0.07; Chi² = 33.72, df = 13 (P = 0.001); I² = 61%
Test for overall effect: Z = 17.63 (P < 0.00001)
Test for subgroup differences: Chi² = 4.86, df = 1 (P = 0.03), I² = 79.4%

B

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
2.1 Waist				
Deflandre et al.,2018a [37]	1.589	0.232	6.8%	1.59 [1.13, 2.04]
Deflandre et al.,2018b [37]	1.127	0.232	6.8%	1.13 [0.67, 1.58]
Deflandre et al.,2018c [37]	2.092	0.232	6.8%	2.09 [1.64, 2.55]
Deflandre et al.,2018d [37]	1.946	0.232	6.8%	1.95 [1.49, 2.40]
Gindre et al.,2016a [43]	1.417	0.232	6.8%	1.42 [0.96, 1.87]
Gindre et al.,2016b [43]	1.792	0.232	6.8%	1.79 [1.34, 2.25]
Gindre et al.,2016c [43]	1.235	0.232	6.8%	1.24 [0.78, 1.69]
Gindre et al.,2016d [43]	1.293	0.232	6.8%	1.29 [0.84, 1.75]
Gindre et al.,2016e [43]	1.305	0.232	6.8%	1.30 [0.85, 1.76]
Gindre et al.,2016f [43]	1.747	0.232	6.8%	1.75 [1.29, 2.20]
Gindre et al.,2016g [43]	1.623	0.232	6.8%	1.62 [1.17, 2.08]
Gindre et al.,2016h [43]	1.313	0.232	6.8%	1.31 [0.86, 1.77]
Subtotal (95% CI)			**81.6%**	**1.54 [1.37, 1.71]**

Heterogeneity: Tau² = 0.04; Chi² = 19.17, df = 11 (P = 0.06); I² = 43%
Test for overall effect: Z = 17.42 (P < 0.00001)

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
2.2 Shoelace				
García-Pinillos et al.,2019a [42]	2.014	0.145	9.2%	2.01 [1.73, 2.30]
García-Pinillos et al.,2019b [42]	2	0.145	9.2%	2.00 [1.72, 2.28]
Subtotal (95% CI)			**18.4%**	**2.01 [1.81, 2.21]**

Heterogeneity: Tau² = 0.00; Chi² = 0.00, df = 1 (P = 0.95); I² = 0%
Test for overall effect: Z = 19.57 (P < 0.00001)

Total (95% CI) **100.0%** **1.63 [1.45, 1.81]**
Heterogeneity: Tau² = 0.07; Chi² = 33.72, df = 13 (P = 0.001); I² = 61%
Test for overall effect: Z = 17.63 (P < 0.00001)
Test for subgroup differences: Chi² = 11.90, df = 1 (P = 0.0006), I² = 91.6%

C

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
3.1 Treadmill				
Deflandre et al.,2018a [37]	1.589	0.232	6.8%	1.59 [1.13, 2.04]
Deflandre et al.,2018b [37]	1.127	0.232	6.8%	1.13 [0.67, 1.58]
Deflandre et al.,2018c [37]	2.092	0.232	6.8%	2.09 [1.64, 2.55]
Deflandre et al.,2018d [37]	1.946	0.232	6.8%	1.95 [1.49, 2.40]
García-Pinillos et al.,2019a [42]	2.014	0.145	9.2%	2.01 [1.73, 2.30]
García-Pinillos et al.,2019b [42]	2	0.145	9.2%	2.00 [1.72, 2.28]
Subtotal (95% CI)			**45.6%**	**1.82 [1.55, 2.08]**

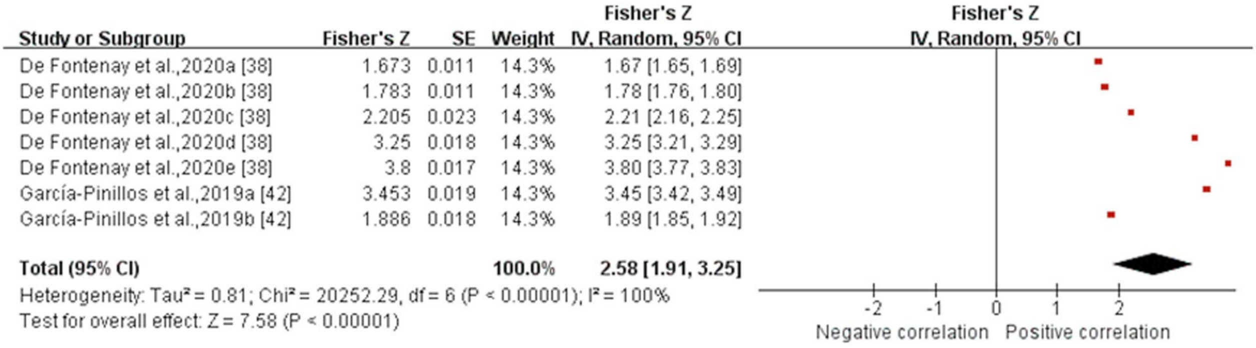
Heterogeneity: Tau² = 0.07; Chi² = 14.54, df = 5 (P = 0.01); I² = 66%
Test for overall effect: Z = 13.34 (P < 0.00001)

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
3.2 Ground				
Gindre et al.,2016a [43]	1.417	0.232	6.8%	1.42 [0.96, 1.87]
Gindre et al.,2016b [43]	1.792	0.232	6.8%	1.79 [1.34, 2.25]
Gindre et al.,2016c [43]	1.235	0.232	6.8%	1.24 [0.78, 1.69]
Gindre et al.,2016d [43]	1.293	0.232	6.8%	1.29 [0.84, 1.75]
Gindre et al.,2016e [43]	1.305	0.232	6.8%	1.30 [0.85, 1.76]
Gindre et al.,2016f [43]	1.747	0.232	6.8%	1.75 [1.29, 2.20]
Gindre et al.,2016g [43]	1.623	0.232	6.8%	1.62 [1.17, 2.08]
Gindre et al.,2016h [43]	1.313	0.232	6.8%	1.31 [0.86, 1.77]
Subtotal (95% CI)			**54.4%**	**1.47 [1.30, 1.63]**

Heterogeneity: Tau² = 0.00; Chi² = 6.41, df = 7 (P = 0.49); I² = 0%
Test for overall effect: Z = 17.87 (P < 0.00001)

Total (95% CI) **100.0%** **1.63 [1.45, 1.81]**
Heterogeneity: Tau² = 0.07; Chi² = 33.72, df = 13 (P = 0.001); I² = 61%
Test for overall effect: Z = 17.63 (P < 0.00001)
Test for subgroup differences: Chi² = 4.89, df = 1 (P = 0.03), I² = 79.5%

Fig. 6 (See legend on previous page.)

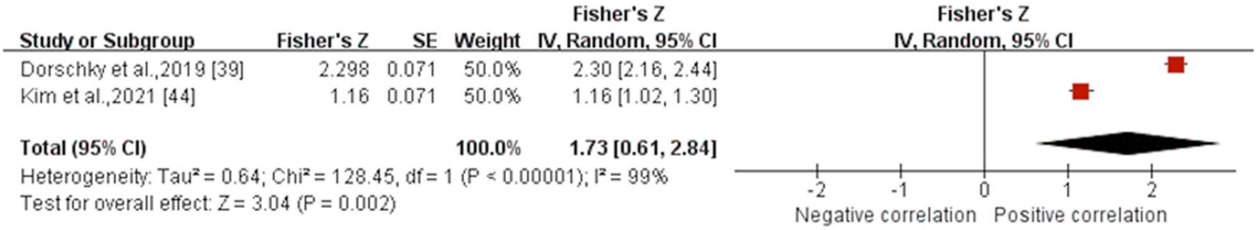


Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
De Fontenay et al.,2020a [38]	1.673	0.011	14.3%	1.67 [1.65, 1.69]
De Fontenay et al.,2020b [38]	1.783	0.011	14.3%	1.78 [1.76, 1.80]
De Fontenay et al.,2020c [38]	2.205	0.023	14.3%	2.21 [2.16, 2.25]
De Fontenay et al.,2020d [38]	3.25	0.018	14.3%	3.25 [3.21, 3.29]
De Fontenay et al.,2020e [38]	3.8	0.017	14.3%	3.80 [3.77, 3.83]
García-Pinillos et al.,2019a [42]	3.453	0.019	14.3%	3.45 [3.42, 3.49]
García-Pinillos et al.,2019b [42]	1.886	0.018	14.3%	1.89 [1.85, 1.92]
Total (95% CI)			**100.0%**	**2.58 [1.91, 3.25]**

Heterogeneity: Tau² = 0.81; Chi² = 20252.29, df = 6 (P < 0.00001); I² = 100%

Test for overall effect: Z = 7.58 (P < 0.00001)

Fig. 7 Forest plot describing the validity of step frequency measured using IMU (Pearson correlation coefficient, *r*). Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. De Fontenay et al.,2020a (IMUs vs optical motion capture system, IMUs: Moov Now™), 2020b (IMUs vs optical motion capture system, IMUs: MilestonePod), 2020c (IMUs vs optical motion capture system, IMUs: RunScribe™), 2020d (IMUs vs optical motion capture system, IMUs: Zoi), 2020e (IMUs vs optical motion capture system, IMUs: TgForce) [38]; García-Pinillos et al.,2019a (IMUs vs optical motion capture system, IMUs: Stryd™), 2019b (IMUs vs optical motion capture system, IMUs: RunScribe™) [42]. *SE* standard error, *IV* inverse variance, *CI* confidence interval

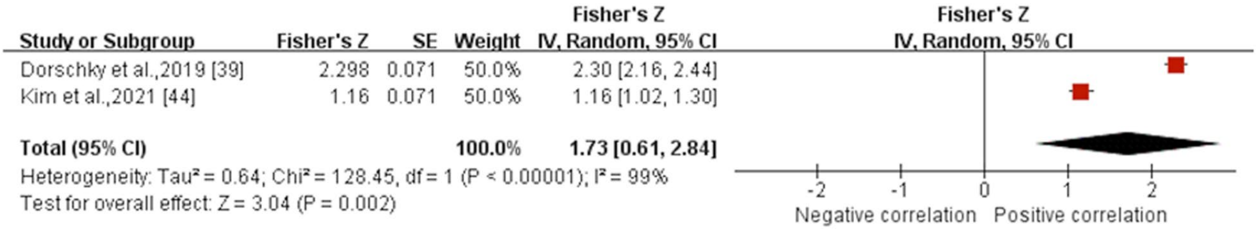


Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
Dorschky et al.,2019 [39]	2.298	0.071	50.0%	2.30 [2.16, 2.44]
Kim et al.,2021 [44]	1.16	0.071	50.0%	1.16 [1.02, 1.30]
Total (95% CI)			**100.0%**	**1.73 [0.61, 2.84]**

Heterogeneity: Tau² = 0.64; Chi² = 128.45, df = 1 (P < 0.00001); I² = 99%

Test for overall effect: Z = 3.04 (P = 0.002)

Fig. 8 Forest plot describing the validity of running speed measured using IMU (Intraclass correlation coefficient, ICC). Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Koldenhoven and Hertel, 2018a (left limb), 2018b (right limb) [45]; Konham et al., 2016a (moderate), 2016b (vigorous) [46]. *SE* standard error, *IV* inverse variance, *CI* confidence interval



Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
Dorschky et al.,2019 [39]	2.298	0.071	50.0%	2.30 [2.16, 2.44]
Kim et al.,2021 [44]	1.16	0.071	50.0%	1.16 [1.02, 1.30]
Total (95% CI)			**100.0%**	**1.73 [0.61, 2.84]**

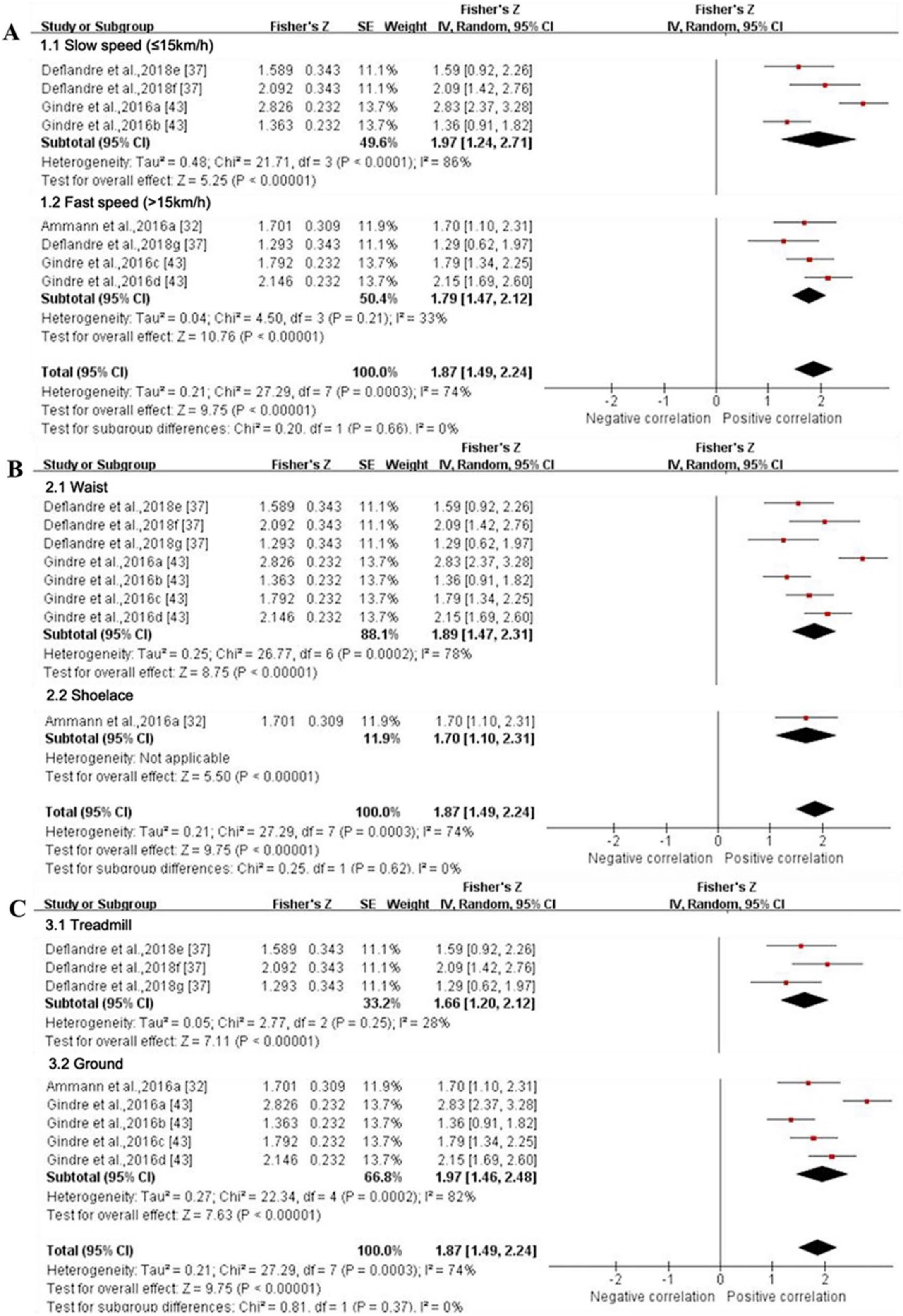
Heterogeneity: Tau² = 0.64; Chi² = 128.45, df = 1 (P < 0.00001); I² = 99%

Test for overall effect: Z = 3.04 (P = 0.002)

Fig. 9 Forest plot describing the validity of ankle angle in the sagittal plane measured using IMU (Pearson correlation coefficient, *r*). Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. *SE* standard error, *IV* inverse variance, *CI* confidence interval

(See figure on next page.)

Fig. 10 Subgroup analysis describing the reliability of stance time measured using IMUs. A based on running speed, B based on location and C based on running surface. Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Deflandre et al.,2018e (8 km/h), 2018f (12 km/h), 2018 g (16 km/h) [37]; Gindre et al.,2016a (12 km/h), 2016b (15 km/h), 2016c (18 km/h), 2016d (21 km/h) [43]. *SE* standard error, *IV* inverse variance, *CI* confidence interval



A

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
1.1 Slow speed (≤15km/h)				
Deflandre et al.,2018e [37]	1.589	0.343	11.1%	1.59 [0.92, 2.26]
Deflandre et al.,2018f [37]	2.092	0.343	11.1%	2.09 [1.42, 2.76]
Gindre et al.,2016a [43]	2.826	0.232	13.7%	2.83 [2.37, 3.28]
Gindre et al.,2016b [43]	1.363	0.232	13.7%	1.36 [0.91, 1.82]
Subtotal (95% CI)			**49.6%**	**1.97 [1.24, 2.71]**
Heterogeneity: Tau² = 0.48; Chi² = 21.71, df = 3 (P < 0.0001); I² = 86%				
Test for overall effect: Z = 5.25 (P < 0.00001)				
1.2 Fast speed (>15km/h)				
Ammann et al.,2016a [32]	1.701	0.309	11.9%	1.70 [1.10, 2.31]
Deflandre et al.,2018g [37]	1.293	0.343	11.1%	1.29 [0.62, 1.97]
Gindre et al.,2016c [43]	1.792	0.232	13.7%	1.79 [1.34, 2.25]
Gindre et al.,2016d [43]	2.146	0.232	13.7%	2.15 [1.69, 2.60]
Subtotal (95% CI)			**50.4%**	**1.79 [1.47, 2.12]**
Heterogeneity: Tau² = 0.04; Chi² = 4.50, df = 3 (P = 0.21); I² = 33%				
Test for overall effect: Z = 10.76 (P < 0.00001)				
Total (95% CI)			**100.0%**	**1.87 [1.49, 2.24]**
Heterogeneity: Tau² = 0.21; Chi² = 27.29, df = 7 (P = 0.0003); I² = 74%				
Test for overall effect: Z = 9.75 (P < 0.00001)				
Test for subgroup differences: Chi² = 0.20, df = 1 (P = 0.66), I² = 0%				

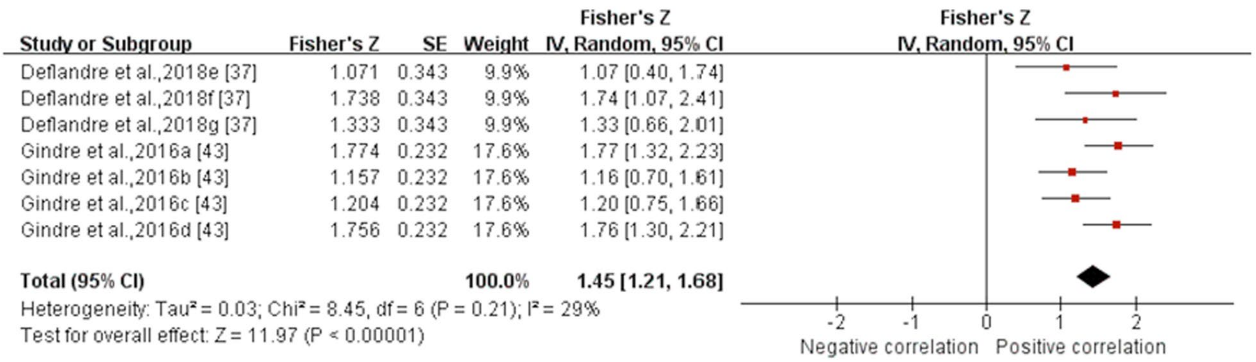
B

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
2.1 Waist				
Deflandre et al.,2018e [37]	1.589	0.343	11.1%	1.59 [0.92, 2.26]
Deflandre et al.,2018f [37]	2.092	0.343	11.1%	2.09 [1.42, 2.76]
Deflandre et al.,2018g [37]	1.293	0.343	11.1%	1.29 [0.62, 1.97]
Gindre et al.,2016a [43]	2.826	0.232	13.7%	2.83 [2.37, 3.28]
Gindre et al.,2016b [43]	1.363	0.232	13.7%	1.36 [0.91, 1.82]
Gindre et al.,2016c [43]	1.792	0.232	13.7%	1.79 [1.34, 2.25]
Gindre et al.,2016d [43]	2.146	0.232	13.7%	2.15 [1.69, 2.60]
Subtotal (95% CI)			**88.1%**	**1.89 [1.47, 2.31]**
Heterogeneity: Tau² = 0.25; Chi² = 26.77, df = 6 (P = 0.0002); I² = 78%				
Test for overall effect: Z = 8.75 (P < 0.00001)				
2.2 Shoelace				
Ammann et al.,2016a [32]	1.701	0.309	11.9%	1.70 [1.10, 2.31]
Subtotal (95% CI)			**11.9%**	**1.70 [1.10, 2.31]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 5.50 (P < 0.00001)				
Total (95% CI)			**100.0%**	**1.87 [1.49, 2.24]**
Heterogeneity: Tau² = 0.21; Chi² = 27.29, df = 7 (P = 0.0003); I² = 74%				
Test for overall effect: Z = 9.75 (P < 0.00001)				
Test for subgroup differences: Chi² = 0.25, df = 1 (P = 0.62), I² = 0%				

C

Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
3.1 Treadmill				
Deflandre et al.,2018e [37]	1.589	0.343	11.1%	1.59 [0.92, 2.26]
Deflandre et al.,2018f [37]	2.092	0.343	11.1%	2.09 [1.42, 2.76]
Deflandre et al.,2018g [37]	1.293	0.343	11.1%	1.29 [0.62, 1.97]
Subtotal (95% CI)			**33.2%**	**1.66 [1.20, 2.12]**
Heterogeneity: Tau² = 0.05; Chi² = 2.77, df = 2 (P = 0.25); I² = 28%				
Test for overall effect: Z = 7.11 (P < 0.00001)				
3.2 Ground				
Ammann et al.,2016a [32]	1.701	0.309	11.9%	1.70 [1.10, 2.31]
Gindre et al.,2016a [43]	2.826	0.232	13.7%	2.83 [2.37, 3.28]
Gindre et al.,2016b [43]	1.363	0.232	13.7%	1.36 [0.91, 1.82]
Gindre et al.,2016c [43]	1.792	0.232	13.7%	1.79 [1.34, 2.25]
Gindre et al.,2016d [43]	2.146	0.232	13.7%	2.15 [1.69, 2.60]
Subtotal (95% CI)			**66.8%**	**1.97 [1.46, 2.48]**
Heterogeneity: Tau² = 0.27; Chi² = 22.34, df = 4 (P = 0.0002); I² = 82%				
Test for overall effect: Z = 7.63 (P < 0.00001)				
Total (95% CI)			**100.0%**	**1.87 [1.49, 2.24]**
Heterogeneity: Tau² = 0.21; Chi² = 27.29, df = 7 (P = 0.0003); I² = 74%				
Test for overall effect: Z = 9.75 (P < 0.00001)				
Test for subgroup differences: Chi² = 0.81, df = 1 (P = 0.37), I² = 0%				

Fig. 10 (See legend on previous page.)



Study or Subgroup	Fisher's Z	SE	Weight	Fisher's Z IV, Random, 95% CI
Deflandre et al.,2018e [37]	1.071	0.343	9.9%	1.07 [0.40, 1.74]
Deflandre et al.,2018f [37]	1.738	0.343	9.9%	1.74 [1.07, 2.41]
Deflandre et al.,2018g [37]	1.333	0.343	9.9%	1.33 [0.66, 2.01]
Gindre et al.,2016a [43]	1.774	0.232	17.6%	1.77 [1.32, 2.23]
Gindre et al.,2016b [43]	1.157	0.232	17.6%	1.16 [0.70, 1.61]
Gindre et al.,2016c [43]	1.204	0.232	17.6%	1.20 [0.75, 1.66]
Gindre et al.,2016d [43]	1.756	0.232	17.6%	1.76 [1.30, 2.21]
Total (95% CI)			**100.0%**	**1.45 [1.21, 1.68]**

Heterogeneity: Tau² = 0.03; Chi² = 8.45, df = 6 (P = 0.21); I² = 29%
Test for overall effect: Z = 11.97 (P < 0.00001)

Fig. 11 Forest plot describing the reliability of step frequency measured using IMUs. Squares represent Fisher's Z; bars indicate 95% confidence intervals and diamonds as pooled data. Deflandre et al., 2018e (8 km/h), 2018f (12 km/h), 2018 g (16 km/h) [37]; Gindre et al., 2016a (12 km/h), 2016b (15 km/h), 2016c (18 km/h), 2016d (21 km/h) [43]. *SE* standard error, *IV* inverse variance, *CI* confidence interval.

Additionally, one study compared different algorithms of IMUs, and the results showed that results obtained by different algorithms vary greatly [54]. This means that the algorithm optimisation is one of the effective methods to improve the validity of measurements from IMUs. Previous studies have pointed out that the data type of IMU data and the corresponding calculation method areimportant factors in measurement errors [53, 64, 65]. IMUs included in the current review were provided by different manufacturers, it is difficult to summarize their calculation procedures, so this review cannot provide strong support for the view. One study compared five commercial IMUs: MilestonePod, Moov Now™, TgForce, Zoi and RunScribe™, although they showed excellent agreement in step frequency ($r \geq 0.955$) [38]. For the kinetic parameters, the results vary considerably (r values range from -0.532 to 0.813) [38]. The finding indicated that different IMUs do have a considerable effect on the measurements. Owing to the validity of the measurement from IMUs is affected by numerous factors, further research is needed in the future.

For homogeneity, the present systematic review and meta-analysis has only reviewed healthy adults. However, measuring subjects with RRIs not only provide evidence that supports the application of IMUs in real life but also allows the optimization of the motion assessment of different people for IMUs. By comparing the kinematic results obtained by the IMUs in healthy and injured subjects during running, the results may not only provide a better understanding of the specific biomechanical mechanisms underlying injuries but also may provide coaches or clinicians with early warning of the occurrence of RRIs. Meanwhile, Bergamini et al. [33] believed that IMUs can be used in monitoring the running of amateur and elite athletes, which was similar to the results in Schmidt et al. [51]. However, one study suggested that the use of IMUs in measuring continuous motion should be considered carefully [41]. The explanation for the contrary conclusion may be that different studies used different types of IMUs and placement. Some studies used IMUs in the continuous analysis of marathon running, and the results showed that IMUs can detect significant changes in running kinematics as mileage increased [66, 67]. These results showed the possibility of applying IMUs in the evaluation of running techniques over a long period of time in specific setting.

The included studies showed low to moderate methodological quality, with scores ranging between 12 and 20 out of 24. The lack of quality research reduces the ability to make any strong conclusions or clear recommendations in this review. Similar to previous study, the sample size was underpowered and/or unjustified in most of the literature which limits the statistical power of the available data [17]. Only two of the studies conducted an a priori power analysis for sample size [41, 52], and more than half of the included studies had a sample size of 12 or less. Furthermore, the paucity of use of appropriate statistical tests was also a prominent issue, with nearly half of the studies not reporting both absolute and relative statistical metrics, or Bland–Altman plots as a visual representation of agreement [68]. Considering these findings, more HQ studies are needed in the future.

Validity

In this review, although all the included studies compared the measurement results derived from IMUs with the reference systems, there were few data that could be quantitatively analysed, and most of them were gait spatiotemporal parameters. Meanwhile, all pooled outcomes had moderate to high heterogeneity ($I^2 \geq 61\%$). The reason for the high heterogeneity is that there were not adequate outcomes that could be subjected to meta-analysis, and most of them were from different velocity situations within the same study.

Among the gait spatiotemporal outcomes, the stance time has the most pooled data with moderate evidence and depicted moderate to excellent agreement. The reason for the relatively low agreement maybe that one of the included studies showed very poor ICCs among gait temporal outcomes, which also occurred in the flight time [37]. Considering the experimental settings and IMUs brands vary greatly among different studies, which is also an important source of heterogeneity, this study was not excluded from the meta-analysis. However, it should be interpreted with caution. Fewer studies evaluated the validity of the stride length, but still found excellent validity in all pooled data with moderate evidence, which is consistent with a previous review [17]. With regard to step frequency, the agreement was excellent and drawn from moderate evidence in the running speed range of 8 km/h to 21 km/h. As for running speed, the results showed a significant asymmetry of the left and right limbs [45], which directly led to running speed only showing good agreement in meta-analysis.

Regarding the validity of lower extremity joint kinematics, only the ankle angle in the sagittal plane can be quantified [17, 18]. It was suggested by moderate evidence that the agreement between the ankle angle in the sagittal plane obtained by the IMUs and the reference systems was excellent. One study showed that the IMUs' measurement results in the hip, knee and ankle joints distinctly improved after offset correction, with RMSEs between 18° and 28° reduced to between 5° and 8° [50]. In addition, compared with walking, the offset between waveforms increased during running, indicating that the motion amplitude would affect the lower limb joint kinematics derived from IMUs [50]. One explanation is that increase in the amplitude of movement may reduce the accuracy of the identification of the initial angle after calibration with the IMUs [50]. For joint discrete parameters, the RMSE of the hip joint (25.1°–36.1°) was greater than that of the knee joint (13.2°–20.0°) and ankle joint (14.4°–19.1°) [50]. Moreover, one study suggested that, compared to the optical motion capture systems, lower limb RMSEs for joint angles calculated using the IMUs data were less than 10° for all axes and more rapid motions involving larger ranges of motion would probably induce greater RMSEs [12]. For rearfoot range of motion, bias increased with velocity on the sagittal plane but had no effect on eversion [47].

Reliability

In contrast to validity, the reliability of measurements derived from IMUs during running was assessed by few studies. Similar to studies assessing validity, high quality research on reliability is lacking. Only stance time and step frequency were included in the meta-analysis. The agreement of stance time was excellent with relatively high heterogeneity ($I^2 = 74\%$). None of the subgroup analyses based on running speed, IMUs' position and running surface were statistically significant ($p \geq 0.37$) and there was excellent agreement across subgroups (summary ICC $\geq$ 0.930), suggesting that IMUs are robust in measuring gait temporal outcomes. Step frequency depicted good reliability with moderate heterogeneity ($I^2 = 29\%$), and sensitivity analysis showed that the results were stable. Only the test–retest reliability (within–tester reliability) of the measurements from IMUs was summarised because all included studies only reported the test–retest reliability, lacking studies on between–tester reliability. Among them, only one study evaluated the reliability between different days (three test sessions, with a time interval of 7 days between each test day) [37], and the remaining five studies only reported the agreement between the test and retest after a short time rest [32, 43, 44, 46, 50].

For reliability, flight time, step length, stride length, running speed and joint kinematics outcomes with the exception of ankle angle in the sagittal plane were only reported in one study, and thus evidence was limited. Furthermore, subgroup analysis based on running speed could not be performed for parameters other than stance time due to a lack of available data. However, for stance time, flight time, step length and stride length, the CV values increased with increasing running velocity [37, 43]. This finding implies that running speed affects the reliability of measurements from IMUs. RMSE did not show results similar to validity for lower extremity joint kinematics. This result indicated a high degree of consistency in the reliability of the hip, knee and ankle measurements obtained by IMUs. However, due to a limited number of studies were included in each parameter, therefore, these findings should be interpreted cautiously.

Calibration (alignment of the IMUs axes with the anatomical segment axes) enables the initial value of the IMUs to be in a prescribed standard state. It is an essential factor affecting IMU reliability, as different calibration protocols may result in substantially different measurements [69, 70]. In this review, only four studies described calibration procedures in detail, including static [48, 50] and functional movements [40, 44, 50]. Thus, it was unable to evaluate and summarise the calibration procedure, as in a previous study [71]. It is generally believed that a standardised measurement process for IMUs is necessary to the enhancement of the comparability among different studies.

Limitations

To enhance quality control, the search was restricted to fully peer-reviewed published articles, and thus relevant conference papers may have been excluded. Only included specific gait spatiotemporal parameters and joint kinematics parameters but did not include acceleration, impact, gait events and foot strike pattern, which may lead to selection bias. In the meta-analysis, only ICC and *r* values were pooled, and the studies that illustrated ICC values without specific values were excluded, and thus the number of articles that could be pooled was reduced. Furthermore, the level of heterogeneity was substantial in most parameters. Thus, these meta-analyses should be interpreted cautiously. In addition, none of the literature in this review was rated as high quality, and thus studies that had higher quality and investigated the validity and reliability of IMUs for lower extremity kinematics during running are needed.

Conclusion

Measuring running kinematics using IMUs helps in efficiently monitoring RRIs and evaluating running techniques in real-world settings. The findings of this review demonstrate that IMUs perform moderate to excellent correlation with gold standard for gait spatiotemporal parameters during running in healthy adults but should be reported with caution when lower extremity joint kinematics are assessed. Further, high quality literature on the validity and reliability of IMUs during running are lacking, and sample sizes seem generally underpowered. Thus, future studies should include more runners with different running skill levels and RRIs, as they may provide evidence that supports the application of IMUs in a variety of specific sports environments and provide the possibility for algorithm improvement. In addition, literature quality should be enhanced, and guidelines for the use of IMUs in running should be developed.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40798-022-00477-0.

Additional file 1. Complete search strategy.

Additional file 2. Complete inclusion and exclusion criteria and definition of spatiotemporal parameters.

Additional file 3. Qualitative summary of validity and reliability for biomechanical parameters.

Acknowledgements
Not applicable.

Author contributions

ZZ, YL and LW designed the research. ZZ and YL conducted the searches and screening process. ZZ, XH, MT and LW completed the full-text screening. YL, MT and LW assessed methodological quality. ZZ and MT extracted the data, which were checked by XH. ZZ and YL performed the statistical analysis and interpreted it. ZZ and YL wrote the manuscript with critical input from LW. All authors read and approved the final manuscript.

Funding
No sources of funding were used to assist in the preparation of this article.

Availability of data and materials

All data generated or analysed during this review are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Ziwei Zeng, Yue Liu, Xiaoyue Hu, Meihua Tang and Lin Wang declare that they have no competing interests.

Received: 23 November 2021 Accepted: 11 June 2022

References

1. Yeh HP, Stone JA, Churchill SM, Brymer E, Davids K. Physical and emotional benefits of different exercise environments designed for treadmill running. Int J Environ Res Public Health. 2017;14(7):752. https://doi.org/10.3390/ijerph14070752.
2. Hespanhol Junior LC, van Mechelen W, Postuma E, Verhagen E. Health and economic burden of running-related injuries in runners training for an event: a prospective cohort study. Scand J Med Sci Sports. 2016;26(9):1091–9. https://doi.org/10.1111/sms.12541.
3. van Gent RN, Siem D, van Middelkoop M, van Os AG, Bierma-Zeinstra SM, Koes BW. Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. Br J Sports Med. 2007;41(8):469–80. https://doi.org/10.1136/bjsm.2006.033548.
4. Saragiotto BT, Yamato TP, Hespanhol Junior LC, Rainbow MJ, Davis IS, Lopes AD. What are the main risk factors for running-related injuries? Sports Med. 2014;44(8):1153–63. https://doi.org/10.1007/s40279-014-0194-6.
5. Newman P, Witchalls J, Waddington G, Adams R. Risk factors associated with medial tibial stress syndrome in runners: a systematic review and meta-analysis. Open Access J Sports Med. 2013;4:229–41. https://doi.org/10.2147/OAJSM.S39331.
6. van Poppel D, van der Worp M, Slabbekoorn A, van den Heuvel SSP, van Middelkoop M, Koes BW, et al. Risk factors for overuse injuries in short- and long-distance running: a systematic review. J Sport Health Sci. 2021;10(1):14–28. https://doi.org/10.1016/j.jshs.2020.06.006.
7. Dingenen B, Blandford L, Comerford M, Staes F, Mottram S. The assessment of movement health in clinical practice: a multidimensional perspective. Phys Ther Sport. 2018;32:282–92. https://doi.org/10.1016/j.ptsp.2018.04.008.
8. Caldas R, Mundt M, Potthast W, de LimaNeto FB, Markert B. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. Gait Posture. 2017;57:204–10. https://doi.org/10.1016/j.gaitpost.2017.06.019.
9. Aroganam G, Manivannan N, Harrison D. Review on wearable technology sensors used in consumer sport applications. Sensors (Basel). 2019;19(9):1983. https://doi.org/10.3390/s19091983.
10. Vienne-Jumeau A, Quijoux F, Vidal PP, Ricard D. Value of gait analysis for measuring disease severity using inertial sensors in patients with multiple sclerosis: protocol for a systematic review and meta-analysis. Syst Rev. 2019;8(1):15. https://doi.org/10.1186/s13643-018-0918-z.

11. Ciuti G, Ricotti L, Menciassi A, Dario P. MEMS sensor technologies for human centred applications in healthcare, physical activities, safety and environmental sensing: a review on research activities in Italy. Sensors (Basel). 2015;15(3):6441–68. https://doi.org/10.3390/s150306441.
12. Mavor MP, Ross GB, Clouthier AL, Karakolis T, Graham RB. Validation of an IMU suit for military-based tasks. Sensors (Basel). 2020;20(15):4280. https://doi.org/10.3390/s20154280.
13. Kok M, Schön TB. Magnetometer calibration using inertial sensors. IEEE Sens J. 2016;16(14):5679–89. https://doi.org/10.1109/JSEN.2016.2569160.
14. Al-Amri M, Nicholas K, Button K, Sparkes V, Sheeran L, Davies JL. Inertial measurement units for clinical movement analysis: reliability and concurrent validity. Sensors (Basel). 2018;18(3):719. https://doi.org/10.3390/s18030719.
15. Wuest S, Masse F, Aminian K, Gonzenbach R, de Bruin ED. Reliability and validity of the inertial sensor-based Timed "Up and Go" test in individuals affected by stroke. J Rehabil Res Dev. 2016;53(5):599–610. https://doi.org/10.1682/JRRD.2015.04.0065.
16. Flachenecker F, Gaßner H, Hannik J, Lee DH, Flachenecker P, Winkler J, et al. Objective sensor-based gait measures reflect motor impairment in multiple sclerosis patients: Reliability and clinical validation of a wearable sensor device. Mult Scler Relat Disord. 2019;39: 101903. https://doi.org/10.1016/j.msard.2019.101903.
17. Kobsar D, Charlton JM, Tse CTF, Esculier JF, Graffos A, Krowchuk NM, et al. Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. J Neuroeng Rehabil. 2020;17(1):62. https://doi.org/10.1186/s12984-020-00685-3.
18. Horsley BJ, Tofari PJ, Halson SL, Kemp JG, Cormack SJ. Does site matter? impact of inertial measurement unit placement on the validity and reliability of stride variables during running: a systematic review and meta-analysis. Sports Med. 2021;51(7):1449–89. https://doi.org/10.1007/s40279-021-01443-8.
19. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372: n160. https://doi.org/10.1136/bmj.n160.
20. Law M, MacDermid J. Evidence-based rehabilitation: a guide to practice. 2nd ed. Thorofare: Slack Inc; 2008.
21. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33(2):363–74.
22. Cohen JA. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46. https://doi.org/10.1177/001316446002000104.
23. van Tulder M, Furlan A, Bombardier C, Bouter L. Editorial board of the cochrane collaboration back review G. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine. 2003;28(12):1290–9. https://doi.org/10.1097/01.BRS.0000065484.95996.AF.
24. Pasciuto I, Ligorio G, Bergamini E, Vannozzi G, Sabatini AM, Cappozzo A. How angular velocity features and different gyroscope noise types interact and determine orientation estimation accuracy. Sensors (Basel). 2015;15(9):23983–4001. https://doi.org/10.3390/s150923983.
25. Portney LG, Watkins MP. Foudantions of clinical research: applications to practice. 3rd ed. New Jersey: Prentice Hall; 2009. p. 524.
26. Higgins J, Thompson SG, Decks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–60. https://doi.org/10.1136/bmj.327.7414.557.
27. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006;11(2):193–206. https://doi.org/10.1037/1082-989X.11.2.193.
28. Harbord RM, Higgins JP. Meta-regression in Stata. Stata J. 2008;8:493–519.
29. Cooper H, Hedges L, Valentine J. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Russell Sage Foundation; 2009.
30. Fisher R. Statistical methods for research worker. 12th ed. New York: Hafner Publishing Company Inc; 1954.
31. Sprague AL, Smith AH, Knox P, Pohlig RT, Gravare SK. Modifiable risk factors for patellar tendinopathy in athletes: a systematic review and meta-analysis. Br J Sports Med. 2018;52(24):1575–85. https://doi.org/10.1136/bjsports-2017-099000.
32. Ammann R, Taube W, Wyss T. Accuracy of PARTwear inertial sensor and Optojump optical measurement system for measuring ground contact time during running. J Strength Cond Res. 2016;30(7):2057–63. https://doi.org/10.1519/JSC.0000000000001299.
33. Bergamini E, Picerno P, Pillet H, Natta F, Thoreux P, Camomilla V. Estimation of temporal parameters during sprint running using a trunk-mounted inertial measurement unit. J Biomech. 2012;45(6):1123–6. https://doi.org/10.1016/j.jbiomech.2011.12.020.
34. Brahms CM, Zhao Y, Gerhard D, Barden JM. Stride length determination during overground running using a single foot-mounted inertial measurement unit. J Biomech. 2018;71:302–5. https://doi.org/10.1016/j.jbiomech.2018.02.003.
35. Cooper G, Sheret I, McMillan L, Siliverdis K, Sha N, Hodgins D, et al. Inertial sensor-based knee flexion/extension angle estimation. J Biomech. 2009;42(16):2678–85. https://doi.org/10.1016/j.jbiomech.2009.08.004.
36. Day EM, Alcantara RS, McGeehan MA, Grabowski AM, Hahn ME. Low-pass filter cutoff frequency affects sacral-mounted inertial measurement unit estimations of peak vertical ground reaction force and contact time during treadmill running. J Biomech. 2021;119: 110323. https://doi.org/10.1016/j.jbiomech.2021.110323.
37. Deflandre D, Miny K, Schwartz C, Dardenne N, Leclerc AF, Bury T. Myotest efficiency in the mechanical analysis of the stride. Gazz Med Ital Arch S. 2018;177(6):293–300. https://doi.org/10.23736/S0393-3660.17.03568-9.
38. De Fontenay BP, Roy JS, Dubois B, Bouyer L, Esculier JF. Validating commercial wearable sensors for running gait parameters estimation. IEEE Sens J. 2020;20(14):7783–91. https://doi.org/10.1109/JSEN.2020.2982568.
39. Dorschky E, Nitschke M, Seifer AK, van den Bogert AJ, Eskofier BM. Estimation of gait kinematics and kinetics from inertial sensor data using optimal control of musculoskeletal models. J Biomech. 2019;95: 109278. https://doi.org/10.1016/j.jbiomech.2019.07.022.
40. Falbriard M, Meyer F, Mariani B, Millet GP, Aminian K. Accurate estimation of running temporal parameters using foot-worn inertial sensors. Front Physiol. 2018;9:610. https://doi.org/10.3389/fphys.2018.00610.
41. Fox JL, O'Grady CJ, Scanlan AT, Sargent C, Stanton R. Validity of the Polar Team Pro Sensor for measuring speed and distance indoors. J Sci Med Sport. 2019;22(11):1260–5. https://doi.org/10.1016/j.jsams.2019.06.012.
42. García-Pinillos F, Latorre-Román PÁ, Soto-Hermoso VM, Párraga-Montilla JA, Pantoja-Vallejo A, Ramírez-Campillo R, et al. Agreement between the spatiotemporal gait parameters from two different wearable devices and high-speed video analysis. PLoS ONE. 2019;14(9): e0222872. https://doi.org/10.1371/journal.pone.0222872.
43. Gindre C, Lussiana T, Hebert-Losier K, Morin J-B. Reliability and validity of the Myotest® for measuring running stride kinematics. J Sports Sci. 2016;34(7):664–70. https://doi.org/10.1080/02640414.2015.1068436.
44. Kim BH, Hong SH, Oh IW, Lee YW, Kee IH, Lee SY. Measurement of ankle joint movements using IMUs during running. Sensors (Basel). 2021;21(12):4240. https://doi.org/10.3390/s21124240.
45. Koldenhoven RM, Hertel J. Validation of a wearable sensor for measuring running biomechanics. Digit Biomark. 2018;2(2):74–8. https://doi.org/10.1159/000491645.
46. Konharn K, Eungpinichpong W, Promdee K, Sangpara P, Nongharnpitak S, Malila W, et al. Validity and reliability of smartphone applications for the assessment of walking and running in normal-weight and overweight/obese young adults. J Phys Act Health. 2016;13(12):1333–40. https://doi.org/10.1123/jpah.2015-0544.
47. Koska D, Gaudel J, Hein T, Maiwald C. Validation of an inertial measurement unit for the quantification of rearfoot kinematics during running. Gait Posture. 2018;64:135–40. https://doi.org/10.1016/j.gaitpost.2018.06.007.
48. Li R, Jumet B, Ren H, Song W, Tse ZTH. An inertial measurement unit tracking system for body movement in comparison with optical tracking. Proc Inst Mech Eng H. 2020;234(7):728–37. https://doi.org/10.1177/0954411920921695.
49. Mo S, Chow DHK. Accuracy of three methods in gait event detection during overground running. Gait Posture. 2018;59:93–8. https://doi.org/10.1016/j.gaitpost.2017.10.009.
50. Nuesch C, Roos E, Pagenstert G, Mundermann A. Measuring joint kinematics of treadmill walking and running: Comparison between an inertial sensor based system and a camera-based system. J Biomech. 2017;57:32–8. https://doi.org/10.1016/j.jbiomech.2017.03.015.
51. Schmidt M, Rheinländer C, Nolte KF, Wille S, Wehn N, Jaitner T. IMU-based determination of stance duration during sprinting. Proc Eng. 2016;147:747–52. https://doi.org/10.1016/j.proeng.2016.06.330.

52. Watari R, Hettinga B, Osis S, Ferber R. Validation of a torso-mounted accelerometer for measures of vertical oscillation and ground contact time during treadmill running. J Appl Biomech. 2016;32(3):306–10. https://doi.org/10.1123/jab.2015-0200.
53. Wouda FJ, Matteo G, Giovanni B, Erik M, Jasper R, Van BBF, et al. Estimation of vertical ground reaction forces and sagittal knee kinematics during running using three inertial sensors. Front Physiol. 2018;9:218. https://doi.org/10.3389/fphys.2018.00218.
54. Zrenner M, Gradl S, Jensen U, Ullrich M, Eskofier BM. Comparison of different algorithms for calculating velocity and stride length in running using inertial measurement units. Sensors (Basel). 2018;18(12):4194. https://doi.org/10.3390/s18124194.
55. Zrenner M, Küderle A, Roth N, Jensen U, Dümler B, Eskofier BM. Does the position of foot-mounted IMU sensors influence the accuracy of spatio-temporal parameters in endurance running? Sensors (Basel). 2020;20(19):5705. https://doi.org/10.3390/s20195705.
56. Hollis CR, Koldenhoven RM, Resch JE, Hertel J. Running biomechanics as measured by wearable sensors: effects of speed and surface. Sports Biomech. 2021;20(5):521–31. https://doi.org/10.1080/14763141.2019.1579366.
57. Ferrari A, Cutti AG, Garofalo P, Raggi M, Heijboer M, Cappello A, et al. First in vivo assessment of "Outwalk": a novel protocol for clinical gait analysis based on inertial and magnetic sensors. Med Biol Eng Comput. 2010;48(1):1–15. https://doi.org/10.1007/s11517-009-0544-y.
58. Napier C, Willy RW, Hannigan BC, McCann R, Menon C. The effect of footwear, running speed, and location on the validity of two commercially available inertial measurement units during running. Front Sports Act Living. 2021;3: 643385. https://doi.org/10.3389/fspor.2021.643385.
59. Edwards S, White S, Humphreys S, Robergs R, O'Dwyer N. Caution using data from triaxial accelerometers housed in player tracking units during running. J Sport Sci. 2019;37(7):810–8. https://doi.org/10.1080/02640414.2018.1527675.
60. Provot T, Chiementin X, Bolaers F, Murer S. Effect of running speed on temporal and frequency indicators from wearable MEMS accelerometers. Sports Biomech. 2021;20(7):831–43. https://doi.org/10.1080/14763141.2019.1607894.
61. Bigelow EM, Elvin NG, Elvin AA, Arnoczky SP. Peak impact accelerations during track and treadmill running. J Appl Biomech. 2013;29(5):639–44. https://doi.org/10.1123/jab.29.5.639.
62. Tao H, Joyce L, Kozak B, Luiken J, Wendt N. Spatiotemporal comparison of overground and treadmill running with pressure sensor insoles in division I collegiate runners. Int J Sports Phys Ther. 2019;14(5):731–9. https://doi.org/10.26603/ijspt20190731.
63. Strohrmann C, Harms H, Kappeler-Setz C, Troster G. Monitoring kinematic changes with fatigue in running using body-worn sensors. IEEE Trans Inf Technol Biomed. 2012;16(5):983–90. https://doi.org/10.1109/TITB.2012.2201950.
64. Scalera GM, Ferrarin M, Rabuffetti M. Gait regularity assessed by wearable sensors: comparison between accelerometer and gyroscope data for different sensor locations and walking speeds in healthy subjects. J Biomech. 2020;113: 110115. https://doi.org/10.1016/j.jbiomech.2020.110115.
65. Prateek GV, Mazzoni P, Earhart GM, Nehorai A. Gait cycle validation and segmentation using inertial sensors. IEEE Trans Biomed Eng. 2020;67(8):2132–44. https://doi.org/10.1109/TBME.2019.2955423.
66. Reenalda J, Maartens E, Homan L, Buurke JHJ. Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics. J Biomech. 2016;49(14):3362–7. https://doi.org/10.1016/j.jbiomech.2016.08.032.
67. Meyer F, Falbriard M, Mariani B, Aminian K, Millet GP. Continuous analysis of marathon running using inertial sensors: hitting two walls? Int J Sports Med. 2021;42(13):1182–90. https://doi.org/10.1055/a-1432-2336.
68. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. J Clin Epidemiol. 2011;64(1):96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002.
69. Mifsud NL, Kristensen NH, Villumsen M, Hansen J, Kersting UG. Portable inertial motion unit for continuous assessment of in-shoe foot movement. Proc Eng. 2014;72:208–13. https://doi.org/10.1016/j.proeng.2014.06.035.
70. Rouhani H, Favre J, Crevoisier X, Aminian K. Measurement of multi-segment foot joint angles during gait using a wearable system. J Biomech Eng. 2012;134(6): 061006. https://doi.org/10.1115/1.4006674.
71. Poitras I, Dupuis F, Bielmann M, Campeau-Lecours A, Mercier C, Bouyer LJ, et al. Validity and reliability of wearable sensors for joint angle estimation: a systematic review. Sensors (Basel). 2019;19(7):1555. https://doi.org/10.3390/s19071555.

Publisher's Note