Bas Van Hooren Paul Willems Guy Plasqui Kenneth Meijer Department of Nutrition and Movement Sciences, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands



The accuracy of commercially available instrumented insoles (ARION) for measuring spatiotemporal running metrics



Revised: 9 May 2023

ORIGINAL ARTICLE



WILEY

The accuracy of commercially available instrumented insoles (ARION) for measuring spatiotemporal running metrics

Bas Van Hooren 💿 | Paul Willems | Guy Plasqui | Kenneth Meijer

Department of Nutrition and Movement Sciences, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands

Correspondence

Bas Van Hooren, Department of Nutrition and Movement Sciences, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Universiteitssingel 50, Maastricht 6229 ER, The Netherlands. Email: basvanhooren@hotmail.com

Funding information

Eurostars, Grant/Award Number: 12912

Spatiotemporal metrics such as step frequency have been associated with running injuries in some studies. Wearables can measure these metrics and provide realtime feedback in-field, but are often not validated. This study assessed the validity of commercially available wireless instrumented insoles (ARION) for quantifying spatiotemporal metrics during level running at different speeds $(2.78-5.0 \text{ m s}^{-1})$ and slopes (3° and 6° up/downhill) to an instrumented treadmill. Mean raw, percentage and absolute percentage error, and limits of agreement (LoA) were calculated. Agreement was statistically quantified using four thresholds: excellent, <5%; good, <10%; acceptable, <15%; and poor, >15% error. Excellent agreement (<5% error) was achieved for stride time across all conditions, and for step frequency across all but one condition with good agreement. Contact time and swing time generally showed at least good agreement. The mean difference across all conditions was -0.95% for contact time, 0.11% for stride time, 0.6% for swing time, -0.11% for step frequency, and -0.09% when averaged across all outcomes and conditions. The accuracy at an individual level was generally good to excellent, being <10% for all but two conditions, with these conditions being <15%. Additional experiments among four runners showed that step length could also be measured with an accuracy of 1.76% across different speeds with an updated version of the insoles. These findings suggests that the ARION wearable may not only be useful for large-scale in-field studies investigating group differences, but also to quantify spatiotemporal metrics with generally good to excellent accuracy for individual runners.

KEYWORDS

accelerometer, agreement, biomechanics, pressure insoles, validity, wearable technology

1 | INTRODUCTION

With approximately 50 million participants in Europe and more than 65 million participants in the United States,

running is one of the most popular sporting activities, and its popularity is still rising. A recent review found the strongest evidence for health benefits of running when compared to other sports.¹ Running is therefore a highly

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. Scandinavian Journal of Medicine & Science In Sports published by John Wiley & Sons Ltd.

effective and low-cost health intervention. Dropout rates of up to 50% have, however, been reported during running intervention programs,² and this dropout increases the risk of developing various adverse health conditions.

Running injuries are the primary cause of dropout,³ and researchers have therefore explored various approaches to reduce running injuries such as load management by means of training programs,^{4,5} online education,^{6,7} and modification of running technique (i.e., gait retraining).⁸ Among these approaches, gait retraining has shown more promising effects than other approaches for reducing running injury risk,^{4–9} although the evidence is limited. Gait retraining is, however, typically applied in a laboratory environment, with expensive motion capture equipment to analyze and provide real-time feedback on running technique. Moreover, participants are typically required to visit the laboratory on multiple occasions. These requirements reduce the ecological validity and applicability of the findings as most runners do not have access to this equipment and may not have time to visit a laboratory multiple times. Further, the modified running technique observed in a laboratory may not fully translate to outdoor running¹⁰ and may also partly return to baseline without periodic gait retraining,¹¹⁻¹³ both of which reduce the effectiveness of laboratory-based interventions.

Wearables offer a promising method to quantify running technique and training intensity outside of the laboratory, and provide real-time feedback on these aspects.¹⁴ This feedback may in turn reduce injury rates and enhance running performance. Although wearables can typically only measure spatiotemporal metrics such as step frequency and contact time, some studies have shown associations between spatiotemporal measures and running injuries,^{14,15} suggesting measurement of these metrics may also be relevant for injury prevention. For example, in a recent randomized controlled trial, Malisoux, Gette, Delattre, Urhausen, and Theisen¹⁶ found that shorter contact times and larger step lengths were associated with a higher overall injury rate. Additionally, Lempke and colleagues showed that individuals with exerciserelated lower leg pain exhibited a longer contact time than symptom-free individuals.¹⁷ Moreover, quantification of spatiotemporal metrics may also be beneficial to optimize performance. Real-time feedback on step frequency may for example improve running economy for runners who are running with a step frequency below their most economical step frequency.^{18–21}

Accurate data and consistent functioning of wearables is, however, important to capture small differences across multiple sessions, and to ensure actual usage by end users.^{14,22,23} However, wearables are often not validated, or validated during conditions that do not well reflect infield use (e.g., only one speed and level running surface). Moreover, some wearables also have limitations for infield use for example because they cause interference with running technique (e.g., for some pressure insoles²⁴), or because they do not provide real-time feedback. Recently, instrumented insoles compromising a tri-axial accelerometer and gyroscope, eight spatially distributed pressure sensors, global positioning system and temperature sensor have been developed by ATO-Gear (Eindhoven, The Netherlands), Figure 1. These insoles can gather data up to 5-6h consecutively and are therefore capable of quantifying spatiotemporal metrics for a large number of steps during prolonged training sessions or competitions. Due to the light-weight (<30 g each) microprocessor attached to the shoe, wireless set-up, thickness (<2 mm), and flexibility of the insoles, the device does likely not interfere with running technique in contrast to other insoles.²⁴ Further, due to the combination with a smartphone application



FIGURE 1 Instrumented insoles by ARION consist of eight pressure sensors located at the medial and lateral heel, lateral mid-foot, first, second, and fourth metatarsal head, phalanges and the hallux, and sample data at 150 Hz (left image). The sampling frequency of the pressure sensors can, however, be configured up to 250 Hz. These insoles are connected to a microprocessor by a small flat cable. This microprocessor comprises an accelerometer, gyroscope, and a temperature sensor and is clipped to the lateral side of the shoe (right image) to avoid interference with the running movement and to measure additional biomechanical parameters. The collected data are transmitted in real time via Bluetooth to a mobile phone for real-time feedback.

or smartwatch, this wearable can be used to provide realtime feedback on spatiotemporal metrics, in contrast to other wearables that typically only provide a summary of metrics after a run. As such, this wearable has the potential to accurately measure and modify spatiotemporal metrics by real-time in-field feedback and thereby overcomes several important limitations of previous wearables. This feedback may in turn lead to less injuries and an improved performance and motivation, eventually leading to less dropout.¹⁴ An older version of ARION has been shown to exhibit a high level of agreement (overall mean difference of 1.2%) on spatiotemporal parameters compared to an instrumented treadmill.²⁵ However, this validation was performed at only one speed $(3.19 \,\mathrm{m\,s}^{-1})$ and with 0° incline. This limits the generalizability of these findings to overground in-field running, which involves a variety of speeds and slopes. The aim of this study was therefore to validate spatiotemporal parameters of the ARION instrumented insoles to an instrumented treadmill during level running at different speeds and during uphill and downhill running to better reflect in-field conditions.

2 | METHODS

2.1 | Participants

Nineteen participants (10 males and 9 females, mean \pm SD age 23.6 \pm 3.7 years, body height 174.9 \pm 9.2 m; body mass 67.2 \pm 10.4 kg) that were free of any moderate (for previous 3 months) or minor (for previous 1 month) musculoskeletal injuries were aged 18–45, comfortable with treadmill running and had a body mass index (BMI) of <26 volunteered to participate in this study. The study was approved by the local ethics committee (nr. 2019-1138) was conducted according to the declaration of Helsinki, and all participants signed informed consent form prior to the measurements. Mean \pm SD weekly training distance was 49.5 \pm 46.5 km (range 0 to 140 km), and training experience was 5.2 \pm 5.2 years (range 0 to 15 year). The sample size calculation is reported in the Supplementary File.

2.2 | General design of the study

All participants completed a single test session and were instructed to avoid strenuous activity for 36h, alcohol for 24h, caffeine for 6h, and a heavy meal 1h before the session. When entering the laboratory, anthropometric measurements were taken using standardized procedures and the participants completed a questionnaire about their weekly training volume, running experience, and seasonal best times. The participants were then equipped

2.3 | Instruments

Each participant wore their own habitual training shoes and were equipped with a pair of appropriate sized ARION insoles (Figure 1) as specified in the Supplementary File. Although the ARION insoles can be configured to sample at frequencies up to 250 Hz, we set the sampling frequency of the pressure sensors within the insoles to 150 Hz to optimize the available bandwidth for the laboratory test conditions with potential radio interference The computer assisted rehabilitation environment (CAREN, Motek, The Netherlands) system combines an instrumented split-belt treadmill (belt length and width 2.15×0.5m, 6.28-kW motor per belt, 60 Hz belt speed update frequency, and $0-18 \,\mathrm{km}\,\mathrm{h}^{-1}$ speed range) with three-dimensional motion capture and was used as golden-standard for spatiotemporal outcomes. Ground reaction forces were collected at 1000 Hz and filtered using a 2nd order Butterworth filter with a low-pass filter of 50 Hz.

2.4 Data collection

Prior to data collection, the participants were instructed to run for 8 min at a fixed-paced speed of $2.78 \,\mathrm{m \, s^{-1}}$ to familiarize themselves with treadmill running.²⁶ This was followed by 4 min of running at $3.33 \,\mathrm{m \, s^{-1}}$. Both conditions were performed with the treadmill at 0° inclination. The participants then completed a series of 1-min runs at different fixed-paced speeds and treadmill slopes (Table SI), with the order of conditions being randomized by an online research randomizer (https://www.randomizer. org/). After these conditions, the participants ran another two trials at $3.33 \,\mathrm{m \, s^{-1}}$, but with a higher and lower step frequency $(\pm 10 \text{ steps min}^{-1})$ compared to their step frequency during the 4 min trial at $3.33 \,\mathrm{m \, s^{-1}}$. The order of these conditions was also randomized. The incline and decline grades were chosen as they are thought to reflect the maximum incline and decline typically encountered by runners in real-world conditions,²⁷ with the speed $(2.78 \,\mathrm{m\,s}^{-1})$ during these conditions reflecting a typical running speed for recreational runners. The level running speeds were chosen to reflect slow (2.78 m s^{-1}) , comfortable (3.0 and 3.33 ms^{-1}), fast (4.0 m s⁻¹), and very fast (5.0 m s^{-1}) training speeds for most recreationally active runners.²⁸ A rest period of 1–2 min was provided between each trial, but longer rest periods were provided when required. Further, participants were instructed to run as if they were running outside and were instructed to focus on the simulated virtual environment (Supplementary File).

2.5 | Data processing

2.5.1 | Insole

Previous research has found similar validity when separately analyzing the left and right leg.^{29,30} Preliminary analysis of the left and right leg data confirmed these findings, and only the right leg was therefore used for assessing the validity of the ARION system. Note that for step frequency data from both the left and right leg were used. Proprietary algorithms computed the following spatiotemporal outcomes: ground contact time (ms), flight time (ms), stride time (ms), and step frequency (steps min^{-1}). Ground contact time represented the time between foot contact and toe-off, swing time represented the time between toe-off and ground contact of the same foot, stride time represented the time from one foot contact to the next contact of the same foot, and step frequency was taken as the number of steps per minute. Although the ARION system also determined step length, the hardware (i.e., range of accelerometer and sampling frequency of gyroscope and accelerometer) and algorithms relevant to this metric were updated after data collection. The accuracy of this metric as measured during the study would therefore not reflect the accuracy of the metric in the current wearable. We therefore performed additional experiments among four participants to determine the accuracy of step length with the updated hardware and algorithms (see Supplementary File for details). The temporal algorithms solely used pressure data, while spatial outcomes also used the accelerometer and gyroscope (both by default at 75 Hz). Note that the accelerometer and gyroscope were also used for activity classification (i.e., running vs. walking).

2.5.2 | CAREN system

The filtered kinetic data were further analyzed using custom-written algorithms in Matlab. Ground contact time was determined using the vertical ground reaction force (vGRF) from force plate data from the CAREN system as this kinetic based approach is widely considered the most accurate way to determine gait events.³¹⁻³³ Footstrike (FS) was identified when vGRF exceeded

50 N and toe-off (TO) was identified when vGRF dropped below 50 N. This threshold has been applied in numerous other studies,³⁴⁻³⁶ and yields very close agreement to a lower threshold (e.g., 20 N) or thresholds that use a percentage of body weight.³⁷ Spatiotemporal parameters were determined using similar procedures as for the insole. ARION and Caren data were synchronized using the vertical ground reaction force impact during the landing of a vertical jump performed prior to each condition. Mean spatiotemporal parameters determined over 60 s of constant-speed running were used for analysis.

2.5.3 | Application of a moving average filter

Primary analyses of spatiotemporal data revealed that the ARION wearable showed substantial differences in spatiotemporal outcomes compared to the Caren system for some conditions of some participants. Inspection of the 60-s time-series data revealed this was often due to outliers in the single-step data (Supplementary File I), potentially related to interruption of the Bluetooth connection between the wearable sensors and data acquisition application. We applied a 10-datapoint moving average filter in Matlab to the spatiotemporal ARION parameters (e.g., contact time and step frequency values; not on the raw data) to remove these outliers, and all presented outcomes are therefore filtered using this filter. Note that this filter has also been implemented in the wearable application after data processing, and the compared data thus reflect the data that would be obtained in-field. A total number of 12111 to 12774 steps were included in analysis depending on the outcome considered, and 0.79%-1.01% of the steps were identified as outliers (corresponding to an absolute number of 97-133 steps).

After application of the filter, one participant still showed a substantial difference in the 60-s averaged data at 5 m s^{-1} . Inspection of the data revealed a substantial difference in the gold-standard (i.e., Caren system) outcomes for this participant as compared to the other participants, likely due to an error in the data acquisition. Based on this observation, the outcomes from this subject at this single speed were removed from the analysis since this would not provide accurate information for comparison.

2.6 | Statistical analysis

All statistical analysis were performed in R studio version 3.6.1. Mean bias between the ARION and CAREN system was assessed using separate Bland–Altman analyses for each outcome and condition in the original units and percentage differences. To objectively assess the agreement between the two systems, we used a statistical approach proposed by Shieh³⁸ with the percentage difference as the unit for comparison. In this method, the mean difference and variability of the difference between the systems are assessed in relation to an a priori determined threshold, whereby a specified proportion of the data should fall within the threshold to declare agreement. We considered four agreement thresholds: A difference of <5% was considered excellent agreement, <10% good, <15% acceptable, and $\geq 15\%$ poor agreement. These thresholds are in line with recommended accuracy thresholds for consumer and clinical wearables.³⁹ The central null-proportion (reflecting the amount of datapoints that should fall within this threshold) was set to 0.95 in line with the widely used 95% limits of agreement, and the alpha level to 0.05. Therefore, if 95% of the data points (i.e., differences) between the systems fell within the specified threshold, the null hypothesis that there is no agreement between the systems was rejected.

Mean absolute percent errors (MAPE) were also calculated to provide an indication of overall measurement error. MAPE were calculated as the average of absolute difference between the wearable system and the criterion measure divided by the criterion measure value, multiplied by 100. This is a more conservative estimate of error that takes into account both overestimation and underestimation because the absolute error value is used in the calculation and may therefore be particularly useful to assess the accuracy at an individual level.

Flexible materials such as the pressure sensors are inherently viscoelastic, making them prone to drift when they are deformed due to intermittent or continuous loading. For example, continuous standing in between the running trials may introduce ink shits and hereby partially mimic the effect of ink shifts that may occur with continuous running. Sensor drift was therefore assessed for all outcomes using the first, middle (6th), and last (11th) condition by assessing the percentage difference across these conditions. An increased difference from the first to the last condition was considered indictive of drift in the wearable. This was assessed by testing if the slope of a repeated-measures mixed model with diagonal unstructured covariate structure and subjects modeled as a random slope significantly differed from zero.

Finally, we also assessed if the difference between the gold-standard system and ARION wearable was consistently higher or lower for each individual relative to others across conditions using a mean rating two-way random model intraclass correlation coefficient (ICC) for consistency (using the "psych" package). In other words, this method allowed us to assess whether the ARION wearable has approximately the same difference across multiple conditions for an individual, despite a potential systematic difference relative to the CAREN system. The ICC was considered <0.69 poor; 0.7–0.79; acceptable; 0.8–0.89, good; and 0.9–0.99, excellent.⁴⁰ 95% confidence intervals were also computed and considered 0 if they were estimated to be negative.⁴¹ Finally, we computed a weighted mean percentage difference over all conditions and outcomes, and also per outcome across conditions.

3 | RESULTS

3.1 Concurrent validity

For all outcomes, mean \pm SD values from the CAREN system and insole as well as the mean \pm SD percentage difference with 95% limits of agreement and results from the agreement test are reported in Tables 1, 2, and SV. The differences in original units are reported in Tables SII–SIV, while Tables SVI–SVIII report the MAPE. Figure 2 shows the mean percentage difference and variability for each outcome and condition. The weighted mean difference across all conditions and outcomes was -0.09%, and -0.95% for contact time, 0.11% for stride time, 0.60% for swing time, and -0.11% for step frequency. All data is also available from the OpenScience framework at DOI 10.17605/OSF.IO/VBE9S.

3.2 Sensor drift

Supplementary File I shows the difference for all outcomes for the first, middle (6th), and last (11th) condition. Overall, there is no clear trend for larger mean differences across all outcomes from the first to the last condition, suggesting no relevant effect of sensor drift during the ~1.5 h experiments. Indeed, the mixed model indicated that the slope of the regression line did not significantly differ from zero for step frequency (p=0.87), swing time (p=0.45), stride time (p=0.58), or contact time (p=0.11).

3.3 | Consistency

ICC values for the consistency of the difference across trials ranged from poor to good and were 0.86 (95% confidence interval 0.77-0.93), <0.01 (<0.01 to 0.49), 0.74 (0.56-0.86), and <0.01 (<0.01-0.48) for contact time, stride time, swing time, and step frequency, respectively.

TABLE 1 Mean \pm SD values and mean \pm SD percentage difference with 95% limits of agreement for the measured spatiotemporal outcomes during level running at different speeds.

Outcome (m s ⁻¹)	n	CAREN (mean±SD)	ARION (mean±SD)	Mean±SD percentage difference [95% limits of agreement]
Contact time (ms)				
2.78	19	257 ± 23	255 ± 23	-0.81 ± 3.62 [-7.90 to 6.29]**
3.00	19	240 ± 21	239 ± 22	$-0.48 \pm 3.78 \ [-7.88 \text{ to } 6.93]^{**}$
3.33	19	238 ± 18	236 ± 19	-1.05 ± 3.82 [-8.53 to 6.44]*
4.0	19	211 ± 18	210 ± 19	$-0.36 \pm 2.08 \ [-4.43 \text{ to } 3.70]^{**}$
5.0	18	186 ± 18	184 ± 16	-0.91 ± 5.31 [-11.3 to 9.51]*
Stride time (ms)				
2.78	19	738 ± 29	738 ± 30	0.02 ± 0.36 [-0.69 to 0.73]***
3.00	19	715 ± 31	720 ± 29	0.77±1.44 [-2.06 to 3.59]***
3.33	19	715 ± 31	715 ± 30	$0.01 \pm 0.35 \ [-0.67 \text{ to } 0.69]^{***}$
4.0	19	689 ± 36	689 ± 36	$0.05 \pm 0.35 \ [-0.64 \text{ to } 0.74]^{***}$
5.0	18	639 ± 36	643±35	$0.59 \pm 1.08 \ [-1.53 \text{ to } 2.72]^{***}$
Swing time (ms)				
2.78	19	480 ± 27	483 ± 28	$0.58 \pm 1.94 [-3.21 \text{ to } 4.38]^{**}$
3.00	19	473 ± 33	480 ± 34	$1.32 \pm 2.23 [-3.05 \text{ to } 5.70]^{**}$
3.33	19	475 ± 30	479±31	0.79±1.61 [-2.37 to 3.95]***
4.0	19	479 ± 34	479 ± 33	$0.07 \pm 0.73 \ [-1.37 \text{ to } 1.51]^{***}$
5.0	18	453 ± 40	459 ± 35	$1.38 \pm 2.59 [-3.70 \text{ to } 6.46]^{**}$
Step frequency (steps min^{-1})				
2.78	19	163 ± 7	163 ± 7	-0.06 ± 0.06 [-0.18 to 0.06]***
3.00	19	168 ± 7	167±7	-0.85 ± 1.36 [-3.52 to 1.82]***
3.33	19	168 ± 7	168 ± 7	-0.04 ± 0.25 [-0.54 to 0.45]***
4.0	19	175±9	175±9	-0.03 ± 0.12 [-0.26 to 0.21]***
5.0	18	189 ± 10	187 ± 10	-0.73 ± 1.71 [-4.08 to 2.61]**

Note: No star indicates poor agreement (≥15% error).

***Excellent agreement (<5% error); **Good agreement (<10% error); *Acceptable agreement (<15% error).

4 | DISCUSSION

-WILEY

The aim of this study was to assess the validity of an instrumented insole for assessing spatiotemporal parameters during running. The primary finding is that the mean percentage difference between ARION and the gold-standard system is generally smaller than the 5% threshold (overall mean difference of -0.09%), reflecting excellent agreement on a group level. Specifically, for all but six outcomes across all conditions, at least an excellent to good level of agreement (<10% error) was achieved. Excellent agreement (<5% error) was achieved for stride time across all conditions, and for step frequency across all but one condition were this outcome showed good agreement. Contact time showed good agreement for 8 of 11 conditions, with the remaining 3 conditions showing acceptable agreement. Swing time showed excellent agreement in six conditions, good agreement in four conditions, and acceptable agreement in one condition. Similar findings

were observed when using the more conservative MAPE instead of percentage difference, which therefore shows that also at an individual level the ARION wearable generally exhibits good to excellent accuracy. In further support of this, outcomes with lower yet still acceptable agreement showed good to acceptable ICC values, indicating that the difference remained largely consistent across conditions for each participant. Additional experiments among four subjects with an updated ARION insole also showed excellent accuracy for step length (<2% error overall).

4.1 | Comparison to other wearables

The overall mean difference of -0.09% is an improvement in comparison with a previous study on an earlier version of ARION, that showed an overall mean difference of 1.2% on spatiotemporal parameters compared to an instrumented treadmill among sixteen individuals.²⁵

FABLE 2	Mean \pm SD values and mean \pm SD percentage difference with 95% limits of agreement for the measured spatiotemporal
outcomes du	ring sloped running.

				Mean \pm SD percentage difference			
Outcome	n	CAREN (mean \pm SD)	ARION (mean \pm SD)	[95% limits of agreement]			
Contact time (ms)							
+3 degrees up	19	261 ± 25	258 ± 22	-0.79±3.08 [-6.83 to 5.25]**			
+6 degrees up	19	263 ± 25	261 ± 23	$-0.68 \pm 3.09 \ [-6.73 \text{ to } 5.36]^{**}$			
-3 degrees down	19	254 ± 27	249 ± 27	$-1.82 \pm 2.76 [-7.23 \text{ to } 3.60]^{**}$			
-6 degrees down	19	249 ± 29	244 ± 29	$-2.26 \pm 3.49 \ [-9.10 \text{ to } 4.58]^*$			
Stride time (ms)							
+3 degrees up	19	725 ± 27	726 ± 29	$0.03 \pm 0.58 [-1.11 \text{ to } 1.18]^{***}$			
+6 degrees up	19	713 ± 32	712 ± 32	-0.05 ± 0.65 [-1.34 to 1.23]***			
-3 degrees down	19	740 ± 32	743 ± 32	$0.36 \pm 1.32 [-2.22 \text{ to } 2.94]^{***}$			
-6 degrees down	19	744±33	744 ± 34	$0.07 \pm 0.45 \ [-0.81 \text{ to } 0.95]^{***}$			
Swing time (ms)							
+3 degrees up	19	465 ± 29	467 ± 29	$0.40 \pm 1.71 [-2.95 \text{ to } 3.76]^{***}$			
+6 degrees up	19	451 ± 28	451 ± 27	$0.11 \pm 1.32 [-2.47 \text{ to } 2.70]^{***}$			
-3 degrees down	19	486 ± 29	493 ± 28	$1.41 \pm 2.51 [-3.52 \text{ to } 6.34]^{**}$			
-6 degrees down	19	495±33	497 ± 31	0.22±3.97 [-7.55 to 7.99]*			
Step frequency (steps \min^{-1})							
+3 degrees up	19	166 ± 7	166 ± 7	$0.00 \pm 0.19 \ [-0.37 \text{ to } 0.37]^{***}$			
+6 degrees up	19	169 ± 8	169 ± 8	$0.05 \pm 0.37 \ [-0.67 \text{ to } 0.77]^{***}$			
-3 degrees down	19	162 ± 7	162 ± 7	-0.33 ± 1.24 [-2.76 to 2.09]***			
-6 degrees down	19	162 ±7	162 ± 7	$-0.02 \pm 0.10 [-0.21 \text{ to } 0.17]^{***}$			

Note: No star indicates poor agreement (≥15% error).

***Excellent agreement (<5% error); **Good agreement (<10% error); *Acceptable agreement (<15% error).

This validation was notably performed at only one speed (3.19 m s^{-1}) and with 0° incline, while our study used a substantially more comprehensive validation protocol. The differences in accuracy likely reflect different approaches to data analysis as well as updates to the proprietary hardware designs and algorithms in the ARION wearable technology. For example, with regard to different data analysis approaches, Mann and co-workers²⁵ used the same approach to determine gait events in both the instrumented treadmill and insole based on the derivate of the pressure signal, while we used an approach based on the vertical ground reaction force in the treadmill, which is considered the gold-standard method to determine initial contact and toe-off in running.^{31–33}

When compared to other insoles, the ARION wearable generally performs slightly better in terms of the mean difference and variability of this difference. For example, Burns, Deneweth Zendler, and Zernicke²⁹ showed that the Loadsol overestimated contact time by about 20 ms (~7%) when compared to an instrumented treadmill during running at 2.78 m s^{-1} . Importantly, the limits of agreement were also relatively broad, ranging from 0 to 40 ms overestimation. In another study, the Loadsol was

compared to a force plate while running at a self-selected speed without shoes⁴² and showed only a very minor overestimation of ground contact time by $2.3 \,\mathrm{ms}$ (0.6%), with limits of agreement from -6.5 to 11 ms (2.4%), indicating high precision on a group and individual level. Another wearable insole system (OpenGo) has been shown to underestimate ground contact time by 10ms (2%) when compared to a force plate during running at a self-selected speed,⁴³ with limits of agreement ranging up to 40 ms (12%). Conversely, the PedarX system overestimated ground contact time by 30 ms (8%) compared to the force plate, with limits of agreement up to 30 ms (8%). The latter study also observed smaller agreement with shorter ground contact times, while our study did not show clear trends for smaller agreement or higher variability with higher speeds (Figure 2 and Table 1). Step frequency represented the most accurate outcome, with a mean raw and percentage differences of 0.29 step min⁻¹ and -0.01%, respectively (MAPE of 0.27% across all conditions). This is in line with other wearables that also reported errors of $\leq 1\%$ for step frequency,^{44,45} although higher errors have also been reported.⁴⁶ Step length was initially not assessed as the hardware and algorithms were updated after data



FIGURE 2 Mean difference and the variability for all outcomes during level running at increasing speeds (left), uphill and downhill running at different slopes (center) and running with a reduced and increased step frequency (right). Top row: contact time; second row: stride time; third row: swing time; fourth row: step frequency. The bold horizontal line depicts the mean group difference, while the white dots represent the average differences over the analyzed period per participant. Colors reflect the threshold for agreement: dark green, excellent agreement (<5%); light green, good agreement (<10%); yellow, acceptable agreement (<15%); and red, poor agreement (≥15%).

collection and the accuracy of step length as measured during the experiments would therefore not reflect the accuracy of this metric in the wearable currently on the market. Instead, we performed additional experiments

among four subjects to assess the accuracy of step length in the updated ARION version (Tables SIX–SX and Figure SV). When averaged over all participants, speeds and different shoes used for each participant, this metric also showed excellent accuracy, with an absolute percentage error of $1.76 \pm 0.49\%$ (corresponding to a difference of -0.74 cm). This error is considerably smaller compared to Runscribe when assessed against 3D motion capture (70– 80 cm underestimation of stride length; corresponding to 13%-15%).³⁰

Although drift can affect the accuracy of wearables over the duration of a session, we found no evidence of sensor drift (i.e., a larger difference between the ARION wearable and instrumented treadmill) from the first to last condition (Supplementary File I), suggesting the accuracy remains similar over the duration of an intermittent session up to ~1.5 h. This is in line with previous research on the LoadSol that for example demonstrated a very small drift of $0.34 \,\mathrm{N}\,\mathrm{step}^{-1}$ in the vertical impulse during running.^{29,42} Future research is, however, required to investigate the effects of drift during a prolonged continuous running session.

4.2 | Analysis of sensor performance characteristics

There are several potential explanations for the differences in spatiotemporal outcomes obtained with the ARION system as compared to the instrumented treadmill. First, although the ARION insoles can be configured to sample at frequencies up to 250 Hz, we set the sampling frequency of the insoles to 150 Hz to optimize the available bandwidth for the laboratory test conditions with potential radio interference. Moreover, this setting also best reflects in-field use were higher sampling rates may reduce battery life. Differences in sampling frequency of the insole (150 Hz) compared to the instrumented treadmill (1000 Hz) are likely compensated by the signal processing methodology (see second paragraph in the perspective section) but may to a limited extend still partly explain differences in the available temporal resolution to detect initial contact and toe-off and hence influence spatiotemporal metrics such as ground contact time for a single step. Moreover, the effective sampling frequency could be further influenced by missing data packets caused by interfering nearby electromagnetic radiation in the laboratory environment.⁴⁷ In support of this, higher sampling frequencies (200 Hz vs. 100 Hz) have been shown to improve the validity of ground reaction force metrics measured by pressure-sensitive insoles.⁴⁷ The differences in sampling frequency may also explain the higher accuracy of the ARION system as compared to other insoles as most other insoles use a sampling frequency that is lower than the ARION system (e.g., 100 Hz for LoadSol^{29,42} and 50 Hz for PedarX and OpenGo⁴³). For step length, we explored the effect of different sampling frequencies by down-sampling

the data from 200 to 30 Hz in 10 Hz steps (Figure SVI). These findings show that the accuracy does not substantially change beyond sampling frequencies of 100 Hz, suggesting this is sufficient to achieve a high accuracy of step length (<2%). Differences in the pressure sensor material (e.g., spacing and ink⁴⁸) between different insoles can also affect the response time of the sensors at a given sampling frequency and hereby impact the accuracy. For example, the I-Scan system (Tekscan[®], South Boston, MA, USA) uses ink-based force-sensing resistors,^{49,50} the ParoTec system (Paromed®, Neubeuern, Germany) uses piezoresistive sensors,^{51,52} and the Pedar system (Novel[®] GmbH, Munich, Germany) uses embedded capacitive sensors.⁵² In contrast, ARION uses force-resistive sensors that include mechanical resistive (for mechanical resistance) and piezoresistive (for electrical resistance) components. Since we found no clear trend for a larger difference or larger variability of the difference with increases in speed, the pressure sensor technology implemented within the ARION system likely exhibits a sufficiently fast sensor response to obtain accurate results at the investigated speeds and when compared with the 50 N threshold in the reference system.

Second, differences in insole and shoe characteristics between studies and participants could also affect the accuracy between studies and among participants within this study. It can for example be argued that a smaller ARION insole length relative to shoe length resulted in a shorter pressure recording (e.g., when the toes were still on the ground, but no pressure was recorded in the case the insole size selected was too small or vice versa with the heel at initial contact) and hence underestimation of ground contact time. However, we found no consistent and strong association between the difference in insole length and the difference in contact time for the lowest or highest investigated speed, suggesting that specification of the difference between original and ARION insole size in the app is unlikely to substantially influence accuracy (Supplementary File I). Another reason for the individual differences could be related to the position of the sensors relative to the midsole cushioning of the shoe. Specifically, the ARION sensor is measuring the contact time of the foot against the shoe, while the Caren system is measuring the contact time of the shoe against the ground. The cushioning of the midsole may induce differences in the measured contact times between both methods, whereby shoes with a high amount of cushioning can delay the impact at the force plate relative to the insole, hereby potentially leading to longer contact times in the insole relative to the force plate. Additional experiments among two participants that ran in a moderately cushioned shoe (Adidas Ultraboost 20) and highly cushioning shoe (Nike Vaporfly) do lend some support to this notion, with both participants

showing shorter contact time recording by ARION in the Adidas shoe, but longer contact times in the Nike shoe relative to the Caren system (Table SXI). Nevertheless, contact times were longer with a similar magnitude when comparing another moderately cushioned shoe (Brooks Glycerin 19) to the highly cushioned Nike shoe. This discrepancy may be related to differences in the mechanical properties of the assessed shoes. As we did not quantify the mechanical properties, future research is required to explore whether shoe cushioning may indeed explain part of the systematic difference in spatiotemporal outcomes between pressure insoles and force plates. In further partial support of the potential influence of shoe cushioning, barefoot running has been shown to result in a smaller mean difference and smaller limits of agreement than shod running with the same insole system.^{29,42}

4.3 | Limitations

There are several limitations to this study that should be considered when interpreting the data. First, although a 20 N threshold has been suggested to accurately identify the stance phase in ballistic movements such as running and has been used as gold-standard to assess kinematic methods of determining ground contact time in several studies,^{31–33} this threshold initially resulted in too many false positive ground contact time detections at higher speeds for the Caren system due to noise in the signal, likely introduced by the treadmill belt. We therefore used a 50N threshold, which could have an effect on the determined differences. When we re-assessed two trials from three random subjects at 2.78 and $5.0 \,\mathrm{m\,s}^{-1}$ using a 20 N threshold, the contact time was 3.6 ms (1.7%) longer compared with the 50N threshold, suggesting this may have some impact on our findings, with the absolute agreement being lower with a 20N threshold. Second, all participants wore their own shoes during the experiments. While this improves the ecological validity of the results, the difference in measured outcomes as compared to the gold-standard system may be smaller when using barefoot conditions.⁴² Thirdly, while drift can affect the accuracy of wearables over the duration of a session,^{29,42} we only assessed sensor drift by investigating if the difference between the ARION wearable and Caren system changed over a duration of approximately 1.5 h of interval running as opposed to continuous running. While further research is therefore required to investigate drift during a continuous running session, we believe the intermittent bouts investigated here also provide relevant information regarding the potential occurrence of short-term sensor drift. The reason for this is that flexible materials such as the pressure sensors are inherently viscoelastic, making

them prone to drift when deformation due to intermittent or continuous loading is induced. For example, continuous standing in between the running trials may introduce ink shits and hereby partially mimic the effect of ink shifts that may occur with continuous running. Fourthly, while we used arbitrary thresholds of <5%, <10%, and <15% to declare excellent, good, and acceptable agreement, respectively, we content that these cutoffs are reasonable for our purposes. For example, differences in contact time between individuals with and without exercise-related lower leg pain have been shown to be 8 ms,¹⁷ which is $\sim 3\%$ of the contact time at a typical recreational running speed of 3 m s^{-1} . Such a small between-subject difference can be detected when the agreement is <5%, "excellent." Further, it has been suggested that wearables need to have an error of <5% for use in clinical trials, while a 10%-15% error may be acceptable for the general population.³⁹ Finally, the chosen terminology of excellent, good, acceptable, and poor is in agreement with terminology often used for validity or reliability purposes in other papers.

4.4 | Perspective

The findings of this study have several practical implications. First, the accuracy on a group level is generally excellent and suggests ARION might be relevant for in-field use in large-scale studies. For example, previous studies reported differences in contact time of 8 ms¹⁷ or 24 ms⁵³ between groups of injured and noninjured runners. These differences are larger than even the largest mean difference of ~5ms observed in our study during downhill running, and considerably larger than the overall difference in contact time of -0.95%. This conclusion is further supported by a low variability of the differences for most outcomes (in all cases still acceptable), and because the wearable was able to detect changes on a group level in spatiotemporal metrics when altering cadence or speed (Tables 1 and SV), thereby opening opportunities for injury prediction and prevention.

Second, the accuracy on an individual level as indicated by the MAPE was generally also good to excellent, with no single subject showing a deviation of >15% across any of the outcomes and conditions. This is important as accurate data and robust functioning is considered important by wearable users.^{14,22} Moreover, the ICC value for consistency was acceptable and good for contact time and swing time, respectively, suggesting that the slightly larger differences for these outcomes in some conditions remains largely constant for each individual across conditions. Real-time feedback relative to an individual's own baseline as currently implemented in the ARION wearable is therefore likely accurate over time. An interesting observation was that stride time and step frequency showed very low ICC values. While this could indicate poor consistency of the difference across trials, this rather reflects a very low between-subject variability relative to within-subject variability (see also the low betweensubject variability for these outcomes in Figure 2), resulting in low ICC values. Indeed, the absolute accuracy as assessed by percentage differences was excellent for both outcomes with low between-subject variability, and this finding therefore highlights the limitations of assessing relative agreement using only ICC values as often done in other studies.

An important consideration when interpreting the accuracy of the ARION wearable is related to the time period over which the spatiotemporal metrics are compared. Specifically, in this study we compared the average spatiotemporal metrics over ~60s between the ARION and Caren system. In such a situation, under- and overestimation of for example contact time over multiple steps due to a lower sampling frequency of the ARION system (150 Hz for the pressure data) will eventually result in an accurate average contact time. Moreover, the average over multiple steps also has a higher precision than the precision that can obtained from a single step as this would be ~7 ms with a sampling frequency of 150 Hz. The accuracy of the ARION system is therefore lower when comparing single steps or a very small number of steps (e.g., 10) due to the lower sampling frequency. Nevertheless, since realtime feedback with this system is provided based on an average value over multiple steps, the applied comparison approach best reflects in-field use.

In the current study, we had to exclude no data due to hardware malfunctioning. This therefore demonstrates acceptable robustness of the wearable in the investigated conditions, and potentially even slightly better performance compared to other insoles on this point. Previous research using the Loadsol for example excluded one of the 30 subjects (~3%) due to equipment challenges during data collection.⁴⁷ Similarly, another study with the Loadsol excluded ten single steps across various conditions due to incorrect re-zeroing of the sensors.⁴² Nevertheless, further research is required on the robustness during in-field conditions, at even higher speed ranges, and on the durability of the equipment.

5 | CONCLUSION

This study shows that the ARION wearable can generally measure spatiotemporal outcomes with excellent to good accuracy during various running speeds and slopes with an overall mean difference of -0.09% at a group level. Moreover, the accuracy at an individual level was also generally good to excellent as indicated by the MAPE. Collectively, these findings suggests that the ARION wearable may be useful for large-scale in-field studies, and also useful to quantify spatiotemporal metrics with generally at least good accuracy for individual runners.

AUTHOR CONTRIBUTIONS

Bas Van Hooren conceived the study, collected and analyzed the data, and wrote the first draft of the manuscript. Paul Willems assisted in data processing. Guy Plasqui and Kenneth Meijer provided comments and edits. All authors approved the final version.

ACKNOWLEDGEMENTS

BVH was funded by an Eurostars grant (ID 12912) awarded by Eureka. We would like to thank Koen Frenken for his assistance with data collection and Andrew Statham from Ato-Gear for his comments on the manuscript.

CONFLICT OF INTEREST STATEMENT

Bas Van Hooren, Paul Willems, Guy Plasqui, and Kenneth Meijer declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OpenScience Framework at https://osf.io/vbe9s/, reference number 10.17605/OSF.IO/VBE9S.

ORCID

Bas Van Hooren **b** https://orcid. org/0000-0001-8163-693X

REFERENCES

- Oja P, Titze S, Kokko S, et al. Health benefits of different sport disciplines for adults: systematic review of observational and intervention studies with meta-analysis. *Br J Sports Med.* 2015;49(7):434-440. doi:10.1136/bjsports-2014-093885
- Baltich J, Emery CA, Whittaker JL, Nigg BM. Running injuries in novice runners enrolled in different training interventions: a pilot randomized controlled trial. *Scand J Med Sci Sports*. 2017;27(11):1372-1383. doi:10.1111/sms.12743
- Fokkema T, Hartgens F, Kluitenberg B, et al. Reasons and predictors of discontinuation of running after a running program for novice runners. *J Sci Med Sport.* 2019;22(1):106-111. doi:10.1016/j.jsams.2018.06.003
- Buist I, Bredeweg SW, van Mechelen W, Lemmink KA, Pepping GJ, Diercks RL. No effect of a graded training program on the number of running-related injuries in novice runners: a randomized controlled trial. *Am J Sports Med.* 2008;36(1):33-39. doi:10.1177/0363546507307505
- Bredeweg SW, Zijlstra S, Bessem B, Buist I. The effectiveness of a preconditioning programme on preventing runningrelated injuries in novice runners: a randomised controlled trial. *Br J Sports Med.* 2012;46(12):865-870. doi:10.1136/ bjsports-2012-091397

- Fokkema T, de Vos RJ, van Ochten JM, et al. Online multifactorial prevention programme has no effect on the number of running-related injuries: a randomised controlled trial. *Br J Sports Med.* 2019;53(23):1479-1485. doi:10.1136/ bjsports-2018-099744
- Cloosterman KLA, Fokkema T, de Vos RJ, et al. Educational online prevention programme (the SPRINT study) has no effect on the number of running-related injuries in recreational runners: a randomised-controlled trial. *Br J Sports Med.* 2022;56(12):676-682. doi:10.1136/bjsports-2021-104539
- Chan ZYS, Zhang JH, Au IPH, et al. Gait retraining for the reduction of injury occurrence in novice distance runners: 1-year follow-up of a randomized controlled trial. *Am J Sports Med.* 2018;46(2):388-395. doi:10.1177/0363546517736277
- Davis IS, Tenforde AS, Neal BS, Roper JL, Willy RW. Gait retraining as an intervention for patellofemoral pain. *Curr Rev Musculoskelet Med.* 2020;13(1):103-114. doi:10.1007/ s12178-020-09605-3
- Zhang JH, Chan ZY, Au IP, An WW, Cheung RT. Can runners maintain a newly learned gait pattern outside a laboratory environment following gait retraining? *Gait Posture*. 2019;69:8-12.
- Clansey AC, Hanlon M, Wallace ES, Nevill A, Lake MJ. Influence of tibial shock feedback training on impact loading and running economy. *Med Sci Sports Exerc.* 2014;46(5):973-981. doi:10.1249/MSS.000000000000182
- 12. Noehren B, Scholz J, Davis I. The effect of real-time gait retraining on hip kinematics, pain and function in subjects with patellofemoral pain syndrome. *Br J Sports Med.* 2011;45(9):691-696. doi:10.1136/bjsm.2009.069112
- Willy RW, Buchenic L, Rogacki K, Ackerman J, Schmidt A, Willson JD. In-field gait retraining and mobile monitoring to address running biomechanics associated with tibial stress fracture. *Scand J Med Sci Sports*. 2016;26(2):197-205. doi:10.1111/ sms.12413
- Van Hooren B, Goudsmit J, Restrepo J, Vos S. Real-time feedback by wearables in running: current approaches, challenges and suggestions for improvements. *J Sports Sci.* 2020;38(2):214-230. doi:10.1080/02640414.2019.1690960
- Ceyssens L, Vanelderen R, Barton C, Malliaras P, Dingenen B. Biomechanical risk factors associated with running-related injuries: a systematic review. *Sports Med.* 2019;49(7):1095-1115. doi:10.1007/s40279-019-01110-z
- 16. Malisoux L, Gette P, Delattre N, Urhausen A, Theisen D. Spatiotemporal and ground-reaction force characteristics as risk factors for running-related injury: a secondary analysis of a randomized trial including 800+recreational runners. *Am J Sports Med.* 2022;50(2):537-544. doi:10.1177/03635465211063909
- Lempke AFD, Hart JM, Hryvniak DJ, Rodu JS, Hertel J. Use of wearable sensors to identify biomechanical alterations in runners with exercise-related lower leg pain. *J Biomech*. 2021;126:110646. doi:10.1016/j.jbiomech.2021.110646
- de Ruiter CJ, Verdijk PW, Werker W, Zuidema MJ, de Haan A. Stride frequency in relation to oxygen consumption in experienced and novice runners. *Eur J Sport Sci.* 2014;14(3):251-258. doi:10.1080/17461391.2013.783627
- van Oeveren BT, de Ruiter CJ, Beek PJ, van Dieen JH. Optimal stride frequencies in running at different speeds. *PLoS ONE*. 2017;12(10):e0184273. doi:10.1371/journal.pone.0184273

- Cavanagh PR, Williams KR. The effect of stride length variation on oxygen uptake during distance running. *Med Sci Sports Exerc.* 1982;14(1):30-35. doi:10.1249/00005768-198201000-00006
- Moore IS, Ashford KJ, Cross C, Hope J, Jones HSR, McCarthy-Ryan M. Humans optimize ground contact time and leg stiffness to minimize the metabolic cost of running. *Front Sports Act Living*. 2019;1:53. doi:10.3389/fspor.2019.00053
- 22. Luczak T, Burch R, Lewis E, Chander H, Ball J. State-of-theart review of athletic wearable technology: what 113 strength and conditioning coaches and athletic trainers from the USA said about technology in sports. *Int J Sports Sci Coach*. 2020;15(1):26-40.
- Sperlich B, Holmberg HC. Wearable, yes, but able ...? It is time for evidence-based marketing claims! *Br J Sports Med.* 2017;51(16):1240. doi:10.1136/bjsports-2016-097295
- Kong PW, De Heer H. Wearing the F-scan mobile in-shoe pressure measurement system alters gait characteristics during running. *Gait Posture*. 2009;29(1):143-145. doi:10.1016/j. gaitpost.2008.05.018
- Mann R, Malisoux L, Brunner R, et al. Reliability and validity of pressure and temporal parameters recorded using a pressuresensitive insole during running. *Gait Posture*. 2014;39(1):455-459. doi:10.1016/j.gaitpost.2013.08.026
- 26. Van Hooren B, Fuller JT, Buckley JD, et al. Is motorized treadmill running biomechanically comparable to overground running? A systematic review and meta-analysis of crossover studies. *Sports Med.* 2020;50(4):785-813. doi:10.1007/ s40279-019-01237-z
- Telhan G, Franz JR, Dicharry J, Wilder RP, Riley PO, Kerrigan DC. Lower limb joint kinetics during moderately sloped running. *J Athl Train.* 2010;45(1):16-21. doi:10.4085/1062-6050-45.1.16
- Kluitenberg B, Bredeweg SW, Zijlstra S, Zijlstra W, Buist I. Comparison of vertical ground reaction forces during overground and treadmill running. A validation study. *BMC Musculoskelet Disord*. 2012;13:235. doi:10.1186/1471-2474-13-235
- Burns GT, Deneweth Zendler J, Zernicke RF. Validation of a wireless shoe insole for ground reaction force measurement. J Sports Sci. 2019;37(10):1129-1138. doi:10.1080/02640414.2018.1 545515
- Koldenhoven RM, Hertel J. Validation of a wearable sensor for measuring running biomechanics. *Digit Biomark*. 2018;2(2):74-78. doi:10.1159/000491645
- Fellin RE, Rose WC, Royer TD, Davis IS. Comparison of methods for kinematic identification of footstrike and toe-off during overground and treadmill running. J Sci Med Sport. 2010;13(6):646-650. doi:10.1016/j.jsams.2010.03.006
- Smith L, Preece S, Mason D, Bramah C. A comparison of kinematic algorithms to estimate gait events during overground running. *Gait Posture*. 2015;41(1):39-43. doi:10.1016/j. gaitpost.2014.08.009
- Milner CE, Paquette MR. A kinematic method to detect foot contact during running for all foot strike patterns. *J Biomech*. 2015;48(12):3502-3505. doi:10.1016/j.jbiomech.2015.07.036
- Chumanov ES, Wille CM, Michalski MP, Heiderscheit BC. Changes in muscle activation patterns when running step rate is increased. *Gait Posture*. 2012;36(2):231-235. doi:10.1016/j. gaitpost.2012.02.023
- 35. Groothuis A, Houdijk H. The effect of prosthetic alignment on prosthetic and total leg stiffness while running with simulated

running-specific prostheses. *Front Sports Act Living*. 2019;1:16. doi:10.3389/fspor.2019.00016

- 36. Weart AN, Miller EM, Freisinger GM, Johnson MR, Goss DL. Agreement between the OptoGait and instrumented treadmill system for the quantification of spatiotemporal treadmill running parameters. *Front Sports Act Living*. 2020;2:571385. doi:10.3389/fspor.2020.571385
- Falbriard M, Meyer F, Mariani B, Millet GP, Aminian K. Accurate estimation of running temporal parameters using footworn inertial sensors. *Front Physiol.* 2018;9:610. doi:10.3389/ fphys.2018.00610
- Shieh G. Assessing agreement between two methods of quantitative measurements: exact test procedure and sample size calculation. *Stat Biopharm Res.* 2020;12(3):352-359. doi:10.1080/1 9466315.2019.1677495
- Johnston W, Judice PB, Molina Garcia P, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. *Br J Sports Med.* 2021;55(14):780-793. doi:10.1136/bjsports-2020-103147
- Hernández G, Domínguez D, Moreno J, et al. Patellar tendon analysis by ultrasound tissue characterization; comparison between professional and amateur basketball players. Asymptomatic versus symptomatic. *Apunts Med Esport*. 2016;52(194):45-52.
- Bartko JJ. On various intraclass correlation reliability coefficients. *Psychol Bull*. 1976;83(5):762-765.
- 42. Seiberl W, Jensen E, Merker J, Leitel M, Schwirtz A. Accuracy and precision of loadsol((R)) insole force-sensors for the quantification of ground reaction force-based biomechanical running parameters. *Eur J Sport Sci.* 2018;18(8):1100-1109. doi:10.1 080/17461391.2018.1477993
- Stoggl T, Martiner A. Validation of Moticon's OpenGo sensor insoles during gait, jumps, balance and cross-country skiing specific imitation movements. *J Sports Sci.* 2017;35(2):196-206. doi:10.1080/02640414.2016.1161205
- Adams D, Pozzi F, Carroll A, Rombach A, Zeni J Jr. Validity and reliability of a commercial fitness watch for measuring running dynamics. *J Orthop Sports Phys Ther.* 2016;46(6):471-476. doi:10.2519/jospt.2016.6391
- Han H, Kim H, Sun W, Malaska M, Miller B. Validation of wearable activity monitors for real-time cadence. *J Sports Sci.* 2020;38(4):383-389. doi:10.1080/02640414.2019.1702281

- 46. de Fontenay BP, Roy J, Dubois B, Bouyer L, Esculier J. Validating commercial wearable sensors for running gait parameters estimation. *IEEE Sens J.* 2020;20(14):7783-7791.
- Renner KE, Williams DS, Queen RM. The reliability and validity of the Loadsol[®] under various walking and running conditions. *Sensors*. 2019;19(2):265. doi:10.3390/s19020265
- Khodasevych I, Parmar S, Troynikov O. Flexible sensors for pressure therapy: effect of substrate curvature and stiffness on sensor performance. *Sensors*. 2017;17(10):2399-2417. doi:10.3390/s17102399
- 49. Tekscan. *I-Scan® System*. Accessed November 15, 2019. https:// www.tekscan.com/pressure-mapping-sensors
- Woodburn J, Helliwell PS. Observations on the F-scan inshoe pressure measuring system. *Clin Biomech (Bristol, Avon)*. 1996;11(5):301-304. doi:10.1016/0268-0033(95)00071-2
- 51. Paromed. *ParoTec*[®] *System*. Accessed November 15, 2019. http://www.paromed.de/
- Razak AH, Zayegh A, Begg RK, Wahab Y. Foot plantar pressure measurement system: a review. *Sensors*. 2012;12(7):9884-9912. doi:10.3390/s120709884
- Bredeweg SW, Kluitenberg B, Bessem B, Buist I. Differences in kinetic variables between injured and noninjured novice runners: a prospective cohort study. *J Sci Med Sport*. 2013;16(3):205-210. doi:10.1016/j.jsams.2012.08.002

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Van Hooren B, Willems P, Plasqui G, Meijer K. The accuracy of commercially available instrumented insoles (ARION) for measuring spatiotemporal running metrics. *Scand J Med Sci Sports*. 2023;00:1-13. doi:10.1111/sms.14424