Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Normalised plantar pressure measures are reliable, whereas absolute force and gait-curve temporal measures are not: a test–retest reliability study of a plantar pressure platform in patients with knee disorders and healthy controls

Wentao Gong
People's Hospital of Ningxia Hui Autonomous Region (Affiliated Hospital of Ningxia Medical University)
He Shang
People's Hospital of Ningxia Hui Autonomous Region (Affiliated Hospital of Ningxia Medical University)
Tao Ma
People's Hospital of Ningxia Hui Autonomous Region, Ningxia Hui Autonomous Region
Yinbin Wang
People's Hospital of Ningxia Hui Autonomous Region (Affiliated Hospital of Ningxia Medical University)
Desheng Chen
charles_cds@163.com
People's Hospital of Ningxia Hui Autonomous Region, Ningxia Hui Autonomous Region

Research Article

Keywords: Test–retest reliability, Plantar pressure, Pedobarography, Gait analysis, Knee, Intraclass correlation coefficient, Standard error of measurement, Minimal detectable change

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11248193/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Plantar pressure platforms are widely used to characterise foot loading in knee disorders, but the measurement error of individual output parameters is rarely reported. Whether a parameter can detect real change depends on its test–retest reliability and on how large that error is relative to the values measured. We assessed the relative and absolute reliability of static, regional and dynamic plantar pressure and gait parameters in patients with knee disorders and healthy controls.

Methods

Thirty-seven patients with knee disorders and 15 healthy controls completed three consecutive trials of quiet standing and barefoot walking on an Emat plantar pressure plate (3doeCom, China; 50 × 50 cm, 4 sensors/cm², 300 Hz). We derived 22 dynamic gait-curve, 60 regional (force, contact area and pressure across ten zones) and 12 static load-distribution parameters, and computed ICC(2,1) with 95% confidence intervals, standard error of measurement, minimal detectable change at 95% (MDC95), within-subject coefficient of variation, and the trials needed to reach ICC 0.75. Reporting followed the GRRAS statement. Physiologically impossible values (negative durations or forces) were treated as missing.

Results

Reliability varied sharply with parameter type. All 12 static percentage load-distribution parameters were reliable (ICC 0.77–0.97), as were regional peak pressures (16/17 ≥ 0.75). Regional force in kilograms reached ICC ≥ 0.75 in only 4 of 20 parameters in patients (13 of 20 in controls), and dynamic gait-curve parameters in only 2 of 22. For unreliable parameters, MDC95 exceeded the group mean, sometimes severalfold (right metatarsal peak force: MDC95 97.3 kg versus a mean of 78.5 kg). Absolute force was measured less reliably in patients than in controls. Most dynamic parameters required three or more trials, and several could not reach acceptable reliability at any number.

Conclusions

Normalised, ratio-derived plantar pressure measures are reliable in knee disorders; absolute force output and the temporal landmarks of the force curve are not. Repeated-measures studies should be built on normalised pressure measures, record several trials and walking speed, and report ICC, SEM and MDC95 for every parameter used.

Background

Plantar pressure platforms are now a routine part of the assessment of lower-limb disorders. In knee osteoarthritis and after knee surgery they are used to describe altered foot loading, to track recovery after total knee arthroplasty, and to assess the effect of rehabilitation or corrective osteotomy [1–5]. Their appeal is straightforward: the test is quick and non-invasive, and it generates a large number of numerical outputs, from whole-foot load distribution to regional forces, contact areas and pressures, and the timing landmarks of the vertical force curve.

That volume of output is also the difficulty. A single session yields dozens of parameters whose reliability is neither uniform nor guaranteed by the reputation of the instrument. Between-day and within-session error has been documented for individual systems [6–8], yet studies of plantar pressure in knee disease still report group means and p values without stating how much of an observed change could simply be measurement noise. Small pre-to-post differences are therefore read as treatment effects when they may lie within the noise floor of the measurement itself [9].

Two distinct concepts are needed to judge this. Relative reliability, expressed as an intraclass correlation coefficient, describes how well participants keep their rank order across repeated measurements. Absolute reliability, expressed as the standard error of measurement and the minimal detectable change, describes how large a real change must be before it can be distinguished from error [10–12]. A parameter may have an acceptable ICC and still be useless for monitoring an individual patient if its MDC95 exceeds the effect one wishes to detect. Both quantities should therefore be reported, as the Guidelines for Reporting Reliability and Agreement Studies make explicit [13].

While establishing a prospective gait-analysis service for patients with knee disorders, we collected repeated plantar pressure measurements in a consecutive series of patients and in healthy volunteers. Rather than report group differences, we used these data to ask a more basic question: which of the parameters this platform produces can actually be trusted in this population? We estimated the relative and absolute reliability of 94 parameters spanning static load distribution, regional force, contact area, regional pressure and dynamic gait-curve landmarks, determined how many trials are needed to reach acceptable reliability, and compared patients with healthy controls.

Methods

Study design and setting

This was a test–retest reliability study conducted in the Department of Joint Surgery, People's Hospital of Ningxia Hui Autonomous Region (Affiliated Hospital of Ningxia Medical University), Yinchuan, China, between August 2024 and December 2024. The study is reported in accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [13] and with the planning questions raised in the published experience of applying them [14]. All measurements were made within a single session, so no intervention and no between-group comparison of outcomes are involved.

Participants

Two groups were recruited. The clinical group comprised 37 consecutive patients attending the Department of Joint Surgery for a knee-related complaint, with diagnoses recorded by the treating clinician as degenerative meniscal or ligamentous pathology, post-traumatic knee injury, or post-surgical follow-up. We did not restrict the group to a single diagnosis, because the aim was to estimate reliability under routine clinical conditions rather than in a narrowly defined disease entity. The control group comprised 15 healthy volunteers with no current knee pain, no previous knee surgery and no condition affecting gait. Participants were excluded if they had an acute foot or ankle injury, a fixed foot-arch abnormality, previous foot surgery, skin lesions precluding barefoot walking, an inability to walk unaided, a major neurological or cognitive disorder, or severe systemic comorbidity. All participants gave written informed consent.

Sex, age, body mass and height were recorded, and body mass index was calculated as body mass divided by squared height. Body mass and height were measured with a calibrated stadiometer and scale.

Instrumentation

Measurements were made with an Emat plantar pressure plate (3doeCom, China). The plate has a 50 cm × 50 cm active acquisition area covered by a resistive sensor array with a density of 4 sensors/cm² (10,000 sensing points) and 8-bit analogue-to-digital resolution, and it samples at 300 Hz [15, 16]. For each sensor it records vertical force, contact area and derived pressure, and aggregates these over ten anatomical zones: medial and lateral heel, medial and lateral midfoot, first, second and lateral metatarsal regions, hallux, second toe and lateral toes. The accompanying FMap software (3doeCom) exports whole-foot load distribution, regional force, contact area and pressure, and the timing landmarks of the vertical force curve, separately for each foot.

Testing protocol

Testing protocol
Each participant attended one session. Testing was performed barefoot, in a room maintained at 22–25°C and 40–60% relative humidity, and all sessions were conducted between 09:00 and 11:00 [16]. After a standardised verbal explanation and visual demonstration, each participant completed two minutes of familiarisation walking on a 10-metre walkway [15, 16]. Three consecutive trials of quiet standing were recorded for the static analysis and three trials of barefoot walking at a self-selected comfortable speed for the dynamic analysis, with a one-minute rest between trials. Participants were asked to look straight ahead, swing their arms naturally and avoid targeting the plate deliberately. A trial was accepted as valid if the foot made complete contact with the acquisition area, if there was no obvious pause, acceleration, turning or targeting, and if the vertical-force and centre-of-pressure traces were continuous, without missing data or abnormal fluctuation [15]. No instructions were given to alter natural walking speed, and gait speed was not recorded; this limitation is addressed in the Discussion. All trials were supervised by the same trained operator, who also judged whether a trial was valid.

Outcome measures

Three families of parameters were extracted for each foot and averaged over the three trials where indicated.

Static parameters (six per foot): whole-foot load as a percentage of total body load; forefoot and hindfoot load as percentages of the load on the same foot; the number of high-pressure points; mean plantar pressure; and total contact area.

Regional parameters (ten anatomical zones, three variables per zone and per foot): peak vertical force in kilograms, contact area in square centimetres, and peak pressure in kilograms per square centimetre.

Dynamic gait-curve parameters (eleven per foot): stance time; the times of the first force peak, the intervening valley and the second force peak; the absolute difference between the two force peaks; heel contact time and heel peak force; midfoot contact time and midfoot peak force; and metatarsal contact time and metatarsal peak force.

Sample size

Reliability studies are sized for the precision of the reliability estimate rather than for power against a clinical endpoint. With 33 to 37 participants contributing three trials, the 95% confidence interval for a point estimate of ICC close to 0.80 spans approximately 0.65 to 0.90, which we considered adequate to separate acceptable from unacceptable parameters. The control group is smaller and its estimates correspondingly less precise; this is stated as a limitation.

Statistical analysis

Relative reliability was estimated with a two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient, ICC(2,1), computed from the three within-session trials [17]. Two-sided 95% confidence intervals were obtained with the Shrout–Fleiss F-based method. Absolute reliability was expressed as the standard error of measurement (SEM = √MSE) and as the minimal detectable change at the 95% level, MDC95 = 1.96 × √2 × SEM [18]. MDC95 was also expressed as a percentage of the group mean so that parameters with different units could be compared. Within-subject variability was summarised as the median coefficient of variation across participants. Agreement between the first two trials was examined with Bland–Altman plots, reporting the mean bias and the 95% limits of agreement [19]. The number of trials required to reach an ICC of 0.75 was estimated from the single-measure ICC with the Spearman–Brown prophecy formula. ICC values were graded as excellent (≥ 0.90), good (0.75 to 0.90), moderate (0.50 to 0.75), poor (0.25 to 0.50) or very poor (< 0.25), following published guidance [17].

All parameters in this dataset are durations, forces, areas, pressures or derived percentages, and are non-negative by definition. Eleven dynamic measurements (0.5% of the dynamic data matrix) were recorded as negative values, which is physiologically impossible; under a rule fixed before analysis these were treated as missing rather than imputed or winsorised. Parameters were analysed for each foot separately, because pooling feet would have inflated the apparent sample size while concealing side-to-side differences in reliability. No imputation was performed; participants without three complete valid trials for a given parameter were excluded from that estimate only, and the number contributing to each estimate is reported. Computations were performed in Python 3.13 using custom code implementing the formulas above; the analysis code is available from the corresponding author on reasonable request.

Use of artificial intelligence tools

A large language model (CodeBuddy, Tencent) was used to assist with data organisation, implementation of the statistical formulas, figure preparation and language editing. All analyses were specified, checked and interpreted by the authors, who take full responsibility for the content of the article. No artificial intelligence tool was used to generate, alter or fabricate patient data.

Results

Participants

Characteristics of the two groups are shown in Table 1. Patients and controls were well matched for age and body mass index. Diagnoses in the clinical group were heterogeneous, as intended: the commonest presentations were unilateral knee degeneration or injury, with a minority of bilateral presentations and a minority recorded as post-surgical.

Table 1
Participant characteristics

Characteristic	Knee disorders (n = 37)	Healthy controls (n = 15)
Number of participants	37	15
Sex (male/female)	16/21	7/8
Age, years (mean ± SD)	52.6 ± 16.1	53.1 ± 15.1
Age range, years	20–80	27–82
Body mass, kg (mean ± SD)	67.7 ± 11.6	66.3 ± 13.8
Body mass index, kg/m² (mean ± SD)	24.5 ± 3.1	24.0 ± 4.5

Diagnoses in the clinical group: degenerative n = 25, traumatic n = 10, post-surgical n = 7.

Reliability of dynamic gait-curve parameters

Dynamic gait-curve parameters were the least reliable family of outputs (Table 2, Fig. 1). Only two of the 22 parameters, stance time and heel contact time on the left, reached an ICC of at least 0.75 (both 0.89). Thirteen of 22 fell below 0.50, and metatarsal contact time on the left was essentially unrelated across trials (ICC 0.15). Reliability was not consistent between sides within a parameter: stance time was good on the left (0.89) but poor on the right (0.40), and the same left–right dissociation appeared for heel contact time (0.89 versus 0.32) and midfoot contact time (0.69 versus 0.39).

Absolute error was large relative to the values being measured. The MDC95 of peak force difference was 54.0 kg on the left against a group mean of 26.3 kg, and 59.3 kg on the right against 29.1 kg; for metatarsal peak force on the right, MDC95 was 97.3 kg against a mean of 78.5 kg. For these parameters, the change needed before one could be confident that a real

change had occurred exceeds the typical value of the parameter itself. Median within-subject coefficients of variation followed the same pattern, exceeding 30% for first peak time, valley time, midfoot peak force and both peak-force-difference parameters.

Table 2
Test–retest reliability of dynamic gait-curve parameters (patients with knee disorders)

Parameter	Foot	n	Mean	ICC (2,1)	95% CI	SEM	MDC95	MDC95%	Median CV %	Trials needed†	Rating
Stance time	Left	33	0.98	0.89	0.82 to 0.94	0.12	0.33	33%	10.0	1	Good
Stance time	Right	33	0.95	0.40	0.19 to 0.61	0.27	0.76	80%	6.2	5	Poor
Heel contact time	Left	33	0.69	0.89	0.81 to 0.94	0.12	0.34	50%	14.1	1	Good
Heel contact time	Right	33	0.63	0.32	0.10 to 0.54	0.28	0.79	125%	13.9	7	Poor
Midfoot contact time	Left	33	0.56	0.69	0.53 to 0.82	0.16	0.45	81%	17.1	2	Moderate
Midfoot contact time	Right	33	0.60	0.39	0.18 to 0.60	0.25	0.70	118%	18.6	5	Poor
Metatarsal contact time	Left	33	0.78	0.15	-0.05 to 0.39	0.31	0.85	109%	8.7	17	Very poor
Metatarsal contact time	Right	32	0.76	0.68	0.51 to 0.81	0.14	0.39	52%	9.5	2	Moderate
First peak time	Left	32	0.31	0.45	0.23 to 0.65	0.18	0.51	163%	34.9	4	Poor
First peak time	Right	31	0.30	0.43	0.21 to 0.64	0.15	0.41	139%	35.8	5	Poor
Valley time	Left	32	0.51	0.43	0.22 to 0.64	0.23	0.64	124%	31.3	4	Poor
Valley time	Right	31	0.50	0.54	0.33 to 0.72	0.16	0.45	89%	18.1	3	Moderate
Second peak time	Left	33	0.73	0.66	0.49 to 0.80	0.16	0.44	60%	13.2	2	Moderate
Second peak time	Right	33	0.74	0.33	0.11 to 0.55	0.18	0.51	69%	8.9	7	Poor
Heel peak force	Left	33	68.79	0.66	0.48 to 0.80	17.31	47.98	70%	19.5	2	Moderate
Heel peak force	Right	33	74.47	0.30	0.08 to 0.53	32.19	89.23	120%	28.0	8	Poor

Parameter	Foot	n	Mean	ICC (2,1)	95% CI	SEM	MDC95	MDC95%	Median CV %	Trials needed†	Rating
Midfoot peak force	Left	33	15.72	0.49	0.28 to 0.68	11.05	30.64	195%	46.2	4	Poor
Midfoot peak force	Right	33	17.54	0.58	0.39 to 0.75	11.22	31.10	177%	47.1	3	Moderate
Metatarsal peak force	Left	33	70.48	0.64	0.45 to 0.78	23.10	64.02	91%	20.1	2	Moderate
Metatarsal peak force	Right	33	78.45	0.40	0.19 to 0.61	35.11	97.31	124%	19.6	5	Poor
Peak force difference	Left	33	26.34	0.41	0.20 to 0.62	19.47	53.96	205%	62.8	5	Poor
Peak force difference	Right	33	29.12	0.48	0.27 to 0.67	21.40	59.31	204%	69.1	4	Poor

ICC, intraclass correlation coefficient (two-way random effects, absolute agreement, single measure). CI, confidence interval. SEM, standard error of measurement. MDC95, minimal detectable change at the 95% level, also expressed as a percentage of the group mean. CV, within-subject coefficient of variation, median across participants. †Trials needed to reach ICC ≥ 0.75 estimated with the Spearman–Brown prophecy formula; 'Not reached' indicates that the point estimate was ≤ 0, from which no finite number of trials can be extrapolated.

Reliability of regional plantar force, contact area and pressure

Regional parameters behaved very differently depending on whether they were expressed as an absolute force or as a normalised pressure (Table 3, Table 4, Fig. 3). Regional peak force in kilograms was poorly reliable in patients: only 4 of 20 zone-force parameters reached an ICC of 0.75 or higher (lateral midfoot bilaterally, lateral heel on the left and lateral metatarsal on the right), while the hallux, second toe and lateral toes were essentially unreliable on at least one side. Regional contact area was intermediate (6 of 20 at 0.75 or higher). Regional peak pressure, by contrast, was reliable almost everywhere it could be evaluated: 16 of 17 evaluable zone-pressure parameters reached an ICC of 0.75 or higher, and heel and midfoot pressures reached 0.84 to 0.95.

Zones that carry very little load were unreliable in every expression. The hallux and the second toe recorded forces of a few hundred grams at most; their ICCs were low and their MDC95 values many times the group mean (for example, hallux force on the right: mean 0.58 kg, MDC95 9.75 kg). Only a handful of participants produced complete three-trial data for hallux pressure, which further reduced the precision of that estimate.

Table 3
Test–retest reliability of regional peak force (kg): patients with knee disorders and healthy controls

Zone	Foot	n	Mean	ICC (2,1)	95% CI	MDC95	MDC95%	n	ICC (2,1)	95% CI
Medial heel (A1)	Left	35	10.01	0.72	0.57 to 0.84	8.58	86%	15	0.66	0.39 to 0.86
Medial heel (A1)	Right	36	8.14	0.70	0.54 to 0.82	8.52	105%	15	0.74	0.50 to 0.89
Lateral heel (A2)	Left	35	7.04	0.79	0.66 to 0.88	6.53	93%	15	0.92	0.81 to 0.97
Lateral heel (A2)	Right	36	7.99	0.63	0.46 to 0.78	9.66	121%	15	0.59	0.29 to 0.82
Medial midfoot (A3)	Left	35	2.39	0.65	0.48 to 0.79	8.46	354%	15	0.97	0.93 to 0.99
Medial midfoot (A3)	Right	36	1.66	0.63	0.46 to 0.78	5.02	303%	15	0.49	0.18 to 0.76
Lateral midfoot (A4)	Left	35	3.37	0.86	0.77 to 0.92	4.24	126%	15	0.94	0.87 to 0.98
Lateral midfoot (A4)	Right	36	3.61	0.84	0.74 to 0.91	4.34	120%	15	0.94	0.86 to 0.98
First metatarsal (A5)	Left	35	2.41	0.47	0.27 to 0.66	5.19	215%	15	0.61	0.32 to 0.83
First metatarsal (A5)	Right	36	1.93	0.56	0.37 to 0.73	3.09	160%	15	0.49	0.18 to 0.76
Second metatarsal (A6)	Left	35	3.66	0.69	0.53 to 0.82	4.01	109%	15	0.84	0.67 to 0.94
Second metatarsal (A6)	Right	36	3.99	0.68	0.52 to 0.81	6.05	152%	15	0.78	0.56 to 0.91
Lateral metatarsal (A7)	Left	35	3.97	0.72	0.57 to 0.83	5.20	131%	15	0.91	0.80 to 0.97
Lateral metatarsal (A7)	Right	36	5.96	0.76	0.63 to 0.86	7.06	118%	15	0.83	0.66 to 0.94
Hallux (A8)	Left	35	0.12	0.55	0.35 to 0.72	0.70	571%	15	0.99	0.97 to 1.00
Hallux (A8)	Right	36	0.58	0.01	-0.16 to 0.22	9.75	1679%	15	0.97	0.94 to 0.99
Second toe (A9)	Left	35	0.12	0.24	0.03 to 0.46	0.81	671%	15	0.87	0.72 to 0.95
Second toe (A9)	Right	36	0.18	0.62	0.44 to 0.76	1.12	619%	15	0.76	0.54 to 0.90
Lateral toes (A10)	Left	35	2.49	0.57	0.38 to 0.73	4.38	176%	15	0.83	0.65 to 0.93
Lateral toes (A10)	Right	36	1.36	0.02	-0.15 to 0.24	8.42	621%	15	0.54	0.23 to 0.79

Left block of columns: patients with knee disorders. Right block (last three columns): healthy controls, for whom force was the only regional variable recorded in this dataset.

Table 4
Test–retest reliability of regional contact area (cm²) and peak pressure (kg/cm²) in patients with knee disorders

Zone	Foot	n	Mean	ICC	MDC95	MDC95%	n	Mean	ICC	MDC95	MDC95%
Medial heel (A1)	Left	35	12.67	0.83	6.68	53%	35	0.83	0.84	0.51	61%
Medial heel (A1)	Right	36	10.84	0.74	6.05	56%	35	0.77	0.92	0.36	47%
Lateral heel (A2)	Left	35	10.92	0.87	5.83	53%	35	0.65	0.89	0.38	59%
Lateral heel (A2)	Right	36	12.24	0.73	8.21	67%	35	0.70	0.94	0.32	45%
Medial midfoot (A3)	Left	35	3.85	0.85	4.26	111%	23	0.51	0.84	0.98	193%
Medial midfoot (A3)	Right	36	3.49	0.87	4.10	117%	22	0.40	0.91	0.37	93%
Lateral midfoot (A4)	Left	35	8.66	0.97	4.26	49%	28	0.41	0.95	0.29	70%
Lateral midfoot (A4)	Right	36	8.78	0.94	5.67	65%	29	0.48	0.79	0.58	120%
First metatarsal (A5)	Left	35	6.18	0.59	7.83	127%	33	0.43	0.69	0.64	149%
First metatarsal (A5)	Right	36	5.47	0.49	7.29	133%	32	0.38	0.84	0.32	84%
Second metatarsal (A6)	Left	35	6.19	0.58	5.99	97%	33	0.67	0.81	0.59	88%
Second metatarsal (A6)	Right	36	6.26	0.60	6.58	105%	34	0.66	0.77	0.82	124%
Lateral metatarsal (A7)	Left	35	7.84	0.59	9.00	115%	33	0.51	0.79	0.44	86%
Lateral metatarsal (A7)	Right	36	9.26	0.68	9.98	108%	33	0.59	0.82	0.37	63%
Hallux (A8)	Left	35	0.32	0.54	1.83	564%	—	—	—	—	—
Hallux (A8)	Right	36	1.09	0.08	9.64	881%	5	0.33	0.82	0.37	113%
Second toe (A9)	Left	35	0.37	0.47	1.70	455%	—	—	—	—	—
Second toe (A9)	Right	36	0.42	0.44	2.08	496%	—	—	—	—	—
Lateral toes (A10)	Left	35	4.68	0.47	7.12	152%	34	0.56	0.82	0.41	74%
Lateral toes (A10)	Right	36	3.86	-0.00	15.23	394%	32	0.32	0.80	0.27	85%

Columns 3 to 7: contact area. Columns 8 to 12: peak pressure. Pressure was computed by the platform software as force divided by contact area and is therefore a normalised, ratio-derived quantity; contact area and force are absolute quantities.

Reliability of static load-distribution parameters

Static parameters were the most reliable family of outputs (Table 5). All 12 static percentage load-distribution parameters reached an ICC of at least 0.75 in patients (0.77 to 0.97), and three of them, mean pressure on the right and contact area bilaterally, were excellent (0.93 to 0.97). Median within-subject coefficients of variation for these parameters lay between 3.8% and 15.1%, roughly an order of magnitude below those of the dynamic force parameters. In the smaller control group

the same parameters were mostly reliable (8 of 12 at 0.75 or higher), with wider confidence intervals, as expected from the sample size.

Table 5
Test–retest reliability of static plantar pressure parameters

Parameter	Foot	Mean	ICC (2,1)	95% CI	SEM	MDC95	Median CV %	ICC (2,1)	95% CI
Whole-foot load (%)	Left	50.86	0.78	0.65 to 0.87	6.55	18.15	5.4	0.55	0.25 to 0.80
Whole-foot load (%)	Right	49.14	0.78	0.65 to 0.87	6.55	18.15	5.4	0.77	0.54 to 0.91
Forefoot load (%)	Left	40.65	0.80	0.69 to 0.88	9.83	27.26	13.2	0.78	0.57 to 0.91
Forefoot load (%)	Right	40.66	0.78	0.66 to 0.87	10.19	28.25	15.1	0.81	0.61 to 0.92
Hindfoot load (%)	Left	58.99	0.77	0.65 to 0.87	10.60	29.40	9.0	0.78	0.57 to 0.91
Hindfoot load (%)	Right	59.28	0.78	0.66 to 0.87	10.18	28.22	8.4	0.81	0.61 to 0.92
Peak pressure point (kg/cm2)	Left	3.39	0.79	0.67 to 0.88	0.76	2.12	8.7	0.63	0.33 to 0.84
Peak pressure point (kg/cm2)	Right	3.16	0.88	0.80 to 0.93	0.76	2.11	9.8	0.69	0.42 to 0.87
Mean pressure (kg/cm2)	Left	0.65	0.88	0.81 to 0.93	0.13	0.35	4.7	0.87	0.73 to 0.95
Mean pressure (kg/cm2)	Right	0.65	0.97	0.95 to 0.98	0.06	0.17	6.2	0.50	0.18 to 0.77
Contact area (cm2)	Left	63.08	0.96	0.92 to 0.98	5.80	16.08	3.8	0.93	0.85 to 0.97
Contact area (cm2)	Right	61.47	0.93	0.88 to 0.96	7.04	19.51	4.4	0.91	0.79 to 0.96

Patients with knee disorders, columns 3 to 8 (n = 37 except as noted); healthy controls, last two columns (n = 15). Static percentages are internally normalised (forefoot and hindfoot loads sum to the load on the same foot; whole-foot loads sum to 100%), which explains part of their superior stability.

Patients versus controls

Where the two groups could be compared directly, patients with knee disorders were measured less reliably than healthy controls for absolute force. Thirteen of 20 regional force parameters reached an ICC of at least 0.75 in controls, against 4 of 20 in patients, and the control group also produced seven excellent estimates against none in patients (Fig. 3). The direction of this difference is consistent with greater stride-to-stride variability in painful knees, although the smaller control sample makes its estimates less precise, and the comparison should be read as indicative rather than definitive.

Number of trials required

For most dynamic parameters, three trials were not enough to reach an ICC of 0.75 when the single-trial estimate lay below that threshold (Table 2). The Spearman–Brown projection indicated that four to eight trials would be needed for several contact-time and force parameters, and 17 for metatarsal contact time on the left; two parameters with non-positive point estimates could not be projected to reach the threshold at any number of trials. Static and regional pressure parameters

required a single trial. Averaging over trials therefore improves dynamic parameters only up to a point, and cannot rescue parameters whose error is not shared across trials.

Agreement between trials

Bland–Altman plots for four representative parameters are shown in Fig. 2. Stance time and heel contact time on the left showed a mean bias close to zero and limits of agreement of roughly ± 0.3 s. Peak force difference on the left showed proportional bias, with the spread of differences widening as the mean increased, and limits of agreement of − 47 to + 69 kg. Metatarsal contact time on the right showed a single extreme outlying pair, illustrating how a small number of aberrant trials can dominate the error of a parameter whose typical value is small.

Discussion

This study asked a question that is usually assumed away: which outputs of a plantar pressure platform can be trusted when the same patient is measured repeatedly? Parameters that are normalised, either internally within a measurement or by division by contact area, were reliable. Parameters expressed as absolute quantities, whether a force in kilograms or a moment in time located on the force curve, were not, and for several of them the error exceeded the quantity being measured.

Principal findings in context

Our estimates for static parameters are consistent with the limited published literature. Gurney and colleagues reported between-day ICCs of 0.84 to 0.99 for regional plantar pressures in healthy adults using a comparable platform [6], and Zammit and colleagues reported within-session ICCs above 0.90 for most regional pressure variables during barefoot walking [7]. Our regional pressure estimates fall in the same range. The novel contribution lies in the contrast between three ways of expressing the same underlying measurement. Regional pressure is force divided by area; regional force and regional area are its numerator and denominator. In our data the ratio was reliable while its components were not, which points to a common multiplicative error, in the force exerted from step to step, that cancels in the ratio but not in the numerator alone. Static percentages behave in the same way for the same reason: forefoot and hindfoot percentages are constrained to sum to the load on the foot, so a trial in which the participant simply loaded the platform more heavily does not shift them.

The poor reliability of the temporal landmarks of the force curve is more striking, and we are not aware of a directly comparable systematic report for this class of platform output. Stance time on one side reached an ICC of 0.89 while the same parameter on the other side reached 0.40, a dissociation that cannot be explained by the measurement device alone, since both sides were recorded simultaneously on the same trials. Temporal landmarks of the force curve are sensitive to small changes in walking speed and in the exact position of the foot on the active sensor area between passes, and several of the parameters we extracted, notably the times of the first peak and of the valley, describe features of a curve whose shape varies from step to step. Temporo-spatial parameters measured over a walkway are known to be highly reliable when speed is controlled [8]; our data suggest that sub-moment landmarks extracted from a pressure platform without speed control do not share that property.

Why patients may be harder to measure than controls

Regional force parameters were less reliable in patients than in controls. We did not design this study to test that difference, and the control group is small, but the direction is biologically plausible: pain and joint instability alter load distribution from step to step, and antalgic gait is by definition variable. This matters in practice. A measurement protocol validated in healthy volunteers may overestimate the reliability achievable in the clinical population it is meant to serve, which argues for validating reliability in the population of interest rather than borrowing estimates from normal subjects.

Implications for research and clinical practice

Three points follow for research and practice. First, repeated-measures designs using this class of platform should be built on normalised measures — whole-foot and regional percentages, and regional peak pressure — because these are the quantities that behave well. Second, absolute force and force-curve timing parameters should either be avoided as primary outcomes or measured with enough trials, and with gait speed recorded and reported, to make the error tolerable; reporting gait speed would also allow others to interpret the timing parameters we found unreliable. Third, the minimal detectable change should be reported alongside any pre-to-post comparison. In our data several dynamic force parameters have MDC95 values larger than their own group means, so any between-occasion difference in them, however statistically significant, is more plausibly noise than effect. This is directly relevant to studies of plantar pressure before and after knee surgery, where changes of the order we found unreliable are routinely interpreted as recovery [2–4].

Strengths and limitations

This study evaluates a larger set of parameters than most reports of its kind, reports relative and absolute reliability together with the number of trials required, and uses a clinical population rather than healthy volunteers alone. Its limitations are equally clear. Reliability was estimated within a single session by a single operator, so our estimates speak to short-term repeatability and say nothing about between-day or between-operator error, which is generally larger [6]. Walking speed was not recorded, which limits both the interpretation of the temporal parameters and the generalisability of our recommendation concerning them. The clinical group was diagnostically heterogeneous, which we regard as appropriate for a measurement study but which means our estimates should not be generalised to any single disease entity without confirmation. The control group was small, so its estimates are imprecise, and the two groups were not individually matched. Finally, the platform and its software are a single commercial system (Emat plate with FMap software, 3doeCom, China), and our conclusions should not be assumed to transfer to other systems without local verification.

Conclusions

In patients with knee disorders and in healthy controls measured with the same plantar pressure platform, normalised ratio-derived measures — static percentage load distribution and regional peak pressure — were reliable within a session, whereas absolute regional force and the temporal landmarks of the vertical force curve were not, with minimal detectable changes that in several cases exceeded the values being measured. Reliability was parameter-specific rather than a property of the platform, and was generally worse in patients than in controls for absolute force. Studies and clinical services using this class of instrument should select outcomes on the basis of measured reliability for the parameters actually used, record walking speed, and report ICC, SEM and MDC95 alongside any repeated-measures comparison.

Abbreviations

CV
coefficient of variation
GRRAS
Guidelines for Reporting Reliability and Agreement Studies
ICC
intraclass correlation coefficient
MDC95
minimal detectable change at the 95% level
SEM
standard error of measurement

Declarations

Ethics approval and consent to participate

The study was approved by the Medical Ethics Committee of the People's Hospital of Ningxia Hui Autonomous Region (Affiliated Hospital of Ningxia Medical University), Yinchuan, China (approval number 2024-KJCG-001). All participants gave written informed consent before inclusion.

Clinical trial number: not applicable.

Consent for publication

Not applicable. No identifiable personal data are presented.

Availability of data and materials

The datasets analysed during the current study are not publicly available, because they contain information that could compromise participant privacy, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Key Scientific and Technological Achievements Transformation Project of Ningxia Hui Autonomous Region (grant number 2024CJE09046). The funder had no role in the design of the study, in the collection, analysis or interpretation of data, or in the writing of the manuscript.

Authors' contributions

WG and DC conceived and designed the study. WG, HS and TM recruited the participants and acquired the plantar pressure and gait data. WG and YW performed data cleaning and the statistical analysis and prepared the figures and tables. WG drafted the manuscript. DC obtained funding and supervised the project. All authors contributed to the interpretation of the results, revised the manuscript critically for important intellectual content, and read and approved the final version. In CRediT terms: conceptualization — WG and DC; data curation — WG, HS and TM; formal analysis — WG and YW; funding acquisition — DC; investigation — WG, HS and TM; methodology — WG and DC; project administration — TM; software — YW; supervision — DC; validation — YW; visualization — YW; writing – original draft — WG; writing – review & editing — DC, YW, HS and TM.

Acknowledgements

The authors thank all participants for their time and cooperation.

Authors' information

Wentao Gong is the first author. He is an attending physician in the Department of Joint Surgery, People's Hospital of Ningxia Hui Autonomous Region, and a postgraduate of the Third Clinical Medical College of Ningxia Medical University. He Shang is a postgraduate of the Third Clinical Medical College of Ningxia Medical University. Tao Ma is an attending physician in the Department of Joint Surgery, People's Hospital of Ningxia Hui Autonomous Region. Yinbin Wang is an attending physician in the Department of Joint Surgery, People's Hospital of Ningxia Hui Autonomous Region, and a postgraduate of the Third Clinical Medical College of Ningxia Medical University. Desheng Chen is Professor and Doctoral Supervisor in the Department of Joint Surgery, People's Hospital of Ningxia Hui Autonomous Region; his research focuses

on clinical and basic research in joint surgery. He is the corresponding author (charles_cds@163.com; Tel: +86 13995006778).

References

1. Lu J, Peng Z, Gu Q, Wu X, Zhu Z, Liu G, Xu H, Tong L. Gait and plantar pressure changes after medial open-wedge high tibial osteotomy combined with arthroscopic partial meniscectomy. Front Bioeng Biotechnol. 2026;14:1760858. https://doi.org/10.3389/fbioe.2026.1760858.
2. Shetty S, Maiya GA, Rao KG, Vijayan S, Shetty SJ, George BM. Alterations in static plantar pressure before and after total knee arthroplasty in individuals with knee osteoarthritis. Prosthesis. 2025;7(5):128. https://doi.org/10.3390/prosthesis7050128.
3. Shetty S, Rao KG, Vijayan S, Maiya GA, George BM. Foot posture as a biomechanical outcome measure following total knee arthroplasty in individuals with severe knee osteoarthritis. Sci Rep. 2025;16(1):1431. https://doi.org/10.1038/s41598-025-31351-0.
4. Szczypiór-Piasecka K, Adamczewska P, Kołodziej Ł, Ziętek P. The temporal-spatial parameters of gait after total knee arthroplasty. J Clin Med. 2025;14(13):4548. https://doi.org/10.3390/jcm14134548.
5. Zhao R, Wei X, Hu S, Zhang Y, Wu H, Li P, Zhao Y. Deficient gait function despite effect index of the Western Ontario and McMaster Universities Osteoarthritis Index score considered cured one year after bilateral total knee arthroplasty. BMC Musculoskelet Disord. 2024;25(1):230. https://doi.org/10.1186/s12891-024-07348-7.
6. Gurney JK, Kersting UG, Rosenbaum D. Between-day reliability of repeated plantar pressure distribution measurements in a normal population. Gait Posture. 2008;27(4):706–9. https://doi.org/10.1016/j.gaitpost.2007.07.002.
7. Zammit GV, Menz HB, Munteanu SE. Reliability of the TekScan MatScan® system for the measurement of plantar forces and pressures during barefoot level walking in healthy adults. J Foot Ankle Res. 2010;3:11. https://doi.org/10.1186/1757-1146-3-11.
8. Menz HB, Latt MD, Tiedemann A, Mun San Kwan M, Lord SR. Reliability of the GAITRite walkway system for the quantification of temporo-spatial parameters of gait in young and older people. Gait Posture. 2004;20(1):20–5. https://doi.org/10.1016/S0966-6362(03)00068-7.
9. Bednarčíková L, Tóth T, Michalíková M, Gajdošová P. Intra-session repeatability of static pedobarometric measurements: analysis of variability in plantar pressure parameters and the spatial localization of the center of pressure and peak pressure. Bioeng (Basel). 2026;13(8):940. https://doi.org/10.3390/bioengineering13080940.
10. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8. https://doi.org/10.1037/0033-2909.86.2.420.
11. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40. https://doi.org/10.1519/15184.1.
12. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033–9. https://doi.org/10.1016/j.jclinepi.2005.10.015.
13. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud. 2011;48(6):661–71. https://doi.org/10.1016/j.ijnurstu.2011.01.016.
14. Gerke O, Möller S, Debrabant B, Halekoh U. Experience applying the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) indicated five questions should be addressed in the planning phase from a statistical point of view. Diagnostics (Basel). 2018;8(4):E69. https://doi.org/10.3390/diagnostics8040069.
15. Xiang Z, Bai R, Wu Y, Wei H, Zhao J, Wang Z, Wei A, Wu J, Chang X, Qin Y. Application of plantar pressure features based on iterative forward feature selection in the auxiliary identification of knee osteoarthritis. Eur J Med Res. 2026;31:1215. https://doi.org/10.1186/s40001-026-04872-1.

16. Shang H, Ma T, Wei Y, Yang T, Liang J, Liu X, Li J, Wang Y, Chen D. Radial extracorporeal shock wave therapy alters plantar pressure distribution in chronic plantar fasciitis: a biomechanical analysis. Front Rehabil Sci. 2026;6:1729608. https://doi.org/10.3389/fresc.2025.1729608.
17. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
18. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10. https://doi.org/10.1016/S0140-6736(86)90837-8.

Figures

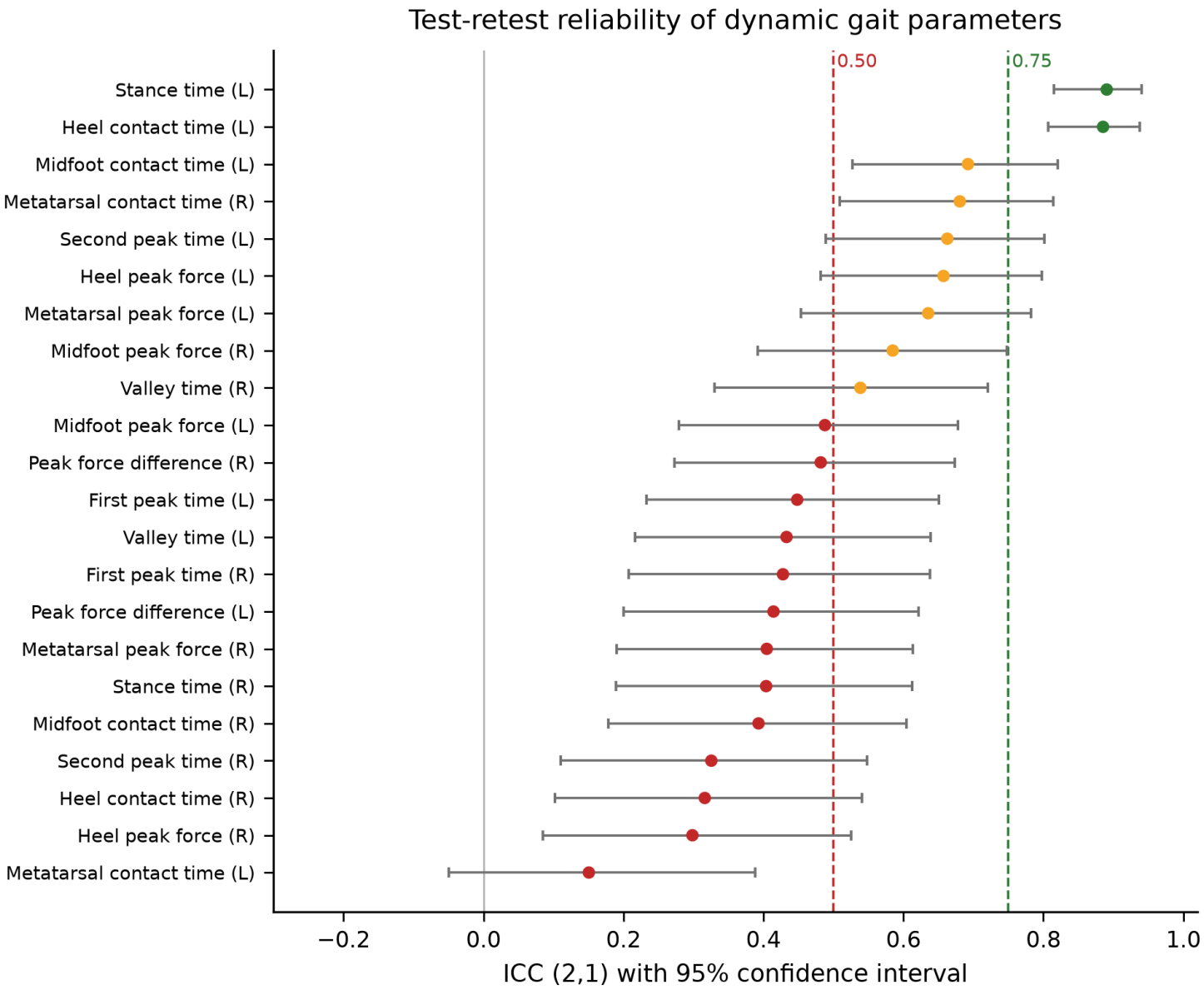


Figure 1

Test–retest reliability of dynamic gait-curve parameters in patients with knee disorders. ICC(2,1) point estimates with 95% confidence intervals, sorted from most to least reliable. Dashed lines mark ICC values of 0.50 and 0.75. L, left; R, right.

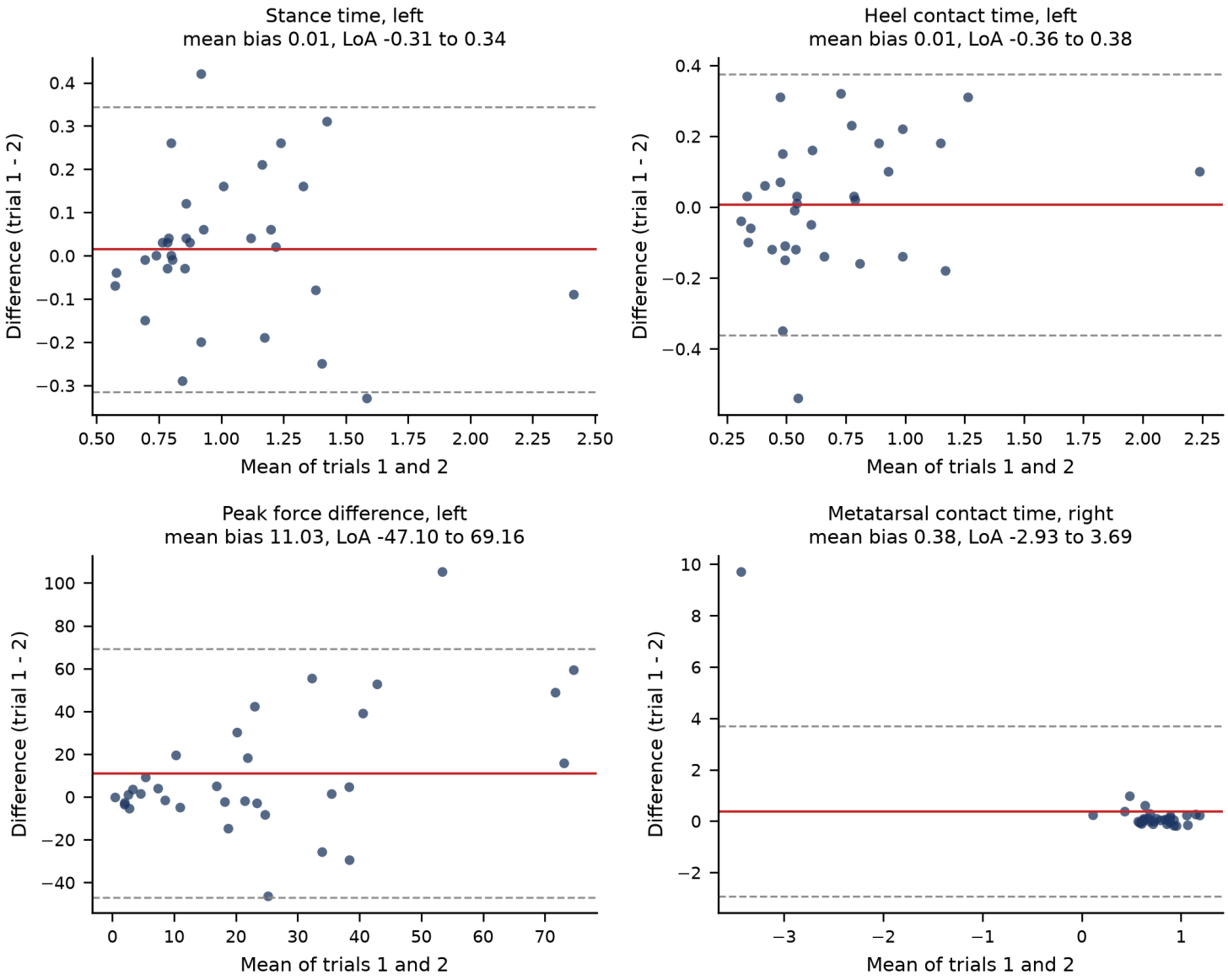


Figure 2

Bland–Altman plots for four representative dynamic parameters, comparing the first and second trials. The solid line is the mean bias and the dashed lines the 95% limits of agreement. Stance time and heel contact time on the left (top row) show narrow, speed-independent limits of agreement; peak force difference on the left and metatarsal contact time on the right (bottom row) show wide limits and, for peak force difference, proportional bias.

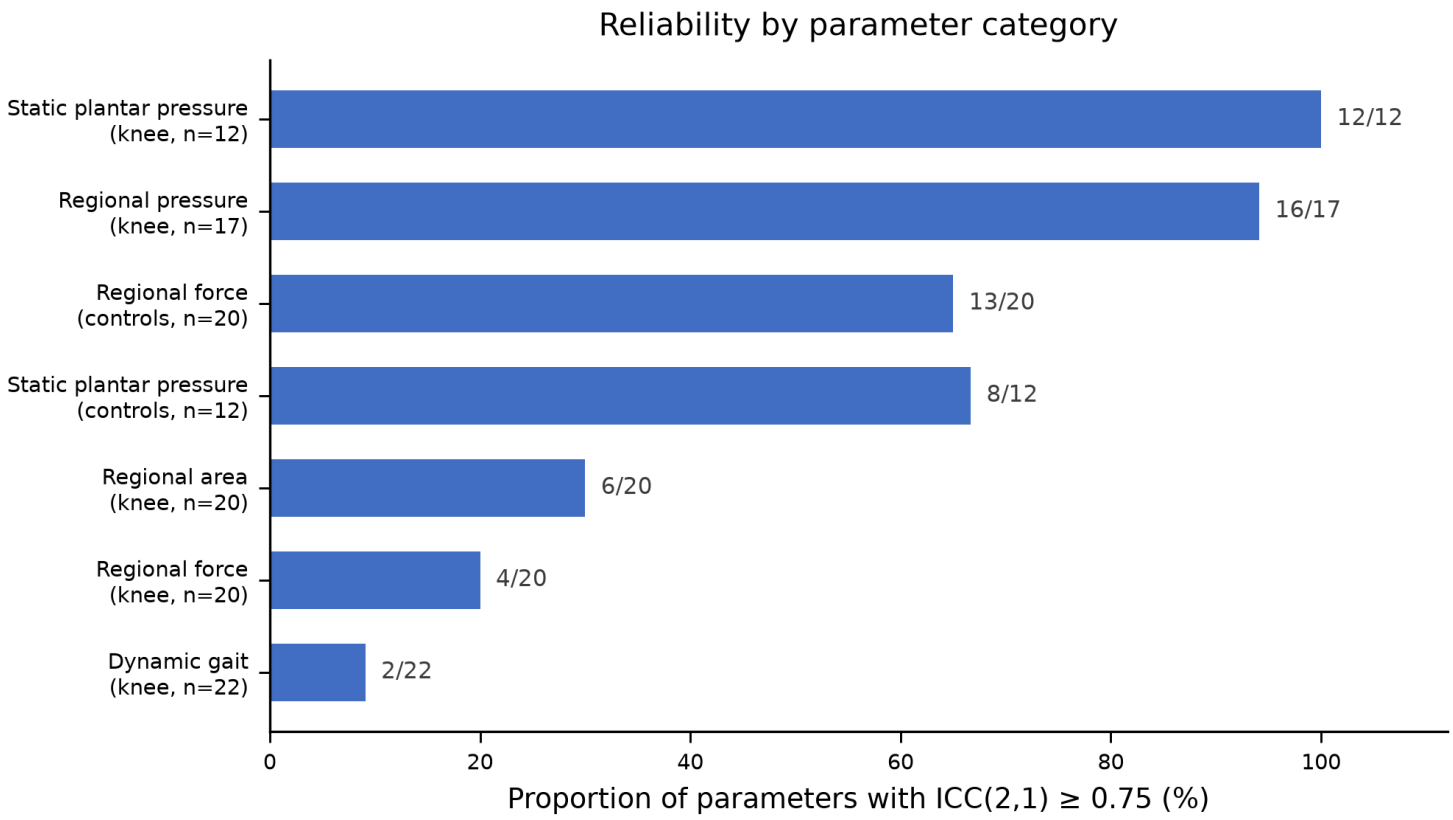


Figure 3

Proportion of parameters reaching ICC(2,1) ≥ 0.75 by parameter category, in patients with knee disorders and in healthy controls. Normalised measures (static percentage load distribution and regional peak pressure) are reliable almost uniformly, whereas absolute force measures and dynamic gait-curve parameters are not.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GRRASChecklist.docx