Portable Biomechanics Laboratory: Clinically Accessible Movement Analysis from a Handheld Smartphone

J.D. Peiffer[1,2], Kunal Shah[1], Irina Djuraskovic[1,3], Shawana Anarwala[1], Kayan Abdou[1], Rujvee Patel[4], Prakash Jayabalan[1,5], Brenton Pennicooke[4], and R. James Cotton[1,5,✉]

[1]Shirley Ryan AbilityLab, Chicago, IL, USA
[2]Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA
[3]Interdepartmental Neuroscience, Northwestern University, Chicago, IL
[4]Department of Neurological Surgery, Washington University School of Medicine, St. Louis, MO, USA
[5]Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Background **The way a person moves is a direct reflection of their neurological and musculoskeletal health, yet it remains one of the most underutilized vital signs in clinical practice. Although clinicians visually observe movement impairments, they lack accessible and validated methods to objectively measure movement in routine care. This gap prevents wider use of biomechanical measurements in practice, which could enable more sensitive outcome measures or earlier identification of impairment.**

Methods **In this work, we present our Portable Biomechanics Laboratory (PBL), which includes a secure, cloud-enabled smartphone app for data collection and a novel algorithm for fitting biomechanical models to this data. We extensively validated PBL's biomechanical measures using a large, clinically representative and heterogeneous dataset with synchronous ground truth. Next, we tested the usability and utility of our system in both a neurosurgery and sports medicine clinic.**

Results **We found joint angle errors within 3 degrees and pelvis translation errors within several centimeters across participants with neurological injury, lower-limb prosthesis users, pediatric inpatients, and controls. In addition to being easy and quick to use, gait metrics computed from the PBL showed high reliability (ICCs > 0.9) and were sensitive to clinical differences. For example, in individuals undergoing decompression surgery for cervical myelopathy, the modified Japanese Orthopedic Association (mJOA) score is a common patient-reported outcome measure; we found that PBL gait metrics not only correlated with mJOA scores but also demonstrated greater responsiveness to surgical intervention than the patient-reported outcomes.**

Conclusions **These findings support the use of handheld smartphone video as a scalable, low-burden tool for capturing clinically meaningful biomechanical data, offering a promising path toward remote, accessible monitoring of mobility impairments in clinical populations. To promote further research and clinical translation, we release the first method for measuring whole-body kinematics from handheld smartphone video validated in clinical populations: https://IntelligentSensingAndRehabilitation.github.io/MonocularBiomechanics/.**

Computer Vision | AI Gait Analysis | Biomechanical Modeling

Correspondence: *rcotton {at} sralab.org*

Introduction

Many clinical conditions have pronounced movement phenotypes which are rarely measured in clinical practice (1–3). For example, temporal parameters, including stride time and swing time, have been shown to predict fall risk in older adults (4). Patients with knee osteoarthritis (KOA) also show altered gait kinematics, which can predict disease progression (5, 6). Gait kinematics can also characterize recovery dynamics after neurological injury and even predict responses to particular interventions (7, 8). The *Stroke Recovery and Rehabilitation Roundtable* emphasized the importance of incorporating movement quantification into clinical trials, while noting that logistical challenges in capturing such data remain a significant barrier (9). Traditionally, gait analysis requires a specialized laboratory equipped with an optical motion capture (OMC) system and force plates, making it expensive, time-consuming, and typically only covered by insurance in limited circumstances (10). In some cases, such as orthopedic surgical planning, force plate measurements and electromyography are required to ensure the precision of marker-based measurements.

Currently, gait is more commonly characterized using the 10-meter walk or 6-minute walk tests, which capture only speed and endurance. Other assessments—such as the Timed Up and Go (TUG) (11), the Berg Balance Scale (BBS) (12), and the Functional Gait Assessment (FGA) (13)—evaluate movement quality using clinical scoring methods that are either subjective or limited to stopwatch-based timing, but do not directly quantify the movement itself. Scalable, accessible movement analysis in the clinic would substantially improve upon the status quo.

Recent advances in AI-based methods offer a promising path forward. Multi-view markerless motion capture (MMMC) requires much less time to acquire and less manual post-processing than OMC. In recent years, numerous studies have demonstrated that MMMC produces similar kinematics to OMC in multiple populations (14–23). OpenCap (24) even allows MMMC to be performed using videos collected from multiple smartphones mounted on tripods. We have also developed methods for MMMC (25–28). A key differentiator of our approaches is the use of MuJoCo (29), a high-

performance physics simulator for machine learning that supports differentiable biomechanical models. This enables us to directly end-to-end optimize the kinematics and skeleton scale from videos. Our end-to-end approach outperforms alternatives that reproduce two-stage marker-based pipelines that first estimate virtual marker locations and then compute inverse kinematics on these trajectories (26, 27). While MMMC is faster than OMC, having multiple synchronized cameras that are spatially calibrated is much less clinically accessible than simply recording with a smartphone.

To make movement analysis more accessible, we previously developed a smartphone app for collecting smartphone videos from clinic (30), which we call our Portable Biomechanics Laboratory (PBL), and validated that we could accurately detect gait events from this data (31). We also tested whether our end-to-end approach to biomechanics can reconstruct kinematics from these smartphone videos and wearable sensors (32). In addition to demonstrating that we could fuse this data, we also found that when the participant is not obscured, we can obtain accurate knee kinematics from smartphone video alone. Fig. 1 provides an overview of our approach. In this study, we validate the accuracy of whole-body kinematics computed with our system and test it in the clinic, showing that the resulting gait metrics are sensitive to the clinical history of our participants. Next, we compare the responsiveness of gait metrics to traditional outcome measures. Specifically, we monitored the gait of a cohort of participants undergoing decompression surgery for cervical myelopathy (CM). We find that gait metrics show higher responsiveness than the modified Japanese Orthopedic Association (mJOA), a commonly used outcome measure for CM (33, 34). Our study demonstrates that quantitative movement analysis can be integrated into a clinic in an accessible and usable workflow, producing accurate and useful results, and that it can be more responsive than traditional clinical outcomes assessments.

Methods

Portable Biomechanics Laboratory. We developed a platform called Portable Biomechanics Laboratory (PBL), which records color video, optional depth video, and internal phone sensor information (accelerometer, gyroscopes, and orientation) from a moving, handheld smartphone (Fig. 1A). The application can optionally be connected to wearable IMUs with all modalities closely synchronized (30, 32). After a recording is finished, PBL uploads the data to a secure cloud platform where data can be pulled into PosePipe (35), an open-source package we developed for managing and processing videos.

Video Preprocessing. Using PosePipe (35), we identified the subject of interest in each video using DeepSortYOLOv4 (36). Next, we extracted 2D and 3D virtual marker locations, termed "keypoints", from each video frame using MeTRAbs-ACAE (37) (Fig. 1B). We previously identified the 87 keypoints from the MoVi dataset (38) to be sufficient for full-body kinematics (39).

Biomechanical Fitting. We obtained kinematics by optimizing the fit of biomechanical models using our end-to-end differentiable biomechanics framework (26, 32). A complete description is in the Supplemental Methods; here we provide a brief overview (Fig. 1C,D). We represented a movement trajectory as a neural network that maps time to pelvis location and joint angles, termed an "implicit function". This network was optimized to minimize error between predicted virtual markers and those extracted from video, using GPU-accelerated physics simulation in MuJoCo (29, 40).

Datasets. We applied our monocular fitting approach to three datasets. The first of these recorded clinical participants with the PBL system and also an MMMC system. The second used iPhone videos and an OMC system, but only contained able-bodied individuals. The third dataset reflects our intended use case of videos recorded from participants seen in various clinics, which we used to demonstrate our method can capture clinically meaningful features of gait. Combined, these datasets contain 21 hours of data.

1. Our MMMC dataset includes 11.7 hours of recordings from 161 sessions with 148 participants performing walking and lower-limb functional tasks. Of these, 121 participants were recorded simultaneously using both the PBL system (30) and our MMMC system (26, 41). This subset included 33 healthy controls, 6 pediatric inpatients, 48 lower limb prosthesis users (LLPUs), and 40 participants with neurological injury. An additional 27 neurological inpatients were recorded using only the MMMC system on a clinical floor. For neurologic patients, we extracted clinical outcome scores—such as the 10 Meter Walk Test (10MWT) and Berg Balance Scale (BBS) (12)—from 54 participants across 95 visits.

2. The OMC dataset used was BML-MoVi (38), a publicly available dataset of 90 healthy controls performing a variety of sports and upper/lower body movements. It contains 3.8 hours of synchronized video (captured with an iPhone 7) and OMC data using a dense 87-marker set. We fit this dataset using our end-to-end differentiable biomechanics approach (26), applying the same biomechanical model used for the MMMC and PBL datasets.

3. Our in-clinic dataset includes 5.5 hours of video from 74 participants recorded with the PBL system during outpatient clinic visits: 55 from a neurosurgery clinic for cervical myelopathy (CM) and 19 from a clinic for knee osteoarthritis (KOA). After their standard physician consultations, participants walked in the clinic hallway while a researcher recorded them using the PBL system. CM participants also completed the Modified Japanese Orthopedic Association (mJOA) questionnaire (33, 34), which scores upper and lower extremity function. Seventeen CM participants were recorded at visits both before and after surgery.

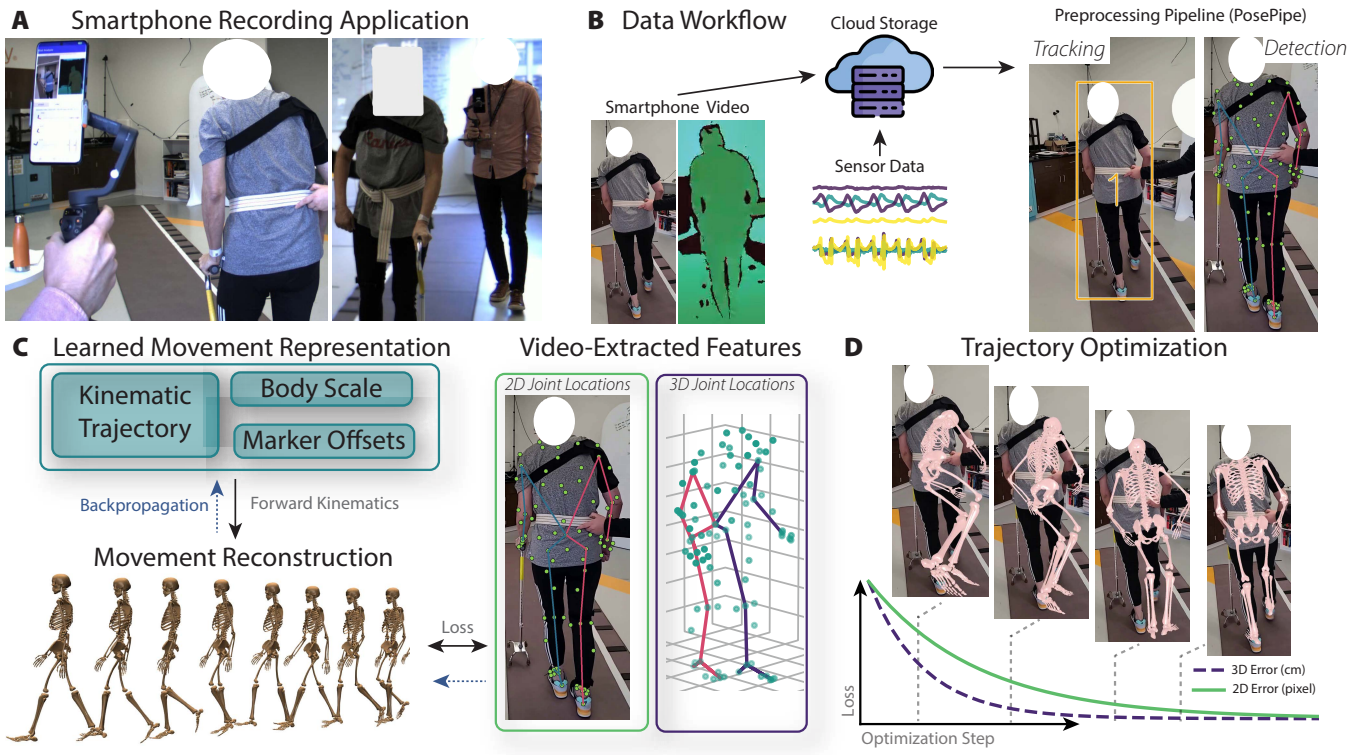


Fig. 1. Methods Overview. We introduce a method for biomechanically grounded movement analysis in clinical settings using a handheld smartphone. **A)** Researchers held a smartphone (optionally with gimbal) while following a participant walking. Our system has no specific requirements regarding viewing angle, distance to subject, or therapist assistance. **B)** Recorded smartphone video and optional wearable sensor data are stored in the cloud, and processed using PosePipe, an open-source package implementing computer vision models for person tracking and keypoint detection. **C)** To reconstruct movement, we represent movement as a function that outputs joint angles, which—combined with body scaling parameters and evaluated through forward kinematics—generate a posed biomechanical model in 3D space. This untrained model is compared to video-extracted joint locations and optionally smartphone sensor data to compute a loss. This loss guides backpropagation to iteratively refine both the kinematic trajectory and body scale. **D)** Initially, the representation lacks knowledge of the person's movements and scale (e.g., height, limb proportions), but after optimization, it typically tracks joint locations within 15 mm in 3D and 5 pixels in 2D.

Accuracy Evaluation. To evaluate the accuracy of the smartphone-based fits compared to OMC and MMMC, we computed the Median Joint Angle Error (MJAE). This is the median of the absolute difference between the two systems over a trial for each joint.

To compare moving (handheld) vs static camera video, we also computed fits from single cameras in the MMMC dataset. As the MMMC videos captured participants from every angle, we examined the effect of camera viewpoints on joint angle error by classifying the camera position relative to the person into front, back, ipsilateral and contralateral.

For static camera and some PBL fits, we report the Root Translation Error (RTE), defined as the average Euclidean error between the pelvis location from the PBL reconstruction and ground truth OMC/MMMC reconstruction. For PBL, we computed the RTE only for fits with camera rotation but no translation (e.g. during a TUG trial).

When aggregating errors over trials, we report the median and normalized interquartile range (nIQR) for MJAE where $nIQR = IQR \cdot 0.7413$.

Gait Metrics and Sensitivity to Clinical Condition. To test whether our portable system collected clinically meaningful information, we calculated the Gait Deviation Index (GDI) (42). Briefly, the GDI uses a dimensionality reduction technique (42, 43) on cycle-aligned joint kinematics and then measures the distance in this subspace to a distribution of normative gait trials. For the normative distribution, we used PBL and MMMC reconstruction from controls in the MMMC dataset as well as an additional 136 participants without gait impairments recruited at the American Society of Biomechanics meeting, resulting in 9,008 normative steps. We also computed double support time and cadence using a model for gait event detection we previously developed and validated (30, 31).

Clinical metrics must be reliable, valid, and responsive. To assess reliability, we computed intra-class correlations (ICCs) between steps taken by the same individual on the same day (44). Construct validity of the GDI was tested by comparing groups with known differences (e.g., transfemoral vs. transtibial amputees) and by correlating kinematic metrics with established clinical scores, including the mJOA, 10MWT, and BBS. To assess responsiveness, we compared the smartphone-based metrics and the mJOA following surgical decompression in the CM dataset using the Standardized Response Mean (SRM) (45, 46). Confidence intervals on SRMs were computed using a jackknife resampling procedure (46, 47).

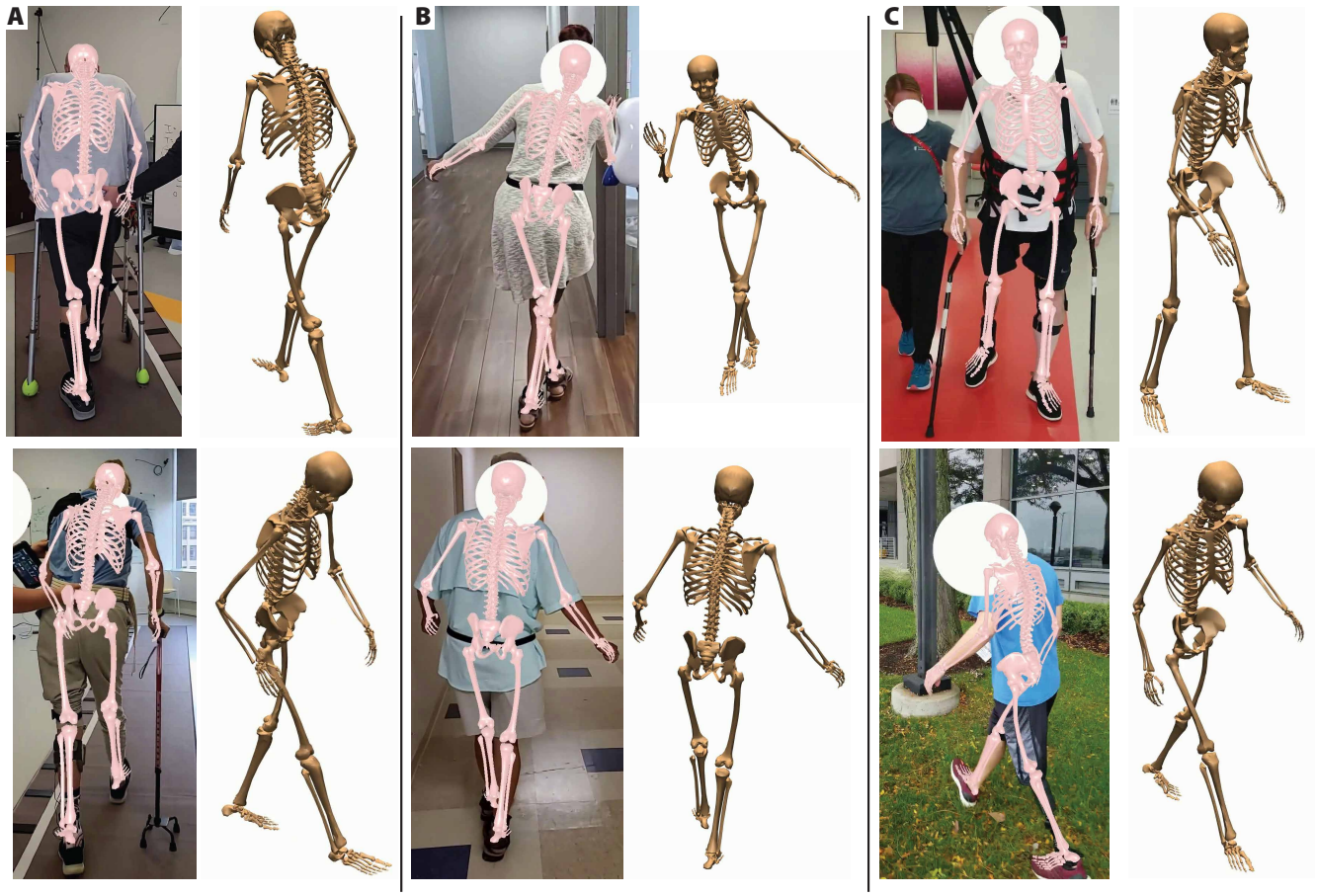


Fig. 2. Biomechanical Reconstructions from Handheld Smartphone. Using only a handheld smartphone, our PBL and end-to-end biomechanical fitting approach robustly captures movement across varied clothing types, assistive devices, and clinical environments. **A)** We first validate this approach on clinical outpatients in our lab, including participants using assistive devices or receiving support from a clinician as needed. **B)** We next deploy this approach in an outpatient clinic for patients with gait impairments, finding this method minimally disrupting clinical workflow while capturing relevant gait features. **C)** This technique allows deployable biomechanical capture in dynamic clinical settings as well as outdoors.

Results

Ground Truth Comparison. In general, kinematics from a single camera closely matched MMMC and OMC. Example overlays demonstrate our method is robust to common mobility aids (Fig. 2) and kinematic traces (Fig. 3A) reflect clinical impairments like neurologic or prosthetic asymmetries. Quantitative ground truth comparisons are presented in Table 1 and Supplement.

For the MMMC dataset, the MJAE between a handheld moving camera and MMMC was 2.79°. This error was slightly higher for static camera fits using a single view from the MMMC system (2.96°), likely due to the greater distances to the person. The error was the greatest in the neurological population, with an MJAE of 3.32 degrees, lower for LLPUs (2.97°), and lowest for controls at 2.51 degrees (Fig. 4B, Table S2). Our method also performed well on the BML-MoVi dataset, showing only 2.74 degrees of MJAE from the monocular fits compared to OMC.

Frontal-plane joint angles (e.g., hip adduction) were best estimated from frontal views, while sagittal-plane joint angles (e.g., hip and knee flexion) were more accurately estimated from sagittal views (one-sided Mann-Whitney U-tests; $p<0.005$ for all comparisons). This difference was greatest for knee flexion at 1.02 degrees (Fig. 3C).

Table 1. Summary of errors for different datasets. MJAE is the median joint angle error (in degrees) and RTE is the median root translation error (in cm). Errors are presented as: (Median (nIQR)).

	MMMC		OMC
	Handheld	Static	Static
MJAE (deg)	2.79 (0.86)	2.96 (0.78)	2.74 (0.50)
RTE (cm)	6.44 (4.77)	5.00 (4.68)	2.41 (0.63)

Visual comparisons of root position between single-camera and multi-view fits are shown in Figure 3D. The clinical dataset using a handheld, rotating smartphone yielded an average RTE of 6.44 cm, while a static camera yielded 5.00 cm. RTE for the BML-MoVi dataset with a static smartphone averaged 2.41 cm (Table 1).

Clinical Feasibility. PBL was accessible and easy to use, causing minimal disruption to routine workflows in both KOA and CM clinics. It required no special clothing, lighting, or calibration—we simply followed participants with a handheld phone as they walked.

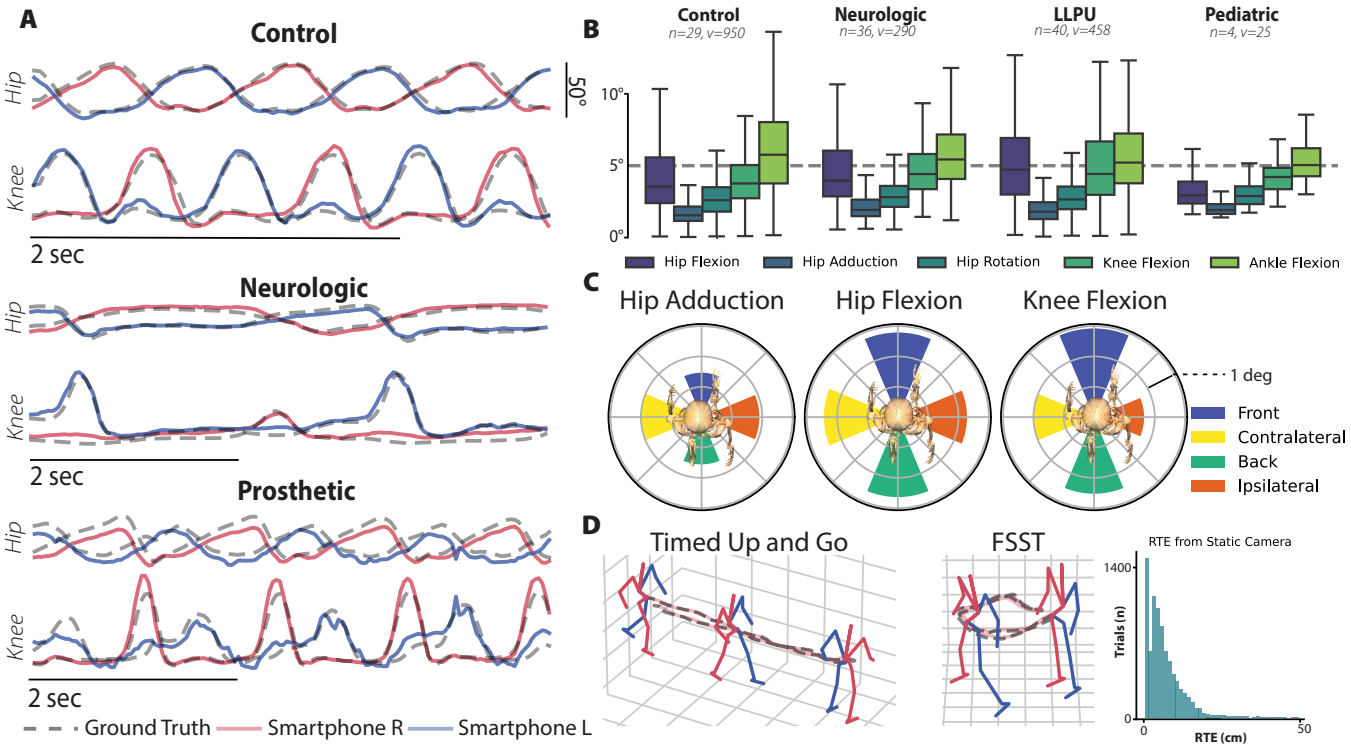


Fig. 3. Quality Measures of Single Camera Fitting A) Kinematic traces from smartphone video (red/blue) compared to ground truth (gray dashed) during walking. **B)** Joint angle errors across populations for select lower limb angles. n denotes the number of unique individuals in each cohort and v denotes the number of total videos for that cohort. **C)** Select joint angle errors with respect to camera view angle show that sagittal plane angles have the lowest error with sagittal camera views, and frontal angles have the lowest error with frontal views. **D)** Pelvis translation (RTE) extracted from handheld smartphone video compared to ground truth during functional gait assessments.

Movement Quality Metrics. The Gait Deviation Index (GDI) provides a holistic measure of movement quality that we use to summarize the information found in PBL-based measures of movement. We test this measure for repeatability, validity, and responsiveness.

GDI Repeatability. We analyzed the >13,500 steps across 222 individuals recorded with the PBL system, and found high repeatability of 0.84 for the GDI of a single step (ICC2) and higher repeatability when averaging multiple steps together: 0.96 (ICC2k) (Table 2). ICCs for spatiotemporal metrics were lower—likely due to timing variability—but remained highly repeatable when averaged (ICC2k).

GDI Construct Validity. We found the GDI measured with our PBL is sensitive to clinical differences between participants (Fig. 5). Performance on the BBS is used to classify fall risk and was measured for our neurologic participants. The GDI was statistically different between the low and high fall-risk participants, and both were different than the control participants (Fig. 5B). The GDI also correlates well with walking speed measured during the 10MWT (Pearson's $r = 0.82$), one of the most commonly measured clinical outcomes for gait (Fig. 5C). It is important to note GDI and 10MWT measure different things, as it is possible to walk faster using a higher cadence with an unchanging kinematic trajectory. Next, the GDI is significantly decreased in transtibial amputees compared to controls and further decreased for transfemoral amputees (Fig. 5D). The KOA population was also statistically different from control with a mean GDI of 83.5 (10.8). Finally, the GDI showed a significant correlation with mJOA ($r = 0.47$) and the lower extremity subscore of the mJOA ($r = 0.48$) (Fig. 5E, Table 2).

GDI Responsiveness. We found that quantitative, PBL-based metrics all had higher responsiveness to surgical intervention than the mJOA, a commonly used clinical outcome measure (Table 2). Notably, the jackknife confidence intervals for GDI and DST did not cross zero while the mJOA did, suggesting greater effect size stability, though not necessarily statistical significance.

Discussion

We present a clinically validated method for extracting accurate biomechanics from smartphone video in clinical settings. Across diverse populations, our single-camera system showed strong agreement with ground truth (typically $< 3^\circ$ error). Gait quality metrics derived from these videos detected clinically relevant group differences and were more responsive than the mJOA following decompression surgery. These findings suggest that a clinician, therapist, or medical assistant could feasibly perform gait analysis in under a minute to obtain an objective measurement of a patient's gait impairment to sensitively track their recovery.

Joint angle errors were generally low across all populations and settings (Table 1). We attribute the slightly higher MJAE in clinical populations (0.46 - 0.81 degrees, Fig. 3B, Fig. 4B, Table S2) to assistive devices such as canes, walkers, ankle foot orthoses, or gait belts (Fig. 2) and the fact that keypoint

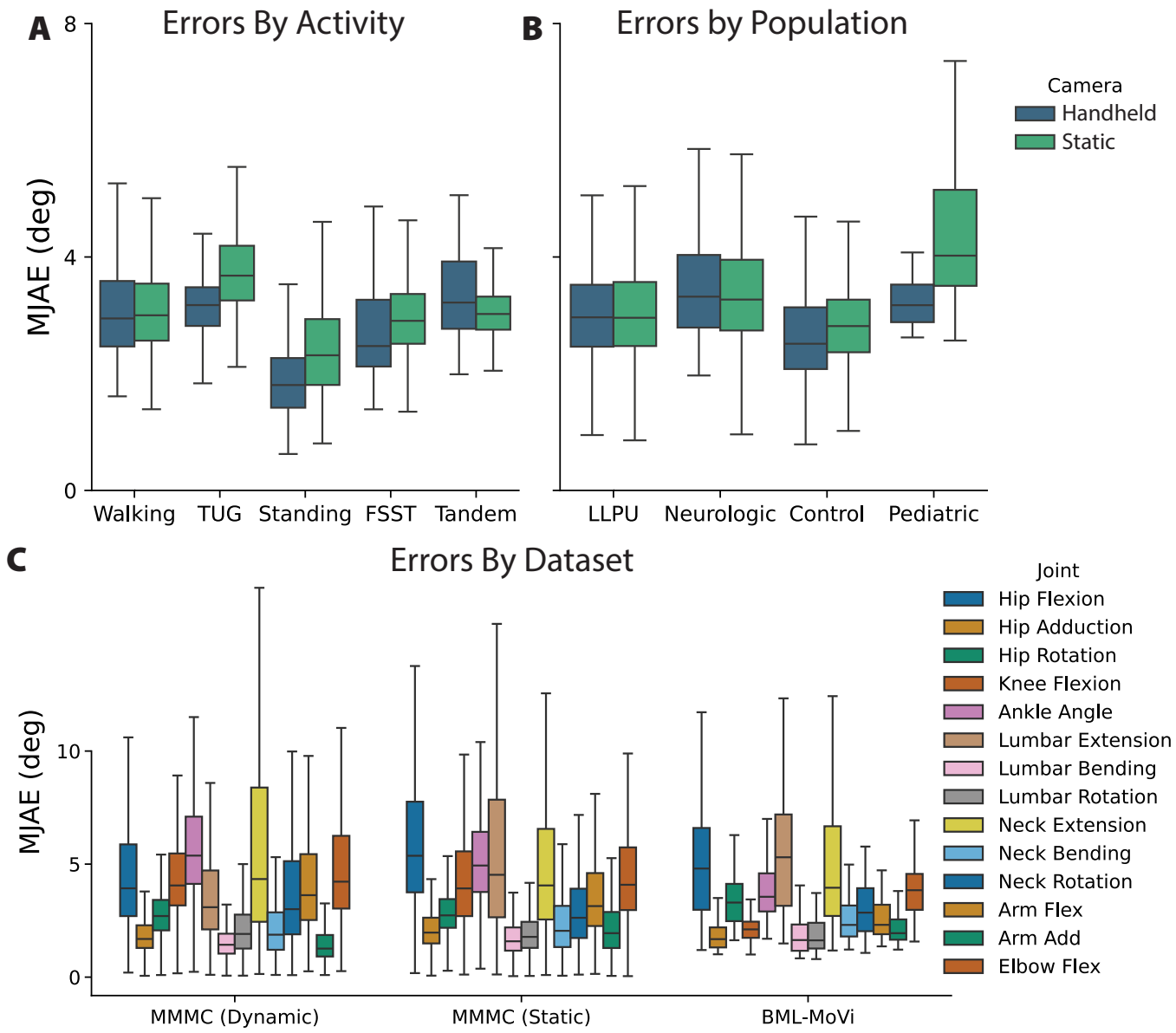


Fig. 4. Joint Angle Error Distributions. A) In the MMMC dataset, MJAE distributions remain relatively stable when aggregated across activity types, participant populations **(B)**, as well as between handheld (moving) and static cameras. **C)** Joint-specific MJAE distributions from both the MMMC and BML-MoVi datasets.

Table 2. Reliability, Validity, and Responsiveness of Smartphone-Based Metrics. ICCs are presented as ICC (95% confidence intervals). mJOA and mJOA-LE show Spearman correlations between gait metric and ordinal ratings. SRM a unitless measure of change in metric before and after surgery, with 95% confidence intervals. * indicates $p < 0.05$.

	ICC2	ICC2k	r mJOA	r mJOA-LE	SRM
mJOA	-	-	-	-	0.20 (-0.38 0.78)
mJOA-LE	-	-	-	-	0.14 (-0.45 0.72)
Cadence	0.58 (0.52 0.64)	0.88 (0.85 0.90)	0.25*	0.39*	0.34 (-0.20 0.88)
DST	0.41 (0.35 0.48)	0.78 (0.72 0.82)	-0.27*	-0.37*	-0.48 (-0.88 -0.07)
GDI	0.84 (0.81 0.86)	0.96 (0.96 0.97)	0.47*	0.48*	0.45 (0.00 0.91)

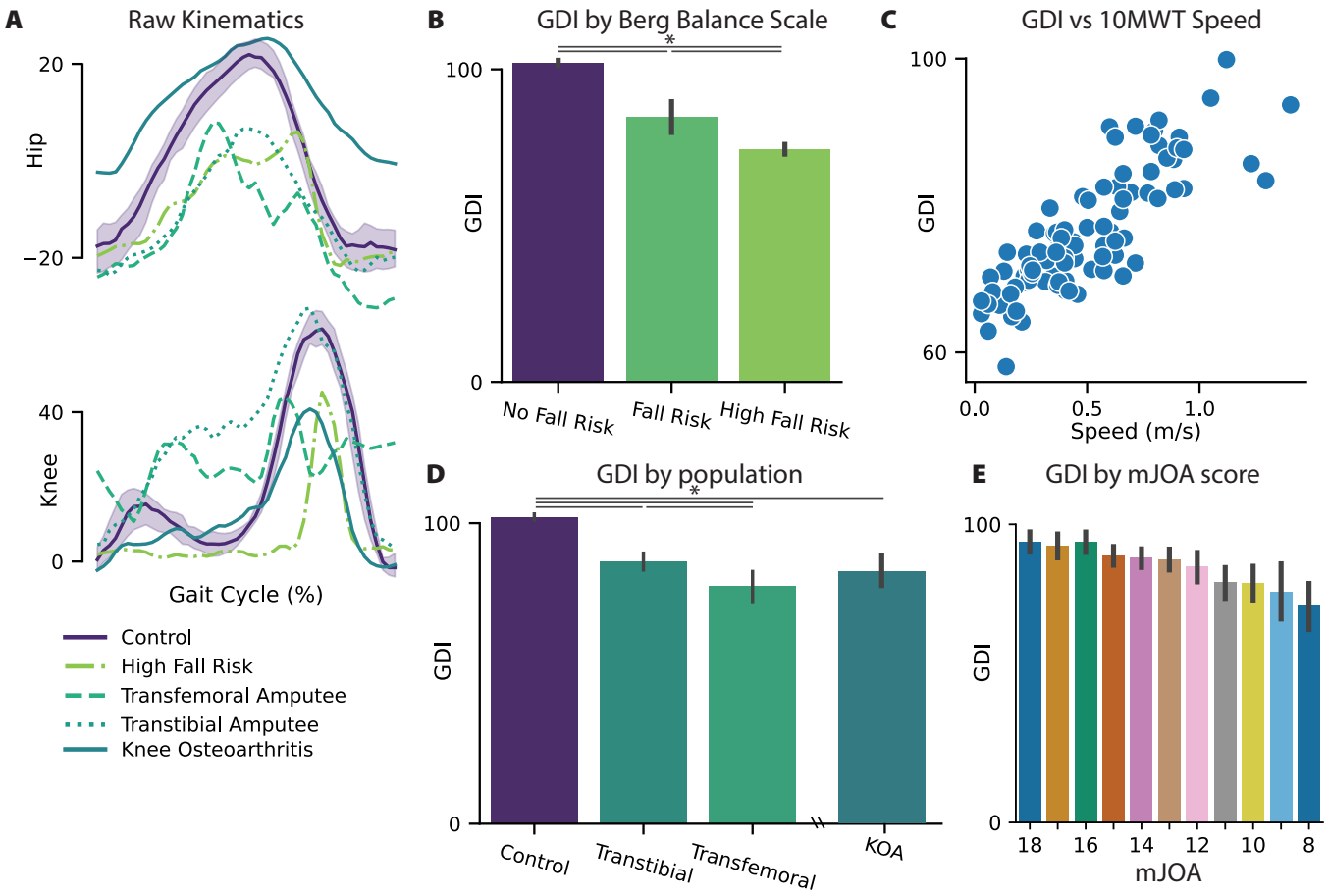


Fig. 5. Clinical Validity of Smartphone-Based Gait Deviation Index. A) Hip and Knee flexion angles of clinical and control groups **B)** GDI separates groups at risk of falls determined by the Berg Balance Scale. **C)** GDI correlates with 10 Meter Walk Test performance $r = 0.82$. **D)** GDI of LLPUs and KOA participants is significantly lower than that of control populations. Further, GDI of Transfemoral amputees is significantly lower than GDI of Transtibial amputees. **E)** GDI collected in clinical settings correlates ($r = 0.47$) with the mJOA, a clinically used ordinal questionnaire.

detection does not perform as well on prosthetic devices (30).

A Responsive and Valid Metric of Functional Status. Our method measured reliable, valid, and responsive features of movement which is arguably more important for clinical practice than joint angle accuracy. GDI demonstrated construct validity for multiple clinical populations, such as being different for inpatients with a high fall risk than low fall risk based on the BBS (Fig. 5B) and correlated highly with the 10MWT (Fig. 5C), both important outcome measures used commonly in clinical practice.

The GDI had a moderate correlation with the mJOA. We note that we do not necessarily expect a strong association between the GDI, a continuous quantitative measure, as the mJOA is a patient-reported ordinal scale. Despite the differences, the GDI and mJOA both responded positively to spinal decompression in the surgical group recorded before/after their surgery (Table 2). Notably, the GDI had a higher SRM (0.45) than both the mJOA (0.20) and mJOA-LE (0.14), with a 95% confidence interval that remained entirely above zero. This indicates that our easy-to-use system could be more sensitive than the widely used mJOA.

We focused on the GDI because it is a widely used summary of gait kinematics. However, we do not claim it is the optimal metric for every application. Instead, we anticipate that scalable gait analysis will enable new research to identify the most clinically meaningful measures. For example, we recently described a method for inferring torques and ground reaction forces from kinematics and saw that the ground reaction forces are sensitive to clinical history (48).

Limitations. The minimally clinically important difference for kinematic features of gait has not yet been established for many clinical conditions. While our accuracy is generally quite good, greater accuracy may be required for certain indications. For example, the 5 degrees of error we typically see at the ankle might be insufficient to detect subtle changes in plantarflexion spasticity or toe clearance from a single camera.

We also note the heavy tails in our error distributions with some trials having errors >10 degrees (Fig. 4C) and our analysis of the reconstruction errors (Fig. S2) indicated little relationship between reconstruction error and MJAE, highlighting the need for reliable confidence scores from markerless motion capture. We have developed methods for both monocular videos without biomechanics (49), and for biomechanics from multi-view video (25). Extending these confidence intervals to monocular biomechanics is an important future direction.

Uniquely Accessible and Validated Approach for Clinical Implementation. Other approaches have proposed deployable systems for human motion capture. Most well-known is OpenCap, which uses two smartphones on static tripods (24) to triangulate joint locations from video and fit biomechanics using the widely used OpenSim framework (50, 51). During walking, the approach presented here shows similar joint angle errors to OpenCap (4.1°) using just a single, moving, camera. OpenCap does go a step further than our approach and produces joint kinetics, which we can incorporate in the future, such as using our Kinetic Twin approach (48). Our experience with PBL shows that the ability to use a single portable camera is critical for easy use in the clinic.

Two recent works estimate biomechanics from monocular video: BioPose (52) and Human Skeleton and Mesh Recovery (HSMR) (53). Neither was evaluated on clinical populations. Our method differs by using trajectory optimization with a single, fixed skeleton scale across trials, and it can readily incorporate wearable sensors (32). HSMR directly regresses joint angles from images, which is fast but likely limited to poses seen in its training data of able-bodied individuals—raising concerns about generalization to clinical movements. It also does not report joint angle errors, making comparisons difficult. On the BML-MoVi dataset, our method performs comparably to BioPose. Ultimately, a proliferation of AI-powered methods for accessible biomechanics will increase use of clinical movement analysis. These large datasets will enable data-driven precision rehabilitation interventions (54).

In conclusion, we present a method for extracting accurate and interpretable joint kinematics from smartphone videos. We validate this method in inpatient, outpatient therapy, and clinical settings, finding it to be highly usable and robust—even in the presence of nearby therapists and assistive devices. Importantly, gait scores extracted from this approach correlate well with commonly used clinical scales such as the 10MWT, BBS, and mJOA. This approach enables increased fidelity in clinical and at-home monitoring of movement impairments, paving the way for a big data revolution in movement science.

ACKNOWLEDGEMENTS

This work was supported by R01HD114776 (RJC), the Restore Center (P2CHD101913), and the Research Accelerator Program of the Shirley Ryan AbilityLab (RJC). JDP is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2234667.

Bibliography

1. M Jacquelin Perry. Gait analysis: normal and pathological function. *New Jersey: SLACK*, 2010.
2. Birol Balaban and Fatih Tok. Gait disturbances in patients with stroke. 6(7):635–642. ISSN 1934-1482, 1934-1563. doi: 10.1016/j.pmrj.2013.12.017.
3. Meg E. Morris, Frances Huxham, Jennifer McGinley, Karen Dodd, and Robert Iansek. The biomechanics and motor control of gait in parkinson disease. 16(6):459–470. ISSN 0268-0033. doi: 10.1016/S0268-0033(01)00035-3.
4. D. Hamacher, N.b. Singh, J.h. Van Dieën, M.o. Heller, and W.r. Taylor. Kinematic measures for assessing gait stability in elderly individuals: a systematic review. 8(65):1682–1698. doi: 10.1098/rsif.2011.0416. Publisher: Royal Society.
5. Tamika L. Heiden, David G. Lloyd, and Timothy R. Ackland. Knee joint kinematics, kinetics and muscle co-contraction in knee osteoarthritis patient gait. 24(10):833–841. ISSN 0268-0033. doi: 10.1016/j.clinbiomech.2009.08.005.
6. L. Sosdian, F. Dobson, T. V. Wrigley, K. Paterson, K. Bennell, M. Dowsey, P. Choong, K. Allison, and R. S. Hinman. Longitudinal changes in knee kinematics and moments following knee arthroplasty: A systematic review. 21(6):994–1008. ISSN 0968-0160. doi: 10.1016/j.knee.2014.09.009.
7. John W. Chow and Dobrivoje S. Stokic. Longitudinal changes in temporospatial gait characteristics during the first year post-stroke. 11(12):1648. ISSN 2076-3425. doi: 10.3390/brainsci11121648. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.
8. Louis N. Awad, Michael D. Lewek, Trisha M. Kesar, Jason R. Franz, and Mark G. Bowden. These legs were made for propulsion: advancing the diagnosis and treatment of post-stroke propulsion deficits. 17(1):139. ISSN 1743-0003. doi: 10.1186/s12984-020-00747-6.
9. Gert Kwakkel, Natasha A Lannin, Karen Borschmann, Coralie English, Myzoon Ali, Leonid Churilov, Gustavo Saposnik, Carolee Winstein, Erwin Eh Van Wegen, Steven L Wolf, John W Krakauer, and Julie Bernhardt. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. 12(5):451–461. ISSN 1747-4930, 1747-4949. doi: 10.1177/1747493017711813.
10. Muhammad Abdullah, Abdul Aziz Hulleck, Rateb Katmah, Kinda Khalaf, and Marwan El-Rich. Multibody dynamics-based musculoskeletal modeling for gait analysis: a systematic review. 21(1):178. ISSN 1743-0003. doi: 10.1186/s12984-024-01458-y.
11. Diane Podsiadlo and Sandra Richardson. The timed "up & go": A test of basic functional mobility for frail elderly persons. 39(2):142–148. ISSN 1532-5415. doi: 10.1111/j.1532-5415.1991.tb01616.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.1991.tb01616.x.
12. Katherine O. Berg, Brian E. Maki, Jack I. Williams, Pamela J. Holliday, and Sharon L. Wood-Dauphinee. Clinical and laboratory measures of postural balance in an elderly population. 73(11):1073–1080. ISSN 0003-9993, 1532-821X. doi: 10.5555/uri:pii:000399939290174U. Publisher: Elsevier.
13. Diane M Wrisley, Gregory F Marchetti, Diane K Kuharsky, and Susan L Whitney. Reliability, internal consistency, and validity of data obtained with the functional gait assessment. 84 (10):906–918. ISSN 0031-9023. doi: 10.1093/ptj/84.10.906.
14. Robert M. Kanko, Elise K. Laende, Elysia M. Davis, W. Scott Selbie, and Kevin J. Deluzio. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. 127:110665, . ISSN 0021-9290. doi: 10.1016/j.jbiomech.2021.110665.
15. Robert M. Kanko, Elise Laende, W. Scott Selbie, and Kevin J. Deluzio. Inter-session repeatability of markerless motion capture gait kinematics. 121:110422, . ISSN 1873-2380. doi: 10.1016/j.jbiomech.2021.110422.
16. Robert M. Kanko, Elise K. Laende, Gerda Strutzenberger, Marcus Brown, W. Scott Selbie, Vincent DePaul, Stephen H. Scott, and Kevin J. Deluzio. Assessment of spatiotemporal gait parameters using a deep learning algorithm-based markerless motion capture system. 122: 110414, . ISSN 1873-2380. doi: 10.1016/j.jbiomech.2021.110414.
17. Sherveen Riazati, Theresa E. McGuirk, Elliott S. Perry, Wandasun B. Sihanath, and Carolynn Patten. Absolute Reliability of Gait Parameters Acquired With Markerless Motion Capture in Living Domains. 16:867474. ISSN 1662-5161. doi: 10.3389/fnhum.2022.867474.
18. Ke Song, Todd J. Hullfish, Rodrigo Scattone Silva, Karin Grävare Silbernagel, and Josh R. Baxter. Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements. 157:111751. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2023.111751.
19. Tishya A. L. Wren, Pavel Isakov, and Susan A. Rethlefsen. Comparison of kinematics between theia markerless and conventional marker-based gait analysis in clinical patients. 104:9–14. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2023.05.029.
20. Brian Horsak, Hans Kainz, and Bernhard Dumphart. Repeatability and minimal detectable change including clothing effects for smartphone-based 3D markerless motion capture. 175: 112281. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2024.112281.
21. Sonia D'Souza, Tobias Siebert, and Vincent Fohanno. A comparison of lower body gait kinematics and kinetics between theia3d markerless and marker-based models in healthy subjects and clinical patients. 14(1):29154. ISSN 2045-2322. doi: 10.1038/s41598-024-80499-8. Publisher: Nature Publishing Group.
22. Junqing Wang, Wei Xu, Zhuoying Wu, Hui Zhang, Biao Wang, Zongke Zhou, Chen Wang, Kang Li, and Yong Nie. Evaluation of a smartphone-based markerless system to measure lower-limb kinematics in patients with knee osteoarthritis. 181:112529. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2025.112529.
23. Yu-Sun Min, Tae-Du Jung, Yang-Soo Lee, Yonghan Kwon, Hyung Joon Kim, Hee Chan Kim, Jung Chan Lee, and Eunhee Park. Biomechanical gait analysis using a smartphone-based motion capture system (OpenCap) in patients with neurological disorders. 11(9):911. ISSN 2306-5354. doi: 10.3390/bioengineering11090911. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.
24. Scott D. Uhlrich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S. Chaudhari, Jennifer L. Hicks, and Scott L. Delp. OpenCap: Human movement dynamics from smartphone videos. 19(10):e1011462. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1011462. Publisher: Public Library of Science.
25. R. James Cotton and Fabian Sinz. Biomechanical reconstruction with confidence intervals from multiview markerless motion capture.
26. R. James Cotton. Differentiable biomechanics unlocks opportunities for markerless motion capture, .
27. Pouyan Firouzabadi, Wendy Murray, Anton R Sobinov, J.D. Peiffer, Kunal Shah, Lee E Miller, and R. James Cotton. Biomechanical arm and hand tracking with multiview markerless motion capture. In *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pages 1641–1648. doi: 10.1109/BioRob60516.2024.10719940. ISSN: 2155-1782.
28. Tim Unger, Arash Sal Moslehian, J. D. Peiffer, Johann Ullrich, Roger Gassert, Olivier Lambercy, R. James Cotton, and Chris Awai Easthope. Differentiable biomechanics for markerless motion capture in upper limb stroke rehabilitation: A comparison with optical motion capture.
29. Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*,

pages 5026–5033. doi: 10.1109/IROS.2012.6386109. ISSN: 2153-0866.
30. Anthony Cimorelli, Ankit Patel, Tasos Karakostas, and R. James Cotton. Validation of portable in-clinic video-based gait analysis for prosthesis users. 14(1):3840. ISSN 2045-2322. doi: 10.1038/s41598-024-53217-7. Publisher: Nature Publishing Group.
31. R. James Cotton, Emoonah McClerklin, Anthony Cimorelli, Ankit Patel, and Tasos Karakostas. Transforming gait: Video-based spatiotemporal gait analysis. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 115–120. IEEE, . ISBN 978-1-7281-2782-8. doi: 10.1109/EMBC48229.2022.9871036.
32. J.D. Peiffer, Kunal Shah, Shawana Anarwala, Kayan Abdou, and R. James Cotton. Fusing uncalibrated IMUs and handheld smartphone video to reconstruct knee kinematics. In *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pages 1275–1282. doi: 10.1109/BioRob60516.2024.10719724. ISSN: 2155-1782.
33. Branko Kopjar, Michael Fehlings, and Beate Hanson. Validity of the Modified Japanese Orthopedic Association Score in Patients with Cervical Spondylotic Myelopathy: The AOSpine North America Multicenter Prospective Study. 11(10):S73–S74. ISSN 1529-9430, 1878-1632. doi: 10.1016/j.spinee.2011.08.183.
34. Lindsay Tetreault, Branko Kopjar, Aria Nouri, Paul Arnold, Giuseppe Barbagallo, Ronald Bartels, Zhou Qiang, Anoushka Singh, Mehmet Zileli, Alexander Vaccaro, and Michael G. Fehlings. The modified Japanese Orthopaedic Association scale: Establishing criteria for mild, moderate and severe impairment in patients with degenerative cervical myelopathy. 26(1):78–84. ISSN 1432-0932. doi: 10.1007/s00586-016-4660-8.
35. R. James Cotton. PosePipe: Open-source human pose estimation pipeline for clinical research, .
36. Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3645–3649. IEEE. ISBN 978-1-5090-2175-8. doi: 10.1109/ICIP.2017.8296962.
37. István Sárándi, Alexander Hermans, and Bastian Leibe. Learning 3d human pose estimation from dozens of datasets using a geometry-aware autoencoder to bridge between skeleton formats. pages 2956–2966.
38. Saeed Ghorbani, Kimia Mahdaviani, Anne Thaler, Konrad Kording, Douglas James Cook, Gunnar Blohm, and Nikolaus F. Troje. MoVi: A large multi-purpose human motion and video dataset. 16(6):e0253157. ISSN 1932-6203. doi: 10.1371/journal.pone.0253157. Publisher: Public Library of Science.
39. R. James Cotton, Allison DeLillo, Anthony Cimorelli, Kunal Shah, J.D. Peiffer, Shawana Anarwala, Kayan Abdou, and Tasos Karakostas. Optimizing trajectories and inverse kinematics for biomechanical analysis of markerless motion capture data. In *2023 International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. IEEE, . ISBN 979-8-3503-4275-8. doi: 10.1109/ICORR58425.2023.10304683.
40. Vittorio Caggiano, Huawei Wang, Guillaume Durandau, Massimo Sartori, and Vikash Kumar. MyoSuite – a contact-rich simulation suite for musculoskeletal motor control.
41. R. James Cotton, Allison DeLillo, Anthony Cimorelli, Kunal Shah, J. D. Peiffer, Shawana Anarwala, Kayan Abdou, and Tasos Karakostas. Markerless motion capture and biomechanical analysis pipeline, .
42. Michael H. Schwartz and Adam Rozumalski. The gait deviation index: A new comprehensive index of gait pathology. 28(3):351–357. ISSN 0966-6362. doi: 10.1016/j.gaitpost.2008.05.001.
43. Michael Marks, Trevor Kingsbury, Richard Bryant, John David Collins, and Marilynn Wyatt. Measuring abnormality in high dimensional spaces with applications in biomechanical gait analysis. 8(1):15481. ISSN 2045-2322. doi: 10.1038/s41598-018-33694-3. Publisher: Nature Publishing Group.
44. John J. Bartko. The intraclass correlation coefficient as a measure of reliability. 19(1):3–11. ISSN 0033-2941. doi: 10.2466/pr0.1966.19.1.3. Publisher: SAGE Publications Inc.
45. J. A. Husted, R. J. Cook, V. T. Farewell, and D. D. Gladman. Methods for assessing responsiveness: a critical review and recommendations. 53(5):459–468. ISSN 0895-4356. doi: 10.1016/s0895-4356(99)00206-1.
46. Matthew H. Liang, Anne H. Fossel, and Martin G. Larson. Comparisons of five health status instruments for orthopedic evaluation. 28(7):632. ISSN 0025-7079.
47. RUPERT G. MILLER. The jackknife-a review. 61(1):1–15. ISSN 0006-3444. doi: 10.1093/biomet/61.1.1.
48. R. James Cotton. Kintwin: Imitation learning with torque and muscle driven biomechanical models enables precise replication of able-bodied and impaired movement from markerless motion capture, 2025.
49. Paweł A. Pierzchlewicz, Caio O. da Silva, R. James Cotton, and Fabian H. Sinz. Platypose: Calibrated zero-shot multi-hypothesis 3d human motion estimation.
50. Scott L. Delp, Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, and Darryl G. Thelen. OpenSim: Open-source software to create and analyze dynamic simulations of movement. 54(11):1940–1950. ISSN 1558-2531. doi: 10.1109/TBME.2007.901024. Conference Name: IEEE Transactions on Biomedical Engineering.
51. Keenon Werling, Nicholas A Bianco, Michael Raitor, Jon Stingel, Steven H Collins, Scott L Delp, and C Karen Liu. AddBiomechanics: Automating model scaling, inverse kinematics, and inverse dynamics from human motion data through sequential optimization. .
52. Farnoosh Koleini, Muhammad Usama Saleem, Pu Wang, Hongfei Xue, Ahmed Helmy, and Abbey Fenwick. BioPose: Biomechanically-accurate 3d pose estimation from monocular videos.
53. Yan Xia, Xiaowei Zhou, Etienne Vouga, Qixing Huang, and Georgios Pavlakos. Reconstructing humans with a biomechanically accurate skeleton.
54. R. James Cotton, Bryant A. Seamon, Richard L. Segal, Randal D. Davis, Amrita Sahu, Michelle M. McLeod, Pablo Celnik, and Sharon L. Ramey. A causal framework for precision rehabilitation, 2024.
55. R. James Cotton and John Rogers. Wearable monitoring of joint angle and muscle activity. 2019:258–263. ISSN 1945-7901. doi: 10.1109/ICORR.2019.8779538.
56. Dimitri Yatsenko, Jacob Reimer, Alexander S. Ecker, Edgar Y. Walker, Fabian Sinz, Philipp Berens, Andreas Hoenselaar, R. James Cotton, Athanassios S. Siapas, and Andreas S. Tolias. Datajoint: managing big scientific data using matlab or python. *bioRxiv*, 2015. doi: 10.1101/031658.
57. Kazuo Yonenobu, Kuniyosi Abumi, Kensei Nagata, Eiji Taketomi, and Kazumasa Ueyama. Interobserver and intraobserver reliability of the japanese orthopaedic association scoring system for evaluation of cervical compression myelopathy. 26(17). ISSN 0362-2436.
58. So Kato, Yasushi Oshima, Hiroyuki Oka, Hirotaka Chikuda, Yujiro Takeshita, Kota Miyoshi, Naohiro Kawamura, Kazuhiro Masuda, Junichi Kunogi, Rentaro Okazaki, Seiichi Azuma, Nobuhiro Hara, Sakae Tanaka, and Katsushi Takeshita. Comparison of the japanese orthopaedic association (JOA) score and modified JOA (mJOA) score for the assessment of cervical myelopathy: A multicenter observational study. 10(4):e0123022. ISSN 1932-6203. doi: 10.1371/journal.pone.0123022.
59. Firas Al-Hafez, Guoping Zhao, Jan Peters, and Davide Tateo. LocoMuJoCo: A comprehensive imitation learning benchmark for locomotion.
60. Samuel R. Hamner, Ajay Seth, and Scott L. Delp. Muscle contributions to propulsion and support during running. 43(14):2709–2716. ISSN 0021-9290. doi: 10.1016/j.jbiomech.2010.06.025.
61. Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: a skinned multi-person linear model. 34(6):1–16. ISSN 0730-0301, 1557-7368. doi: 10.1145/2816795.2818013.
62. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need.
63. Kevin M. Lynch and Frank C. Park. *Modern robotics: mechanics, planning, and control*. Cambridge University Press. OCLC: ocn983881868.
64. Keenon Werling, Michael Raitor, Jon Stingel, Jennifer L. Hicks, Steve Collins, Scott L. Delp, and C. Karen Liu. Rapid bilevel optimization to concurrently solve musculoskeletal scaling, marker registration, and inverse kinematic problems for human motion reconstruction, . Pages: 2022.08.22.504896 Section: New Results.
65. Patrick Kidger and Cristian Garcia. Equinox: neural networks in JAX via callable PyTrees and filtered transformations.
66. DeepMind, Igor Babuschkin, Kate Baumli, Alison Bell, Surya Bhupatiraju, Jake Bruce, Peter Buchlovsky, David Budden, Trevor Cai, Aidan Clark, Ivo Danihelka, Antoine Dedieu, Claudio Fantacci, Jonathan Godwin, Chris Jones, Ross Hemsley, Tom Hennigan, Matteo Hessel, Shaobo Hou, Steven Kapturowski, Thomas Keck, Iurii Kemaev, Michael King, Markus Kunesch, Lena Martens, Hamza Merzic, Vladimir Mikulik, Tamara Norman, George Papamakarios, John Quan, Roman Ring, Francisco Ruiz, Alvaro Sanchez, Laurent Sartran, Rosalia Schneider, Eren Sezener, Stephen Spencer, Srivatsan Srinivasan, Miloš Stanojević, Wojciech Stokowiec, Luyu Wang, Guangyao Zhou, and Fabio Viola. The DeepMind JAX Ecosystem, 2020.
67. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization.
68. S. S. SHAPIRO and M. B. WILK. An analysis of variance test for normality (complete samples)†. 52(3):591–611. ISSN 0006-3444. doi: 10.1093/biomet/52.3-4.591.

Supplementary Methods

Datasets.

MMMC Dataset. This study was approved by the Northwestern University Institutional Review Board. Data were collected in our laboratory using handheld smartphone video (30), an instrumented GaitRite walkway, and an MMMC system (41). The dataset includes recordings from 33 control participants, 48 lower limb prosthesis users (LLPUs), and 73 individuals with a history of neurological injury, including 6 pediatric cases. All recording sessions consisted of overground walking along the GaitRite walkway (8 m) while a researcher followed the participant with smartphone.

Control participants typically completed three recording sessions. The first session included nine walking trials, while subsequent sessions incorporated additional assessments such as the Four Square Step Test (FSST), Postural Sway Test (PST), Timed Up and Go (TUG), and Tandem Gait, averaging 15 recordings per session. LLPUs were generally recruited from an outpatient clinic for a single session including the additional tasks if the participant felt comfortable. The neurologic participants were largely outpatients (40) collected with both PBL and MMMC. We also analyzed single-view videos from MMMC data collected from 27 participants admitted to inpatient rehabilitation using static cameras installed in therapy gyms at Shirley Ryan AbilityLab. These participants walked at a self selected speed using whatever assistive device they generally used (e.g. walker, cane, ankle-foot orthosis, gait belt).

Portable Biomechanics Laboratory (PBL) All smartphone RGB video was captured on a Samsung Galaxy S8 or Samsung Galaxy S20 Ultra at 30 Hz and 1080 $\times$ 1920 resolution (30) using a custom application. This application simultaneously logs smartphone RGB video, and if available, depth video, internal phone sensor logs (gyroscope, magnetometer, accelerometer, and fused orientation estimate), and wearable sensor data (55). During walking trials, a researcher held the phone in portrait orientation mounted on a 3-axis gimbal and followed the participant from behind at a mostly sagittal, but partially oblique view. The gimbal was used solely to improve visualizations; the method remains accurate without it, though the resulting video may be slightly shakier. FSST was collected at a frontal view in front of the participant, and TUG was collected from a sagittal view with the camera rotating to follow the participant.

The video and sensor data from the smartphone application was uploaded to a secure cloud server for storage and later downloaded into PosePipe (35), an open source application implementing a variety of computer vision tasks such as person detection, annotation, and 2D and 3D keypoint detection. PosePipe tracks all data and analyses in a MySQL database using DataJoint (56). This makes it feasible to manage the 100s of thousands of videos we have collected.

Multi-View Markerless Motion Capture (MMMC) Recordings in the laboratory dataset employed the Multi-View Markerless Motion Capture (MMMC) system which consists of 8-12 FLIR BlackFly S GigE cameras which acquire synchronized RGB video at 29 fps. Cameras were arranged such that at least three cameras covered participants at all points during a recording. A detailed description of this system can be found in (41). For this system, we reconstructed biomechanical fits using methods described in (26). In total, the dataset includes 1,267,586 frames across 2,206 trials from 143 participants—comprising 1,056,861 frames from 2,028 trials of 116 participants recorded with both the PBL and MMMC systems, and an additional 210,725 frames from 178 trials of 27 participants recorded with a single MMMC camera—amounting to 11.7 hours of video.

OMC Dataset: BML-MoVi. The publicly available BML-MoVi (38) dataset contains synchronous optical motion capture (OMC) data and smartphone RGB video data of 90 able-bodied participants performing a variety of everyday and sports movements. The OMC data contains a dense set of 87 markers, suitable for reconstructing biomechanics recorded at 120 Hz.

Smartphone video was collected at 30 Hz and 1080 x 1920 resolution using an iPhone 7. The smartphone video totals 415,060 frames for 3.8 hours of video.

For the OMC data, we fit our model using only the 3D keypoint loss as the markers had already been triangulated and cleaned, following the approach from (26). For the smartphone data, we apply the 3D and 2D keypoint loss following the method we use on the PBL data and describe below.

In-Clinic Dataset. This study was approved by the Northwestern University and Washington University in St. Louis Institutional Review Boards. To validate our methods in a real-world clinical environment, we deployed our PBL system (30) in two outpatient clinics. The first was a neurosurgery clinic treating patients with cervical myelopathy (CM)—a degenerative condition of the cervical spine that often impairs gait and balance. During routine pre- and post-operative visits for spinal decompression, participants were asked to walk three times at a self-selected pace and three times at a fast pace down the clinic hallway while being recorded from behind by a clinician. At each visit, patients completed the modified Japanese Orthopedic Association (mJOA) questionnaire (57, 58) to assess their symptoms. The CM cohort includes 911 videos from 55 patients collected at multiple timepoints: pre-surgery, 6 weeks post-surgery, 3 months post-surgery, and 1 year post-surgery. 17 of these participants have sessions before and after surgery, which were used in our sensitivity analysis.

The second cohort consisted of 110 videos of 19 participants with knee osteoarthritis (KOA) coming to clinic for corticosteroid injection. As before, these videos were filmed in the hallway before or after the visit.

In total, this dataset contains 589,557 frames or 5.5 hours of video in clinical settings.

Keypoint Detection. We employed MeTRAbs-ACAE (37) to detect 2D and 3D virtual marker locations from RGB videos. As recommended by the author, we approximated confidence by measuring the standard deviation of each 3D joint location estimated from 10 different augmented versions of each video frame. This was converted to a confidence estimate using a sigmoid function with a half maximum at 30 mm and a width of 10 mm. This results in a confidence score $c(t, j) \in [0, 1]$ for each timepoint t and joint j. For training, we only used frames where the person was fully in the video view, allowing for short periods ($<$1.0) seconds of partial coverage, reducing total dataset size to 12.4 hours.

Differentiable Biomechanical Model. Smartphone and MMMC reconstructions utilized a biomechanically-grounded model implemented in Mujoco from LocoMujoco (59), originally based on an OpenSim model (60). Mujoco's GPU acceleration engine, MJX, supports parallelizing forward kinematic passes of this model. We have previously optimized site locations for the 87 MoVi keypoints on a similar model (39). We also removed collisions other than the feet, added a neck joint with 3 degrees of freedom, and extended the hip flexion and extension ranges. Model scaling is controlled by 8 scaling parameters: overall size, the pelvis, left thigh, left leg and foot, right thigh, right leg and foot, the left arm, and the left leg. We follow notation used by the skinned multi-person linear model (SMPL) (61) and represent our GPU-accelerated, biomechanically-ground, forward kinematic equation as:

$$\mathbf{x} = \mathcal{M}(\theta, \beta) \tag{1}$$

where $\theta \in \mathbb{R}^{40}$ are joint angles that pose the model, $\beta \in \mathbb{R}^{8+87\times3}$ are the 8 scaling parameters and an offset for each of the 87 keypoints. $\mathbf{x} \in \mathbb{R}^{87\times3}$ are the marker locations following the scaling, site offset and forward kinematics.

Implicit Representation. We represent the kinematic trajectory for each trial as a learned implicit function, f_θ, implemented as a multi-layer perceptron that takes time as an input using sinusoidal positional encoding (62) and outputs joint angles $\theta(t)$. In addition to the 40 joint angles output from the final layer, we also output 3 parameters representing the orientation of the smartphone in global space $\mathbf{r}$.

$$f_\phi : t \rightarrow (\theta, \mathbf{r}) \tag{2}$$

The outputs of f_ϕ corresponding to rotations (i.e., not the pelvis location) are passed through a tanh nonlinearity to limit it to $(-1, 1)$ followed by scaling this range to match the biomechanical model joint limits. Joint angles, $\theta(t)$, are then passed through the forward kinematic equation above (Eq. 1) to obtain the 3D joint locations, $\mathbf{x}(t)$, at that time.

Reference frames. We define the global reference frame $\{n\}$ for our model output. Since the PBL system tracks changes in phone orientation from the smartphone IMUs, we are able to relate the orientation of the camera frame $\{c\}$ to the orientation of the world frame at any point in time using $R_{nc} \in SO(3)$. We follow the convention presented in (63) for rotations where R_{ab} represents frame $\{b\}$ relative to frame $\{a\}$.

Model Fitting. Each trial is represented as a unique implicit function. We jointly optimize the parameters of the implicit representations for N trials $\{\phi_0, \phi_1, ..., \phi_{N-1}\}$, the body scaling, and site offset parameters β, to follow 2D and 3D keypoints extracted from video as well as orientation data recorded from the smartphone. This approach of jointly learning kinematics and body scaling and marker offset for the entire session has been called bilevel optimization (64). For example, a session consisting of 10 trials would jointly optimize 10 implicit functions f_{θ_i} and a single set of body scaling parameters β.
For each training step, we evaluated every implicit function with a batch of 300 time samples for each trial to extract the predicted joint angles at those time points $\hat{\theta}(t)$, which is performed in parallel across all trials. Using these poses and scaling parameters, we performed GPU accelerated forward kinematic passes of our model (Eq. 1) to predict 3D joint locations for that training batch, $\hat{\mathbf{x}} = \mathcal{M}(\hat{\theta}(t), \hat{\beta})$. Next, we describe the losses used during this optimization. To reduce notational clutter, we drop the explicit time and trial dependence in the loss definitions.

3D Keypoint Loss. We obtain a pure video estimate of the 3D keypoint locations in the camera reference frame x_c using MeTRAbs-ACAE (37). We can rotate these keypoints into the global frame using $x_n = \hat{R}_{nc} x_c$ and define a loss function on the Euclidean distance in centimeters between keypoints from the model $\hat{x}_n$ with the video keypoints as:

$$\mathcal{L}_{3D}(\phi, \vec{\beta}) = \frac{1}{J} \sum_{j \in J} c(j) g(\|\hat{x}_n - \hat{R}_{nc} x_c\|_2) \tag{3}$$

where $c(j) \in [0, 1]$ is the confidence score for joint keypoint j, which also varies with time, and $g(\cdot)$ is a Huber loss which is quadratic within 10 cm and linear after, necessary for stabilizing early training. This loss is computed between the 3D keypoints set with their mean translation removed.

2D Keypoint Loss. MeTRAbs-ACAE also produces 2D keypoints in the image frame $\vec{u} \in \mathbb{R}^{87\times 2}$. The 3D keypoints produced by our model $\hat{x}_n$ can be rotated into the camera frame using $\hat{R}_{nc}$ and projected through a camera model Π with the calibrated camera intrinsics, to compute the error in pixels with detected 2D keypoints:

$$\mathcal{L}_{2D}(\phi, \vec{\beta}) = \frac{1}{J}\sum_{j\in J} c(j)g(\|\Pi(\hat{R}_{nc}^{-1}\hat{x}_n) - \vec{u}\|_2) \tag{4}$$

In this loss, $g(\cdot)$ is a Huber loss which is quadratic within 5 pixels and linear after, which reduces the sensitivity of the fits to outliers.

Phone orientation loss. For videos collected with our PBL platform that include phone orientation data, we evaluate the implicit functions at the phone IMU sample points and extract the predicted phone orientation trajectory $\hat{R}_{nc}(t)$. We found representing R_{nc} as a rotation vector led to more stable training; however, the phone orientation output from internal sensors was represented as a quaternion, so we convert R_{nc} to a quaternion for calculating the angular difference between the predicted and measured orientation, measured in degrees with the following loss:

$$\mathcal{L}_{phone} = \frac{180}{\pi}\cdot 2\cdot \arctan\left(\frac{\sqrt{q_1^2+q_2^2+q_3^2}}{|q_0|}\right) \tag{5}$$

Total loss and optimization. These terms were combined with hyperparameters ($\lambda_1 = 1, \lambda_2 = 1e-1, \lambda_3 = 1$), to control their relative weights and provide an overall loss:

$$\mathcal{L} = \lambda_1\mathcal{L}_{3D} + \lambda_2\mathcal{L}_{2D} + \lambda_3\mathcal{L}_{phone} \tag{6}$$

We defined implicit functions and computed the loss using Equinox (65) and JAX (66), optimizing with the Adam optimizer for 25,000 iterations of gradient descent. Weight decay of $1e{-}5$ was applied to the implicit function parameters. With Optax, we set the learning rate to start at $1e{-}3$ and decay exponentially to $1e{-}6$ (66, 67).
For MMMC (26), we found random initialization of the network sufficient for optimizing implicit representations. However, for monocular videos, this often failed to converge if the subject wasn't initially visible to the camera. To address this, we adjusted each trial's implicit function by biasing the final layer to place the pelvis 1.5 meters from the camera and initializing R_{nc} outputs to the median of observed R_{nc} values.

Gait Deviation Index. The GDI maps cycle-aligned joint angles to a lower-dimensional subspace in which we can measure distance from a normative population. We computed gait event timings for every self-selected gait trial using a pretrained transformer from our prior work (31), after discarding the first and last step. This generated 13,609 gait cycles from the PBL system. Of these, 1,267 were of control participants and we use these and 4,675 gait cycles from our data collection at the American Society of Biomechanics meeting and 3,066 cycles from our MMMC system as normative data.
As PBL most accurately captures hip flexion, hip adduction, and knee flexion angles, we only used these kinematic traces in the GDI. Each kinematic trace was interpolated to 50 points at 2% increments across the gait cycle and concatenated to a 150×1 column vector. We extract right-sided waveforms aligned between right initial contact events and aligned left-sided waveforms to the left foot initial contact.
While the original GDI (42) used singular value decomposition to reduce this high dimensional data, more recent findings (43) demonstrate that principal component analysis (PCA) finds a more suitable subspace. We fit PCA to our $150 \times 21{,}350$ matrix, finding that 12 components accounted for 95% of the variance.
This resulted in a GDI value for every step a subject took. We averaged a participant's GDI over all steps (left and right) they took on a given day to compute a singular GDI number for each participant and session.

Statistics. To compare population-level GDIs, we first assessed normality using the Shapiro-Wilk test (68). For comparisons between two normally distributed populations, we used Student's t-test. When at least one distribution was non-Gaussian, we used the Mann-Whitney U test to assess group differences.
To evaluate relationships between smartphone-based metrics and clinical scales, we used different correlation measures based on variable type. For associations between two continuous variables (e.g., GDI and 10MWT speed), we computed the Pearson correlation coefficient. For associations between a continuous variable and an ordinal or categorical variable (e.g., GDI and mJOA), we used the Spearman rank correlation coefficient.
To assess the responsiveness of various measures to surgical intervention in the CM cohort, we computed the Standardized Response Mean (SRM) (45, 46), defined as:

$$SRM = \frac{\bar{D}_x}{SD(D_x)} \tag{7}$$

Supplementary Table 1. Median Joint Angle Errors (median (nIQR)) for clinical and BML-MoVi datasets.

Dataset	Clinical		BML-MoVi
Camera	Handheld	Static	Static
Hip Flexion	3.93 (2.35)	5.37 (2.97)	4.80 (2.68)
Hip Adduction	1.69 (0.74)	1.97 (0.84)	1.68 (0.66)
Hip Rotation	2.70 (1.00)	2.73 (0.94)	3.30 (1.22)
Knee Angle	4.05 (1.70)	3.92 (2.12)	2.11 (0.53)
Ankle Angle	5.38 (2.20)	4.94 (1.97)	3.55 (1.25)
Lumbar Extension	3.09 (1.93)	4.53 (3.86)	5.30 (3.00)
Lumbar Bending	1.43 (0.65)	1.59 (0.76)	1.64 (0.86)
Lumbar Rotation	1.91 (1.11)	1.78 (0.85)	1.63 (0.84)
Neck Extension	4.34 (4.40)	4.06 (2.97)	3.95 (2.94)
Neck Bending	1.88 (1.23)	2.05 (1.35)	2.31 (1.02)
Neck Rotation	3.01 (2.40)	2.63 (1.62)	2.85 (1.41)
Arm Flex	3.63 (2.16)	3.14 (1.74)	2.31 (0.97)
Arm Add	1.26 (0.70)	1.94 (1.18)	1.94 (0.67)
Elbow Flex	4.22 (2.38)	4.09 (2.06)	3.85 (1.18)
All	2.79 (0.86)	2.96 (0.78)	2.74 (0.50)

where D_x is the vector of pre- to post-surgery differences in measure x, $\bar{D}_x$ is the mean of these differences, and $SD(D_x)$ is their standard deviation. We calculated SRMs for the mJOA, cadence, double support time, and GDI to compare their sensitivity to change.

To estimate confidence intervals for the SRM, we employed the jackknife technique (47) as was done in (46). This nonparametric resampling method systematically recomputes the SRM by omitting one observation at a time, producing a set of jackknife replicates. These replicates are then used to compute pseudo-values, which approximate the influence of each observation on the overall SRM. The mean of these pseudo-values serves as a bias-corrected point estimate of the SRM, and their variability provides an estimate of its standard error. A t-distribution is then used to construct confidence intervals, enabling inference about the SRM's precision without assuming normality of the underlying data.

We defined significance as $p < 0.05$ in all cases.

PBL vs MMMC GDI. To assess the impact of recording system (MMMC vs. PBL) on GDI, we conducted a separate analysis using only sessions recorded simultaneously with both systems. For each session, we computed GDI twice: once using PBL data with a PBL-based normative distribution, and once using MMMC data with an MMMC-based normative distribution. We then compared the resulting GDI values for each participant across the two systems.

Fitting Residuals. We measure the 2D Fitting Residual and 3D fitting residual to evaluate the closeness of biomechanical fits to video-based joint locations. The 2D fitting residual evaluates the reconstructed model's agreement between the video-based 2D keypoints by projecting the reconstruction through a known camera model and comparing the Euclidean distance in the pixel space. The 3D Fitting Residual compares Euclidean error between 3D keypoints of the biomechanical reconstruction and video-based 3D keypoints or 3D markers.

Supplementary Results

Detailed Analysis of Joint Angle Errors. Overall, median joint angle error (MJAE) was approximately 3° across datasets, populations, and activities. Below, we highlight key trends and deviations from this general pattern, offering possible explanations and practical takeaways for users. However, we do not perform extensive statistical testing across the many variables examined, as the focus of this work is on enabling accessible motion capture rather than achieving the most precise measurement possible.

Across datasets, the BML-MoVi dataset exhibited the lowest MJAE (2.74°; Table S1, Fig. 4C), which is expected since it included only control participants. The MMMC dataset showed similar errors when using a handheld camera (2.79°), but higher errors with a static camera (2.96°). This increase is likely due to the greater distance between the participant and camera in the static setup, whereas the handheld condition involved a researcher following the participant more closely (Fig. 2). Some of the added error in the MMMC dataset may also be attributable to noise in the MMMC system itself, despite the use of multiple cameras.

When comparing clinical populations within the MMMC dataset, control participants showed the lowest MJAE (2.51°), while participants with neurological conditions showed the highest (3.32°; Table S2). This is consistent with expectations: many

Supplementary Table 2. Median Joint Angle Error (median (nIQR)) for handheld smartphone reconstruction across different clinical populations.

Joint	MJAE (nIQR)			
	Control	Neurological	LLPU	Pediatric
Hip Flexion	3.49 (2.27)	3.92 (2.58)	4.69 (2.10)	3.06 (1.32)
Hip Adduction	1.55 (0.69)	1.93 (0.78)	1.80 (0.80)	1.92 (0.27)
Hip Rotation	2.66 (1.17)	2.83 (0.79)	2.70 (0.93)	2.74 (0.61)
Knee Flexion	3.77 (1.44)	4.36 (1.48)	4.68 (2.08)	3.89 (0.90)
Ankle Angle	5.49 (2.75)	5.41 (1.60)	5.28 (1.89)	5.30 (1.24)
Lumbar Extension	3.13 (2.22)	3.29 (1.85)	2.89 (1.51)	4.09 (2.28)
Lumbar Bending	1.38 (0.64)	1.58 (0.65)	1.44 (0.66)	1.50 (0.37)
Lumbar Rotation	1.71 (1.13)	2.19 (1.08)	2.06 (1.02)	1.95 (0.34)
Neck Extension	3.62 (2.76)	10.12 (6.58)	4.27 (4.49)	7.16 (3.60)
Neck Bending	1.62 (1.12)	2.49 (1.46)	1.88 (1.11)	2.37 (2.29)
Neck Rotation	2.78 (2.11)	5.00 (3.78)	2.63 (1.90)	3.95 (2.54)
Arm Flex	3.01 (1.61)	4.33 (3.03)	4.43 (2.68)	4.00 (1.47)
Arm Add	1.06 (0.54)	1.71 (1.28)	1.47 (0.73)	1.74 (0.95)
Elbow Flex	3.59 (2.09)	4.79 (2.47)	4.90 (2.54)	5.18 (2.46)
All	2.51 (0.78)	3.32 (0.92)	2.97 (0.79)	3.17 (0.48)

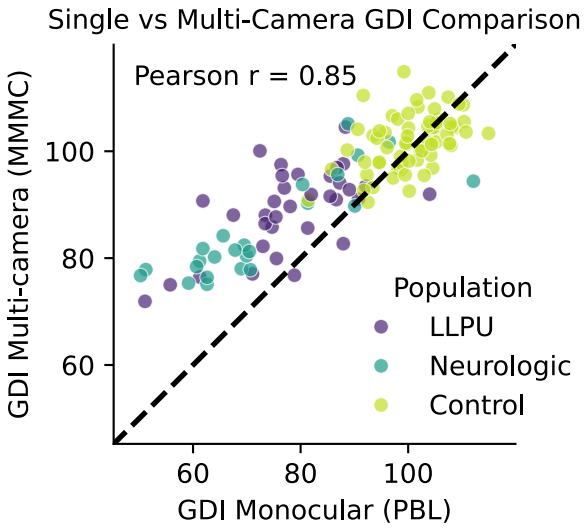
neurological participants used assistive devices—such as canes, walkers, or ankle-foot orthoses—which can obstruct joint visibility and hinder keypoint detection. Additionally, physical assistance from a therapist using a gait belt may contribute to further occlusion and error. Lower limb prosthesis users (LLPUs) also exhibited elevated MJAE (2.97°), reflecting known limitations of keypoint detectors when applied to prosthetic limbs (30).

Across different activities, MJAE remained relatively stable (Fig. 4A). Unsurprisingly, standing had the lowest MJAE (1.80°), followed by the Four Square Step Test (FSST; 2.47°), both of which involve minimal occlusion. In contrast, activities such as the Timed Up and Go (3.17°), tandem gait (3.22°), and walking (2.95°) yielded slightly higher errors—likely due to partial leg occlusions caused by the oblique handheld filming angle used during much of these trials.

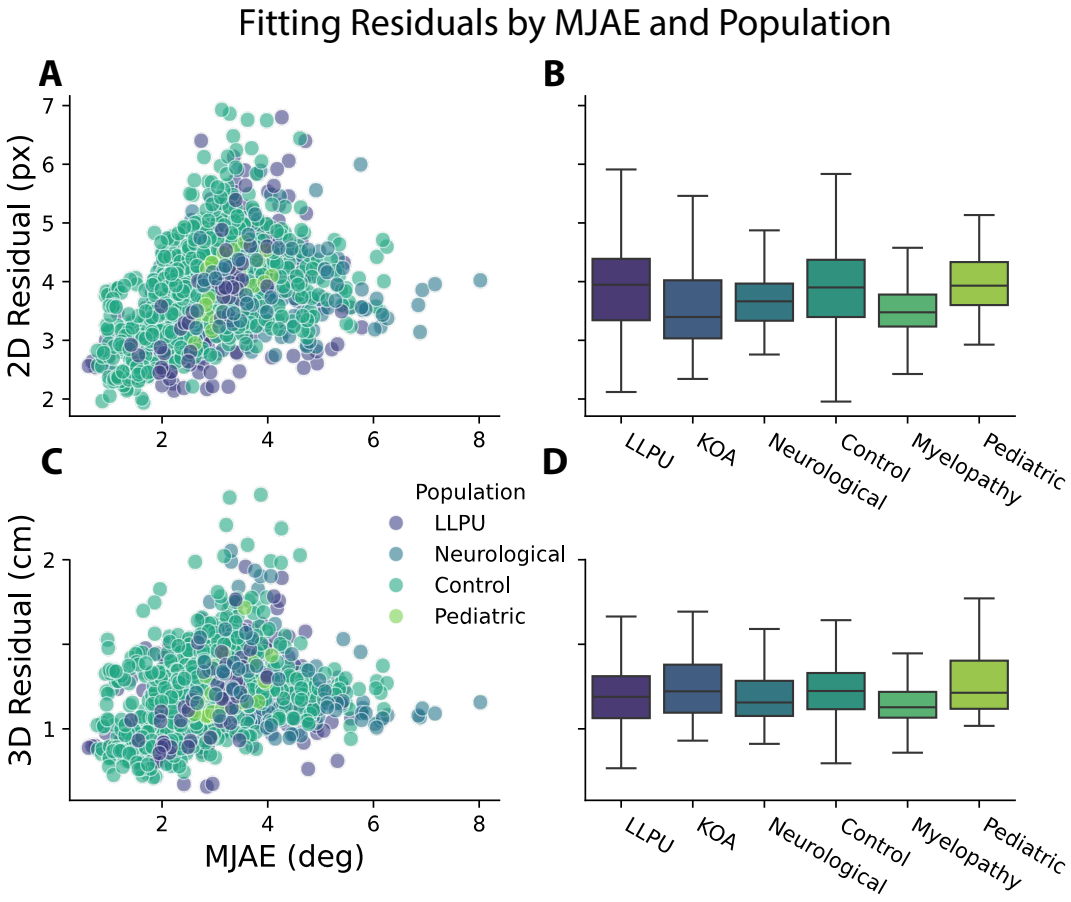
Joint-level trends in MJAE were largely consistent across populations (Table S2). Notably, ankle MJAE exceeded 5° in all populations, indicating a clear need for improved foot keypoint detection. Elbow flexion error approached 5° in clinical cohorts, likely because one arm is frequently occluded during walking. Hip flexion error was higher than expected in the BML-MoVi and static-camera MMMC datasets (Table S1). These datasets featured more frontal views of the participant, which—as shown in our viewpoint analysis (Fig. 3C)—can increase sagittal plane error. Interestingly, this often manifested as a consistent offset rather than an error in the joint's range of motion. A more detailed investigation of these effects, including marker placement and viewpoint-dependent bias in keypoint detection, is left for future work.

PBL vs MMMC GDI. GDI values computed from PBL and MMMC recordings of the same sessions were highly consistent, despite differences in recording modality and normative datasets. We found a strong, significant, correlation between the two different modalities (0.85) demonstrating that PBL and MMMC both likely capture the same features needed for downstream gait analysis tasks (Fig. S1). We note that the PBL-based GDI shows a greater change than the MMMC system, which could suggest some of the deviations in gait detected by the PBL system arise from measurement error from the PBL system; this is more impactful for greater gait impairments.

Fitting residuals. Across all datasets and modalities, 2D and 3D residuals between reconstructed and detected joint locations were generally low at <5 pixels for 2D reprojection errors and <1.5 cm for 3D marker errors. Correlations between fitting residuals and MJAE were significant but low: 0.35 and 0.24 for 2D and 3D residuals, respectively (Fig. S2A,C). While this suggests there may be some relationship between the closeness of our model fit to the actual joint angle accuracy, it is quite weak, at least at the levels most of our reconstructions achieved. This likely highlights the fact that the underlying computer vision algorithm does not have well calibrated confidence scores, something we aim to address in the future (25, 49). Importantly, we do not see any obvious issues with our approach generalizing to clinical settings as there were not large differences in either residual between in-lab and in-clinic cohorts (Fig. S2B,D).



Supplementary Figure 1. Gait Deviation Index Comparison between Monocular and Multi-View Fits GDI values were computed separately using either PBL (monocular) or MMMC (multi-view) recordings from the same sessions. For each computation, we used only the data and normative distribution corresponding to the respective modality (e.g., monocular GDI was calculated using single-camera fits and a single-camera-based normative distribution).



Supplementary Figure 2. Fitting Residuals. Correlation between 2D and 3D fitting residuals and MJAE is shown in (A), (C), respectively. 2D and 3D fitting residuals distributed across clinical populations is shown in (B) and (D), respectively.