

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

J.D. Peiffer^{1,2}, Kunal Shah¹, Irina Djuraskovic^{1,3}, Shawana Anarwala¹, Kayan Abdou¹, Rujvee Patel⁴, Prakash Jayabalan^{1,5}, Brenton Pennicooke⁴, and R. James Cotton^{1,5,✉}

¹Shirley Ryan AbilityLab, Chicago, IL, USA

²Department of Biomedical Engineering, Northwestern University, Evanston, IL, USA

³Interdepartmental Neuroscience, Northwestern University, Chicago, IL

⁴Department of Neurological Surgery, Washington University School of Medicine, St. Louis, MO, USA

⁵Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Movement directly reflects neurological and musculoskeletal health, yet objective biomechanical assessment is rarely available in routine care. We introduce Portable Biomechanics Laboratory (PBL), a secure platform for fitting biomechanical models to video collected with a handheld, moving, smartphone. We validate this approach on over 15 hours of data synchronized to ground truth motion capture, finding mean joint-angle errors $< 3^\circ$ and pelvis-translation errors of a few centimeters across patients with neurological-injury, lower-limb prosthesis users, pediatric in-patients, and controls. In > 5 hours of prospective deployments to neurosurgery and sports-medicine clinics, PBL was easy to setup, yielded highly reliable gait metrics (ICC > 0.9), and detected clinically relevant differences. For cervical-myelopathy patients, its measurement of gait quality correlated with modified Japanese Orthopedic Association (mJOA) scores and were responsive to clinical intervention. Handheld smartphone video can therefore deliver accurate, scalable, and low-burden biomechanical measurement, enabling greatly increased monitoring of movement impairments. We release the first clinically-validated method for measuring whole-body kinematics from handheld smartphone video at <https://IntelligentSensingAndRehabilitation.github.io/MonocularBiomechanics/>.

Correspondence: [rcotton {at} sralab.org](mailto:rcotton@at.sralab.org)

Introduction

Many clinical conditions have pronounced movement phenotypes which are rarely measured in clinical practice [1, 2, 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 also 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, 15,

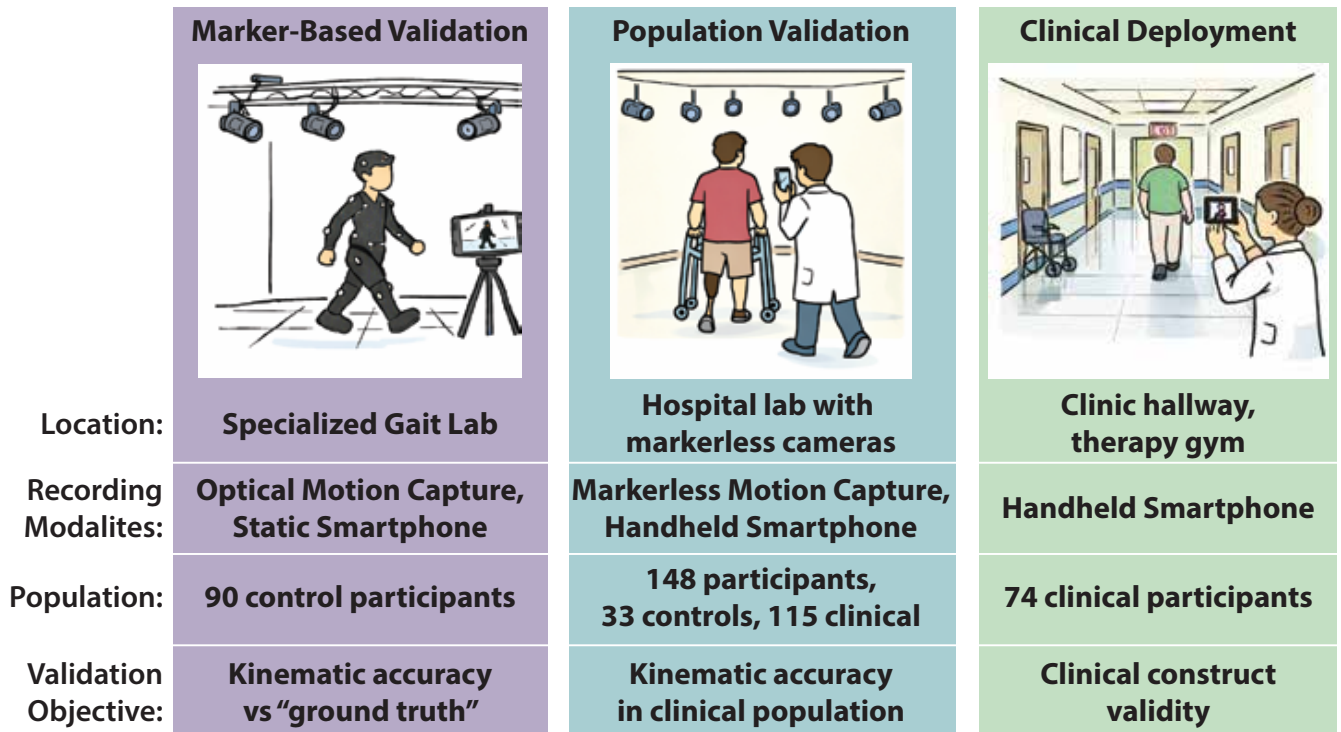


Fig. 1. Study Design Overview. We first validated the Portable Biomechanics Laboratory (PBL) against an optical motion capture dataset in which control participants performed a range of upper- and lower-limb activities. We then evaluated PBL in clinical populations by recording participants performing lower-limb tasks using both multi-camera markerless motion capture and PBL in hospital laboratories and on clinical floors. Finally, we deployed PBL in clinic hallways and therapy gyms, recording participants after clinical visits to assess the relationship between PBL-derived gait quality scores and standard clinical measures.

16, 17, 18, 19, 20, 21, 22, 23]. We have developed methods for MMMC [24, 25, 26, 27] using MuJoCo [28], 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 multi-stage marker-based pipelines that first estimate virtual marker locations and then compute inverse kinematics on these trajectories [25, 26]. Although MMMC is faster than OMC, installing and calibrating several high-performance cameras remains a significant barrier to clinical implementation.

OpenCap substantially lowers the technical burden of MMMC by using two or three tripod-mounted smartphones, achieving < 5 degrees of joint-angle error relative to OMC [29] and detecting movement patterns that classify disease [30, 23]. While OpenCap is a significant and widely used contribution, the requirement for two tripods placed at ± 45 degrees is not always feasible in clinical settings, which often include small exam rooms and tight hallways. In these contexts, a single handheld camera offers a more practical alternative. Monocular video analysis has been used clinically with people with Parkinson’s disease [31], pediatric patients with cerebral palsy [32], and stroke [33]. However, these approaches infer 2D or 3D joint locations, or non-biomechanically-grounded mesh representations, rather than biomechanical joint angles used in traditional movement analysis. Recent methods [34, 35] have explored data-driven biomechanical reconstruction from a single camera, but have not been trained or validated on clinical populations. To enable data capture in real-world clinical settings, we previously developed a smartphone application for collecting handheld videos in clinics [36], which we refer to as the Portable Biomechanics Laboratory (PBL), and validated gait event detection from these recordings [37]. We subsequently demonstrated our end-to-end biomechanical pipeline can reconstruct biomechanically-grounded kinematics from handheld smartphone video, with or without wearable sensors, achieving accurate knee kinematics from video alone [38].

The central goal of the present study is to evaluate whether this smartphone-based biomechanics pipeline can produce accurate kinematics across diverse patient populations, movement tasks, and clinical environments that show construct validity against clinical metrics. We perform this validation by first evaluating our system against OMC in control populations, next against MMMC in clinical populations, and finally, our intended use case, collect data in actual clinic hallways to compare against clinical scores (Fig. 1). Overall, we find PBL has high joint-angle

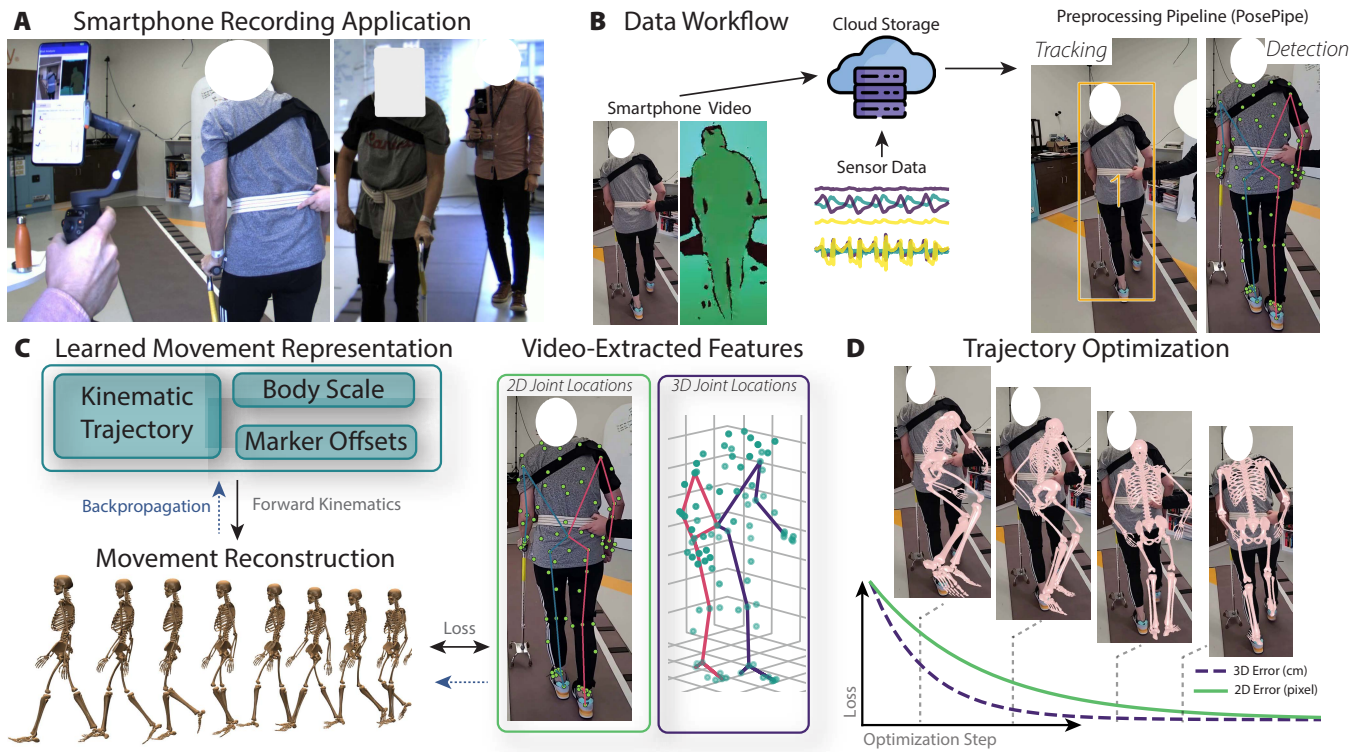


Fig. 2. Biomechanical Reconstruction Overview. We introduce a method for biomechanically grounded movement analysis in clinical settings using a handheld smartphone. **A)** Researchers held a smartphone (optionally with gimbal) while following a participant walking. Our system has no specific requirements 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 compared to OMC and MMMC, concordance with clinical scores, and is responsive to clinical intervention. Together, these evaluations demonstrate that quantitative movement analysis can be integrated into routine clinical workflows using a single handheld smartphone for accurate and clinically relevant biomechanical measurement.

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. 2A). The application can optionally be connected to wearable IMUs with all modalities closely synchronized [36, 38]. After a recording is finished, PBL uploads the data to a secure cloud platform where data can be pulled into PosePipe [39], an open-source package we developed for managing and processing videos.

Video Preprocessing. Using PosePipe [39], we identified the subject of interest in each video using DeepSortYOLOv4 [40]. Next, we extracted 2D and 3D virtual marker locations, termed “keypoints”, from each video frame using MeTRAbs-ACAE [41] (Fig. 2B). We previously identified the 87 keypoints from the MoVi dataset [42] to be sufficient for full-body kinematics [43].

Biomechanical Fitting. We obtained kinematics by optimizing the fit of biomechanical models using our end-to-end differentiable biomechanics framework [38, 25]. We note that this study estimated joint kinematics only; no

inverse dynamics or kinetic estimation was performed. A complete description is in the Supplemental Methods; here we provide a brief overview (Fig. 2C,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 [28, 44].

Run Time. Preprocessing a 6 second video takes 10 seconds on a server-grade GPU, 1 minute on a laptop-grade GPU and scales linearly with video length. Biomechanical fitting takes 1 minute on a server-grade GPU, <5 minutes on a laptop-grade GPU, and does not scale with video length.

Datasets. We applied our monocular fitting approach to three datasets (Fig. 1). The first is a publicly available dataset that used iPhone videos and an OMC system, but only contained able-bodied individuals. The second of these recorded clinical participants with the PBL system and also an MMMC system. 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. The OMC dataset used was BML-MoVi [42], 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 both modalities of this dataset using our end-to-end differentiable biomechanics approach [25], applying the same biomechanical model used for the MMMC and PBL datasets.
2. 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 [36] and our MMMC system [45, 25]. 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. For the BBS, we classify a score of ≤ 50 as a fall risk [46] and ≤ 40 as high fall risk [47].
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 [48, 49], which scores upper and lower extremity function. Seventeen CM participants were recorded at visits both before and after surgery.

Accuracy Evaluation. To evaluate the accuracy of the smartphone-based fits compared to OMC and MMMC, we computed the Median Absolute Error, defined as the median of joint-angle errors within each trial followed by the median of these per-trial medians across trials. We used Median Absolute Error because both the within-trial and across-trial error distributions were non-Gaussian. When aggregating errors across trials, we report the median and the normalized interquartile range (nIQR), where $nIQR = IQR \cdot 0.7413$.

To compare moving (handheld) vs. static camera video, we also computed monocular fits from single cameras in the MMMC dataset. Because MMMC records participants from all directions, we examined the influence of camera viewpoint on joint-angle error by classifying viewpoint as front, back, ipsilateral, or contralateral.

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

For comparison to OpenCap and other methods, we also compute the Mean Absolute Error following the procedure in [29], detailed in the Supplementary Methods.

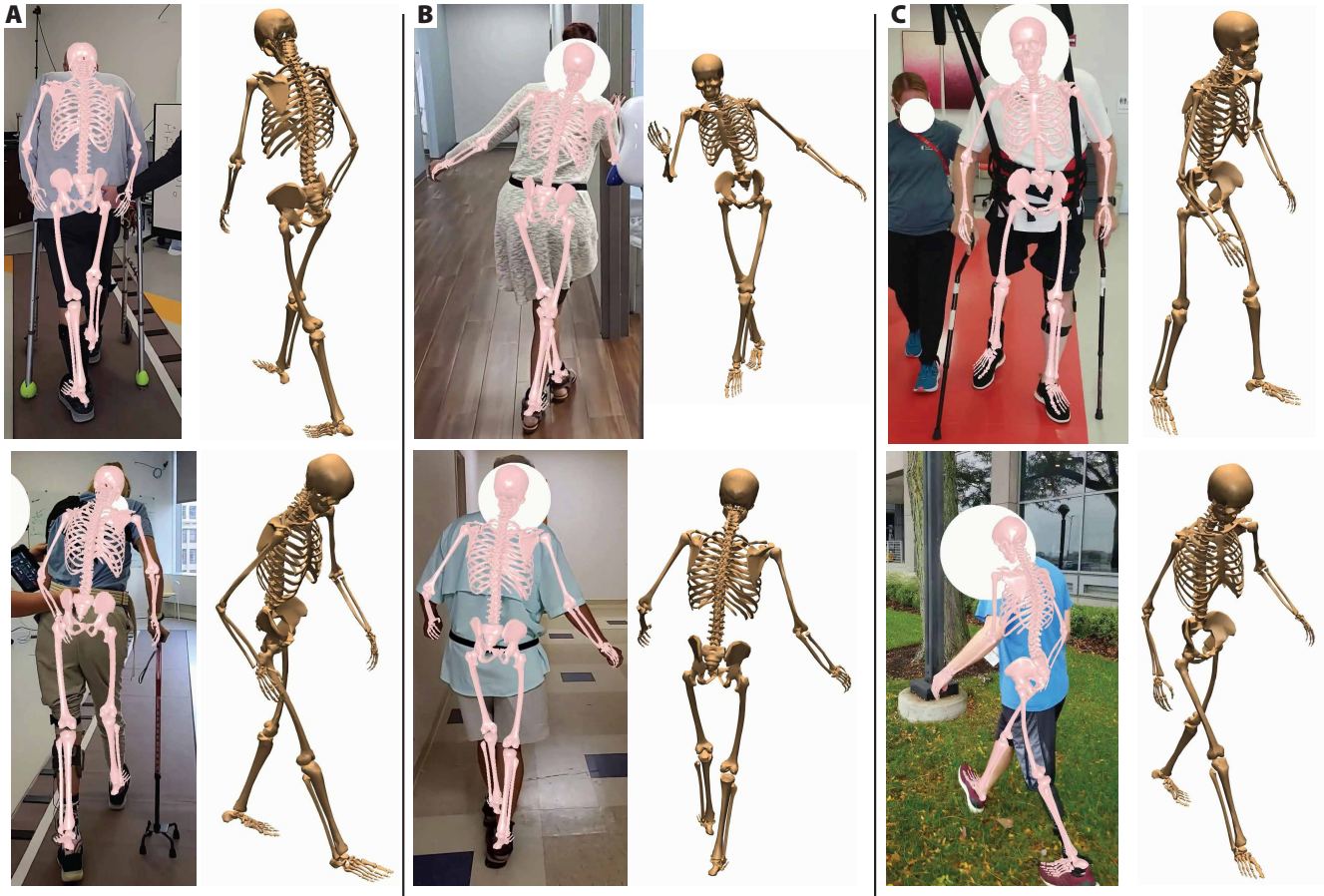


Fig. 3. Biomechanical Reconstructions from Handheld Smartphone. Using only a handheld smartphone, our PBL and end-to-end biomechanical fitting approach robustly captures movement across varied clothing types, assistive devices, and clinical environments. **A)** We first validate this approach on clinical 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.

Clinical Construct Validity of Smartphone-Derived Gait Metrics.

Gait Deviation Index. To reduce kinematic timeseries to an interpretable score representing gait quality, we calculated the Gait Deviation Index (GDI) [50]. Briefly, the GDI uses a dimensionality reduction technique [50, 51] on cycle-aligned joint kinematics and then measures the distance in this subspace to a distribution of normative gait trials (for which we used 9,008 steps of 206 control participants from both monocular and MMMC systems). We computed GDI, measured between 0 and 100 with 100 being normative gait, on every gait cycle for a participant and then averaged GDI over a session. Finally, we also computed double support time (DST) and cadence using a model for gait event detection that we previously developed and validated [37, 36].

Construct Validity. Clinical metrics must be repeatable, valid, and responsive. We assessed the repeatability of kinematics from our system by computing intra-class correlations (ICCs) of the GDI between steps taken by the same individual on the same day [52]. We tested construct validity of our system using t-tests to separate GDI distributions between groups. To further test construct validity, we also compared the GDI to established clinical scores, including the mJOA, 10MWT, and BBS. To evaluate the responsiveness of smartphone-based gait metrics to surgical intervention, we analyzed longitudinal pre- and post-operative changes in mJOA, cadence, DST, and GDI in a cohort of cervical myelopathy participants ($n = 17$) that were recorded before and after surgery. Paired changes were first assessed for statistical significance using the Wilcoxon signed-rank test. The magnitude and direction of change were then quantified using the corresponding Wilcoxon effect size, which provides a standardized measure

Table 1. Summary of errors for different datasets. Median and Mean Absolute Errors are for joint angles aggregated across all joints, participants, and trials. RTE is the median root translation error (in cm). Errors are presented as: (Median (nIQR)).

	MMC		OMC
	Handheld	Static	Static
Median Absolute Error (deg)	2.79 (0.86)	2.96 (0.78)	2.74 (0.50)
Mean Absolute Error (deg)	4.10 (3.71)	4.49 (1.99)	3.89 (1.66)
RTE (cm)	6.44 (4.77)	5.00 (4.68)	2.41 (0.63)

on a $[-1, 1]$ scale [53]. Ninety-five percent confidence intervals for the effect sizes were estimated via bootstrapping. In addition, changes in raw knee kinematics before and after surgery were assessed using paired SPM1D t-tests [54, 55].

Results

Ground Truth Comparison. In general, kinematics from a single camera closely matched MMC and OMC. Example overlays demonstrate our method is robust to common mobility aids (Fig. 3) and kinematic traces (Fig. 4A) reflect clinical impairments like neurologic or prosthetic asymmetries. Quantitative ground truth comparisons are presented in Table 1 and Supplement.

For the MMC dataset, the Median Absolute Error between a handheld moving camera and MMC was 2.79° . This error was slightly higher for static camera fits using a single view from the MMC system (2.96°), likely due to the greater distances to the person. The error was the greatest in the neurological population, with a Median Absolute Error of 3.32° , lower for LLPUs (2.97°), and lowest for controls at 2.51° (Fig. S1, Table S2). Our method also performed well on the BML-MoVi dataset, showing only 2.74° of Median Absolute Error from the monocular fits compared to OMC. Error distributions are shown for lower limb joints in Figure 4B and broken down across all joints in Tables S1, S2 and Figure S1. Notably wide error distributions are Neck Extension, Lumbar Extension, Hip Flexion (static camera), with third quartiles above 5° .

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° (Fig. 4C).

Visual comparisons of root position between single-camera and multi-view fits are shown in Figure S3. 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.

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. Wilcoxon Effect Size presented with 95% confidence intervals. * indicates $p < 0.05$.

		Repeatability		Validity		Responsiveness
		ICC2	ICC2k	r mJOA	r mJOA-LE	Effect Size
Patient-Reported	mJOA	-	-	-	-	0.24 (-0.35 0.76)
	mJOA-LE	-	-	-	-	0.22 (-0.41 0.91)
Smartphone-Based	Cadence	0.58 (0.52 0.64)	0.88 (0.85 0.90)	0.25*	0.39*	0.39 (-0.09 0.77)
	DST	0.41 (0.35 0.48)	0.78 (0.72 0.82)	-0.27*	-0.37*	-0.52 (-0.84 -0.11)*
	GDI	0.84 (0.81 0.86)	0.96 (0.96 0.97)	0.47*	0.48*	0.44 (-0.02 0.76)*

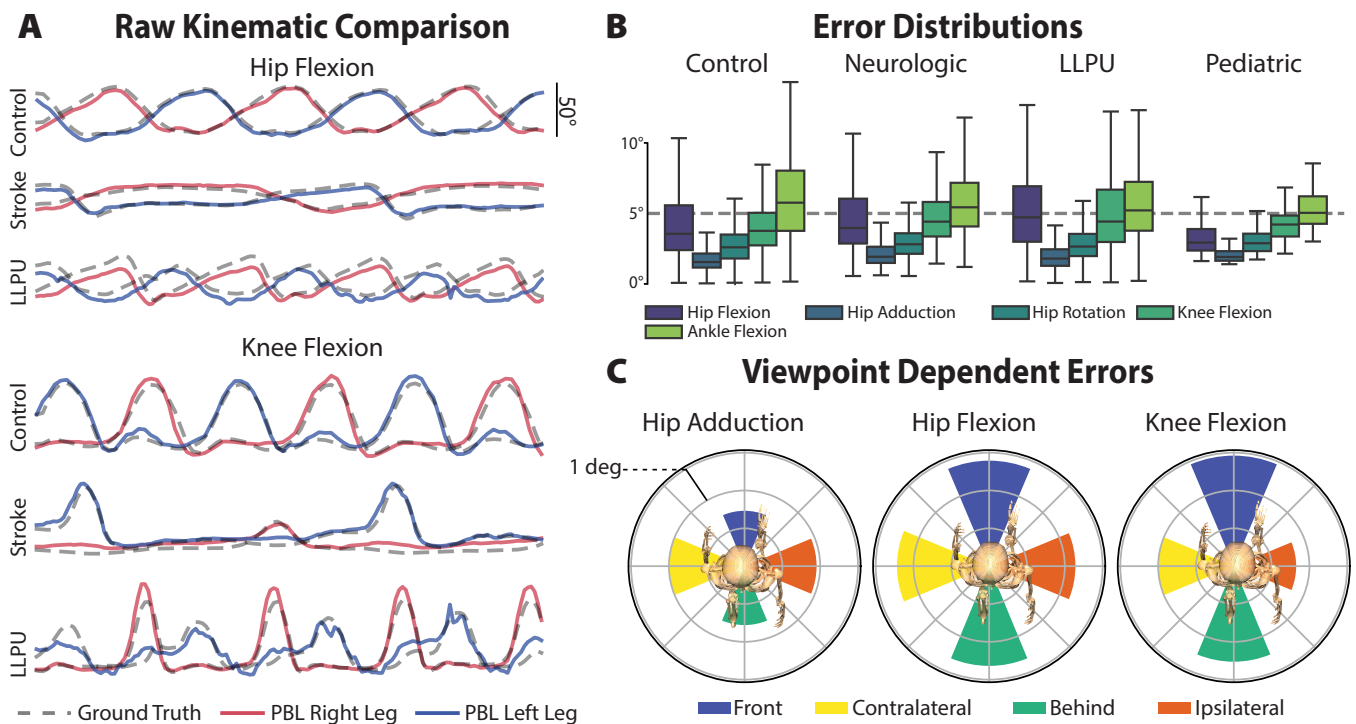


Fig. 4. Quality Measures of Single Camera Fitting. **A)** Kinematic traces from smartphone video (red/blue) compared to ground truth (gray dashed) during walking. **B)** Joint angle errors across populations for select lower limb angles. Control: $n=29$ participants, $v=950$ videos, Neurologic: $n=36$, $v=290$, LLPU: $n=40$, $v=458$, Pediatric: $n=4$, $v=25$. **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.

Clinical Inference from Smartphone-Based Kinematics. Extracting valid, repeatable, and responsive measurements from our smartphone-based system is arguably more important than accuracy alone. We computed the GDI as a measure of gait quality, testing it for repeatability, validity and responsiveness (Figure 5, Table 2). We also test for differences in raw kinematic traces pre/post surgery using the paired SPM1D t-tests [54, 55].

Gait Measurement Clinical Validity. GDI separated groups with expected differences, for instance GDI was significantly lower for participants scoring higher fall risk on the Berg Balance Scale and Transfemoral LLPUs had a significantly lower GDI than Transtibial LLPUs (Fig. 5B,D). The GDI also correlates well with the 10MWT (Pearson's $r = 0.82$), one of the most commonly measured clinical outcomes for gait (Fig. 5C). 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).

Gait Measurement 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 (cadence, DST) were lower-likely due to timing variability—but remained highly repeatable when averaged (ICC2k).

Responsiveness of Gait Measurement to Clinical Intervention. Smartphone-derived gait metrics (GDI and DST) demonstrated significant postoperative changes in individuals with cervical myelopathy (Wilcoxon signed-rank test, $p<0.05$), whereas patient-reported clinical scales (mJOA and mJOA-LE) did not show significant change (Table 2). Wilcoxon effect sizes were 0.24 for mJOA, 0.22 for mJOA-LE, -0.52 for DST, and 0.44 for GDI. In addition, paired SPM1D t-tests revealed a significant postoperative increase in knee flexion during initial swing (64–68 % of the gait cycle).

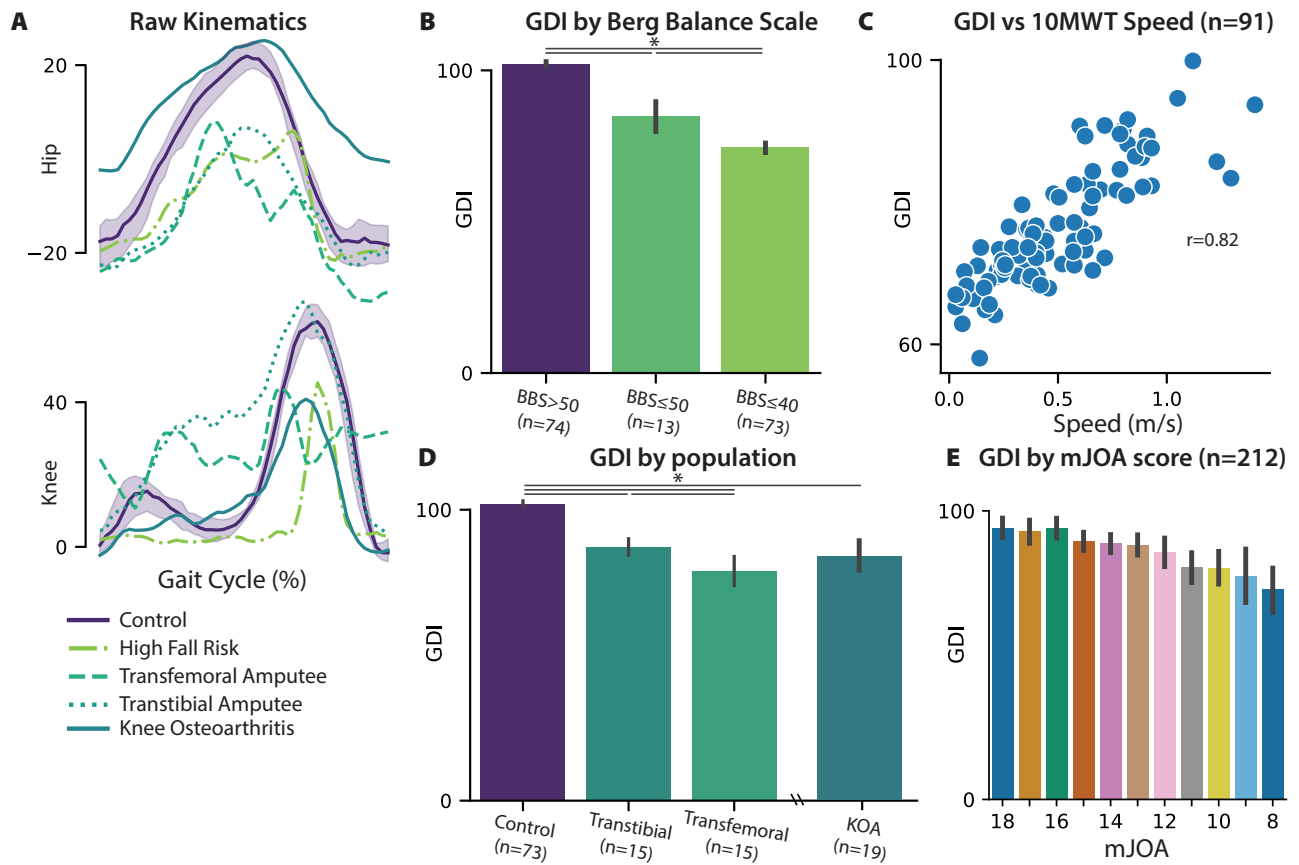


Fig. 5. Clinical Validity of Smartphone-Based Gait Deviation Index. **A)** Hip and Knee flexion angles of clinical and control groups **B)** GDI separates groups at risk of falls determined by the Berg Balance Scale. **C)** GDI correlates with 10 Meter Walk Test performance $r = 0.82$. **D)** GDI of 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.

Discussion

We present a clinically validated method for extracting biomechanically accurate kinematics from smartphone video in clinical settings. Across diverse populations, our single-camera system showed strong agreement with ground truth (typically $< 3^\circ$ Median Absolute Error). Gait quality metrics derived from these videos detected clinically relevant group differences and were responsive to clinical intervention. 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 track their recovery.

Joint angle errors were generally low across all populations and settings (Table 1). We attribute the slightly higher Median Absolute Error in clinical populations (0.46 - 0.81 degrees, Fig. 4B, Fig. S1B, Table S2) to assistive devices such as canes, walkers, ankle foot orthoses, or gait belts (Fig. 3) and the fact that keypoint detection does not perform as well on prosthetic and orthotic devices [36]. Studies using OpenCap also report increases in joint angle errors in clinical populations [22]. Intuitively, sagittal-plane angles were most accurate when recorded from sagittal viewpoints, and frontal-plane angles were most accurate from frontal viewpoints (Fig. 4C), giving clinicians a simple and actionable strategy for mitigating joint-angle error during data collection.

Our only dataset comparing PBL to Optical Motion Capture (OMC) is BML-MoVi, which only contains able-bodied controls. To further validate the biomechanical generalization of PBL, we also included a dataset with diverse etiologies of gait impairments assessed with our previously validated MMMC system [45, 25]. Although MMMC differs from OMC, multiple independent studies have demonstrated strong agreement between the two approaches [14, 16, 17, 18, 19, 20, 21, 27]. Nonetheless, neither method constitutes a true biomechanical gold standard relative to biplanar fluoroscopy, due to known limitations including marker placement variability [56] and soft-tissue artifact [57]. The consistently strong correspondence observed between PBL and both OMC and MMMC across a diverse

participant population supports the biomechanical accuracy and generalizability of the proposed approach (Table 1). While monocular video-based analysis is expected to exhibit some loss of accuracy relative to multiview systems, it substantially improves accessibility in clinical environments. Thus, the latter part of this study evaluates whether PBL collected in clinical settings is sufficiently accurate to demonstrate construct validity against established clinical metrics.

To demonstrate this, we showed that movement quality scores from PBL: (1) change with expected population changes, (2) correlate with clinical scores, and (3) are responsive to clinical intervention. First, GDI from PBL was decreased for inpatients with fall risk assessed by the BBS and further decreased for inpatients classified as high fall risk (Fig. 5B). Similarly, GDI was decreased for Transtibial LLPU and further decreased for Transfemoral LLPU, an intuitive result. Second, GDI correlated highly with the 10MWT (Fig. 5C), a widely used clinical outcome measure. The GDI also had a moderate correlation with the mJOA (Fig 5E, Table 2), which is expected considering the GDI is a continuous quantitative measure and the mJOA is a patient-reported ordinal scale. Regarding responsiveness, SPM1D analysis of kinematic traces showed significant changes in knee flexion following surgical intervention, establishing that our system can be used to analyze specific changes in movement and quantitatively track recovery. Quantitative metrics (DST and GDI) extracted from our smartphone-based system showed significant postoperative changes, whereas the mJOA did not (Table 2). This may suggest that smartphone-derived kinematic measures capture postoperative kinematics changes not reflected in patient-reported scores, but it is important to highlight these are measuring fundamentally different constructs. Thus, quantitative gait metrics (e.g., GDI, DST) complement patient-reported outcomes such as the mJOA, and we defer to future studies to investigate whether they may precede and causally influence patient outcomes [58].

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 [59].

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 high, 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 observe heavy-tailed error distributions, with some trials exceeding 10° of error (Fig. S1C). Although occasional large errors and failure cases warrant further investigation, these errors occur infrequently and do not introduce systematic bias. Aggregated kinematic trajectories and summary metrics remain strongly correlated with clinically relevant measures (Table 2, Fig. 5), supporting the method's reliability for population-level analysis and leaving subject-level clinical decision-making to future work. We tested whether fitting residuals—2D reprojection error or 3D keypoint error—could serve as reliability metrics. Their correlations with joint-angle error were weak (0.35 and 0.24, respectively; Fig. S5), underscoring the need for better confidence estimates in markerless motion capture. We have previously developed calibrated confidence intervals for monocular pose estimation without biomechanics [60] and for multi-view biomechanical estimation [24]; extending these ideas to monocular biomechanics is an important direction for future work.

Other approaches have proposed deployable systems for human motion capture. Most well-known is OpenCap [29], which uses two or three smartphones on static tripods to triangulate joint locations from video and fit biomechanics using the widely used OpenSim framework [61, 62]. OpenCap reports a Mean Absolute Error of 4.1° during walking in 10 controls using two static cameras. In our control cohort (n=29), PBL achieves a comparable Mean Absolute Error of 4.32 (2.06)° using a single handheld moving camera. While this error was computed on similar populations and activities, we note this is not a direct comparison and leave that to future work. While more cameras will generally increase accuracy [24], our results suggest that the difference between OpenCap with 2 camera on tripods and a single handheld smartphone is now small. Apart from accuracy, OpenCap and our PBL have different advantages: OpenCap provides a very user friendly platform, uses the widely accepted OpenSim package [61], and calculates joint kinetics. Our PBL approach presented here is markedly easier to deploy in clinical settings such as narrow hallways, small exam rooms, and busy gym spaces. The same simplicity also enables potential use in patients' homes with recordings performed by caregivers. Extending our method to include joint kinetics is feasible

with future work such as [59]. While our pipeline is not real-time, per-trial runtimes are compatible with same-day or visit-level clinical analysis.

Two recent approaches estimate biomechanics from monocular video: BioPose [34] and Human Skeleton and Mesh Recovery (HSMR) [35]. Both regress pose directly from images, making them faster than our pipeline. However, neither was trained or evaluated on clinical populations, so their performance on clinical gait—often far outside the distribution of healthy movement (Fig. 5A)—is uncertain. Our method instead uses trajectory optimization with in-loop biomechanical model scaling applied consistently across trials, aligning more closely with established biomechanical reconstruction practices. It can also integrate wearable sensors [38] to address occlusion and high-speed motion. For healthy controls, our kinematic accuracy is similar to BioPose ($<3.9^\circ$ Mean Absolute Error on MoVi dataset), though the comparison is incomplete because BioPose does not specify which joints contribute to its reported errors. HSMR does not report joint-angle errors. Ultimately, broader development of AI-driven biomechanics tools will expand clinical movement analysis, and large resulting datasets will support data-driven, precision rehabilitation strategies [58].

In conclusion, we present a method for extracting accurate and biomechanically 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, and respond to clinical intervention. 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.

References

1. Jacquelin Perry, M. Gait analysis: normal and pathological function. *New Jersey: SLACK* (2010).
2. Balaban, B. & Tok, F. Gait Disturbances in Patients With Stroke. *PM&R* **6**, 635–642 (July 2014).
3. Morris, M. E., Huxham, F., McGinley, J., Dodd, K. & Iansek, R. The biomechanics and motor control of gait in Parkinson disease. *Clinical Biomechanics* **16**, 459–470 (July 1, 2001).
4. Hamacher, D., Singh, N., Van Dieën, J., Heller, M. & Taylor, W. Kinematic measures for assessing gait stability in elderly individuals: a systematic review. *Journal of The Royal Society Interface* **8**. Publisher: Royal Society, 1682–1698 (Aug. 31, 2011).
5. Heiden, T. L., Lloyd, D. G. & Ackland, T. R. Knee joint kinematics, kinetics and muscle co-contraction in knee osteoarthritis patient gait. *Clinical Biomechanics* **24**, 833–841 (Dec. 1, 2009).
6. Sosdian, L. *et al.* Longitudinal changes in knee kinematics and moments following knee arthroplasty: A systematic review. *The Knee* **21**, 994–1008 (Dec. 1, 2014).
7. Chow, J. W. & Stokic, D. S. Longitudinal Changes in Temporospatial Gait Characteristics during the First Year Post-Stroke. *Brain Sciences* **11**. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, 1648 (Dec. 2021).
8. Awad, L. N., Lewek, M. D., Kesar, T. M., Franz, J. R. & Bowden, M. G. These legs were made for propulsion: advancing the diagnosis and treatment of post-stroke propulsion deficits. *J NeuroEngineering Rehabil* **17**, 139 (Dec. 2020).
9. Kwakkel, G. *et al.* Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *International Journal of Stroke* **12**, 451–461 (July 2017).
10. Abdullah, M., Hulleck, A. A., Katmah, R., Khalaf, K. & El-Rich, M. Multibody dynamics-based musculoskeletal modeling for gait analysis: a systematic review. *Journal of NeuroEngineering and Rehabilitation* **21**, 178 (Oct. 5, 2024).
11. Podsiadlo, D. & Richardson, S. The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons. *Journal of the American Geriatrics Society* **39**. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.1991.tb01616.x>, 142–148 (1991).

12. Berg, K. O., Maki, B. E., Williams, J. I., Holliday, P. J. & Wood-Dauphinee, S. L. Clinical and laboratory measures of postural balance in an elderly population. *Archives of Physical Medicine and Rehabilitation* **73**. Publisher: Elsevier, 1073–1080 (Nov. 1, 1992).
13. Wrisley, D. M., Marchetti, G. F., Kuharsky, D. K. & Whitney, S. L. Reliability, Internal Consistency, and Validity of Data Obtained With the Functional Gait Assessment. *Physical Therapy* **84**, 906–918 (Oct. 1, 2004).
14. Kanko, R. M., Laende, E. K., Davis, E. M., Selbie, W. S. & Deluzio, K. J. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of Biomechanics* **127**, 110665 (Oct. 11, 2021).
15. Kanko, R. M., Laende, E., Selbie, W. S. & Deluzio, K. J. Inter-Session Repeatability of Markerless Motion Capture Gait Kinematics. *Journal of Biomechanics* **121**, 110422 (May 24, 2021).
16. Kanko, R. M. *et al.* Assessment of Spatiotemporal Gait Parameters Using a Deep Learning Algorithm-Based Markerless Motion Capture System. *Journal of Biomechanics* **122**, 110414 (June 9, 2021).
17. Riazati, S., McGuirk, T. E., Perry, E. S., Sihanath, W. B. & Patten, C. Absolute Reliability of Gait Parameters Acquired With Markerless Motion Capture in Living Domains. *Frontiers in Human Neuroscience* **16**, 867474 (June 16, 2022).
18. Song, K., Hullfish, T. J., Scattone Silva, R., Silbernagel, K. G. & Baxter, J. R. Markerless motion capture estimates of lower extremity kinematics and kinetics are comparable to marker-based across 8 movements. *Journal of Biomechanics* **157**, 111751 (Aug. 1, 2023).
19. Wren, T. A. L., Isakov, P. & Rethlefsen, S. A. Comparison of kinematics between Theia markerless and conventional marker-based gait analysis in clinical patients. *Gait & Posture* **104**, 9–14 (July 1, 2023).
20. Horsak, B., Kainz, H. & Dumphart, B. Repeatability and Minimal Detectable Change Including Clothing Effects for Smartphone-Based 3D Markerless Motion Capture. *Journal of Biomechanics* **175**, 112281 (Oct. 1, 2024).
21. D’Souza, S., Siebert, T. & Fohanno, V. A comparison of lower body gait kinematics and kinetics between Theia3D markerless and marker-based models in healthy subjects and clinical patients. *Sci Rep* **14**. Publisher: Nature Publishing Group, 29154 (Nov. 25, 2024).
22. Wang, J. *et al.* Evaluation of a smartphone-based markerless system to measure lower-limb kinematics in patients with knee osteoarthritis. *Journal of Biomechanics* **181**, 112529 (Mar. 1, 2025).
23. Min, Y.-S. *et al.* Biomechanical Gait Analysis Using a Smartphone-Based Motion Capture System (OpenCap) in Patients with Neurological Disorders. *Bioengineering* **11**. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, 911 (Sept. 2024).
24. Cotton, R. J. & Sinz, F. Biomechanical Reconstruction with Confidence Intervals from Multiview Markerless Motion Capture. *Annu Int Conf IEEE Eng Med Biol Soc* **2025**, 1–7 (July 2025).
25. Cotton, R. J. *Differentiable Biomechanics Unlocks Opportunities for Markerless Motion Capture* in 2025 *International Conference On Rehabilitation Robotics (ICORR)* 2025 International Conference On Rehabilitation Robotics (ICORR). ISSN: 1945-7901 (May 2025), 44–51.
26. Firouzabadi, P. *et al.* *Biomechanical Arm and Hand Tracking with Multiview Markerless Motion Capture* in 2024 *10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)* 2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob). ISSN: 2155-1782 (Sept. 2024), 1641–1648.
27. Unger, T. *et al.* Differentiable Biomechanics for Markerless Motion Capture in Upper Limb Stroke Rehabilitation: A Comparison With Optical Motion Capture. *IEEE Transactions on Medical Robotics and Bionics*, 1–1 (2025).
28. Todorov, E., Erez, T. & Tassa, Y. *MuJoCo: A physics engine for model-based control* in 2012 *IEEE/RSJ International Conference on Intelligent Robots and Systems* 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. ISSN: 2153-0866 (Oct. 2012), 5026–5033.
29. Uhlich, S. D. *et al.* OpenCap: Human movement dynamics from smartphone videos. *PLOS Computational Biology* **19**. Publisher: Public Library of Science, e1011462 (Oct. 19, 2023).
30. Ruth, P. S. *et al.* Video-Based Biomechanical Analysis Captures Disease-Specific Movement Signatures of Different Neuromuscular Diseases. *NEJM AI* **2**. Publisher: Massachusetts Medical Society, AIoa2401137 (Aug. 28, 2025).

31. Acevedo, G. *et al.* VisionMD: an open-source tool for video-based analysis of motor function in movement disorders. *npj Parkinsons Dis.* **11**. Publisher: Nature Publishing Group, 27 (Feb. 4, 2025).
32. Segado, M., Prosser, L., Duncan, A. F., Johnson, M. J. & Kording, K. P. *Data-Driven Early Prediction of Cerebral Palsy Using AutoML and interpretable kinematic features* Feb. 12, 2025.
33. Lonini, L. *et al.* Video-Based Pose Estimation for Gait Analysis in Stroke Survivors during Clinical Assessments: A Proof-of-Concept Study. *Digit Biomark* **6**, 9–18 (Jan. 13, 2022).
34. Koleini, F. *et al.* BioPose: Biomechanically-Accurate 3D Pose Estimation from Monocular Videos in 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). ISSN: 2642-9381 (Feb. 2025), 6330–6339.
35. Xia, Y., Zhou, X., Vouga, E., Huang, Q. & Pavlakos, G. *Reconstructing Humans with a Biomechanically Accurate Skeleton in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2025), 5355–5365.
36. Cimorelli, A., Patel, A., Karakostas, T. & Cotton, R. J. Validation of portable in-clinic video-based gait analysis for prosthesis users. *Sci Rep* **14**. Publisher: Nature Publishing Group, 3840 (Feb. 15, 2024).
37. Cotton, R. J., McClerklin, E., Cimorelli, A., Patel, A. & Karakostas, T. *Transforming Gait: Video-Based Spatiotemporal Gait Analysis in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, Glasgow, Scotland, United Kingdom, July 11, 2022), 115–120.
38. Peiffer, J., Shah, K., Anarwala, S., Abdou, K. & Cotton, R. J. *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)* ISSN: 2155-1782 (Sept. 2024), 1275–1282.
39. Cotton, R. J. PosePipe: Open-Source Human Pose Estimation Pipeline for Rehabilitation Research. *Archives of Physical Medicine and Rehabilitation* **103**, e161–e162 (Dec. 1, 2022).
40. Wojke, N., Bewley, A. & Paulus, D. *Simple online and realtime tracking with a deep association metric in 2017 IEEE International Conference on Image Processing (ICIP)* 2017 IEEE International Conference on Image Processing (ICIP) (IEEE, Beijing, Sept. 2017), 3645–3649.
41. Sáráandi, I., Hermans, A. & Leibe, B. *Learning 3D Human Pose Estimation From Dozens of Datasets Using a Geometry-Aware Autoencoder To Bridge Between Skeleton Formats in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2023), 2956–2966.
42. Ghorbani, S. *et al.* MoVi: A large multi-purpose human motion and video dataset. *PLOS ONE* **16**. Publisher: Public Library of Science, e0253157 (June 17, 2021).
43. Cotton, R. J. *et al.* Optimizing Trajectories and Inverse Kinematics for Biomechanical Analysis of Markerless Motion Capture Data in 2023 International Conference on Rehabilitation Robotics (ICORR) 2023 International Conference on Rehabilitation Robotics (ICORR) (IEEE, Singapore, Singapore, Sept. 24, 2023), 1–6.
44. Caggiano, V., Wang, H., Durandau, G., Sartori, M. & Kumar, V. MyoSuite: A contact-rich simulation suite for musculoskeletal motor control. *arXiv preprint arXiv:2205.13600* (2022).
45. Cotton, R. J. *et al.* Markerless Motion Capture and Biomechanical Analysis Pipeline. *arXiv preprint arXiv:2303.10654* (2023).
46. Lusardi, M. M. *et al.* Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. *Journal of Geriatric Physical Therapy* **40**, 1–36 (Jan. 2017).
47. Shumway-Cook, A., Baldwin, M., Polissar, N. L. & Gruber, W. Predicting the Probability for Falls in Community-Dwelling Older Adults. *Physical Therapy* **77**, 812–819 (Aug. 1, 1997).
48. Kopjar, B., Fehlings, M. & Hanson, B. Validity of the Modified Japanese Orthopedic Association Score in Patients with Cervical Spondylotic Myelopathy: The AOSpine North America Multicenter Prospective Study. *The Spine Journal* **11**, S73–S74 (Oct. 1, 2011).
49. Tetreault, L. *et al.* The Modified Japanese Orthopaedic Association Scale: Establishing Criteria for Mild, Moderate and Severe Impairment in Patients with Degenerative Cervical Myelopathy. *European Spine Journal: Official Publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* **26**, 78–84 (Jan. 2017).
50. Schwartz, M. H. & Rozumalski, A. The gait deviation index: A new comprehensive index of gait pathology. *Gait & Posture* **28**, 351–357 (Oct. 1, 2008).

51. Marks, M., Kingsbury, T., Bryant, R., Collins, J. D. & Wyatt, M. Measuring Abnormality in High Dimensional Spaces with Applications in Biomechanical Gait Analysis. *Scientific Reports* **8**. Publisher: Nature Publishing Group, 15481 (Oct. 19, 2018).
52. Bartko, J. J. The Intraclass Correlation Coefficient as a Measure of Reliability. *Psychol Rep* **19**. Publisher: SAGE Publications Inc, 3–11 (Aug. 1, 1966).
53. Fritz, C. O., Morris, P. E. & Richler, J. J. Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General* **141**, 2–18 (2012).
54. Pataky, T. C. Generalized n -dimensional biomechanical field analysis using statistical parametric mapping. *Journal of Biomechanics* **43**, 1976–1982 (July 20, 2010).
55. Pataky, T. C. One-dimensional statistical parametric mapping in Python. *Computer Methods in Biomechanics and Biomedical Engineering* **15**. Publisher: Taylor & Francis_eprint: <https://doi.org/10.1080/10255842.2010.527837>, 295–301 (Mar. 1, 2012).
56. Uchida, T. K. & Seth, A. Conclusion or Illusion: Quantifying Uncertainty in Inverse Analyses From Marker-Based Motion Capture due to Errors in Marker Registration and Model Scaling. *Front. Bioeng. Biotechnol.* **10**. Publisher: Frontiers (May 25, 2022).
57. Akbarshahi, M. *et al.* Non-invasive assessment of soft-tissue artifact and its effect on knee joint kinematics during functional activity. *Journal of Biomechanics* **43**, 1292–1301 (May 2010).
58. Cotton, R. J. *et al.* A Causal Framework for Precision Rehabilitation. *arXiv preprint arXiv:2411.03919* (2024).
59. Cotton, R. J. KinTwin: Imitation Learning with Torque and Muscle Driven Biomechanical Models Enables Precise Replication of Able-Bodied and Impaired Movement from Markerless Motion Capture. *arXiv preprint arXiv:2505.13436* (2025).
60. Pierzchlewicz, P. A., Silva, C. O. d., Cotton, R. J. & Sinz, F. H. Platypose: Calibrated Zero-Shot Multi-Hypothesis 3D Human Motion Estimation. *arXiv preprint arXiv:2403.06164* (2024).
61. Delp, S. L. *et al.* OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement. *IEEE Transactions on Biomedical Engineering* **54**. Conference Name: IEEE Transactions on Biomedical Engineering, 1940–1950 (Nov. 2007).
62. Werling, K. *et al.* AddBiomechanics: Automating model scaling, inverse kinematics, and inverse dynamics from human motion data through sequential optimization. *PLOS ONE* **18**. Publisher: Public Library of Science, e0295152 (Nov. 30, 2023).
63. Cotton, R. J. & Rogers, J. Wearable Monitoring of Joint Angle and Muscle Activity. *IEEE Int Conf Rehabil Robot* **2019**, 258–263 (June 2019).
64. Yatsenko, D. *et al.* DataJoint: managing big scientific data using MATLAB or Python. *bioRxiv* (2015).
65. Yonenobu, K., Abumi, K., Nagata, K., Taketomi, E. & Ueyama, K. Interobserver and Intraobserver Reliability of the Japanese Orthopaedic Association Scoring System for Evaluation of Cervical Compression Myelopathy. *Spine* **26** (2001).
66. Kato, S. *et al.* Comparison of the Japanese Orthopaedic Association (JOA) Score and Modified JOA (mJOA) Score for the Assessment of Cervical Myelopathy: A Multicenter Observational Study. *PLoS One* **10**, e0123022 (Apr. 2, 2015).
67. Al-Hafez, F., Zhao, G., Peters, J. & Tateo, D. *LocoMuJoCo: A Comprehensive Imitation Learning Benchmark for Locomotion* Nov. 30, 2023.
68. Hamner, S. R., Seth, A. & Delp, S. L. Muscle contributions to propulsion and support during running. *Journal of Biomechanics* **43**, 2709–2716 (Oct. 19, 2010).
69. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. & Black, M. J. SMPL: a skinned multi-person linear model. *ACM Trans. Graph.* **34**, 1–16 (Nov. 4, 2015).
70. Vaswani, A. *et al.* Attention is All you Need in *Advances in Neural Information Processing Systems* (eds Guyon, I. *et al.*) **30** (Curran Associates, Inc., 2017).
71. Lynch, K. M. & Park, F. C. *Modern robotics: mechanics, planning, and control* OCLC: ocn983881868. 528 pp. (Cambridge University Press, Cambridge, UK, 2017).
72. Kidger, P. & Garcia, C. *Equinox: neural networks in JAX via callable PyTrees and filtered transformations* Oct. 30, 2021.
73. DeepMind *et al.* *The DeepMind JAX Ecosystem* 2020.

74. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization* in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (eds Bengio, Y. & LeCun, Y.) (2015).
75. SHAPIRO, S. S. & WILK, M. B. An analysis of variance test for normality (complete samples)[†]. *Biometrika* **52**, 591–611 (Dec. 1, 1965).
76. Husted, J. A., Cook, R. J., Farewell, V. T. & Gladman, D. D. Methods for assessing responsiveness: a critical review and recommendations. *Journal of Clinical Epidemiology* **53**, 459–468 (May 2000).
77. Liang, M. H., Fossel, A. H. & Larson, M. G. Comparisons of Five Health Status Instruments for Orthopedic Evaluation. *Medical Care* **28**, 632 (July 1990).
78. Werner, C. *et al.* Concurrent Validity, Test-Retest Reliability, and Sensitivity to Change of a Single Body-Fixed Sensor for Gait Analysis during Rollator-Assisted Walking in Acute Geriatric Patients. *Sensors* **20**. Publisher: Multidisciplinary Digital Publishing Institute, 4866 (Jan. 2020).
79. MILLER, R. G. The jackknife-a review. *Biometrika* **61**, 1–15 (Apr. 1, 1974).

Supplementary Methods

Datasets.

OMC Dataset: BML-MoVi. The publicly available BML-MoVi [42] 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 [25]. For the smartphone data, we apply the 3D and 2D keypoint loss following the method we use on the PBL data and describe below.

MMMC Dataset. This study was approved by the Northwestern University Institutional Review Board. Participants provided informed consent for their videos to be stored, processed, and de-identified videos to be shared in publication. Data were collected in our laboratory using handheld smartphone video [36], an instrumented GaitRite walkway, and an MMC system [45]. 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 MMC. We also analyzed single-view videos from MMC 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 × 1920 resolution [36] 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 [63]. 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 [39], 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 [64]. This makes it feasible to manage the 100s of thousands of videos we have collected.

Multi-View Markerless Motion Capture (MMC) Recordings in the laboratory dataset employed the Multi-View Markerless Motion Capture (MMC) 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 [45]. For this system, we reconstructed biomechanical fits using methods described in [25]. 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 MMC systems, and an additional 210,725 frames from 178 trials of 27 participants recorded with a single MMC camera—amounting to 11.7 hours of video.

In-Clinic Dataset. This study was approved by the Northwestern University and Washington University in St. Louis Institutional Review Boards. Participants provided informed consent for their videos to be stored, processed, and de-identified videos to be shared in publication. To validate our methods in a real-world clinical environment, we deployed our PBL system [36] 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 [65, 66] 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 [41] 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. For a 6 second video with one person in frame, MeTRAbs-ACAE took 10 seconds to run on a NVIDIA A100 GPU and 1 minute to run on a NVIDIA GeForce RTX 3050 Laptop GPU. This processing time will scale linearly with video time.

Differentiable Biomechanical Model. Smartphone and MMMC reconstructions utilized a biomechanically-grounded model implemented in Mujoco from LocoMujoco [67], originally based on an OpenSim model [68]. 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 [43]. 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 right leg. We follow notation used by the skinned multi-person linear model (SMPL) [69] and represent our GPU-accelerated, biomechanically-ground, forward kinematic equation as:

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

where $\theta \in \mathbb{R}^{40}$ are joint angles that pose the model, $\beta \in \mathbb{R}^{8+87 \times 3}$ are the 8 scaling parameters and an offset for each of the 87 keypoints. $\mathbf{x} \in \mathbb{R}^{87 \times 3}$ 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 [70] 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 $\mathbf{r}$.

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

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 [71] 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 $\{\phi_0, \phi_1, \dots, \phi_{N-1}\}$, 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 [62]. 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 . A single trial required 1 minute to optimize on a NVIDIA A100 GPU and 4.5 minutes to optimize on a consumer-grade (NVIDIA GeForce RTX 3050 Laptop) GPU. Trial length has little effect on runtime, and additional trials add only an incremental processing cost.

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, $\hat{x} = \mathcal{M}(\hat{\theta}(t), \hat{\beta})$. 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 [41]. 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:

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

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:

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

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:

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

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:

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

We defined implicit functions and computed the loss using Equinox [72] and JAX [73], 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 $1e-6$ [74, 73].

For MMMC [25], 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.

Validation Metrics. We evaluated normality of errors at two levels using the Shapiro–Wilk test [75]: (1) the distribution of joint angle errors within each trial and (2) the distribution of trial-level error summaries across trials. Both levels were non-Gaussian. Consequently, we report the Median Absolute Error, computed by first taking the median joint-angle error within each trial and then taking the median of these per-trial medians across trials. The Median Absolute Error includes the following degrees of freedom: hips (2×3), knees (2×1), ankles (2×1), lumbar spine (3), neck (3), shoulders (2×2), and elbows (2×1).

When aggregating results across trials, we also report the normalized interquartile range, $nIQR = IQR \cdot 0.7413$.

For comparison to prior work, including OpenCap, we additionally report the Mean Absolute Error following the procedure in [29]. This Mean Absolute Error calculation includes the following joints: pelvis rotations (3), hips (2×3), knees (2×1), ankles (2×2), and lumbar (3).

To elucidate where in the gait cycle errors occur, we extracted 1,971 gait cycles from our MMMC dataset with paired MMMC (multi-camera) and PBL (single-camera) recordings and performed a paired SPM1D t-test [54, 55] between hip and knee flexion angles.

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 [37], 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 [50] used singular value decomposition to reduce this high dimensional data, more recent findings [51] 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 [75]. 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.

In addition to the Wilcoxon effect size calculated to assess responsiveness to surgical intervention, we computed the Standardized Response Mean (SRM) [76, 77, 78], defined as:

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

Supplementary Table 1. Median Absolute Error (median (nIQR)) for clinical and BML-MoVi datasets.

Dataset Camera	Clinical		BML-MoVi
	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 estimate confidence intervals for the SRM, we employed the jackknife technique [79] as was done in [77]. 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.

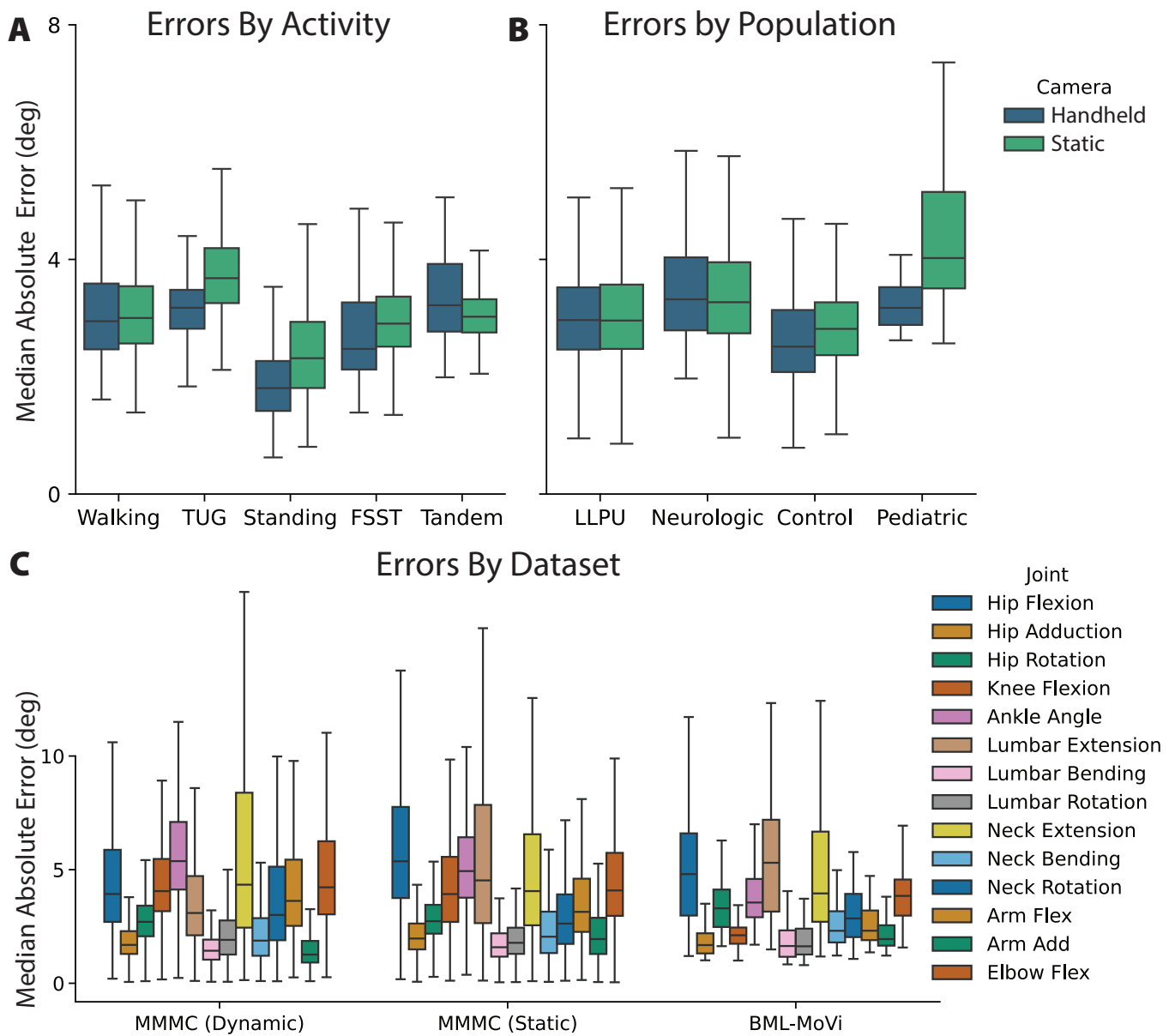
PBL vs MMMC GDI. To assess the impact of recording system (MMMC vs. PBL) on GDI, we conducted a separate analysis using only sessions recorded simultaneously with both systems. For each session, we computed GDI twice: once using PBL data with a PBL-based normative distribution, and once using MMMC data 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 Absolute Error of joint angles 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 Median Absolute Error (2.74° ; Table S1, Fig. 1C), which is expected since it included only control participants. The MMMC dataset showed similar errors when



Supplementary Fig. 1. Joint Angle Error Distributions. **A)** In the MMMC dataset, Median Absolute Error distributions remain relatively stable when aggregated across activity types, participant populations **(B)**, as well as between handheld (moving) and static cameras. **C)** Joint-specific Median Absolute Error distributions from both the MMMC and BML-MoVi datasets.

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. 3). 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 Median Absolute Error (2.51°), while participants with neurological conditions showed the highest (3.32° ; Table S2). This is consistent with expectations: many 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 Median Absolute Error (2.97°), reflecting known limitations of keypoint detectors when applied to prosthetic limbs [36].

Across different activities, Median Absolute Error remained relatively stable (Fig. S1A). Unsurprisingly, standing had the lowest Median Absolute Error (1.80°), followed by the Four Square Step Test (FSST; 2.47°), both of

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

Joint	Median Absolute Error (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)

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 hand-held filming angle used during much of these trials.

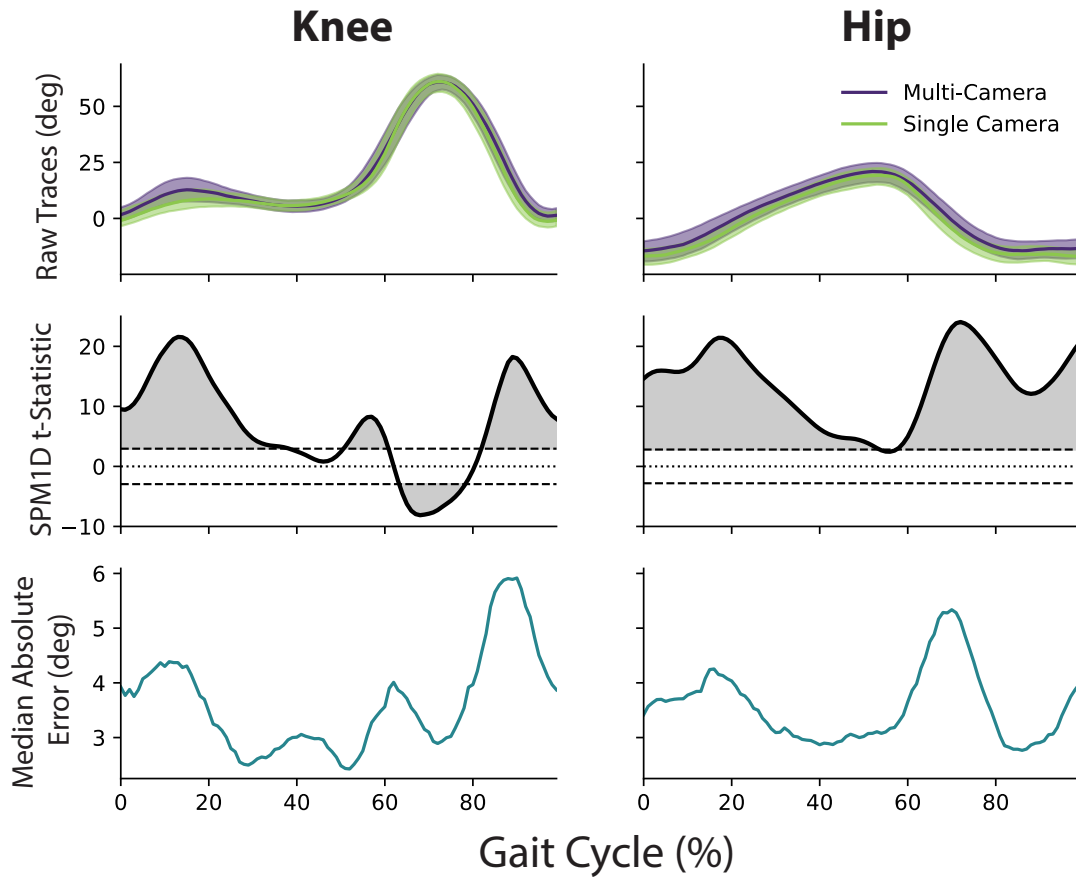
Joint-level trends in Median Absolute Error were largely consistent across populations (Table S2). Notably, ankle Median Absolute Error exceeded 5° in all populations, indicating a clear need for improved foot keypoint detection. Elbow flexion error approached 5° in clinical cohorts, likely because one 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. 4C)—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.

Supplementary Table 3. Lower Body Mean Absolute Errors. Computed over trials (n = 1702) and participants (n = 107), and the reported mean is an average over activities and degrees of freedom (three for pelvis orientation, three for the lumbar, three per hip, one per knee, and two per ankle.)

(a) Across Activity						
Activity	FSST	Standing	TUG	Tandem	Walking	All
Mean Absolute Error (deg)	4.14 (2.13)	2.19 (1.41)	5.17 (1.47)	4.91 (2.67)	4.46 (2.07)	4.18

(b) Across Population					
Population	Control	LLPU	Neurologic	Pediatric	All
Mean Absolute Error (deg)	3.90 (1.82)	4.29 (2.09)	4.49 (1.94)	4.11 (1.81)	4.20

Mean Absolute Error. Mean Absolute Errors calculated with the same joints as used in OpenCap [29] were below 5 degrees for most activities and populations (Table S3). TUG had the highest Mean Absolute Error at 5.17°, and the neurologic population had the highest Mean Absolute Error at 4.49°. For the most faithful comparison to OpenCap, we computed Mean Absolute Error on the control population during walking, finding 4.32 (2.06)° of error, which approaches OpenCap’s reported error of 4.1° on the same population and activity.



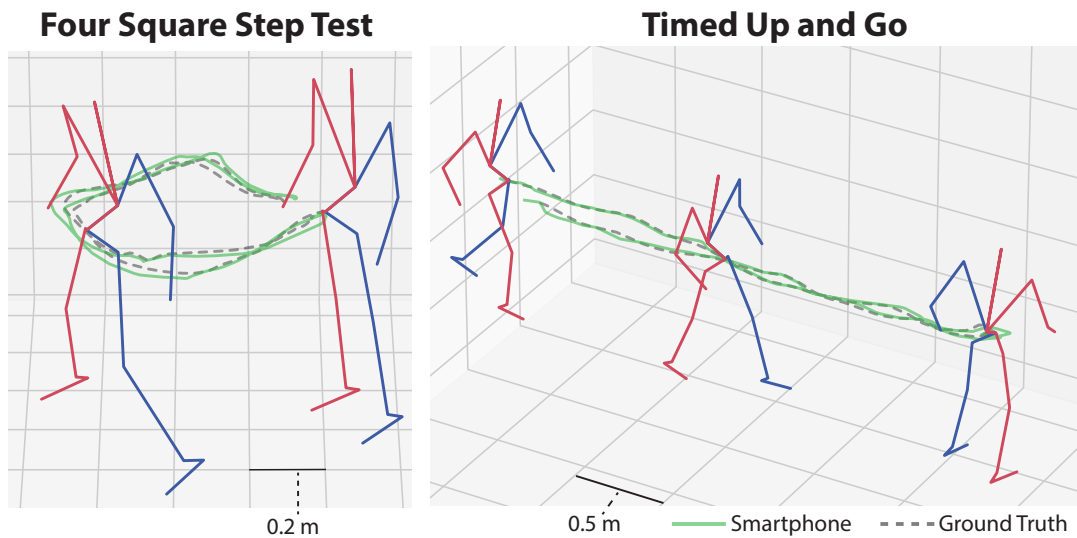
Supplementary Fig. 2. SPM1D Comparison of Multi Camera and Monocular. Raw kinematic traces ($n=1971$) compared between the Multi-Camera (MMMMC) and Single Camera (PBL) modalities of the MMMC dataset are compared using paired SPM1D t-tests, revealing significant differences at almost every timepoint in the gait cycle. Dashed lines in second row represent $p < 0.05$ significance threshold.

Supplementary Table 4. Standardized Response Mean (SRM) due to Surgical Intervention SRM is a unitless measure of change in metric before and after surgery, with 95% confidence intervals.

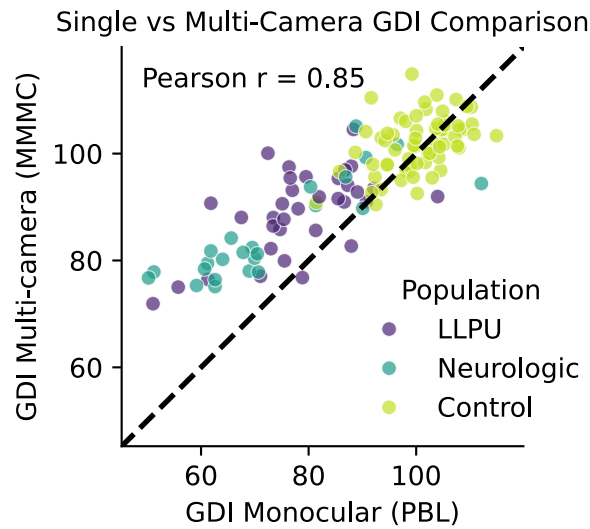
	Metric	SRM
Patient-Reported	mJOA	0.20 (-0.38 0.78)
	mJOA-LE	0.14 (-0.45 0.72)
Smartphone-Based	Cadence	0.34 (-0.20 0.88)
	DST	-0.48 (-0.88 -0.07)
	GDI	0.45 (0.00 0.91)

SPM1D Comparison. Paired SPM1D t-tests on per-step kinematics showed significant differences in knee and hip angles across nearly the entire gait cycle (Fig. S2). Given the large sample size ($n = 1,971$), these results are expected, and the absolute differences remain small. Hip flexion was consistently underestimated by the single-camera method, suggesting a potential pelvis-tilt bias, possibly driven by viewpoint-dependent errors (Fig. 4C), as most recordings were captured from behind. Knee flexion was also underestimated during stance and slightly overestimated near peak flexion. Despite the statistical significance, the magnitude of these errors is modest.

Standardized Response Mean (SRM). SRM results (Table S4) aligned closely with the effect sizes reported in Table 2. For the smartphone-based metrics, SRMs for pre/post-operative change were 0.34 for cadence, -0.48 for DST, and 0.45 for GDI. Patient-reported outcomes showed smaller SRMs of 0.20 (mJOA) and 0.14 (mJOA-LE). Notably, the confidence intervals for DST and GDI did not cross zero, whereas all other intervals did.



Supplementary Fig. 3. Root Translation Error. Body translation in 3D space reconstructed from smartphone (green, solid) closely mirrored ground truth (grey, dashed). Note that our method accounts for smartphone rotation, but not translation.

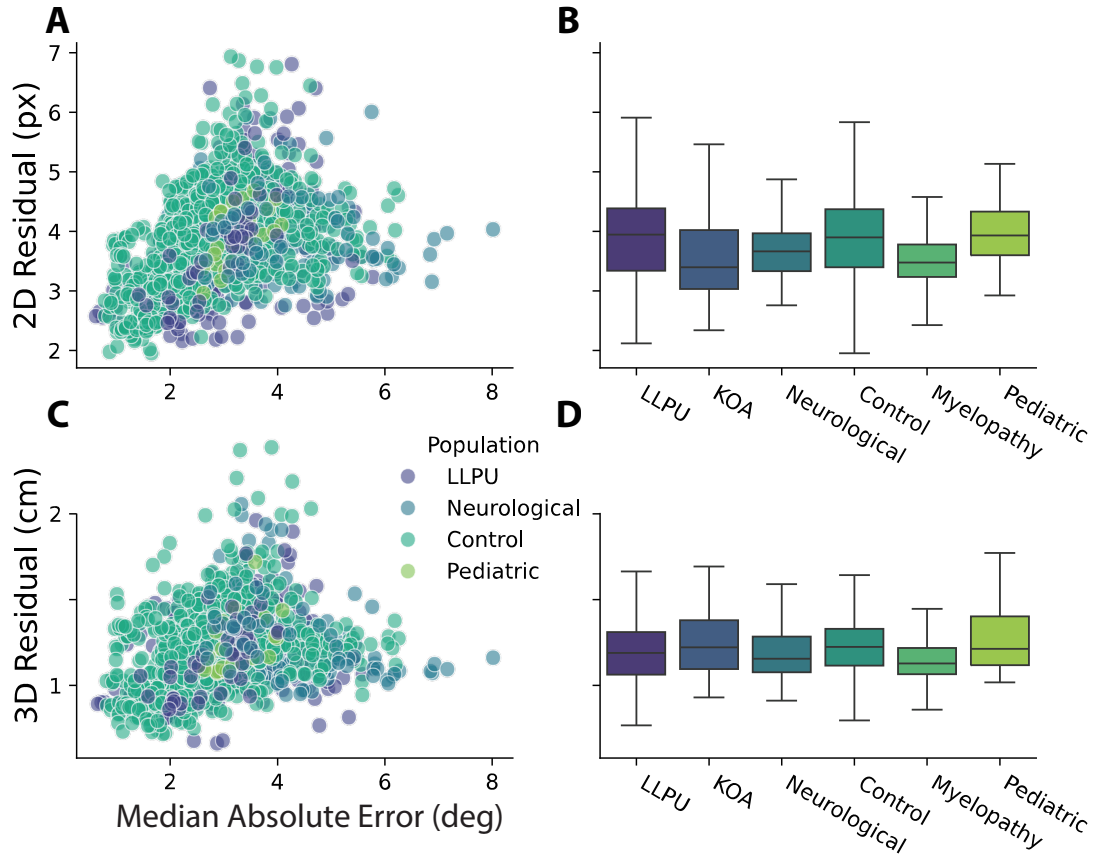


Supplementary Fig. 4. Gait Deviation Index Comparison between Monocular and Multi-View Fits GDI values were computed separately using either PBL (monocular) or MMC (multi-view) recordings from the same sessions. For each computation, we used only the data and normative distribution corresponding to the respective modality (e.g., monocular GDI was calculated using single-camera fits and a single-camera-based normative distribution).

PBL vs MMC GDI. GDI values computed from PBL and MMC 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 MMC both likely capture the same features needed for downstream gait analysis tasks (Fig. S4). We note that the PBL-based GDI shows a greater change than the MMC 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 Median Absolute Error were significant but low: 0.35 and 0.24 for 2D and 3D residuals, respectively (Fig. S5A,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

Fitting Residuals by MJAE and Population



Supplementary Fig. 5. Fitting Residuals. Correlation between 2D and 3D fitting residuals and Median Absolute Error is shown in (A), (C), respectively. 2D and 3D fitting residuals distributed across clinical populations is shown in (B) and (D), respectively.

scores, something we aim to address in the future [60, 24]. 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. S5B,D).