Comparison of a Scaled Cadaver-Based Musculoskeletal Model With a Clinical Upper Extremity Model

Nagaraja, Vikranth H.; Bergmann, Jeroen H.M.; Andersen, Michael S.; Thompson, Mark S.

Published in:
Journal of Biomechanical Engineering

DOI (link to publication from Publisher):
10.1115/1.4056172

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

Comparison of a Scaled Cadaver-based Musculoskeletal Model with a Clinical Upper Extremity Model

Vikranth H. Nagaraja*
Department of Engineering Science,
Institute of Biomedical Engineering,
University of Oxford, Oxford OX1 3PJ, UK
Email: vikranth.harthikotenagaraja@eng.ox.ac.uk

Jeroen H.M. Bergmann
Department of Engineering Science,
Institute of Biomedical Engineering,
University of Oxford, Oxford OX1 3PJ, UK
Email: jeroen.bergmann@eng.ox.ac.uk

Michael S. Andersen
Department of Materials and Production,
Aalborg University, Fibigerstraede 16,
DK-9220 Aalborg East, Denmark
Email: msa@mp.aau.dk

Mark S. Thompson
Department of Engineering Science,
Institute of Biomedical Engineering,
University of Oxford, Oxford OX1 3PJ, UK
Email: mark.thompson@eng.ox.ac.uk

Abstract *– Reliably and accurately estimating joint/segmental kinematics from optical motion capture data has remained challenging. Studies objectively characterizing human movement patterns have typically involved inverse kinematics and inverse dynamics techniques. Subsequent research has included scaled cadaver-based musculoskeletal (MSK) modeling for non-invasively estimating joint and muscle loads. As one of the ways to enhance confidence in the validity of MSK model predictions, the kinematics from the preceding step that drives such a model needs to be checked for agreement or compared with established/widely-used models.*

This study rigorously compares the upper-extremity joint kinematics calculated by the Dutch Shoulder Model implemented in the AnyBody Managed Model Repository (involving Multibody Kinematics Optimization) with those estimated by the Vicon Plug-in Gait model (involving Single-body Kinematics Optimization). Ten subjects performed three trials of (different types of) Reaching tasks in a three-dimensional marker-based optical motion capture laboratory setting. Joint angles, processed marker trajectories, and reconstruction residuals corresponding to both models were compared. Scatter plots and Bland-Altman plots were used to assess the agreement between the two model outputs.

Results showed the largest differences between the two models for shoulder, followed by elbow and wrist, with all root-mean-squared differences less than 10° (although this limit might be unacceptable for clinical use). Strong-to-excellent Spearman's rank correlation co-efficients were found between the two model outputs. The Bland-Altman plots showed a good agreement between most of the outputs. In conclusion, results indicate that these two models with different kinematic algorithms broadly agree with each other, albeit with few key differences.

Keywords:
Agreement, Comparison, Dutch Shoulder Model, Kinematics, Multibody Kinematics Optimization, Musculoskeletal modeling, Plug-in Gait model, Single-body Kinematics Optimization, Upper extremity

1 Introduction

Movement analysis involving three-dimensional stereophotogrammetry [1] has proven indispensable in objective evaluation, outcomes research, and clinical treatment planning. Although, accurately estimating the underlying segments' pose from an optical motion capture system has remained challenging [2]. Due to its robustness and computational efficiency, most musculoskeletal (MSK) modeling applications typically involve inverse dynamics analysis of input motion capture data (and external forces, if relevant). Such models use standard anatomical (or cadaver-based) models with simple parametric scaling to match individual anthropometry [3]. As the first step of MSK modeling, kinematic variables are estimated, which subsequently enables the calculation of *in vivo* joint and muscle loads [4, 5]. Thus, it is imperative to ensure the accuracy of kinematics as any errors/uncertainties introduced at earlier stages can propagate through the process and subsequently affect accuracies further down the movement analysis pipeline [6, 7, 8]. A systematic review emphasized that the model-derived kinematics largely depend on the modeling choices, and the joint kinematics accuracy is not well-established [9]. Particularly, the choice of marker set [10], differences in joint coordinate system definitions between models [7], and soft tissue artifacts (STA) [11, 12] were found to affect joint kinematics, influencing subsequent calculations. As one of the means to enhance confidence in the validity of predictions from an MSK model, the kinematics that drives an MSK model need to be checked for agreement or compared with established/widely-used models (especially in the

*Address all correspondence to this author.

upper extremity (UE), where such studies have lagged its lower-extremity counterpart) before they can subsequently be applied to estimate *in vivo* joint and muscle loads [9, 13]. In addition, compared to clinical gait analysis, UE motion analyses have been impeded by numerous issues, such as the complex nature of the movements involved, extremely large range of motion, methodological variability, as well as a lack of consensus and standardization thereof [14, 15].

The two most widely-used algorithms for estimating kinematic parameters are *Single-body Kinematics Optimization (SKO)* and *Multibody Kinematics Optimization (MKO)*. SKO assumes that the experimental markers are rigidly attached to bones or segments, and the joint kinematics are calculated between the adjacent segments defined from the markers' positions but *without* considering joint constraints [16]. Contrarily, MKO—a widely-used step in MSK-modeling exercises—employs a skeletal-joint model with a set of modeled markers rigidly attached to the model [9, 17, 18, 19, 20]. This technique embeds a least-squares approach and articular constraints. It has been increasingly used since the first application of the Lu and O'Connor algorithm [17] to the UE by Roux et al. [20]. MKO computes joint kinematics by adjusting the model coordinates to attain the best match between the positions of model-determined (also called 'virtual') markers and experimental markers on a subject's skin (from the motion capture), thereby accepting marker residuals to enforce the joint constraints of the skeletal-joint model (thus avoiding joints dislocation as well as non-physiological/unrealistic bone positions and orientations). Performing inverse kinematics and inverse dynamics analyses using a model representing physiological joints and scaled to the subject's anthropometry has been recommended [13].

It has been widely suggested that the MKO approach minimizes the effects of STA [20, 21]; and compensating for STA in relation to the underlying bones (especially while non-invasively modeling the kinematics of a complex joint such as the shoulder using skin-mounted markers [21]) has remained one of the key problems in obtaining accurate and reliable skeletal kinematics from motion capture [12, 22, 23]. MKO applied to an MSK model enables analyses such as estimation of muscle-tendon length [24], muscle moment-arm [25], muscle-tendon force [26], and joint contact forces [27], which may provide valuable information for clinical decision-making for people with movement disorders (e.g., cerebral palsy). There are additional benefits of the MKO approach reported, such as the ability to use a reduced marker set [28], consistent inter-trial variability across different marker sets [10], improved inter- and intra-observer repeatability [29], avoidance of joint dislocations [17], and robustness to marker mislocation [30] and computations more robust to noise [17]. Notably, the number of UE studies reporting the application of MKO has been fewer than that for the lower extremity [9, 21].

Begon et al. [9] have provided recommendations to systematically evaluate MKO algorithms that adhere to the "best practices" proposed for MSK modeling [13, 31]. Among other forms of assessment that remain relevant for validation and performance, comparing MKO with other techniques (i.e., SKO and non-optimized kinematics) for estimating joint kinematics has been a popular, non-invasive choice [9]. For instance, many lower-extremity studies have non-invasively compared the Plug-in Gait model (SKO approach) with other established models to assess biomechanical parameters (e.g., [32]). Furthermore, as the first step of formal validation, Hicks et al. [13] have recommended comparing the model's outputs to as many independent datasets as possible and previously validated and published studies. Nevertheless, compared to the lower extremity, very few studies have compared the UE kinematics between SKO and MKO algorithms [9, 33, 34, 35].

Therefore, it is critical to determine whether different algorithms (e.g., SKO and MKO) produce different joint kinematics. There seems to be a need for comparing UE kinematics estimated by the cadaver-based MSK models with those calculated by established or widely-used (clinical) models before they can confidently be used for clinical decision-making. Thus, this study's objective is to qualitatively and quantitatively compare and check the agreement of joint kinematics calculated by the MKO model (available with the AnyBody Managed Model Repository; known as the AnyBody model henceforth) and the clinically widely-used SKO model (available as the Vicon Plug-in Gait model; known as the Plug-in Gait model hereafter) during select task execution in non-disabled individuals. Based on the differences between these two models, we hypothesize that the estimated kinematics would differ between the two models.

2 Methods

2.1 Ethics

This study was approved by the local Research Ethics Committee (Ref no. 16/SC/0051), and informed consent was obtained from each participant before data capture.

2.2 Study participants

Ten right-handed non-disabled volunteers (Three females and seven males; Age: 26.4 $\pm$ 3.0 years; Height: 1.7 $\pm$ 0.1 m; Weight: 67.9 $\pm$ 12.5 kg; Body Mass Index: 23.6 $\pm$ 3.6 kg/m^2) participated in this study. Notably, eleven subjects were recruited initially; however, data captured for one subject was excluded due to poor quality. All volunteers were reimbursed for their participation time.

2.3 Equipment

The data was acquired in a laboratory setting using a 16-camera Vicon T40S Series system (Vicon Motion Systems, Oxford, UK) at a sampling frequency of 100 Hz. A custom-built apparatus was developed that could be adjusted to suit the anthropometric requirements of different participants (Figure 1) and facilitate the execution of *Reaching* tasks. The Reaching tasks involved the participant reaching various pre-defined points as instructed. The reach point ('Front', 'Right', or 'Left' points in Figure 1) was within 90% of an individual's arm length (i.e., acromion to middle-fingertip length, with the arm hanging down) to minimize the contribution of trunk movement for task execution [36]. Participants performed the task in a seated position on a height-adjustable chair located behind a height-adjustable table with the test apparatus on top.

2.4 Protocol

Detailed instructions and a couple of practice sessions were provided for each subject before data collection. First,

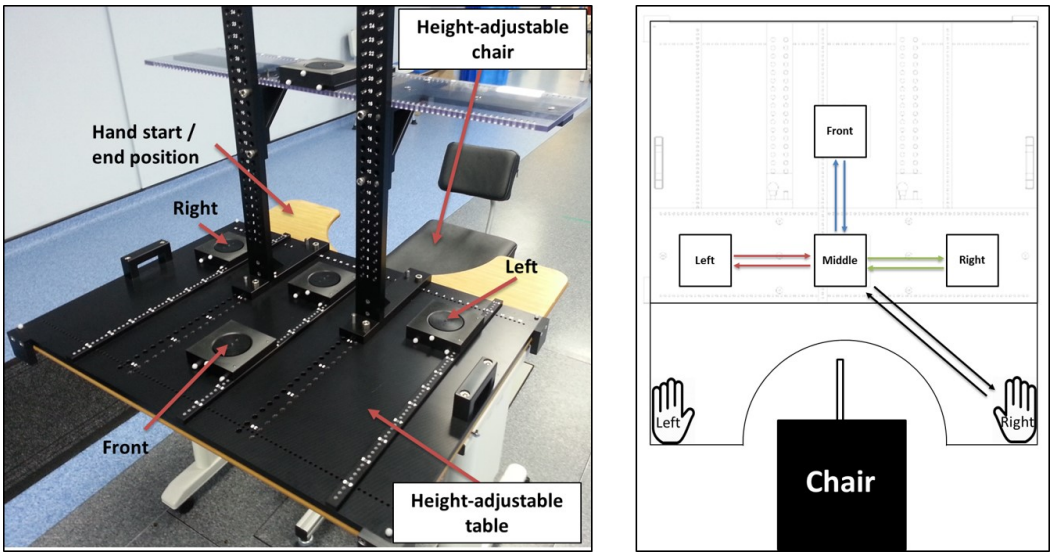


Fig. 1. Left side: Custom-built apparatus for *Reaching* task execution; Right-side: Reaching task setup

participants performed a 'static' trial (by standing in a stationary neutral or anatomical position in the center of the capture volume), followed by three 'dynamic' trials of Reaching task with their dominant hand at a self-selected pace. Reaching tasks were executed in three different directions: 'Reach-to-Front (RF task),' 'Reach-to-Right (RR task),' and 'Reach-to-Left (RL task).' Execution of the RF task involved moving the hand from the *'Start/End'* position to touch a fixed point in the *'Middle'* block with the index finger, followed by touching a similar point in the *'Front'* block (*'Right'* or *'Left'* block for the RR or RL task, respectively), then the hand would follow the same instructions/path in the return phase, and then the hand would return to the *'Start/End'* location.

2.5 Data collection

Passive, retro-reflective markers (⌀ 9.5 mm) were placed on the participant's skin using a double-sided hypoallergenic tape per the Plug-in Gait marker set [37]. RIC and LIC markers were added to this protocol (used to reconstruct the position and orientation of the Pelvis segment in the AnyBody model [38]) to mitigate the potential issue of RASI and/or LASI marker's occlusion by the table during task execution. Furthermore, anthropometric aspects such as weight, height, arm length, shoulder depth, elbow joint width, wrist joint width, and hand thickness were collected to facilitate personalization (i.e., subject-specific scaling) of the two models. Marker placement and data capture for all subjects was performed by the same tester (VHN) to remove inter-tester variability as an error source.

2.6 Data processing

2.6.1 Pre-processing

Reconstruction of raw data, marker labeling, and gap-filling steps were performed using the Vicon Nexus v.2.5 software (Vicon Motion Systems, Oxford, UK) [39]. Missing marker data were interpolated using a rigid-body hypothesis or filled using other marker gap-filling options (selected as appropriate to the segment) available in the Vicon Nexus software. For each trial, the identification of 'Start' and 'End' frames (hand leaving and returning at the 'Hand Start/End' position (Figure 1), respectively) were performed manually. The raw marker data were filtered using a 4^{th} order, Zero-lag, Low-pass Butterworth filter with a cut-off frequency of 6 Hz. The data for different trials and participants were resampled to a fixed length for temporal normalization and to facilitate comparison.

2.6.2 Adaptation and assumptions in both models

Details pertinent to the two models under consideration can be found elsewhere [37, 38]. The following assumptions and adaptations were implemented—if they were not already available—in both the Plug-in Gait and AnyBody models to match study requirements:

– ***Common across both models:*** The MSK system is assumed to be a rigid body, and only a unilateral (right-handed) UE body model is considered. Motions of the head and neck are neglected, and the joints are assumed to be ideal and frictionless. Inter-individual differences in carrying angle (between the upper and lower arm) are overlooked. The internal/external rotation at the wrist is ignored, and the motions of the metacarpals and phalanges are omitted from the hand's complexity.

– ***Plug-in Gait model:*** The UE is divided into the torso, upper arm, forearm, and hand. The gleno-humeral joint is modeled as a spherical joint, and the elbow joint is modeled as a revolute joint. The wrist joint is created as two revolute joints where the axes of rotations are not coincident. In the Plug-in Gait model, six degrees of freedom exist for each segment. The joint kinematics are calculated between the adjacent segments defined from the experimental markers' positions but without considering joint constraints.

– ***AnyBody model:*** The UE is divided into the torso, scapula, clavicle, upper arm, radius, ulna, and hand. The scapulothoracic joint between the scapula and the surface of the thorax is modeled as an ellipsoid. The elbow joint is modeled as two non-orthogonal revolute joints — one for flexion/extension and one for pronation/supination. The wrist joint is created as two revolute joints — one for flexion/extension and one for radial/ulnar deviation where the axes of rotations are not coincident.

2.6.3 Plug-in Gait Modeling (SKO approach)

The Plug-in Gait model [37] is one of the most widely-used computational models adopting the SKO approach in movement analysis and is well adopted by the clinical gait analysis community [40, 41], and perhaps for researchers evaluating upper-limb movements during gait. This model is an implementation of the *Newington Children's Hospital* model [42] and the *Helen-Hayes* model [43]. The Plug-in Gait model is one of the widely-used UE models—if not *the most* widely-used model—for the following reasons:

1. The demonstrated reliability of kinematic and kinetic data from the Plug-in Gait model suggests that the magnitude of the errors estimated by this model is clinically reasonable [44].
2. The Plug-in Gait model is popular partly since it is distributed in Vicon's software packages, e.g., Vicon Nexus [39]. Hence, many laboratories worldwide rely on the commercial version of the Plug-in Gait model for both upper- and lower-extremity studies.
3. The accompanying Plug-in Gait marker set [37] is a popular marker protocol widely used in research for motion analysis and MSK modeling.
4. Finally, attempts have recently been made (e.g., [45]) to enable easy integration and improvements in the Plug-in Gait model code and 'disconnect' marker locations of this model from the commercial software. Such efforts are likely to facilitate cross-platform compatibility and comparison studies with the Plug-in Gait model in the future.

Thus, the Plug-in Gait model [37], available with the Vicon Nexus software, was used for SKO modeling, which employs a minimal number of markers to compute three-dimensional segment and joint kinematics. Using the pre-processed experimental markers in the Plug-in Gait model, all the joint angles were calculated by comparing the relative orientations of the two segments (Supplementary Table 1) [46, 37]. Technical segment frames or segment local coordinate systems for the Plug-in Gait model are detailed in Supplementary Table 2 [37]. The Plug-in Gait model was adapted (see Section 2.6.2) to suit study requirements to generate the 'Plug-in Gait model output' as shown in Figure 2. This endeavor yielded Plug-in Gait model-processed markers, segment origins, and joint angles.

2.6.4 AnyBody Modeling (MKO approach)

The AnyBody Modeling System is one of the few dedicated MSK engines available currently, and their human body model is used widely in research and across a broad range of industries. The AnyBody model is a widely-used computational model involving the MKO approach in movement analysis and MSK modeling communities. For each segment and joint, the AnyBody and Plug-in Gait models have different joint coordinate systems, sequence of angle outputs, etc. [37, 38]. Therefore, to establish an equivalence of the upper-extremity joint coordinate systems and kinematic outputs between the two models, numerous steps (detailed later in this sub-section) were carried out. By ensuring the same marker set, joint coordinate system, etc., our study facilitates evaluating the difference in joint kinematics that can be solely attributed to the different kinematic methods under consideration (MKO versus SKO).

Raw data (cleaned/labeled marker trajectories from the Vicon Nexus software) were exported and processed in the commercially available AnyBody Modeling System v.6.0.6 (AnyBody Technology A/S, Aalborg, Denmark). The UE of the *Standard MoCap* model, which is an implementation of the Dutch Shoulder Model (i.e., the Shoulder Arm Model) [47] and the Spine Model [48] available with the AnyBody Managed Model Repository v.1.6.5 was adapted (see Section 2.6.2) to suit study requirements. Model scaling and segment tracking were performed using experimental markers and anthropometric dimensions (detailed in Section 2.5) according to the scaling algorithm of Andersen et al. [49] for the static trial (also known as the 'calibration trial'). Generally, tracking the dynamic trials for the MKO approach requires manually fitting weights for the individual markers to account for different STA that occurs at each marker [5, 9, 50]. In this study, the markers have equal weight [9, 16] since understanding the effect of marker weight on the tracking of segments was not within the current scope.

New markers (corresponding to the Plug-in Gait marker set) are defined and registered onto the various segments of the scaled AnyBody model (Figure 2). Using these markers, local reference frames for each segment (i.e., thorax, upper arm, etc.) are defined to match the technical segment frames of the Plug-in Gait model as shown in Supplementary Table 2 [37]. Next, the experimental markers are mapped onto these newly-registered markers. This step was followed by matching the rotation sequence along the different axes of the segment reference frames with the absolute Euler angle sequences of the corresponding segment in the Plug-in Gait model [37]. Finally, the joint angles along the various degrees of freedom (for shoulder, elbow, and wrist) are calculated as the relative angles between these segments as per the corresponding Euler angle sequences of the Plug-in Gait model joints (per Supplementary Table 1) [37]. Tracking for subsequent dynamic trials was performed using an algorithm developed by Andersen et al. [19]. The location of the processed markers that were rigidly attached to AnyBody model segments and the scaled segment lengths—after performing MKO within the AnyBody Modeling System—are saved. These processed markers were then used to create a new set of Coordinate 3D (.C3D) files corresponding to each trial using MATLAB® R2016b (Mathworks Inc., Natick, MA, USA) and *Biomechanical Toolkit* [51]. These files were subsequently used to perform Plug-in Gait modeling in the Vicon Nexus software to create 'AnyBody model output' with the joint angle sequences (per Supplementary Table 2 [37]) that were consistent with the Plug-in Gait model output created directly using the experimental data (in Section 2.6.3).

Notably, there are two different marker sets rigidly attached to the AnyBody model: (i) a set of experimental markers corresponding to those placed by the tester (VHN) on the bony landmarks that are used for tracking, and (ii) registration of the true marker location onto the scaled segments that are then exported for analysis in the Vicon Nexus software for comparing the two models. This endeavor yielded AnyBody model-processed markers, Plug-in Gait model segment origins registered to the AnyBody model in the standing reference trial, and AnyBody model-estimated joint angles with identical definitions to Plug-in Gait model outputs.

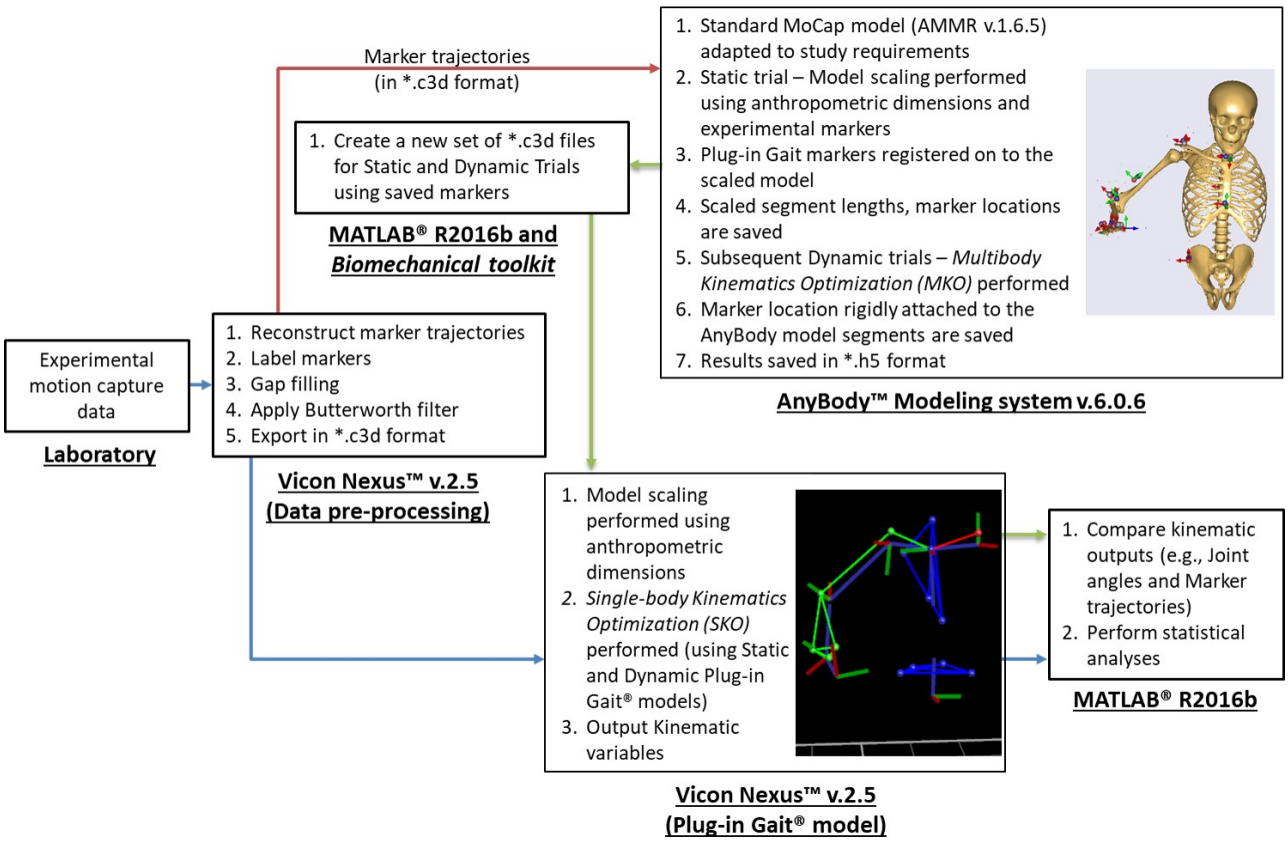


Fig. 2. Pipeline to estimate AnyBody and Plug-in Gait model outputs and facilitate comparison

2.7 Data/Statistical analysis

Data reduction techniques, statistical analysis, and comparison of both model outputs were carried out with MATLAB® R2016b. Outcome variables of interest estimated by the two models were the joint angles (in degrees) as well as processed marker trajectories and segment origins (in mm). A Jarque-Bera test [52] was used to verify normality, and failure to satisfy parametric assumptions led to the use of median and inter-quartile range (median (IQR)), as well as the Wilcoxon rank-sum test [53] for descriptive statistics. Statistical significance was detected at $p < 0.05$ for all relevant tests. Sign convention for the joint angles was — Flexion (Fl), Abduction (Ab), Radial deviation (Rd), and Internal rotation (Int) are positive (+ve); Extension (Ex), Adduction (Ad), Ulnar deviation (Ud), and External rotation (Ext) are negative (-ve).

Qualitative analyses were performed via visual assessment to assess the match in median (IQR) ensemble plots ($n = 10$; three trials) for both the model outputs for the three tasks. We quantitatively assessed consistency between the two models through Difference Analysis, Wilcoxon rank-sum test [53] of estimated joint range-of-motion angles, and assessing reconstruction residuals through box-plots. Besides, the degree of agreement was evaluated through Scatter plots and Bland-Altman analysis [54, 55]. While undertaking kinematics comparison of MKO methods with various noninvasive methods, differences are usually expressed in terms of root-mean-square differences (RMSD) (e.g., [56]). For Difference analysis, we chose 5° — 10° RMSD between the two models as acceptable for comparing the UE kinematics based on a similar lower-extremity study non-invasively comparing the consistency between SKO and MKO techniques [56].

We report reconstruction residuals as recommended earlier [9]. Reconstruction residuals for markers were calculated as the Euclidean distance between AnyBody model-processed markers (that were subsequently modeled using the Plug-in Gait model, Figure 2) and experimental markers. Similarly, the reconstruction residual for segment origins (Section 2.6.4) was calculated as the Euclidean distance between the segments created in the Plug-in Gait model using AnyBody markers and those created in Plug-in Gait model using the experimental markers (Figure 2).

The data for scatter plots were down-sampled by a factor of 100 to have a minimum set of points to be able to visually analyze the strength, direction, and form of the relationship between kinematic variables corresponding to the two models. Spearman's rank correlation coefficients (ρ) were calculated and reported for the scatter plots (see Supplementary Table 4 and Supplementary Figure 3 (a)–(c)) and categorized similarly to Taylor [57], as "weak" ($\rho \leq 0.35$), "moderate" ($0.35 < \rho \leq 0.67$), "strong" ($0.67 < \rho \leq 0.90$), and "excellent" ($\rho > 0.90$).

To assess agreement/consistency between kinematic outputs of the AnyBody and Plug-in Gait models, the Bland-Altman plots [54] for each model were compared. The reference kinematics, determined using the Plug-in Gait model, assumed to be a reference, was set as an invariant for comparison of the AnyBody model outputs. The data were down-sampled by a factor of 100 to have a minimum set of points to be able to analyze the differences and biases in the Bland-Altman plots visually. Bias and 95% lower and upper limits of agreement are reported in the respective graphs (see Supplementary Table 5 and Sup-

plementary Figure 4 (a)–(c)).

3 Results

The following sub-sections provide the set of results to compare the two model outputs qualitatively and quantitatively. The results also aim to check the agreement between the outputs and identify if there is any statistical significance between these two sets of kinematic outputs.

3.1 Qualitative comparison

Characteristic UE joint excursions are represented as median (IQR) ensemble plots in Figure 3 for the three joints (i.e., shoulder, elbow, and wrist). For qualitative comparison via visual assessment of the match between the AnyBody model outputs (in red) and Plug-in Gait model outputs (in blue), it can be noticed in Figure 3 (A) that there is a good match between the outputs of the two models for the shoulder Fl/Ex angles for all three tasks (i.e., RF, RR, and RL tasks). However, the match between the two models for the shoulder Ab/Ad and shoulder Int/Ext angles seems relatively poor for RF and RL tasks. The AnyBody model tends to overestimate the shoulder Ab/Ad angles and underestimate the shoulder Int/Ext angles for the RF task, and vice versa for the RL task. On the contrary, the trends for shoulder Ab/Ad angles and shoulder Int/Ext angles were similar between models for the RR task. A possible *cross-talk* between shoulder Ab/Ad and shoulder Int/Ext angles (i.e., one joint rotation getting interpreted as another [58]; in other words, the under- and over-estimation of the two joint angles when compared with their corresponding AnyBody model outputs) calculated by the Plug-in Gait model can be observed for the RL task, although this has not been quantified in our study. The elbow Fl/Ex angles (in Figure 3 (B)) agree well between the two models for all three tasks, especially for Fl angles less than 90°. The wrist Fl/Ex and wrist Rd/Ud angles (in Figure 3 (C)) match well in terms of the median; however, there is a large difference in their IQR. Interestingly, the differences between most of the two model outputs—across joints and degrees of freedom—are at their highest mid-way during the task (i.e., 50% of task completion), which is when maximum joint excursions in the selected tasks are generally expected.

3.2 Quantitative comparison

3.2.1 Difference analysis

The RMSD for the median for all the angles and tasks were less than 9° (Table 1), indicating a good match between the models. Besides, various Difference measures calculated between the two model outputs (for the three joints along their respective degrees of freedom) are reported in Supplementary Table 3. A few notable differences are: the absolute maximum difference between the AnyBody and Plug-in Gait models was the highest for shoulder Ab/Ad with median (IQR) values of 14.6° (19.2°) and 20.1° (25.4°) for RF and RL tasks, respectively. In this study, the maximum difference values were the same as the absolute maximum difference values. The highest minimum difference value of -11.9° (18.9°) was found for the shoulder Int/Ext angles for the RL task. The maximum difference range was at the shoulder Ab/Ad angles (19.3° (17.6°)) for the RF task, and shoulder Ab/Ad angles (23.8° (23.4°)) and shoulder Int/Ext angles (14.8° (16.3°)) for the RL task.

3.2.2 Wilcoxon rank-sum test

Wilcoxon rank-sum test [53] was carried out to compare the joint range-of-motion angles estimated by the two models (Table 2). Significant differences were detected for: (i) shoulder Ab/Ad angles for the RF task, (ii) shoulder Fl/Ex and wrist Rd/Ud angles for the RR task, and (iii) all shoulder angles and wrist Fl/Ex angles for the RL task.

3.2.3 Reconstruction residuals

Supplementary Figures 1 and 2 correspond to boxplots for reconstruction residual for markers and segment origins, respectively. It can be noted that T10, STRN, RSHO, and RFIN markers have numerous outliers. The marker reconstruction residual was highest in C7 and RELB markers for all three tasks. The segment origin reconstruction residual (Supplementary Figure 2) was the highest for the humerus, followed by the hand. These reconstruction residual values in the Plug-in Gait model are theoretically zero as there are no joint constraints, i.e., the technical reference frames and joint kinematics are defined based on the markers for every frame.

3.3 Agreement between the two models

3.3.1 Scatter plots and Correlation

Both the model outputs have a linear (or near-linear) relation for all the angles and each of the three tasks, as seen in the scatter plots in Supplementary Figure 3 (a)–(c). Some of the notable exceptions are described here. For the RF task, shoulder Ab/Ad, shoulder Int/Ext, wrist Fl/Ex, and wrist Rd/Ud angles have a large deviation from linearity for some of the participants. In the case of the RL task, shoulder Ab/Ad, shoulder Int/Ext, and wrist Rd/Ud angles were not linearly related for some subjects, especially for Ad angles. Scatter plots show a strong to excellent correlation (via Spearman's rank correlation co-efficient) between the two model outputs (Supplementary Table 4).

3.3.2 Bland-Altman analysis

Bland-Altman bias plots are provided for all three tasks in Supplementary Figure 4 (a)–(c), with a corresponding summary in Supplementary Table 5. Notably, the Bland-Altman plot method only defines the intervals of agreements; it does not say whether those limits are acceptable or not. Acceptable limits must be defined *a priori*, based on clinical necessity, biological considerations or other goals [59]. For the shoulder Fl/Ex angles for all three tasks, it can be noticed that the AnyBody model tends to overestimate Fl angles and underestimate Ex angles. The spread for the RR task is relatively less for all three shoulder angles. The elbow Fl/Ex angles are consistently overestimated by the AnyBody model for all three tasks, as evident by the value median differences. For the remaining outputs, Bland-Altman plots revealed no clear trends, so they are not highlighted.

4 Discussion

The effect of biomechanical model choice on the UE kinematic estimates need to be better understood as per

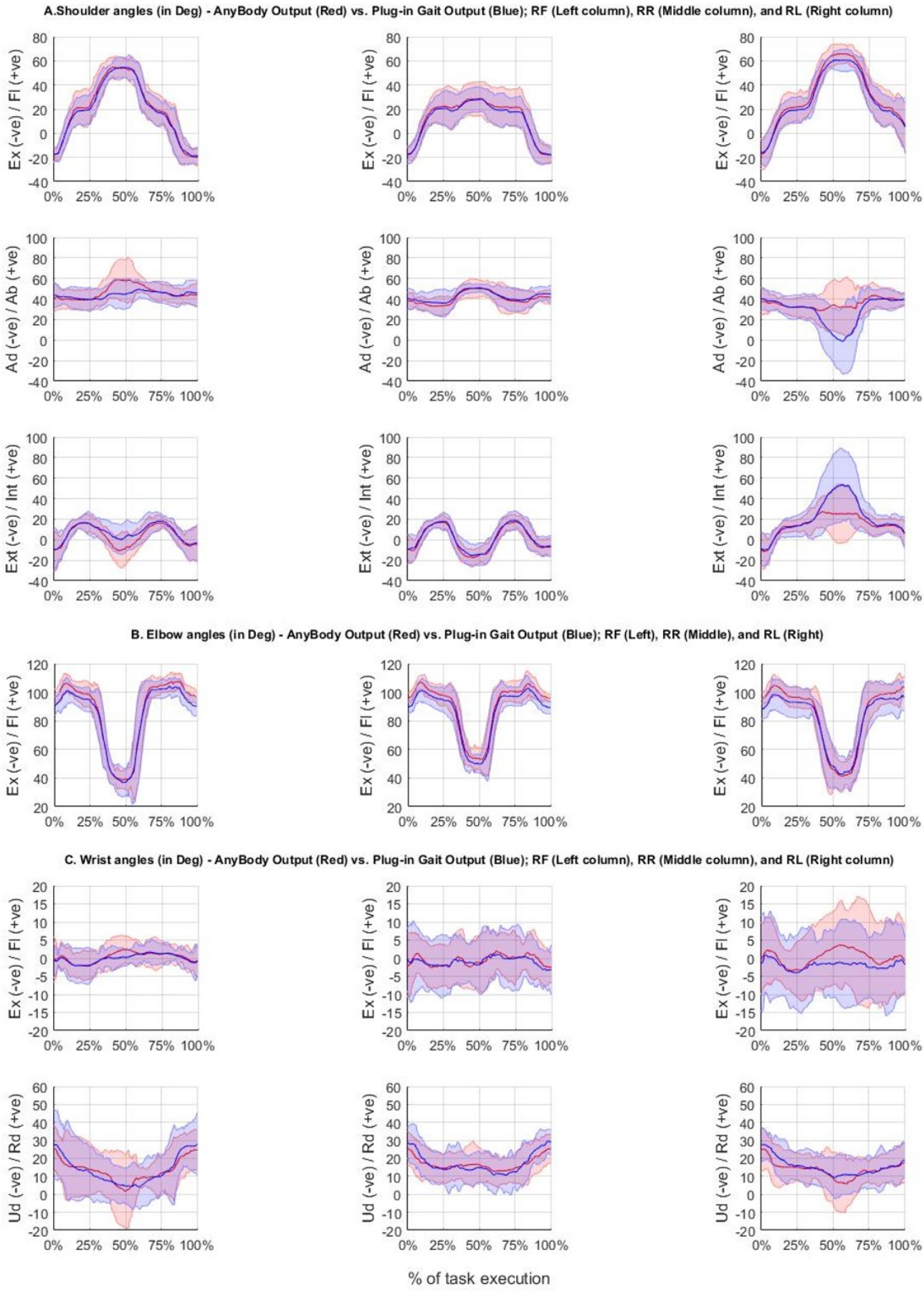


Fig. 3. Ensemble median (IQR) plots for (A). Shoulder, (B). Elbow, and (C). Wrist angles (in degrees); AnyBody model outputs (in Red) versus Plug-in Gait model outputs (in Blue); RF (left column), RR (middle column), and RL tasks (right column); three trials & n = 10.

the suggested modeling and simulation "best practice" of comparing results from studies that used different models [9, 13]. By constraining the joint kinematics to physiological motions by the MKO approach, the feasible re-

Difference Type (Between AnyBody & Plug-in Gait model outputs)	Joint – Degree of Freedom	Reach-to-Front (RF) task Median (IQR) Degrees	Reach-to-Right (RR) task Median (IQR) Degrees	Reach-to-Left (RL) task Median (IQR) Degrees
Root-mean-squared-difference (RMSD)	Shoulder – Fl/Ex	2.7 (0.8)	2.9 (2.4)	3.3 (1.4)
	Shoulder – Ab/Ad	6.4 (5.7)	5.3 (5.0)	8.6 (8.3)
	Shoulder – Int/Ext	5.8 (4.9)	2.3 (2.3)	5.5 (8.4)
	Elbow – Fl/Ex	3.6 (2.7)	3.9 (2.1)	3.8 (2.1)
	Wrist – Fl/Ex	2.9 (1.2)	2.3 (2.4)	3.5 (0.9)
	Wrist – Rd/Ud	4.2 (2.9)	3.5 (2.0)	3.7 (4.1)

Table 1. Difference between AnyBody and Plug-in Gait model-estimated joint angle outputs in degrees; three trials & n = 10.

Joint – Degree of Freedom	*Reach-to-Front (RF) task*			*Reach-to-Right (RR) task*			*Reach-to-Left (RL) task*		
	AnyBody	Plug-in Gait	p-Value	AnyBody	Plug-in Gait	p-Value	AnyBody	Plug-in Gait	p-Value
	RoM (in Median (IQR))			RoM (in Median (IQR))			RoM (in Median (IQR))		
Shoulder – Fl/Ex	77.8 (11.1)	76.6 (11.1)	0.819	49.8 (6.3)	47.4 (7.0)	0.044*	88.4 (10.8)	82.2 (8.4)	0.004*
Shoulder – Ab/Ad	27.4 (14.5)	16.4 (9.9)	0.00024*	18.8 (6.7)	16.6 (4.4)	0.473	26.7 (21.2)	50.4 (34.9)	0.001*
Shoulder – Int/Ext	33.2 (8.5)	31.4 (10.2)	0.297	35.7 (4.3)	35.9 (4.9)	0.888	41.7 (23.2)	62.6 (32.1)	0.002*
Elbow – Fl/Ex	72.5 (6.4)	71.0 (9.5)	0.379	53.7 (6.3)	53.1 (8.1)	0.600	66.5 (12.5)	66.0 (17.4)	0.181
Wrist – Fl/Ex	13.4 (5.7)	10.7 (4.8)	0.171	13.0 (4.7)	11.1 (3.3)	0.223	12.0 (7.5)	9.8 (4.6)	0.014*
Wrist – Rd/Ud	22.7 (11.7)	24.9 (6.8)	0.252	16.6 (5.7)	21.3 (5.6)	0.0003*	20.3 (13.2)	20.6 (9.6)	0.865

*Statistical significance ($p < 0.05$)

Table 2. Results of Wilcoxon rank-sum test of AnyBody and Plug-in Gait model-estimated joint range of motion (RoM) in degrees along different degrees of freedom; three trials & n = 10.

gion is smaller, and the solution is less sensitive to STA [9]. Numerous bone pose estimators, including MKO, proposed to reduce the effects of STA on estimated joint kinematics have been listed [11]. However, no consensus has been reached on the best available estimator or the maximum errors associated with these different methods [60, 61]. Our study complements the work of previous such investigations into motion capture and estimation of UE joint/segmental kinematics. The following sub-sections discuss the qualitative and quantitative results and place these in a broader context within this field.

4.1 Qualitative comparison

Qualitatively, the ensemble plots generally showed a good match between the two model outputs for several degrees of freedom across the three joints. Differences were higher for tasks requiring higher joint angles (such as RF and RL tasks). This could possibly be attributed to STA since, compared to the Plug-in Gait model (SKO approach), the AnyBody model (MKO approach) minimizes the effects of STA [9, 21]. Additionally, these differences could also be high because the Plug-in Gait model is relatively simple. In contrast, the AnyBody model (involving the Delft Shoulder Model) is more complex, imposes joint constraints, and hence, is generally considered more accurate. Therefore, a ground-truth bone-pin [33, 34, 62] or fluoroscopy study [35] could help reveal the cause of these differences. Elbow flexion alone is a strong indicator of a patient's ability to perform an activity of daily living [63], and in Figure 3, similar elbow angles across the two models are a good indication of the match. The match for the medians of the wrist angles (both Fl/Ex and Rd/Ud angles) is good but has a high IQR value, possibly due to the RFIN marker being prone to STA during task execution due to its placement on the 2nd metacarpophalangeal joint.

Joint angles estimated during motion analysis are often subject to error caused by kinematic *cross-talk*, i.e., one joint rotation (e.g., flexion) getting interpreted as another (e.g., abduction) [58]. Several studies have used cross-talk as a metric to compare and judge the quality of their models, assuming the smaller the cross-talk value, the better the model (e.g., [29, 58]). In Figure 3, a possible cross-talk between shoulder Ab/Ad and Int/Ext angles calculated by the Plug-in Gait model can be observed for the RL task, although this has not been quantified in our study. This could also be due to the simpler shoulder joint of the Plug-in Gait model compared to the more complex Dutch Shoulder Model (i.e., AnyBody model). Laitenberger et al. [64] recommend introducing a more complex shoulder mechanism that may improve the kinematic cross-talk during functional tasks. It could be possible that the choice of the joint coordinate systems and gimbal lock positions might influence the kinematic estimates in some tasks. Chin et al. [65] have suggested that apart from STA, mislocation of shoulder and elbow joint centers may cause elbow Fl/Ex cross-talk to enter into upper-arm Int/Ext rotation. A better understanding of these aspects would be a worthwhile pursuit for future studies.

The kinematic cross-talk effect is an error that could be caused by two major sources leading to modification of the joint coordinate system orientation: (i) inaccuracy inherent in the motion capture system [66] used to measure the locations of points/markers that define joint coordinate system axes; and (ii) errors associated with palpation of external anatomical landmarks—whose locations may be challenging to determine with sufficient accuracy—which can lead

to inter- and intra-examiner repeatability issues and variability due to individual anatomical differences [67]. However, the MKO approach offers improved inter- and intra-observer repeatability [29] and robustness to marker mislocation [30], and thus, it might be better suited to handle cross-talk than the SKO technique.

4.2 Quantitative comparison

4.2.1 Difference analysis

Using different biomechanical models for analysis may produce differing kinematic results, leading to differences in clinical interpretation, such as misclassification of non-disabled or pathological movement patterns. Relatively larger differences between the AnyBody and Plug-in Gait model outputs were observed at the shoulder than at the elbow and wrist, as indicated in a similar lower-extremity study [32], where the largest errors were observed for the hip than the knee and ankle joints. Our results showed the largest differences between the AnyBody and Plug-in Gait models for the shoulder, followed by elbow and wrist, with all RMSD values less than 10°, indicating an acceptable match (although this limit might be unacceptable for clinical use). The Fl/Ex angles, with a generally larger range of motion, have lower differences of ∼3° for the three tasks and all three joints (Table 1). However, it is acknowledged that the shoulder joint (which has an extremely large range of motion) should have different limits for acceptability than the other upper-limb joints. A metric such as normalized RMSD (e.g., normalized to the data's range or median) could help address this issue. The results suggest that the differences between the kinematic outputs of the two models are task-dependent, joint-dependent, and vary depending on the joint's degree of freedom. Nevertheless, this aspect needs to be scrutinized in future studies.

Placement and motion of markers relative to anatomical landmarks can introduce measurement errors while calculating joint kinematics from three-dimensional marker-based motion capture [6, 66]. In the MKO framework, open-loop chain models, including mostly universal and spherical joints, present the advantage of avoiding non-physiological joint dislocation, which appears when using SKO [21]. Roux et al. [20] showed a significant reduction in errors and variability due to skin movement using an MKO approach. However, it should be noted that the Plug-in Gait model does not attempt to correct STA. Notably, in the absence of joint constraints, MKO is equivalent to SKO [9]. It has been noted that most of the differences in joint kinematics can be attributed not only to the joint constraints but also to the different definitions of anatomical frames associated with each method [16,32]. To mitigate the effect of STA, Begon et al. [33] assessed the performance of MKO during the measurement of humeral rotational kinematics for shoulder Fl/Ex and Ab/Ad angles while concurrently recording bone-pin markers for use as the 'gold standard' comparison; humerus orientation errors were halved to less than 5° because of their novel algorithm. Begon et al. [33] also reported that the MKO approach improved the accuracy by 40–50%. When validated against ground-truth bone kinematics, typical errors for the model-derived glenohumeral rotations were between 3° and 10° during arm flexion and abduction, and the errors were maximal for internal–external rotation [33, 35].

The limit for differences and their interpretation often depends on the application – the acceptable limits for errors could vary with the joint and the corresponding degree of freedom under scrutiny [20]. Begon et al. [9] note that very few studies have compared the UE kinematics estimated by the MKO approach with other methods involving simultaneous data acquisition with fluoroscopy and motion capture [35] and bone-pin markers [33, 34]. Charbonnier et al. [35] considered an RMS error of less than 4° as good and acceptable for clinical use in the study of shoulder pathology; however, this study focused on estimating the accuracy of the MKO approach. Conversely, our study focuses on the agreement between the two models and not on a comparison against a 'gold standard.'

Additionally, the calculated differences might likely not be consistent across all subjects, joints, and tasks due to different subject anthropometric dimensions, measurement errors, STA, etc. Some subjects could have resulted in greater differences between the two models than others; however, quantifying inter- and intra-subject variations in differences was not within this study's scope. Even though the same experimental marker data, joint coordinate system, and angle sequence were used for both models, joint kinematics varied slightly between models. One possible cause of these differences in kinematics could be the choice of marker weights (i.e., how closely the virtual marker's location should match the experimental marker's location) in the MKO approach because these weightings account for different STA corresponding to each marker [50]. Finally, the performance of MKO algorithms depends on the joint constraints [18], model parameters (e.g., segment length [68]) and the markers/sensors involved [28]. Unfortunately, no optimal marker set or sensor placement exists yet [15].

4.2.2 Wilcoxon rank-sum test

Accurate measurement of UE movement during functional task execution provides an objective measure of functional outcome and is valuable information for evaluation. Statistical significance was observed for the range-of-motion angles estimated by the two models for some joint degrees of freedom and/or tasks. The hypothesis that the estimated joint kinematics would differ between models was partially confirmed. In summary, this study had significant differences for some joint range-of-motion angles (Table 2); however, the qualitative trends matched reasonably well. Results indicate that the two models with different kinematic algorithms (SKO and MKO) result in broadly consistent solutions, albeit with some key differences.

4.2.3 Reconstruction residuals

Numerous experimental studies have reported the mean residual error on the marker positions [9], and a large range of residual values was found (4–40 mm) according to the complexity of the kinematic chain model (e.g., upper-limb or full-body model), the degree of personalization as well as the movement of interest. Marker reconstruction residuals have been considered a criterion for accuracy in bone and joint kinematics, with the underlying assumption that a better kinematic chain would result in smaller marker reconstruction errors [69]. For systematically evaluating MKO models, Begon et al. [9] have recommended ensuring that the marker residuals remain within the STA range reported in the literature [23].

The peak values of reconstruction residual through the AnyBody model (MKO approach) in our study are less than 28 mm, as also found in an earlier study [64]. The large difference of $\sim$ 35 mm for the humerus segment origin may be due to the large Ab/Ad and Int/Ext angles involved during select task execution. Such high values for the humerus segment may also be due to high reconstruction residual values for the RELB marker, as seen in Supplementary Figure 1. The markers have equal weight since understanding the effect of marker weight on the tracking of segments is not within the scope of this study. However, adjusting the weight factor for fitting the markers is crucial to account for STA, and the lower arm and hand should require fitting differently in future studies. It has been reported in the literature that marker weights are subject- and movement-specific, which reinforces that STA is subject- and movement-specific [50]. Following this, the hand segment had a high reconstruction residual, which may be attributed to the unintended movement of the markers used for describing hand kinematics, i.e., RWRA, RWRB, and RFIN markers. Numerous markers and segment origins (Supplementary Figures 1 and 2, respectively) have outliers, which may be due to the assumption that the AnyBody model is scaled precisely to match each participant.

Reconstruction accuracy is vital as errors in the definition of a joint center may negatively influence calculations of joint angles and, subsequently, joint and muscle loads [70]. It has been discussed in the literature that marker residuals may represent the amount of STA but also the adequacy between the kinematic chain (i.e., number of degrees of freedom and geometrical parameters) and the skeleton [71]. Besides, as recommended by Begon et al. [33], marker residuals were in the same range as values of STA observed for the shoulder and upper limbs [72]. However, due to the current study's lack of a 'gold standard', we cannot comment further on reconstruction residuals to determine model performance and accuracy.

4.3 Agreement between the two models

4.3.1 Scatter plots and Correlation

The relation between both model outputs is linear in most cases. The results were similar to those estimated by a similar study in the lower extremity [32]. Scatter plots show a strong to excellent correlation (via Spearman's rank correlation co-efficient) between the two model outputs (Supplementary Table 4). However, correlation does not offer the necessary information to check the agreement between the model outputs, which necessitated undertaking Bland-Altman analysis.

4.3.2 Bland-Altman analysis

In a similar study for the lower extremity (one of the very few studies of its kind across upper and lower extremities), to evaluate the degree of agreement between model-derived and measured joint kinematics, Richard et al. [73] have used Bland-Altman analysis. The Bland-Altman plots are a graphical method to quantify the agreement between two measurements by studying the bias and constructing limits of agreement [54] without assuming a prior underlying relationship (e.g., linear). Furthermore, the Bland-Altman plot analysis is a simple way to evaluate a bias between the mean differences and estimate an agreement interval, within which 95% of the differences of the second method, compared to the first one, fall [59]. Inspection of the Bland-Altman plots (Supplementary Figure 4(a)–(c) and Supplementary Table 5) revealed low systematic biases (i.e., median differences), which generally have been within $\pm$ 2.5° for shoulder, 2.65° — 3.65° for elbow, and nearly $\pm$ 2° for the wrist. Negative and positive bias for the two different Bland-Altman plots, respectively, indicates systematic under- and over-estimations by the AnyBody model. Further, a lower limit of agreement on the Bland-Altman plots means a good agreement between the models and vice versa. The plots show narrower limits of agreement for shoulder Fl/Ex and wrist Fl/Ex angles than other outputs. Notably, Bland-Altman plots revealed no clear trends as positive and negative bias values as well as wider and narrower limits of agreement—across tasks, joints, and/or joint degrees of freedom—indicate that the level of agreement between the two models might be affected by how the two models deal with STA [9] as well as model complexity (e.g., presence/absence of joint constraints). STA is typically task-dependent, joint-dependent, and varies depending on the joint's degree of freedom [12]. Data can be analyzed as unit differences plot and percentage differences in Bland-Altman analysis [59]. Future studies should focus on adopting percentage difference for Bland-Altman analysis because joints like the shoulder have a much higher joint incursion than other joints.

5 Study limitations

Several limitations are acknowledged that reduce the generalizability and ready adoption of our findings. First, the UE aspect of the Plug-in Gait model is not yet clinically validated (Jacques G, personal communication, April 01, 2016). Besides, this framework does not allow the calculation of forearm pronation/supination angles [46]. Although the Plug-in Gait model is validated for clinical gait analysis, a few issues have been listed that are directly associated with the calculation methods of the Plug-in Gait model that are also relevant for the UE [45]: (i) as a direct kinematic and hierarchal method, a proximal origin's frame definition influences more distal segments [30, 43], e.g., the definition of the radius local coordinate system (distal) relies on the correct definition of the humerus (proximal) local coordinate system; (ii) incorrect placement of markers along predefined anatomical landmarks is a known source of error in the Plug-in Gait model joint kinematic outputs [30]; (iii) as with other marker-based SKO models, the Plug-in Gait model is prone to errors from STA [17, 22, 74]; and (iv) in practice, another issue has been the location of the shoulder and elbow joint centers having been generated by a proprietary formula known as the "Chord function" [37].

This study deals with ten non-disabled participants – a small sample size to form any notable conclusions. Currently, the AnyBody model is scaled linearly based on a few anthropometric dimensions and the static trial; however, scaling based on dynamic trials and functional tasks could enhance AnyBody model (MKO approach) personalization further to improve the accuracy of predictions [75]. Scaling MKO models as close as possible to functional anatomy based on high-resolution imaging data in clinical applications has been sought as they can be adapted to pathological cases, and a better STA reduction can be ex-

pected [21]. Further, a recent study quantified the effects of anatomical errors on shoulder kinematics and showed that participants' anthropometry and sex could indirectly affect kinematic outcomes [76]. For UE models, the joint centers and axes of the AnyBody and Plug-in Gait models must be defined per the *International Society of Biomechanics* recommendations for kinematics [77] to facilitate standardization and comparison with other published literature.

6 Recommended future work

The main limitation of comparative studies such as this study comes from the lack of a 'gold standard' [9]. While comparing the two models in our research, due to the absence of an independent (or a reference) 'gold standard' technique, it was not possible to strictly conclude on the most suitable model for determining kinematic variables. It has been concluded by Begon et al. [9] that the accuracy of both MKO and other approaches (i.e., SKO and non-optimized kinematics) has not been extensively validated against ground-truth bone kinematics.

The validity of UE kinematic models could be better examined [13] by simultaneously monitoring the movement of the experimental skin markers and the underlying bone during dynamic tasks using 'gold standard' techniques such as intracortical pins [33, 62], percutaneous bone tracking devices [74], or bio-imaging methods such as fluoroscopy [35], X-rays [78], magnetic resonance imaging [32], computed tomography [79], and ultrasound [80]. Nevertheless, these studies—limited in number—have been reported to typically deal with relatively small sample sizes, which can be attributed to these techniques often being cumbersome, expensive, time-consuming, and requiring highly-specific and multidisciplinary expertise (bio-engineers, orthopedics, physiotherapists, etc.) [11]. Besides, most of these experimental techniques are invasive or can cause ionizing radiation exposure. Finally, much of this equipment is rarely available in a clinical setting (e.g., [35]). Addressing this barrier, studies aimed at developing 'gold standard' datasets which include comprehensive, high-fidelity experimental data, are collected using techniques inaccessible to many researchers [13] and have been made publicly available [11].

Future work could focus on recreating upper-limb movement tasks like that listed by Cereatti et al. [11] for driving AnyBody and Plug-in Gait models. This publicly-available dataset includes motion of the skin markers measured *in vivo* and *ex vivo* using stereophotogrammetry as well as the movement of the underlying bones measured using ground-truth invasive (i.e., bone-pin) or bio-imaging techniques (i.e., fluoroscopy) for tasks such as arm elevation and functional upper-limb movements [35, 39]. As also commended earlier [9], this methodology could be used to compare better—and perhaps, blindly validate—the MKO and SKO algorithms. Such technical studies should be used as a reference in the future to determine whether the variability introduced by choice of the kinematic model impacts a study's clinical relevance [18]. While Charlton et al. [29] showed that MKO outperforms commonly used models (e.g., Plug-in Gait model with an SKO approach) in terms of intra-subject and inter-observer repeatability, further studies were recommended to confirm the clinical relevance of an MKO approach [9].

In a review article focusing on clinical gait analysis, McGinley et al. [44] have argued that in most common clinical situations, error $\leq 2^\circ$ is highly likely to be widely considered acceptable, as such errors are probably too small to require explicit consideration during data interpretation. Errors ranging between 2° – 5° are also likely to be considered reasonable but may require care in data interpretation. Additionally, McGinley et al. [44] suggest that errors $> 5^\circ$ should raise concern and may be large enough to mislead clinical interpretation. However, some of these key differences between the two models in our study deter readily using the models in a clinical context. Hence, suitable model validation against a 'gold standard' is recommended, and care should be taken when comparing our results with other UE models at this stage.

Finally, quantifying disparities between the two models on a kinetic level is warranted as changes in kinematics could affect the estimation of muscle activities, as shown earlier [7, 8]. Additionally, there are numerous quantifiable sources of inaccuracy in the input variables for inverse dynamics calculations (e.g., errors in body segment parameter estimates, joint center of rotation locations, motion capture system measurements, force plate measurements (if applicable), and segment angle calculations due to STA) [6].

7 Conclusions

This study reports a rigorous comparison of joint kinematics estimated by a scaled cadaver-based MSK model implementation in the AnyBody Managed Model Repository (i.e., AnyBody model with the MKO approach) with the widely-used Plug-in Gait model (i.e., SKO approach). However, neither represents a 'gold standard.' Moreover, few studies have systematically analyzed the independent impact of computational methods on estimated UE joint kinematics to the authors' knowledge. Therefore, this study facilitates evaluating the difference in joint kinematics that can be solely attributed to the different kinematic methods under consideration (MKO versus SKO) by ensuring equivalence of the marker set, joint coordinate system, sequence of joint angle outputs, etc. Consequently, this study creates new research avenues for collaboration and data reuse, especially whilst comparing MKO and SKO model-estimated kinematic outputs.

Relatively, the largest differences between the two models were found in the shoulder, followed by the elbow and wrist. Differences were lower for tasks involving smaller joint angles, such as the Reach-to-Right (RR) task involving using the dominant hand by right-handed participants for task execution. The RMS differences were within the acceptable level of 10°; although this limit might be unacceptable for clinical use. It was observed that the relation between the two outputs is near-linear, with strong-to-excellent correlation co-efficients. Joint kinematics broadly match qualitatively and quantitatively, albeit with few key differences as noticed in the joint range-of-motion angles, as well as bias and wider limits of agreement on the Bland-Altman plots.

Acknowledgements

This work was supported by the *Research Councils UK (RCUK) Digital Economy Programme grant num-*

ber EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation). Additional funding through a *Wellcome Trust Affordable Healthcare in India Award 103383/B/13/Z* is gratefully acknowledged. The research team acknowledges the support of the National Institute for Health Research Clinical Research Network (NIHR CRN).

References

[1] Cappozzo, A., Della Croce, U., Leardini, A., and Chiari, L., 2005. "Human movement analysis using stereophotogrammetry: Part 1: Theoretical background". *Gait & Posture,* **21**(2), pp. 186–196.

[2] Cappozzo, A., Catani, F., Leardini, A., Benedetti, M., and Della Croce, U., 1996. "Position and orientation in space of bones during movement: Experimental artefacts". *Clinical Biomechanics,* **11**(2), pp. 90–100.

[3] Erdemir, A., McLean, S., Herzog, W., and van den Bogert, A. J., 2007. "Model-based estimation of muscle forces exerted during movements". *Clinical Biomechanics,* **22**(2), pp. 131–154.

[4] Damsgaard, M., Rasmussen, J., Christensen, S. T., Surma, E., and De Zee, M., 2006. "Analysis of musculoskeletal systems in the AnyBody Modeling System". *Simulation Modelling Practice and Theory,* **14**(8), pp. 1100–1111.

[5] Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E., and Thelen, D. G., 2007. "OpenSim: Open-source software to create and analyze dynamic simulations of movement". *IEEE Transactions on Biomedical Engineering,* **54**(11), pp. 1940–1950.

[6] Riemer, R., Hsiao-Wecksler, E. T., and Zhang, X., 2008. "Uncertainties in inverse dynamics solutions: A comprehensive analysis and an application to gait". *Gait & Posture,* **27**(4), pp. 578–588.

[7] Roelker, S. A., Caruthers, E. J., Baker, R. K., Pelz, N. C., Chaudhari, A. M., and Siston, R. A., 2017. "Interpreting musculoskeletal models and dynamic simulations: Causes and effects of differences between models". *Annals of Biomedical Engineering,* **45**(11), pp. 2635–2647.

[8] Myers, C. A., Laz, P. J., Shelburne, K. B., and Davidson, B. S., 2015. "A probabilistic approach to quantify the impact of uncertainty propagation in musculoskeletal simulations". *Annals of Biomedical Engineering,* **43**(5), pp. 1098–1111.

[9] Begon, M., Andersen, M. S., and Dumas, R., 2018. "Multibody Kinematics Optimization for the estimation of upper and lower limb human joint kinematics: A systematized methodological review". *Journal of Biomechanical Engineering,* **140**(3), p. 030801.

[10] Mantovani, G., and Lamontagne, M., 2017. "How different marker sets affect joint angles in inverse kinematics framework". *Journal of Biomechanical Engineering,* **139**(4).

[11] Cereatti, A., Bonci, T., Akbarshahi, M., Aminian, K., Barré, A., Begon, M., Benoit, D. L., Charbonnier, C., Dal Maso, F., Fantozzi, S., et al., 2017. "Standardization proposal of soft tissue artefact description for data sharing in human motion measurements". *Journal of Biomechanics,* **62**, pp. 5–13.

[12] Camomilla, V., Dumas, R., and Cappozzo, A., 2017. "Human movement analysis: The soft tissue artefact issue". *Journal of Biomechanics,* **62**, pp. 1–4.

[13] Hicks, J. L., Uchida, T. K., Seth, A., Rajagopal, A., and Delp, S. L., 2015. "Is my model good enough? Best practices for verification and validation of musculoskeletal models and simulations of movement". *Journal of Biomechanical Engineering,* **137**(2).

[14] Rau, G., Disselhorst-Klug, C., and Schmidt, R., 2000. "Movement biomechanics goes upwards: From the leg to the arm". *Journal of Biomechanics,* **33**(10), pp. 1207–1216.

[15] Anglin, C., and Wyss, U. P., 2000. "Arm motion and load analysis of sit-to-stand, stand-to-sit, cane walking and lifting". *Clinical Biomechanics,* **15**(6), pp. 441–448.

[16] Lathrop, R. L., Chaudhari, A. M., and Siston, R. A., 2011. "Comparative assessment of bone pose estimation using point cluster technique and OpenSim". *Journal of Biomechanical Engineering,* **133**(11), pp. 1–6.

[17] Lu, T.-W., and O'connor, J., 1999. "Bone position estimation from skin marker co-ordinates using global optimisation with joint constraints". *Journal of Biomechanics,* **32**(2), pp. 129–134.

[18] Duprey, S., Cheze, L., and Dumas, R., 2010. "Influence of joint constraints on lower limb kinematics estimation from skin markers using global optimization". *Journal of Biomechanics,* **43**(14), pp. 2858–2862.

[19] Andersen, M. S., Damsgaard, M., and Rasmussen, J., 2009. "Kinematic analysis of over-determinate biomechanical systems". *Computer Methods in Biomechanics and Biomedical Engineering,* **12**(4), pp. 371–384.

[20] Roux, E., Bouilland, S., Godillon-Maquinghen, A.-P., and Bouttens, D., 2002. "Evaluation of the global optimisation method within the upper limb kinematics analysis". *Journal of Biomechanics,* **35**(9), pp. 1279–1283.

[21] Duprey, S., Naaim, A., Moissenet, F., Begon, M., and Chèze, L., 2017. "Kinematic models of the upper limb joints for Multibody Kinematics Optimisation: An overview". *Journal of Biomechanics,* **62**, pp. 87–94.

[22] Leardini, A., Chiari, L., Della Croce, U., and Cappozzo, A., 2005. "Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation". *Gait & Posture,* **21**(2), pp. 212–225.

[23] Peters, A., Galna, B., Sangeux, M., Morris, M., and Baker, R., 2010. "Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review". *Gait & Posture,* **31**(1), pp. 1–8.

[24] Riley, P. O., Franz, J., Dicharry, J., and Kerrigan, D. C., 2010. "Changes in hip joint muscle–tendon lengths with mode of locomotion". *Gait & Posture,* **31**(2), pp. 279–283.

[25] Arnold, A. S., Salinas, S., Hakawa, D. J., and Delp, S. L., 2000. "Accuracy of muscle moment arms estimated from MRI-based musculoskeletal models of the lower extremity". *Computer Aided Surgery,* **5**(2), pp. 108–119.

[26] Morgan, K. D., Donnelly, C. J., and Reinbolt, J. A.,

2014. “Elevated gastrocnemius forces compensate for decreased hamstrings forces during the weight-acceptance phase of single-leg jump landing: Implications for anterior cruciate ligament injury risk”. *Journal of Biomechanics,* **47**(13), pp. 3295–3302.

[27] Modenese, L., Gopalakrishnan, A., and Phillips, A., 2013. “Application of a falsification strategy to a musculoskeletal model of the lower limb and accuracy of the predicted hip contact force vector”. *Journal of Biomechanics,* **46**(6), pp. 1193–1200.

[28] Begon, M., Wieber, P.-B., and Yeadon, M. R., 2008. “Kinematics estimation of straddled movements on high bar from a limited number of skin markers using a chain model”. *Journal of Biomechanics,* **41**(3), pp. 581–586.

[29] Charlton, I. W., Tate, P., Smyth, P., and Roren, L., 2004. “Repeatability of an optimised lower body model”. *Gait & Posture,* **20**(2), pp. 213–221.

[30] Groen, B. E., Geurts, M., Nienhuis, B., and Duysens, J., 2012. “Sensitivity of the OLGA and VCM models to erroneous marker placement: Effects on 3d-gait kinematics”. *Gait & Posture,* **35**(3), pp. 517–521.

[31] Lund, M. E., de Zee, M., Andersen, M. S., and Rasmussen, J., 2012. “On validation of multibody musculoskeletal models”. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine,* **226**(2), pp. 82–94.

[32] Kainz, H., Modenese, L., Lloyd, D., Maine, S., Walsh, H., and Carty, C., 2016. “Joint kinematic calculation based on clinical direct kinematic versus inverse kinematic gait models”. *Journal of Biomechanics,* **49**(9), pp. 1658–1669.

[33] Begon, M., Bélaise, C., Naaim, A., Lundberg, A., and Chèze, L., 2017. “Multibody Kinematics Optimization with marker projection improves the accuracy of the humerus rotational kinematics”. *Journal of Biomechanics,* **62**, pp. 117–123.

[34] Seth, A., Matias, R., Veloso, A. P., and Delp, S. L., 2016. “A biomechanical model of the scapulothoracic joint to accurately capture scapular kinematics during shoulder movements”. *PLoS ONE,* **11**(1), p. e0141028.

[35] Charbonnier, C., Chague, S., Kolo, F., Chow, J., and Lädermann, A., 2014. “A patient-specific measurement technique to model shoulder joint kinematics”. *Orthopaedics & Traumatology: Surgery & Research,* **100**(7), pp. 715–719.

[36] Nagaraja, V. H., Bergmann, J. H., Andersen, M. S., and Thompson, M. S., 2018. “Compensatory movements involved during simulated upper limb prosthetic usage: Reach Task vs. Reach-to-Grasp Task”. In XV ISB International Symposium on 3-D Analysis of Human Movement, ISB.

[37] Vicon Motion Systems | Plug-in Gait Reference Guide – Vicon Documentation. `https://docs.vicon.com/display/Nexus212/Plug-in+Gait+Reference+Guide`. (Accessed on 12/04/2022).

[38] AnyBody tutorials. `https://anyscript.org/tutorials/`. (Accessed on 11/20/2021).

[39] Vicon Nexus | Software for Motion Capture in Life Sciences | Vicon. `https://www.vicon.com/software/nexus/?section=downloads`. (Accessed on 12/04/2022).

[40] Pinzone, O., Schwartz, M. H., Thomason, P., and Baker, R., 2014. “The comparison of normative reference data from different gait analysis services”. *Gait & Posture,* **40**(2), pp. 286–290.

[41] Schwartz, M. H., Rozumalski, A., and Trost, J. P., 2008. “The effect of walking speed on the gait of typically developing children”. *Journal of Biomechanics,* **41**(8), pp. 1639–1650.

[42] Davis III, R. B., Ounpuu, S., Tyburski, D., and Gage, J. R., 1991. “A gait analysis data collection and reduction technique”. *Human Movement Science,* **10**(5), pp. 575–587.

[43] Kadaba, M., Ramakrishnan, H., and Wootten, M., 1990. “Measurement of lower extremity kinematics during level walking”. *Journal of Orthopaedic Research,* **8**(3), pp. 383–392.

[44] McGinley, J. L., Baker, R., Wolfe, R., and Morris, M. E., 2009. “The reliability of three-dimensional kinematic gait measurements: A systematic review”. *Gait & Posture,* **29**(3), pp. 360–369.

[45] Schwartz, M., and Dixon, P. C., 2018. “The effect of subject measurement error on joint kinematics in the conventional gait model: Insights from the open-source pyCGM tool using high performance computing methods”. *PLoS ONE,* **13**(1), p. e0189984.

[46] Angle Definitions – Nexus 2.5 Documentation – Vicon Documentation. `https://docs.vicon.com/display/Nexus25/Angle+definitions`. (Accessed on 12/04/2022).

[47] Van der Helm, F. C., 1994. “A finite element musculoskeletal model of the shoulder mechanism”. *Journal of Biomechanics,* **27**(5), pp. 551–569.

[48] De Zee, M., Hansen, L., Wong, C., Rasmussen, J., and Simonsen, E. B., 2007. “A generic detailed rigid-body lumbar spine model”. *Journal of Biomechanics,* **40**(6), pp. 1219–1227.

[49] Andersen, M. S., Damsgaard, M., MacWilliams, B., and Rasmussen, J., 2010. “A computationally efficient optimisation-based method for parameter identification of kinematically determinate and over-determinate biomechanical systems”. *Computer Methods in Biomechanics and Biomedical Engineering,* **13**(2), pp. 171–183.

[50] Begon, M., Dal Maso, F., Arndt, A., and Monnet, T., 2015. “Can optimal marker weightings improve thoracohumeral kinematics accuracy?”. *Journal of Biomechanics,* **48**(10), pp. 2019–2025.

[51] Barre, A., and Armand, S., 2014. “Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data”. *Computer Methods and Programs in Biomedicine,* **114**(1), pp. 80–87.

[52] Jarque, C. M., and Bera, A. K., 1980. “Efficient tests for normality, homoscedasticity and serial independence of regression residuals”. *Economics Letters,* **6**(3), pp. 255–259.

[53] Wilcoxon, F., 1992. “Individual comparisons by ranking methods”. In *Breakthroughs in Statistics*. Springer, pp. 196–202.

[54] Bland, J. M., and Altman, D. G., 1986. “Statistical methods for assessing agreement between two methods of clinical measurement”. *The Lancet,* **327**(8476), pp. 307–310.

[55] McLaughlin, P., 2013. "Testing agreement between a new method and the gold standard—how do we test?". *Journal of Biomechanics,* **46**(16), pp. 2757–2760.

[56] Robinson, M. A., Donnelly, C. J., Tsao, J., and Vanrenterghem, J., 2013. "Impact of knee modeling approach on indicators and classification of ACL injury risk". *Medicine & Science in Sports & Exercise,* **46**, pp. 1269–1276.

[57] Taylor, R., 1990. "Interpretation of the correlation coefficient: A basic review". *Journal of Diagnostic Medical Sonography,* **6**(1), pp. 35–39.

[58] Piazza, S. J., and Cavanagh, P. R., 2000. "Measurement of the screw-home motion of the knee is sensitive to errors in axis alignment". *Journal of Biomechanics,* **33**(8), pp. 1029–1034.

[59] Giavarina, D., 2015. "Understanding Bland Altman analysis". *Biochemia Medica,* **25**(2), pp. 141–151.

[60] Cereatti, A., Della Croce, U., and Cappozzo, A., 2006. "Reconstruction of skeletal movement using skin markers: Comparative assessment of bone pose estimators". *Journal of NeuroEngineering and Rehabilitation,* **3**(1), pp. 1–12.

[61] Stagni, R., Fantozzi, S., and Cappello, A., 2009. "Double calibration vs. global optimisation: Performance and effectiveness for clinical application". *Gait & posture,* **29**(1), pp. 119–122.

[62] Andersen, M. S., Benoit, D. L., Damsgaard, M., Ramsey, D. K., and Rasmussen, J., 2010. "Do kinematic models reduce the effects of soft tissue artefacts in skin marker-based motion analysis? An in vivo study of knee kinematics". *Journal of Biomechanics,* **43**(2), pp. 268–273.

[63] Magermans, D., Chadwick, E., Veeger, H., and Van Der Helm, F., 2005. "Requirements for upper extremity motions during activities of daily living". *Clinical Biomechanics,* **20**(6), pp. 591–599.

[64] Laitenberger, M., Raison, M., Périé, D., and Begon, M., 2015. "Refinement of the upper limb joint kinematics and dynamics using a subject-specific closed-loop forearm model". *Multibody System Dynamics,* **33**(4), pp. 413–438.

[65] Chin, A., Lloyd, D., Alderson, J., Elliott, B., and Mills, P., 2010. "A marker-based mean finite helical axis model to determine elbow rotation axes and kinematics in vivo". *Journal of Applied Biomechanics,* **26**(3), pp. 305–315.

[66] Chiari, L., Della Croce, U., Leardini, A., and Cappozzo, A., 2005. "Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors". *Gait & Posture,* **21**(2), pp. 197–211.

[67] Della Croce, U., Leardini, A., Chiari, L., and Cappozzo, A., 2005. "Human movement analysis using stereophotogrammetry: Part 4: Assessment of anatomical landmark misplacement and its effects on joint kinematics". *Gait & Posture,* **21**(2), pp. 226–237.

[68] El Habachi, A., Duprey, S., Cheze, L., and Dumas, R., 2013. "Global sensitivity analysis of the kinematics obtained with a multi-body optimisation using a parallel mechanism of the shoulder". *Computer Methods in Biomechanics and Biomedical Engineering,* **16**(sup1), pp. 61–62.

[69] Puchaud, P., Hybois, S., Lombart, A., Bascou, J., Pillet, H., Fodé, P., and Sauret, C., 2019. "On the influence of the shoulder kinematic chain on joint kinematics and musculotendon lengths during wheelchair propulsion estimated from multibody kinematics optimization". *Journal of Biomechanical Engineering,* **141**(10).

[70] Stagni, R., Leardini, A., Cappozzo, A., Benedetti, M. G., and Cappello, A., 2000. "Effects of hip joint centre mislocation on gait analysis results". *Journal of Biomechanics,* **33**(11), pp. 1479–1487.

[71] Ojeda, J., Martínez-Reina, J., and Mayo, J., 2016. "The effect of kinematic constraints in the inverse dynamics problem in biomechanics". *Multibody System Dynamics,* **37**(3), pp. 291–309.

[72] Matsui, K., Shimada, K., and Andrew, P. D., 2006. "Deviation of skin marker from bone target during movement of the scapula". *Journal of Orthopaedic Science,* **11**(2), pp. 180–184.

[73] Richard, V., Lamberto, G., Lu, T.-W., Cappozzo, A., and Dumas, R., 2016. "Knee kinematics estimation using multi-body optimisation embedding a knee joint stiffness matrix: A feasibility study". *PLoS ONE,* **11**(6), p. e0157010.

[74] Holden, J. P., Orsini, J. A., Siegel, K. L., Kepple, T. M., Gerber, L. H., and Stanhope, S. J., 1997. "Surface movement errors in shank kinematics and knee kinetics during gait". *Gait & Posture,* **5**(3), pp. 217–227.

[75] Lund, M. E., Andersen, M. S., de Zee, M., and Rasmussen, J., 2015. "Scaling of musculoskeletal models from static and dynamic trials". *International Biomechanics,* **2**(1), pp. 1–11.

[76] Lavaill, M., Martelli, S., Gilliland, L., Gupta, A., Kerr, G., and Pivonka, P., 2022. "The effects of anatomical errors on shoulder kinematics computed using multi-body models". *Biomechanics and Modeling in Mechanobiology*, pp. 1–12.

[77] Wu, G., Van der Helm, F. C., Veeger, H. D., Makhsous, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A. R., McQuade, K., Wang, X., et al., 2005. "Isb recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part ii: Shoulder, elbow, wrist and hand". *Journal of Biomechanics,* **38**(5), pp. 981–992.

[78] Dumas, R., and Duprey, S., 2022. "Subject-specific model-derived kinematics of the shoulder based on skin markers during arm abduction up to 180°-assessment of 4 gleno-humeral joint models". *Journal of Biomechanics*, p. 111061.

[79] Marra, M. A., Vanheule, V., Fluit, R., Koopman, B. H., Rasmussen, J., Verdonschot, N., and Andersen, M. S., 2015. "A subject-specific musculoskeletal modeling framework to predict in vivo mechanics of total knee arthroplasty". *Journal of Biomechanical Engineering,* **137**(2), p. 020904.

[80] Jia, R., Mellon, S., Monk, P., Murray, D., and Noble, J. A., 2016. "A computer-aided tracking and motion analysis with ultrasound (CAT & MAUS) system for the description of hip joint kinematics". *International Journal of Computer Assisted Radiology and Surgery,* **11**(11), pp. 1965–1977.

Supplementary Information: Comparison of a Scaled Cadaver-based Musculoskeletal Model with a Clinical Upper Extremity Model

Vikranth H. Nagaraja*, Jeroen H.M. Bergmann, Michael S. Andersen, and Mark S. Thompson

October 12, 2022

2 Methods

2.6.3 Plug-in Gait Modeling (SKO approach)

Angles	Goniometric	Order	Positive Rotation	Axis	Direction
RThoraxAngles	Absolute	1	Backward Tilt	Prg.Fm. Y	Clockwise
		2	Left Tilt	Prg.Fm. X'	Clockwise
		3	Left Rotation	Prg.Fm. Z''	Anti-clockwise
RShoulderAngles	Relative	1	Flexion	Thorax Y	Anti-clockwise
		2	Abduction	Thorax X'	Clockwise
		3	Internal Rotation	Thorax Z''	Clockwise
RElbowAngles	Relative	1	Flexion	Humeral Y	Clockwise
		2	-	Humeral X'	-
		3	-	Humeral Z''	-
RWristAngles	Relative	1	Ulnar Deviation	Radius X	Anti-clockwise
		2	Extension	Radius Y'	Clockwise
		3	Internal Rotation	Radius Z''	Anti-clockwise

Table 1: Plug-in Gait model joint angle output sequence; Adapted with permission from Vicon Plug-in Gait Reference Guide.

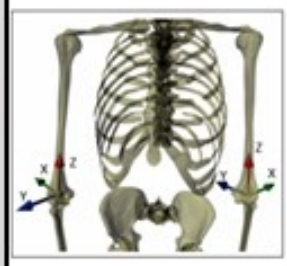
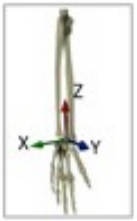
Technical Segment Frames for the Plug-in Gait® model			
Segment	**Aspect**	**Description**	**Illustration**
Thorax	Origin	CLAV marker with an offset of half a marker diameter backwards along the Second defining axis	
	First defining axis	Mid-point of the STRN and T10 to the mid-point of CLAV and C7 surface markers	
	Second defining axis	Mid-point of C7 and T10 to the mid-point of CLAV and STRN surface markers	
	Third defining axis	Mutually perpendicular to the other two axes	
Humerus	Origin	Elbow Joint Center	
	First defining axis	Elbow Joint Center to Shoulder Joint Center	
	Second defining axis	A secondary line between the Elbow Joint Center and the Wrist Joint Center is described and a cross product between this line and the Z-axis of the Humerus is calculated to define the Y-axis of the Humerus.	
	Third defining axis	Mutually perpendicular to the other two axes	
Radius	Origin	Wrist Joint Center	
	First defining axis	Elbow Joint Center to Wrist Joint Center	
	Second defining axis	Y-axis of the Humerus segment	
	Third defining axis	Mutually perpendicular to the other two axes	
Hand	Origin	Defined with the chord function using the Wrist Joint Center, RFIN surface marker, and Hand Offset value	
	First defining axis	Wrist Joint Center to Hand origin	
	Second defining axis	Direction of the line joining the wrist markers	
	Third defining axis	Mutually perpendicular to the other two axes	

Table 2: Technical segment frames for Plug-in Gait model; Segment images adapted with permission from Vicon Plug-in Gait Reference Guide. **Note**: All segment axis-systems are *Right-handed*; Green → X-axis; Blue → Y-axis; Red → Z-axis.

3 Results

3.2 Quantitative comparison

3.2.1 Difference analysis

Difference Type (Between AnyBody & Plug-in Gait model outputs)	Joint – Degree of Freedom	Reach-to-Front (RF) task Median (IQR) Degrees	Reach-to-Right (RR) task Median (IQR) Degrees	Reach-to-Left (RL) task Median (IQR) Degrees
Absolute maximum difference	Shoulder – Fl/Ex	7.2 (1.9)	4.0 (3.7)	5.9 (2.4)
	Shoulder – Ab/Ad	**14.6 (19.2)**	4.4 (15.9)	**20.1 (25.4)**
	Shoulder – Int/Ext	7.0 (10.3)	1.4 (2.2)	1.6 (1.9)
	Elbow – Fl/Ex	6.7 (4.4)	6.5 (3.3)	7.0 (3.4)
	Wrist – Fl/Ex	4.4 (3.0)	3.6 (6.1)	6.4 (3.5)
	Wrist – Rd/Ud	7.2 (4.7)	6.7 (4.0)	6.8 (9.6)
Maximum difference	Shoulder – Fl/Ex	3.9 (3.7)	4.0 (3.7)	5.9 (2.4)
	Shoulder – Ab/Ad	**14.6 (19.2)**	3.6 (19.3)	**20.1 (25.5)**
	Shoulder – Int/Ext	1.3 (6.9)	0.8 (3.0)	1.6 (2.1)
	Elbow – Fl/Ex	6.7 (4.5)	6.5 (3.3)	6.4 (3.8)
	Wrist – Fl/Ex	4.3 (4.1)	2.9 (1.5)	6.0 (3.0)
	Wrist – Rd/Ud	3.3 (4.1)	1.6 (5.9)	2.5 (4.2)
Minimum difference	Shoulder – Fl/Ex	-2.0 (4.5)	-1.3 (3.1)	-0.8 (2.8)
	Shoulder – Ab/Ad	-4.2 (5)	-3.4 (4.1)	-3.8 (2.9)
	Shoulder – Int/Ext	-8.3 (11.2)	-3.8 (6.9)	**-11.9 (18.9)**
	Elbow – Fl/Ex	-0.9 (1.5)	1.8 (3.6)	-2.4 (1.9)
	Wrist – Fl/Ex	-1.9 (5.4)	-3.2 (8.6)	-1.7 (4.8)
	Wrist – Rd/Ud	-6.8 (5.7)	-5.2 (4.3)	-6.8 (10.2)
Root-mean-squared-difference (RMSD)	Shoulder – Fl/Ex	2.7 (0.8)	2.9 (2.4)	3.3 (1.4)
	Shoulder – Ab/Ad	6.4 (5.7)	5.3 (5.0)	8.6 (8.3)
	Shoulder – Int/Ext	5.8 (4.9)	2.3 (2.3)	5.5 (8.4)
	Elbow – Fl/Ex	3.6 (2.7)	3.9 (2.1)	3.8 (2.1)
	Wrist – Fl/Ex	2.9 (1.2)	2.3 (2.4)	3.5 (0.9)
	Wrist – Rd/Ud	4.2 (2.9)	3.5 (2.0)	3.7 (4.1)
Mean absolute difference	Shoulder – Fl/Ex	2.3 (0.7)	2.6 (1.9)	3.1 (1.5)
	Shoulder – Ab/Ad	4.9 (3.2)	4.2 (3.2)	5.5 (4.2)
	Shoulder – Int/Ext	5.5 (3.9)	2.2 (2.1)	3.9 (5.0)
	Elbow – Fl/Ex	2.9 (2.4)	3.3 (2.6)	3.3 (1.8)
	Wrist – Fl/Ex	2.3 (1.4)	2.0 (1.3)	2.8 (1.0)
	Wrist – Rd/Ud	3.7 (2.3)	3.2 (1.6)	3.3 (2.6)
Difference range	Shoulder – Fl/Ex	6.4 (1.6)	6.4 (2.3)	7.6 (3.1)
	Shoulder – Ab/Ad	**19.3 (17.6)**	5 (14.9)	**23.8 (23.4)**
	Shoulder – Int/Ext	9.5 (10.7)	4.3 (6.6)	**14.8 (16.3)**
	Elbow – Fl/Ex	8.7 (4.4)	7.9 (3.8)	9.4 (3.8)
	Wrist – Fl/Ex	6.5 (4.1)	5.2 (3.3)	7.7 (6.3)
	Wrist – Rd/Ud	9.9 (6.5)	8.1 (4.1)	8.8 (5.8)

Note: Flexion (Fl), Abduction (Ab), Radial wrist deviation (Rd), and Internal rotation (Int) are *positive (+ve)*; Extension (Ex), Adduction (Ad), Ulnar wrist deviation (Ud), and External rotation (Ext) are *negative (-ve)*.

Table 3: Differences between AnyBody and Plug-in Gait model-estimated kinematic outputs in degrees (three trials & n = 10). **Note**: Values $> 10°$ are in **Bold**.

3.2.3 Reconstruction residual

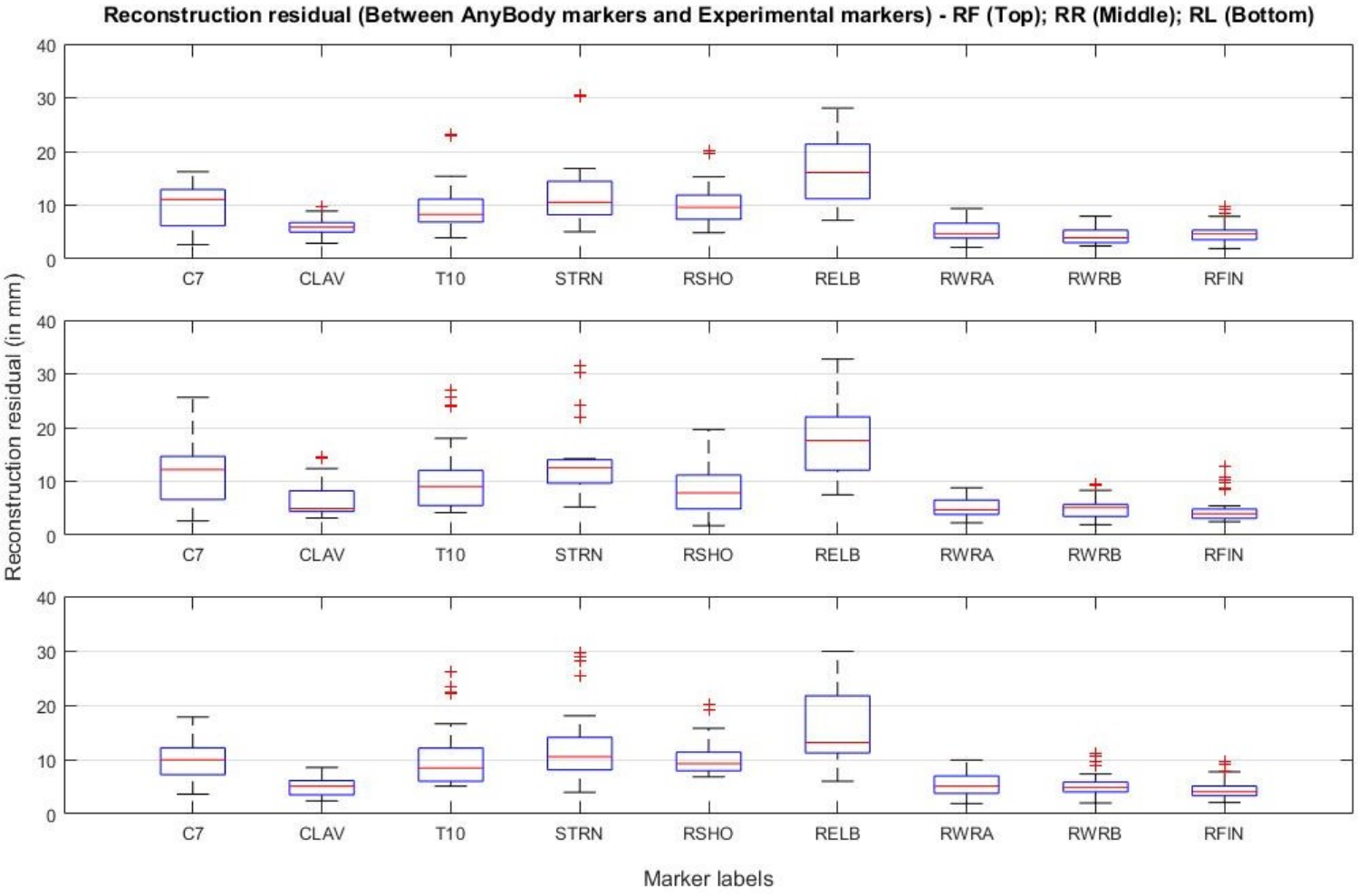


Figure 1: Box-plots for reconstruction residual (in mm) between AnyBody markers and Experimental markers (three trials & n = 10). **Note**: On each box, central red mark indicates the median, and the bottom and top edges of the box indicate the 25^{th} and 75^{th} percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol in red.

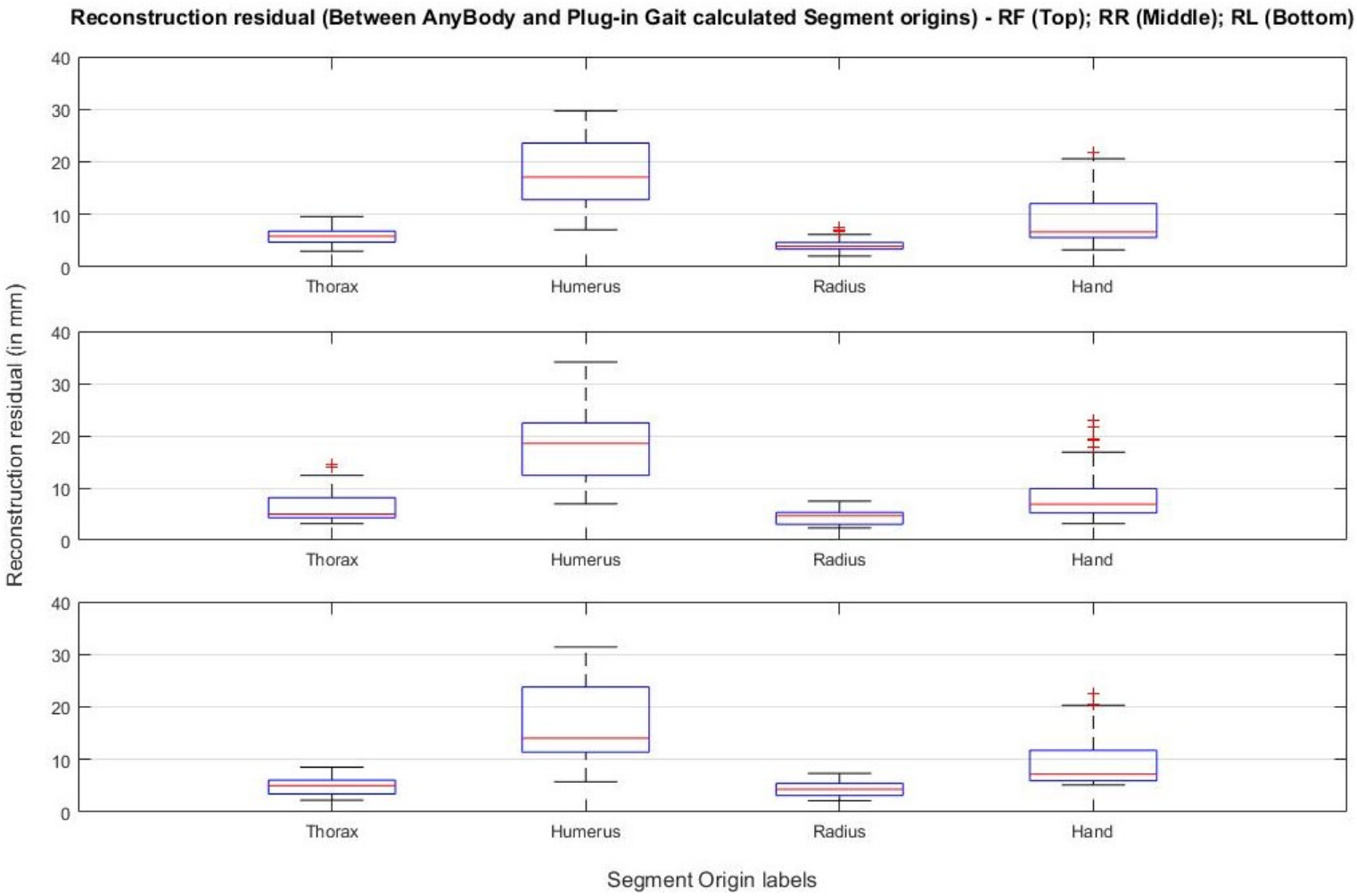


Figure 2: Box-plots for reconstruction residual (in mm) between AnyBody and Plug-in Gait model segment origins (three trials & n = 10); **Note**: On each box, central red mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol in red.

3.3 Agreement between the two models

3.3.1 Scatter plots and Correlation

Joint – Degree of Freedom	Spearman's rank correlation co-efficient (ρ)		
	Reach-to-Front (RF) task	*Reach-to-Right (RR) task*	*Reach-to-Left (RL) task*
Shoulder – Fl/Ex	0.996	0.98	0.997
Shoulder – Ab/Ad	0.79	0.91	0.75
Shoulder – Int/Ext	0.93	0.98	0.96
Elbow – Fl/Ex	0.97	0.96	0.96
Wrist – Fl/Ex	0.89	0.8995	0.89
Wrist – Rd/Ud	0.95	0.86	0.92

Legend for classifying correlation coefficients (ρ)	
Weak	$\rho \leq 0.35$
Moderate	$0.35 < \rho \leq 0.67$
Strong	$0.67 < \rho \leq 0.90$
Excellent	$\rho > 0.90$

Table 4: Classification of Spearman's rank correlation co-efficients (ρ) corresponding to scatter plots (Taylor 1990)

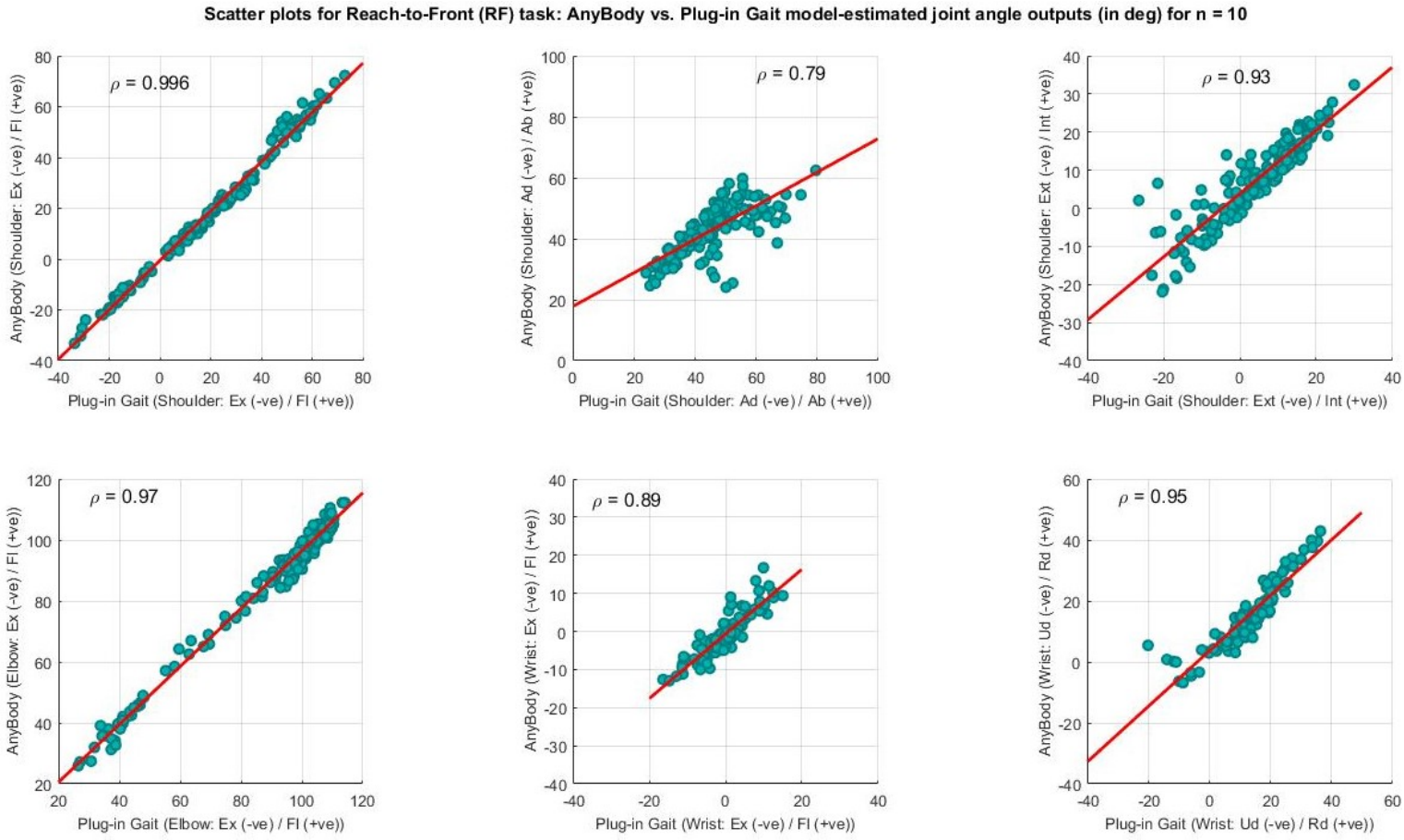


Figure 3: (a) Scatter plots between AnyBody and Plug-in Gait model-estimated joint angle outputs (in degrees) for three trials of *Reach-to-Front (RF)* task (three trials & n = 10). **Note**: Spearman's rank correlation coefficient (ρ) corresponding to each scatter plot are reported. Also, the red line is a least-squares line on each scatter plot.

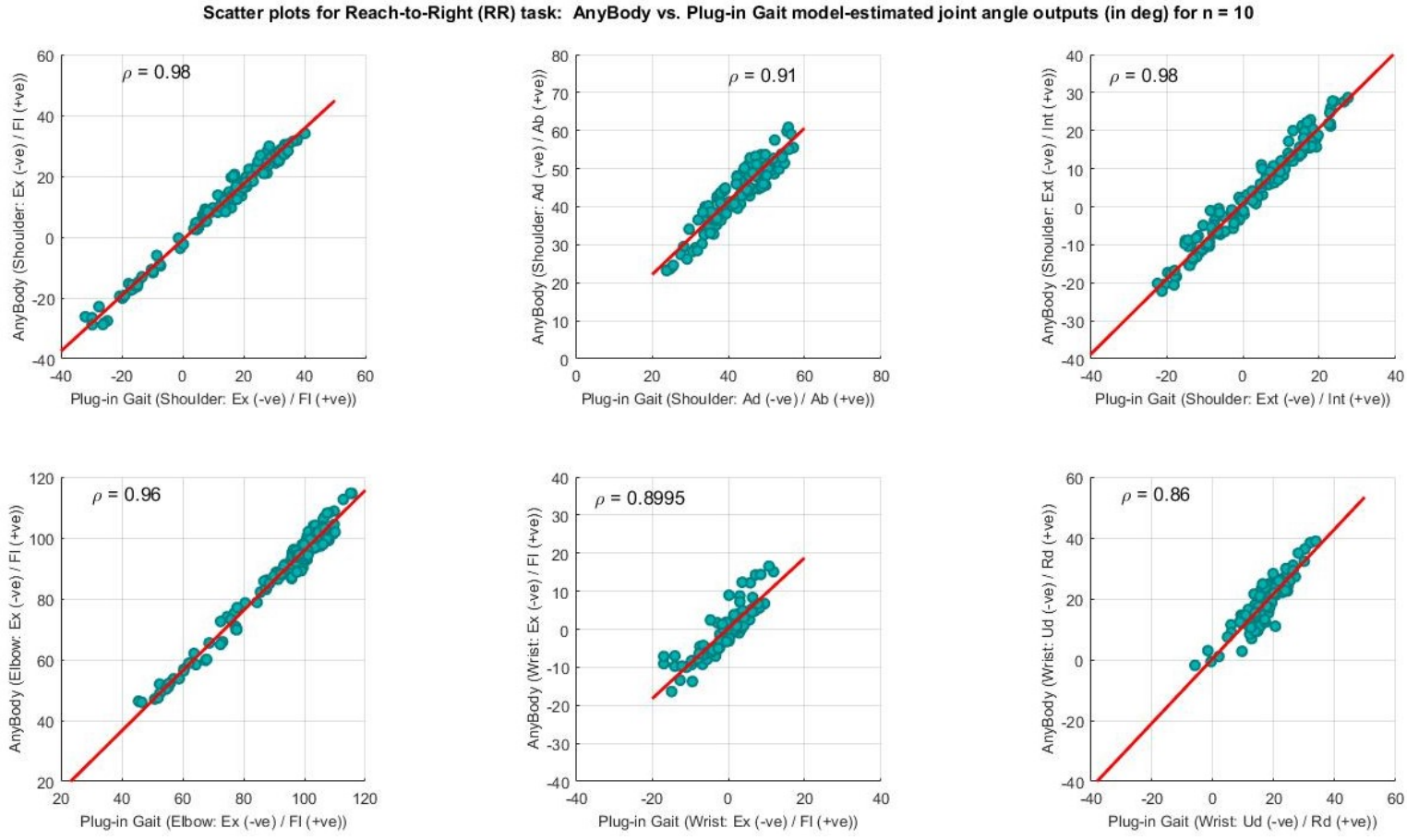


Figure 3: (b) Scatter plots between AnyBody and Plug-in Gait model-estimated joint angle outputs (in degrees) for three trials of *Reach-to-Right (RR)* task (three trials & n = 10). **Note**: Spearman's rank correlation coefficient (ρ) corresponding to each scatter plot are reported. Also, the red line is a least-squares line on each scatter plot.

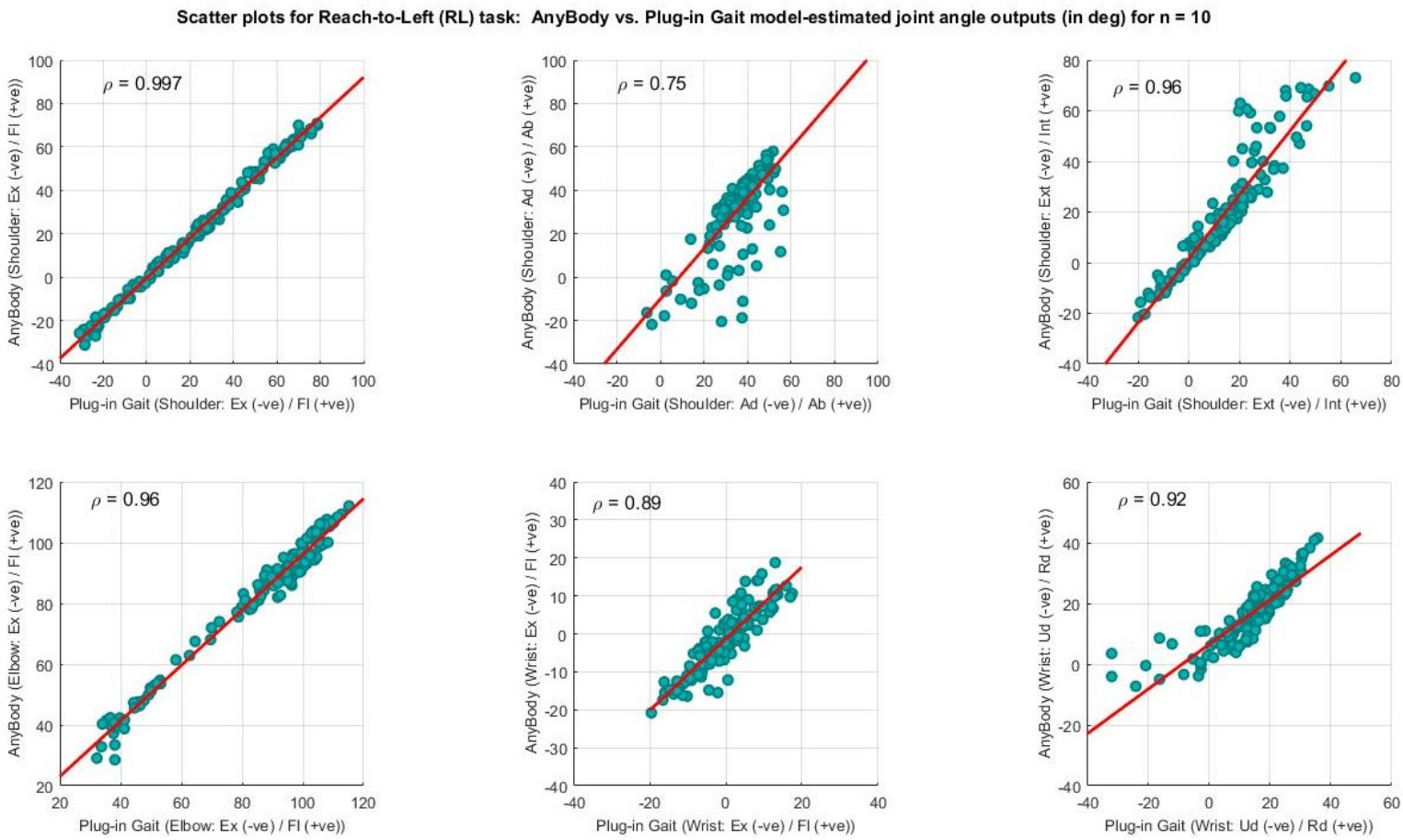


Figure 3: (c) Scatter plots between AnyBody and Plug-in Gait model-estimated joint angle outputs (in degrees) for three trials of *Reach-to-Left (RL)* task (three trials & n = 10). **Note**: Spearman's rank correlation coefficient (ρ) corresponding to each scatter plot are reported. Also, the red line is a least-squares line on each scatter plot.

3.3.2 Bland-Altman analysis

Joint – Degree of Freedom	Bland-Altman test summary					
	Reach-to-Front (RF) task		*Reach-to-Right (RR) task*		*Reach-to-Left (RL) task*	
	Systematic bias (Median difference between outputs)	95% Limits of Agreement (LoA)	Systematic bias (Median difference between outputs)	95% Limits of Agreement (LoA)	Systematic bias (Median difference between outputs)	95% Limits of Agreement (LoA)
Shoulder – Fl/Ex	1.42	-4.00 to 6.03	1.65	-3.53 to 6.24	2.48	-4.34 to 8.60
Shoulder – Ab/Ad	0.64	-5.54 to 21.19	-0.02	-5.86 to 3.59	0.04	-6.19 to 43.59
Shoulder – Int/Ext	-1.53	-15.08 to 2.11	0.09	-6.23 to 2.73	-1.41	-34.79 to 2.70
Elbow – Fl/Ex	2.65	-1.59 to 9.22	3.65	-0.69 to 8.85	2.84	-3.24 to 9.59
Wrist – Fl/Ex	1.21	-6.62 to 11.05	0.71	-8.71 to 3.71	1.92	-6.63 to 8.16
Wrist – Rd/Ud	-1.60	-14.77 to 7.56	-1.77	-7.87 to 5.97	-2.09	-20.32 to 4.25

Table 5: Bland-Altman test summary (Bland & Altman 1986)

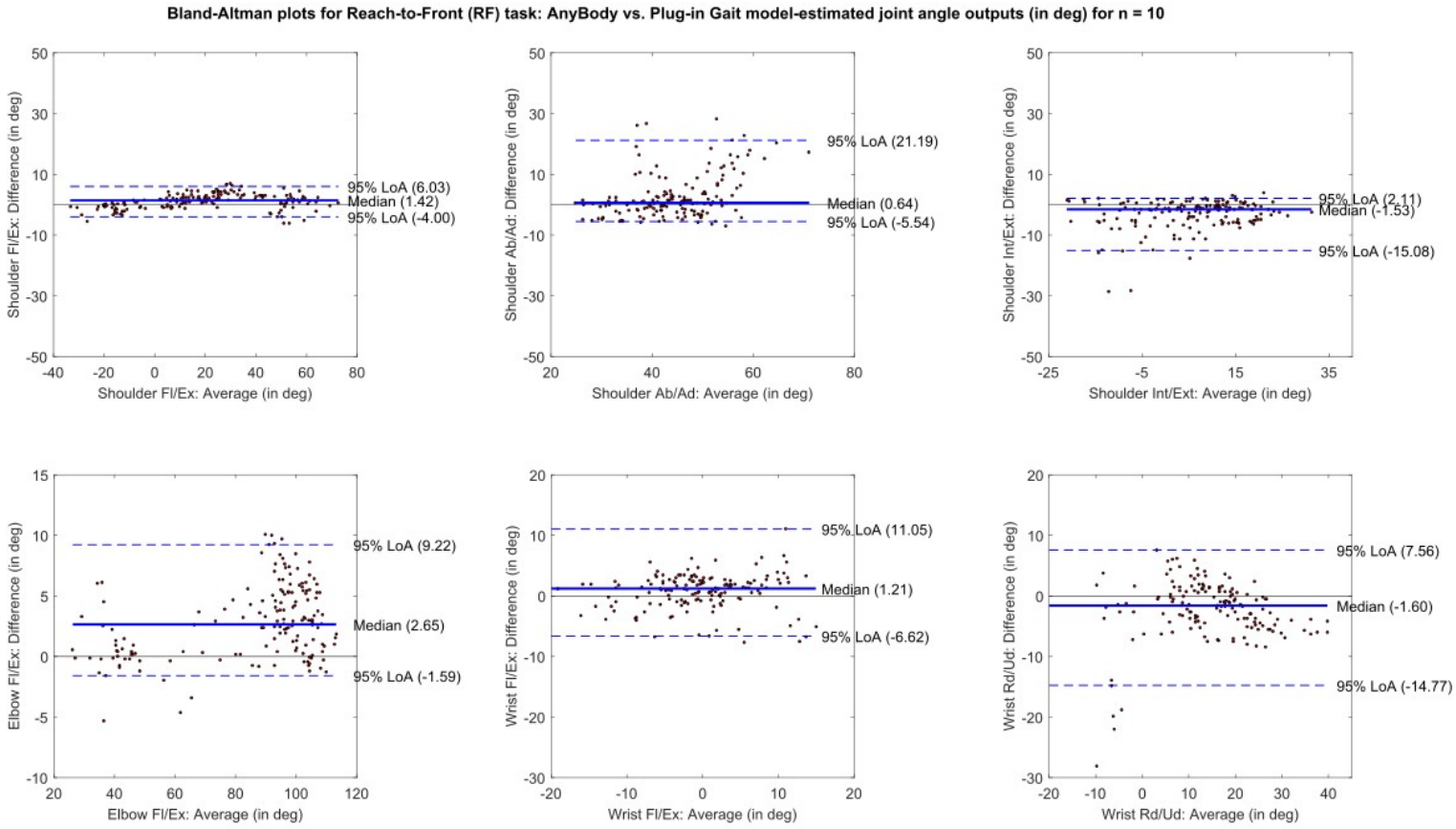


Figure 4: (a) Bland-Altman plots for checking agreement between the AnyBody and Plug-in Gait model-estimated joint angle outputs for three trials of *Reach-to-Front (RF)* task (three trials & n = 10). $Difference = (AnyBody - Plug-in\ Gait)$ (in degrees) and $Average = \frac{(AnyBody\ +\ Plug-in\ Gait)}{2}$ (in degrees). **Note**: Thick blue line is the median; the bottom and top dotted blue lines correspond to 2.5^{th} and 97.5^{th} percentiles, respectively; Average bias is estimated as the median of differences, i.e., 95% Limits of Agreement (LoA); and the horizontal black line at $y = 0$ represents a line of equality to establish systematic differences.

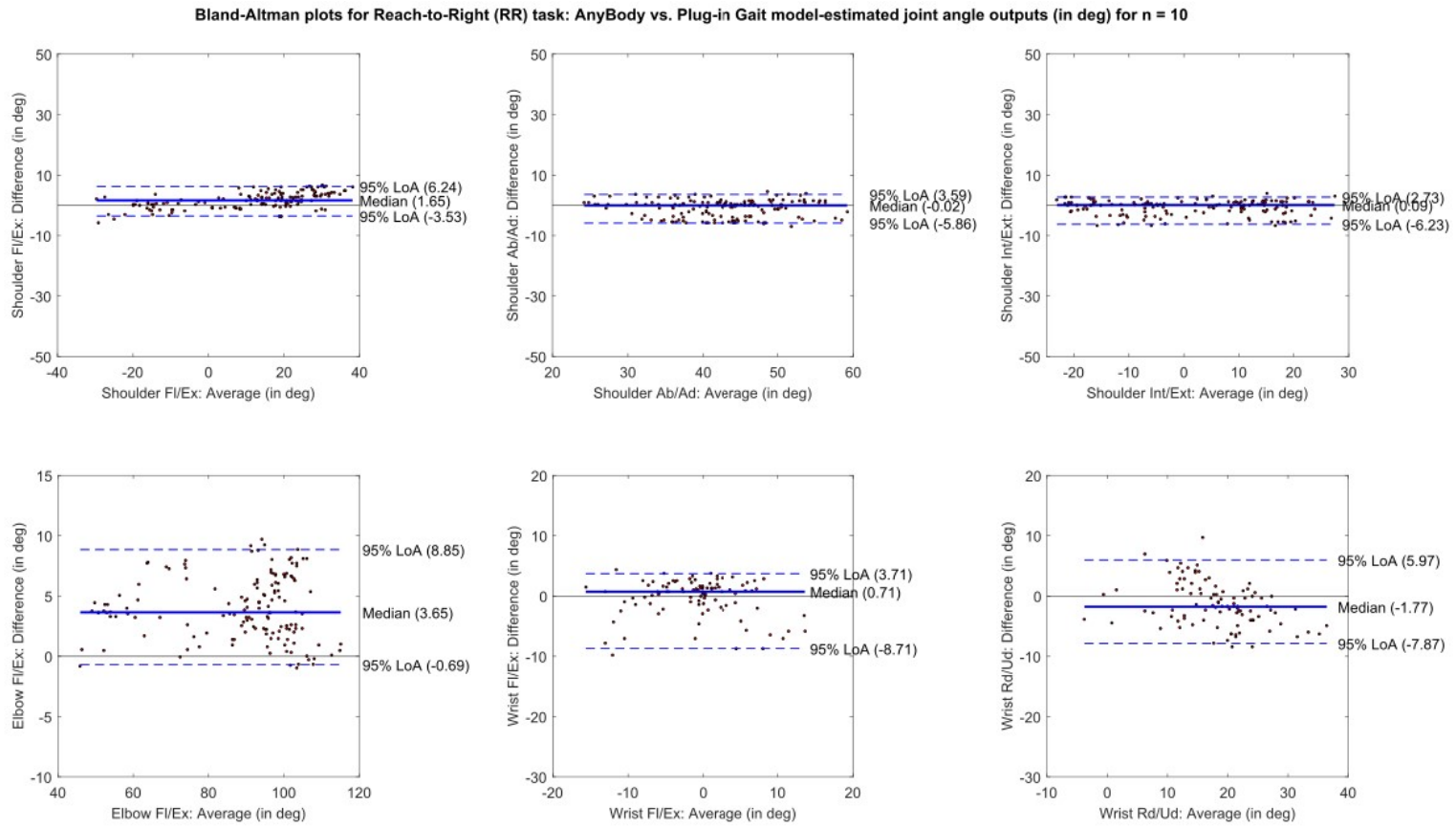


Figure 4: (b) Bland-Altman plots for checking agreement between the AnyBody and Plug-in Gait model-estimated joint angle outputs for three trials of *Reach-to-Right (RR)* task (three trials & n = 10). $Difference = (AnyBody - Plug-in\ Gait)$ (in degrees) and $Average = \frac{(AnyBody\ +\ Plug-in\ Gait)}{2}$ (in degrees). **Note**: Thick blue line is the median; the bottom and top dotted blue lines correspond to 2.5^{th} and 97.5^{th} percentiles, respectively; Average bias is estimated as the median of differences, i.e., 95% Limits of Agreement (LoA); and the horizontal black line at $y = 0$ represents a line of equality to establish systematic differences.

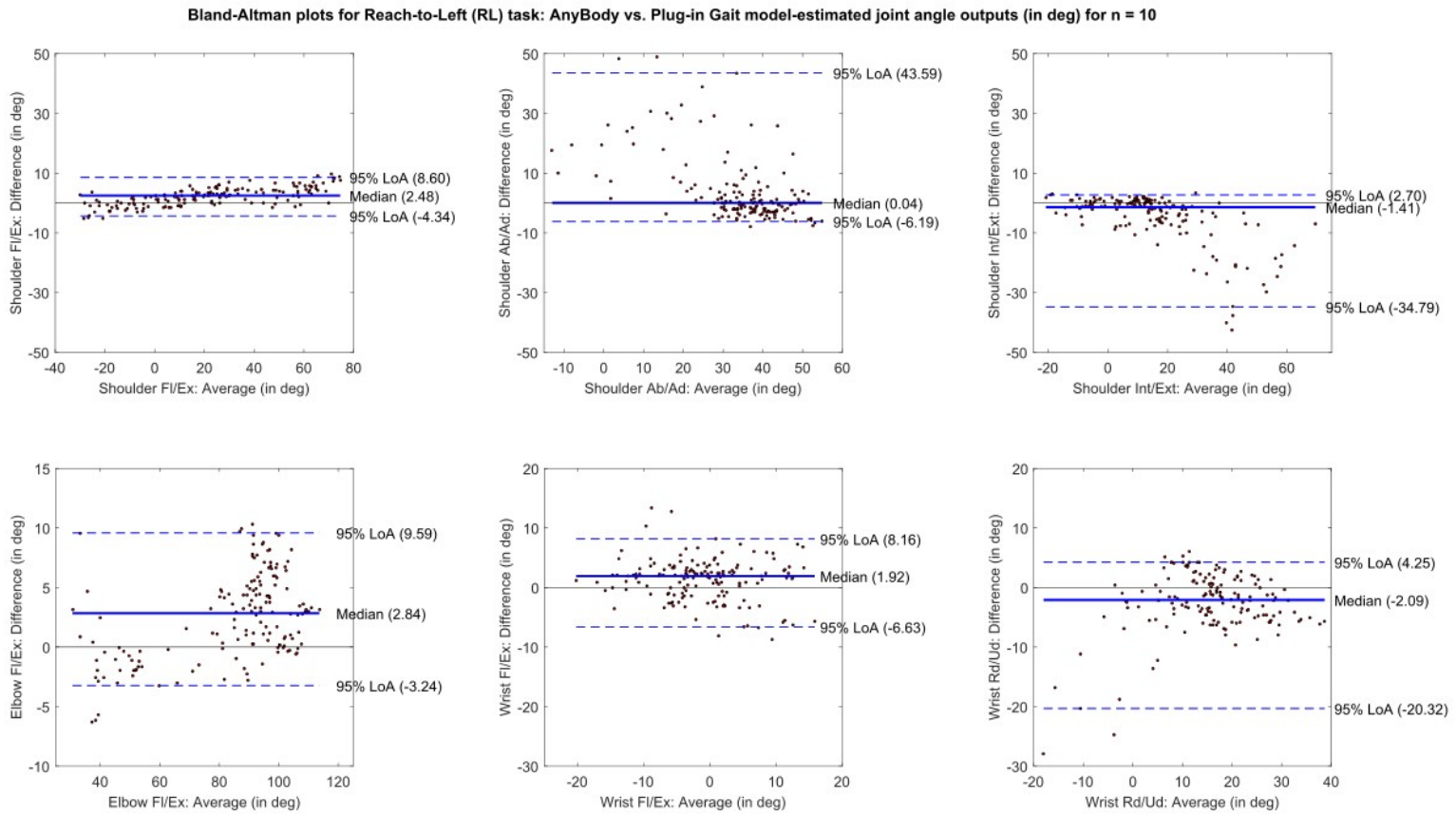


Figure 4: (c) Bland-Altman plots for checking agreement between the AnyBody and Plug-in Gait model-estimated joint angle outputs for three trials of *Reach-to-Left (RL)* task (three trials & n = 10). $Difference = (AnyBody - Plug-in\ Gait)$ (in degrees) and $Average = \frac{(AnyBody\ +\ Plug-in\ Gait)}{2}$ (in degrees). **Note**: Thick blue line is the median; the bottom and top dotted blue lines correspond to 2.5^{th} and 97.5^{th} percentiles, respectively; Average bias is estimated as the median of differences, i.e., 95% Limits of Agreement (LoA); and the horizontal black line at $y = 0$ represents a line of equality to establish systematic differences.