

# **Ultrasound-based Dynamic Bone Tracking to Enhance Clinical Assessments of Knee Kinematics**

Matthew B. Blomquist<sup>1</sup>, Joshua D. Roth<sup>2,3</sup>

<sup>1</sup>Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI;

<sup>2</sup>Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI;

<sup>3</sup>Department of Orthopedics and Rehabilitation, University of Wisconsin-Madison, Madison, WI

## **Correspondence:**

Joshua D. Roth

Department of Orthopedics and Rehabilitation

University of Wisconsin-Madison

1111 Highland Ave., WIMR 5037, Madison, WI 53705, USA.

Email: roth@ortho.wisc.edu

## **Keywords**

Knee Kinematics, Bone Tracking, Motion Tracking, Ultrasound Tracking, Cross-Correlation

## **Abstract**

**Purpose:** Measuring joint kinematics in the clinic is important for diagnosing injuries, tracking healing, and guiding treatments. However, current methods for measuring joint kinematics are limited by accuracy and/or feasibility of widespread clinical adoption. Therefore, the purpose of this study was to develop and validate an ultrasound-based bone-tracking algorithm to track bone motion and assess kinematics during simulated clinical assessments.

**Methods:** We mimicked four standard laxity assessments (i.e., anterior, posterior, varus, valgus) on five human cadaver knees using a robotic testing system. We simultaneously collected B-mode cine loops with an ultrasound transducer aligned to image in the plane of the applied load. We used the bone-tracking algorithm to estimate the change in knee kinematics throughout each laxity test to assess the potential of using ultrasound to estimate dynamic knee kinematics. Additionally, we conducted additional studies to test the repeatability of measuring laxity with the transducer at the same position and the reproducibility of measuring laxity at different transducer positions.

**Results:** Using the bone-tracking algorithm to estimate changes in knee kinematics, we computed the maximum root-mean-square errors of our bone-tracking algorithm to be 2.2 mm and 1.2° for the anterior-posterior and varus-valgus laxity assessments, respectively. Repeated laxity assessments proved to have good to excellent repeatability, while ICCs from repositioning the transducer varied more widely, ranging from poor to good reproducibility.

**Conclusions:** Ultrasound is an imaging modality capable of tracking knee kinematics in both anterior-posterior and varus-valgus degrees-of-freedom. Since ultrasound is widely used in both clinical and research settings, our ultrasound-based bone-tracking algorithm has the potential to assess knee kinematics in a variety of applications, such as diagnosing disorders, monitoring healing, and informing rehabilitation.

## Introduction

Current techniques for assessing joint motion are limited by accuracy and/or feasibility of widespread clinical adoption. In the clinic, the most common assessment of joint motion is a qualitative assessment of range of motion or joint laxity to diagnose ligament injuries or disease severity [22]. However, because they are not quantitative, they are limited in their ability to (1) diagnose more subtle changes in kinematics due to partial injury, disease progression, or healing, and (2) inform treatment planning (e.g., pre-operative planning for knee arthroplasty). Arthrometers, such as the commonly used KT-1000 or custom arthrometers (e.g., [33]), are also used in the clinic to measure joint motion. However, these devices are limited in their accuracy, in part due to soft-tissue compliance [10, 35], and thus are used more so to distinguish between normal and abnormal knee states (e.g., ruptured and intact ligament states [32, 35]) rather than to diagnose partial injuries or mild diseases, or to longitudinally track healing or disease progression. Stress radiographs [17] and other novel arthrometers [26] use x-rays to directly image changes in relative pose of the bones between loaded and unloaded states. However, while they overcome the limitations of soft-tissue compliance, they expose the patient to ionizing radiation [21] and are limited to static poses.

In a research setting, optical motion capture is commonly used to capture six degree-of-freedom kinematics. However, skin-motion artifacts can introduce large errors in measured joint motion [3, 28]. Fluoroscopy (or dynamic x-ray) is the gold standard for measuring joint motion in the research setting, but it requires costly custom hardware, image-processing expertise, and exposes patients to ionizing radiation, which make large-scale clinical implementation, use in longitudinal studies, and/or use in sensitive populations (e.g., pediatric patients [4, 12]) challenging. Thus, there is a critical need for a novel method to track joint motion that is both accurate and can be implemented for a broad range of applications in the clinic. To achieve the

latter requirement, this method should utilize hardware that is already commonly available in both clinical and research environments.

Ultrasound may be a suitable alternative imaging modality to assess knee motion in clinical and research settings. It is advantageous over radiography and fluoroscopy because it does not expose patients to ionizing radiation and is already broadly familiar to both clinicians and researchers. Previous studies have shown that ultrasound has the potential to track increases in joint gaps, demonstrated in both the tibiofemoral joint during a valgus loading test [34] and the patellofemoral joint during a lateral patellar glide test [5]. However, both groups manually measured bone motion during post-processing, which limits the ability to track bone motion across a large-scale of activities and/or subjects. Constructing a method to automatically track bone motion and convert it to joint kinematics during dynamic testing would increase the feasibility of using ultrasound to assess bone motion, and thus measure kinematics, in both clinical and research applications.

Therefore, the purpose of this study was to develop and validate an ultrasound-based bone-tracking algorithm to track bone motion during dynamic movement. Specifically, we used clinical laxity exams on the knee as an example use case. Accordingly, our two objectives for the present study were to (1) develop an ultrasound-based bone-tracking algorithm and determine the errors in using our algorithm to compute kinematics of the knee and (2) assess the repeatability and reproducibility of using our algorithm to measure kinematics of the knee.

## **Methods**

### ***Specimen Preparation***

We prepared five fresh-frozen cadaveric knees (1F/4M,  $66.2 \pm 3.4$  years,  $178.3 \pm 6.4$  cm,  $70.3 \pm 11.0$  kg) for robotic testing. We did not screen knees prior to testing, but following

dissection, we found two knees to have small to moderately small osteophyte formation (i.e., WORMS of 2-3 [27]) on the medial aspects of the femur and tibia. For each knee, we first fixed the fibula to the tibia using a transverse screw inserted 12 cm distally from the joint line to maintain the rigidity of the proximal fibular-tibial joint, and then transected the fibula just distal to this screw. Second, we transected the femur and tibia 24 cm proximal and distal from the joint line, respectively. Third, we removed all the tissue from the proximal and distal 10 cm of the femur and tibia, respectively. Fourth, we potted the femur and tibia within aluminum tubes (Bondo Body Filler, 3M) to enable rigid mounting of each bone to our six-degree-of-freedom (DOF) robotic testing system (KR300-R2700, KUKA Robotics) (**Figure 1a**). Fifth, we attached arrays of motion capture markers to the shafts of the femur and tibia using screws. Sixth, after mounting the specimen to the robot, we digitized anatomical landmarks of the knee using an optical motion capture system (Optitrack Prime 13x, NaturalPoint, Inc.) to establish an initial joint coordinate system [13]. Seventh, we defined a functional coordinate system that minimized kinematic cross-talk [15] between the primary DOFs (i.e., flexion-extension and internal-external rotation) and secondary DOFs (i.e., varus-valgus rotation, and anterior-posterior, medial-lateral, and compression-distraction translation) during prescribed passive flexion-extension and internal-external rotation paths [25].

### ***Robotic Testing Protocol***

We used a six-DOF robotic system controlled with simVITRO® software (Cleveland Clinic) to perform clinical assessments of anterior, posterior, varus, and valgus laxity. We performed all laxity assessments with two different loading trajectories (**Figure 2**): (1) uni-directional loading to simulate an ideal laxity exam with minimized off-axis loads, except a constant 50 N compressive load, and (2) multi-directional loading to simulate a manual laxity exam with small, representative off-axis loads, in addition to a constant 50 N compressive load. For the multi-directional loading simulations, we determined the load and flexion angle

trajectories by tracking joint kinematics and kinetics during manual laxity assessments on a knee surrogate mounted to the robotic system (**Supplement A**). Briefly, we mounted the tibia of the surrogate to a universal force-torque sensor and then manually applied a load or torque to the femur to mimic a laxity exam performed by a clinician [1]. We recorded joint kinematics and kinetics using simVITRO® software (Cleveland Clinic) in the same functional coordinate system [25] that we use for cadaveric knees so these loading profiles could be directly applied to the cadaver knees. The maximum forces and torques for both trajectories were  $\pm 89$  N (at a rate of 4.5 N/s) for anterior-posterior (A-P) assessments and  $\pm 15$  Nm (at a rate of 0.75 Nm/s) for varus-valgus (V-V) assessments. We performed all assessments at 0°, 20°, and 45° of knee flexion. For each knee, we randomized the order of the four loading directions, and then we randomized the order of the flexion angles and used this random order for each loading direction. Further details on the loading trajectories and robotic tracking errors for each assessment are shown in **Supplement B**.

### ***Tracking of Bone Motion***

To track bone motion during each laxity assessment, we collected ultrasound B-mode cine loops that were synchronized with the robotic testing system's data acquisition of knee kinematics and kinetics. We placed the ultrasound transducer (LF11-5H60-A3, ArtUS, TELEMED) in a custom holder that included four motion capture markers (**Figure 1b**). We used these markers to track the transducer's position and orientation relative to the knee throughout testing. During testing, we secured the transducer in the custom holder to the knee using self-adherent wrap (Coban, 3M), which mimics a fixation strategy that could be translated into the clinic. During A-P loading, we initially positioned the transducer over the patellar tendon, and then to avoid imaging over the patella, we moved the transducer laterally around the periphery of the knee until we obtained a quality image of both the femur and tibia without contacting the patella or patellar tendon (**Figure 1c**). During V-V loading, we positioned the transducer over

the lateral collateral ligament for varus assessments and the medial collateral ligament for valgus assessments (**Figure 1c**).

We developed and implemented a custom algorithm to track bone motion in B-mode cine loops throughout each trial (MATLAB 2023a, Mathworks, Natick, MA). In this algorithm, we used the first B-mode frame to define two rectangular regions of interest (ROIs) that we wanted to track over time: one ROI around the tibia and one ROI around the femur (**Figure 3a**). We conducted a supplemental study to determine the reproducibility of estimating kinematics when different users select the initial ROIs and found good reproducibility using our method (**Supplement C**). In the second B-mode frame, our custom algorithm performed a normalized cross-correlation to find the peak correlation between each bone ROI in the first and second B-mode frames (**Figure 3b**). The code repeated this cross-correlation between each successive pair of B-mode frames to track both ROIs throughout the trial. We converted the changes in ROI positions to distances using the resolution of the B-mode images, which was 0.051 mm per pixel for all trials in this study. Finally, to eliminate high-frequency noise caused by small tracking errors from frame to frame, we filtered the distance changes using a second-order Butterworth low-pass filter with a cutoff frequency of 1 Hz. We also filtered the robot kinematics using the same method.

### ***Evaluating tracking at clinically relevant movement speeds***

To obtain accurate tracking of the kinetics using the robotic system, we conducted each laxity assessment at a speed much slower than typical clinical assessments. Therefore, in addition to the speed at which we conducted the experiment (Case 1), we also evaluated our bone-tracking algorithm with assessments conducted at faster velocities to mimic joint speeds during different clinically relevant movements. Specifically, we evaluated three additional cases: a standard laxity assessment of 0.5 seconds of loading followed by 0.5 seconds of unloading (Case 2), the relative velocity between the femur and tibia during healthy gait (4.2 mm/s for A-P

motion [7] and 14.8 °/s for V-V motion [11], Case 3), and the relative velocity between the femur and tibia in a pathologic state (e.g., 28.6 mm/s to mimic an ACL injury for A-P [7] and 32.2 °/s to mimic varus thrust for V-V [6], Case 4).

To mimic each of these additional cases, we downsampled the ultrasound data in Case 1 by different factors to replicate the relative velocities reported in each of the other cases (**Table 1**). For example, for the uni-directional V-V trials during the experiment, we loaded the knee to 15 Nm over 20 seconds. Thus, to instead simulate 15 Nm over 0.5 seconds, we downsampled the ultrasound data by a factor of 40 (i.e., used every 40<sup>th</sup> frame of the Case 1 data as the Case 2 data). For Cases 3 and 4, we computed the experimental velocity for each knee by dividing the maximum motion from the experimental trials by the time elapsed to reach this maximum motion. Then, we divided reported velocities of the knee during healthy and pathologic gait, respectively, by these experimental velocities.

### ***Conversion of measured bone motion to joint kinematics***

To improve the interpretability of our measurements, we converted the measured motion between the bones using our bone-tracking algorithm into either A-P or V-V kinematics. For each frame in each A-P assessment, we computed the angle between the face of the ultrasound transducer and the anterior-posterior axis of the tibial plateau, defined as the line orthogonal to the superior-inferior axis through the tibial shaft and the line between the most medial and lateral points on the tibial plateau [25]. Then, we divided the measured motion between the bones by the cosine of this angle to account for the oblique imaging plane necessary to avoid disrupting patellar motion (**Figure 4a**). For V-V assessments, we assumed the varus or valgus rotation occurred about an axis through the middle of the tibial plateau [8]. We calculated the medial-lateral width of each tibia as the distance between the most medial and lateral extremes of the tibial plateau. Using a small angle approximation, we divided the measured motion using our algorithm by half the width of the tibial plateau to compute the

corresponding rotation. Finally, we calculated the angle between the face of the ultrasound transducer and the axis of the tibial plateau in the medial-lateral direction for each frame. We divided the corresponding rotation by the cosine of this angle to account for the potential oblique view from the transducer (**Figure 4b**).

### ***Repeatability and reproducibility studies***

We collected repeated trials on each knee to assess both the repeatability and reproducibility of our bone-tracking algorithm. First, we conducted three additional uni- and multi-directional loading assessments at 20° of knee flexion for all DOFs. However, instead of prescribing kinetics like the first trial, we prescribed the resulting kinematics recorded during the first, kinetic-controlled trial for each of the three additional tests. For example, after we completed a uni-directional varus loading assessment in kinetic control, we used the resulting kinematics from that assessment to prescribe the kinematics for the three repeatability tests of varus loading. Second, we repositioned the transducer to conduct additional tests. For A-P trials, we repositioned the transducer once more: instead of initially placing the transducer anterior to the knee and moving laterally, we instead moved the transducer medially until a quality image of the femur and tibia appeared. For V-V assessments, we repositioned the transducer twice more: one anterior to the original position (~2 cm) and one posterior to the original position (~2 cm). For each position, we conducted three repeatability tests. Thus, we conducted six tests per A-P assessment (two positions for A-P, each with three trials) and nine tests per V-V assessment (three positions for V-V, each with three trials) for each specimen.

For the repeatability study, we analyzed the data within each transducer position to assess the repeatability of our algorithm when neither the transducer placement nor the true knee kinematics changed. For the reproducibility study, we analyzed the data between transducer placements to characterize the errors introduced by repositioning the ultrasound

transducer. For each of these additional laxity assessments, we used our US-based bone-tracking algorithm to measure kinematics.

### **Statistical Analysis**

To address the first objective, we computed the errors in the US-measured kinematics by taking the difference between the US- and the robot-measured kinematics at each ultrasound frame. Following processing, we excluded one trial (a varus, multi-directional loading assessment at 45° of knee flexion) because the femur dropped out of the ultrasound frames. Additionally, we computed the errors in the maximum excursions (i.e., laxity) measured with ultrasound compared with the maximum excursions measured with the robot. We pooled the errors across specimens and flexion angles to compute the bias (mean), precision (standard deviation), and root-mean-square error (RMSE) for each DOF [2]. We repeated this process for all four cases of the different loading rates.

To address the second objective, we performed two different analyses on (1) the repeatability of measuring joint laxity in the same position and (2) the reproducibility of measuring joint laxity after repositioning the ultrasound transducer. First, we computed the repeatability and reproducibility variances of our ultrasound-based bone-tracking algorithm using ISO 5725 [16] (equations found in **Supplement D**). Then, to put these variances into context of the measured kinematics, we computed the coefficient of variation for each repeatability and reproducibility variance using **Equation 1**:

$$CV = \frac{\sqrt{\sigma^2}}{\mu} \quad \text{Equation 1}$$

Where  $CV$  is the repeatability or reproducibility coefficient of variation,  $\sigma^2$  is the repeatability or reproducibility variance, and  $\mu$  is the mean of the measured kinematics over all the trials for that specimen's laxity assessment. Finally, we computed the median coefficient of variation for each laxity assessment to summarize the coefficients of variation across the specimens.

For the second analysis, we computed the intra-class correlation coefficients (ICC) to assess the repeatability and reproducibility of measuring laxity. We computed ICCs using two-way random effects models for absolute agreement among measurements [24]. To compute ICCs for the repeatability study, we treated each position as independent, so we had 10 (for A-P) or 15 (for V-V) positions, each with three measurements. To compute ICCs for the reproducibility study, we treated each measurement within a specimen as independent, so we had five specimens, each with six (for A-P) or nine (for V-V) measurements. We interpreted ICC values according to Koo et al. [20] where ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.9 are indicative of poor, moderate, good, and excellent repeatability/reproducibility, respectively. Due to data collection or post-processing challenges (e.g., femur dropped out of the ultrasound frames), we excluded two positions of the multi-directional-loading anterior test, and one position each for a multi-directional-loading varus test, uni-directional-loading anterior test, and uni- and multi-directional-loading posterior test.

## Results

When using all of the ultrasound frames (Case 1), the largest RMSEs for translations and rotations were 2.2 mm and 1.2°, respectively, for uni-directional loading and 2.1 mm and 1.3°, respectively, for multi-directional loading (**Figure 5**). Our algorithm tended to underestimate kinematics in each degree of freedom, except for the varus assessments. When downsampling to simulate different clinically relevant joint speeds (Cases 2-4), the errors largely remained unchanged when simulating a one-second laxity exam, but increased, more in A-P than in V-V, when simulating relative velocity of the knee during healthy and pathological gait. We reported errors in tabular format in **Supplement E**. When computing joint laxity, the largest RMSEs for translations and rotations were 2.6 mm and 1.6°, respectively, for uni-directional

loading and 2.1 mm and 1.3°, respectively, for multi-directional loading (**Figure 6**). The errors largely remained unchanged when downsampling to simulate different joint speeds.

To assess the repeatability of measuring laxity with the transducer in the same position, we computed the median coefficients of variation to be between 2.6 and 21.1% (**Figure 7**). In terms of ICC values, ICCs ranged from 0.81 to 0.99 (**Table 2**). To assess the reproducibility of measuring laxity with the transducer at different positions, we computed the median coefficients of variation to be between 8.9 and 47.3% (**Figure 7**). In terms of ICC values, ICCs ranged from 0.21 to 0.89 (**Table 2**).

## Discussion

Current techniques for assessing joint motion are limited by accuracy and/or feasibility of widespread clinical adoption. Thus, our objectives in the present study were to (1) develop an ultrasound-based bone-tracking algorithm and determine the errors in using our algorithm to compute kinematics of the knee and (2) assess the repeatability and reproducibility of using our algorithm to compute kinematics of the knee. Our first key finding was that the RMSEs for measuring A-P and V-V kinematics using our bone-tracking algorithm were less than 2.2 mm and 1.3°, respectively, for the experimental and the one-second loading rates. The second key finding was that our bone-tracking algorithm was highly repeatable when the transducer was in approximately the same position (median coefficients of variation between 2.6 and 21.1%, ICC = 0.81 to 0.99), but not as reproducible when the transducer was placed in different positions (median coefficient of variation between 8.9 and 47.3%, ICC = 0.21 to 0.89).

Regarding the first key finding that RMSEs for measuring A-P and V-V kinematics using our bone-tracking algorithm were less than 2.2 mm and 1.3°, these errors fall between those reported for optical motion capture and fluoroscopy during walking. We used kinematic tracking

errors during walking for comparison to our results because we are not aware of comparable kinematic tracking errors for laxity assessments. Reinschmidt et al. compared rotational kinematics from optical motion capture to intracortical bone pins during walking [29]. In three subjects, the range of root-mean-square errors in V-V were 2.1 to 2.8°. Similarly, Benoit et al. quantified the errors in motion capture to bone pins in eight subjects during walking [3]. Absolute error values during foot-strike, mid-stance, and toe-off ranged between 2.5 to 4.4° in V-V. Other studies have quantified the errors of using biplane fluoroscopy to track kinematics, which is the current gold standard method. For example, Guan et al. measured the accuracy of a mobile biplane fluoroscopy system in one healthy subject during walking [14]. Root-mean-square errors resulted in 0.77° for V-V. Therefore, errors in V-V are lower than values previously reported with optical motion capture and slightly larger than values previously reported with biplane fluoroscopy.

When computing joint laxity, the largest RMSEs for translations and rotations were 2.6 mm and 1.6°, respectively, for the experimental and the one-second loading rates. To put these errors into better context, we can compare them to other studies that have used different modalities to estimate joint laxity. For instance, a previous study [10] compared KT-1000 measurements of total A-P laxity to the gold-standard, Roentgen Stereophotogrammetric Analysis, in 15 patients following anterior cruciate ligament reconstruction. They reported an RMSE of 5.0 mm, which is larger than the RMSEs of the total A-P laxity from our ultrasound measurements (2.8 and 2.3 mm for the uni- and multi-directional loading assessments, respectively). For V-V measurements, we are not aware of any studies that have reported the errors in V-V laxity compared to a gold-standard method.

We were encouraged that the trials were similar when the data was downsampled to replicate a one-second laxity exam, with the only notable change in RMSE being an improvement from 2.1 to 1.5 mm in the anterior, multi-directional loading assessments. When

simulating joint motion velocities reported in healthy and pathologic gait, all but valgus uni- and multi-directional-loading and varus multi-directional-loading assessments increased in errors, up to 141% for the posterior, uni-directional-loading assessments during healthy gait. These results suggest that to obtain more accurate tracking of bone motion during functional activities, a faster frame rate than used in the current study (~40 Hz) is needed, but can be realistically obtained.

The second key finding was that our bone-tracking algorithm was highly repeatable when the transducer was in approximately the same position (median coefficients of variation between 2.6 and 21.1%, ICC = 0.81 to 0.99). These ICC values were comparable to values previously reported when using devices during clinical assessments, such as the KT-1000 for A-P laxity (0.47-0.94 [19, 31, 36]), or an instrumented V-V laxity device (0.84-0.93 [9]) or stress radiographs (0.97-1.0 [18]) for V-V laxity. In terms of the reproducibility study, we found measuring kinematics when the transducer is in different positions to not be reproducible, with median coefficients of variation between 8.9 and 47.3% and ICC values ranging from 0.21 to 0.89. These ICC values were comparable, but slightly lower than values previously reported when using devices during clinical assessments (KT-1000: 0.14-0.75 [30–32, 36], V-V instrumented laxity device: 0.65-0.88 [9], stress radiographs: 0.98-0.99 [18]).

These results suggest repositioning the transducer could lead to high variability in results, which is troublesome for studies consisting of repeated measurements over time. However, because we used anatomical landmarks to initially position the transducer before the reproducibility study, and repeated measurements in approximately the same location produced excellent repeatability, we suggest positioning the transducer over anatomical landmarks (LCL for varus, MCL for valgus, and the first quality image of the femur and tibia laterally to the patellar tendon) to obtain repeatable measurements. This is easily achievable because these structures are visible on the US images.

Three limitations should be considered when interpreting our findings. First, our sample size of five knees is relatively small. However, the focus of this paper was to demonstrate the ultrasound-based bone-tracking algorithm and its use on a subset of knees. Because we were not drawing statistical comparisons from these data, our sample size was not a limiting factor in our conclusions. Second, our cross-correlation algorithm is a relatively easy motion-tracking method, but artifacts found within the bone ROIs (e.g., shape of the bones changing or surrounding soft tissue that is in close proximity to the bones) can lead to poor tracking. For further details on how these artifacts can cause poor tracking of the bones, see **Supplement F**. Our ongoing work is focused on more robust tracking approaches to mitigate these artifacts. Third, discrepancies between the assumed planes of motion and the actual planes of motion could lead to errors during the conversion of bone motion to joint kinematics. For A-P assessments, tracking bone motion at an oblique angle to the A-P axis could increase the errors in A-P kinematics. However, we wanted to place the transducer in a position that could track bone motion during dynamic, functional activities, which caused us to place the transducer lateral of the patella. For V-V trials, we assumed a rotation axis about the center of the joint [8]. In reality, the rotation axis likely changes with different V-V angles [8], and potentially different knee geometries (e.g., cartilage wear in one compartment). As a supplement, we processed the A-P and V-V data while varying the transducer angle and location of the V-V axis, respectively, to show their effects on predicting kinematics (**Supplement G**).

In conclusion, this study showed that our ultrasound-based bone-tracking algorithm can track bone motion with RMSEs between 0.6 and 2.2 mm of A-P translation and 0.4° and 1.3° of V-V rotation during a range of loading scenarios from experimental speeds to clinical laxity assessments. Overall, this study is a promising first step in showing our ultrasound-based bone-tracking algorithm can measure joint kinematics in the clinic, which could enhance a variety of

clinical applications including diagnosing joint dysfunction, monitoring healing, and informing rehabilitation.

## Tables

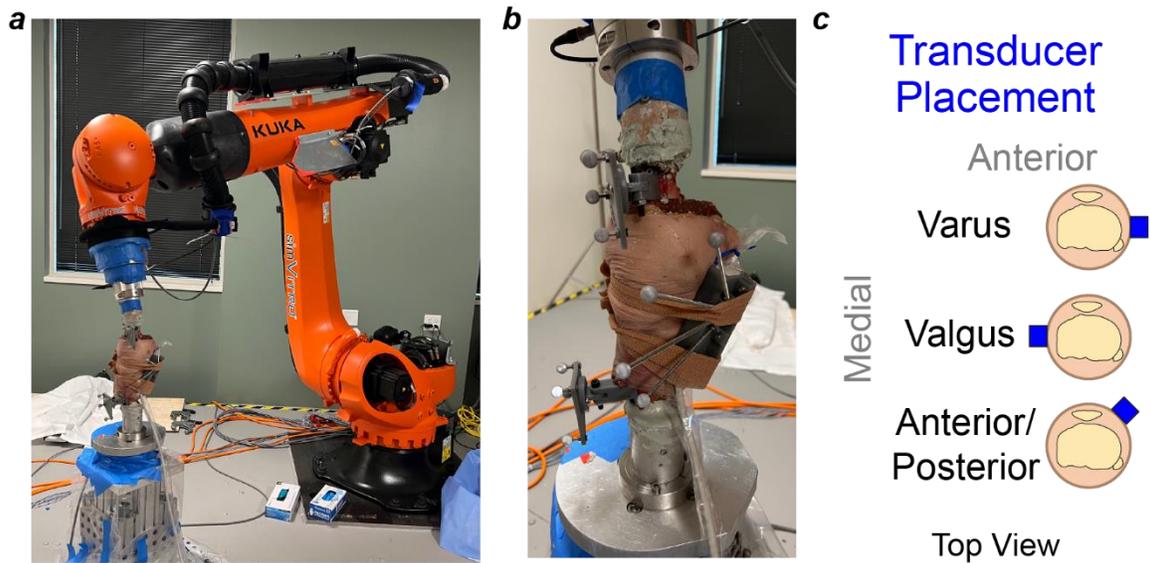
**Table 1:** Downsampling factors used to mimic different velocities of joint motion during different movements. For example, to mimic Case 2 velocities (a standard laxity exam of 0.5 seconds of loading followed by 0.5 seconds of unloading), the anterior-posterior uni-directional loading trials were downsampled by a factor of 40 (i.e., we used the 1<sup>st</sup>, 41<sup>st</sup>, 81<sup>st</sup>, etc. frames). Cases 3 and 4 are reported as ranges as the downsampling factor was adjusted for each knee because the maximum excursion of each trial differed from knee to knee. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

		Case 1 (Experiment)	Case 2 (Laxity Exam)	Case 3 (Healthy Gait)	Case 4 (Pathologic Gait)
Anterior-Posterior	Uni	1	40	14-30	95-204
	Multi	1	54	21-33	144-224
Varus-Valgus	Uni	1	40	81-113	177-245
	Multi	1	58	167-227	364-494

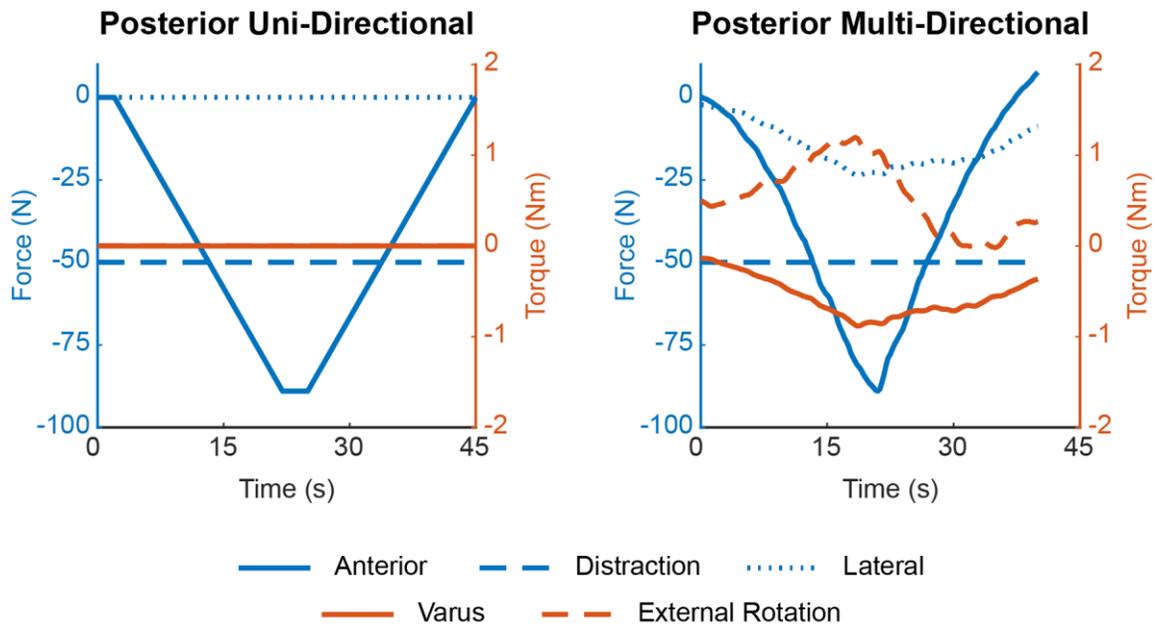
**Table 2:** Intra-class correlation coefficients (ICCs) to determine the repeatability in measuring laxity in the same transducer position ( $ICC_{\text{Repeatability}}$ ) and the reproducibility when the ultrasound transducer is in a new position ( $ICC_{\text{Reproducibility}}$ ). Also reported are 95% confidence intervals (95% CI). Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments. Note: due to data collection or post-processing challenges (e.g., femur dropped out of the ultrasound frames), we excluded two positions of the multi-directional-loading anterior test, and one position each for a multi-directional-loading varus test, uni-directional-loading anterior test, and uni- and multi-directional-loading posterior test.

		$ICC_{\text{Repeatability}}$ (95% CI)	$ICC_{\text{Reproducibility}}$ (95% CI)
Anterior	Uni	0.99 (0.96-1.00)	0.89 (0.67-0.99)
	Multi	0.97 (0.90-0.99)	0.19 (-0.09-0.94)
Posterior	Uni	0.99 (0.95-1.00)	0.81 (0.49-0.98)
	Multi	0.99 (0.96-1.00)	0.73 (0.29-0.98)
Varus	Uni	0.92 (0.83-0.97)	0.55 (0.24-0.92)
	Multi	0.81 (0.62-0.93)	0.55 (0.22-0.95)
Valgus	Uni	0.89 (0.77-0.96)	0.34 (0.10-0.83)
	Multi	0.91 (0.81-0.97)	0.21 (-0.01-0.76)

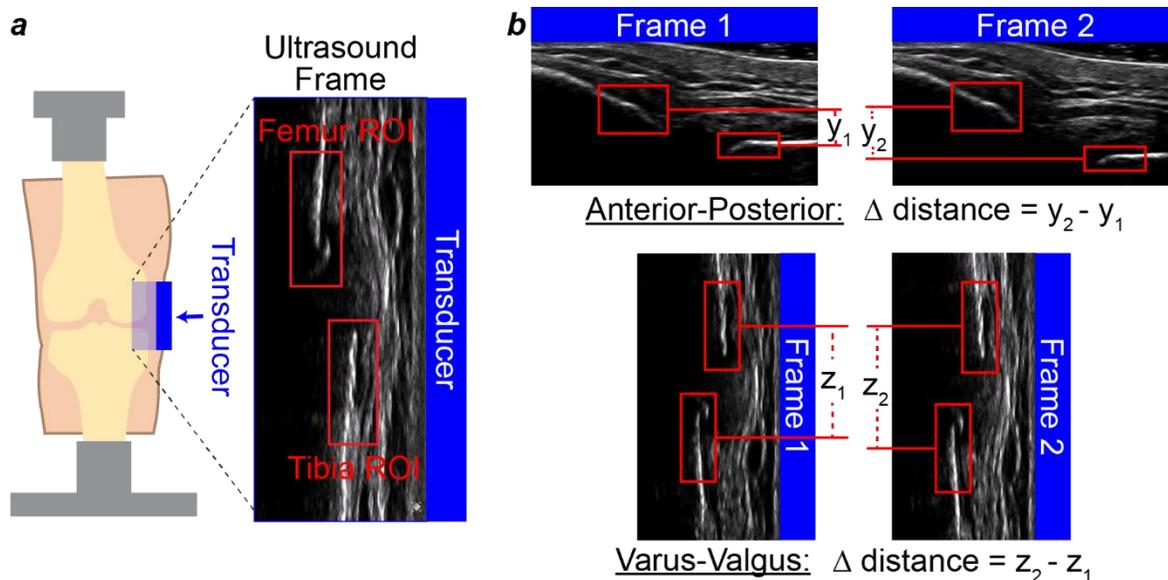
## Figures



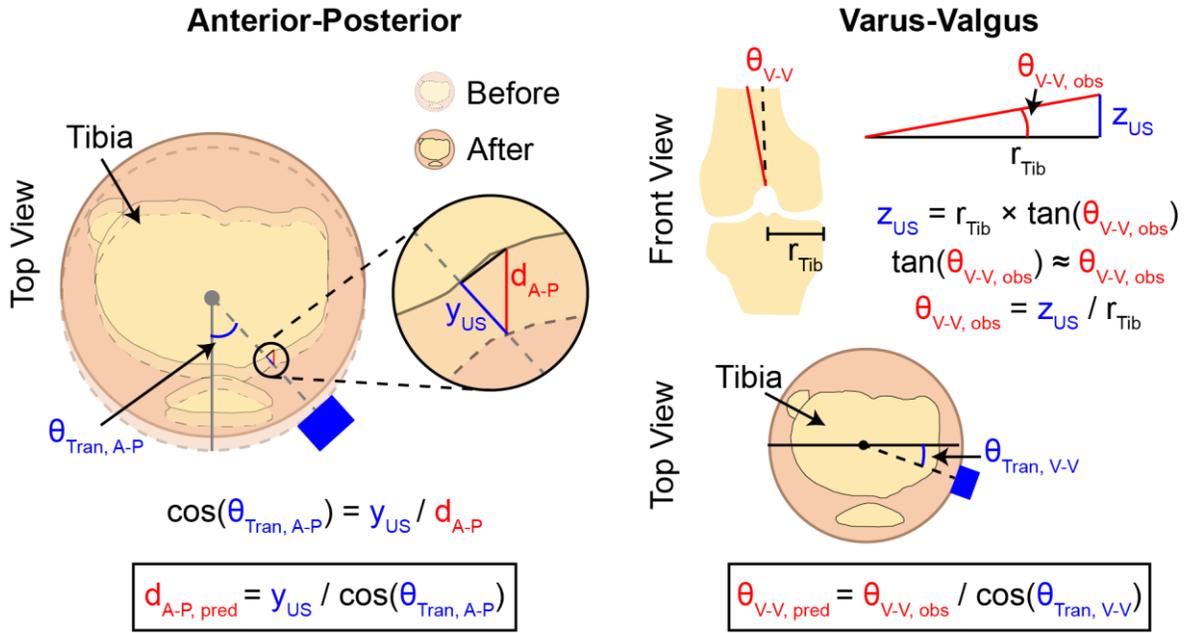
**Figure 1:** (a) We used a six-degree-of-freedom robotic testing system to perform laxity assessments in human cadaver knees. (b) We placed an ultrasound transducer over the joint line of the knee to track relative motion of the femur and tibia during each assessment. (c) We placed the transducer over the lateral collateral ligament during varus loading, over the medial collateral ligament during valgus loading, and lateral to the patellar tendon during anterior and posterior loading.



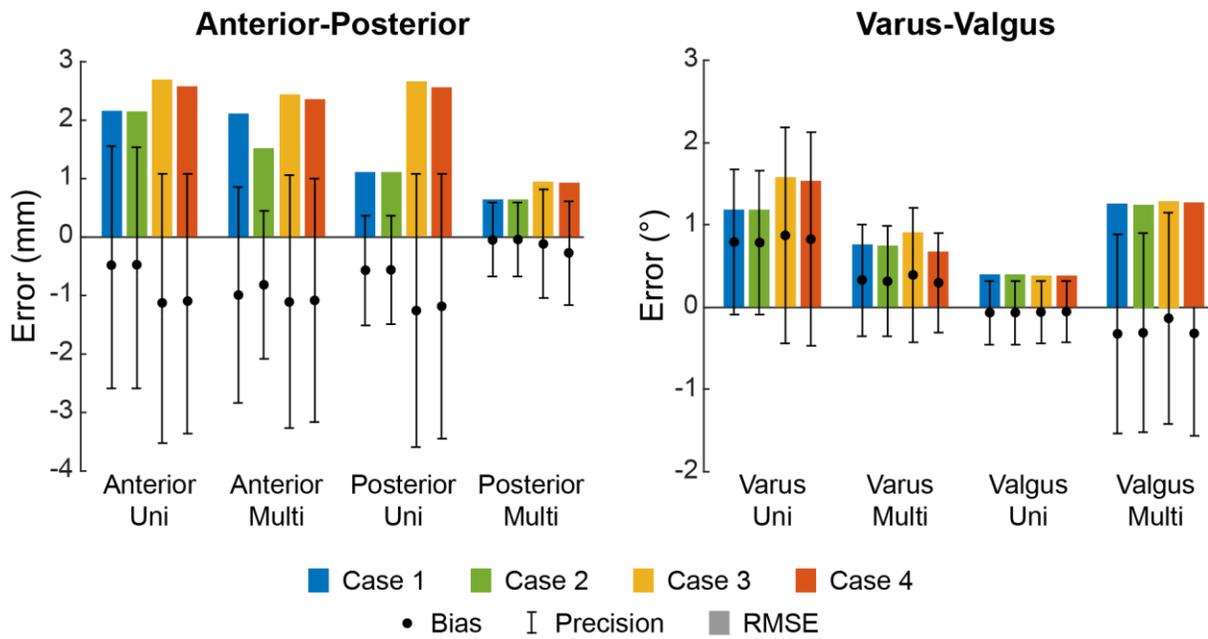
**Figure 2:** Example uni-directional and multi-directional loading trajectories for posterior assessments. For all uni-directional loading trajectories, the load of interest was ramped to maximum force/torque, held for five seconds, and then ramped back to zero. For all multi-directional loading trajectories, we manually applied loads/torques to a knee surrogate to mimic a laxity exam performed by a clinician [1]. In addition to these forces/torques, we also prescribed the knee flexion angle for each trajectory.



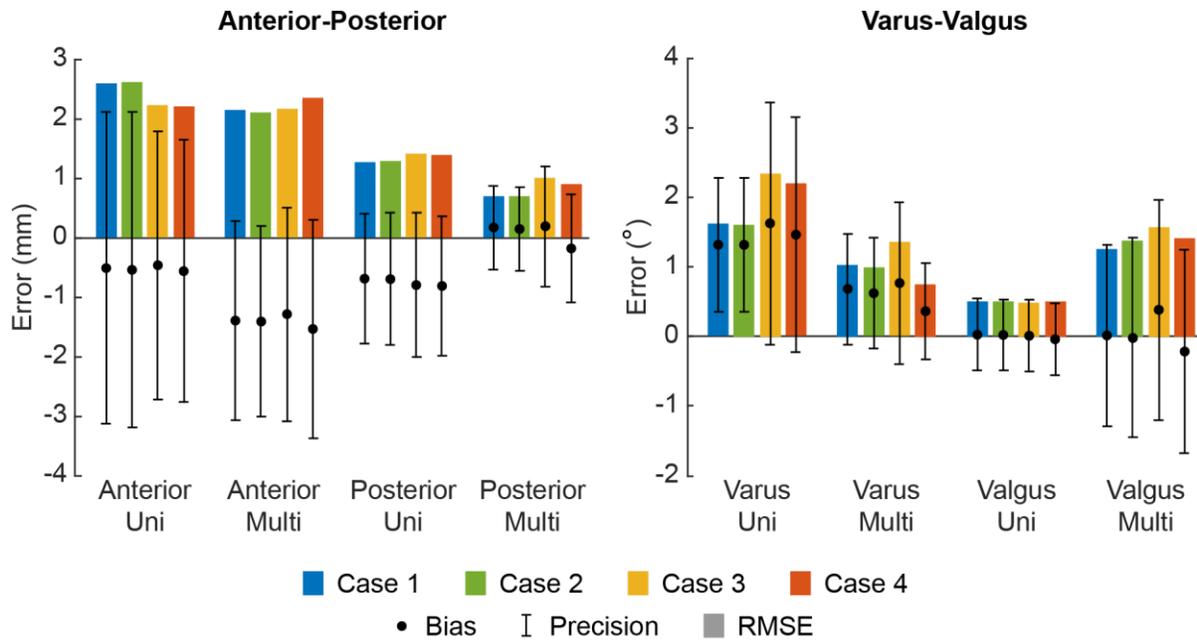
**Figure 3:** (a) We defined a rectangular region of interest (ROI) around both the tibia and the femur in the first B-mode frame. Our bone-tracking algorithm finds the maximum correlation of these ROIs in subsequent frames to track bone motion over time. (b) For anterior-posterior assessments, we computed distance changes ( $\Delta$  distance) as the changes in the ROI positions that are perpendicular to the transducer plane. For varus-valgus assessments, we computed distance changes ( $\Delta$  distance) as the changes in the ROI positions that are parallel to the transducer plane.



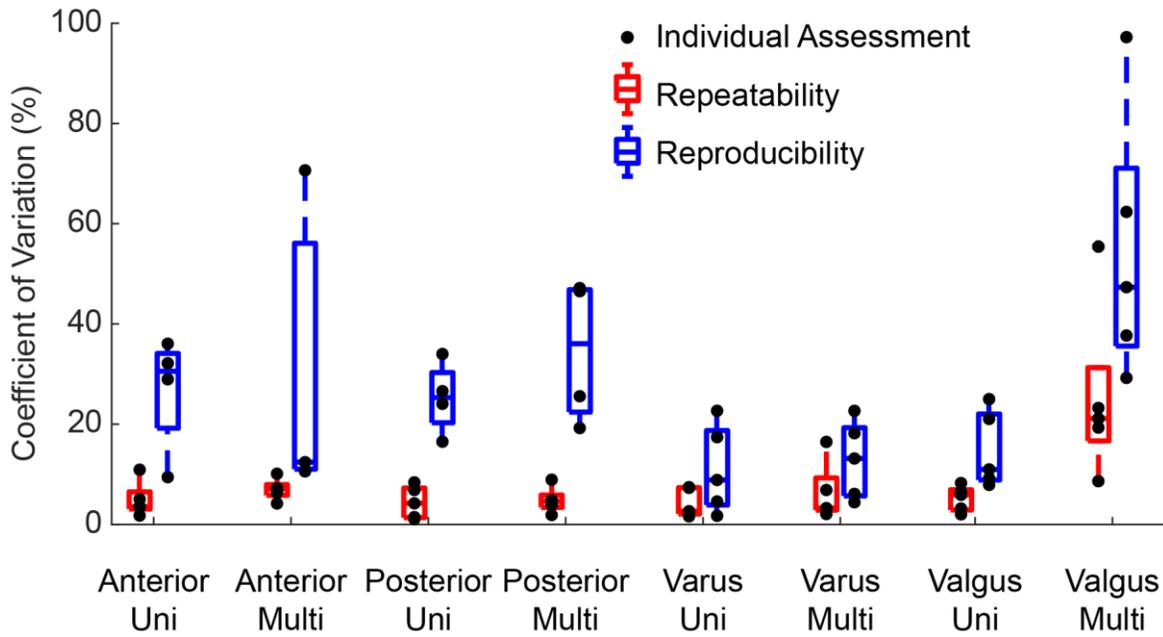
**Figure 4:** We converted the bone motion measured using ultrasound to anterior-posterior (A-P) or varus-valgus (V-V) kinematics using the equations above. For A-P assessments, we accounted for the oblique placement of the ultrasound transducer relative to the A-P axis of the tibia. For V-V assessments, we accounted for the V-V rotation axis, as well as the placement of the transducer relative to the medial-lateral axis of the tibia. **Abbreviations:**  $d_{\text{A-P}}$  = A-P translation measured by the robot,  $y_{\text{US}}$  = bone displacement measured by ultrasound (A-P),  $\theta_{\text{Tran, A-P}}$  = angle between the transducer and the A-P axis of the tibia,  $d_{\text{A-P, pred}}$  = ultrasound prediction of A-P displacement,  $\theta_{\text{V-V}}$  = V-V rotation measured by the robot,  $r_{\text{Tib}}$  = half the medial-lateral width of the tibial plateau,  $z_{\text{US}}$  = bone displacement measured by ultrasound (V-V),  $\theta_{\text{V-V, obs}}$  = ultrasound prediction of V-V angular displacement at oblique angle,  $\theta_{\text{Tran, V-V}}$  = angle between transducer and medial-lateral axis of the tibia,  $\theta_{\text{V-V, pred}}$  = ultrasound prediction of V-V angular displacement.



**Figure 5:** We computed the errors in kinematics at each frame when using our ultrasound kinematics compared to the gold-standard robot kinematics. We measured kinematics using each frame (Case 1), as well as downsampling the data to mimic a one-second laxity exam (Case 2), the relative velocity between the femur and tibia during healthy gait (Case 3), and the relative velocity between the femur and tibia in a pathologic state (Case 4). The bar plots represent root-mean-square errors, the scatter points represent the bias errors, and the error bars represent the precision errors. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments. Note: we excluded one trial (varus, multi-directional loading assessment at 45° of knee flexion) because the femur dropped out of the ultrasound frames during this trial.



**Figure 6:** We computed the errors in measuring laxity using our ultrasound-based tracking by comparing values to the gold-standard robot measurements during the experiment. We measured laxity as the maximum excursion of the kinematics when using each frame (Case 1), as well as downsampling the data to mimic a one-second laxity exam (Case 2), the relative velocity between the femur and tibia during healthy gait (Case 3), and the relative velocity between the femur and tibia in a pathologic state (Case 4). The bar plots represent root-mean-square errors, the scatter points represent the bias errors, and the error bars represent the precision errors. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments. Note: we excluded one trial (varus, specimen 3, multi-directional loading assessment at 45° of knee flexion) because the femur dropped out of the ultrasound frames during this trial.

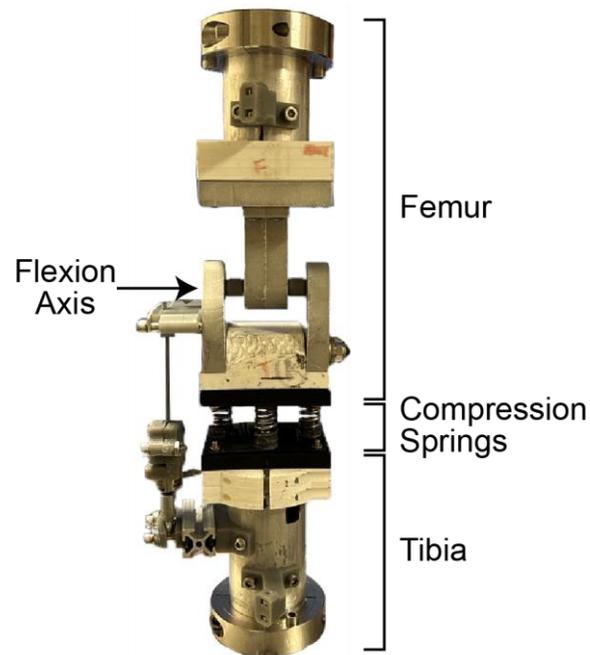


**Figure 7:** We computed the coefficients of variation of measuring laxity with the transducer at the same position (Repeatability) and at different positions (Reproducibility). The CVs of each specimen are plotted as scatter points, and box plots are created to obtain the medians and the 25% and 75% percentiles. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments. Note: due to data collection or post-processing challenges (e.g., femur dropped out of the ultrasound frames), we excluded two positions of the multi-directional-loading anterior test, and one position each for a multi-directional-loading varus test, uni-directional-loading anterior test, and uni- and multi-directional-loading posterior test.

## Supplemental Sections

### Supplement A: Knee Surrogate

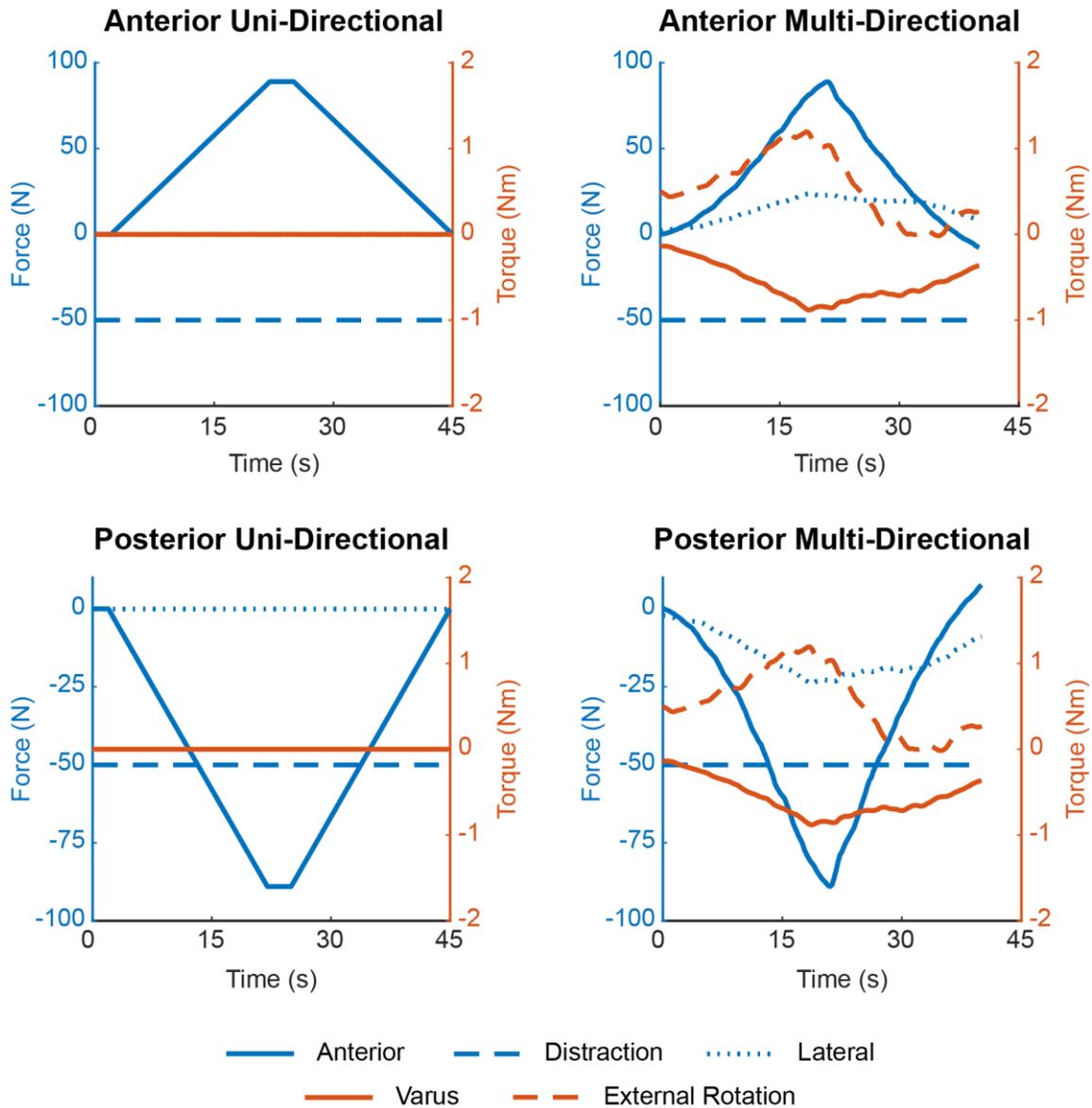
We used a surrogate knee model to obtain joint kinematics and kinetics during a manual laxity assessment, similar to a laxity assessment performed by a clinician. Our knee surrogate consists of two rigid bodies connected by compression springs to provide comparable stiffness to a typical knee (**Figure S.1**). Each rigid body consists of a combination of aluminum and 3D printed material (PLA material). The compression springs allowed for six-degree-of-freedom movement between the two rigid bodies, mimicking movement of a knee.



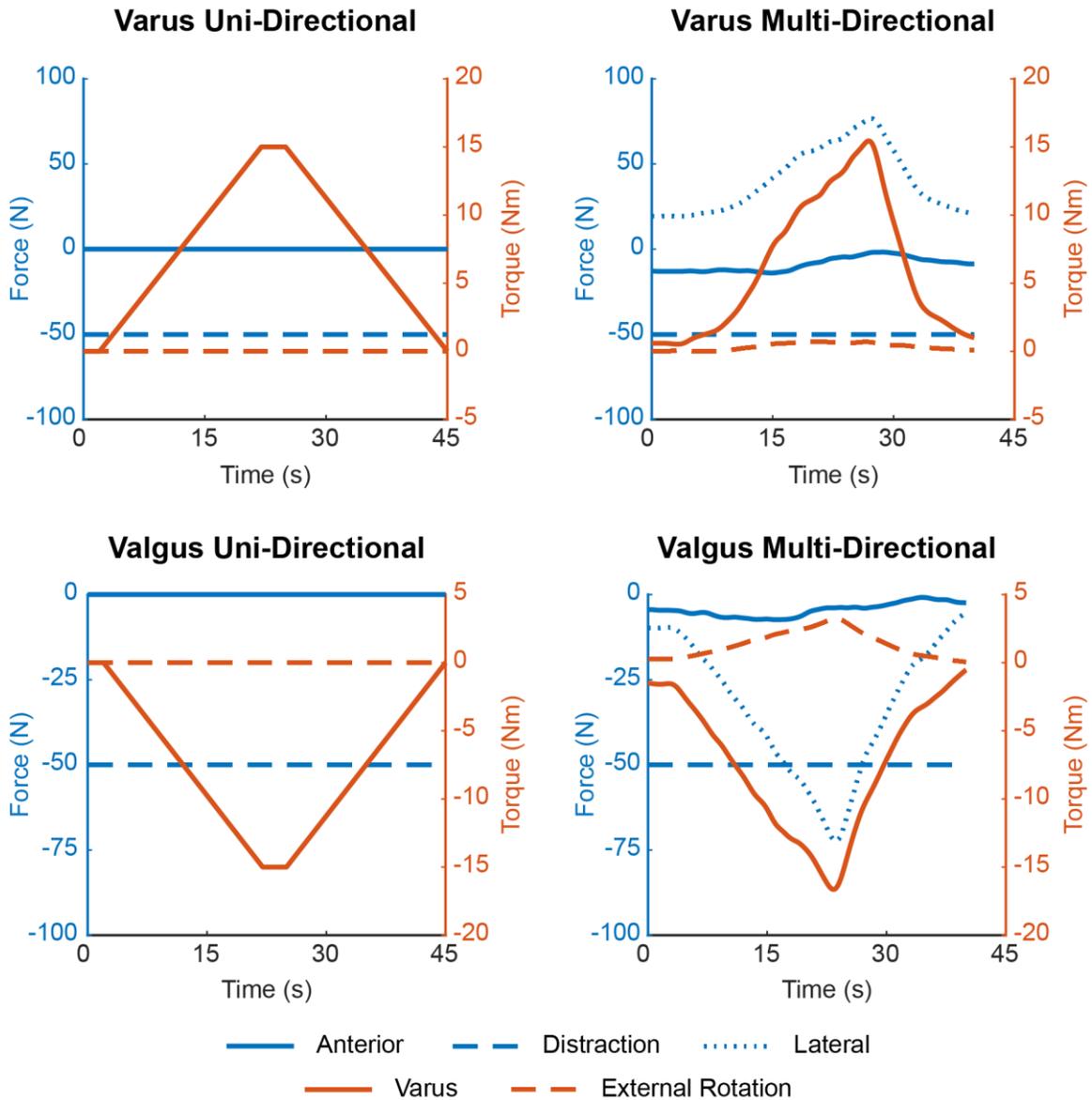
**Figure S.1:** We used a surrogate knee model to mimic a laxity assessment performed by a clinician. The surrogate consisted of a femur, tibia, and compression springs between the two rigid bodies to allow for six-degree-of-freedom movement.

## **Supplement B: Loading Trajectories and Load Tracking Errors**

We performed all laxity assessments with two different loading trajectories: (1) uni-directional loading to simulate an ideal laxity exam with minimized off-axis loads, except a constant 50 N compressive load, and (2) multi-directional loading to simulate a manual laxity exam with small, representative off-axis loads, in addition to a constant 50 N compressive load. For the multi-directional loading simulations, we determined the load and flexion angle trajectories by tracking joint kinematics and kinetics during manual laxity assessments on a knee surrogate mounted to the robotic system (**Supplement A**). The loading trajectories that we prescribed for the anterior and posterior laxity assessments are shown in **Figure S.2**, and the varus and valgus laxity assessments are shown in **Figure S.3**.

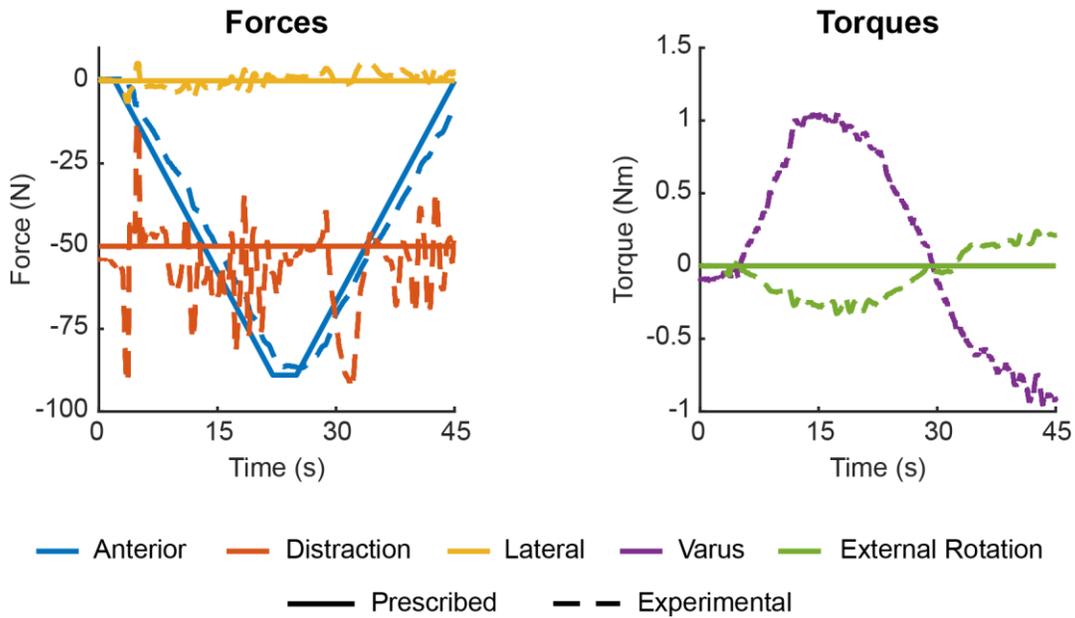


**Figure S.2:** We prescribed these forces and torques for the uni-directional and multi-directional loading trajectories for anterior and posterior laxity assessments. For all uni-directional loading trajectories, one load of interest was ramped to maximum force/torque, held for five seconds, and then ramped back to zero. For all multi-directional loading trajectories, we manually applied loads/torques to a knee surrogate to mimic a laxity exam performed by a clinician [1]. In addition to the forces/torques, we also prescribed knee flexion angle for each trajectory.



**Figure S.3:** We prescribed these forces and torques for the uni-directional and multi-directional loading trajectories for varus and valgus laxity assessments. For all uni-directional loading trajectories, one load of interest was ramped to maximum force/torque, held for five seconds, and then ramped back to zero. For all multi-directional loading trajectories, we manually applied loads/torques to a knee surrogate to mimic a laxity exam performed by a clinician [1]. In addition to the forces/torques, we also prescribed the knee flexion angle for each trajectory.

We also computed the errors in prescribing forces/torques during the experiment by taking the difference between the prescribed loads and the actual applied loads at each time point (**Figure S.4**). We pooled the errors across specimens and flexion angles to compute the root-mean-square error (RMSE) for each load in each laxity assessment [2] (**Table S.1**).



**Figure S.4:** Example plots of the prescribed loads (solid lines) versus the actual (i.e., experimental) loads (dashed lines) during a posterior, uni-directional loading assessment.

**Table S.1:** Root-mean-square errors (RMSE) for prescribing loads compared to the actual loads applied for each laxity assessment. Abbreviations: A-P = anterior-posterior force, C-D = compression-distraction force, M-L = medial-lateral force, V-V = varus-valgus torque, I-E = internal-external rotation torque, F-E = flexion-extension angle, Uni = uni-directional-loading assessments, Multi = multi-directional loading assessments.

		A-P (N)	C-D (N)	M-L (N)	V-V (Nm)	I-E (Nm)	F-E (°)
Anterior	Uni	7.4	13.3	2.4	0.7	0.1	0.2
	Multi	7.9	21.5	3.4	0.8	0.4	0.2
Posterior	Uni	6.7	12.3	2.0	0.7	0.2	0.2
	Multi	6.2	21.7	5.0	0.5	0.3	0.2
Varus	Uni	1.8	26.6	2.6	3.0	0.3	0.2
	Multi	3.0	29.7	4.5	2.1	0.4	0.3
Valgus	Uni	2.8	28.3	3.5	2.6	0.3	0.2
	Multi	3.6	32.9	6.1	2.3	0.5	0.2

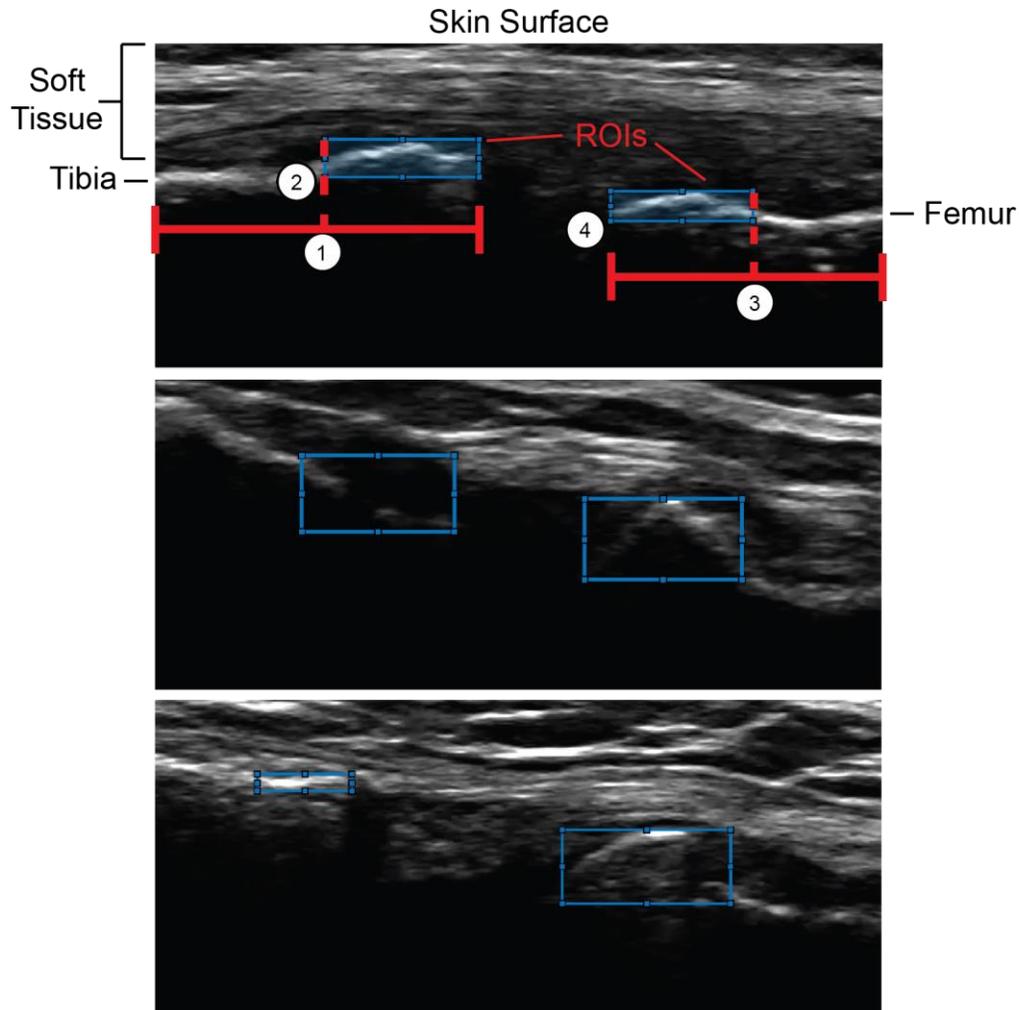
We computed RMSEs in tracking loads to be under 8 N in both anterior-posterior and medial-lateral loads. However, errors in compression-distraction (C-D) tracking were larger, with RMSEs up to 32.9 N. This is likely due to the much larger sensitivity of C-D, as even translations of less than a millimeter can cause increases/decreases of hundreds of newtons. For torques, we computed RMSEs to be under 3 Nm for varus-valgus and 0.5 Nm for internal-external rotation. Interestingly, errors in V-V were larger for the V-V uni-directional loading assessments compared to the multi-directional loading assessments, whereas this relationship was opposite for all of the other forces/torques for the V-V assessments. Our hypothesis is the multi-directional loading case is the more natural path for the knee to go into V-V torque, with small off-axis loads, rather than the pure V-V angular rotation. This more natural trajectory could be less stiff than the pure V-V angular rotation, allowing for better tracking of the prescribed loads. Finally, errors in prescribing flexion-extension were low, with maximum RMSEs reaching only 0.3° for all assessments.

## Supplement C: Reproducibility of ROI Selection

To determine the repeatability and reproducibility variances of the within-user and between-user variance, respectively, of selecting the bone ROIs, we used the uni- and multi-directional-loading assessments at 20° of knee flexion for all DOFs (eight assessments per specimen). The lead author (MBB) gave high-level directions on the method to select the ROIs to three users (**Figure S.5**). Each user had up to three practice trials of selecting ROIs in a random assessment to ensure they could correctly identify the bones. We duplicated the eight assessments two more times to obtain three trials per assessment. We put the 24 trials per specimen in a random order and each user independently selected the initial ROIs for the bone-tracking algorithm. For each of these additional trials, we used our US-based bone-tracking algorithm to estimate kinematics.

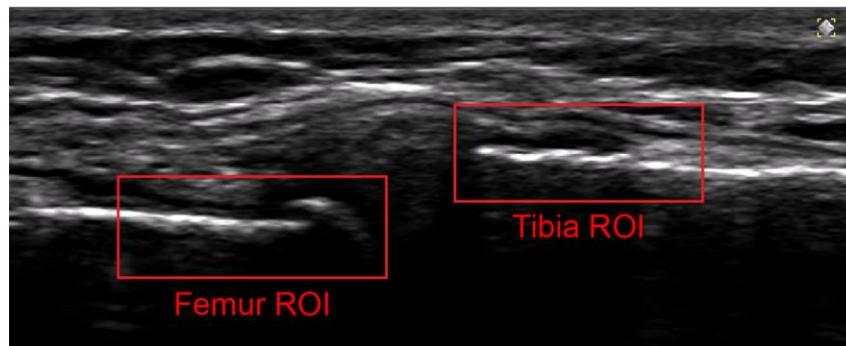
The goal is to enclose the inside halves of each bone

- ① Visually split left bone in half in left-right direction
- ② Click and drag from one corner to the other to enclose the inside half of the bone
- ③ Repeat step 1 for right bone
- ④ Repeat step 2 for right bone



**Figure S.5:** We showed the following directions and example ROI selections to each user before the user selected the bone ROIs for each assessment. Then, we ran our ultrasound-based bone-tracking algorithm to compute the laxity in each assessment.

We chose to create ROIs that split the bones in the middle in the left-right direction in case a trial had a large V-V movement that would cause the ROI to leave the total frame. Additionally, in order to capture the contrast of the edges of the bones, we increased the ROIs in the top (superficial), bottom (deep), and inside (towards the joint line) directions by 10, 20, and 20 pixels, respectively. As a reminder, the pixel resolution in both directions was 0.051 mm/pixel for each trial. We chose these values during pilot testing to give a balance of capturing the bright/dark contrast while minimizing the amount of soft tissue around the bones that were captured in the ROIs (**Figure S.6**).



**Figure S.6:** After selecting the regions of interest (ROIs) around the bones, we expanded them by 10, 20, and 20 pixels in the top (superficial), bottom (deep), and inside (towards the joint line) directions to capture the bright/dark contrast of the frame while minimizing the amount of soft tissue around the bones in the ROIs.

We computed the repeatability variance and reproducibility variance for each assessment to assess repeatability of measuring joint laxity when selecting bone ROIs within a user and between users, respectively. To compute the repeatability variance, we used

**Equation S.1** from the ISO 5725 standard:

$$s_{rj}^2 = \frac{\sum_{i=1}^p (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^p (n_{ij} - 1)} \quad \text{Equation S.1}$$

Where  $s_{rj}^2$  is the repeatability variance for each laxity assessment ( $j$ ),  $p$  is the number of users (i.e., 3),  $n_i$  is the number of repeated trials per laxity assessment (i.e., 3), and  $s_{ij}^2$  is the variance in laxity measurements of the three repeated trials in a laxity assessment.

To compute the between-user variance, we used **Equation S.2** from the ISO 5725 standard:

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{n}_j} \quad \text{Equation S.2}$$

Where  $s_{Lj}^2$  is the between-user variance for each laxity assessment ( $j$ ),  $s_{rj}^2$  is the repeatability variance for each laxity assessment (**Equation S.1**), and  $s_{dj}^2$  and  $\bar{n}_j$  are calculated in **Equation S.3** and **Equation 4**:

$$s_{dj}^2 = \frac{1}{p-1} \sum_{i=1}^p n_{ij} (\bar{y}_{ij} - \bar{y}_j)^2 \quad \text{Equation S.3}$$

$$\bar{n}_j = \frac{1}{p-1} \left[ \sum_{i=1}^p n_{ij} - \frac{\sum_{i=1}^p n_{ij}^2}{\sum_{i=1}^p n_{ij}} \right] \quad \text{Equation S.4}$$

Where  $\bar{y}_{ij}$  is the mean of the measured kinematics over the three repeated trials in a laxity assessment and  $\bar{y}_j$  is the mean of the measured kinematics over all the trials (i.e., for all three users) for that laxity assessment.

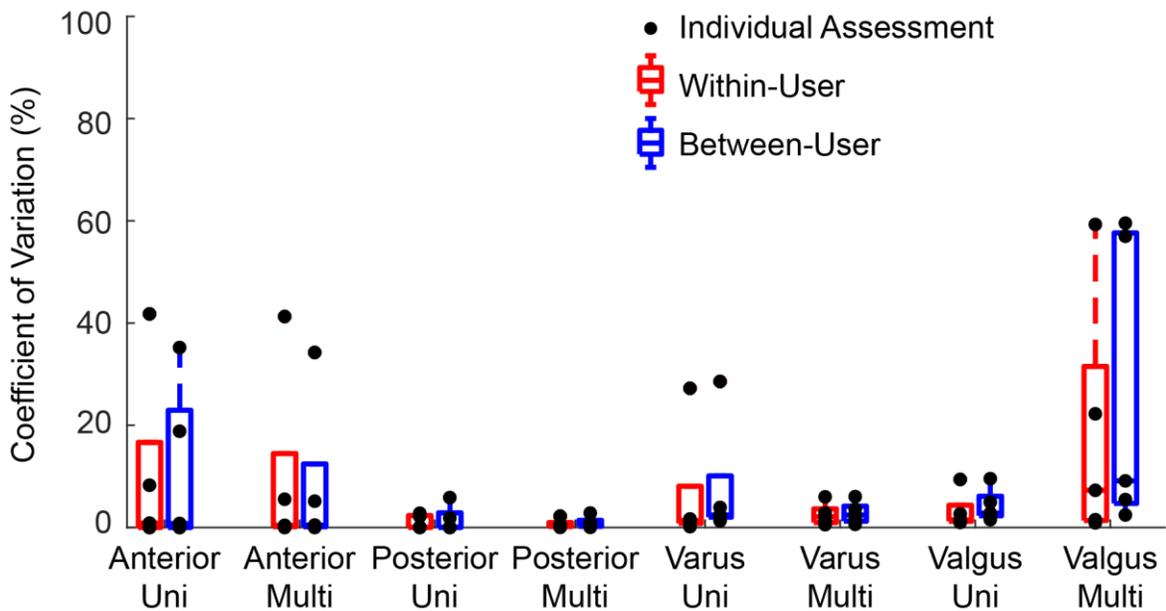
Next, we computed the reproducibility variance for each assessment to assess the reproducibility of measuring joint laxity between different users (**Equation S.5**):

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \quad \text{Equation S.5}$$

Finally, to put these variances into context of the measured kinematics, we computed the coefficient of variation for each repeatability and reproducibility variance using **Equation S.6**:

$$CV_{repeatability,j} = \frac{\sqrt{s_{rj}^2}}{\bar{y}_j} \text{ and } CV_{reproducibility,j} = \frac{\sqrt{s_{Rj}^2}}{\bar{y}_j} \quad \text{Equation S.6}$$

Where  $CV_{repeatability,j}$  is the coefficient of variation within a user for each laxity assessment,  $s_{rj}^2$  is the repeatability variance for each laxity assessment,  $CV_{reproducibility,j}$  is the coefficient of variation between users for each laxity assessment, and  $s_{Rj}^2$  is the reproducibility variance for each laxity assessment. The results are shown in **Figure S.7** below.



**Figure S.7:** We computed the coefficients of variances (CV) of measuring laxity when selecting regions of interest within a user (Within-User) and between different users (Between-User). The CVs of each specimen and each assessment are plotted as scatter points, and box plots are created to obtain the medians and the 25% and 75% percentiles. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

Generally, the coefficients of variation (CVs) within a user were very similar to those between users, as a majority of the CVs were very low with the median coefficients of variation

to be between 0.04 and 7.3% within a user and 0.04 and 9.1% between users. These results indicate selecting ROIs are largely robust to the user selecting the ROI, even after the user has a very brief training on how to select the ROIs.

#### Supplement D: Equations to Compute Reproducibility

We computed the repeatability variance and reproducibility variance for each assessment to assess repeatability of measuring joint laxity in the same position and to assess the reproducibility of measuring joint laxity when the transducer position changes, respectively. To compute the repeatability variance, we used **Equation S.7** from the ISO 5725 standard:

$$s_{rj}^2 = \frac{\sum_{i=1}^p (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^p (n_{ij} - 1)} \quad \text{Equation S.7}$$

Where  $s_{rj}^2$  is the repeatability variance for each laxity assessment ( $j$ ),  $p$  is the number of positions (i.e., 3 for V-V and 2 for A-P),  $n_i$  is the number of trials per position (i.e., 3), and  $s_{ij}^2$  is the variance in laxity measurements of the three trials at the same position for each combination of loading trajectory-loading direction combination.

To compute the between-position variance, we used **Equation S.8** from the ISO 5725 standard:

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{n}_j} \quad \text{Equation S.8}$$

Where  $s_{Lj}^2$  is the between-position variance for each laxity assessment ( $j$ ),  $s_{rj}^2$  is the repeatability variance for each laxity assessment (**Equation S.7**), and  $s_{dj}^2$  and  $\bar{n}_j$  are calculated in **Equation S.9** and **Equation S.10**:

$$s_{dj}^2 = \frac{1}{p-1} \sum_{i=1}^p n_{ij} (\bar{y}_{ij} - \bar{y}_j)^2 \quad \text{Equation S.9}$$

$$\bar{n}_j = \frac{1}{p-1} \left[ \sum_{i=1}^p n_{ij} - \frac{\sum_{i=1}^p n_{ij}^2}{\sum_{i=1}^p n_{ij}} \right] \quad \text{Equation S.10}$$

Where  $\bar{y}_{ij}$  is the mean of the measured kinematics over the three trials at the same position and  $\bar{y}_j$  is the mean of the measured kinematics over all the trials for that laxity assessment.

Next, we computed the reproducibility variance for each assessment to assess the reproducibility of measuring joint laxity when repositioning the ultrasound transducer (**Equation S.11**):

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \quad \text{Equation S.11}$$

Finally, to put these variances into context of the measured kinematics, we computed the coefficient of variation for each repeatability and reproducibility variance using **Equation S.12**:

$$CV_{repeatability,j} = \frac{\sqrt{s_{rj}^2}}{\bar{y}_j} \text{ and } CV_{reproducibility,j} = \frac{\sqrt{s_{Rj}^2}}{\bar{y}_j} \quad \text{Equation S.12}$$

Where  $CV_{repeatability,j}$  is the coefficient of variation within a position for laxity assessment,  $s_{rj}^2$  is the repeatability variance for each laxity assessment,  $CV_{reproducibility,j}$  is the coefficient of variation between positions for each laxity assessment, and  $s_{Rj}^2$  is the reproducibility variance for each laxity assessment.

## Supplement E: Bias, Precision, and Root-Mean-Square Errors for Predicting Kinematics

We computed values of using our ultrasound-based bone-tracking algorithm to estimate kinematics of laxity assessments at different joint velocities. Specifically, we evaluated at the experimental speed (Case 1; **Table S.2**), the relative velocity during a standard, one-second

laxity assessment (Case 2, **Table S.3**), the relative velocity between the femur and tibia during healthy gait (Case 3, **Table S.4**), and the relative velocity between the femur and tibia in a pathologic state (Case 4, **Table S.5**).

**Table S.2:** Bias, precision, and root-mean-square errors (RMSE) for estimating kinematics for each loading trajectory during loading velocities at the experimental speed. The last column contains the maximum translations/rotations measured by the robot (mean  $\pm$  standard deviation) for each loading trajectory. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

		Bias	Precision	RMSE	Maximum Motion
Anterior	Uni	-0.46 mm	2.10 mm	2.15 mm	4.4 $\pm$ 2.3 mm
	Multi	-1.00 mm	1.84 mm	2.10 mm	5.6 $\pm$ 1.9 mm
Posterior	Uni	-0.58 mm	0.94 mm	1.10 mm	3.4 $\pm$ 1.6 mm
	Multi	-0.06 mm	0.63 mm	0.64 mm	3.0 $\pm$ 1.2 mm
Varus	Uni	0.79°	0.88°	1.18°	3.3 $\pm$ 1.3°
	Multi	0.33°	0.68°	0.76°	1.9 $\pm$ 0.6°
Valgus	Uni	-0.07°	0.39°	0.40°	2.9 $\pm$ 0.7°
	Multi	-0.32°	1.21°	1.25°	2.3 $\pm$ 1.2°

**Table S.3:** Bias, precision, and root-mean-square errors (RMSE) for estimating kinematics for each loading trajectory during loading velocities mimicking a one-second laxity assessment. The last column contains the maximum translations/rotations measured by the robot (mean  $\pm$  standard deviation) for each loading trajectory. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

		Bias	Precision	RMSE	Maximum Motion
Anterior	Uni	-0.46 mm	2.09 mm	2.14 mm	4.4 $\pm$ 2.3 mm
	Multi	-0.82 mm	1.27 mm	1.51 mm	5.6 $\pm$ 1.9 mm
Posterior	Uni	-0.57 mm	0.94 mm	1.09 mm	3.4 $\pm$ 1.6 mm
	Multi	-0.05 mm	0.64 mm	0.64 mm	3.0 $\pm$ 1.2 mm
Varus	Uni	0.79°	0.88°	1.18°	3.3 $\pm$ 1.3°
	Multi	0.32°	0.67°	0.74°	1.9 $\pm$ 0.6°
Valgus	Uni	-0.06°	0.39°	0.39°	2.9 $\pm$ 0.7°
	Multi	-0.31°	1.21°	1.25°	2.3 $\pm$ 1.2°

**Table S.4:** Bias, precision, and root-mean-square errors (RMSE) for estimating kinematics for each loading trajectory during loading velocities mimicking relative motion of the knee during healthy gait. The last column contains the maximum translations/rotations measured by the robot (mean  $\pm$  standard deviation) for each loading trajectory. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

		Bias	Precision	RMSE	Maximum Motion
Anterior	Uni	-1.13 mm	2.44 mm	2.68 mm	4.4 $\pm$ 2.3 mm
	Multi	-1.12 mm	2.16 mm	2.43 mm	5.6 $\pm$ 1.9 mm
Posterior	Uni	-1.27 mm	2.33 mm	2.65 mm	3.4 $\pm$ 1.6 mm
	Multi	-0.12 mm	0.93 mm	0.94 mm	3.0 $\pm$ 1.2 mm
Varus	Uni	0.87°	1.31°	1.57°	3.3 $\pm$ 1.3°
	Multi	0.39°	0.82°	0.91°	1.9 $\pm$ 0.6°
Valgus	Uni	-0.06°	0.38°	0.39°	2.9 $\pm$ 0.7°
	Multi	-0.13°	1.28°	1.29°	2.3 $\pm$ 1.2°

**Table S.5:** Bias, precision, and root-mean-square errors (RMSE) for estimating kinematics for each loading trajectory during loading velocities mimicking relative motion of the knee during pathologic gait. The last column contains the maximum translations/rotations measured by the robot (mean  $\pm$  standard deviation) for each loading trajectory. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

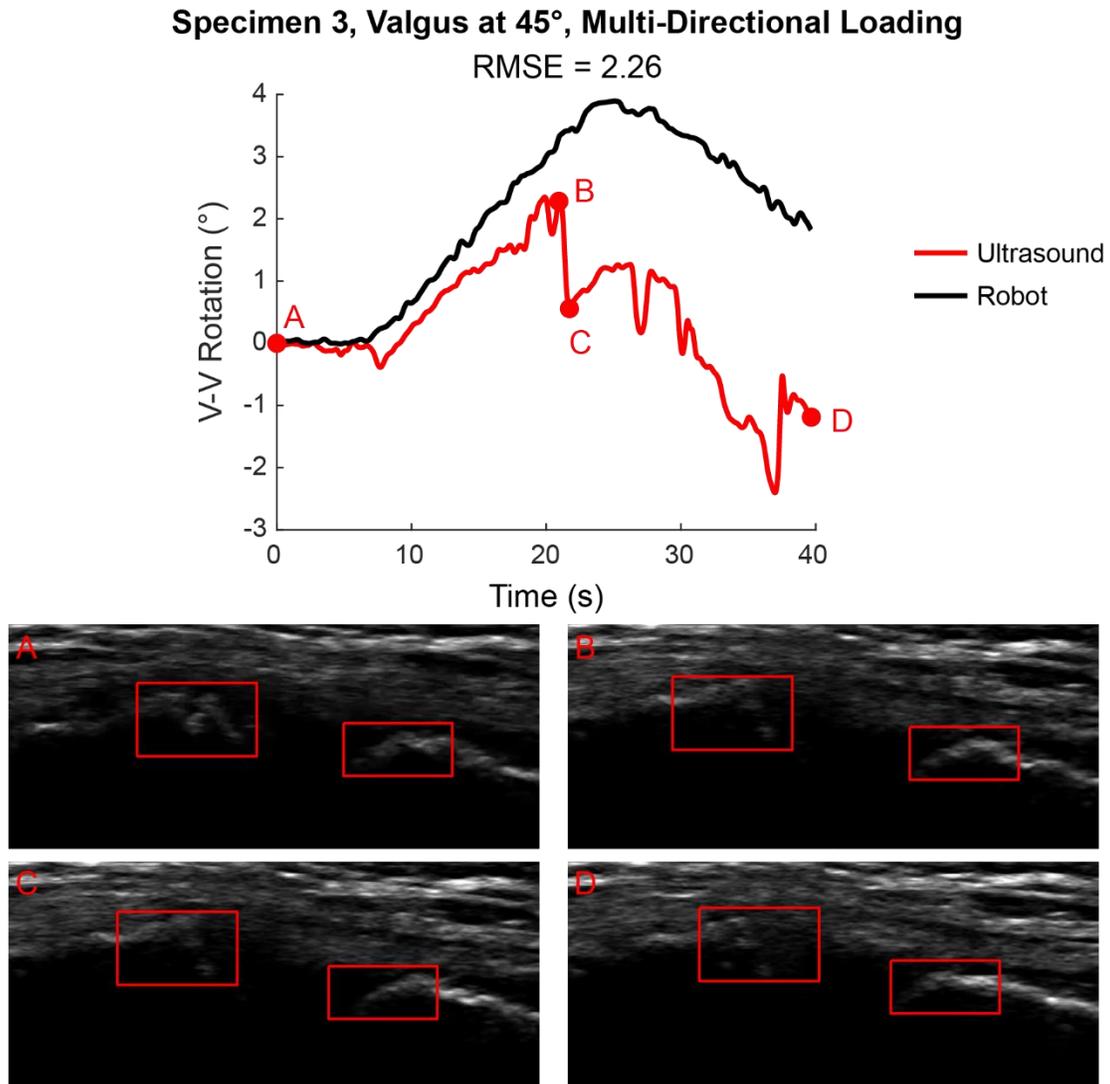
		Bias	Precision	RMSE	Maximum Motion
Anterior	Uni	-1.05 mm	2.35 mm	2.57 mm	4.4 $\pm$ 2.3 mm
	Multi	-1.09 mm	2.09 mm	2.35 mm	5.6 $\pm$ 1.9 mm
Posterior	Uni	-1.19 mm	2.26 mm	2.55 mm	3.4 $\pm$ 1.6 mm
	Multi	-0.28 mm	0.88 mm	0.92 mm	3.0 $\pm$ 1.2 mm
Varus	Uni	0.83°	1.30°	1.54°	3.3 $\pm$ 1.3°
	Multi	0.30°	0.60°	0.67°	1.9 $\pm$ 0.6°
Valgus	Uni	-0.05°	0.38°	0.38°	2.9 $\pm$ 0.7°
	Multi	-0.32°	1.24°	1.27°	2.3 $\pm$ 1.2°

### Supplement F: Further Details into Trials with High Errors

We took a more detailed look into the trials with higher errors in order to help us understand what could be the cause of the high errors. As mentioned in the limitations of the **Discussion**, some trials results in high errors potentially due to artifacts within the bone ROIs that led to poor cross-correlation tracking. Below is a more detailed analysis of three trials, with plots of robot and ultrasound kinematics, along with images of some ultrasound frames at specific time points.

(1) Example Trial 1

Below is an example trial of a valgus, multi-directional-loading assessment in specimen 3 (**Figure S.8**). This trial had a root-mean-square error of  $2.26^\circ$ , and likely did not track the bones well throughout the trial due to the contours of the bone changing shape as well as the close proximity of surrounding tissue around the bones.



**Figure S.8:** (Top) Ultrasound vs robot kinematics for a trial that had poor tracking due to shape changes of the bones. (Bottom) We extracted specific frames (Frames A-D) from the

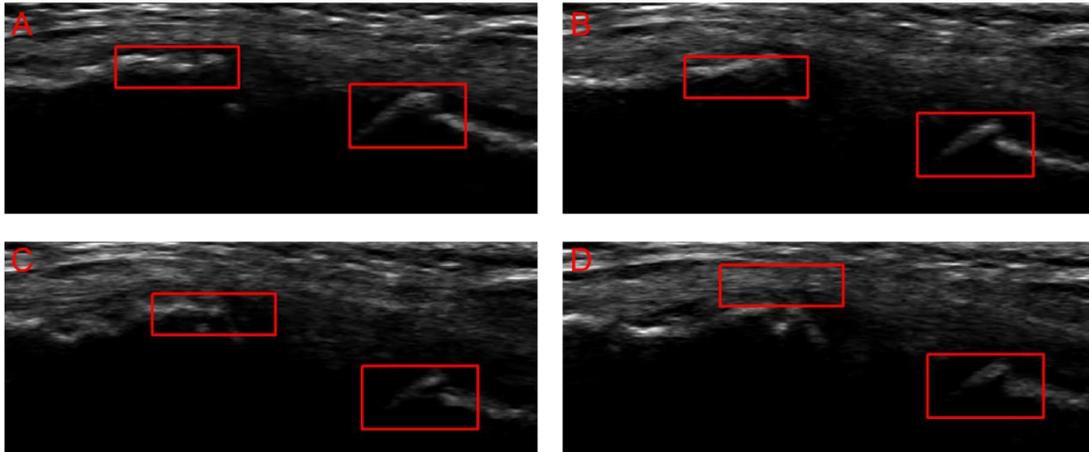
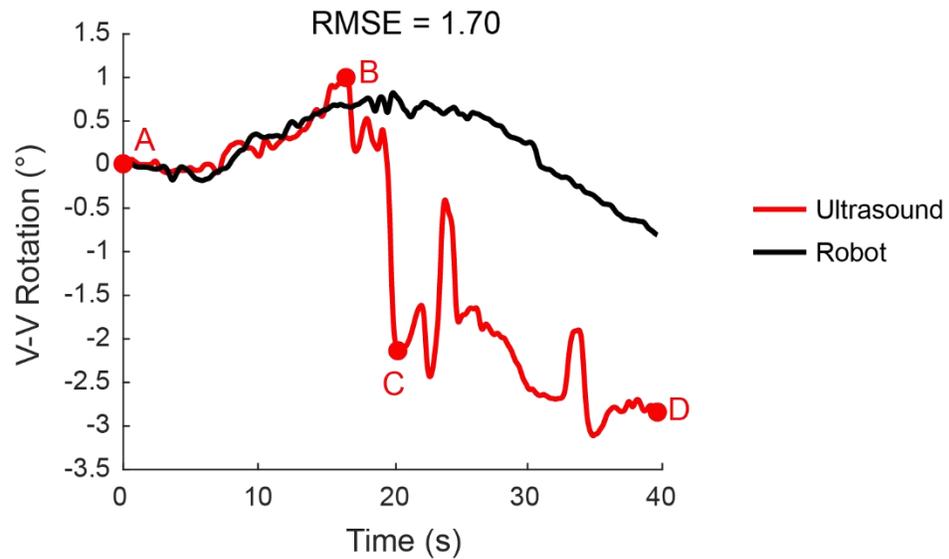
ultrasound cine loop to identify potential reasons that the errors in estimating kinematics were high.

In frame A, we create the ROIs. By frame B, the tibia (left ROI) and femur (right ROI) are starting to change shape. At frame C, the femur ROI shifts inward (you may notice the larger black space in the inside of the right ROI), leading to a sharp decrease in estimated V-V rotation. Additionally, the tibia's ROI is obscured with the superficial tissue in close proximity to the tibia, which also shifts the tibia ROI inward. These ROIs never get back to looking like they did in frame A, as frame D still predicts a negative rotation.

### *(2) Example Trial 2*

Below is another example trial of a valgus, multi-directional-loading assessment in specimen 3 (**Figure S.9**). This trial had a root-mean-square error of  $1.70^\circ$ , and likely did not track the bones well throughout the trial due to the contours of the bone changing shape as well as close proximity of surrounding tissue around the bones.

### Specimen 3, Valgus at 20°, Multi-Directional Loading



**Figure S.9:** (Top) Ultrasound vs robot kinematics for a trial that had poor tracking due to surrounding soft tissue around the bones. (Bottom) We extracted specific frames (Frames A-D) from the ultrasound cine loop to identify potential reasons that the errors in estimating kinematics were high.

In frame A, we create the ROIs. By frame B, the tibia (left ROI) is starting to shift downward, potentially due to the close proximity of the superficial tissue. By frame C, the tibia ROI shifts inward (you may notice the larger black space in the inside of the left ROI), leading to a sharp decrease in estimated V-V rotation. The ROIs never get back to surrounding the bones

as they do in frame A, and the tibia ROI shifts entirely to the superficial tissue by frame D, leading to poor estimates of V-V kinematics.

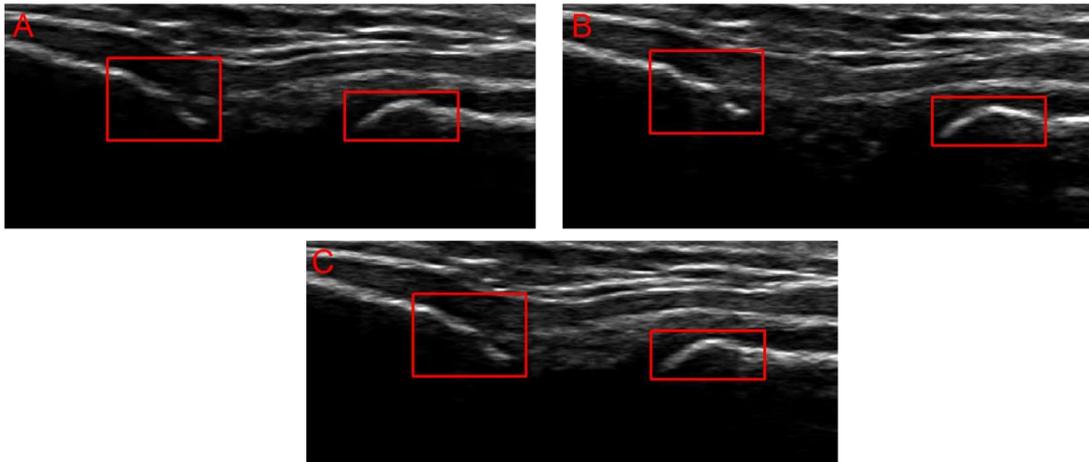
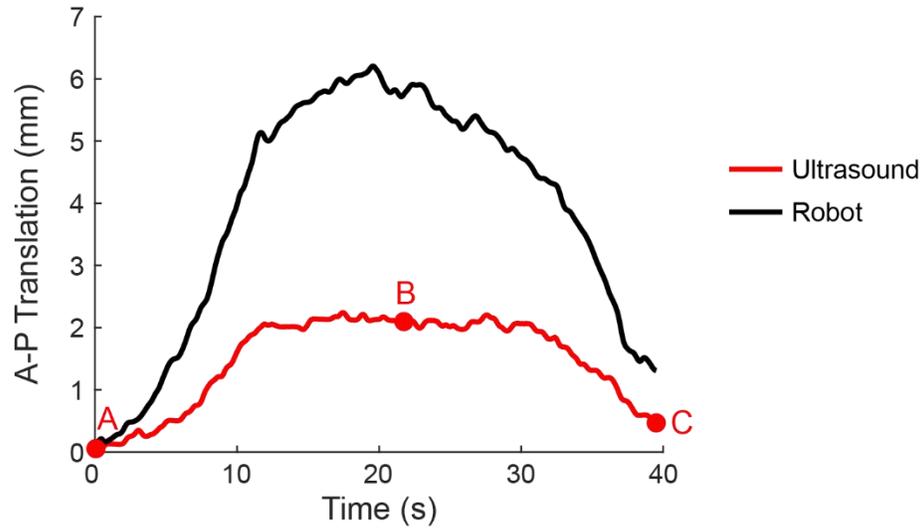
### *(3) Example Trial 3*

Finally, some trials resulted in high errors even though the cross-correlation method tracked the bones well. A potential source of this error may be in converting distance changes in the B-mode ultrasound frames to kinematics of the joint. For A-P assessments, we needed to account for the oblique angle of the transducer relative to the A-P axis since we had to move the transducer to the side of the patella to image the femur and the tibia. Therefore, to convert the distance changes measured from ultrasound to A-P kinematics, we divided by the angle between the transducer and the anterior-posterior axis of the tibia. Likely, tracking bone motion at an oblique angle to the A-P axis increases the errors in A-P kinematics. For V-V trials, we assumed a rotation axis about the center of the joint [8]. In reality, though, the rotation axis likely changes with different V-V angles [8], and potentially different knee geometries (e.g., cartilage wear in one compartment).

Below is an anterior assessment with a root-mean-square error of  $2.68^\circ$  that tracked the bones well throughout the trial, but still had larger errors (**Figure S.10**).

### Specimen 1, Anterior at 45°, Multi-Directional Loading

RMSE = 2.68

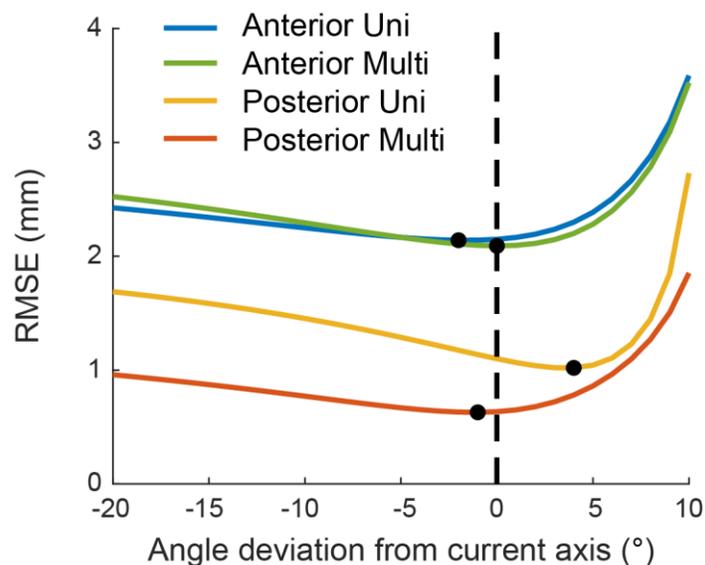


**Figure S.10:** (Top) Ultrasound vs robot kinematics for a trial that had good tracking, but high errors in kinematics. (Bottom) We extracted specific frames (Frames A-D) from the ultrasound cine loop to identify potential reasons that the errors in estimating kinematics were high.

In frame A, we create the ROIs. In frames B and C, the shape of the bone are similar to the first frame, and the ROIs around the bone are still surrounding the same portion of the bone. Thus, this is a trial that tracked the bone well, but still under-predicted the A-P kinematics.

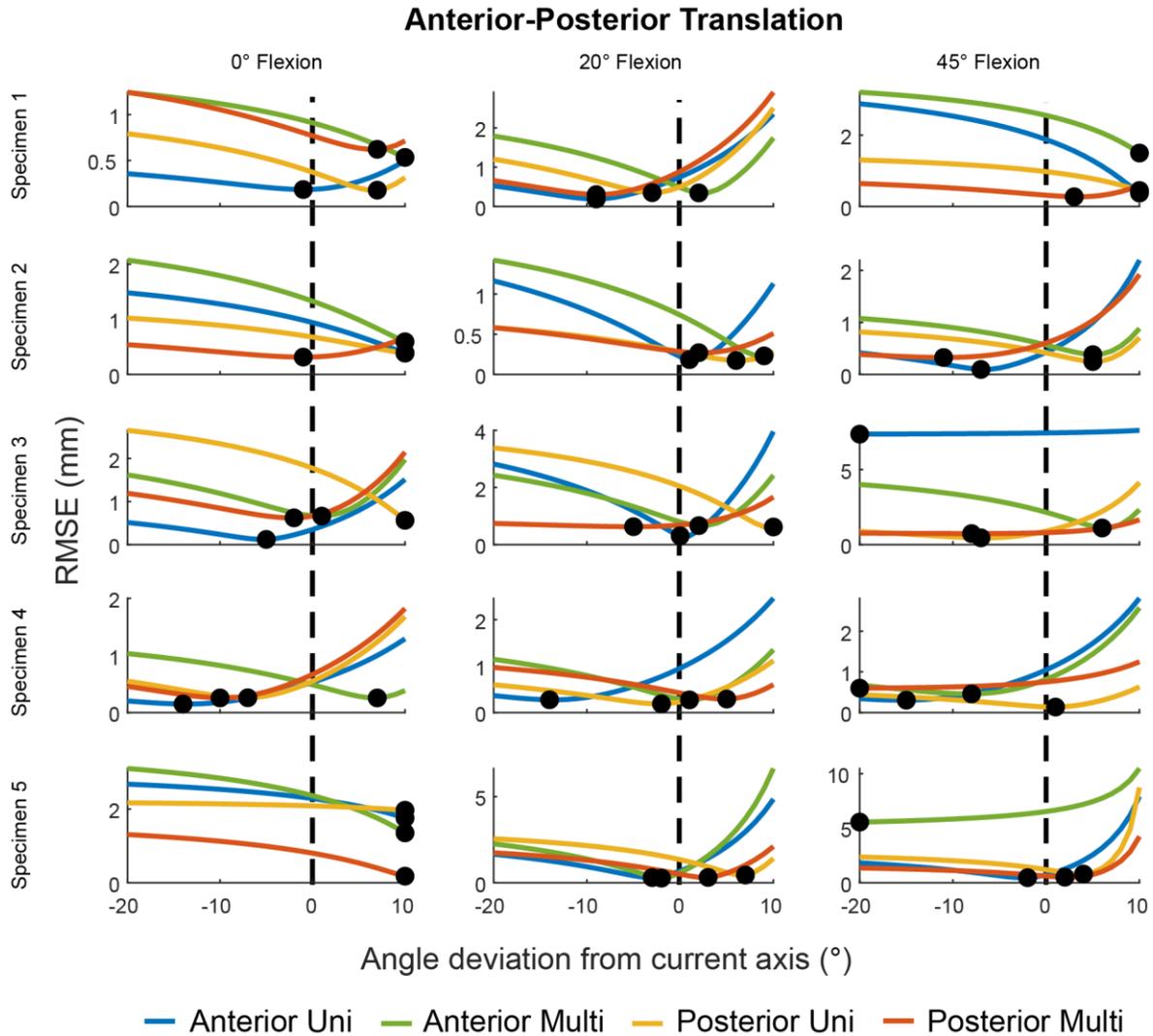
## Supplement G: Effects of Altering Planes of Motion

A potential source of error in our ultrasound-based bone-tracking algorithm is the conversion of distance changes to kinematics. For A-P assessments, we needed to account for the oblique angle of the transducer relative to the A-P axis since we had to move the transducer to the side of the patella to image the femur and the tibia. Therefore, to convert the distance changes measured from ultrasound to A-P kinematics, we divided by the angle between the transducer and the anterior-posterior axis of the tibia. However, we wanted to explore processing the data at varying angles from the transducer angle to observe if a trend could be obtained that would lower the errors. Thus, we reprocessed the data, but when accounting for the angle of the transducer and the A-P axis of the tibia, we added an offset angle. We plotted the root-mean-square errors as a function of this angle deviation, both when all angles and specimens were pooled within an assessment (**Figure S.11**) and for each individual trial (**Figure S.12**).



**Figure S.11:** We plotted root-mean-square errors as a function of the angle deviation from the current transducer axis. Data was pooled by assessment, so fifteen different trials are combined for each curve (three angles and five specimens per assessment). Black scatter

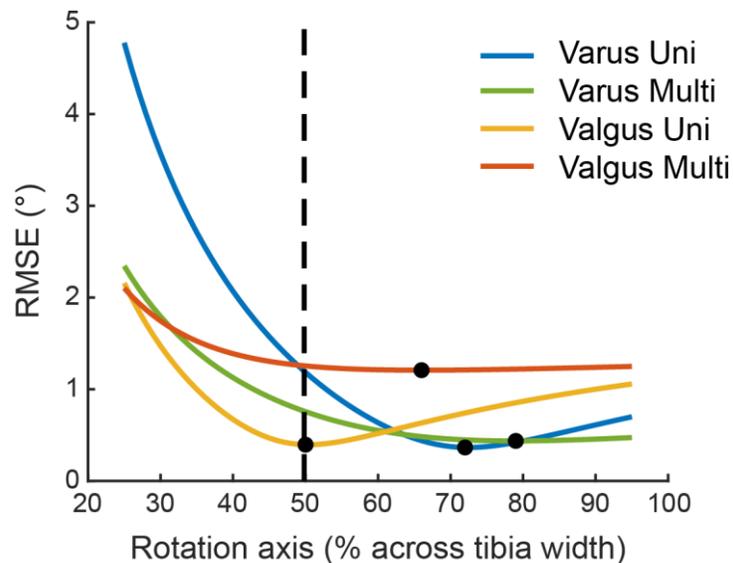
points represent the minimum values of each curve. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.



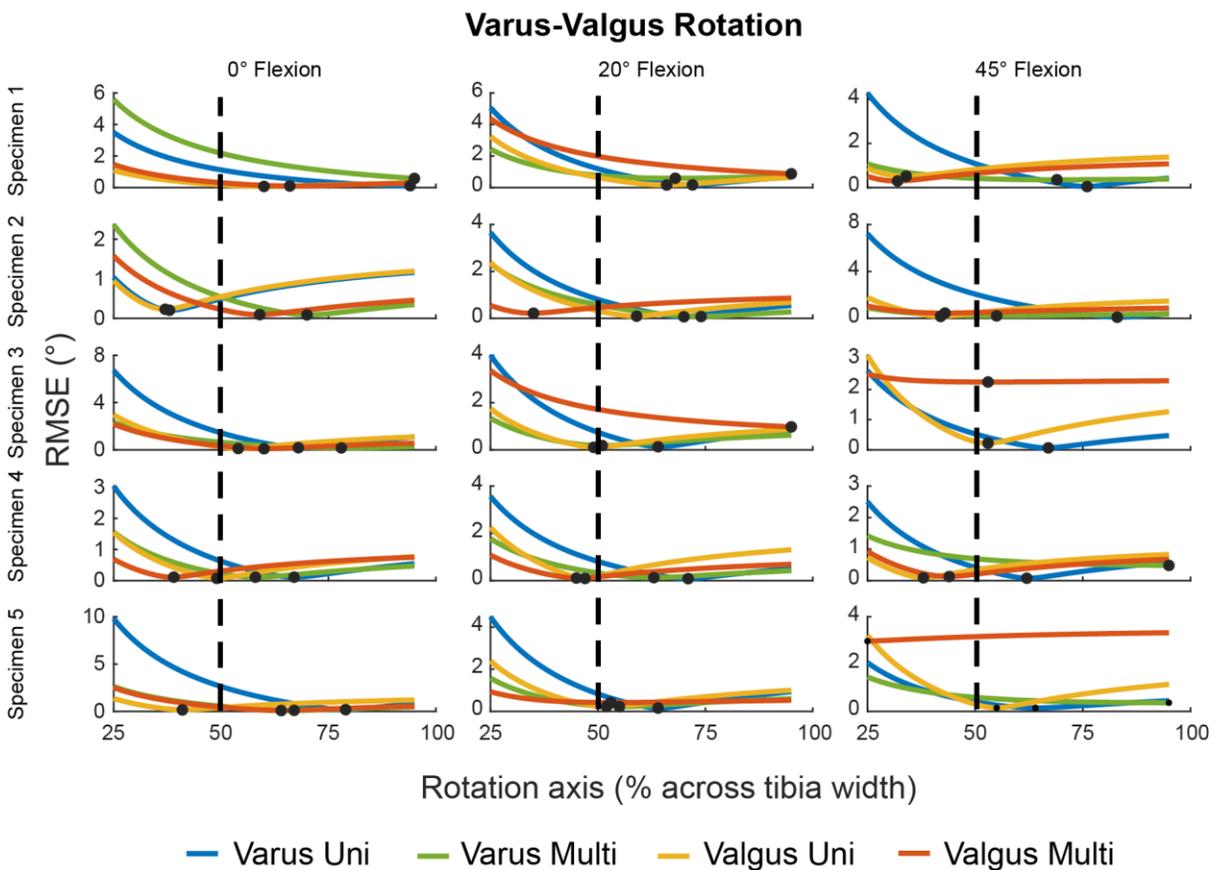
**Figure S.12:** We plotted root-mean-square errors as a function of the angle deviation from the current transducer axis for each trial. Each row represents a different specimen, and each column represents a different flexion angle. Black scatter points represent the minimum values of each curve. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

Although the minimum RMSE of each assessment differed slightly from the 0° angle deviation used previously, the differences were negligible. Additionally, no trends were observed between different flexion angles and different specimens.

For V-V trials, we assumed a rotation axis about the center of the joint [8]. In reality, though, the rotation axis likely changes with different V-V angles [8], and potentially different knee geometries (e.g., cartilage wear in one compartment). Thus, we conducted a supplemental study to show the effect of using different locations of the V-V axis on estimating V-V kinematics. To do this, we reprocessed the data and altered the rotation axis about the joint as a function of the percent distance across the tibia joint. A value of 50 represents the midpoint of the tibial plateau and a value above 50 represents a rotation axis about the opposite compartment that is opening (i.e., medial compartment for varus and lateral compartment for valgus). We plotted the root-mean-square errors as a function of this rotation axis, both when all angles and specimens were pooled within an assessment (**Figure S.13**) and for each individual trial (**Figure S.14**).



**Figure S.13:** We plotted root-mean-square errors as a function of the V-V rotation axis for each trial. Data was pooled by assessment, so fifteen different trials are combined for each curve (three angles and five specimens per assessment). Black scatter points represent the minimum values of each curve. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.



**Figure S.14:** We plotted root-mean-square errors as a function of the V-V rotation axis for each trial. Each row represents a different specimen, and each column represents a different flexion angle. Black scatter points represent the minimum values of each curve. Uni = uni-directional-loading assessments and Multi = multi-directional loading assessments.

When pooling the data within each laxity assessment, RMSEs generally decreased as the rotation axis across the tibia width increased. Therefore, a rotation axis around 75% (i.e., approximately in the center of the medial compartment during varus loading and approximately in the center of the lateral compartment during valgus loading) would give lower errors than the 50% location in all assessments except for the valgus uni-directional loading assessments. However, as previously stated, the rotation axis likely changes based on many factors (e.g., specimen, V-V angle), as seen in **Figure S.14**.

## References

1. Arant LR, Laudon ME, Roth JD (2023) Shear Wave Tensiometers Can Measure In Situ Ligament Tension and Engagement Length With Low Bias Errors. Orthopaedic Research Society Conference Dallas, Texas
2. ASTM E (2008) Standard practice for use of the terms precision and bias in ASTM test methods. ASTM International
3. Benoit DL, Ramsey DK, Lamontagne M, Xu L, Wretenberg P, Renström P (2006) Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo. *Gait & Posture* 24(2):152–164
4. Bhargavan M (2008) TRENDS IN THE UTILIZATION OF MEDICAL PROCEDURES THAT USE IONIZING RADIATION. *Health Physics* 95(5):612–627
5. Bhimani R, Ashkani-Esfahani S, Mirochnik K, Lubberts B, DiGiovanni CW, Tanaka MJ (2022) Utility of Diagnostic Ultrasound in the Assessment of Patellar Instability. *Orthop J Sports Med* 10(5):23259671221098748
6. Chang AH, Chmiel JS, Moisio KC, Almagor O, Zhang Y, Cahue S, Sharma L (2013) Varus thrust and knee frontal plane dynamic motion in persons with knee osteoarthritis. *Osteoarthritis and Cartilage* 21(11):1668–1673
7. Chen C-H, Li J-S, Hosseini A, Gadikota HR, Gill TJ, Li G (2012) Anteroposterior stability of the knee during the stance phase of gait after anterior cruciate ligament deficiency. *Gait & Posture* 35(3):467–471
8. Dhaher YY, Francis MJ (2006) Determination of the abduction–adduction axis of rotation at the human knee: Helical axis representation. *Journal of Orthopaedic Research* 24(12):2187–2200
9. van der Esch M, Steultjens M, Ostelo RWJG, Harlaar J, Dekker J (2006) Reproducibility of instrumented knee joint laxity measurement in healthy subjects. *Rheumatology* 45(5):595–599
10. Fleming BC, Brattbakk B, Peura GD, Badger GJ, Beynon BD (2002) Measurement of anterior–posterior knee laxity: a comparison of three techniques. *Journal of Orthopaedic Research* 20(3):421–426
11. Foroughi N, Smith RM, Lange AK, Baker MK, Singh MAF, Vanwanseele B (2010) Dynamic alignment and its association with knee adduction moment in medial knee osteoarthritis. *The Knee* 17(3):210–216
12. Giordano BD, Grauer JN, Miller CP, Morgan TL, Rehtine GRI (2011) Radiation Exposure Issues in Orthopaedics\*. *JBJS* 93(12):e69
13. Grood ES, Suntay WJ (1983) A Joint Coordinate System for the Clinical Description of Three-Dimensional Motions: Application to the Knee. *Journal of Biomechanical Engineering* 105(2):136–144
14. Guan S, Gray HA, Keynejad F, Pandy MG (2016) Mobile Biplane X-Ray Imaging System for Measuring 3D Dynamic Joint Motion During Overground Gait. *IEEE Transactions on Medical Imaging* 35(1):326–336
15. Hull ML (2020) Coordinate system requirements to determine motions of the tibiofemoral joint free from kinematic crosstalk errors. *Journal of Biomechanics* 109:109928
16. International Standards Organization (1994) Accuracy (Trueness and Precision) of Measurement Methods and Results - Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method.
17. James EW, Williams BT, LaPrade RF (2014) Stress Radiography for the Diagnosis of Knee Ligament Injuries: A Systematic Review. *Clin Orthop Relat Res* 472(9):2644–2657

18. Kappel A, Mortensen JF, Nielsen PT, Odgaard A, Laursen M (2020) Reliability of stress radiography in the assessment of coronal laxity following total knee arthroplasty. *The Knee* 27(1):221–228
19. Klasan A, Putnis SE, Kandhari V, Oshima T, Fritsch BA, Parker DA (2020) Healthy knee KT1000 measurements of anterior tibial translation have significant variation. *Knee Surg Sports Traumatol Arthrosc* 28(7):2177–2183
20. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15(2):155–163
21. Lopes C, Vilaca A, Rocha C, Mendes J (2023) Knee positioning systems for X-ray environment: a literature review. *Phys Eng Sci Med* 46(1):45–55
22. Malanga GA, Andrus S, Nadler SF, McLean J (2003) Physical examination of the knee: A review of the original test description and scientific validity of common orthopedic tests. *Archives of Physical Medicine and Rehabilitation* 84(4):592–603
23. Malcom LL, Daniel DM, Stone ML, Sachs R (1985) The measurement of anterior knee laxity after ACL reconstructive surgery. *Clin Orthop Relat Res*
24. McGraw KO, Wong SP Forming Inferences About Some Intraclass Correlation Coefficients.
25. Nagle TF, Erdemir A, Colbrunn RW (2021) A generalized framework for determination of functional musculoskeletal joint coordinate systems. *Journal of Biomechanics* 127:110664
26. Pedersen D, Vanheule V, Wirix-Speetjens R, Taylan O, Delpont Hp, Scheys L, Andersen Ms (2019) A novel non-invasive method for measuring knee joint laxity in four dof: In vitro proof-of-concept and validation. *Journal of Biomechanics* 82:62–69
27. Peterfy CG, Guermazi A, Zaim S, Tirman PFJ, Miaux Y, White D, Kothari M, Lu Y, Fye K, Zhao S, Genant HK (2004) Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis and Cartilage* 12(3):177–190
28. Peters A, Galna B, Sangeux M, Morris M, Baker R (2010) Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review. *Gait & Posture* 31(1):1–8
29. Reinschmidt C, van den Bogert AJ, Lundberg A, Nigg BM, Murphy N, Stacoff A, Stano A (1997) Tibiofemoral and tibio-calcaneal motion during walking: external vs. skeletal markers. *Gait & Posture* 6(2):98–109
30. Robnett NJ, Riddle DL, Kues JM (1995) Intertester Reliability of Measurements Obtained With the KT-1000 on Patients With Reconstructed Anterior Cruciate Ligaments. *J Orthop Sports Phys Ther* 21(2):113–119
31. Runer A, Roberti di Sarsina T, Starke V, Ilchev A, Felmet G, Braun S, Fink C, Csapo R (2021) The evaluation of Rolimeter, KLT, KiRA and KT-1000 arthrometer in healthy individuals shows acceptable intra-rater but poor inter-rater reliability in the measurement of anterior tibial knee translation. *Knee Surg Sports Traumatol Arthrosc* 29(8):2717–2726
32. Sernert N, Kartus J, Köhler K, Ejerhed L, Karlsson J (2001) Evaluation of the reproducibility of the KT-1000 arthrometer. *Scandinavian Journal of Medicine & Science in Sports* 11(2):120–125
33. Shultz SJ, Shimokochi Y, Nguyen A-D, Schmitz RJ, Beynon BD, Perrin DH (2007) Measurement of varus–valgus and internal–external rotational knee laxities in vivo—part i: assessment of measurement reliability and bilateral asymmetry. *Journal of Orthopaedic Research* 25(8):981–988
34. Slane LC, Slane JA, Scheys L (2017) The measurement of medial knee gap width using ultrasound. *Arch Orthop Trauma Surg* 137(8):1121–1128
35. Stäubli H-U, Jakob RP (1991) Anterior knee motion analysis: Measurement and simultaneous radiography. *Am J Sports Med* 19(2):172–177

36. Wiertsema SH, van Hooff HJA, Migchelsen LAA, Steultjens MPM (2008) Reliability of the KT1000 arthrometer and the Lachman test in patients with an ACL rupture. *The Knee* 15(2):107–110