Standardizing CT-Based Clinical Measurements of the Sacroiliac Joint: Assessment of a Novel Joint Width Method's Single-Rater Reliability and Variability.

Joshua Tandio[1], Sarah Mayer[1], Oluwatodimu R. Raji[1,2], and Jeremi M. Leasure[1,3]

[1]Medical Device Development, San Francisco, CA, USA, [2]UCSF Health St. Mary's Hospital, San Francisco, CA, USA, [3]CSU Department of Mechanical Engineering, Northridge, CA, USA

Email of Presenting Author: joshuatandio@mdevdev.com

Disclosures: The authors have no relevant financial or other competing interests to declare.

INTRODUCTION: The sacroiliac joint (SIJ) is central to load transfer between the spine and pelvis, and its dysfunction has been linked to low back pain. Standardized methods for quantifying SIJ width do not exist, and the reliability and variability of this measurement remain unknown. This study aimed to measure the variability and reproducibility of a CT-based workflow for SIJ width measurement by evaluating its intra-rater reliability in a larger effort to standardize these measurements. We hypothesized our method would have a good test-retest reliability with an intraclass correlation coefficient (ICC) score greater than 0.75 and good variability with a within-subject coefficient of variation (CV) lower than between-subjects CV.

METHODS: 62 Sacroiliac joint samples from 31 CT scans (18 male and 13 female) with spatial resolution in the axial plane ranging from 0.60 mm to 1.07 mm and slice thickness ranging from 0.4 mm to 3.0 mm were obtained from accredited tissue banks. A structured workflow (Figure 1a) using open-source software (3D Slicer) was used for pelvic alignment, SIJ regional marking (anterior-superior AS, anterior-inferior AI, posterior P) (Figure 1b), and cortical gap measurement (Figure 1c). A single rater performed three independent measurement trials at least 1 day apart to minimize recall bias. Intra-rater reliability was assessed using ICCs calculated from a two-way mixed effects model. Intra-rater variability was assessed by comparing within-subject and between-subjects CVs. Statistical analyses were performed in R statistical analysis software.

RESULTS: The intra-rater ICCs were 0.92 (AS), 0.85 (AI), and 0.84 (P). Figure 2 visualizes the agreement between pairs of trials in each region. The within-subject CVs were 8.6 ± 5.9 % (AS), 11.3 ± 8.6 % (AI), and 14.7 ± 8.9 % (P). The between-subjects CVs were 33.7% (AS), 30.8% (AI), and 37.2% (P). The within-subjects CV was lower than the between-subjects CV for each region, as displayed in Figure 3.

DISCUSSION: The standardized workflow yielded highly consistent results within a single rater, shown by the ICC scores greater than 0.75 and within-subject CVs lower than between-subjects CVs. The strengths of this study include a large clinically relevant sample size using software freely available to clinicians and researchers. Limitations include the absence of inter-rater testing, which will be necessary before broader standardization.

SIGNIFICANCE/CLINICAL RELEVANCE: Evaluating the reliability and variability of a novel SIJ width measurement protocol is critical for advancing both clinical assessment and research of the joint. This study provides a workflow for single raters and provides a benchmark for those raters to be considered proficient at measuring sacroiliac joint width. A reliable single-rater workflow provides a foundation for future multi-rater assessments and may enhance diagnostic accuracy and comparability across SIJ studies.
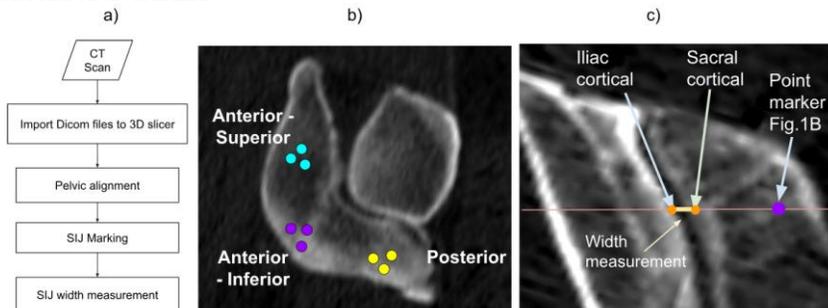
IMAGES AND TABLES:

Figure 1. a) SIJ measurement workflow; b) three zones of the SIJ and the three measurement markers; and c) SIJ width measurement example. Joint width was repeatedly measured at each of the three markers over three independent sessions.
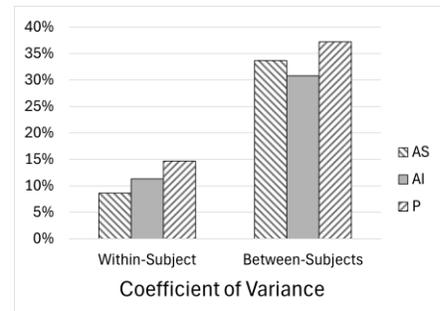

Figure 3. Variability of the SIJ width measurement. All data are represented as means. AS = Anterior-Superior, AI = Anterior-Inferior, and P = Posterior.
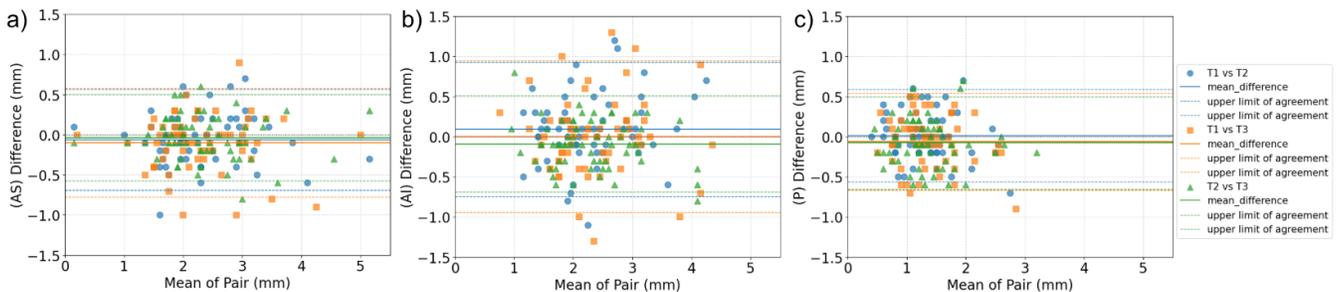

Figure 2. Bland-Altman plot for the intra-rater reliability: a) AS region width; b) AI region width; P region width. The plot illustrates the agreement between the three measurement trials presented as the pairwise comparison. AS = Anterior-Superior, AI = Anterior-Inferior, and P = Posterior. T1 = Session 1 joint width measures, T2 = Session 2 joint width measures, and T3 = Session 3 joint width measures.