Reliability Study For Assessing Proximal Femur Shape Variations In Cam-Type FAI With Statistical Shape Modeling

Siri Harish[1,2], Steven P Mell[1], Jorge Chahla[1], Shane J Nho[1], Catherine Yuh[1]
[1]Rush University, Chicago, IL, [2]University of Illinois at Chicago, Chicago, IL
catherine_yuh@rush.edu

Disclosures: SH (N), SPM (N), JC (3B- Arthrex Inc., CONMED Linvatec, Ossur, Smith & Nephew; 9- AAOS, AANA, ISAKOS), SJN (3B; Stryker. 5; Athletico. 7A; Allosource, Arthrex, Inc, Smith & Nephew, DJ Orthopaedics, Linvatec, Ossur, Stryker. 7B; Springer. 9; AOSSM, AANA), CY (N)

INTRODUCTION: Statistical shape modeling (SSM) is a technique that has seen increased use over the past decade to assess shape variations in the context of musculoskeletal pathology, such as cam femoroacetabular impingement syndrome (FAI). Existing literature on SSM for cam-type FAI has focused on assessing the spatial distribution of the cam deformity, length of cam protrusion, and differences in mean shape between cam FAI patients versus control subjects [1-12]. Often, principal component analysis (PCA) is used to assess variations in shape within a cohort of subjects. However, the sample sizes reported for 3D SSM studies in the context of FAI are highly variable, ranging from less than 30 to greater than 100 samples. The variability of sample sizes reported presents a challenge in interpreting SSM data across different groups, and determining whether findings are generalizable to the greater patient population. To date, no study has robustly assessed model sensitivity to sample size in SSM for cam FAI. Here, we present a random sampling based reliability study using a large cohort of FAI proximal femur models, by testing multiple iterations of random samples at varying sample size. The goal of this study is to determine a benchmark for sample size, when using SSM with PCA to assess variability in proximal femur morphology in cam FAI.

METHODS: For this institutional review board approved study, N=112 (32 Male/80 Female) pre-operative MRI scans of patients with FAI were used for this reliability study. 1.5T MRI scans were obtained using a 3D gradient dual-echo MRI sequence with two separate echo times. Using exported DICOMs of the pre-operative proximal femur MRI data, segmentations were performed in 3D Slicer to generate 3D surface meshes for each subject. In a previous study [13], the meshes were pre-processed in Geomagic Wrap and imported into ShapeWorks [14] as .stl files, where 2048 correspondence particles were optimized per subject. Procrustes scaling was applied to minimize the effect of femoral size in the statistical shape model (Figure 1). In this reliability study, the effect of sample size for statistical shape modeling was assessed by conducting randomized sampling experiments from the original set of 112 subjects. An initial set of experiments tested sample sizes of N=10, 25, 50, 75, and 100, with 10 randomized iterations per group. Based on preliminary results, additional experiments were performed at N=30, 35, 40, and 45 to more precisely determine the sample size at which model convergence was achieved. For each sample size and iteration, principal component analysis (PCA) was conducted, and the percentage of variance explained by each mode was recorded. In terms of convergence criteria, this was defined as the following: 1) for the first 10 principal components for each sample size group, if the individual explained variances of each principal component were all within ±1%, then convergence was achieved for those sample sizes, and 2) for the first principal component, the standard deviation of the 10 iterations ran for each sample size group has to be less than 1% to meet convergence criteria.

RESULTS: Each sub-sampled model produced N – 1 principal components, as expected (Figure 2). At smaller sample sizes (N=10, 25), the proportion of variance explained by PC1 varied widely across iterations, ranging from 65.10% to 84.23% at N=10 and 76.32% to 84.29% at N=25, reflecting instability in these model. An initial set of experiments at N=10, 25, 50, 75, and 100 demonstrated that N=50 was the lowest sample size where results were noticeably more consistent, with PC1 values clustered around 80% (range 78.69 - 81.23%, standard deviation 0.94%) across 10 iterations, and nearly indistinguishable from N=100. Based on these initial observations, additional runs were performed at N=30, 35, 40, and 45. These intermediate sampling sizes showed inconsistent behavior, with variability similar to the initial smaller sample sizes (e.g., N=45 range 74.46 - 81.90%, standard deviation 2.42%), further showing instability of models generated with fewer than 50 subjects. Collectively, these findings suggest that about 50 samples are required to achieve the convergence criteria defined for this study in PCA outcomes, as measured by reproducibility of the first principal component across iterations (Figure 3).

DISCUSSION: This study demonstrates that PCA-based SSM of the proximal femur in cam FAI patients is sensitive to cohort size, and that models built from fewer than 50 subjects may produce unstable estimates of the modes of variation. The high variability of the proportion of explained variance captured by PC1 at smaller sample sizes indicates that small cohorts may be prone to overemphasizing undesired features such as segmentation noise or outliers, rather than population-level morphology. The relative stability of PC1, and indeed all principal components at sample sizes greater than 50 imply that for this dataset and for this processing methodology, approximately 50 independent samples are required to reliably determine the principal modes of shape variation relevant to cam morphology. Several factors are likely to influence this finding: 1) Procrustes scaling reduces size-driven variance and shifts the explained variance towards shape features such as cam size and location, 2) number of particles, 3) image modality and resolution, 4) segmentation protocol, and 5) cohort demographics and characteristics. Future work will assess how these modeling choices influence our findings. Importantly, our study used only pre-operative cam FAI subjects from a single surgeon. Inclusion of controls or a wider range of subjects may shift the required number of subjects and is planned. Future work will also analyze directly PC loadings for stability and not just explained variance, ensuring that the modes themselves are stable, and represent consistent shape variability. Our initial findings nevertheless suggest that a minimum of 50 independent subjects is needed to properly capture the shape variability within a cohort of cam FAI femurs when PCA derived shape mode stability is the goal. This reliability study provides a benchmark and workflow for future researchers utilizing SSM in FAI. Studies with small sample sizes risk reporting unstable and non-generalizable shape modes.

SIGNIFICANCE/CLINICAL RELEVANCE: Statistical shape modeling has been increasingly used for the assessment of joint shape in the context of musculoskeletal pathology, such as FAI. However, no assessment has been reported, in the context of FAI, to determine an appropriate sample size that is representative of the larger patient population.

REFERENCES: 1 Harris, M. *et al. J. Hip Preserv. Surg* 11, 132-139 (2024*)* **2** Bugeja, J. *et al. JOR* 42, 385-394 (2024*)* **3** Bugeja, J. *et al. Quant. Imaging. Med. Surgy* 12, 4924-4941 (2022*)* **4** Van Buuren, M. M. A. *et al. Osteoathr. Cartil. Open* 4, 100275 (2022) **5** Keating, T. *et al. ASMAR* 2, 91-95 (2020) **6** Atkins, P. *et al. CORR* 477, 242-253 (2019) **7** Sochacki, K. *et al. OJSM* 6, 2325967118791790 (2018) **8** Vellieux, N. *et al. JOR* 36, 3161-3168 (2018) **9** Atkins, P. *et al. CORR* 475, 1977-1986 (2017) **10** Atkins, P. *et al. J*OR 35, 1743-1753 (2017) **11** Khanduja, V. *et al Int J Med Robot* 12, 765-772 (2016) **12** Harris, M. *et al. JOR* 31, 1620-1626 (2013) **13** Gerhold, C. *et al.* in *Proceedings of the Orthopaedic Research Society 2025* (2025) **14** Cates, J. *et al.* in *Statistical Shape and Deformation Analysis* (eds. Zheng, G. et al.) 257–298 (Academic Press, 2017)

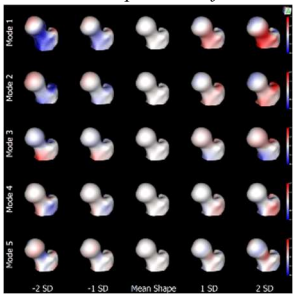

Figure 1: PCs 1-5 from an SSM (N=112 FAI femurs)

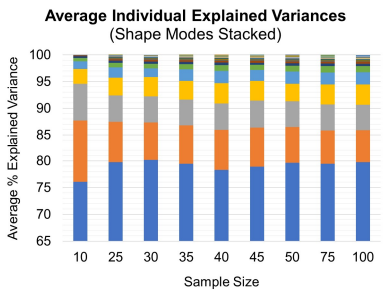

Figure 2: Stacked plot of PCs for each sample size (PC1 starts at bottom)

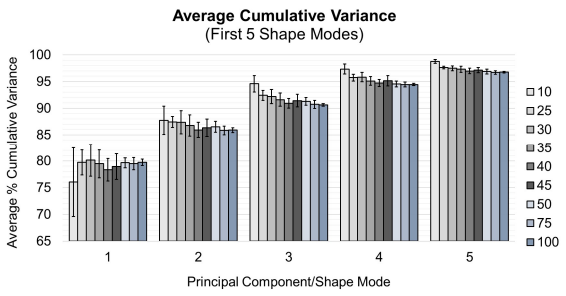

Figure 3: Cumulative explained variances for PCs 1-5 for each sample size tested. Error bars represent standard deviations.