# Automated Risk Stratification of Hip Osteoarthritis Development in Patients With Femoroacetabular Impingement Using an Unsupervised Clustering Algorithm

## A Study From the Rochester Epidemiology Project

Sunho Ko,* MD, Ayoosh Pareek,[†] MD, Changwung Jo,* MD, Hyuk-Soo Han,* MD, Myung Chul Lee,* MD, Aaron J. Krych,[†] MD, and Du Hyun Ro,*[‡] MD
*Investigation performed at the Department of Sports Medicine Orthopedic Surgery, Mayo Clinic, Rochester, Minnesota, USA*

**Background:** Studies evaluating the natural history of femoroacetabular impingement (FAI) are limited.

**Purpose:** To stratify the risk of progression to osteoarthritis (OA) in patients with FAI using an unsupervised machine-learning algorithm, compare the characteristics of each subgroup, and validate the reproducibility of staging.

**Study Design:** Cohort study (prognosis); Level of evidence, 2.

**Methods:** A geographic database from the Rochester Epidemiology Project was used to identify patients with hip pain between 2000 and 2016. Medical charts were reviewed to obtain characteristic information, physical examination findings, and imaging details. The patient data were randomly split into 2 mutually exclusive sets: train set (70%) for model development and test set (30%) for validation. The data were transformed via Uniform Manifold Approximation and Projection and were clustered using Hierarchical Density-based Spatial Clustering of Applications with Noise.

**Results:** The study included 1071 patients with a mean follow-up period of 24.7 ± 12.5 years. The patients were clustered into 5 subgroups based on train set results: patients in cluster 1 were in their early 20s (20.9 ± 9.6 years), female dominant (84%), with low body mass index (<19); patients in cluster 2 were in their early 20s (22.9 ± 6.7 years), female dominant (95%), and pincer-type FAI (100%) dominant; patients in cluster 3 were in their mid 20s (26.4 ± 9.7) and were mixed-type FAI dominant (92%); patients in cluster 4 were in their early 30s (32.7 ± 7.8), with high body mass index (≥29), and diabetes (17%); and patients in cluster 5 were in their early 30s (30.0 ± 9.1), with a higher percentage of males (43%) compared with the other clusters and with limited internal rotation (14%). Mean survival for clusters 1 to 5 was 17.9 ± 0.6, 18.7 ± 0.3, 17.1 ± 0.4, 15.0 ± 0.5, and 15.6 ± 0.5 years, respectively, in the train set. The survival difference was significant between clusters 1 and 4 ($P = .02$), 2 and 4 ($P < .005$), 2 and 5 ($P = .01$), and 3 and 4 ($P < .005$) in the train set and between clusters 2 and 5 ($P = .03$) and 3 and 4 ($P = .01$) in the test set. Cluster characteristics and prognosis was well reproduced in the test set.

**Conclusion:** Using the clustering algorithm, it was possible to determine the prognosis for OA progression in patients with FAI in the presence of conflicting risk factors acting in combination.

**Keywords:** femoroacetabular impingement; osteoarthritis; machine learning; clustering; UMAP; FAI

Femoroacetabular impingement (FAI) involves abnormal contact between femoral head and acetabulum leading to labral damage and cartilaginous injury.[4] It can occur in active young adults with hip pain and functional limitation.[17] The most common structural deformities include abnormal femoral head with increasing radius (cam-type lesion), or general or local anterior overcoverage of acetabulum (pincer-type lesion), or combined deformity.[6] The awareness of the symptomatic FAI has been increasing, demonstrated to be 54.4 per 100,000 person-years, which increased consistently between 2000 and 2016. The number of operations including hip arthroscopy, surgical hip dislocation, and anteverting periacetabular osteotomy has increased concordantly.[9]

Patients with hip FAI are known to be at an increased risk of hip osteoarthritis (OA),[2,3,6] yet little is known about the risk factors for hip OA. Recently, Melugin et al conducted a large cohort study of 1104 patients in the Rochester Epidemiology Project (REP) database analyzing risk factors for hip OA in patients with FAI without prior surgery.[16] REP is a unique population health resource and a collaborative initiative established in 1966 and originating in Olmsted County, Minnesota.[15] The candidate risk factors include characteristic information, radiologic measurements, and physical examination findings. The rate of OA was 13.5% and total hip arthroplasty (THA) was performed in 4% of patients. Male sex, body mass index (BMI) >29 kg/m$^2$, and increased age were risk factors for OA.

Independent risk factors for disease are analyzed using conventional statistics in many studies. However, patients have multiple risk factors and the relationship between risk factors is not independent. Therefore, prediction models based on machine learning have been developed to calculate the overall effect of these risk factors in orthopaedics, classifying patients into high- and low-risk groups.[11,13] Certain risk factors are not predominantly seen in the entire patient group, but in specific groups, they can have a great effect on outcomes. However, it is difficult to identify these risk factors with existing conventional models. If patients are clustered based on topological relationship between variables, the risk stratification and the dominant risk factors in each cluster can be studied.

The purpose of this study was to stratify the risk of progression to OA in patients with hip FAI using an unsupervised machine-learning algorithm, compare the characteristics of each subgroup, and validate the reproducibility of staging performed via clustering analysis using a geographic population with long-term follow-up. We hypothesized that (1) risk stratification for progression to OA of patients with hip FAI can be achieved with unsupervised machine learning. (2) There are several subgroups of patients with hip FAI for progression to OA, and there will be distinct risk factors for each subgroup. (3) The staging done by clustering algorithm is reproducible.

## METHODS

### Study Population

This study was approved by an institutional review board. Patients who presented to physician with an International Classification of Diseases, 9th Revision (ICD-9) or 10th Revision (ICD-10), diagnostic code of hip pain, hip impingement, or hip joint disorders between January 2000 and December 2016 from the REP were reviewed. The REP is a medical record linkage system representing a population-based data resource combining the medical records of our institution and other community providers in the city of Rochester and in Olmsted County, Minnesota. Patients between ages of 14 and 50 years were included. For all patients, time from initial hip pain to follow-up was calculated. The patient data were monitored until December 2019 (minimum follow-up of 3 years). Patients with a history of other hip disorders (avascular necrosis, neuromuscular disorder, trochanteric bursitis, hip fracture, pelvic fracture, and hip dislocation) or a history of hip arthroplasty or hip preservation procedures were excluded.

All patients provided antero-posterior views and at least 1 lateral view (cross-table, frog-leg, or 45° Dunn) hip radiograph during their initial assessment. All radiographs were reviewed by attending or senior resident-level orthopaedic surgeons. The classification of FAI (cam, pincer, mixed) was based on radiographic criteria proposed by Clohisy et al.[5] The radiographic reviews were evaluated by 2 orthopaedic surgeons to confirm the consistency of radiographic measurement. Clinical charts were reviewed to determine the physical examination findings. Patients with symptoms consistent with hip FAI defined by Warwick agreement,[7] clinical signs, and radiographic findings were included. Patient characteristic factors were identified during chart review.

Symptomatic hip OA was defined as symptoms requiring treatment and Tönnis grade 1 or higher on hip radiographs. Symptoms included pain, hip clicking, catching, locking, stiffness, restricted range of motion, or giving way.[7]

The data collection and preprocessing steps were performed as described by Melgun et al.[16]

### Statistical Analysis

The statistical analysis was performed using R 4.0 (R Core Team, http://www.R-project.org/), and Python 3.7 (Python Software Foundation, https://www.python.org).

A descriptive statistical analysis of characteristic, radiographic, and physical examination features was performed. Chi-square and *t* tests were used to compare OA and non-OA groups. One-way analysis of variance was used for multiple group comparison. Survival curves were plotted using

‡Address correspondence to Du Hyun Ro, MD, Department of Orthopedic Surgery, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul, 03080, Republic of Korea (email: duhyunro@gmail.com); or Aaron J. Krych, MD, Department of Orthopedic Surgery, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA (email: krych.aaron@mayo.edu).

*Department of Orthopedic Surgery, Seoul National University Hospital, Seoul, Republic of Korea.
†Department of Orthopedic Surgery, Mayo Clinic, Rochester, Minnesota, USA.
S.K. and A.P. contributed equally to this work.
A.J.K and D.H.R. contributed equally to this work.

Initial data
(1,893 patients,
40 features)

→ Exclusion (822 patients)

Feature selection
by univariate analysis

Processed data
(1,071 patients,
16 features)

Preprocessing pipeline

Train group
(70%, N=749)

Test group
(30%, N=322)

UMAP
↓
HDBSCAN

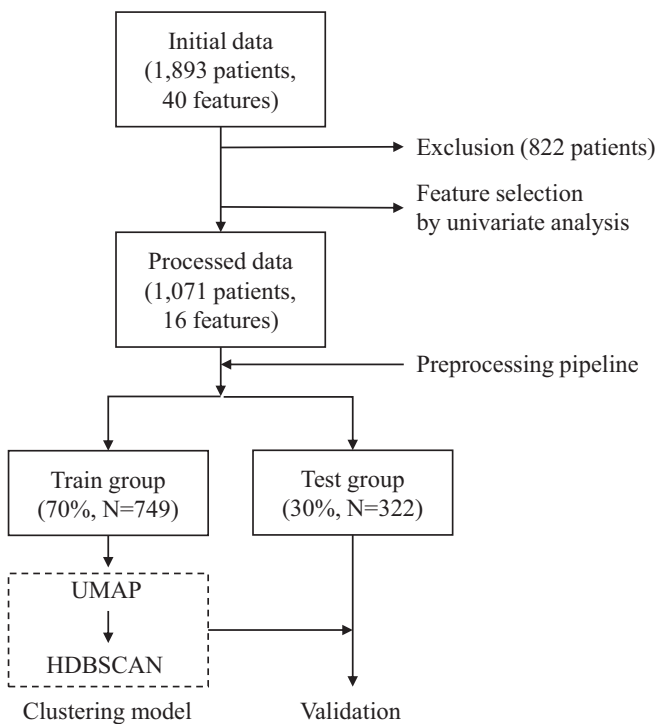Clustering model          Validation

**Figure 1.** The study population. HDBSCAN, hierarchical density-based spatial clustering of applications with noise; UMAP, uniform manifold approximation and projection for dimension reduction.

the Kaplan-Meier method, and log-rank test was used for comparison. Missing data were replaced by median value for numerical features and constant value for categorical features (0 for binary categorical variable, otherwise, closest category to the median value of the feature)

## Model Creation

Initial feature selection was performed to screen the OA and non-OA groups based only on significant differences in features to cluster the patients based on the risk factors for OA progression. The patient data were randomly split into 2 mutually exclusive sets: train set (70%) for model development and test set (30%) for validation. The features of train set and test set were compared to assess the adequacy of split.

## Clustering

In unsupervised machine learning, the training data are not labelled, and thus the system learns without any guidance. Clustering can be used to group the patients without any labels first (train group), then we could use clustering to label the test group for validation, or to find out which cluster a new patient belongs to. In this study, the unsupervised machine-learning algorithm, dimension reduction, and the clustering algorithm were used to stage the risk of progression to OA in patients with hip FAI.

Uniform manifold approximation and projection for dimension reduction (UMAP) is a powerful, nonlinear dimension reduction technique.[14] UMAP projects the data into low-dimensional space via conservation of its original topological structure. Hierarchical density-based spatial clustering of applications with noise (HDBSCAN) is an enhanced version of density-based spatial clustering of applications with noise (DBSCAN), which reflects the local density and the hierarchical structure of the data. The data were first projected into the 2-dimensional space via UMAP and then clustered by HDBSCAN.

## Comparison with Binary-Classification Machine-Learning Model

The clustering-based model was compared with the binary-classification machine-learning model. The binary-classification model was developed according to the gradient boosting machine algorithm. The model classifies patients into high-risk and low-risk groups. For comparison with the clustering technique, the test set of the previous study was clustered and visualized with the algorithm developed in this study.

## RESULTS

Overall, 1071 of 1893 patients met the inclusion criteria. Patients who had undergone hip arthroscopy or hip preservation procedures were excluded (n = 208). The flowchart of process is shown in Figure 1. Mean follow-up time was 24.7 ± 12.5 years. Mean age at the onset of pain was 28.5 ± 9.3 years, and males constituted 29.6% of the cohort. Progression to OA was detected in 149 (13.9%) patients over an average follow-up of 40.1 months. The average follow-up of non-OA patients was 95.4 months.

Of the 37 variables, 14 were initially selected via univariate analysis. Full list of 37 variables can be found in Appendix Table A1. The characteristics of OA and non-OA groups are listed in Table 1. Follow-up duration of OA (40.1 ± 52.9 months) and non-OA (95.4 ± 58.9 months) groups were significantly different. Labral tear on magnetic resonance imaging (MRI) of OA (72%) and non-OA (52%) groups were also significantly different. Since MRI was taken electively, 740 (69%) patients do not have MRI data. Physical examination data also have large portion of missing values ranging from 26% to 33%. There were no significant differences between train and test sets (Appendix Table A2). Additionally, no difference in survival outcome was detected between the 2 sets (Appendix Figure A1) indicating excellent split between train and test datasets.

### Clustering Outcome

Figure 2 shows the clusters of the train and test dataset via UMAP and DBSCAN. The train set was clustered into 5 subgroups (Figure 2A). Figure 2B annotates the patients who progressed to OA in the train set. Figure 2C shows the projection of test patients by the trained model. Figure 2D annotates the patients progressed to OA in the test set.

TABLE 1
Patient Characteristics and Physical Examination and Imaging Findings[a]

| Variable | All Patients (N = 1071) | OA Group (n = 149) | Non-OA Group (n = 922) | P | Missing |
|---|---|---|---|---|---|
| Follow-up, mo | 87.7 ± 61.2 | 40.1 ± 52.9 | 95.4 ± 58.9 | <.001 | 0 (0) |
| Age at onset of pain, y | 28.9 ± 9.3 | 35.7 ± 5.6 | 27.8 ± 9.3 | <.001 | 0 (0) |
| Male sex | 327 (31) | 66 (44) | 261 (28) | <.001 | 0 (0) |
| BMI | | | | <.001 | 19 (2) |
| <19 | 59 (6) | 2 (1) | 57 (6) | | |
| 19-23 | 354 (34) | 22 (15) | 332 (37) | | |
| 24-28 | 284 (27) | 42 (29) | 242 (27) | | |
| 29-33 | 202 (19) | 39 (27) | 163 (18) | | |
| ≥34 | 153 (15) | 42 (29) | 111 (12) | | |
| Diabetes | 84 (8) | 27 (18) | 57 (6) | <.001 | 1 (0) |
| Pain: anterior/groin | 605 (56) | 106 (71) | 499 (54) | <.001 | 1 (0) |
| Pain: back | 72 (7) | 21 (14) | 51 (6) | <.001 | 1 (0) |
| Flexion limited | 69 (10) | 25 (23) | 44 (7) | <.001 | 351 (33) |
| IR limited | 83 (11) | 31 (29) | 52 (8) | <.001 | 311 (29) |
| ER limited | 45 (6) | 21 (18) | 24 (4) | <.001 | 279 (26) |
| Labral tear on MRI scan | 186 (56) | 49 (72) | 137 (52) | .003 | 740 (69) |
| Tönnis grade | | | | <.001 | 2 (0) |
| 0 | 382 (36) | 23 (15) | 359 (39) | | |
| 1 | 622 (58) | 91 (61) | 531 (58) | | |
| 2 | 59 (6) | 31 (21) | 28 (3) | | |
| 3 | 6 (1) | 4 (3) | 2 (0) | | |
| Type of impingement | | | | | 0 (0) |
| Pincer | 209 (20) | 16 (11) | 193 (21) | .001 | |
| Cam | 106 (10) | 24 (16) | 82 (9) | | |
| Mixed | 755 (70) | 109 (73) | 646 (70) | | |

[a]Data are reported as mean ± SD or n (%). All variables were statistically significantly different between the study groups (P < .05). BMI, body mass index; ER, external rotation; IR, internal rotation; MRI, magnetic resonance imaging; OA, osteoarthritis.

Distribution of patients and progression to OA was well reproduced in the test set.

The characteristics and progression to OA of the clusters in the train set are listed in Table 2 and shown in Figure 3. Cluster 1 had a mean age of 20.9 years, was 16% male, all patients had a BMI of <19 kg/m², and the mean Tönnis grade was 0.63. Cluster 2 had a mean age of 22.9 years, was 5% male, all patients had a BMI of 19 to 24 kg/m², and the mean Tönnis grade was 0.74. Cluster 3 had a mean age of 26.4 years, was 29% male, all patients had a BMI of 19 to 24 kg/m², and the mean Tönnis grade was 0.75. Cluster 4 had a mean age of 32.7 years, was 24% male, all patients had a BMI of 29 to 34 kg/m² in 58% and ≥34 kg/m² in 42%, and mean Tönnis grade was 0.10. Cluster 5 had a mean age of 30.0 years, was 43% male, all patients had a BMI of 24 to 29 kg/m², and the mean Tönnis grade was 0.81. The corresponding characteristics and progression to OA of the test-set clusters are shown in Table 3 and Figure 4.

### Comparison with Binary Classification Model

Figure 5 shows the difference between a binary-classification model and a clustering-based model. The patient groups based on clustering are clearly distinguished from other groups in terms of the risk or rate of OA progression and represent the group of patients who did not appear in the existing binary-classification, machine-learning model.

### Long-term Prognosis of Clusters

The progression to OA in each group is presented in Figure 6. The percentage of progression to OA (survivals) of each cluster are listed in Table 4. In the train set, the mean survival for clusters 1 to 5 were 17.9 ± 0.6, 18.7 ± 0.3, 17.1 ± 0.4, 15.0 ± 0.5, 15.6 ± 0.5 years, respectively. According to the log-rank test, there were significant differences in survival between train-set clusters 1 and 4 (P = .02), 2 and 4 (P < .005), 2 and 5 (P = .01), and 3 and 4 (P < .005), likewise for the test-set clusters 2 and 5 (P = .03), and 3 and 4 (P = .01). The long-term prognosis of each group was clearly distinguished in both train and test groups in a similar fashion.

### DISCUSSION

Using the dimension reduction and clustering algorithms, patients with hip FAI are separated into 5 clusters. Characteristics and survival of each cluster were evaluated, and each cluster has a different risk for OA progression.

### Cluster Characteristics

Table 2 clearly shows the difference between BMI and type of impingement between clusters in the train set. The best prognosis among the 5 clusters was seen in the cluster 2 patients, characterized by age in the early 20s (22.9 ± 6.7 years),
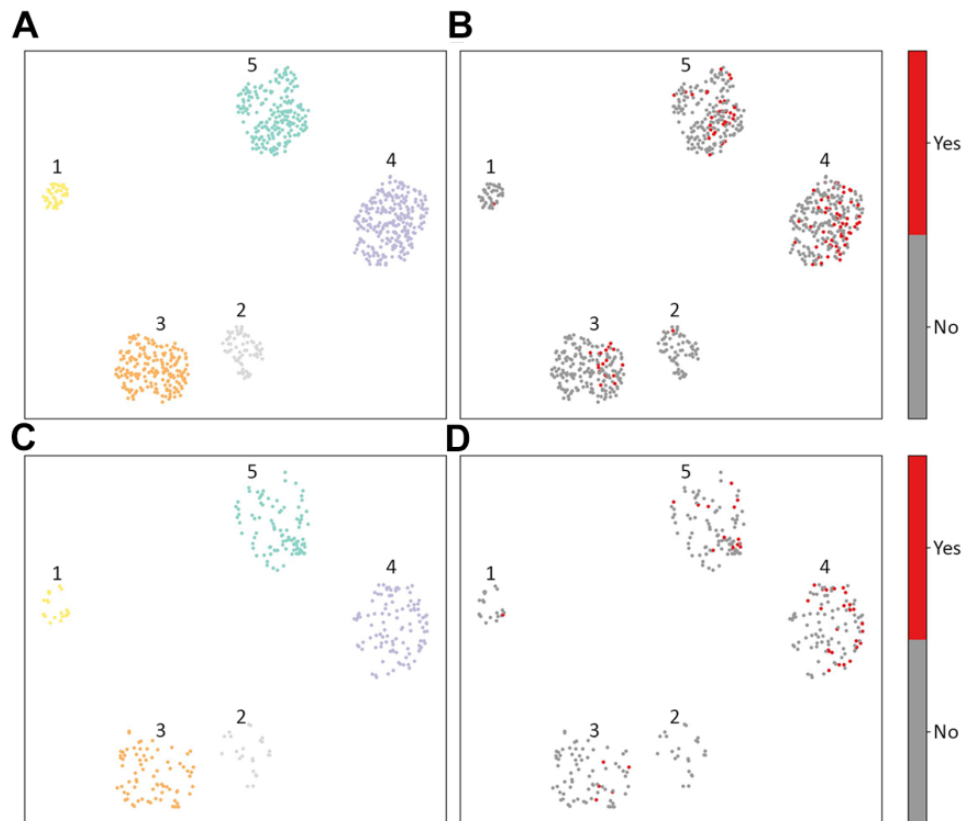
**Figure 2.** (A) Clustering of patient data for the train set via UMAP-enhanced DBSCAN. (B) Progression to osteoarthritis (OA) in each cluster (characterized by a different color). The rate of OA progression is clearly distinguished within each cluster. (C) Clustering of patient data for the test set via UMAP-enhanced DBSCAN. (D) Progression to OA in each cluster (characterized by a different color). The test set shows the same distribution as the train set, and the rate of OA progression is also similar. DBSCAN, density-based spatial clustering of applications with noise; UMAP, uniform manifold approximation and projection for dimension reduction.

female dominant (95%), and pincer-type FAI (100%) dominant, with a mean survival of $18.7 \pm 0.3$ years. On the other hand, the worst prognosis was in the cluster 4 patients, characterized by age in the early 30s ($32.7 \pm 7.8$ years), high BMI ($\geq 29$ kg/m$^2$), and diabetes (17%), with a mean survival of $15.0 \pm 0.5$ years. Patients in cluster 3 were in their mid-20s ($26.4 \pm 9.7$ years), and were mixed-type FAI (92%) dominant, with a mean survival of $17.1 \pm 0.4$ years. Patients in cluster 1 were characterized by age in the early 20s ($20.9 \pm 9.6$ years), female dominant (84%), and low BMI ($<19$ kg/m$^2$), with a mean survival of $17.9 \pm 0.6$ years. Cluster 5 patients were in their early 30s ($30.0 \pm 9.1$ years) and were male (43%) dominant than the other clusters, with limited internal rotation (14%) and a mean survival of $15.6 \pm 0.5$ years.

The relationship between the characteristics of each cluster and its prognosis can be explained by former study results. BMI $>29$ kg/m$^2$, increased age, and male sex have all been identified as risk factors for OA progression.[10,16,18]

The difference in the type of impingement between clusters may originate in the sex difference. Pincer type is predominant in females, while cam-type lesion is dominant in young male athletes.[1,4,8] This can be seen in cluster 2, which was female dominant (95%) and pincer-type

dominant (100%), while cluster 3 had a higher percentage of males (29%) than cluster 2 and cam (8%), and were mixed-type (92%) impingement.

There was no significant difference between clusters in MRI labral tear or physical examination data, despite their clinical importance. MRI labral tears or physical examination data are binary categorical features with large missing value. For binary categorical features, the imputation of missing values can be done only by constant value—usually and in this case zero, resulting in only small portion among the whole patients having positive values. Because the clustering algorithm divides subgroups based on features that 'group' the 'whole patients', influence of features with only small percentage of positive value inevitably decreases in clustering modeling.

## Modeling Strengths

Classification models predicting high-risk groups and survival regression models such as random survival forest was mainly used in prognostic models using supervised machine-learning algorithm.[11-13] The classification model facilitates the interpretation of the result clearly. Risk status, sensitivity, and specificity data provide clear insight

TABLE 2
Characteristics of the Clusters in the Train Set[a]

| | All Train Set (N = 749) | Cluster 1 (n = 38) | Cluster 2 (n = 65) | Cluster 3 (n = 192) | Cluster 4 (n = 250) | Cluster 5 (n = 204) | P | Missing |
|---|---|---|---|---|---|---|---|---|
| Progression to OA | 104 (14) | 1 (3) | 1 (2) | 16 (8) | **56 (22)** | 30 (15) | <.001 | 0 (0) |
| Follow-up, mo | 87.4 ± 61.0 | 82.6 ± 44.9 | 80.4 ± 54.5 | 86.5 ± 61.4 | 94.6 ± 65.3 | 83.4 ± 59.6 | .249 | 0 (0) |
| Age at onset of pain, y | 28.9 ± 9.4 | 20.9 ± 9.6 | 22.9 ± 6.7 | 26.4 ± 9.7 | **32.7 ± 7.8** | 30.0 ± 9.1 | <.001 | 0 (0) |
| Male sex | 214 (29) | 6 (16) | **3 (5)** | 56 (29) | 61 (24) | 88 (43) | <.001 | 0 (0) |
| BMI | | | | | | | <.001 | 14 (2) |
| <19 | 38 (5) | 38 (100) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | | |
| 19-23 | 257 (35) | 0 (0) | 65 (100) | 192 (100) | 0 (0) | 0 (0) | | |
| 24-28 | 190 (26) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 190 (100) | | |
| 29-33 | 145 (20) | 0 (0) | 0 (0) | 0 (0) | 145 (58) | 0 (0) | | |
| ≥34 | 105 (14) | 0 (0) | 0 (0) | 0 (0) | **105 (42)** | 0 (0) | | |
| Diabetes | 64 (9) | 0 (0) | 1 (2) | 7 (4) | **42 (17)** | 14 (7) | <.001 | 1 (0) |
| Pain: anterior/groin | 415 (55) | 19 (50) | 39 (60) | 101 (53) | 139 (56) | 117 (57) | .481 | 0 (0) |
| Pain: back | 53 (7) | 1 (3) | 2 (3) | 5 (3) | 32 (13) | 13 (6) | <.001 | 0 (0) |
| Flexion limited | 42 (8) | 1 (4) | 0 (0) | 6 (5) | 21 (13) | 14 (9) | .016 | 241 (32) |
| IR limited | 55 (10) | 2 (8) | 2 (4) | 8 (5) | 23 (13) | 20 (14) | .041 | 207 (27) |
| ER limited | 30 (5) | 2 (9) | 1 (2) | 5 (3) | 15 (9) | 7 (4) | .124 | 188 (25) |
| Labral tear on MRI scan | 125 (53) | 3 (38) | 9 (45) | 36 (55) | 35 (51) | 42 (58) | .538 | 514 (68) |
| Tönnis grade | | | | | | | <.001 | 0 (0) |
| 0 | 264 (35) | **14 (37)** | **58 (89)** | 58 (30) | 69 (28) | 65 (32) | | |
| 1 | 443 (59) | 23 (61) | 7 (11) | **123 (64)** | **164 (66)** | **126 (62)** | | |
| 2 | 39 (5) | 1 (3) | 0 (0) | 10 (5) | 17 (7) | 11 (5) | | |
| 3 | 3 (0) | 0 (0) | 0 (0) | 1 (1) | 0 (0) | 2 (1) | | |
| Type of impingement | | | | | | | <.001 | 0 (0) |
| Pincer | 147 (20) | 9 (24) | **65 (100)** | 0 (0) | 34 (14) | 39 (19) | | |
| Cam | 71 (10) | 0 (0) | 0 (0) | 16 (8) | 24 (10) | 31 (15) | | |
| Mixed | 530 (71) | **29 (76)** | 0 (0) | **176 (92)** | **192 (77)** | 133 (65) | | |

[a]Data are reported as n (%) or mean ± SD. Bolded values indicate significant differentiating factors among clusters. BMI, body mass index; ER, external rotation; IR, internal rotation; MRI, magnetic resonance imaging; OA, osteoarthritis.
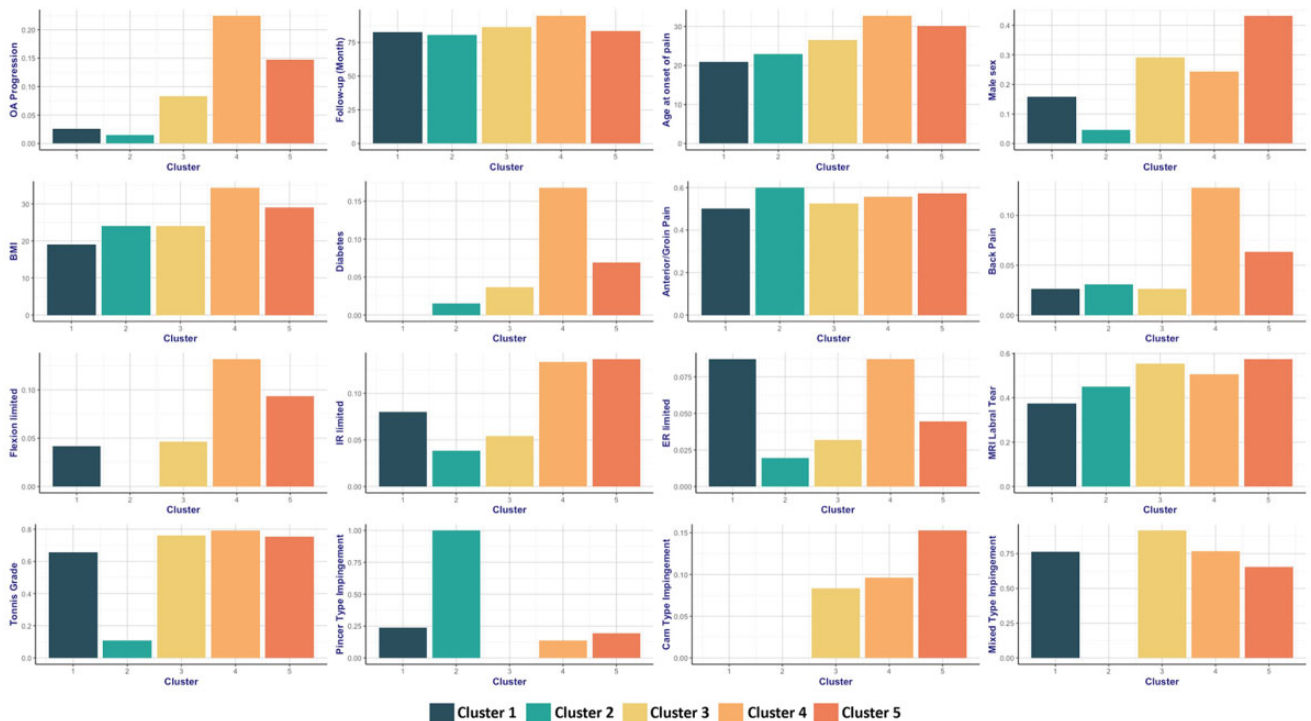


**Figure 3.** Characteristics of clusters in the train set as visualized using bar graphs. BMI, body mass index; OA, osteoarthritis.

TABLE 3
Characteristics of Clusters in the Test Set[a]

| | All Test Set (N = 322) | Cluster 1 (n = 21) | Cluster 2 (n = 23) | Cluster 3 (n = 74) | Cluster 4 (n = 105) | Cluster 5 (n = 99) | P | Missing |
|---|---|---|---|---|---|---|---|---|
| Progression to OA | 45 (14) | 1 (5) | 0 (0) | 5 (7) | **25 (24)** | 14 (14) | .002 | 0 (0) |
| Follow-up, mo | 87.8 ± 61.5 | 108.1 ± 65.2 | 76.3 ± 54.2 | 88.5 ± 55.8 | 89.1 ± 64.6 | 84 ± 63 | .487 | 0 (0) |
| Age at onset of pain, y | 28.9 ± 9.1 | 21.1 ± 9.9 | 23.1 ± 8.4 | 25.0 ± 9.4 | **32.4 ± 7.5** | 31 ± 8 | <.001 | 0 (0) |
| Male sex | 113 (35) | 6 (29) | **1 (4)** | 29 (39) | 38 (36) | 39 (39) | .023 | 0 (0) |
| BMI | | | | | | | <.001 | 5 (2) |
| <19 | 21 (7) | 21 (100) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | | |
| 19-23 | 97 (31) | 0 (0) | 23 (100) | 74 (100) | 0 (0) | 0 (0) | | |
| 24-28 | 94 (30) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 94 (100) | | |
| 29-33 | 57 (18) | 0 (0) | 0 (0) | 0 (0) | 57 (54) | 0 (0) | | |
| ≥34 | 48 (15) | 0 (0) | 0 (0) | 0 (0) | 48 (46) | 0 (0) | | |
| Diabetes | 20 (6) | 0 (0) | 0 (0) | 1 (1) | **16 (15)** | 3 (3) | <.001 | 0 (0) |
| Pain: anterior/groin | 190 (59) | 11 (55) | 13 (57) | 40 (54) | 63 (60) | 63 (64) | .765 | 1 (0) |
| Pain: back | 19 (6) | 0 (0) | 0 (0) | 2 (3) | 11 (11) | 6 (6) | .092 | 1 (0) |
| Flexion limited | 27 (13) | 2 (17) | 0 (0) | 6 (12) | 11 (16) | 8 (11) | .615 | 110 (34) |
| IR limited | 28 (13) | 0 (0) | 0 (0) | 6 (11) | 11 (17) | 11 (16) | .202 | 104 (32) |
| ER limited | 15 (7) | 0 (0) | 0 (0) | 1 (2) | 7 (10) | 7 (9) | .205 | 91 (28) |
| Labral tear on MRI scan | 61 (64) | 2 (50) | 1 (50) | 19 (56) | 22 (79) | 17 (61) | .396 | 226 (70) |
| Tönnis grade | | | | | | | <.001 | 2 (1) |
| 0 | 118 (37) | **9 (45)** | **21 (91)** | 24 (32) | 29 (28) | 35 (36) | | |
| 1 | 179 (56) | 10 (50) | 2 (9) | **47 (64)** | **61 (58)** | **59 (60)** | | |
| 2 | 20 (6) | 1 (5) | 0 (0) | 3 (4) | 14 (13) | 2 (2) | | |
| 3 | 3 (1) | 0 (0) | 0 (0) | 0 (0) | 1 (1) | 2 (2) | | |
| Type of impingement | | | | | | | <.001 | 0 (0) |
| Pincer | 62 (19) | 5 (24) | **23 (100)** | 0 (0) | 12 (11) | 22 (22) | | |
| Cam | 35 (11) | 2 (10) | 0 (0) | 7 (10) | 19 (18) | 7 (7) | | |
| Mixed | 225 (70) | **14 (67)** | 0 (0) | **67 (91)** | **74 (71)** | **70 (71)** | | |

Data are reported as n (%) or mean ± SD. Bolded values indicate differentiating factors among clusters. BMI, body mass index; ER, external rotation; IR, internal rotation; MRI, magnetic resonance imaging; OA, osteoarthritis.
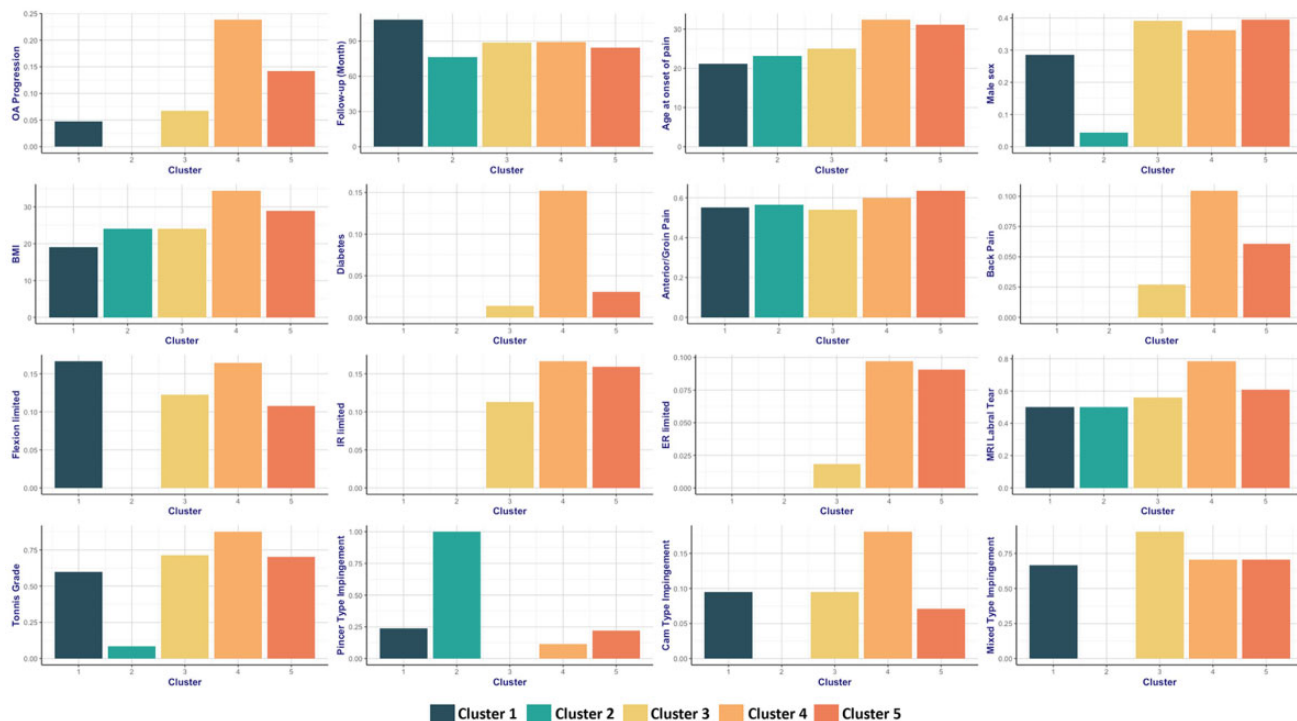


**Figure 4.** Characteristics of clusters in the test set as visualized using bar graphs. BMI, body mass index; OA, osteoarthritis.
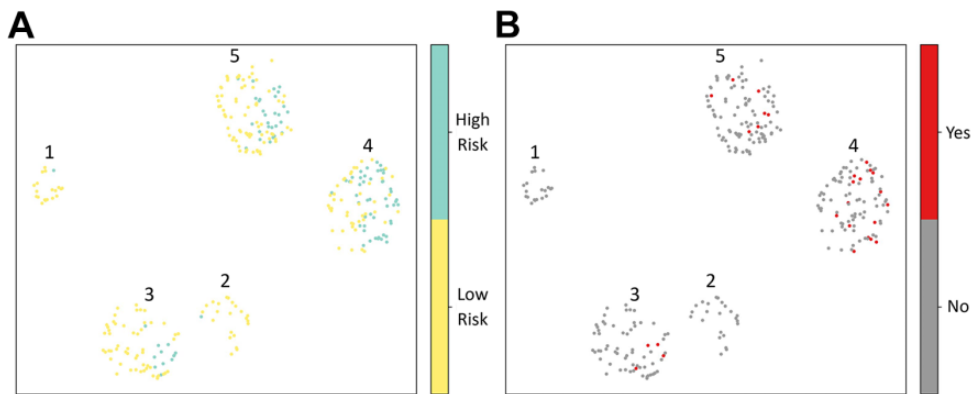
**Figure 5.** Progression to osteoarthritis (OA) as shown using the binary-classification machine-learning model. (A) Green and yellow represent patients determined by machine-learning models as being at high risk and low risk, respectively. High-risk patients determined by the machine-learning model are distributed across several clusters. (B) OA progression in each cluster. The unique subgroups of patients identified using the clustering algorithm are undetected in the existing binary-classification machine-learning model.
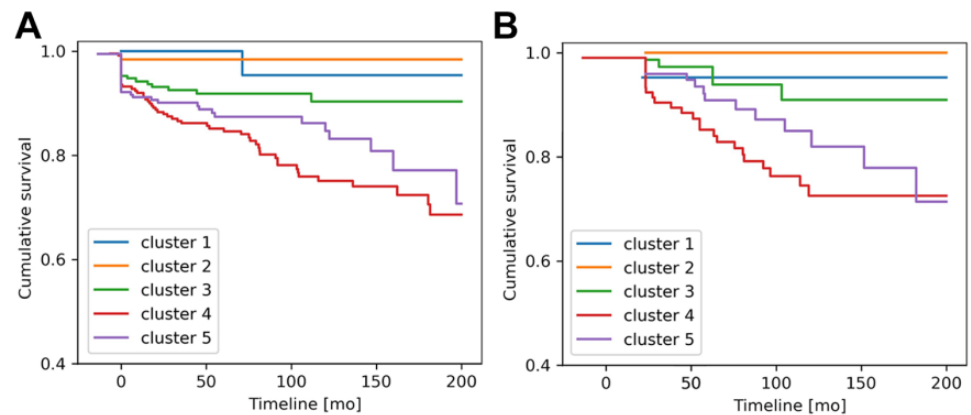


**Figure 6.** Survival functions for the (A) train set and (b) test set. The long-term prognosis of each group was clearly distinguished in the train and test groups similarly.

TABLE 4
Percentage of Progression to Osteoarthritis (Survivals) of Clusters for the Train Set and Test Set

| Train Set | Cluster 1 (n = 38) | Cluster 2 (n = 65) | Cluster 3 (n = 192) | Cluster 4 (n = 250) | Cluster 5 (n = 204) |
|---|---|---|---|---|---|
| 2 y | 1.000 | 0.985 | 0.932 | 0.883 | 0.901 |
| 5 y | 1.000 | 0.985 | 0.919 | 0.852 | 0.874 |
| 10 y | 0.954 | 0.985 | 0.904 | 0.751 | 0.862 |
| 15 y | 0.954 | 0.985 | 0.904 | 0.723 | 0.771 |
| Test Set | Cluster 1 (n = 21) | Cluster 2 (n = 23) | Cluster 3 (n = 74) | Cluster 4 (n = 105) | Cluster 5 (n = 99) |
| 2 y | 0.952 | 1.000 | 0.972 | 0.884 | 0.959 |
| 5 y | 0.952 | 1.000 | 0.938 | 0.817 | 0.891 |
| 10 y | 0.952 | 1.000 | 0.909 | 0.725 | 0.820 |
| 15 y | 0.952 | 1.000 | 0.909 | 0.725 | 0.713 |

into patient status. However, the follow-up time is not considered in modeling of long-term follow-up data. In the case of survival regression model, a quantified survival curve

can be obtained. However, the c-index, a widely used evaluation metric, determines the order of events (progression to OA), but not the exact survival time. However, the

advantage of the clustering technique is that there is no need to consider follow-up time as it does not require outcome variable in the modeling process, and the result interpretation is also intuitive and clear.

## Limitations

There are limitations in the patient population. Although the mean age of the group was young, patients up to 55 years old were included in the study. The risk of OA is expected to increase with age. The patients were from single location, and the dominant race was White. Additional multiregional studies are needed to determine whether this model can be generalized to patients in other regions.

There may be disagreement on the definition of symptomatic hip OA. In this study, symptomatic hip OA was defined as symptoms requiring treatment and Tönnis grade $\geq 1$ on hip radiographs. Some may argue that including Tönnis grade 1 is too stringent in determining significant OA. The indication for THA was not standardized in all patients, so this group was excluded and not analyzed separately.

There are limitations in the modeling algorithm. Because the clustering method does not provide model interpretability (black box), the modeling process must be presumed based on the characteristics of each cluster. Therefore, another statistical analysis is required to determine the effect of specific factors.

The model may not provide cluster information for some patients. Although all patients in the test set were successfully classified in to 5 clusters, a patient from different cohort can have different characteristic, and may not belong to any clusters.

UMAP clusters the data based on the topological relationship between input data—which means that causal relationship between the features is not considered. Therefore, domain knowledge is required to interpret the causal relationship between the features.

Application of clustering algorithm to other patient groups cannot stratify prognosis clearly if risk factors in each cluster shift favorably or unfavorably. However, this also has clinical implications in that it is possible to determine the prognosis in the presence of conflicting risk factors acting in combination in actual patients. Therefore, the clustering algorithm is of high clinical value.

## CONCLUSION

The candidate risk factors for OA progression in patients with FAI were selected; then, unsupervised machine-learning algorithm was applied for stratifying the risk of OA progression. The clusters were characterized by BMI, type of impingement, and sex, which were identified as independent risk factors for OA progression by conventional statistics.

## REFERENCES

1. Allen D, Beaule PE, Ramadan O, Doucette S. Prevalence of associated deformities and hip pain in patients with cam-type femoroacetabular impingement. *J Bone Joint Surg Br*. 2009;91(5):589-594.
2. Bardakos NV, Villar RN. Predictors of progression of osteoarthritis in femoroacetabular impingement: a radiological study with a minimum of ten years follow-up. *J Bone Joint Surg Br*. 2009;91(2):162-169.
3. Beck M, Kalhor M, Leunig M, Ganz R. Hip morphology influences the pattern of damage to the acetabular cartilage: femoroacetabular impingement as a cause of early osteoarthritis of the hip. *J Bone Joint Surg Br*. 2005;87(7):1012-1018.
4. Bedi A, Kelly BT. Femoroacetabular impingement. *J Bone Joint Surg Am*. 2013;95(1):82-92.
5. Clohisy JC, Carlisle JC, Beaule PE, et al. A systematic approach to the plain radiographic evaluation of the young adult hip. *J Bone Joint Surg Am*. 2008;90(4):47-66.
6. Ganz R, Parvizi J, Beck M, et al. Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res*. 2003;417:112-120.
7. Griffin DR, Dickenson EJ, O'Donnell J, et al. The Warwick Agreement on femoroacetabular impingement syndrome (FAI syndrome): an international consensus statement. *Br J Sports Med*. 2016;50(19):1169-1176.
8. Hack K, Di Primio G, Rakhra K, Beaule PE. Prevalence of cam-type femoroacetabular impingement morphology in asymptomatic volunteers. *J Bone Joint Surg Am*. 2010;92(14):2436-2444.
9. Hale RF, Melugin HP, Zhou J, et al. Incidence of femoroacetabular impingement and surgical management trends over time. *Am J Sports Med*. 2021;49(1):35-41.
10. Jiang L, Rong J, Wang Y, et al. The relationship between body mass index and hip osteoarthritis: a systematic review and meta-analysis. *Joint Bone Spine*. 2011;78(2):150-155.
11. Jo C, Ko S, Shin WC, et al. Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. *Knee Surg Sports Traumatol Arthrosc*. 2020;28(6):1757-1764.
12. Kim DW, Lee S, Kwon S, et al. Deep learning-based survival prediction of oral cancer patients. *Sci Rep*. 2019;9(1):6994.
13. Ko S, Jo C, Chang CB, et al. A web-based machine-learning algorithm predicting postoperative acute kidney injury after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. Published online September 3, 2020. doi:10.1007/s00167-020-06258-0
14. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018: arXiv:1802.03426. https://ui.adsabs.harvard.edu/abs/2018arXiv180203426M
15. Melton LJ III. History of the Rochester Epidemiology Project. *Mayo Clin Proc*. 1996;71(3):266-274.
16. Melugin HP, Hale RF, Zhou J, et al. Risk factors for long-term hip osteoarthritis in patients with femoroacetabular impingement without surgical intervention. *Am J Sports Med*. 2020;48(12):2881-2886.
17. Sankar WN, Nevitt M, Parvizi J, et al. Femoroacetabular impingement: defining the condition and its role in the pathophysiology of osteoarthritis. *J Am Acad Orthop Surg*. 2013;21(1):S7-S15.
18. Singer SP, Dammerer D, Krismer M, Liebensteiner MC. Maximum lifetime body mass index is the appropriate predictor of knee and hip osteoarthritis. *Arch Orthop Trauma Surg*. 2018;138(1):99-103.

APPENDIX

TABLE A1
Full List of Variables$^a$

| Characteristic Variables | |
| --- | --- |
| | Age at onset of pain |
| | Sex |
| | Race |
| | Education |
| | BMI |
| | Smoker |
| | Diabetes |
| | Ehlers-Danlos (preoperative) |
| Radiographic variables | |
| | Cam deformity |
| | Joint space (mm) |
| | LCEA |
| | Tönnis angle |
| | Tönnis grade |
| | Crossover sign |
| | Posterior wall sign |
| | Ischial spine sign |
| | Acetabular protrusio |
| | Os acetabuli |
| | Type of impingement (choice = pincer) |
| | Type of impingement (choice = cam) |
| | Type of impingement (choice = mixed) |
| | MRI labral tear |
| Physical examination variables | |
| | Anterior/groin pain |
| | Posterior/buttocks pain |
| | Lateral pain |
| | Back pain |
| | Clicking/locking/instability |
| | Stinchfield test |
| | Sitting pain |
| | FADIR test |
| | FABER test |
| | Iliopsoas snapping test |
| | TFL snapping test |
| | Athletic pubalgia test |
| | Flexion limited |
| | Flexion pain |
| | Internal rotation limited |
| | Internal rotation pain |
| | External rotation limited |
| | External rotation pain |

$^a$BMI, body mass index; FABER, flexion, abduction, external rotation; FADIR, flexion, adduction and internal rotation; LCEA, lateral center edge angle; MRI, magnetic resonance imaging; TFL, tensor fasciae latae.

TABLE A2
Comparison of Train and Test Set Groups[a]

| | Total (N = 1071) | Train (n = 749) | Test (n = 322) | P (Train vs Test) | Missing |
|---|---|---|---|---|---|
| Progression to OA | 149 (13.9) | 104 (13.9) | 45 (14.0) | .969 | 0 (0) |
| Follow-up, mo | 87.7 ± 61.2 | 87.6 ± 61.1 | 87.8 ± 61.5 | | 0 (0) |
| Age at onset of pain, y | 28.9 ± 9.3 | 28.9 ± 9.4 | 28.8 ± 9.1 | .947 | 0 (0) |
| Male sex | 327 (31) | 214 (29) | 113 (35) | .034 | 0 (0) |
| BMI | | | | .442 | 19 (2) |
|   <19 | 59 (6) | 38 (5) | 21 (7) | | |
|   19-23 | 354 (34) | 257 (35) | 97 (30) | | |
|   24-28 | 284 (27) | 190 (26) | 94 (29) | | |
|   29-33 | 202 (19) | 145 (20) | 57 (18) | | |
|   ≥34 | 153 (15) | 105 (14) | 48 (15) | | |
| Diabetes | 84 (8) | 64 (9) | 20 (6) | .191 | 1 (0) |
| Pain: anterior/groin | 605 (56) | 415 (55) | 190 (59) | .253 | 1 (0) |
| Pain: back | 72 (7) | 53 (7) | 19 (6) | .489 | 1 (0) |
| Flexion limited | 69 (10) | 42 (8) | 27 (13) | .063 | 351 (33) |
| IR limited | 83 (11) | 55 (10) | 28 (13) | .281 | 311 (29) |
| ER limited | 45 (6) | 30 (5) | 15 (6) | .527 | 279 (26) |
| Labral tear on MRI scan | 186 (56) | 125 (53) | 61 (64) | .085 | 740 (69) |
| Tönnis grade | | | | .902 | 2 (0) |
|   0 | 382 (36) | 264 (35) | 118 (37) | | |
|   1 | 622 (58) | 443 (59) | 179 (56) | | |
|   2 | 59 (6) | 39 (5) | 20 (6) | | |
|   3 | 6 (1) | 3 (0) | 3 (1) | | |
| Type of impingement | | | | .787 | 0 (0) |
|   Pincer | 209 (20) | 147 (20) | 65 (20) | | |
|   Cam | 106 (10) | 71 (9) | 35 (11) | | |
|   Mixed | 755 (70) | 530 (71) | 225 (70) | | |

[a]Data are reported as n (%) or mean ± SD. BMI, body mass index; ER, external rotation; IR, internal rotation; MRI, magnetic resonance imaging; OA, osteoarthritis.
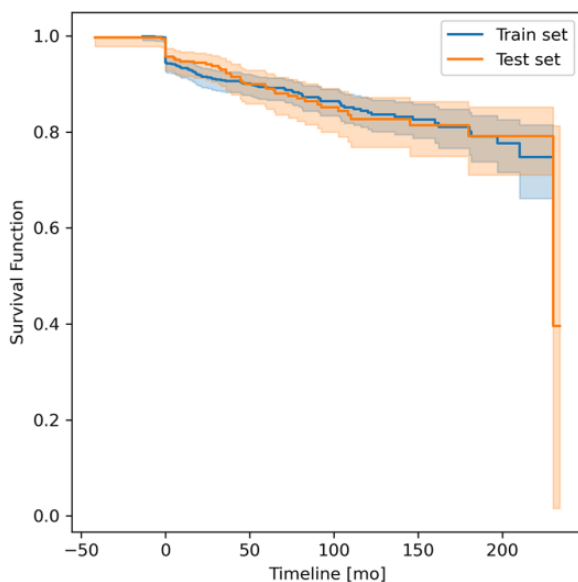


**Figure A1.** Kaplan-Meier curve of train and test sets. Shading represents 95% CI.