Seminars in Arthritis and Rheumatism 58 (2023) 152140

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Unsupervised machine-learning algorithms for the identification of clinical phenotypes in the osteoarthritis initiative database

David Demanse [a,#], Franziska Saxer [b,f,#], Patrick Lustenberger [c,#], László B. Tankó [d,#], Philipp Nikolaus [c], Ilja Rasin [c], Damian F. Brennan [c], Ronenn Roubenoff [b], Sumehra Premji [a,c], Philip G Conaghan [e], Matthias Schieker [b,*]

[a] Novartis Pharma AG, 4002, Basel, Switzerland
[b] Novartis Institutes for Biomedical Research, Novartis Campus, 4002, Basel, Switzerland
[c] IBM Switzerland AG, Vulkanstrasse 106, 8048, Zürich, Switzerland
[d] Bayer Pharmaceuticals, 4002, Basel Basel, Schweiz
[e] Leeds Institute of Rheumatic & Musculoskeletal Medicine, University of Leeds and NIHR Leeds Biomedical Research Centre, UK
[f] Medical Faculty, University of Basel, 4002, Basel, Switzerland

ARTICLE INFO

Keywords:
Knee osteoarthritis
Machine learning
Cluster analysis
Clinical phenotypes
Patient segments
Precision medicine

ABSTRACT

Objectives: Osteoarthritis (OA) is a complex disease comprising diverse underlying patho-mechanisms. To enable the development of effective therapies, segmentation of the heterogenous patient population is critical. This study aimed at identifying such patient clusters using two different machine learning algorithms.
Methods: Using the progression and incident cohorts of the Osteoarthritis Initiative (OAI) dataset, deep embedded clustering (DEC) and multiple factor analysis with clustering (MFAC) approaches, including 157 input-variables at baseline, were employed to differentiate specific patient profiles.
Results: DEC resulted in 5 and MFAC in 3 distinct patient phenotypes. Both identified a "comorbid" cluster with higher body mass index (BMI), relevant burden of comorbidity and low levels of physical activity. Both methods also identified a younger and physically more active cluster and an elderly cluster with functional limitations, but low disease impact. The additional two clusters identified with DEC were subgroups of the young/physically active and the elderly/physically inactive clusters. Overall pain trajectories over 9 years were stable, only the numeric rating scale (NRS) for pain showed distinct increase, while physical activity decreased in all clusters. Clusters showed different (though non-significant) trajectories of joint space changes over the follow-up period of 8 years.
Conclusion: Two different clustering approaches yielded similar patient allocations primarily separating complex "comorbid" patients from healthier subjects, the latter divided in young/physically active vs elderly/physically inactive subjects. The observed association to clinical (pain/physical activity) and structural progression could be helpful for early trial design as strategy to enrich for patients who may specifically benefit from disease-modifying treatments.

Introduction

Knee osteoarthritis (KOA) is the most common arthritis of weight-bearing joints, and is a leading cause of disability worldwide with high unmet medical need [1]. Its increasing prevalence with ageing of societies and the epidemic of obesity poses a growing economic burden on health care systems [2]. One major challenge is that various underlying pathologies may contribute to the clinical and radiologic picture of OA. Segmentation of this heterogenous patient population may increase the probability of meeting both patient-centered (pain/function) and joint-structural endpoints in clinical trials. This concept has sparked interest in phenotype research to allow enrichment in clinical trials

* Corresponding author.
E-mail addresses: david.demanse@novartis.com (D. Demanse), franziska.saxer@novartis.com (F. Saxer), patrick.lustenberger@ibm.com (P. Lustenberger), ilja.rasin@ch.ibm.com (I. Rasin), ronenn.roubenoff@novartis.com (R. Roubenoff), sumehra.premji@novartis.com (S. Premji), p.conaghan@leeds.ac.uk (P.G. Conaghan), matthias.schieker@novartis.com (M. Schieker).
[#] These authors have contributed equally to the analytic work and drafting of the manuscript.

https://doi.org/10.1016/j.semarthrit.2022.152140

Available online 19 November 2022

based on observable traits (phenotypes). Studies have evaluated various traits such as radiographic features, laboratory biomarkers and clinical characteristics [3,4]. The assumption is that phenotypes are associated with specific endotypes, i.e., molecular pathways of the disease, which would allow the development of more specific treatments. Ultimately endotype identification could support the recognition of genotypes opening vast opportunities for screening, diagnostics and targeted treatment [5,6].

Several groups have employed phenotyping as segmentation strategy. With various approaches diverging results have been observed. Broadly approaches can be separated according to methodology and the variables employed to derive phenotypes. On one hand data have been analyzed based on pre-specified often mechanistic disease concepts, e.g., stratification of atrophic vs hypertrophic OA [7], the evaluation of joint space narrowing over time as phenotypic trait [8] or the assessment of serum markers expected to be present in potential inflammatory, bone or cartilage driven phenotypes [9,10].

Some groups have focused on outcomes like changes in cartilage volume [11], OA related changes in conventional radiography [12] or clinical outcomes [13–15] and correlated these to clinical or laboratory patient characteristics.

Others have evaluated phenotyping based on specific factors assumed to be associated with different presentations of OA like pain susceptibility [16] psychological profile [17] or clinical risk factors [18] and analysed outcome trajectories for the resulting clusters. These approaches have increased our understanding of OA phenotypes, but are often limited by the number of patients, a certain selection bias and the limited number of input variables. In addition, the direction of analysis and underlying assumptions may influence the results.

Advances in machine learning (ML) have opened new possibilities [19–21] including the analysis of big data and unbiased analyses using ML-based clustering methods. These have been implemented in genome wide analyses [22,23] but increasingly also for imaging [24] and clinical data analyses [25–28]. We aimed to apply DEC (deep embedding clustering) [29] and multiple factor analysis with clustering (MFAC) [30–32] to the Osteoarthritis Initiative (OAI) dataset [33,34], using multiple clinical input variables to determine if distinct patient groups could be discerned, providing a basis for above-described enrichment strategies in future OA trials.

Subjects and methods

Study cohort

The OAI is a multi-center, longitudinal (2004-2016), prospective observational cohort-study of KOA including 4796 subjects. The database includes an ethnically diverse sample of both sexes aged 45–79 years, categorized according to their baseline characteristics:

- Progression-cohort (n=1390, 29%): frequent knee symptoms and radiographic signs of tibio-femoral KOA.
- Incidence-cohort (n=3284, 68%): frequent knee symptoms without radiographic evidence of OA, plus $\geq$ 2 risk factors i.e., Heberden's nodes (both hands), increased body weight, previous knee operation or injury, family history of OA.
- Control-cohort (n=122, 3%): no symptoms or risk factors for KOA. These were excluded in the present exercise.

The OAI database contains numerous variables such as patient characteristics, lifestyle preferences (e.g., smoking habits, nutrition), generic and disease-specific patient reported outcomes (PROs), physical activity (patient reported) and comorbidity scores supplemented by physical examinations, functional tests, imaging and serum/urine samples. The protocol and data definitions are publicly available [35].

For this analysis, we focused on 157 features (from the 1698 available) based on their clinical relevance, potential ease of collection and availability of the information for at least 90% of subjects, i.e., a max. of 10% of missing values for the overall population (see Supplemental Table 1 for all input-variables and proportion of missing values). We purposefully excluded radiographic input variables, since imaging is not primarily recommended for the diagnosis of OA in patients with typical presentation and therefore not necessarily a routine variable [36]. In addition, the avoidance of imaging related variables as input avoids the introduction of bias for imaging results as outcome. For the longitudinal analysis the more severely affected knee was included. For the cluster analysis both knees were evaluated, thereby mirroring the actual patient profile.

Description of the analytic workflow

a) Deep embedded clustering (DEC)

Fig. 1A visualizes the process of data processing for DEC, see supplement for more information. To identify the number of clusters, the k-means algorithm [37] was used, minimizing the sum of the squared variance per cluster from the cluster-center (Supplemental Fig. 3A). A DEC model using an auto-encoder for dimensionality reduction and a clustering layer for cluster identification was implemented [29]. This basically implies the stepwise unsupervised reduction of input variables to a set of representative essential features. The model training was conducted in three steps (allowing also to overcome limitations due to asymmetric/unbalanced representation of variables in the input space). In a layer-wise pre-training and across-layer fine-tuning of the auto-encoder, the model weights were initialized. To initialize the cluster centres (centroids) for the final clustering training, the data was passed through the initialized DNN (deconvolutional neural network) and standard k-means clustering (Student's t-distribution was used to measure the similarity between embedded point and the cluster centroids), was performed on the embedded data points. The clustering was then refined by training the DEC model optimizing the Kullback-Leibler divergence. All 3 training steps were performed using the same independent train partition of the data. Validation of the clustering algorithm was performed using 10 random initializations and comparing the yielded clusters by assignment-overlap and medical meaningfulness. To assess the clustering performance, a t-SNE plot [38] was created (Supplemental Fig. 1).

b) Multiple factor analysis with clustering (MFAC)

Fig. 1B visualizes the process of data processing for MFAC. Missing data for continuous and categorical variables was imputed via a regularized iterative principal components algorithm [39]. The number of components leading to the smallest mean squared error of prediction was retained using k-fold cross-validation. Multiple factor analysis (MFA) [32] as an extension of principal component analysis (PCA) assigns weights to variables balancing the influence of groups of similar variables on the global analysis. The 157 features were integrated in 12 groups corresponding to different clinical domains (Supplemental Table 1) for MFA and one group as illustrative (total numerical scores). One-hundred repeats and the 95%CI of the bootstrapped eigenvalues distribution identified the number of MFA components to retain for clustering. To define the number of clusters, Hierarchical Clustering on Principal Components (HCPC) [40] was performed using the Ward's criterion and an Euclidian distance metric on the selected MFA principal components. The inter-cluster inertia gain was used to select the optimal level of division (Supplemental Fig. 3B). Graphical MFAC outputs for the first two principal components (PC1 and PC2) are provided (Figs. 2 and 3).

Descriptive summary of the clusters

Clustering was performed in the incidence and progression cohort.

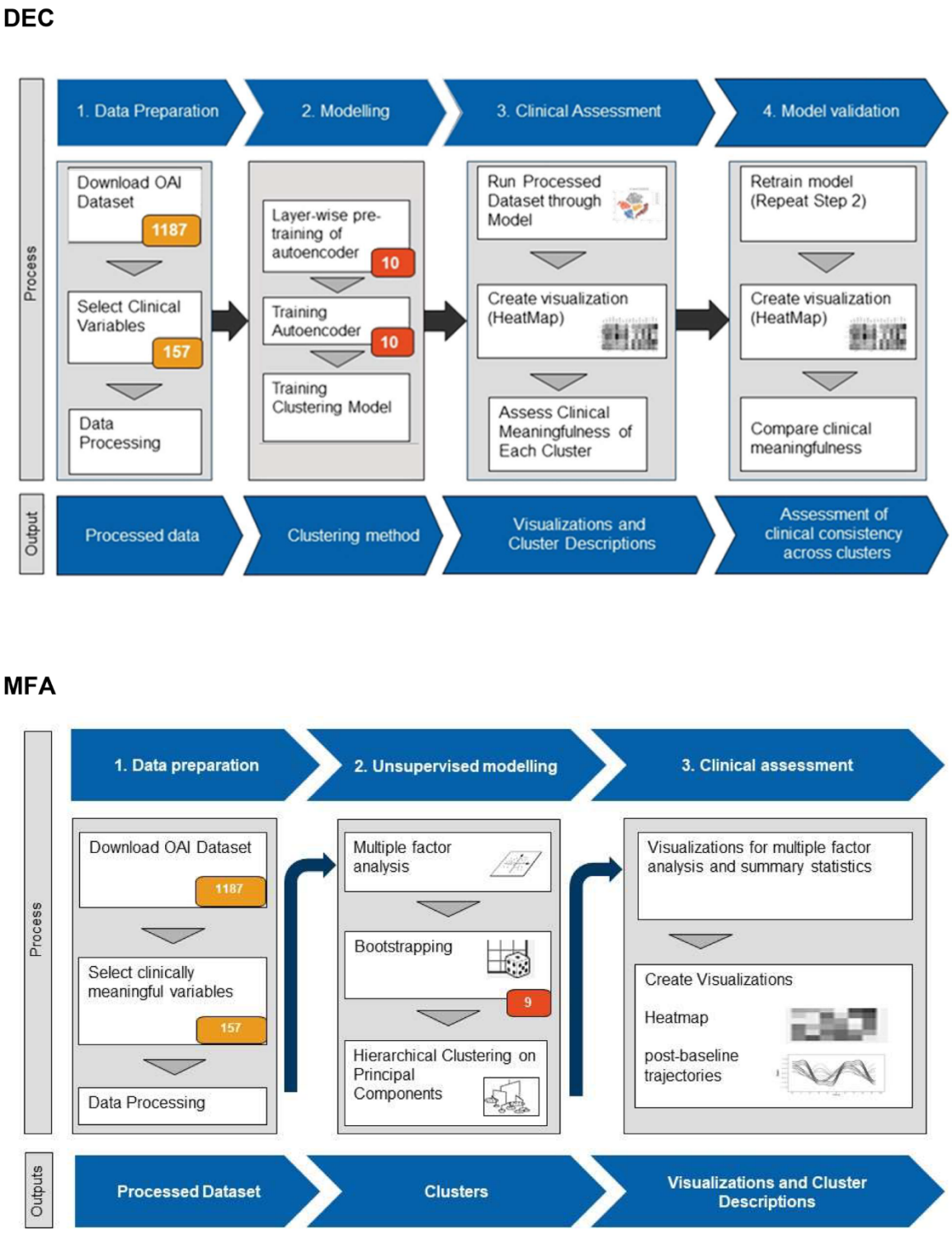

Fig. 1. A schematic diagram of data management from download to analysis (methods visualization).
See text and supplement for more detailed information.

To identify statistically significant differences between cohorts or clusters at baseline, nonparametric Wilcoxon-Mann-Whitney or Kruskal-Wallis rank sum and chi-squared contingency table tests were performed for continuous and categorical variables, respectively. Forty-nine variables were used to compare differences between cohorts and clusters. In addition, cluster pairwise comparisons using 29 baseline variables were performed with the Steel-Dwass-Critchlow-Fligner test [41]. All p-values were adjusted for multiple testing using the Benjamini-Hochberg procedure for differences between cluster or cohort and for post-hoc pairwise comparisons [42]. To characterize and further explore the different clusters' disease severity and progression, i.e., key structural, functional, and pain parameters (PROs with 7-day recall), these were analyzed for differences between clusters both cross-sectionally (at baseline) and longitudinally (over 8 years for radiographic and 8 or 9 years for PRO changes). A linear mixed-effects model with random intercept was used to model the OA progression

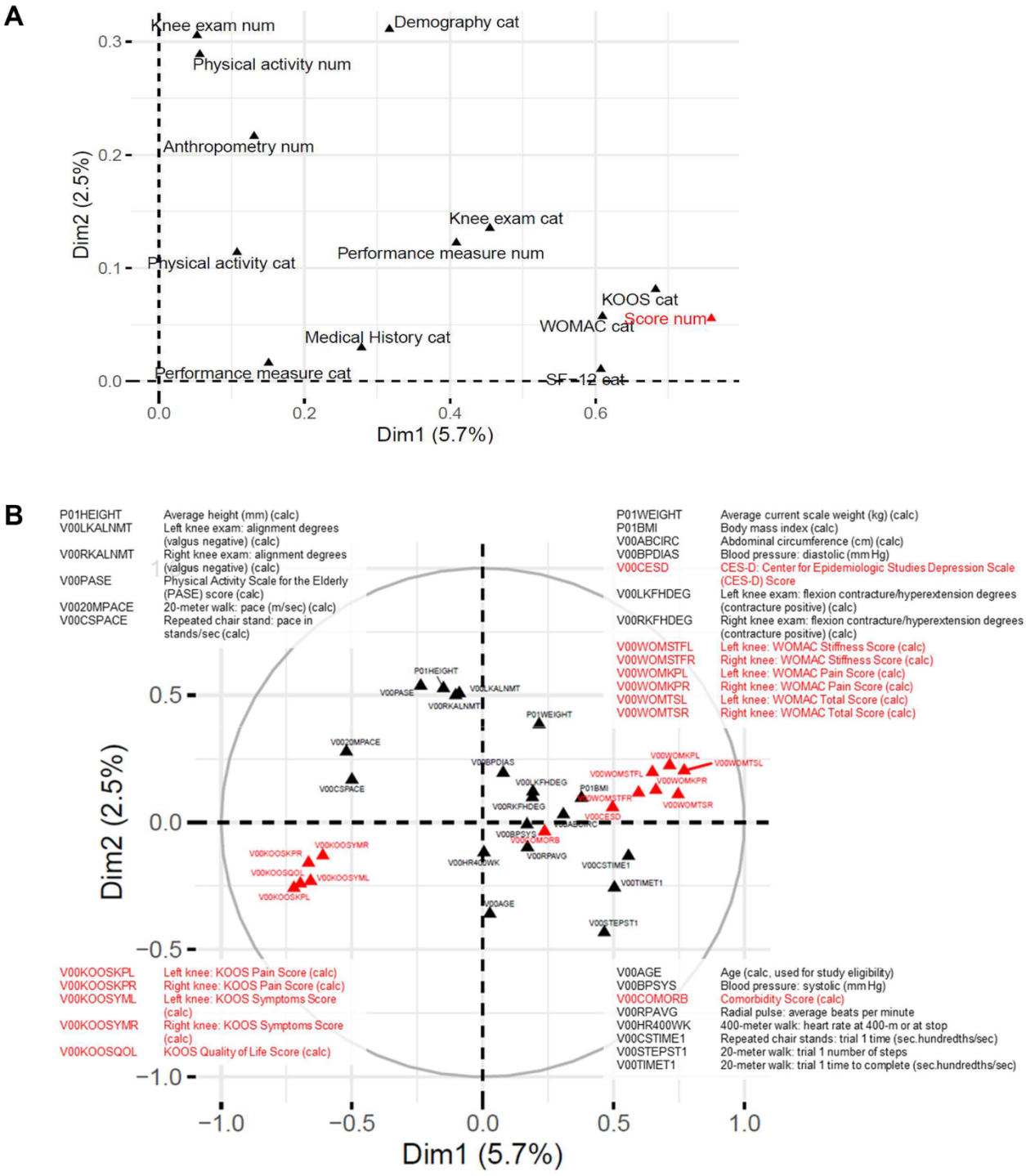

Fig. 2. Variables projection according to MFA. Group representations are displayed in A. Continuous variables are illustrated in B. Dim 1, principal component 1; Dim 2, principal component 2; Dim=dimension or principal component (PC). As PC1 and PC2 represent the first two main factors of MFA they could be interpreted as follows: PC1 summarizes characteristics of disease perception; PC2 summarizes characteristics of clinical picture/patient profile. Variables in black have been used for the construction of the MFA dimensions, red colors correspond to illustrative (supplementary) variables. num=numerical variables, cat=categorical variables.

over time. The random-effects part describes the time course of the different endpoints modeled (PROs or radiographic like fixed medial (fm) joint space width (JSW) at 0.225 and fixed lateral (fl) JSW at 0.775 [43]) for each patient and takes into account the within-subject correlation of different measurements. The model was adjusted for age, gender, BMI (body mass index), abdominal circumference, diabetes, Charlson comorbidity index (CCI), CES-D (Center for Epidemiological Studies Depression) score, presence of multi-site joint pain, pain killer use (for definitions see Table 1a), the PASE (Physical Activity Scale for the Elderly) and knee alignment (valgus/varus/neutral following [44]).

DEC was performed with Python 3.6 for model training and evaluation (using the following packages: numpy==1.12.1, pandas==0.22.0, matplotlib==2.2.2, scikit-learn==0.19.1, Keras==2.1.5, tensorflow==1.1.0, h5py==2.7.1, scipy==1.1.0) and R version 4.1.0 [45] for data visualization. All other statistical computations were carried out in R version 4.1.0 (2021-05-18) [45] using RStudio version 1.4.1717-3 environment [46]. MFA and HCPC were performed using the factoMineR package in R [47]. Missing data was imputed using the

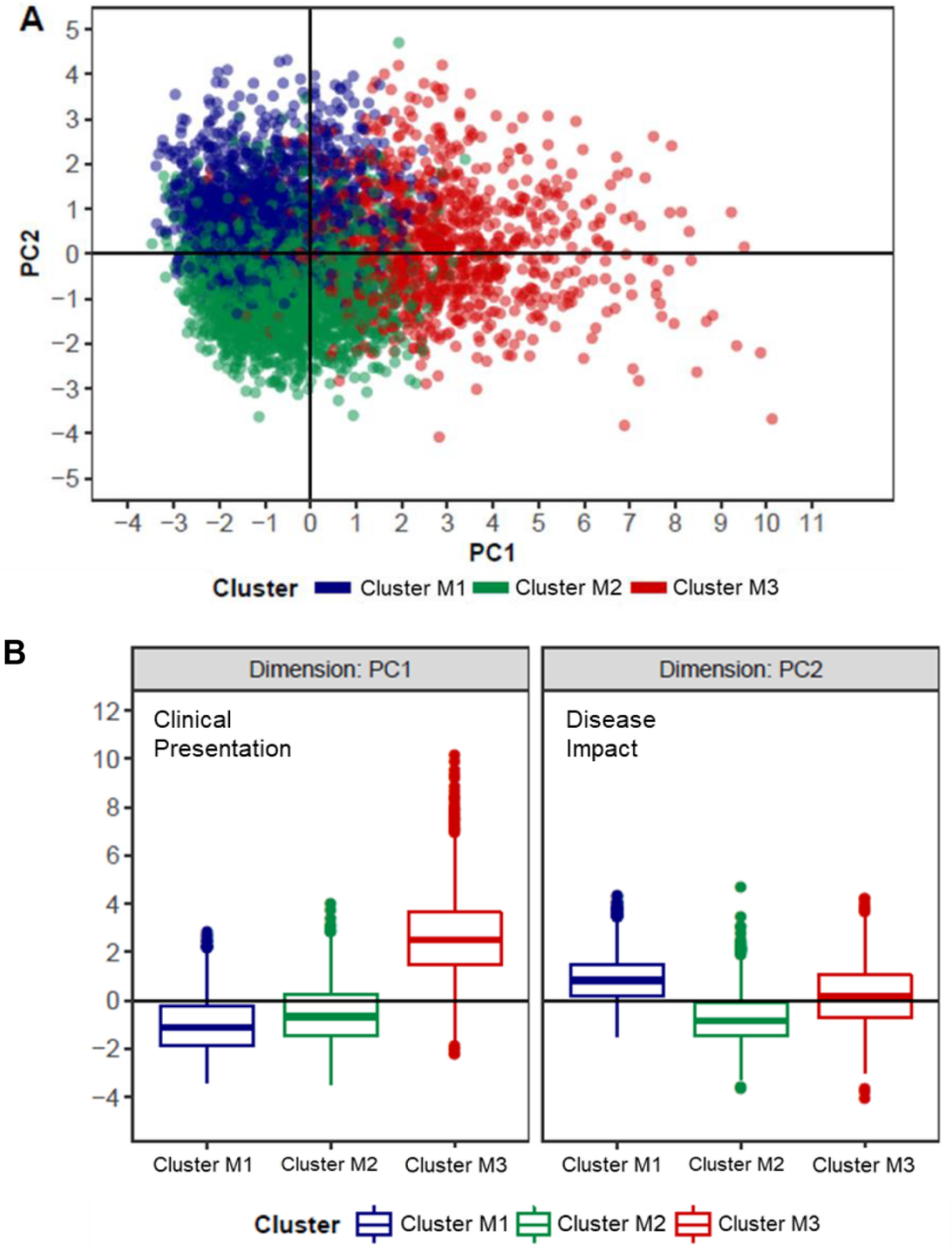

Fig. 3. Subjects projection and distribution according to MFA. A: Subjects are presented as points on the scatter plot created with the first two main dimensions of MFA. Each individual is colored following the cluster type Cluster M1 (blue), Cluster M2 (green) and Cluster M3 (red); PC1 principal component 1; PC2, principal component 2. Cluster M1 and M2 included evaluations with low values of PC1 but high/low values on PC2 for M1 and M2 respectively. M3 included evaluations with particularly high values of PC1 without differences compared to the overall for component 2. B: Boxplots are drawn for PC1 and PC2 in Cluster 1 (red), Cluster 2 (green) and Cluster 3 (blue). Solid line within each box represents the median, and upper and lower hinges represent the 75th and 25th percentiles, respectively. The upper error bar extends from the hinge to the largest value, no further than 1.5 times the inter-quartile range from the hinge. Data beyond the end of the error bar are determined as outlying points and are therefore plotted individually.

missMDA package in R [39].

Results

I Development cohort

The demographic characteristics of both the progression and incidence cohort were largely comparable. There was however a significant difference in BMI and waist circumference and a higher burden of comorbidity and functional impairment in the progression-cohort (Supplemental Table 2). The percentage of missing in the 157 input features was lower than 10% (Supplemental Table 1).

II Exploration of the presence of clusters in the entire OAI population

The analysis indicated five DEC (D1-5) and three MFAC (M1-3) clusters. Supplemental Fig. 2 outlines the results of DEC as a 2D t-SNE (t-Distributed Stochastic Neighbor Embedding) plot, demonstrating separated clusters.

Fig. 2 visualizes the results of MFAC for the first two dimensions. While the first dimension was dominated by data derived from PROs (Fig. 2A), the second dimension contained predominantly variables derived from knee examination, physical activity and anthropometric measures. The component loadings indicated correlations between the principal components and the quantified variables as displayed in Fig. 2B (see supplement for more detail). Nine significant MFA dimensions were retained during the bootstrapping procedure and after the HCPC on these components, subjects were grouped in three clusters based on the inter-cluster inertia gains (Supplemental Fig. 3). A map of subjects with respect to PC1 and 2 is depicted in Fig. 3.

III Separation of the clusters

Table 1a
Characterization of clusters according to patient characteristics at baseline.

Clusters	D1 (N=619)	D2 (N=849)	D3 (N=860)	D4 (N=785)	D5 (N=1551)	p-value	M1 (N=1524)	M2 (N=2146)	M3 (N=1004)	p-value
Age, years	59.0 (9.1)	63.0 (9.3)	60.7 (8.9)	59.3 (8.9)	62.8 (9.0)	< 0.001[1]	60.4 (9.2)	62.4 (9.1)	60.3 (9.1)	< 0.001[1]
No. of females, n (%)	344 (56%)	481 (57%)	635 (74%)	425 (54%)	837 (54%)	< 0.001[2]	146 (10%)	1841 (86%)	742 (74%)	< 0.001[2]
BMI, kg/m^2	27.7 (4.4)	28.9 (4.5)	30.6 (5.4)	27.7 (4.7)	28.5 (4.6)	< 0.001[1]	29.0 (4.0)	27.3 (4.6)	31.3 (5.2)	< 0.001[1]
Waist circumference, cm	98.8 (12.0)	104.7 (12.5)	105.8 (14.0)	99.3 (12.2)	102.7 (12.2)	< 0.001[1]	103.6 (11.1)	99.5 (12.8)	107.3 (13.7)	< 0.001[1]
Charlson comorbidity index	0.3 (0.7)	0.4 (0.9)	0.6 (1.0)	0.3 (0.7)	0.3 (0.8)	< 0.001[1]	0.3 (0.8)	0.3 (0.8)	0.6 (1.0)	< 0.001[1]
Presence of Diabetes, n (%)	38 (6%)	74 (9%)	92 (11%)	35 (5%)	121 (8%)	< 0.001[2]	96 (6%)	134 (6%)	131 (14%)	< 0.001[2]
CES-D score	7.3 (7.8)	6.4 (6.6)	10.1 (8.8)	5.8 (6.4)	5.1 (5.0)	< 0.001[1]	4.5 (4.3)	6.3 (6.7)	10.8 (8.9)	< 0.001[1]
Multi-OAP[3], n (%)	216 (37%)	454 (56%)	439 (56%)	260 (35%)	580 (39%)	< 0.001[2]	604 (42%)	768 (38%)	581 (63%)	< 0.001[2]
PASE score	202.8 (83.2)	153.0 (78.8)	141.4 (78.9)	180.4 (85.0)	148.0 (77.1)	< 0.001[1]	183.7 (86.3)	147.5 (74.2)	152.5 (86.4)	< 0.001[1]
Pain killer use[4], n (%)	243 (39%)	409 (48%)	501 (59%)	269 (34%)	484 (31%)	< 0.001[2]	502 (33%)	781 (36%)	627 (63%)	< 0.001[2]
Total number of medications	3.3 (2.3)	3.8 (2.6)	4.3 (2.9)	3.4 (2.4)	3.5 (2.4)	< 0.001[1]	3.2 (2.2)	3.6 (2.5)	4.3 (2.9)	< 0.001[1]
WOMAC total score (left)	10.5 (15.1)	15.0 (17.2)	24.4 (19.5)	8.3 (12.6)	6.6 (9.4)	< 0.001[1]	7.0 (9.7)	6.5 (8.9)	32.5 (18.8)	< 0.001[1]
WOMAC total score (right)	10.8 (13.6)	14.8 (15.0)	25.4 (17.8)	7.7 (10.3)	7.0 (9.2)	< 0.001[1]	8.1 (10.4)	8.2 (9.9)	28.2 (17.9)	< 0.001[1]
WOMAC pain (left)	2.1 (3.4)	2.9 (3.7)	4.8 (4.4)	1.7 (2.8)	1.2 (2.0)	< 0.001[1]	1.4 (2.1)	1.2 (1.9)	6.5 (4.3)	< 0.001[1]
WOMAC pain (right)	2.3 (3.1)	3.1 (3.5)	5.1 (4.0)	1.6 (2.3)	1.4 (2.1)	< 0.001[1]	1.8 (2.5)	1.7 (2.3)	5.6 (4.1)	< 0.001[1]
WOMAC stiffness (left)	1.3 (1.6)	1.7 (1.8)	2.5 (1.8)	1.1 (1.5)	1.0 (1.3)	< 0.001[1]	1.0 (1.2)	1.0 (1.2)	3.2 (1.8)	< 0.001[1]
WOMAC stiffness (right)	1.4 (1.5)	1.9 (1.7)	2.8 (1.7)	1.2 (1.4)	1.1 (1.3)	< 0.001[1]	1.2 (1.4)	1.3 (1.4)	3.0 (1.7)	< 0.001[1]
KOOS pain (left)	86.3 (18.1)	81.6 (19.6)	72.2 (21.8)	88.5 (15.3)	91.2 (11.6)	< 0.001[1]	89.7 (12.7)	91.4 (11.1)	63.2 (21.2)	< 0.001[1]
KOOS pain (right)	84.8 (16.8)	80.1 (18.2)	69.7 (19.9)	88.6 (12.8)	90.2 (11.6)	< 0.001[1]	87.4 (14.3)	88.5 (12.7)	67.3 (20.2)	< 0.001[1]
KOOS symptoms (left)	88.3 (15.2)	82.8 (17.9)	77.1 (18.5)	89.1 (13.5)	91.8 (10.8)	< 0.001[1]	90.8 (11.9)	91.4 (10.4)	69.5 (18.9)	< 0.001[1]
KOOS symptoms (right)	87.9 (12.3)	82.4 (16.3)	75.3 (17.0)	89.7 (11.3)	91.4 (9.8)	< 0.001[1]	89.4 (12.4)	89.4 (11.2)	73.4 (17.1)	< 0.001[1]
KOOS sport recreation	73.7 (24.5)	64.1 (26.1)	48.6 (26.4)	79.3 (21.3)	81.2 (20.3)	< 0.001[1]	77.3 (22.1)	78.5 (21.0)	41.6 (24.1)	< 0.001[1]
KOOS QoL	69.2 (20.8)	60.3 (22.1)	48.7 (21.1)	73.2 (19.3)	74.3 (18.7)	< 0.001[1]	70.3 (20.1)	73.2 (18.7)	44.7 (19.7)	< 0.001[1]

M1-M3 and p-value columns for MFA clustering. Data are mean (SD) or n (%). CES-D: Center for Epidemiologic Studies Depression. PASE: Physical Activity Scale for the Elderly

[1] Kruskal-Wallis rank sum test (adjusted for multiple comparisons)

[2] Pearson's Chi-squared test (adjusted for multiple comparisons)

[3] Osteoarthritis/degenerative arthritis in at least 2 locations from hip, hand/fingers, knee, back/neck or some other joint plus knee X-ray confirming grade 2 or higher at either one or both sides.

[4] Used of tylenol, (non)-prescription NSAIDS, COXIBS, narcotics, SAMe (S-adenosylmethionine), MSM (methylsulfonylmethane) or Doxycycline for joint pain or arthritis.

DEC clusters

Demographics

Table 1a reports key demographic characteristics of the five clusters obtained by DEC. Subjects in D2 and D5 were significantly older than those in other clusters (Supplemental Table 3, $p<0.001$). D3 showed a skewed gender-distribution (74% females), in the others the gender-distribution was balanced.

Comorbidities and pain

D3-subjects had a mean BMI above 30 kg/m^2 with significant differences compared to all other clusters ($p<0.001$). Overall D3 comprised a "comorbid" phenotype, as indicated by higher waist circumference, a trend toward higher burden of diabetes and other comorbidities (summarized as CCI) and lower activity (PASE), shown in Fig. 4A and Table 1a. D3-subjects suffered the highest pain levels, more multi-site joint pain, had the highest proportion of subjects on analgesics, and the highest mean CES-D scores ($p<0.001$).

D2- and D5-subjects were older (63y average), were relatively physically inactive and had a low burden of comorbidity. They differed in waist circumference, multi-site joint pain, use of analgesic medication and CES-D, with D2 displaying higher values in these areas. While subjects in D5 could be characterized as physically inactive but otherwise relatively healthy elderly, those in D2 were borderline to the "comorbid" phenotype.

Finally, D1- and D4-subjects were younger (59y average), with low pain levels and a higher degree of physical activity. D1 stood out with the highest activity level and frequency of squatting/kneeling as main distinguishing feature from D4.

In summary, the clusters could be categorized as active (D1/D4

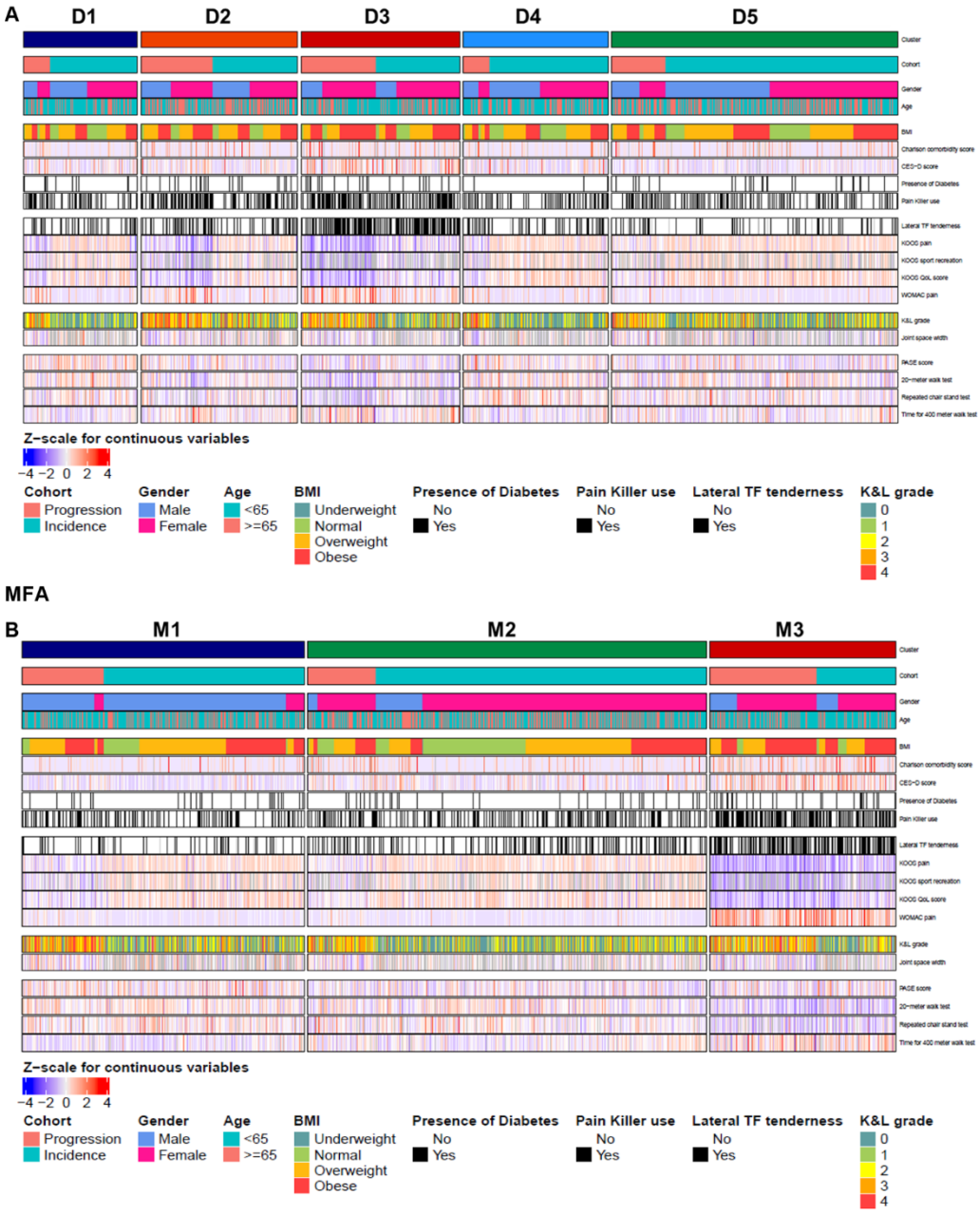

Fig. 4. Heat maps for DEC and MFA clusters. Visualization of key patient characteristics

separated by squatting activities in D1), comorbid (D2/3 separated by a higher prevalence of effusion in D2 and a higher burden of comorbidity in D3) elderly, inactive (D5).

Knee-clinical results

As depicted in Table 1b, D2- and D3-subjects suffered most from knee pain. They showed a tendency for periarticular pain, discomfort while walking and multi-site joint pain, reflected in the lowest functional performance ($p<0.001$). This was associated with an impaired quality of life (KOOS QoL) for D3-subjects, compared to all other clusters ($p<0.001$ – Table 1b). The distinguishing feature in D2 was a high rate of subjects with joint effusion (potentially indicative of an inflammatory phenotype).

Subjects in D1, D5 and D4 were similar in their clinical presentation. D1-subjects showed periarticular symptoms and a comparatively low KOOS QoL, in spite of their preserved physical activity.

Imaging

The fmJSW and flJSW was smallest in D2 and D3 (4.7mm and 4.6mm medially, 6.5mm and 6.1mm laterally). All clusters had a mean fmJSW above 4.5mm with relatively small differences, while the baseline variability laterally was higher with a mean flJSW above 6.2mm. The different clusters identified by DEC did not reveal major differences in the prevalence of malalignment. Varus malalignment was reported for 43% of patients (27% valgus, 30% neutral).

Clinical progression

The trajectories for pain levels were dependent on the measure. Overall KOOS sports/recreation, KOOS QoL and global rating scale (GLRS) [48] were stable over time (Fig. 5). As shown in Fig. 6, KOOS pain and WOMAC pain after an initial improvement, were stable with a maximum change of 11.4 and 1.03 points respectively, below the usually reported minimal clinically important differences (MCID) [49]. For the pain numeric rating scale (NRS), an increase between 0.57 and 0.83

Table 1b
Characterization of clusters for clinical and imaging results at baseline.

Clusters	D1 (N=619)	D2 (N=849)	D3 (N=860)	D4 (N=785)	D5 (N=1551)	p-value	M1 (N=1524)	M2 (N=2146)	M3 (N=1004)	p-value
Clinical information										
K&L grade (left), n (%)						< 0.001[2]				< 0.001[2]
0	276 (47%)	199 (25%)	236 (31%)	368 (49%)	590 (40%)		560 (39%)	896 (44%)	218 (24%)	
1	111 (19%)	132 (17%)	114 (15%)	148 (20%)	275 (19%)		257 (18%)	403 (20%)	122 (14%)	
2	140 (24%)	250 (32%)	232 (31%)	144 (19%)	370 (25%)		333 (23%)	494 (24%)	310 (35%)	
3	54 (9%)	153 (19%)	150 (20%)	67 (9%)	201 (14%)		233 (16%)	203 (10%)	189 (21%)	
4	10 (2%)	56 (7%)	28 (4%)	18 (2%)	32 (2%)		61 (4%)	25 (1%)	58 (6%)	
K&L grade (right), n (%)						< 0.001[2]				< 0.001[2]
0	264 (45%)	176 (22%)	207 (27%)	367 (49%)	566 (39%)		521 (36%)	845 (42%)	216 (24%)	
1	111 (19%)	128 (16%)	132 (17%)	132 (18%)	266 (18%)		260 (18%)	380 (19%)	131 (15%)	
2	141 (24%)	270 (34%)	263 (35%)	157 (21%)	402 (28%)		372 (26%)	528 (26%)	335 (38%)	
3	65 (11%)	166 (21%)	123 (16%)	72 (10%)	182 (13%)		217 (15%)	220 (11%)	172 (19%)	
4	9 (2%)	56 (7%)	32 (4%)	14 (2%)	39 (3%)		70 (5%)	42 (2%)	39 (4%)	
Medial Joint space width, mm (at 0.225mm)*	4.9 (1.4)	4.7 (1.6)	4.6 (1.4)	5.0 (1.3)	4.8 (1.4)	< 0.001[1]	5.0 (1.7)	4.8 (1.2)	4.5 (1.5)	< 0.001[1]
Lateral Joint space width, mm (at 0.775mm)*	6.8 (1.6)	6.1 (2.0)	6.5(1.8)	6.7 (1.7)	6.7 (1.7)	< 0.001[1]	7.2 (1.8)	6.3 (1.6)	6.0 (1.9)	< 0.001[1]
Knee alignment (left), n (%)						< 0.001[2]				< 0.001[2]
Neither	211 (34%)	201 (24%)	255 (30%)	231 (30%)	476 (31%)		457 (30%)	635 (30%)	287 (29%)	
Varus	244 (39%)	408 (48%)	390 (46%)	337 (43%)	639 (42%)		453 (30%)	1093 (51%)	473 (47%)	
Valgus	163 (26%)	233 (28%)	207 (24%)	213 (27%)	421 (27%)		601 (40%)	403 (19%)	237 (24%)	
Knee alignment (right), n (%)						< 0.001[2]				< 0.001[2]
Neither	212 (34%)	200 (24%)	247 (29%)	234 (30%)	471 (31%)		438 (29%)	630 (30%)	299 (30%)	
Varus	236 (38%)	413 (49%)	387 (46%)	328 (42%)	619 (41%)		447 (30%)	1088 (51%)	452 (46%)	
Valgus	170 (28%)	229 (27%)	213 (25%)	215 (28%)	434 (28%)		624 (41%)	405 (19%)	234 (24%)	
Flexion tenderness (left), n (%)	98 (16%)	249 (30%)	272 (32%)	109 (14%)	129 (8%)	< 0.001[2]	208 (14%)	220 (10%)	432 (43%)	< 0.001[2]
Flexion tenderness (right), n (%)	85 (14%)	236 (28%)	281 (33%)	77 (10%)	124 (8%)	< 0.001[2]	213 (14%)	260 (12%)	337 (34%)	< 0.001[2]
Patellar quadriceps tendinitis (left), n (%)	96 (16%)	141 (17%)	310 (36%)	86 (11%)	72 (5%)	< 0.001[2]	76 (5%)	217 (10%)	417 (42%)	< 0.001[2]
Patellar quadriceps tendinitis (right), n (%)	90 (15%)	136 (16%)	297 (35%)	89 (11%)	72 (5%)	< 0.001[2]	89 (6%)	261 (12%)	341 (34%)	< 0.001[2]
Medial TF tenderness (left), n (%)	169 (27%)	257 (31%)	453 (53%)	152 (19%)	201 (13%)	< 0.001[2]	181 (12%)	495 (23%)	562 (56%)	< 0.001[2]
Medial TF tenderness (right), n (%)	156 (25%)	241 (29%)	456 (54%)	148 (19%)	175 (11%)	< 0.001[2]	188 (12%)	492 (23%)	500 (51%)	< 0.001[2]
Lateral TF tenderness (left), n (%)	119 (19%)	160 (19%)	354 (41%)	115 (15%)	144 (9%)	< 0.001[2]	119 (8%)	339 (16%)	439 (44%)	< 0.001[2]
Lateral TF tenderness (right), n (%)	127 (21%)	193 (23%)	395 (47%)	120 (15%)	167 (11%)	< 0.001[2]	141 (9%)	434 (20%)	431 (44%)	< 0.001[2]
Anserine bursa tenderness (left), n (%)	144 (23%)	271 (32%)	413 (48%)	163 (21%)	206 (13%)	< 0.001[2]	183 (12%)	557 (26%)	462 (46%)	< 0.001[2]
Anserine bursa tenderness (right), n (%)	127 (21%)	244 (29%)	410 (48%)	139 (18%)	172 (11%)	< 0.001[2]	150 (10%)	532 (25%)	415 (42%)	< 0.001[2]
Effusion bulge sign positive (left), n (%)	62 (10%)	501 (60%)	14 (2%)	49 (6%)	0 (0%)	< 0.001[2]	219 (15%)	219 (10%)	191 (19%)	< 0.001[2]
Effusion bulge sign positive (right), n (%)	86 (14%)	579 (69%)	5 (1%)	27 (3%)	0 (0%)	< 0.001[2]	244 (16%)	265 (13%)	192 (20%)	< 0.001[2]
Effusion patellar tap positive (left), n (%)	20 (3%)	111 (13%)	29 (3%)	23 (3%)	29 (2%)	< 0.001[2]	50 (3%)	88 (4%)	74 (8%)	< 0.001[2]
Effusion patellar tap positive (right), n (%)	24 (4%)	136 (16%)	21 (2%)	26 (3%)	31 (2%)	< 0.001[2]	60 (4%)	99 (5%)	79 (8%)	< 0.001[2]
Anamnestic information										
Discomfort walking, n (%)	76 (13%)	138 (17%)	267 (33%)	79 (10%)	180 (12%)	< 0.001[2]	112 (7%)	317 (15%)	314 (34%)	< 0.001[2]
Frequent squatting, n (%)	478 (77%)	25 (3%)	15 (2%)	100 (13%)	26 (2%)	< 0.001[2]	212 (14%)	309 (14%)	124 (12%)	0.304[2]
Frequent kneeling, n (%)	591 (95%)	146 (17%)	149 (17%)	55 (7%)	218 (14%)	< 0.001[2]	413 (27%)	564 (26%)	182 (18%)	< 0.001[2]
Performance										
20-meter walk test, m/sec	1.4 (0.2)	1.3 (0.2)	1.2 (0.2)	1.4 (0.2)	1.3 (0.2)	< 0.001[1]	1.4 (0.2)	1. 3 (0.2)	1.2 (0.2)	< 0.001[1]
Repeated chair stand test, stands/sec	0.5 (0.2)	0.5 (0.1)	0.4 (0.1)	0.5 (0.1)	0.5 (0.1)	< 0.001[1]	0.5 (0.1)	0.5 (0.1)	0.4 (0.1)	< 0.001[1]
Time for 400 meter walk test, sec	295.0 (45.5)	316.1 (64.4)	330.3 (71.3)	292.5 (45.6)	302.2 (49.9)	< 0.001[1]	288.2 (42.6)	305.3 (48.1)	341.9 (78.3)	< 0.001[1]

M1-M3 and p-value columns for MFA clustering. Data are mean (SD) or n (%). K&L: Kellgren–Lawrence. TF: Tibio-femoral, *values for the more affected side
[1] Kruskal-Wallis rank sum test (adjusted for multiple comparisons)
[2] Pearson's Chi-squared test (adjusted for multiple comparisons)

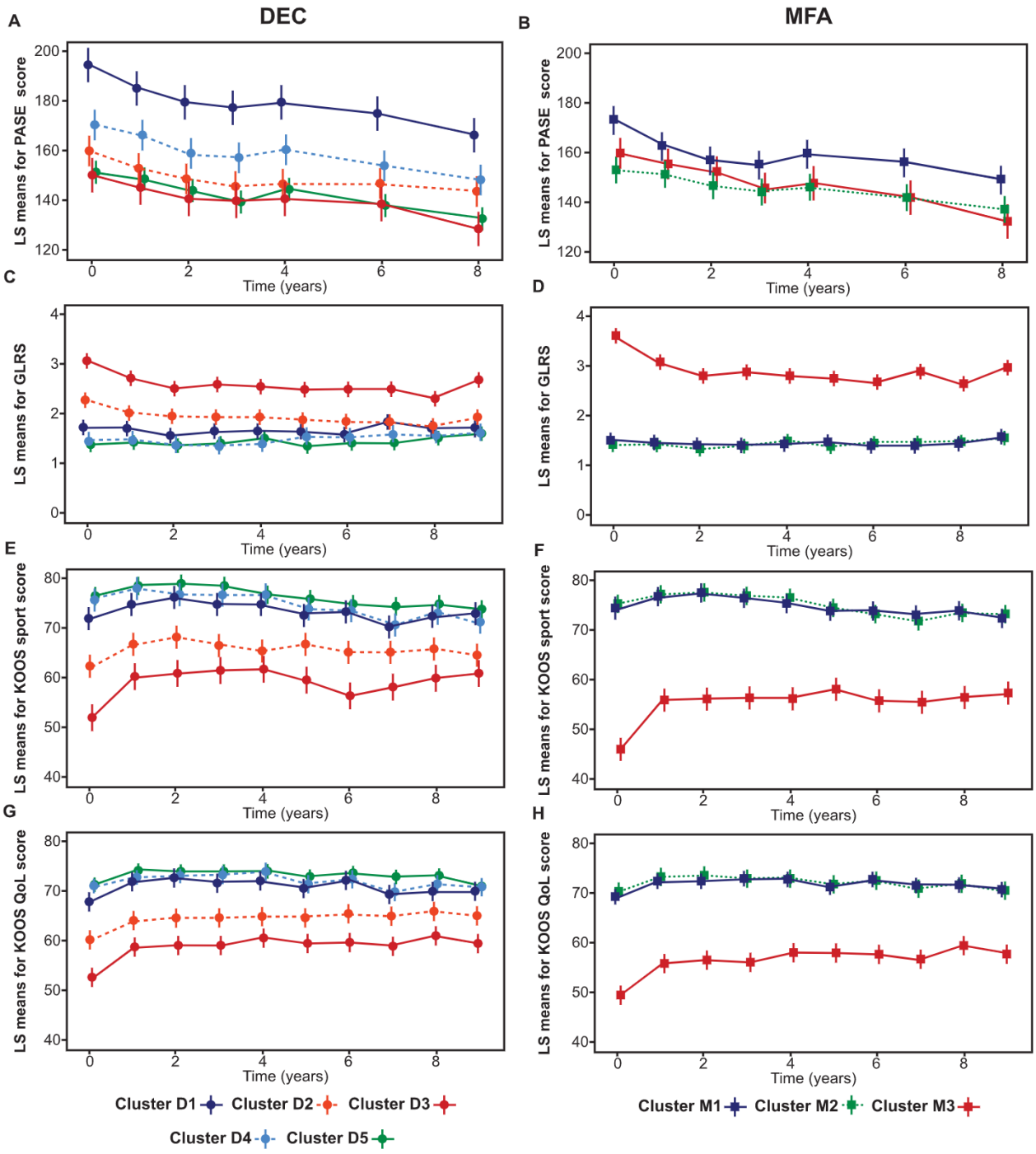

Fig. 5. Longitudinal course of activity and measures of disease perception over 8 and 9 years respectively. Linear mixed models for DEC (left panel) and MFA (right panel) including clusters, time points, gender, diabetes, presence of multi-site joint pain, pain killer use and knee alignment as factors and age, BMI, abdominal circumference, Charlson comorbidity index, CES-D and with or without PASE as continuous covariates. Clusters by time points was included as an interaction term in the model. Dots/squares and error bars represent the estimated mean and the 95% confidence limits, see supplement for individual comparison of clusters and confidence intervals. Further information is provided in the supplemental material.

comparatively close to the usually reported MCID of 1 [50] was observed for all clusters except D3 (0.33). In parallel, PASE (activity) decreased in all clusters. For PASE the MCID has been reported to be between 17-25 points (depending on methodology [51]), which was reached by all clusters except D2. Detailed statistics are provided in the supplemental material.

Radiographic progression

Radiographic progression medially was 0.56-0.68mm over 8 years for all clusters. D2 was most affected medially with a 0.68mm loss, laterally with 0.7mm, while the average in the other clusters laterally was between 0.36-0.55mm (Figs. 7 and supplemental material for detailed statistical evaluation).

MFAC clusters

Demographics

As shown in Table 1a and supplemental Table 4, subjects in M2 were

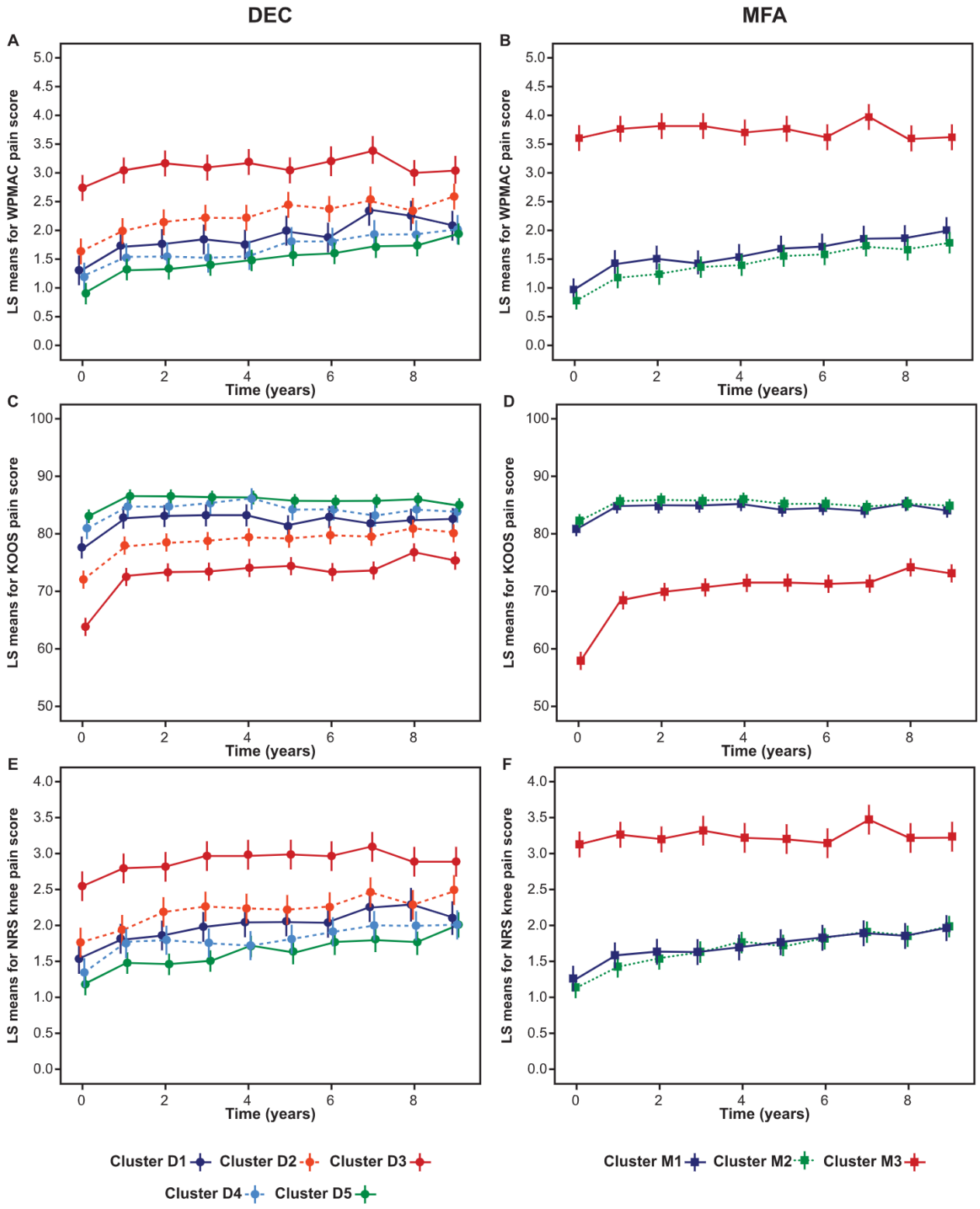

Fig. 6. Longitudinal course of pain over 9 years. Linear mixed models for DEC (left panel) and MFA (right panel) including clusters, time points, gender, diabetes, presence of multi-site joint pain, pain killer use and knee alignment as factors and age, BMI, abdominal circumference, Charlson comorbidity index, CES-D and PASE as continuous covariates. Clusters by time points was included as an interaction term in the model. Dots/squares and error bars represent the estimated mean and the 95% confidence limits, see supplement for individual comparison of clusters and confidence intervals. Further information is provided in the supplemental material.

significantly older than subjects in the two other clusters ($p<0.001$). M2 and M3 included predominantly women (86% and 74%, respectively), while M1 mainly consisted of men (90%).

Comorbidities and pain

Similar to D3, M3-subjects showed a higher mean BMI, higher waist circumference, a higher incidence of diabetes, comorbidities, multi-site joint pain and high CES-D-scores ($p<0.001$) as visualized in Fig. 4B. The proportion of frequent pain-killer use was two-fold higher in M3 than in

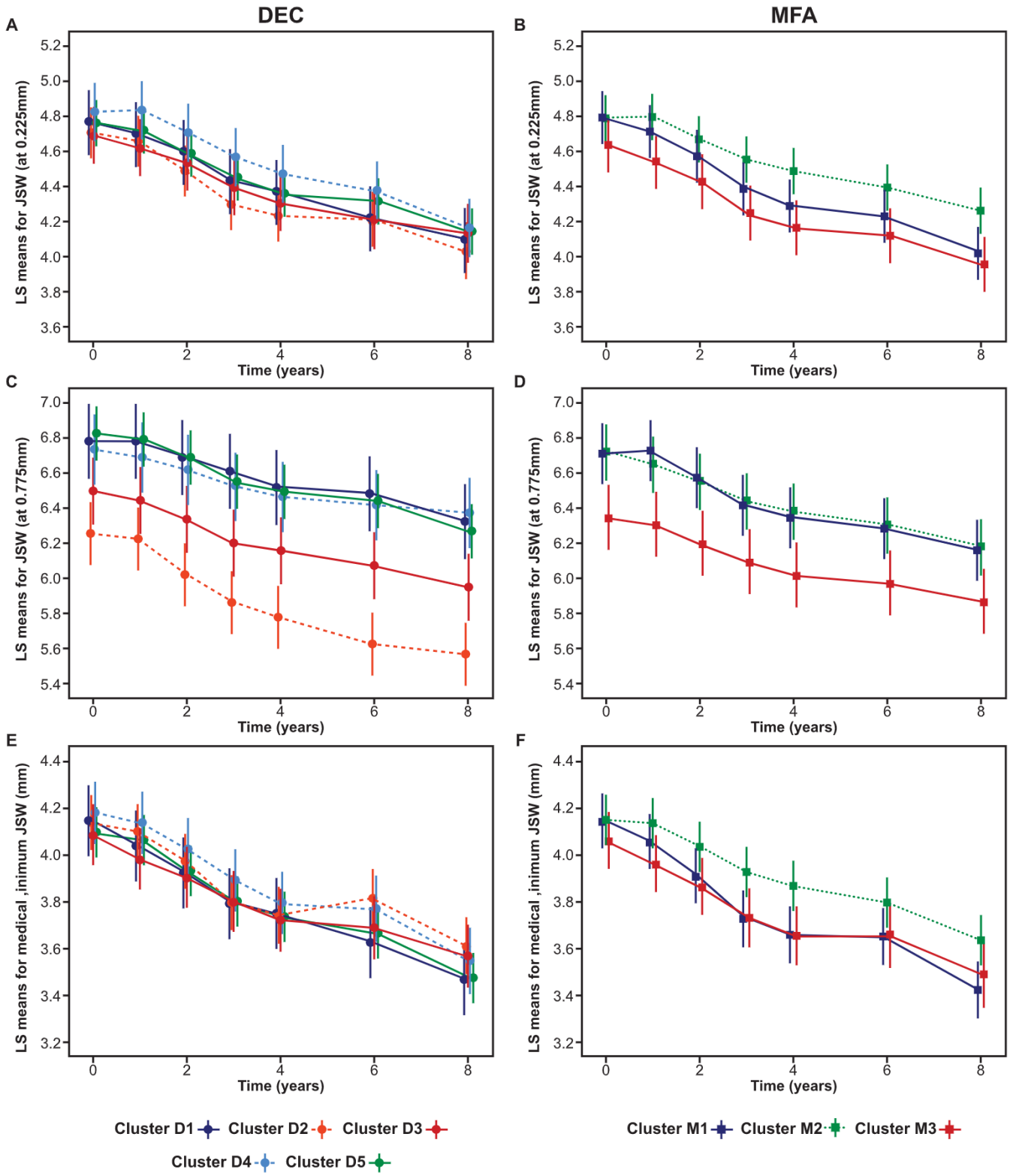

Fig. 7. Structural progression over 8 years. Linear mixed models for DEC (left panel) and MFA (right panel) including clusters, time points, gender, diabetes, presence of multi-site joint pain, pain killer use and knee alignment as factors and age, BMI, abdominal circumference, Charlson comorbidity index, CES-D and PASE as continuous covariates. Clusters by time points was included as an interaction term in the model. Dots/squares and error bars represent the estimated mean and the 95% confidence limits. Further information is provided in the supplemental material.

the others (Table 1b). M2- and M3-subjects had similarly low levels of activity (PASE), while M1 proved most active. In summary, the clusters could be categorized as male active (M1), elderly, inactive (M2) and comorbid (M3).

Knee-clinical results

Comparable to D3, M3-subjects suffered from periarticular symptoms and showed relevant functional impairments ($p<0.001$). M1 and M2 did not differ based on the clinical examination; M2-subjects

performed slightly less well in the functional tests. Noteworthy though, M1 comprising predominantly physically active men showed a significantly higher prevalence of valgus malalignment compared to the others ($p<0.001$).

Imaging

Both the fmJSW and flJSW were smallest in M3 with approx. 4.8mm and 6.0mm, respectively. The baseline values for M1 and M2 were 5.0mm and 4.5mm medially, and 7.2mm and 6.3 mm laterally.

Clinical progression

While global measures were stable over time, pain measures again changed depending on the measure used (Figs. 5 and 6). As described for DEC, the clusters were stable or below the MCID for KOOS pain and WOMAC pain. Changes in NRS pain were more pronounced in M1 and M2 (0.7 and 0.85), while there were no changes for M3 (0.09, at a higher pain level). The limited changes in pain levels were associated with a decrease in activity that reached MCID for M1 and M3 (24.07 and 27.08).

Radiographic progression

All clusters also showed structural progression over time with changes in fmJSW of 0.77mm, 0.53mm and 0.68mm for M1-3 respectively and in flJSW around 0.5mm (Fig. 7, and supplement).

Discussion

The study evaluated the utility and relevance of two unsupervised ML-based methods for identifying patient clusters based on clinical information in a large population with knee pain. Both methods have identified similar clusters discriminated by obesity, comorbidities, physical activity, and disease impact. The advantage of using ML-based methods to come to this conclusion is the opportunity of an unbiased appraisal of a large quantity of various types of data difficult to handle with standard approaches.

The two clustering approaches have yielded comparable results. Differences to previous reports is likely relate to the use of different input variables, rather than the techniques applied. Previous reports were based on imaging data and WOMAC in a selected patient population [25] from the incident [26] or progression cohort [18]. A certain robustness of the above approach is implied by the similarity to the clusters delineated by Knoop et al. [27] who unlike other groups chose input variables from various dimensions (demographics, pain, biomechanics, etc.). This underlines the impact of multidimensionality, selection and underlying assumptions in clustering (for a review see [52]). The prognostic value of the above stated baseline characteristics on changes in mJSW as described previously [26,53] could not be confirmed in the present analysis, possibly due to the inclusion of both, incident- and progression-cohort. Neither could we detect a clear correlation between pain and structural changes for any cluster, similar to previous reports [26].

Both methods described here delineate a “comorbid” D3/M3 comprised of patients with multiple comorbidities, multi-site joint pain, and peri-articular soft tissue knee pain. This profile suggests that KOA may not be the real symptom-driver in these patients. Abdominal obesity has a known metabolic and pro-inflammatory impact [54] increasing the risk of symptomatic OA disproportionately to corresponding joint loads [27,28,55,56], and is often associated with insulin resistance, muscle atrophy, accelerated atherogenesis and poor microcirculation, [54,57-59]. Their characteristics suggest that D3/M3-subjects may be less likely to benefit from structure-modifying OA drugs, especially from intra-articular compounds. In addition, detecting changes in pain/function might present a challenge in these patients given the variability in pain reduction observed even for substances with proven analgesic effect [60]. D3/M3 patients might however benefit from systemic anti-inflammatory treatment [61].

Cluster D2 had the highest degree of baseline joint space loss and radiographic progression especially laterally. This cluster is characterized by a high proportion of participants with knee effusion at baseline. This could imply both, degeneration as inflammatory trigger and inflammation as trigger for disease progression. Most probably the observation is the manifestation of a vicious circle alternatingly perpetuated by both. Further prospective analyses in similar patients are needed to clarify the association, especially since relevant loss of JSW is typically defined as 0.5mm within shorter timeframes [62]. Assuming though a clinically relevant association between inflammation and degradation in D2-type patients they might be specifically susceptible to treatments with combined local anti-inflammatory and possibly chondroanabolic effects.

D1/4 and M1 present attractive targets for DMOAD trials with chondro-anabolic or anti-catabolic modes of action. Although changes are less pronounced in D1 and D4 compared to M1, the higher activity and increased knee bending activities (D1) might make them eligible for intra-articular DMOAD treatment. Increased bending is known to be associated with frequent presence of cartilage lesions, accelerated progression of cartilage and meniscal lesions, already present at early stages of the disease [63–66]. Compared to D3/M3, these active clusters (D1/M1 and D4) indeed seem to have a slightly faster progression of joint space narrowing, which was more pronounced in MFAC. Earlier effects might be observable with more sensitive imaging methods, e.g., quantitative MRI, meniscal structure assessments or other, thereby also contributing to a reduction in trial duration.

The initial improvement in pain observed in all clusters does not necessarily mirror the individual experience, but might be a regression to the mean effect, which has been previously described for the OAI dataset [67] given the high variability of values. Another explanation is a Hawthorne effect [68,69], with study entry prompting subjects to seek medical care or implement life-style modifications, especially given the detailed assessment of nutritional habits and physical activity at baseline.

Long-term pain trajectories show diverging results with stable levels for KOOS and WOMAC and an increase in NRS, underlining the multidimensionality of pain. The differences in PRO results also imply a varying sensitivity to change, while the concomitant decrease in activity as evidenced by PASE highlight the importance of assessing potentially confounding factors, to draw valid conclusions from pain assessments [70]. This raises the fundamental question of which assessment, or battery of assessments to use to capture patients’ perception of disease burden with reasonable effort and probability of detecting true change.

Limitations

The reliance on the OAI dataset implies several limitations. The dataset is limited in size necessitating the preselection of input-variables to adequately train and test potential algorithms. In addition, the validity of conventional radiographic measurements might be limited. Furthermore, the OAI database may suffer from selection bias, e.g., less than 10% of subjects suffer from diabetes, which does not necessarily reflect the global population of patients with OA. A further limitation is the relatively small proportion of subjects entering with confirmed clinical and radiographic KOA (29.7%). This study only evaluates one cohort, limiting the generalizability of the conclusions.

There are also limitations to our clustering methods. In general, the clustering for phenotype identification is limited by the amount and diversity of patient data points, meaning that the clusters will represent the differences in the available data and cannot extrapolate to medical relations that are not represented in the data. Further the prespecified number of clusters was identified using statistical methods, therefore their interpretation and meaning must be put into context by clinicians and could lead to the combination of clusters based on their characteristics depending on the use case. A validation of the results was achieved by using two independent clustering approaches to compare the

resulting clusters as well as by putting the observed results in context with the current state of knowledge. Further multiple random initializations were conducted to proof convergence of the clustering. To evaluate the quality of the findings and clinical utility of these clusters external validation is needed. The choice of the variables used for MFA groups, the number of components retained to perform the clustering as well as the determination of the optimal number of clusters (inertia gain ratio for MFA vs. other criteria to assess internal and stability measures like Dunn index and average proportion of non-overlap) may change the proportion of patients in the different cluster. Other clustering approaches may also have been considered (model-based clustering). Despite these limitations, our study confirms that both methods identified a distinct and potentially clinically relevant group of patients D3/M3.

Conclusion

The study demonstrates the possibility of using ML-based methods to identify different potentially meaningful phenotypes in patients with KOA from complex multi-dimensional datasets such as OA registries or longitudinal studies. While individual clusters could be correlated with clinical or structural progression, the results also underline the complexity of measuring pain and patient-centered outcomes with regards to their inherent variability. Further studies are needed to test the applicability of algorithms on other datasets and evaluate their prognostic and predictive value alone or in combination with laboratory, biomechanical or imaging biomarkers.

Author contributions

DD, PL, LT, RR, PC and MS have mainly contributed to conception and design of the study. All authors have been involved in the analysis and interpretation of the data. Statistical and data science expertise was provided by DD, PL, PN, IR, DB and SP. DD, FS, PL, LT and MS have drafted the manuscript. All authors have contributed to, critically revised and approved the final version of the manuscript. Franziska Saxer (franziska.saxer@novartis.com) and Matthias Schieker (matthias.schieker@novartis.com) take responsibility for the integrity of the work as a whole, from inception to finished article.

Source of data and code

The clinical data, patient reported outcomes and additional analyses are publicly available after registration from https://nda.nih.gov/oai/ (access October 18th, 2022). The source code for DEC is available from https://github.com/piiswrong/dec (access October 24th, 2022) adapted to facilitate working with the Keras Package instead of the Caffe Package as reported by Xie et al.[29], for MFA refer to Le et al 2008 [47].

Role of the funding source

The analysis was founded by the Novartis Institutes of Biomedical Research (BASICHR0042). The Funder had no influence on the study design, data interpretation or publication strategy.

Declaration of Competing Interest

David Demanse is employee and shareholder of Novartis. Franziska Saxer is employee and shareholder of Novartis, she is affiliated to the University Basel and member of the European Union Medical Devices - Expert Panel section Orthopaedics, traumatology, rehabilitation, rheumatology. Patrick Lustenberger is employee of IBM, Switzerland and shareholder of Novartis. László B. Tankó was employee of Novartis and held shares, he is current employee of Bayer Pharmaceuticals. Philipp Nikolaus was employee of IBM, Switzerland, he is current employee of MSCI Inc. Ilja Rasin is employee of IBM, Switzerland. Damian F. Brennan is a former employee of IBM, Switzerland and in this employment consultant to Novartis. He currently is employee of the Macquarie Group, Australia. Ronenn Roubenoff is employee and shareholder of Novartis. Sumehra Premji is a former employee of IBM Switzerland and in this employment, consultant and account partner to Novartis. She is current employee and shareholder of Novartis. Philip G Conaghan reports fees for speaker's bureaus (AbbVie, Novartis) and consultancies (AstraZeneca, BMS, Eli Lilly, Galapagos, Genascence, GSK, Merck, Novartis, Pfizer, Regeneron, Stryker and UCB). Matthias Schieker is employee and shareholder of Novartis and owner LivImplant GmbH.

Acknowledgments

The authors like to thank the participants, investigators and funders of the OAI database, a public–private partnership comprising five contracts (N01-AR-2–2258; N01-AR-2–2259; N01-AR-2–2260; N01-AR-2–2261; N01-AR-2–2262) funded by the NIH, and conducted by the OAI Study Investigators. Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI), a data repository housed within the NIMH Data Archive (NDA).

The authors further would like to acknowledge Alexander Goehler and the Joint Bone and Tendon Group for their throughout review of and valuable suggestions on the manuscript, Rajeeb Gosh for his constant medical writing support and Priyanka Malla, Dhanya Mukundan and John Gallager for their prompt support on the visualizations.

Philip G Conaghan is supported in part through the NIHR Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2022.152140.

References

[1] Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019;393:1745–59.
[2] Wenham CY, Conaghan PG. New horizons in osteoarthritis. Age Ageing 2013;42:272–8.
[3] Dell'Isola A, Allan R, Smith SL, Marreiros SS, Steultjens M. Identification of clinical phenotypes in knee osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord 2016;17:425.
[4] Deveza LA, Melo L, Yamato TP, Mills K, Ravi V, Hunter DJ. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. Osteoarthritis Cartilage 2017;25:1926–41.
[5] Mobasheri A, Saarakkala S, Finnila M, Karsdal MA, Bay-Jensen AC, van Spil WE. Recent advances in understanding the phenotypes of osteoarthritis. F1000Res 2019;8.
[6] Mobasheri A, van Spil WE, Budd E, Uzieliene I, Bernotiene E, Bay-Jensen AC, et al. Molecular taxonomy of osteoarthritis for patient stratification, disease management and drug development: biochemical markers associated with emerging clinical phenotypes and molecular endotypes. Curr Opin Rheumatol 2019;31:80–9.
[7] Roemer FW, Guermazi A, Niu J, Zhang Y, Mohr A, Felson DT. Prevalence of magnetic resonance imaging-defined atrophic and hypertrophic phenotypes of knee osteoarthritis in a population-based cohort. Arthritis Rheum 2012;64:429–37.
[8] Bartlett SJ, Ling SM, Mayo NE, Scott SC, Bingham 3rd CO. Identifying common trajectories of joint space narrowing over two years in knee osteoarthritis. Arthritis Care Res (Hoboken) 2011;63:1722–8.
[9] Doss F, Menard J, Hauschild M, Kreutzer HJ, Mittlmeier T, Muller-Steinhardt M, et al. Elevated IL-6 levels in the synovial fluid of osteoarthritis patients stem from plasma cells. Scand J Rheumatol 2007;36:136–9.
[10] Otterness IG, Swindell AC, Zimmerer RO, Poole AR, Ionescu M, Weiner E. An analysis of 14 molecular markers for monitoring osteoarthritis: segregation of the markers into clusters and distinguishing osteoarthritis at baseline. Osteoarthritis Cartilage 2000;8:180–5.
[11] Berry PA, Maciewicz RA, Cicuttini FM, Jones MD, Hellawell CJ, Wluka AE. Markers of bone formation and resorption identify subgroups of patients with clinical knee osteoarthritis who have reduced rates of cartilage loss. J Rheumatol 2010;37:1252–9.
[12] Blumenfeld O, Williams FM, Hart DJ, Spector TD, Arden N, Livshits G. Association between cartilage and bone biomarkers and incidence of radiographic knee

osteoarthritis (RKOA) in UK females: a prospective study. Osteoarthritis Cartilage 2013;21:923–9.
[13] Berry PA, Maciewicz RA, Wluka AE, Downey-Jones MD, Forbes A, Hellawell CJ, et al. Relationship of serum markers of cartilage metabolism to imaging and clinical outcome measures of knee joint structure. Ann Rheum Dis 2010;69:1816–22.
[14] Egsgaard LL, Eskehave TN, Bay-Jensen AC, Hoeck HC. Arendt-Nielsen L. Identifying specific profiles in patients with different degrees of painful knee osteoarthritis based on serological biochemical and mechanistic pain biomarkers: a diagnostic approach based on cluster analysis. Pain 2015;156:96–107.
[15] Holla JF, van der Leeden M, Heymans MW, Roorda LD, Bierma-Zeinstra SM, Boers M, et al. Three trajectories of activity limitations in early symptomatic knee osteoarthritis: a 5-year follow-up study. Ann Rheum Dis 2014;73:1369–75.
[16] Carlesso LC, Neogi T. Identifying pain susceptibility phenotypes in knee osteoarthritis. Clin Exp Rheumatol 2019;37(Suppl 120):96–9.
[17] Cruz-Almeida Y, King CD, Goodin BR, Sibille KT, Glover TL, Riley JL, et al. Psychological profiles and pain characteristics of older adults with knee osteoarthritis. Arthritis Care Res (Hoboken) 2013;65:1786–94.
[18] Waarsing JH, Bierma-Zeinstra SM, Weinans H. Distinct subtypes of knee osteoarthritis: data from the Osteoarthritis Initiative. Rheumatology (Oxford) 2015;54:1650–8.
[19] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001;23:89–109.
[20] Jamshidi A, Leclercq M, Labbe A, Pelletier J-P, Abram F, Droit A, et al. Identification of the most important features of knee osteoarthritis structural progressors using machine learning methods. Therapeutic Adv Musculoskeletal Dis 2020;12. 1759720X2093346.
[21] Jamshidi A, Pelletier J-P, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. Nat Rev Rheumatol 2019;15:49–60.
[22] Fernandez-Tajes J, Soto-Hermida A, Vazquez-Mosquera ME, Cortes-Pereira E, Mosquera A, Fernandez-Moreno M, et al. Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. Ann Rheum Dis 2014;73:668–77.
[23] Attur M, Belitskaya-Levy I, Oh C, Krasnokutsky S, Greenberg J, Samuels J, et al. Increased interleukin-1beta gene expression in peripheral blood leukocytes is associated with increased pain and predicts risk for progression of symptomatic knee osteoarthritis. Arthritis Rheum 2011;63:1908–17.
[24] Kinds MB, Marijnissen AC, Viergever MA, Emans PJ, Lafeber FP, Welsing PM. Identifying phenotypes of knee osteoarthritis by separate quantitative radiographic features may improve patient selection for more targeted treatment. J Rheumatol 2013;40:891–902.
[25] Du Y, Almajalid R, Shan J, Zhang M. A Novel Method to Predict Knee Osteoarthritis Progression on MRI Using Machine Learning Methods. IEEE Trans Nanobiosci 2018;17:228–36.
[26] Halilaj E, Le Y, Hicks JL, Hastie TJ, Delp SL. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2018;26:1643–50.
[27] Knoop J, van der Leeden M, Thorstensson CA, Roorda LD, Lems WF, Knol DL, et al. Identification of phenotypes with different clinical outcomes in knee osteoarthritis: data from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2011;63:1535–42.
[28] van der Esch M, Knoop J, van der Leeden M, Roorda L, Lems W, Knol D, et al. Clinical Phenotypes in Patients with Knee Osteoarthritis: A Study in the Amsterdam Osteoarthritis Cohort. Osteoarthritis Cartilage 2015;23:A367–8.
[29] Xie J, Girshick R, Farhadi A. Unsupervised Deep Embedding for Clustering Analysis. Eds.. In: Maria Florina B, Kilian QW, editors. Proceedings of The 33rd International Conference on Machine Learning. 48; 2016. p. 478–87. Proceedings of Machine Learning Research: PMLR
[30] Abdi H, Williams LJ, Valentin D. Multiple factor analysis: principal component analysis for multitable and multiblock data sets. Wiley Interdisciplinary Rev 2013;5:149–79.
[31] Escofier B, Pagès J. Multiple factor analysis (AFMULT package). Comput Stat Data Anal 1994;18:121–40.
[32] Escofier B, Pagès J. Analyses factorielles simples et multiples. Objectifs méthodes et interprétation. Dunod 2008.
[33] Fawaz-Estrup F. The osteoarthritis initiative: an overview. Med Health R I 2004;87:169–71.
[34] Felson DT, Nevitt MC. Epidemiologic studies for osteoarthritis: new versus conventional study design approaches. Rheum Dis Clin North Am 2004;30:783–97. vii.
[35] Nevitt MC, Felson DT, Lester G. OAI Protocol Osteoarthritis Initiative: A Knee Health Study. vol. 2021, Last modified on Oct 18, 2017 ed: The Osteoarthritis Initiative.
[36] Sakellariou G, Conaghan PG, Zhang W, Bijlsma JWJ, Boyesen P, D'Agostino MA, et al. EULAR recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis. Ann Rheum Dis 2017;76:1484–94.
[37] Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. J R Stat Soc Series C (Applied Statistics) 1979;28:100–8.
[38] van der Maarten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9:2579–605.
[39] Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. J Statist Software 2016;70.
[40] Husson F, Josse J, Pagès J. Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data?. Technical Report of the Applied Mathematics Department (Agrocampus). Rennes: Applied Mathematics Department; 2010. p. 1–17.
[41] Critchlow ED, Fligner AM. On distribution-free multiple comparisons in the one-way analysis of variance. Commun Stat 1991;20:12.
[42] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B-Stat Methodol 1995;57:289–300.
[43] Neumann G, Hunter D, Nevitt M, Chibnik LB, Kwoh K, Chen H, et al. Location specific radiographic joint space width for osteoarthritis progression. Osteoarthritis Cartilage 2009;17:761–5.
[44] Cooke TD, Sled EA, Scudamore RA. Frontal plane knee alignment: a call for standardized measurement. J Rheumatol 2007;34:1796–801.
[45] R Core Team. R: A Language and Environment for Statistical Computing. vol. 2022. https://www.R-project.org2021.
[46] RStudio Team. RStudio: Integrated Development Environment for r. vol. 2022. http://www.rstudio.com/2021.
[47] Le S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. J Statist Software 2008;25:1–18.
[48] Kamper SJ, Maher CG, Mackay G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. J Manual Manipulative Therapy 2009;17:163–70.
[49] Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. What is the Minimum Clinically Important Difference for the WOMAC Index After TKA? Clin Orthop Relat Res 2018;476:2005–14.
[50] Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004;8:283–91.
[51] Granger CL, Parry SM, Denehy L. The self-reported Physical Activity Scale for the Elderly (PASE) is a valid and clinically applicable measure in lung cancer. Support Care Cancer 2015;23:3211–8.
[52] Kokkotis C, Moustakidis S, Papageorgiou E, Giakas G, Tsaopoulos DE. Machine learning in knee osteoarthritis: A review. Osteoarthr Cartil Open 2020;2:100069.
[53] Bastick AN, Belo JN, Runhaar J, Bierma-Zeinstra SMA. What Are the Prognostic Factors for Radiographic Progression of Knee Osteoarthritis? A Meta-analysis. Clin Orthopaed Rel Res 2015;473:2969–89.
[54] Osteoarthritis Duclos M. obesity and type 2 diabetes: The weight of waist circumference. Ann Phys Rehab Med 2016;59:157–60.
[55] Lee S, Kim TN, Kim SH, Kim YG, Lee CK, Moon HB, et al. Obesity, metabolic abnormality, and knee osteoarthritis: A cross-sectional study in Korean women. Mod Rheumatol 2015;25:292–7.
[56] Lee BJ, Yang S, Kwon S, Choi KH, Kim W. Association between Metabolic Syndrome and Knee Osteoarthritis: A Cross-Sectional Nationwide Survey Study. J Rehab Med 2019;51:464–70.
[57] Kluzek S, Newton JL, Arden NK. Is osteoarthritis a metabolic disorder? Br Med Bull 2015;115:111–21.
[58] Sanchez-Santos MT, Judge A, Gulati M, Spector TD, Hart DJ, Newton JL, et al. Association of Metabolic Syndrome with Knee and Hand Osteoarthritis: A Community-Based Study of Women. Osteoarthr Cartil 2018;26:S233. -S233.
[59] Bierma-Zeinstra SMA, Waarsing JH. The role of atherosclerosis in osteoarthritis. Best Pract Res Clin Rheumatol 2017;31:613–33.
[60] Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. Pain 2016;157:1851–71.
[61] Schieker M, Conaghan PG, Mindeholm L, Praestgaard J, Solomon DH, Scotti C, et al. Effects of Interleukin-1beta Inhibition on Incident Hip and Knee Replacement: Exploratory Analyses From a Randomized, Double-Blind, Placebo-Controlled Trial. Ann Intern Med 2020;173:509–15.
[62] Schiratti JB, Dubois R, Herent P, Cahane D, Dachary J, Clozel T, et al. A deep learning method for predicting knee osteoarthritis radiographic progression from MRI. Arthritis Res Ther 2021;23:262.
[63] Virayavanich W, Alizai H, Baum T, Nardo L, Nevitt MC, Lynch JA, et al. Association of Frequent Knee Bending Activity With Focal Knee Lesions Detected With 3T Magnetic Resonance Imaging: Data From the Osteoarthritis Initiative. Arthritis Care Res 2013;65:1441–8.
[64] Hovis KK, Stehling C, Souza RB, Haughom BD, Baum T, Nevitt M, et al. Physical Activity Is Associated With Magnetic Resonance Imaging-Based Knee Cartilage T2 Measurements in Asymptomatic Subjects With and Those Without Osteoarthritis Risk Factors. Arthritis Rheumatol 2011;63:2248–56.
[65] Martin KR, Kuh D, Harris TB, Guralnik JM, Coggon D, Wills AK. Body mass index, occupational activity, and leisure-time physical activity: an exploration of risk factors and modifiers for knee osteoarthritis in the 1946 British birth cohort. BMC Musculoskelet Disord 2013;14.
[66] Palmer KT. Occupational activities and osteoarthritis of the knee. Br Med Bull 2012;102:147–70.
[67] Ashbeck E, Bedrick EJ, Kwoh CK. The "placebo effect" in osteoarthritis clinical trials: challenging the narrative. Osteoarthr Cartil 2021;29:S13–4.
[68] McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014;67:267–77.
[69] Pripp AH. Hawthorne effects. Tidsskr Nor 2020;140:1475. -1475.
[70] Patel KV, Amtmann D, Jensen MP, Smith SM, Veasley C, Turk DC. Clinical outcome assessment in clinical trials of chronic pain treatments. Pain Rep 2021;6:e784.