Pain Trajectory Groups in Persons With, or At High Risk of, Knee Osteoarthritis: Findings from the Knee Clinical Assessment Study and the Osteoarthritis Initiative

Web Appendix

(1) Osteoarthritis Initiative datasets
(2) Description of the method to select the optimal latent class growth model
(3) Description of sensitivity analyses undertaken
(4) Description of the method to sample participants from the OAI cohort
(5) Web Tables 1, 2, 3
(6) Web Figures 1, 2, 3
WebAppendix 1. Osteoarthritis Initiatives datasets

The following datasets were downloaded from http://www.oai.ucsf.edu/ for the current study:

“enrollees; version 17”, “allclinical00; version 0.2.2”, “allclinical01”; version 1.2.1, “allclinical03”; version 3.2.1, “allclinical05”; version 5.2.1, “allclinical06”; version 6.2.1, “allclinical07”; version 7.2.1, “allclinical08”; version 8.2.1, “kxr_sq_bu00”; version 0.5, “outcomes99”; version 3

Exclusion criteria for OAI: inflammatory arthritis, severe joint space narrowing in both knees, a unilateral total knee arthroplasty and severe joint space narrowing in the other knee, a bilateral total knee arthroplasty or one planned in the next 3-years, any contraindications to magnetic resonance imaging scan, were pregnant, could not provide a blood sample, were highly dependent on ambulatory aids, had a comorbid condition affecting participation, did not plan to reside in the area for more than 3 years, were already taking part in a clinical trial, or were unwilling to give informed consent.
Web Appendix 2. Description of the method to select the optimal latent class growth model

The statistical indices used to identify the optimum latent class growth model included the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC) and the sample-size adjusted BIC (ABIC) (with lower values indicating better fit). Two likelihood ratio tests - the Vuong-Lo-Mendall-Rubin likelihood ratio test (VLMR-LRT) and the non-parametric bootstrap likelihood ratio test (B-LRT) - were used to test the null hypothesis that a model with \( k-1 \) classes was a better fit to the data than one with \( k \) classes\(^1\). For these tests, a non-significant \( p \)-value (\( p > 0.05 \), 2-tailed test) was desirable to indicate that the model with \( k-1 \) classes was an optimum fit to the data. Preferred models would also have higher entropy values (range: 0-1), a class prevalence >1% in the smallest class\(^2\), and class-specific average posterior probabilities > 0.7\(^3\).

Parameter estimates for all models were estimated using full information maximum likelihood estimation (which assumes data are missing at random) and the expectation-maximization algorithm. Associated confidence intervals were calculated using robust standard errors if WOMAC scores had a skewed distribution. Prior to reporting any parameter estimates, model solutions were checked to ensure that a global solution had been reached in the estimation algorithm. This was achieved by re-running the model using 5,000 different starting values to check that the same model log likelihood was achieved regardless of starting value used. A
global solution was concluded if the highest log likelihood value was repeated in at least two final stage solutions. All models were fitted using Mplus software version 6.1.

References

1 Muthen LK, Muthen BO. (1998-2010). Mplus User's Guide. 6th ed. Los Angeles: Muthen & Muthen, 2010.

2 Jung T Wickrama KAS. An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. Social and Personality Psychology Compass. 2008;2(1):302-17. (doi: 10.1111/j.1751-9004.2007.00054.x).

3 Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010;6:109-38. (doi: 10.1146/annurev.clinpsy.121208.131413; 10.1146/annurev.clinpsy.121208.131413).
WebAppendix 3. Sensitivity analyses

1) Testing whether a linear model is optimal for the CAS-K data

Quadratic and cubic terms were added to the optimal linear model in CAS-K to test whether their inclusion improved model fit and/or greatly influenced trajectory shape. It was also explored whether the decision regarding the optimal number of classes would have differed if a quadratic or cubic model had been assumed from the outset of the analysis. This analysis was only completed for participants in CAS-K as it was this dataset that was used to derive the model that was then later tested for reproducibility in the OAI data.

2) Testing whether trajectory shape is sensitive to a) the number of non-missing measurement points required for a participant to be included in the analysis and b) exclusion of the first time-point in the analysis

A sensitivity analysis was conducted to explore whether model results would have differed if the optimal model in CAS-K were applied to all participants, regardless as to how much missing WOMAC data participants had, and also to those who had WOMAC data at all time-points. This analysis was applied to both the CAS-K data and OAI data (with the matching process repeated separately for each analysis i.e. OAI participants were sampled from a pool of participants after any selection for non-missing WOMAC data had been completed). In addition, it was explored whether removal of data at the first time point influenced trajectory shape in both datasets to explore whether findings were greatly influenced by potential regression to the mean that could occur after the initial recruitment visit.
Web Appendix 4

OAI participants were selected from those aged 50 years and over and reporting knee pain in the last 12-months at baseline using propensity score matching. A propensity score was generated and was defined as the predicted values from a logistic regression model predicting cohort membership (0=OAI, 1=CAS-K) from the matching variables (with CAS-K and OAI participants required to have a WOMAC Pain score at baseline and for at least two follow-up time points and complete data on all matching variables). The matching process was applied by taking each person in CAS-K and sampling (without replacement) a person in the OAI database with the same (or as close as possible) propensity score as for the CAS-K participant (1:1 nearest neighbour matching). The matching procedure was fitted in STATA v12.0 using the user-written program psmatch2. As an additional check, the robustness of the matching processes was tested by repeating the matching process on five separate occasions to ensure that any model results derived did not depend on the particular subset of OAI participants that had been sampled (i.e. using five different random number seeds in the matching algorithm).

References

1. Leuven E SB. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Version 4.0.6. ed., 2003.
**WebTable 1:** Characteristics of Key Participant Groups in the Knee Clinical Assessment Study (CAS-K)

|                             | Attended baseline clinic | Analysed participants | Excluded participants |
|-----------------------------|--------------------------|-----------------------|----------------------|
|                             | N=819                    | N=570                 | N=249                |
| Age (years): mean (SD)      | 66 (8.7)                 | 64 (8.0)              | 68 (9.5)             |
| Female gender               | 440 (54)                 | 309 (54)              | 131 (53)             |
| Body mass index: mean (SD)  | 29.6 (5.2)               | 29.5 (5.0)            | 29.9 (5.7)           |
| Time since problem onset    |                          |                       |                      |
| < 1 year                    | 99 (12)                  | 57 (10)               | 42 (17)              |
| 1-5 years                   | 284 (35)                 | 201 (35)              | 83 (33)              |
| 5+ years                    | 436 (53)                 | 312 (55)              | 124 (50)             |
| WOMAC function: median (IQR)| 19.0 (7.0, 32.9)         | 17.0 (6.0, 30.0)      | 26.0 (12.0, 37.0)    |
| Kellgren-Lawrence score in tibiofemoral joint | | | |
| 0                           | 408 (53)                 | 308 (54)              | 100 (48)             |
| 1                           | 68 (9)                   | 54 (9)                | 14 (7)               |
| 2                           | 109 (14)                 | 82 (14)               | 27 (13)              |
| 3                           | 89 (11)                  | 70 (12)               | 19 (9)               |
| 4                           | 103 (13)                 | 56 (10)               | 47 (23)              |
| WOMAC Pain: median (IQR)    |                          |                       |                      |
| Baseline                    | 6.0 (3.0, 10.0)          | 5.0 (2.0, 9.0)        | 8.0 (5.0, 11.0)      |
| 1.5-years                   | 5.0 (2.0, 9.0)           | 5.0 (2.0, 9.0)        | 6.0 (2.0, 10.0)      |
| 3-years                     | 6.0 (2.0, 10.0)          | 6.0 (2.0, 10.0)       | 5.0 (1.0, 9.0)       |
| 4.5-years                   | 5.0 (2.0, 9.0)           | 5.0 (2.0, 9.0)        | 5.0 (1.0, 9.0)       |
| 6-years                     | 6.3 (3.0, 10.0)          | 6.0 (3.0, 10.0)       | 7.0 (3.0, 10.0)      |

Abbreviations: IQR, Inter-quartile range; SD, standard deviation; WOMAC, Western Ontario & McMaster Universities Osteoarthritis Index. Numbers are N (%) unless otherwise stated. Analysis sample defined as participants with WOMAC follow-up data for at least 2 follow-up time points, with no diagnosis of inflammatory disease and with complete data on all matching variables i.e. age, gender, body mass index, WOMAC Pain and Function scores and Kellgren-Lawrence score in the tibiofemoral joint of the patients’ most problematic (index) knee.
**Web Table 2: Goodness-of-fit Statistics for Quadratic and Cubic WOMAC Pain Models**

| Classes | AIC   | BIC   | ABIC  | Entropy | VLMR LRT | Adjusted LMR LRT | PB LRT | Class N      | Average posterior probability |
|---------|-------|-------|-------|---------|----------|------------------|--------|--------------|-----------------------------|
| **Quadratic** |       |       |       |         |          |                  |        |              |                             |
| 1       | 14683 | 14718 | 14692 | 0.84    | P<0.001  | P<0.001         | P<0.001| 570          | 1.0                         |
| 2       | 13686 | 13738 | 13700 | 0.84    | P<0.001  | P<0.001         | P<0.001| 343, 227     | 0.96, 0.93                  |
| 3       | 13359 | 13429 | 13378 | 0.83    | P<0.001  | P<0.001         | P<0.001| 119, 233, 218| 0.94, 0.90, 0.94            |
| 4       | 13261 | 13348 | 13285 | 0.86    | P<0.001  | P<0.001         | P<0.001| 14, 224, 213, 119| 0.93, 0.90, 0.94, 0.93     |
| 5       | 13237 | 13341 | 13265 | 0.80    | p = 0.675| p = 0.685      | P<0.001| 343, 227     | 0.94, 0.95, 0.80, 0.71, 0.92|
| 6       | 13215 | 13337 | 13248 | 0.78    | p = 0.119| p = 0.124      | P<0.001| 143, 27, 199, 113, 14, 74| 0.77, 0.73, 0.92, 0.93, 0.93, 0.71|
| **Cubic** |       |       |       |         |          |                  |        |              |                             |
| 1       | 14683 | 14722 | 14694 | 0.84    | P<0.001  | P<0.001         | P<0.001| 570          | 1.0                         |
| 2       | 13688 | 13749 | 13704 | 0.84    | P<0.001  | P<0.001         | P<0.001| 343, 227     | 0.96, 0.93                  |
| 3       | 13364 | 13446 | 13386 | 0.83    | P<0.001  | P<0.001         | P<0.001| 235, 119, 216| 0.90, 0.94, 0.95            |
| 4       | 13267 | 13371 | 13295 | 0.86    | P<0.001  | P<0.001         | P<0.001| 225, 14, 212, 119| 0.90, 0.92, 0.94, 0.93     |
| 5       | 13244 | 13370 | 13278 | 0.80    | p = 0.887| p = 0.890      | P<0.001| 121, 167, 204, 14, 64| 0.93, 0.80, 0.95, 0.93, 0.70|
| 6       | 13220 | 13368 | 13260 | 0.77    | p = 0.064| p = 0.066      | P<0.001| 141, 82, 190, 35, 14, 108| 0.77, 0.72, 0.92, 0.72, 0.93, 0.93|

Abbreviations: AIC, Akaike Information Criteria; ABIC, Sample-size adjusted BIC; BIC, Bayesian Information Criteria; LMR LRT, Lo-Mendell-Rubin likelihood ratio test; PBLRT, parametric bootstrapped likelihood ratio test; VLMR LRT, Vuong-Lo-Mendell-Rubin likelihood ratio test
**WebTable 3**: Comparison of Trajectory Group Membership for the 5-class Linear and 5-class Quadratic Model in CAS-K

| WOMAC 5-class Linear model | Mild, non-progressive | Progressive | Moderate | Improving | Severe, non-improving | Total |
|-----------------------------|-----------------------|-------------|----------|-----------|-----------------------|-------|
| Mild, non-progressive      | 200                   | 1           | 0        | 0         | 0                     | 201   |
| Progressive                 | 3                     | 154         | 0        | 5         | 0                     | 162   |
| Moderate                    | 0                     | 3           | 121      | 0         | 0                     | 124   |
| Improving                   | 0                     | 10          | 0        | 58        | 0                     | 68    |
| Severe, non-improving      | 0                     | 0           | 1        | 0         | 14                    | 15    |
| **Total**                   | **203**               | **168**     | **122**  | **63**    | **14**                | **570**|

Footnote: A plot of 10 participants allocated to the “Improv ers” group in the linear model and the “Progressive” group in the quadratic model showed a trend of improvement over time, therefore they were more appropriately classified by the linear, rather than quadratic, model. Bold indicates participants that would be allocated to the same trajectory group in the linear and quadratic model (96% of participants). Class allocation was identical for the quadratic and cubic models except for three participants that moved from the “Progressive” group in the quadratic model to either the “Mild, non-progressive”, or “Improving” groups in the cubic model, or who moved from the “Moderate” group in the quadratic model to the “Progressive” group in the cubic model; the cubic term was not significant in the model ($P > 0.05$).
**WebFigure 1**: WOMAC Pain Scores by Trajectory Group Membership for (A) CAS-K and (B) a Respective Matched OAI Sample - Participant Pool Restricted to those with WOMAC Pain Data at all Follow-up Time Points (N=348)

### A)

| Trajectory Group            | Sample Size | PYRC (95% CI) |
|-----------------------------|-------------|---------------|
| Mild, non-progressive       | (N=116)     | 0.05 (-0.03, 0.12) |
| Progressive                 | (N=107)     | 0.56 (0.27, 0.85) |
| Moderate                    | (N=74)      | 0.12 (-0.03, 0.28) |
| Improving                   | (N=44)      | -0.55 (-1.07, -0.03) |
| Severe, non-improving       | (N=7)       | -0.16 (-0.58, 0.26) |

### B)

| Trajectory Group            | Sample Size | PYRC (95% CI) |
|-----------------------------|-------------|---------------|
| Mild, non-progressive       | (N=150)     | -0.10 (-0.17, -0.03) |
| Moderate (A)                | (N=97)      | -0.14 (-0.30, 0.01) |
| Moderate (B)                | (N=63)      | -0.02 (-0.28, 0.23) |
| Moderate (C)                | (N=30)      | -0.28 (-0.48, -0.08) |
| Severe, non-improving       | (N=8)       | -0.15 (-0.40, 0.10) |

**Abbreviations**: PYRC = Per-year rate of change in WOMAC points; 95% confidence interval in brackets. Solid lines are fitted lines from the latent class growth model. Dashed lines are raw data points that are plotted to show the amount of individual variation in each trajectory group.
WebFigure 2: WOMAC Pain Scores by Trajectory Group Membership for (A) CAS-K and (B) a Respective Matched OAI Sample – Participants with any Amount of WOMAC Pain Follow-up Data were Included in the Analysis (N=691)

A) Mild, non-progressive (N=235) PYRC = 0.00 (-0.07, 0.08) Progressive (N=186) PYRC = 0.49 (0.27, 0.71) Moderate (N=171) PYRC = 0.10 (-0.03, 0.24) Improving (N=78) PYRC = -0.67 (-1.05, -0.28) Severe, non-improving (N=21) PYRC = -0.02 (-0.38, 0.34)

B) Mild, non-progressive (N=307) PYRC = -0.09 (-0.15, -0.03) Moderate (A) (N=197) PYRC = -0.11 (-0.77, 0.54) Moderate (B) (N=112) PYRC = -0.03 (-0.63, 0.57) Improving (N=18) PYRC = -1.59 (-2.89, -0.29) Severe, non-improving (N=57) PYRC = 0.06 (-0.27, 0.40)

Abbreviations: PYRC = Per-year rate of change in WOMAC points; 95% confidence interval in brackets. Solid lines are fitted lines from the latent class growth model. Dashed lines are raw data points that are plotted to show the amount of individual variation in each trajectory group.
**WebFigure 3:** WOMAC Pain Scores by Trajectory Group Membership for (A) CAS-K and (B) a Respective Matched OAI Sample – Baseline Time Point Excluded From the Analysis (N=570)

| Trajectory Group             | N    | PYRC (95% CI)          |
|------------------------------|------|------------------------|
| **Mild, non-progressive**    | 222  | 0.08 (-0.05, 0.20)     |
| **Moderate (A)**             | 179  | -0.15 (-0.56, 0.27)    |
| **Progressive**              | 33   | 1.37 (0.34, 2.41)      |
| **Severe, non-improving**    | 14   | 0.07 (-0.31, 0.45)     |

| Trajectory Group             | N    | PYRC (95% CI)          |
|------------------------------|------|------------------------|
| **Mild, non-progressive**    | 248  | -0.04 (-0.10, 0.03)    |
| **Moderate (A)**             | 136  | 0.03 (-0.22, 0.28)     |
| **Moderate (B)**             | 107  | 0.09 (-0.22, 0.40)     |
| **Moderate (C)**             | 65   | 0.19 (-0.17, 0.54)     |
| **Severe, non-improving**    | 14   | 0.03 (-0.57, 0.63)     |

Abbreviations: PYRC = Per-year rate of change in WOMAC points; 95% confidence interval in brackets. Solid lines are fitted lines from the latent class growth model. Dashed lines are raw data points that are plotted to show the amount of individual variation in each trajectory group.
