The following content was supplied by the authors as supporting material and has not been copy-edited or verified by JBJS.

Appendix 1

Materials and Methods

Difference between conventional growth models, LCGA and GMM

Where conventional growth models (e.g. random-effects models) assume that all patients are drawn from a single population and that the use of one intercept (initial status) and one slope (change over time) sufficiently describes overall growth in that population, LCGA and GMM assume that there are two or more unobserved subgroups with each their own characteristics of initial status and change. These unobserved subgroups are accordingly defined by different growth parameters (i.e. intercept and slope). The difference between LCGA and GMM lies in the within-group variability: LCGA assumes that there is no variability in growth factors within subgroups (i.e. all individuals within a certain subgroup are assumed to have the same initial level and amount/pattern of change), where GMM does allow within-group variability in growth factors. For a clear, more detailed explanation on both approaches, we recommend the papers by Jung and Wickrama[1] and Berlin et al.[2].

Model specification

Experts advise to use theory, previous empirical findings and initial examinations of the data to guide model specification and selection[2-4]. To assess the overall degree of heterogeneity between patients we started with a conventional growth model where the intercept and slope variance was estimated as well as the covariance in our sample as a whole (see Jung and Wickrama[1]).

As it is unknown how many recovery trajectories after THA may exist, we fitted 1-class to 6-class LCGA and GMMs and compared the results to our conventional growth model. In both the LCGA and

GMMs we estimated the pattern of change and means of the growth factors per class, and free residual variances in the overall model only. In the LCGA models, variance and covariance are naturally restricted to zero. In the GMMs, we estimated variance and covariance for the overall model only, not per class.

All models were run with 500 random starting values and 20 final iterations, and subsequently rerun with 2000 random starting values and 400 final iterations to ensure the optimal solution was found.

Model selection

As advised (see Ram and Grimm[3]), we based our model selection on a combination of 1) visual inspection of the plots and parsimony, interpretability and clinical meaningfulness of the model (e.g. a model with a few classes with distinct change patterns may be more meaningful than a model with a higher number of classes that exhibit slight variations on the same change pattern), 2) the relative fit statistics Bayesian Information Criteria (BIC), Akaike Information Criteria (AIC) and Adjusted BIC, where lower values indicate a better fit, and 3) entropy, where a higher entropy indicates a higher confidence in the correct classification of individuals. More specifically, we first considered the BIC, AIC and Adjusted BIC and used plots of the values to aid in the interpretation. We did not use a predefined cut-off value of the relative fit statistics to determine which model would be best. Instead, we subsequently scrutinized the plots of the models and debated the interpretability and clinical meaningfulness of the models. On the basis of these considerations, we chose one final model that had the lowest relative fit statistics of the models that still had adequate interpretability and clinical meaningfulness, as well as an adequate entropy. We used this final model to further explore patient- and surgical characteristics associated with the different trajectories of recovery.

References

1. Jung T, Wickrama K. An introduction to latent class growth analysis and growth mixture modeling. Soc Personal Psychol Compass. 2008;2(1):302-17.

2. Berlin KS, Parra GR, Williams NA. An introduction to latent variable mixture modeling (part 2): longitudinal latent class growth analysis and growth mixture models. J Pediatr Psychol. 2014;39(2):188-203.

3. Ram N, Grimm KJ. Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. Int J Behav Dev. 2009;33(6):565-76.

4. Van De Schoot R, Sijbrandij M, Winter SD, Depaoli S, Vermunt JK. The GRoLTS-checklist: guidelines for reporting on latent trajectory studies. Structural Equation Modeling: A Multidisciplinary Journal. 2017;24(3):451-67.

http://dx.doi.org/10.2106/JBJS.19.00234

Appendix 2

Results

Selection of the final model

The conventional one-class growth model showed a large amount of variability in preoperative OHS and longitudinal change. When adding classes, the BIC, adjusted BIC and AIC all continued to improve up to the six-class model in both the LCGA and GMMs, although Figure 1 in this appendix shows that this decrease starts to flatten somewhat after the three-class models. The entropy (Table II of the main article) decreased slightly for every class added to the models, but remained sufficiently high (>0.80 for all models)[1].

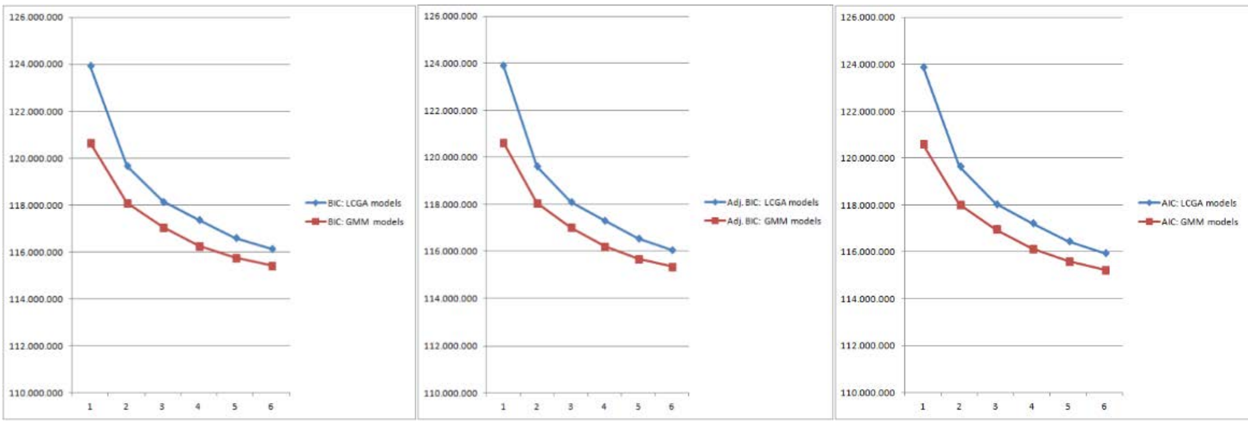
The largest class was always fairly homogeneous. The smaller classes were more heterogeneous in the LCGA than in the GMMs. Seeing this heterogeneity, combined with worse fit statistics, we carried on with the GMMs.

Up to the 3-class GMM, each new class added a distinctly different type of trajectory. From the 4-class model upwards, the new classes were mostly slight variations on the three distinct trajectories. Furthermore, the smallest classes became even smaller (up to 1.7%), thereby limiting clinical meaningfulness. Hence our decision to choose the 3-class GMM as our final model.

We subsequently evaluated the classification accuracy of our final model by investigating whether the estimated probability of subgroup membership corresponded closely to the proportion classified in that subgroup based on the highest posterior probability, and by evaluating the confidence intervals around the estimated probabilities. Furthermore, we also evaluated the average posterior probability (AvePP) of subgroup membership for individuals to each subgroup and the odds of correct classification (OCC). Nagin[2] recommends that the AvePP exceeds 0.7 and the OCC exceeds 5. Table 2 in this appendix shows the results of these evaluations which indicated good classification accuracy of the 3-class model.

References

1. Ram N, Grimm KJ. Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups. Int J Behav Dev. 2009;33(6):565-76.
2. Nagin DS, NAGIN D. Group-based modeling of development: Harvard University Press; 2005.



Appendix 2, Figure 1. Plots of BIC, Adjusted BIC and AIC of the LCGA and GMMs.

Appendix 2, TABLE 1 Comparison of Preoperative Patient Characteristics between Patients with No, Some and All OHS missing

Variable	**No OHS missing (N=6030)**	**1 or 2 OHS missing (N=19328)**	**All OHS missing (N=48926)**
Age *mean (SD)*	68.6 (8.99)	69.6 (9.55)	69.6 (9.89)
Sex			
Female	63.9 %	65.8 %	66.4 %
Male	36.1 %	34.2 %	33.6 %
BMI			
Underweight	0.5 %	0.5 %	0.7 %
Normal weight	32.9 %	31.3 %	30.9 %
Overweight	43.1 %	43.4 %	43.4 %
Obesity	23.6 %	24.7 %	25.1 %
ASA score			
ASA I	22.8 %	18.6 %	18.5%
ASA II	62.9 %	66.1 %	67 %
ASA III-IV	14.3 %	15.2 %	14.5 %
Charnley class			
A	46.7 %	45.7 %	45.4 %
B1	29.5 %	30.2 %	30.4 %
B2	21.1 %	21.9 %	22.2 %
C	2.7 %	2.1 %	2 %
Smoking			
No	90.3 %	89.2 %	88.1 %
Yes	9.7 %	10.8 %	11.9 %

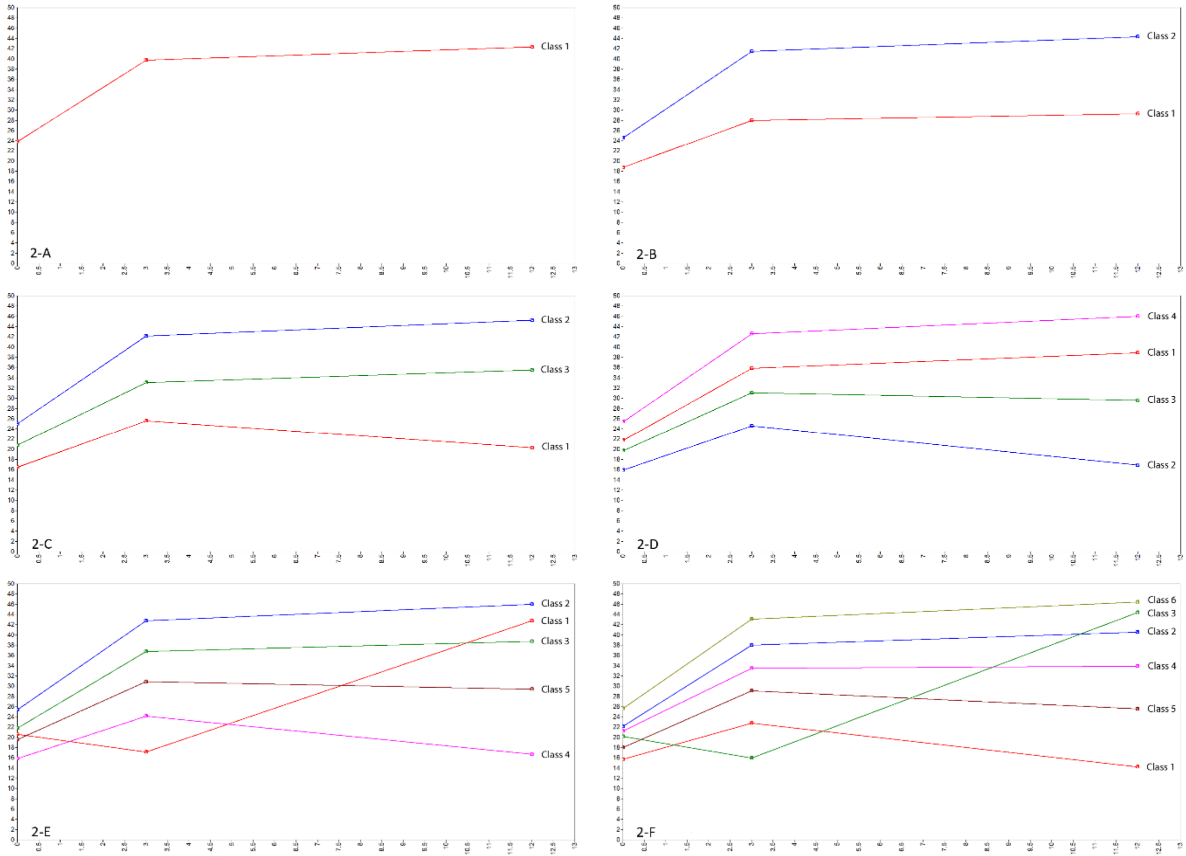
OHS = Oxford Hip Score

Appendix 2, TABLE 2 Classification Diagnostics for the Final 3-Class Model

Class	**Estimated probability of subgroup membership**	**95% CI***	**Proportion classified in subgroup based on highest posterior probability**	**AvePP**	**OCC**
Slow Starters	0.052	0.037 – 0.069	0.046	0.863	113.96
Late Dippers	0.078	0.068 – 0.088	0.077	0.913	123.48
Fast Starters	0.869	0.852 – 0.886	0.877	0.979	7.01

*Bias-corrected bootstrap 95% confidence interval, AvePP = Average Posterior Probability, OCC = Odds of Correct Classification

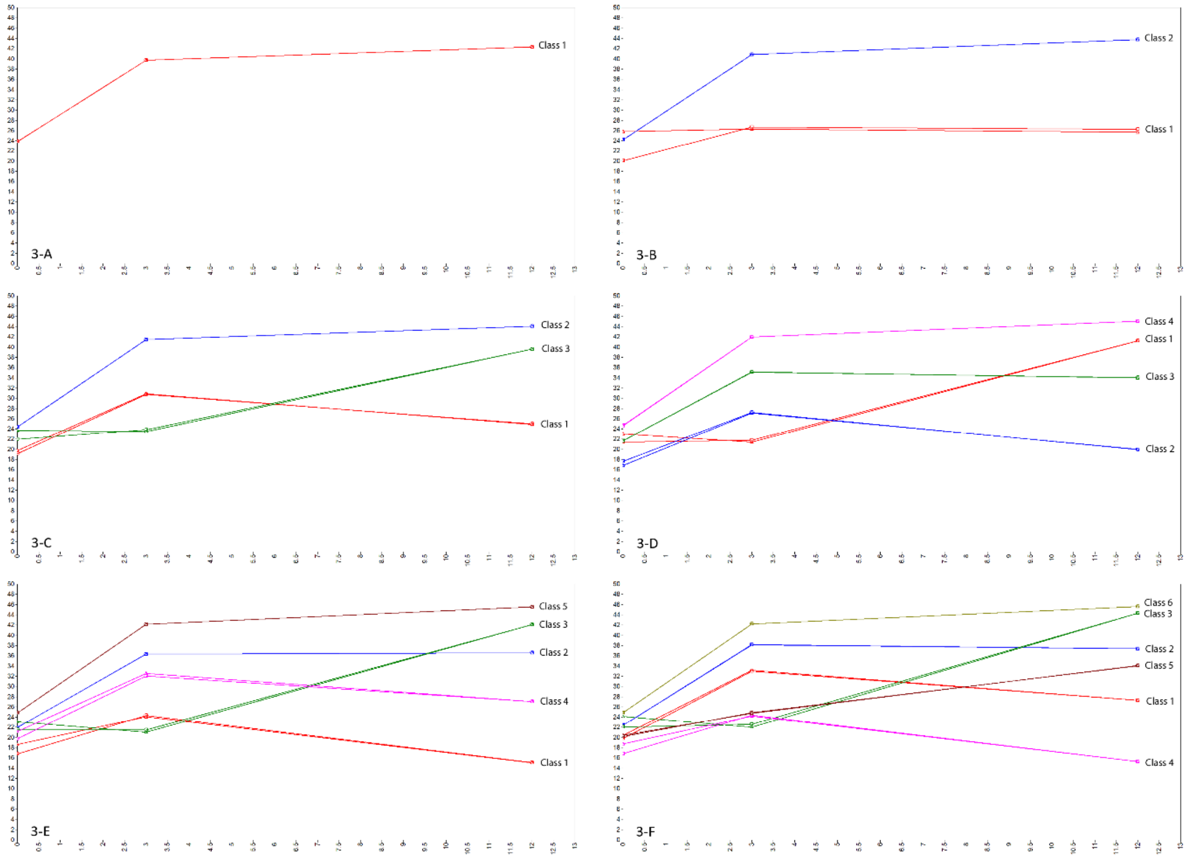
Appendix 3



Appendix 3, Figures 2-A through 2-F. Estimated means and sample means of the LCGA models.

Fig. 2-A = 1-class model. **Fig. 2-B** = 2-class model. **Fig. 2-C** = 3-class model. **Fig. 2-D** = 4-class model. **Fig. 2-E** = 5-class model. **Fig. 2-F** = 6-class model.

http://dx.doi.org/10.2106/JBJS.19.00234



Appendix 3, Figs. 3-A through 3-F. Estimated means and sample means of the GMMs. **Fig. 3-A** = 1-class model. **Fig. 3-B** = 2-class model. **Fig. 3-C** = 3-class model. **Fig. 3-D** = 4-class model. **Fig. 3-E** = 5-class model. **Fig. 3-F** = 6-class model.