Fast Starters, Slow Starters, and Late Dippers

Trajectories of Patient-Reported Outcomes After Total Hip Arthroplasty: Results from a Dutch Nationwide Database

Hesseling, B.; Mathijssen, N. M.C.; van Steenbergen, L. N.; Melles, M.; Vehmeijer, S. B.W.; Porsius, J. T.

DOI
10.2106/JBJS.19.00234

Publication date
2019

Document Version
Accepted author manuscript

Published in
The Journal of bone and joint surgery. American volume

Article title

Fast Starters, Slow Starters and Late Dippers: Trajectories of Patient Reported Outcomes after Total Hip Arthroplasty - results from a nationwide database

Authors' names

B. Hesseling[1], MSc; N.M.C. Mathijssen[1], PhD; L.N. van Steenbergen[2], PhD; M. Melles[3], PhD; S.B.W. Vehmeijer[1], M.D., PhD; J.T. Porsius[3,4,5], PhD

Institution at which the work was performed

Reinier de Graaf Hospital, Delft, The Netherlands

Institution with which each author is affiliated

[1]: Department of Orthopedic Surgery, Reinier de Graaf Hospital (Delft, The Netherlands)

[2]: Dutch Arthroplasty Register , Landelijke Registratie Orthopedische Implantaten (LROI) ('s-Hertogenbosch, The Netherlands)

[3]: Faculty of Industrial Design Engineering, Delft University of Technology (Delft, The Netherlands)

[4]: Department of Plastic, Reconstructive and Hand Surgery, Erasmus MC, Rotterdam, the Netherlands

[5]: Department of Rehabilitation Medicine, Erasmus MC, Rotterdam, the Netherlands

Corresponding author

B. Hesseling, MSc, Department of Orthopedic Surgery, Reinier de Graaf Groep, Reinier de Graafweg 5, 2625 AD Delft, The Netherlands

Email: b.hesseling@rdgg.nl

Fast Starters, Slow Starters and Late Dippers: Trajectories of Patient Reported Outcomes after Total Hip Arthroplasty - results from a nationwide database

Abstract

Background

To explore whether subgroups of patients with different functional recovery trajectories after THA can be discerned, as well as their predictors, using data from the Dutch Arthroplasty Register (LROI).

Methods

We retrospectively reviewed prospectively collected Oxford Hip Scores (OHS) up to one year postoperatively of 6030 primary THA patients. Latent growth curve modeling (LGCM) was used to classify groups of patients according to trajectory of functional recovery represented by their OHS scores. We used multivariable multinomial logistic regression analysis to explore factors associated with class membership.

Results

LGCM identified Fast Starters (fast initial improvement, high 12-month scores, 87.7%), Slow Starters (no initial change and subsequent improvement, 4.6%) and Late Dippers (initial improvement and subsequent deterioration, 7.7%).

Factors associated with Slow Starters (OR, 95% CI) were female sex (1.63, 1.14-2.33), smoking (1.95, 1.26-3.03) and anterior approach (0.47, 0.29-0.78).

Factors associated with Late Dippers (OR, 95% CI) were age > 75 years (1.62, 1.22-2.15), smoking (1.68, 1.17-2.42), ASA ≥ II (1.41, 1.05-1.91), obesity (1.96, 1.43-2.69), EQ-5D Self-Care (1.41, 1.10-1.82 ('some problems') and 2.90, 1.39-6.03 ('unable')), EQ-5D Anxiety/Depression (1.31, 1.00-1.71 ('moderately') and 1.86, 1.06-3.24 ('extremely')), EQ-5D VAS (0.91, 0.86-0.97 per 10 points), direct lateral approach (2.18, 1.58-3.02) and hybrid fixation (1.79, 1.00-3.21).

Conclusions

We discerned Fast Starters, Slow Starters and Late Dippers after THA. Female sex, older age, obesity, higher ASA scores and worse EQ-5D scores were associated with **a less favorable** response to THA **(although all groups experienced functional improvement following THA)**, as well as anterior and direct lateral approach and hybrid fixation.

Level of Evidence

Level of Evidence: II

Introduction

While the majority of patients with end-stage hip OA (osteoarthritis) respond well to THA (total hip arthroplasty), a reported 7-23% of patients do not respond as favorably, indicating that some degree of heterogeneity in recovery after THA exists[1-4]. To further improve outcomes of THA it is important to better understand differences between patients in how they respond to THA. Several previous studies have examined recovery after THA in terms of reported pain and functioning and found on average a clinically meaningful, non-linear improvement where most of the improvement occurred in the first 3 months[1-5]. These studies did not investigate the degree of variation between patients in recovery. However, several studies do report associations between pre- and perioperative factors, such as BMI or surgical approach, and postoperative outcomes[6-8], suggesting that variation in recovery trajectories may exist.

A suitable method to investigate heterogeneity in change patterns is latent class growth modelling[9-11] (LCGM). This is an extension to latent growth curve modelling, or its often used mathematical equivalence, the mixed or multilevel model[12]. A mixed model applied to longitudinal data allows for estimating the degree of heterogeneity between patients in recovery trajectories by estimating the random slope variance (see Laird and Ware[13] for an explanation of random slope models). Porsius et al.[14] used LCGM to analyze subgroups of patients according to their hip function trajectory during the first six weeks after THA. To our knowledge only one other study used such a model to examine change in patient reported outcomes after THA, but they do not report on their random effects and used a small sample of only 80 patients[15]. The advantage of LCGM is that heterogeneity can be addressed by modeling different recovery patterns for different subgroups of patients[9, 16, 17]. Previous successful applications have for instance illustrated the wide variety in patients' responses to total knee arthroplasty[18] or cardiac rehabilitation[19].

To properly study heterogeneity in recovery after THA it is important to study a large nationally representative sample of patients. In the present study we apply LCGM to outcomes as gathered by the LROI (Dutch Arthroplasty Register, in Dutch: Landelijke Registratie Orthopedische Implantaten)[20]. We aimed to characterize subgroups of patients according to their hip function and pain trajectory, as measured with the OHS (Oxford Hip Score), and determine associations with pre- and perioperative characteristics.

Materials and Methods

Data source

Data for this study was extracted from the LROI. This national web-based longitudinal database contains data on primary and revision arthroplasty procedures since its start in 2007 and on PROMs (Patient Reported Outcome Measures) since 2014. Large-scale registration of hip PROMs started in 2015. In 2016, data on surgeries (e.g. patient characteristics and surgical variables) was provided by up to 100 hospitals and clinics, with a completeness of registration of 99% of the total number of performed arthroplasties. Data on PROMs was provided by up to 80 centers[20].

Data collection

We retrospectively obtained prospectively collected data from the LROI, from patients who received a primary THA between January 1st, 2014 and December 31st, 2016 and who had a primary diagnosis of osteoarthritis. For the purpose of the present study we selected all patients who completed the OHS preoperatively (maximum of 182 days before surgery), at 3 months (63 – 110 days after surgery) and at 12 months (323 – 407 days after surgery). Obtained data comprised patient characteristics (age, sex, smoking, ASA (American Society of Anesthesiologists) score and Charnley status, BMI (body mass index) and previous surgeries on the affected hip), surgery details (approach, fixation, articulation and femoral head diameter), revision status and PROMs (Numeric Rating Scale for pain, OHS, HOOS-PS[21] and EQ-5D[22] (EuroQoL-5D-3L).

Outcome

The outcome of interest consisted of reported severity of problems with the operated hip as measured with the commonly used OHS[23]. The OHS is calculated by summing the answers of 12 questions related to pain and functional problems regarding the operated hip. Higher scores indicate better functioning and less pain (range = 0 – 48)[24]. Anchor-based methods have revealed that a change score of ~ 11 on the OHS indicates a meaningful improvement at the group-level[25].

Predictors

Predictors of interest that were extracted from the database included preoperative patient characteristics (age (dichotomized to ≤75 years and >75 years), sex, smoking, ASA, Charnley score, BMI, previous surgery on the affected hip and all preoperative EQ-5D items and EQ-5D VAS (Visual Analog Scale) scores as predictors, except for the item mobility since the OHS itself already contains items regarding walking) and surgery specific factors (approach, fixation and femoral head diameter).

Statistical analyses

We used IBM SPSS Statistics version 21.0 (Armonk, NY: IBM Corp.) for data cleaning and providing descriptives of our overall sample. To investigate whether different subgroups could be distinguished in our sample based on the trajectories of OHS scores, we used Mplus Version 8.1 (Los Angeles, CA: Muthén & Muthén[26]) to perform 1-class to 6-class LCGM analyses in the form of LCGA (latent class growth analysis) and GMM (growth mixture modeling) in addition to a conventional growth model. See Appendix 1 for a description of the differences between a conventional growth model and LCGA and GMM models, and for a detailed description of our model specification and selection.

As previous research has demonstrated a non-linear growth pattern after THA[1-5], we specified for all models a latent basis model for the growth pattern [9, 17]; the first (preoperative) and last (12-month postoperatively) measurement were fixed to respectively 0 and 1, and the second measurement (3 months postoperatively) was estimated freely. As such, the estimated average slopes in our models represent the amount of change between the first and last measurement and the estimated factor loading of the second measurement represents how much of that change occurred at 3 months. All models were unconditional models, meaning that the latent class probabilities were independent from other variables.

Subsequently, we used the r3step procedure in Mplus to perform both univariable and multivariable multinomial logistic regression analyses where we compared the smaller subgroups of patients to the largest group of patients.

Source of funding

This study was funded by the Van Rens Fonds (Foundation) (VRF2017-005), The Netherlands.

Results

Patient characteristics

A total of 6030 patients (8.12%) of 74284 THA patients in the study period had OHS scores on all three measurements and were therefore included in the analysis. 48.926 patients (65.86%) had no OHS scores on any measurement, 7336 patients (9.88%) only had preoperative scores and 11.992 patients (16.14%) mere missing one of the three OHS scores.

Patients who had no missing OHS scores (and were thus included in our analysis) were slightly younger, slightly more often male, slightly more often non-smokers and had somewhat better weight, Charnley scores and ASA scores compared to patients who were missing one or more OHS scores (see also Appendix 2, Table 1). These differences are similar to those found by the LROI, who compared patients who completed preoperative and 3 month PROMs in 2016 to the entire THA population of 2016[27].

Table 1 displays patient characteristics of the entire sample, as well as characteristics of each class in our final model.

Selection of the final model

The model fit statistics are summarized in Table 2, as well as the model parameters (i.e. factor loading, intercept, slope and class size).

We chose the 3-class GMM model as our final model (Figure 1); we based this on the combination of the distinct trajectories, entropy, class sizes and the fact that although fit statistics continued to decrease up to the 6-class model, this decrease started to flatten out from the 3-class model. See Appendix 3 for figures of the LCGM (Figs. 2-A through 2-F) and GMM models from

1 to 6 classes (Figs. 3-A through 3-F); the conventional growth model is identical to the 1-class GMM model.

For a detailed explanation of the selection process of the final model we refer to the Appendix 2, as well as for detailed model results that show our final model demonstrated good classification accuracy.

Trajectory patterns

Figure 1 shows the estimated trajectory in combination with the observed individual trajectories of each class.

The largest class consisted of 5290 patients and is portrayed by a steep improvement in OHS scores during the first three months, after which OHS scores level out. We labeled this class 'Fast Starters'. At three months, the Fast Starters reached 86.8% of the total amount of change at one year after surgery.

The class labeled as 'Late Dippers' (463 patients) demonstrate an initial, more modest improvement in OHS scores and subsequently a decline towards the one year mark, although there is still improvement at one year compared to preoperatively. At three months, the Late Dippers reached up to 216.4% of their overall change between preoperative and one year OHS scores.

The smallest class, consisting of 277 patients and labeled as 'Slow Starters', is characterized by virtually no change at the three month mark (-1.8% of their overall change), followed by an improvement in OHS scores at one year after surgery.

We refer to Table 2 for the exact values of initial status and overall change for each class.

Differences between classes in patient characteristics

For all analyses, the 'Fast Starters' class was chosen as the reference category. The results of the univariable and multivariable multinomial logistic regression analyses are shown in Table 3 and Table 4, respectively.

In the univariable analysis, the following variables were statistically significant for membership to class 'Slow Starters': female sex, smoking, ASA scores ≥ III, obesity (BMI ≥ 30), Charnley score B2, EQ-5D items 'Self-Care', 'Usual Activities', 'Pain / Discomfort' and 'Anxiety / Depression', EQ-5D VAS per 10 points, and anterior approach.

The variables that were statistically significant for membership to class 'Late Dippers' were: age > 75 years, female sex, smoking, ASA scores ≥ III, obesity (BMI ≥ 30), Charnley score C, EQ-5D items 'Self-Care', 'Usual Activities', 'Pain / Discomfort' and 'Anxiety / Depression', EQ-VAS per 10 points, anterior and direct lateral approach, uncemented fixation and femoral head diameter of 22 – 28 mm.

In the multivariable analysis, the following variables remained statistically significant for membership to class 'Slow Starters': female sex, smoking, obesity (BMI ≥ 30) and anterior approach.

The variables that remained statistically significant for membership to class 'Late Dippers' were: age > 75 years, smoking, ASA scores ≥ III, obesity (BMI ≥ 30), EQ-5D items 'Self-Care' and 'Anxiety / Depression', EQ-5D VAS per 10 points, direct lateral approach and hybrid fixation (cemented acetabulum).

Discussion

We found three subgroups with different functional recovery trajectories in our large sample of 6030 THA patients: Fast Starters, Slow Starters and Late Dippers. Based on our results using the OHS as outcome measure, Fast Starters can be seen as the most favorable trajectory for THA patients. Late Dippers can be seen as the least favorable response trajectory.

In addition, we found in a multivariable analysis that female sex, age above 75 years, obesity, ASA score III-IV, lower pre-operative perceived health, direct lateral approach and hybrid fixation (cemented acetabulum) were associated with not being classified as Fast Starter.

Our study corroborates other studies that found an association of certain variables with an unfavorable response. Systematic reviews by Buirs et al.[6], Hofstede et al.[7] and Lungu et al.[8] found that functional outcomes were associated with, amongst other variables, BMI, comorbidities and general mental health. Accordingly, in our study, patients with obesity, higher ASA scores, lower EQ-5D VAS scores and higher scores on the EQ-5D items 'Self-Care' and 'Anxiety/Depression' were more likely to be classified into the Late Dippers subgroup. Interestingly, problems with self-care, anxiety/depression and overall quality of life were not just markers for general health in our sample, but appear to have had an independent effect on the outcomes after THA; even after correcting for age, smoking, ASA and BMI, these items still increased the odds of becoming Late Dippers.

The subgroups and trajectories we found in our study differ from those in the study of Lenguerrand et al.[15]. This may be due to the different statistical approach: Lenguerrand et al.[15] predefined two groups (i.e. high or low preoperative scores) and used a random effects model to estimate one trajectory per group. In contrast, we did not predefine subgroups but used LCGM to explore if and how many different subgroups could be distinguished and, although we hypothesized that trajectories of the subgroups could differ qualitatively, we did not impose specific shapes of trajectories. This gave us the advantage of letting previously unknown subgroups emerge from the data.

We find it interesting that no subgroup marking ‘no improvement’ or ‘decline’ in outcomes emerged. Visual inspections of the plots suggest that these trajectories are very uncommon in our large sample and are therefore incorporated in the smaller, more heterogeneous, subgroups, instead of forming a separate subgroup; even in the 6-class models no such trajectory emerged.

Unfortunately, we could not define any factors that clearly distinguished between Late Dippers and Slow Starters. We find it likely that the difference between Late Dippers and Slow Starters is determined by other patient-related factors that were not measured in the national database.

For example, psychological factors such as preoperative expectations might influence how patients perceive pain and functional outcomes[28, 29]. Expectations, and other psychological factors such as pain catastrophizing[30] and illness perceptions, were not measured in the national database; therefore, we could not investigate its role in the subgroups we found.

Moreover, the subjective nature of our outcome scores may be amplified in the trajectories. For example, Late Dippers may be quite pleased at the 3 month mark with the progress so far in relation to their starting point, and thus offer an optimistic valuation of the OHS scores. Subsequently, this elation may wear off after time continues (while perhaps the progress stagnates), allowing for a more realistic (or even pessimistic) valuation of OHS scores at 12 months.

The greatest strength of this study is the uniquely large sample size combined with the analysis of recovery trajectories, therefore adding to the current literature a more detailed understanding of the degree of variation between patients in the recovery after THA.

There are also some important potential drawbacks associated with our study. One limitation is that the database did not contain more detailed patient-related information (e.g. coping style), probably one of the reasons we could not differentiate between Late Dippers and Slow Starters. Furthermore, while the three observations of OHS scores are sufficient to employ a latent basis model, it is possible that the true underlying trajectories could be described more accurately with more observations.

Another drawback is that we analyzed data from patients who had complete OHS scores for all time points; since large-scale registration of PROMs started in 2015, not all hospitals were registering PROMs for the patients in this study. In addition, the methods of collecting PROMs differ between hospitals; this may also affect completeness of the PROMs. Consequently, **our results represent the outcomes of 8% of all primary THAs performed during our study period and** no findings are available for patients who underwent revision within the first postoperative year or who had missing OHS scores. **However, a**lthough we found slight differences between patients that were included in our analysis and those that were excluded, comparable to the differences found by the LROI[27], **the extent to which these small differences affect the generalizability of our results to the entire Dutch THA population is uncertain.** Moreover, the heterogeneity in collecting PROMs was also present in the LROI investigation of differences between patients.

In conclusion, we discerned Fast Starters, Slow Starters and Late Dippers after THA. Sex, age, BMI, ASA scores and EQ-5D scores were associated with **a less favorable** response to THA, as well as approach and fixation**, although all groups experienced functional improvement following THA**. Our findings enable surgeons to more accurately estimate which patients are at risk of **a less favorable** recovery. In turn, this will improve the capability of surgeons to provide tailored expectation management to patients undergoing THA.

References

1. Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, et al. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. Osteoarthritis Cartilage. 2001;9(2):137-46.

2. Davis AM, Perruccio AV, Ibrahim S, Hogg-Johnson S, Wong R, Streiner DL, et al. The trajectory of recovery and the inter-relationships of symptoms, activity and participation in the first year following total hip and knee replacement. Osteoarthritis Cartilage. 2011;19(12):1413-21.

3. Kennedy DM, Stratford PW, Robarts S, Gollish JD. Using outcome measure results to facilitate clinical decisions the first year after total hip arthroplasty. J Orthop Sports Phys Ther. 2011;41(4):232-9.

4. Naylor JM, Harmer AR, Heard RC, Harris IA. Patterns of recovery following knee and hip replacement in an Australian cohort. Aust Health Rev. 2009;33(1):124-35.

5. Halket A, Stratford PW, Kennedy DM, Woodhouse LJ. Using hierarchical linear modeling to explore predictors of pain after total hip and knee arthroplasty as a consequence of osteoarthritis. J Arthroplasty. 2010;25(2):254-62.

6. Buirs LD, Van Beers LW, Scholtes VA, Pastoors T, Sprague S, Poolman RW. Predictors of physical functioning after total hip arthroplasty: a systematic review. BMJ Open. 2016;6(9):e010725.

7. Hofstede SN, Gademan MG, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: a systematic review. BMC Musculoskelet Disord. 2016;17:212.

8. Lungu E, Maftoon S, Vendittoli PA, Desmeules F. A systematic review of preoperative determinants of patient-reported pain and physical function up to 2 years following primary unilateral total hip arthroplasty. Orthop Traumatol Surg Res. 2016;102(3):397-403.

9. 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.

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

11. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010;6:109-38.

12. Bauer DJ. Estimating multilevel linear models as structural equation models. Journal of Educational and Behavioral Statistics. 2003;28(2):135-67.

13. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982:963-74.

14. Porsius JT, Mathijssen NM, Klapwijk-Van Heijningen LC, Van Egmond JC, Melles M, Vehmeijer SB. Early recovery trajectories after fast-track primary total hip arthroplasty: the role of patient characteristics. Acta Orthop. 2018:1-7.

15. Lenguerrand E, Wylde V, Gooberman-Hill R, Sayers A, Brunton L, Beswick AD, et al. Trajectories of Pain and Function after Primary Hip and Knee Arthroplasty: The ADAPT Cohort Study. PLoS One. 2016;11(2):e0149306.

16. Ram N, Grimm K. Using simple and complex growth models to articulate developmental change: Matching theory to method. International Journal of Behavioral Development. 2007;31(4):303-16.

17. 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.

18. Dowsey MM, Smith AJ, Choong PFM. Latent Class Growth Analysis predicts long term pain and function trajectories in total knee arthroplasty: a study of 689 patients. Osteoarthritis Cartilage. 2015;23(12):2141-9.

19. Arthur HM, Blanchard C, Gunn E, Kodis J, Walker S, Toner B. Exercise trajectories of women from entry to a 6-month cardiac rehabilitation program to one year after discharge. Biomed Res Int. 2013;2013:121030.

20. Dutch Arthroplasty Register (in Dutch: LROI). Online LROI-Rapportage 2016: Blik op uitkomsten. http://www.lroi-rapportage.nl/media/pdf/PDF%20Online%20LROI-Rapportage%202016.pdf. Accessed 2018 Nov 2.

21. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16(5):551-9.

22. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33(5):337-43.

23. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br. 1996;78(2):185-90.

24. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and knee scores. J Bone Joint Surg Br. 2007;89(8):1010-4.

25. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol. 2015;68(1):73-9.

26. Muthén, L.K. and Muthén, B.O. (1998-2017). MPlus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén).

27. Dutch Arthroplasty Register (in Dutch: LROI). LROI magazine 2017. https://issuu.com/novlroi/docs/lroi_magazine_2017. Accessed 2019 Jun 4.

28. Gonzalez Saenz de Tejada M, Escobar A, Bilbao A, Herrera-Espineira C, Garcia-Perez L, Aizpuru F, et al. A prospective study of the association of patient expectations with changes in health-related quality of life outcomes, following total joint replacement. BMC Musculoskelet Disord. 2014;15:248.

29. Jain D, Bendich I, Nguyen LL, Nguyen LL, Lewis CG, Huddleston JI, et al. Do Patient Expectations Influence Patient-Reported Outcomes and Satisfaction in Total Hip Arthroplasty? A Prospective, Multicenter Study. J Arthroplasty. 2017;32(11):3322-7.

30. Riddle DL, Wade JB, Jiranek WA, Kong X. Preoperative pain catastrophizing predicts pain outcome after knee arthroplasty. Clin Orthop Relat Res. 2010;468(3):798-806.

Figure Legends

Figs 1-A through 1-D. Plots of the 3-class GMM model.

Fig. 1-A = estimated means and sample means. **Fig. 1-B** = estimated means and observed individual values for class 1. **Fig. 1-C** = estimated means and observed individual values for class 2. **Fig. 1-D** = estimated means and observed individual values for class 3.

Appendix Figure Legends

Appendix 2:

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

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.

Appendix 3:

Appendix 3, Figs. 3-A through 3-F. Estimated means and sample means of the GMM models.

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.

Table 1

TABLE I Descriptive Statistics of **Preoperative Patient Characteristics and Surgery Characteristics of** the Entire Sample and of the Three Separate Classes

Variable	**Entire Sample (N = 6030)**	**Slow Starters (N = 277)**	**Late dippers (N = 463)**	**Fast Starters (N = 5290)**
Age *(mean (SD) [95% CI])*	68.64 (8.99) [68.42 - 68.87]	68.19 (9.25) [67.10 – 69.29]	70.44 (9.47) [69.57 - 71.30]	68.51 (8.92) [68.27 - 68.75]
Age *(no. (%))*				
≤75 yrs	4644 (77%)	207 (75%)	316 (68%)	4121 (78%)
>75 yrs	1384 (23%)	70 (25%)	147 (31%)	1167 (22%)
Sex *(no. (%))*				
Male	2175 (36%)	78 (28%)	135 (29%)	1962 (37%)
Female	3849 (64%)	199 (72%)	328 (71%)	3322 (63%)
Smoking *(no. (%))*				
No	5045 (84%)	221 (80%)	371 (80%)	4453 (84%)
Yes	544 (9%)	37 (13%)	54 (12%)	453 (9%)
ASA *(no. (%))*				
Class I or II	5163 (86%)	226 (82%)	350 (76%)	4587 (87%)
Class III or higher	859 (14%)	50 (18%)	113 (24%)	696 (13%)
BMI *(no. (%))*				
Normal weight	1998 (33%)	78 (28%)	119 (26%)	1801 (34%)
Overweight (BMI 25 – 30)	2573 (43%)	109 (40%)	172 (38%)	2292 (44%)
Obese (BMI ≥ 30)	1405 (23%)	88 (32%)	165 (36%)	1152 (22%)
Previous surgery *(no. (%))*				
No	5909 (98%)	272 (99%)	451 (97%)	5186 (98%)
Yes	103 (2%)	4 (1%)	12 (3%)	87 (2%)
Charnley score *(no. (%))*				
A	2784 (46%)	112 (41%)	198 (44%)	2474 (47%)
B1	1760 (29%)	86 (31%)	137 (30%)	1537 (30%)
B2	1255 (21%)	68 (25%)	100 (22%)	1087 (21%)
C	162 (3%)	8 (3%)	20 (4%)	134 (3%)
Pain at rest *(mean (SD) [95% CI])*	4.97 (2.52) [4.91 - 5.03]	5.39 (2.40) [5.10 - 5.67]	5.87 (2.31) [5.66 - 6.08]	4.87 (2.53) [4.80 - 4.94]
Pain during activity *(mean (SD) [95% CI])*	7.04 (2.07) [6.99 - 7.09]	7.35 (1.97) [7.11 - 7.59]	7.67 (1.79) [7.50 - 7.83]	6.97 (2.08) [6.91 - 7.03]
EQ-5D item ‘Mobility’ *(no. (%))*				
No problems	316 (5%)	9 (3%)	23 (5%)	284 (5%)
Some problems in walking about	5666 (94%)	263 (95%)	432 (94%)	4971 (94%)
Confined to bed	39 (1%)	4 (1%)	6 (1%)	29 (1%)

EQ-5D item 'Self-Care' *(no. (%))*				
No problems	3630 (60%)	149 (54%)	204 (44%)	3277 (62%)
Some problems washing or dressing	2322 (39%)	122 (44%)	239 (52%)	1961 (37%)
Unable to wash or dress	65 (1%)	5 (2%)	18 (4%)	42 (1%)
EQ-5D item 'Usual Activities' *(no. (%))*				
No problems	805 (13%)	24 (9%)	29 (6%)	752 (14%)
Some problems performing usual activities	4671 (78%)	217 (78%)	346 (75%)	4108 (78%)
Unable to perform usual activities	547 (9%)	36 (13%)	87 (19%)	424 (8%)
EQ-5D item 'Pain / Discomfort' *(no. (%))*				
No pain or discomfort	314 (5%)	10 (4%)	9 (2%)	295 (6%)
Moderate pain or discomfort	4184 (70%)	173 (63%)	259 (56%)	3752 (71%)
Extreme pain or discomfort	1520 (25%)	92 (34%)	191 (42%)	1237 (23%)
EQ-5D item 'Anxiety / Depression' *(no. (%))*				
Not anxious or depressed	4470 (74%)	190 (69%)	278 (60%)	4002 (76%)
Moderately anxious or depressed	1382 (23%)	69 (25%)	153 (33%)	1160 (22%)
Extremely anxious or depressed	160 (3%)	17 (6%)	31 (7%)	112 (2%)
EQ-5D VAS *(mean (SD) [95% CI])*	66.95 (19.65) [66.45 - 67.45]	63.85 (21.04) [61.33 – 66.38]	59.5 (21.08) [57.55 - 61.44]	67.77 (19.29) [67.24 - 68.29]
Approach *(no. (%))*				
Posterolateral	3819 (63%)	191 (69%)	294 (64%)	3334 (63%)
Anterior	1368 (23%)	38 (14%)	63 (14%)	1267 (24%)
Anterolateral	214 (4%)	8 (3%)	19 (4%)	187 (4%)
Direct lateral	626 (10%)	39 (14%)	86 (19%)	501 (10%)
Other	1 (0%)	0 (0%)	1 (0%)	0 (0%)
Fixation *(no. (%))*				
Cemented	1233 (21%)	61 (22%)	139 (30%)	1033 (20%)
Hybrid – cemented acetabulum	169 (3%)	6 (2%)	26 (6%)	137 (3%)
Hybrid – cemented femur	273 (5%)	19 (7%)	25 (5%)	229 (4%)
Uncemented	4347 (72%)	191 (69%)	273 (59%)	3883 (74%)
Articulation *(no. (%))*				
Ceramic-on-PE	3549 (59%)	194 (70%)	259 (57%)	3096 (59%)
Metal-on-PE	1839 (31%)	60 (22%)	163 (36%)	1616 (31%)
Ceramic-on-ceramic	157 (3%)	10 (4%)	8 (2%)	139 (3%)
Oxidized zirconium-on-PE	413 (7%)	12 (4%)	28 (6%)	373 (7%)
Other	6 (0%)	0 (0%)	0 (0%)	6 (0%)

Femoral head diameter *(no. (%))*				
22 – 28 mm	1362 (23%)	61 (22%)	144 (31%)	1157 (22%)
32 mm	3429 (57%)	159 (58%)	234 (51%)	3036 (58%)
≥36 mm	1223 (20%)	56 (20%)	85 (18%)	1082 (21%)

PE = polyethylene

Table 2

TABLE II: Fit and Model Parameters for the GMM and LCGA Models

Fit statistics								Model parameters				
Model	LL	BIC	AIC	Adjusted BIC	BLRT	Entropy	Number of Free Parameters		Factor Loading OHST1	Intercept (S.E.)	Slope (S.E.)	Patients per Class *(n (%))*
GMM												
1 class	-60290.731	120659.803	120599.462	120631.203	-	-	9	*Class 1*	0.856	23.84 (0.109)	18.52 (0.121)	6030 (100%)
2 class	-58998.010	118109.178	118022.019	118067.867	p = 0.000	0.945	13	*Class 1*	-3.38	25.82 (0.471)	-0.143 (0.178)	460 (7.6%)
								Class 2	0.851	24.17 (0.114)	19.59 (0.122)	5570 (92.4%)
3 class	-58463.891	117075.758	116961.781	117021.737	p = 0.000	0.928	17	*Class 1*	2.164	19.72 (0.518)	5.16 (0.484)	463 (7.7%)
								Class 2	0.868	24.37 (0.119)	19.71 (0.126)	5290 (87.7%)
								Class 3	-0.018	23.72 (0.755)	15.86 (0.986)	277 (4.6%)
4 class	-58047.954*	116278.703	116137.908	116211.970	p = 0.000	0.913	21	*Class 1*	-0.088	23.00 (0.888)	18.22 (1.135)	176 (2.9%)
								Class 2	3.287	16.84 (0.851)	3.13 (0.346)	232 (3.8%)
								Class 3	1.095	21.65 (0.334)	12.35 (0.444)	865 (14.3%)
								Class 4	0.847	24.66 (0.126)	20.42 (0.134)	4757 (78.9%)
5 class	-57775.794*	115769.200	115601.587	115689.757	p = 0.000	0.911	25	*Class 1*	-1.516	18.65 (0.816)	-3.59 (0.399)	102 (1.7%)
								Class 2	0.983	21.99 (0.295)	14.57 (0.381)	1028 (17%)
								Class 3	-0.108	23.07 (0.877)	19.00 (1.070)	152 (2.5%)
								Class 4	1.938	21.14 (0.992)	5.88 (1.161)	325 (5.4%)
								Class 5	0.838	24.84 (0.132)	20.69 (0.134)	4423 73.3%)
6 class	-57589.403*	115431.236	115236.806	115339.082	p = 0.000	0.906	29	*Class 1*	1.863	20.43 (0.630)	6.84 (0.850)	189 (3.1%0
								Class 2	1.052	22.40 (0.311)	14.96 (0.436)	314 (5.2%)
								Class 3	-0.106	24.18 (1.006)	20.17 (1.117)	945 (15.7%)
								Class 4	-1.552	18.78 (0.906)	-3.48 (0.390)	4351 (72.2%)
								Class 5	0.319	20.42 (0.847)	13.65 (1.049)	124 (2.1%)
								Class 6	0.837	24.87 (0.136)	20.74 (0.135)	107 (1.8%)
LCGA												
1 class	-61941.774	123935.774	123895.547	123916.708	-	-	6	*Class 1*	0.856	23.84 (0.109)	18.52 (0.121)	6030 (100%)
2 class	-59796.150	119679.346	119647.569	119612.301	p = 0.000	0.919	10	*Class 1*	0.875	18.75 (0.342)	10.53 (0.514)	765 (12.7%)

								Class 2	0.855	24.60 (0.120)	19.70 (0.125)	5265 (87.3%)
3 class	-59013.834	118149.532	118055.669	118105.044	p = 0.000	0.897	14	*Class 1*	2.371	16.47 (0.574)	3.84 (0.870)	238 (3.9%)
								Class 2	0.846	25.00 (0.132)	20.25 (0.131)	4666 (77.4%)
								Class 3	0.831	20.79 (0.271)	14.74 (0.392)	1126 18.7%)
4 class	-58600.489*	117357.660	117236.979	117300.461		0.885	18	*Class 1*	0.822	21.72 (0.247)	17.16 (0.368)	1383 (22.9%)
								Class 2	8.871	15.91 (0.734)	0.97 (0.957)	146 (2.4%)
								Class 3	1.153	19.69 (0.456)	9.86 (0.606)	458 (7.6%)
								Class 4	0.836	25.34 (0.146)	20.63 (0.141)	4043 (67%)
5 class	-58206.217*	116603.933	116456.434	116534.023		0.898	22	*Class 1*	-0.157	20.60 (0.989)	22.17 (1.469)	86 (1.4%)
								Class 2	0.844	25.39 (0.149)	20.59 (0.140)	4015 (66.6%)
								Class 3	0.884	21.79 (0.254)	16.95 (0.410)	1336 (22.1%)
								Class 4	9.165	15.82 (0.770)	0.91 (1.045)	140 (2.3%)
								Class 5	1.154	19.56 (0.495)	9.84 (0.707)	453 (7.5%)
6 class	-57949.465*	116125.247	115950.930	116042.626		0.893	26	*Class 1*	-4.779	15.70 (0.846)	-1.48 (1.455)	91 (1.5%)
								Class 2	0.859	22.16 (0.252)	18.40 (0.559)	1393 (23.1%)
								Class 3	-0.175	20.19 (1.151)	24.16 (1.841)	62 (1%)
								Class 4	0.966	21.21 (0.592)	12.69 (0.626)	604 (10%)
								Class 5	1.467	17.99 (0.656)	7.58 (1.068)	261 (4.3%)
								Class 6	0.840	25.64 (0.170)	20.75 (0.146)	3619 (60%)

* Although the best loglikelihood value was replicated in these classes, solutions from subsequent loglikelihood values revealed different parameter estimates and/or class sizes, or produced errors. Therefore, the results of these models may not be trustworthy[1]

LL = loglikelihood, BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion, BLRT = Bootstrapped Likelihood Ratio Test, S.E. = Standard Error

1. Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. Psychol Methods. 2006;11(1):36.

Table 3

TABLE III Univariable Multinomial Logistic Regression Analysis with Class Membership as Dependent Variable (Class 'Fast Starters' as Reference Class)

	Slow Starters vs. Fast Starters		Late Dippers vs. Fast Starters	
	OR (95% CI)	*p*	OR (95% CI)	*p*
Age >75 yrs (vs. ≤75 yrs)	1.21 (0.88 – 1.68)	0.247	1.72 (1.37 – 2.16)	<0.001
Female (vs. male)	1.61 (1.17 – 2.22)	0.004	1.49 (1.18 – 1.87)	0.001
Smoking (vs. no smoking)	1.76 (1.17 – 2.66)	0.007	1.48 (1.06 – 2.05)	0.020
ASA III-IV (vs. I-II)	1.52 (1.05 – 2.19)	0.026	2.27 (1.78 – 2.90)	<0.001
BMI				
Normal weight (ref.)	1.0	-	1.0	-
Overweight (BMI 25 – 30)	1.11 (0.78 – 1.59)	0.551	1.15 (0.88 – 1.51)	0.307
Obese (BMI ≥30)	1.90 (1.31 – 2.74)	0.001	2.33 (1.77 – 3.07)	<0.001
Had previous surgery	0.84 (0.24 – 2.89)	0.779	1.65 (0.86 – 3.18)	0.133
Charnley score				
A (ref.)	1.0	-	1.0	-
B1	1.28 (0.91 – 1.80)	0.154	1.12 (0.88 – 1.45)	0.360
B2	1.46 (1.01 – 2.09)	0.042	1.16 (0.88 – 1.53)	0.286
C	1.36 (0.57 – 3.25)	0.492	1.97 (1.16 – 3.33)	0.012
EQ-5D item 'Self-Care'				
No problems (ref.)	1.0	-	1.0	-
Some problems washing or dressing	1.42 (1.06 – 1.89)	0.018	2.10 (1.69 – 2.61)	<0.001
Unable to wash or dress	2.88 (0.96 – 8.62)	0.059	8.08 (4.37 – 14.95)	<0.001
EQ-5D item 'Usual Activities'				
No problems (ref.)	1.0	-	1.0	-
Some problems with performing usual activities	1.80 (1.05 – 3.08)	0.031	2.44 (1.53 – 3.90)	<0.001
Unable to perform usual activities	3.04 (1.60 – 5.78)	0.001	6.40 (3.82 – 10.71)	<0.001
EQ-5D item 'Pain / Discomfort'				
No pain or discomfort (ref.)	1.0	-	1.0	-
Moderate pain or discomfort	1.42 (0.64 – 3.16)	0.385	2.64 (1.10 – 6.32)	0.030
Extreme pain or discomfort	2.43 (1.08 – 5.48)	0.033	6.35 (2.64 – 15.29)	<0.001
EQ-5D item 'Anxiety / Depression'				
Not anxious or depressed (ref.)	1.0	-	1.0	-
Moderately anxious or depressed	1.28 (0.92 – 1.78)	0.149	2.02 (1.61 – 2.54)	<0.001
Extremely anxious or depressed	3.66 (2.01 – 6.65)	<0.001	4.50 (2.88 – 7.06)	<0.001
EQ-5D VAS (per 10 points)	0.90 (0.83 – 0.97)	0.003	0.81 (0.78 – 0.84)	<0.001
Approach				
Posterolateral (ref.)	1.0	-	1.0	-
Anterior	0.46 (0.30 – 0.73)	0.001	0.53 (0.38 – 0.73)	<0.001
Anterolateral	0.70 (0.29 – 1.70)	0.433	1.17 (0.69 – 1.99)	0.560
Direct lateral	1.40 (0.93 – 2.11)	0.110	2.06 (1.56 – 2.72)	<0.001
Other	*	*	*	*
Fixation				
Cemented (ref.)	1.0	-	1.0	-
Hybrid – cemented acetabulum	0.68 (0.23 – 2.02)	0.486	1.44 (0.89 – 2.34)	0.138
Hybrid – cemented femur	1.49 (0.81 – 2.73)	0.201	0.80 (0.49 – 1.30)	0.363
Uncemented	0.82 (0.58 – 1.17)	0.274	0.49 (0.39 – 0.62)	<0.001
Femoral head diameter				
32 mm (ref.)	1.0	-	1.0	-
22 - 28 mm	0.99 (0.69 – 1.42)	0.965	1.69 (1.33 – 2.15)	<0.001
≥36 mm	0.99 (0.68 – 1.42)	0.939	1.02 (0.77 – 1.36)	0.884

* No reliable odds ratio could be determined

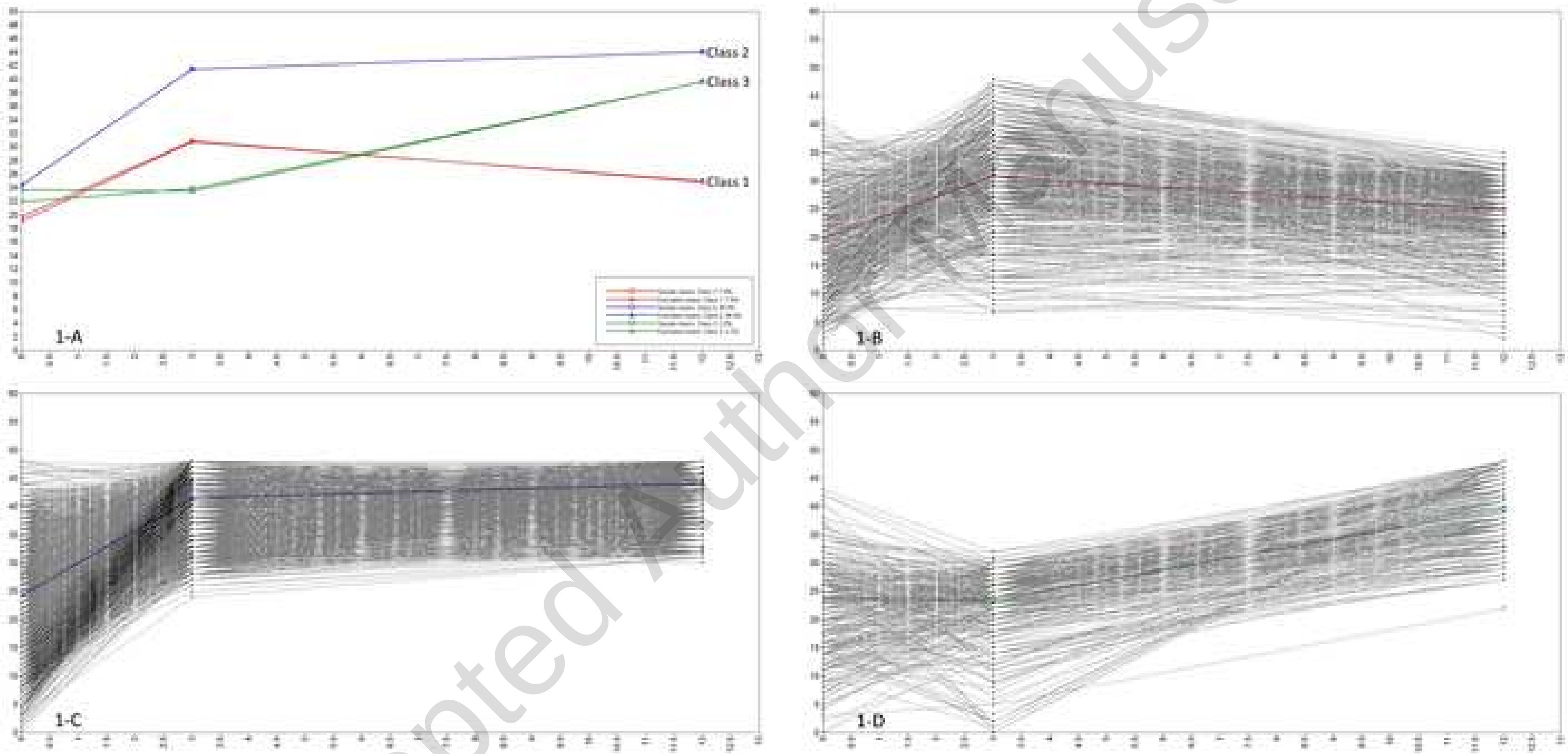
OR = Odds Ratio, CI = Confidence Interval

Table 4

TABLE IV Multivariable Multinomial Logistic Regression Analysis With Class Membership as Dependent Variable (Class 'Fast Starters' as Reference Class)

	Slow Starters vs. Fast Starters		Late Dippers vs. Fast Starters	
	OR (95% CI)	*p*	OR (95% CI)	*p*
Age >75 yrs (vs. ≤75 yrs)	1.22 (0.83 – 1.79)	0.310	1.62 (1.22 – 2.15)	0.001
Female (vs. male)	1.63 (1.14 – 2.33)	0.007	1.22 (0.94 – 1.59)	0.132
Smoking (vs. no smoking)	1.95 (1.26 – 3.03)	0.003	1.68 (1.17 – 2.42)	0.005
ASA III-IV (vs. I-II)	1.20 (0.78 – 1.87)	0.405	1.41 (1.05 – 1.91)	0.023
BMI				
Normal weight (ref.)	1.0	-	1.0	-
Overweight (BMI 25 – 30)	1.19 (0.82 – 1.74)	0.360	1.17 (0.86 – 1.58)	0.320
Obese (BMI >30)	1.54 (1.02 – 2.33)	0.041	1.96 (1.43 – 2.69)	<0.001
Had previous surgery	0.82 (0.19 – 3.51)	0.788	1.13 (0.51 – 2.51)	0.764
Charnley score				
A (ref.)	1.0	-	1.0	-
B1	1.41 (0.97 – 2.04)	0.070	1.14 (0.86 – 1.52)	0.353
B2	1.46 (0.98 – 2.19)	0.066	1.26 (0.93 – 1.71)	0.130
C	1.39 (0.57 – 3.36)	0.467	1.57 (0.82 – 3.00)	0.171
EQ-5D item 'Self-Care'				
No problems (ref.)	1.0	-	1.0	-
Some problems washing or dressing	0.98 (0.70 – 1.37)	0.887	1.41 (1.09 – 1.82)	0.008
Unable to wash or dress	1.40 (0.42 – 4.70)	0.590	2.90 (1.39 – 6.03)	0.004
EQ-5D item 'Usual Activities'				
No problems (ref.)	1.0	-	1.0	-
Some problems with performing usual activities	1.50 (0.81 – 2.77)	0.201	1.36 (0.81 – 2.27)	0.242
Unable to perform usual activities	1.89 (0.86 – 4.13)	0.112	1.53 (0.83 – 2.83)	0.175
EQ-5D item 'Pain / Discomfort'				
No pain or discomfort (ref.)	1.0	-	1.0	-
Moderate pain or discomfort	1.23 (0.50 – 3.04)	0.660	1.79 (0.71 – 4.48)	0.215
Extreme pain or discomfort	1.65 (0.64 – 4.26)	0.305	2.47 (0.96 – 6.33)	0.060
EQ-5D item 'Anxiety / Depression'				
Not anxious or depressed (ref.)	1.0	-	1.0	-
Moderately anxious or depressed	0.93 (0.63 – 1.36)	0.699	1.31 (1.00 – 1.71)	0.048
Extremely anxious or depressed	1.84 (0.92 – 3.71)	0.086	1.86 (1.06 – 3.24)	0.030
EQ-5D VAS (per 10 points)	0.96 (0.89 – 1.04)	0.366	0.91 (0.86 – 0.97)	0.003
Approach				
Posterolateral (ref.)	1.0	-	1.0	-
Anterior	0.47 (0.29 – 0.78)	0.003	0.71 (0.50 – 1.01)	0.057
Anterolateral	0.64 (0.25 – 1.60)	0.335	1.13 (0.63 – 2.02)	0.689
Direct lateral	1.39 (0.86 – 2.24)	0.176	2.18 (1.58 – 3.02)	<0.001
Other	1.00 (1.00 – 1.00)	1.000	1.00 (1.00 – 1.00)	1.000
Fixation				
Cemented (ref.)	1.0	-	1.0	-
Hybrid – cemented acetabulum	0.75 (0.21 – 2.70)	0.665	1.79 (1.00 – 3.21)	0.049
Hybrid – cemented femur	1.97 (0.97 – 4.01)	0.060	1.54 (0.91 – 2.63)	0.110
Uncemented	1.19 (0.77 – 1.82)	0.435	0.89 (0.67 – 1.19)	0.431
Femoral head diameter				
32 mm (ref.)	1.0	-	1.0	-
22 - 28 mm	0.95 (0.62 – 1.45)	0.814	1.29 (0.98 – 1.71)	0.066
≥36 mm	1.16 (0.77 – 1.76)	0.479	1.00 (0.72 – 1.40)	0.998

OR = Odds Ratio, CI = Confidence Interval


Class 2
Class 3
Class 1
1-A
1-B
1-C
1-D

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 groups 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 groups (i.e. all individuals within a certain group 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 GMM models and compared the results to our conventional growth model. In both the LCGA and GMM models 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 GMM models, 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.

Appendix 2

Results

Selection of the final model

The conventional one-class growth model showed a large amount of variability in preoperative OHS scores 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 GMM models, although **Figure 1 in this appendix** shows that this decrease starts to flatten somewhat after the three-class models. The entropy (Table 2 **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 models than in the GMM models. Seeing this heterogeneity, combined with worse fit statistics, we carried on with the GMM models.

Up to the 3-class GMM model, 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 model as our final model.

We subsequently evaluated the classification accuracy of our final model by investigating whether the estimated probability of group membership corresponded closely to the proportion classified in that group 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 group membership for individuals to each group 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, TABLE 1 Comparison of Preoperative Patient Characteristics between Patients with No, Some and All OHS scores missing

Variable	**No OHS scores missing (N=6030)**	**1 or 2 OHS scores missing (N=19328)**	**All OHS scores 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 score			
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 %

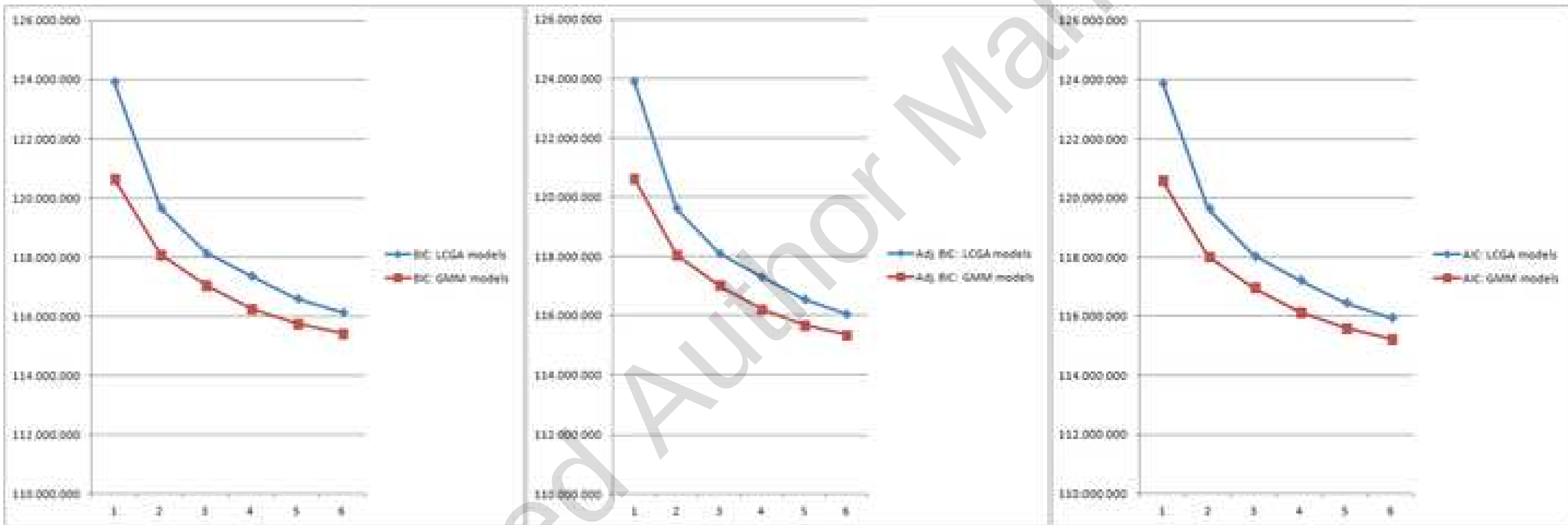
OHS = Oxford Hip Score

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

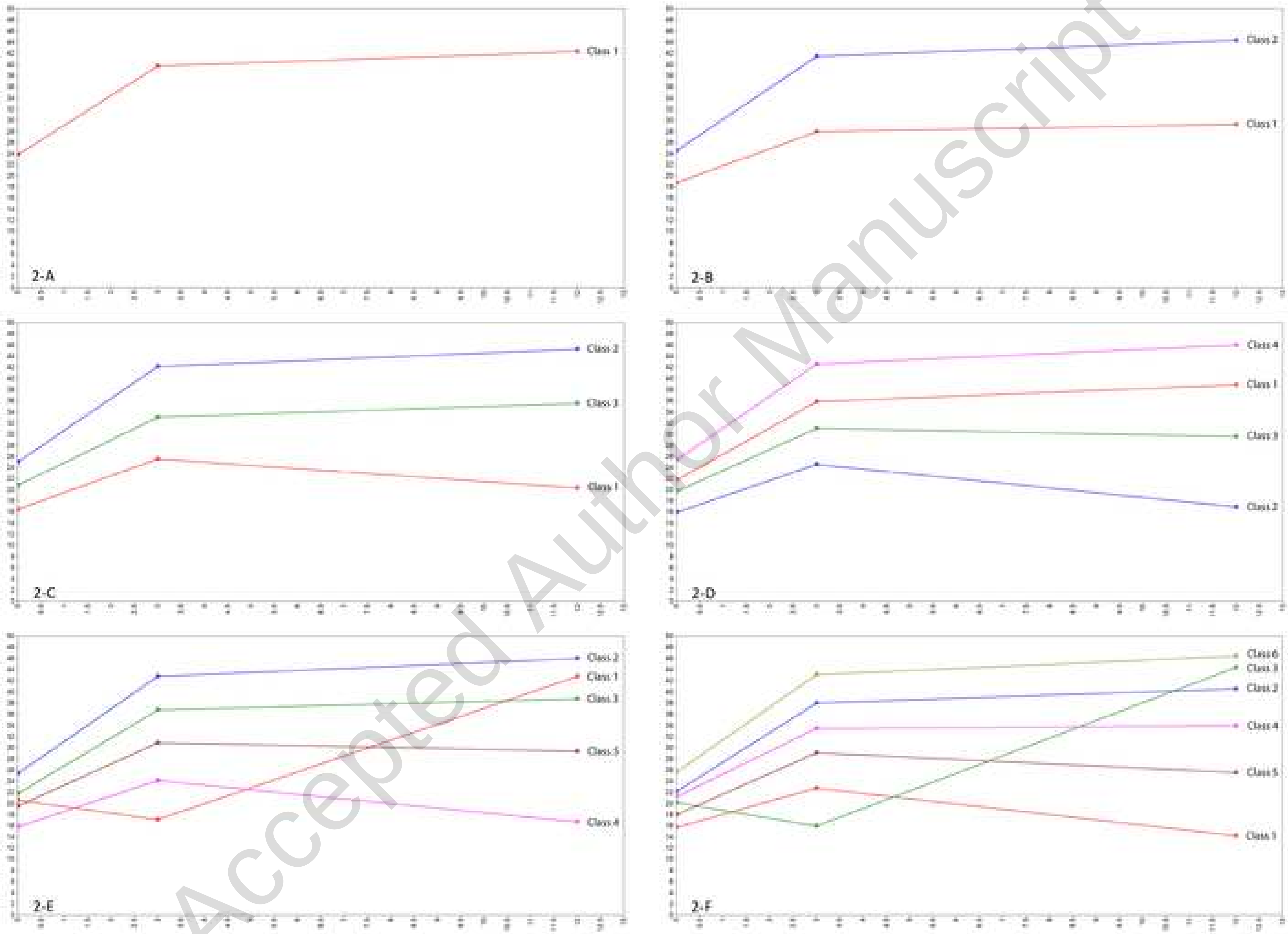
Class	Estimated probability of group membership	95% CI*	Proportion classified in group based on highest posterior probability	AvePP	OCC
Slow Starters	0.052	0.037 – 0.069	0.046	0.863	113.96
Late Dippers	0.078	0.068 – 0.088	0.077	0.913	123.48
Fast Starters	0.869	0.852 – 0.886	0.877	0.979	7.01

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

Appendix 2, Figure 1



Appendix 3, Figure 2-A through 2-F



Appendix 3, Figure 3-A through 3-F