Chapter 9

Development of preoperative prediction models for pain and functional outcome after total knee arthroplasty using The Dutch Arthroplasty Register (LROI) data

Jaap J. Tolk, J. (Erwin) H. Waarsing, Rob P.A. Janssen, Liza N. van Steenbergen, Sita M.A. Bierma-Zeinstra, Max Reijman.

Journal of Arthroplasty. 2019 Oct 18. [Epub ahead of print]

Abstract

Background

One of the main determinants of treatment satisfaction after total knee arthroplasty (TKA) is the fulfillment of preoperative expectations. For optimal expectation management, it is useful to accurately predict the treatment result. Multiple patient factors registered in the Dutch Arthroplasty Register (LROI) can potentially be utilized to estimate the most likely treatment result. The aim of the present study was to create and validate models that predict residual symptoms for patients undergoing primary TKA for knee osteoarthritis.

Methods

Data was extracted from the LROI of all TKA patients who had pre- and postoperative PROMs registered. Multivariable logistic regression analyses were performed to construct predictive algorithms for satisfaction, treatment success, and residual symptoms concerning pain at rest and during activity, sit-to-stand movement, stair negotiation, walking, performance of activities of daily living, kneeling and squatting. We assessed predictive performance by examining measures of calibration and discrimination.

Results

Data of 7071 patients could be included for data analysis. Residual complaints on kneeling (female 72% / male 59%) and squatting (female 71% /male 56%) were reported most frequently, and least residual complaints were scored for walking (female 16% / male 12%) and pain at rest (female 18% / male 14%). The predictive algorithms were presented as clinical calculators that present the probability of residual symptoms for an individual patient. The models for residual symptoms concerning sit-to-stand movement, stair negotiation, walking, ADL and treatment success showed acceptable discriminative values (Area Under the Curve (AUC) 0.68 – 0.74). The algorithms for residual complaints regarding kneeling, squatting, pain and satisfaction showed less favorable results (AUC 0.58 – 0.64). The calibration curves showed adequate calibration for most of the models.

Conclusion

A considerable proportion of patients have residual complaints after TKA. The present study showed that demographic and PROMs data collected in the LROI can be used to predict the probability of residual symptoms after TKA. The models developed in the present study predict the chance of residual symptoms for an individual patient on 10 specific items concerning treatment success, functional outcome and pain relief. This prediction can be useful for individualized expectation management in patients planned for TKA.

Introduction

The rate of satisfaction after total knee arthroplasty (TKA) is consistently reported around 80%, leaving 1 in 5 patients unsatisfied to some extent after their knee surgery.[1] One of the main determinants of treatment satisfaction is the fulfillment of preoperative expectations.[1–3] In this light, expectation management in TKA patients resulting in more realistic expectations, is thought to be advantageous to achieve optimal patient satisfaction.[2,4]

Individualized education about postoperative outcome should lead to more realistic patient expectations.[4–6] Previous research has identified specific expectations of TKA patients that are considered most important,[7,8] and expectations that are most often not fulfilled in TKA patients.[9] Expectations from both these categories are useful to address in pre-operative education, and include expectations on pain-relief and improvement in walking, stair negotiation, performance of daily activities, change in position, kneeling and squatting. [8,9]

It has been shown that a useful prediction on postoperative outcome can be made based on demographic factors, pre-operative pain scores and patient-reported outcome measure (PROM) scores.[10–12] However, existing outcome prediction tools mainly focus on identifying patients at risk of not benefiting from TKA in general.[13–15] Specific information on pain and functional outcome to guide pre-operative expectation management is not provided by these tools.[13–15] Furthermore, a patient might improve in general, but could experience residual complaints on some specific items which might not be distinguished by these tools. Therefore, for use in an expectation management intervention, there is a need for an outcome prediction model that provides specific information on the probability of residual pain and functional limitations after TKA. [10–12]

Multiple patient factors that are obtained from the Dutch Arthroplasty Register (Landelijke Registratie Orthopedische Implantaten [LROI]) can potentially be utilized to estimate the most likely outcome on pain and functional outcome for an individual patient.[16] If this data can be used to make a prediction of the treatment result, the data would not only be useful for measurement of quality of care on a group level, but could be of direct value for the individual patient.

Therefore, the aim of the present study was to create and validate models that predict the chance of having residual symptoms on 10 specific outcome parameters at 12-month follow-up for individual patients undergoing primary TKA for knee osteoarthritis (OA). Since there is a known difference between men and women in functional outcome after TKA, the development of the prediction models was stratified by gender.[17,18]

Material and Methods

Study setting and data collection

Data included in this study was obtained from the LROI. The LROI is a Dutch nationwide registry that collects data on all joint arthroplasties. All hospitals in the Netherlands participate in this registry, that was founded by the Dutch Orthopedic Association in 2007. The registry had a completeness rate of 99% for primary TKAs in 2016.[19]

Patients

Data was extracted from the registry on patients with a primary TKA for osteoarthritis, who had pre-operative and 12-month follow-up PROMs registered in the period January 2015 until July 2017, (n=7071). Only data of hospitals that provided PROM data on at least 25 patients with preoperative and postoperative PROMs questionnaires was included (53 hospitals).[20]

For assessment of generalizability, data of patients that met the inclusion criteria during the study period, but did not have PROMs registered were extracted (n=31022). A comparison for patient characteristics of the groups with and without PROMs registered was made using a chi-square test for ordinal parameters and student's t-test for continuous variables.

Data collection

Patient characteristics that are included in the LROI are age, sex, operation side, general health using the American Society of Anesthesiologists (ASA) classification system (dichotomized for regression analysis purposes; ASA I-II vs ASA III-IV),[21] body mass index (BMI) in kg/m^2, smoking (yes/no), previous operation on the affected joint (yes/no) and orthopedic vitality using the Charnley score (dichotomized for regression analysis purposes; Charnley A vs B & C).[22] Surgical variables such as type of procedure and type of implant are registered.

Patient-Reported Outcome Measures

For primary TKA patients, the Knee injury and Osteoarthritis Outcome Score - Physical Function Short Form (KOOS-PS), Oxford Knee Score (OKS), EuroQol 5D-3L and Numerical Rating Scale (NRS) on pain and satisfaction are obtained pre-operatively, at 6- and 12-months follow-up. The individual questions and composite scores of each PROM were available for analysis.

The KOOS-PS Dutch version assesses physical function on 7 different activities. These are scored on a 5-point Likert scale ranging from none to extreme difficulty. A normalized score can be calculated ranging from 0 indicating no symptoms and 100 indicating extreme symptoms.[23] KOOS-PS has shown to be reliable, valid and responsive to change in knee OA patients.[23–26]

The Dutch version OKS is a PROM on pain and function after TKA.[27] Twelve items are scored on a 5-point Likert scale, leading to a total score ranging from 0 to 48. Lower scores indicate more symptoms.[27] It has good measurement properties in knee OA patients.[26,27] The cut-off value for treatment success after TKA is set at >32.5 points on the OKS total score.[28]

The Dutch version of the EuroQol 5D-3L is a PROM on health-related quality of life (HRQoL).[29] The score consists of 5 questions (EQ-5D index) and the EQ visual analog scale (EQ-VAS). The EQ-VAS records the respondent's self-rated health on a vertical, visual analog scale where the endpoints are labeled 'Best imaginable health state' and 'Worst imaginable health state'. The 5 questions are scored on a 3-point Likert scale, from which the EQ-5D index can be calculated. The outcome scores range from -0.333 to 1.0, where 1.0 represents perfect HRQoL. This questionnaire has good reliability and validity in knee OA patients and has been validated for the Dutch population.[30,31]

Pain scores were measured using an NRS for pain during activity and in rest. A score of 0 represented 'no pain' and a score of 10 represented 'worst imaginable pain'. NRS pain values of ≤3 correspond with none or mild pain, and NRS pain score >3 represents moderate to severe pain.[32] The NRS has good reliability and responsiveness.[24]

Satisfaction with treatment results at 12-month follow-up is scored on an NRS scale ranging from 0 (not satisfied at all) to 10 (best possible satisfaction). 'Very satisfied with treatment result' is defined as an NRS satisfaction score of ≥8.

Outcome of interest

The developed prediction tools provide a probability of persistent complaints on items that have been identified as important and/or most often unfulfilled in TKA patients.[7,9,33] The items addressed are residual symptoms concerning pain at rest and during activity, sit-to-stand movement, stair negotiation, walking, performance of activities of daily living, kneeling and squatting. Furthermore, satisfaction and treatment success in general were included as outcome parameters of interest in the predictive model development. For each of these factors, a corresponding question from the available PROMs in the LROI data set was identified and a threshold for remaining symptoms was chosen (Table 1).

Statistical analysis

The LROI data was randomly divided into two sets using SPSS Statistics version 24.0 (IBM Corporation). Part one of the dataset (70% of patients) was used for building the prediction models, and the remaining 30% of the dataset for validation of the models.

Model development

Gender differences play an important role in outcome after TKA, therefore the prediction model development was performed for men and women separately.[17,18] Patient

characteristics and baseline PROM scores (questions and total scores) were used as candidate predictors. Categorical variables were presented using frequency and percentage, and continuous variables as mean with standard deviation (SD). Candidate variables with more than 25% missing data were excluded. Patterns of missingness were investigated to assess the presence of a non-random element to the missing data. For the remaining variables, multiple imputations were performed to estimate the missing values, resulting in 5 imputed data sets.[34]

A logistic regression analysis was performed for each of the 10 dependent variables listed in Table 1, for men and women separately. Potential predictors were identified in the univariable analyses, with significance set at $p < 0.15$. Potential predictors identified in the univariable analysis were entered into the multivariable model. This resulted in a predictive algorithm for each outcome variable of interest.

We assessed the predictive performance of these algorithms by examining measures of calibration and discrimination. Discrimination is the ability of the prediction model to distinguish between patients that have residual complaints after TKA from patients that do not have complaints. This was assessed by calculating the area under the receiver operating characteristic curve statistic (AUC) and Nagelkerke's R^2 as a measure of explained variation. Calibration considers the agreement between the predicted and the actual outcome. This was assessed using calibration plots, in which patients were classified by tenths of the predicted risk, augmented by a locally estimated scatterplot smoothed (loess) line over the entire predicted probability range.[35] Predictions of a perfect model should lie on the 45-degree line for agreement with the actual outcome.

All statistical analyses were performed using SPSS statistics version 24.0 (IBM Corporation). This study was conducted and reported in line with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines.[36]

Results

Between January 2015 and July 2017, 38093 patients were registered in the LROI with a primary TKA for knee osteoarthritis. Of these patients, 7071 had had pre-operative and 12-month follow-up PROMs available (flowchart figure 1). There were no candidate predictors with >25% of missing data in this cohort, and there was no non-random element to the missing data. Therefore, multiple imputation could be performed as planned.[34] When compared to the group of patients that did not have PROMs available, the cohort of patients that did have PROMs available was slightly younger (68.4 (+/-8.5) vs 68.8 (+/-9.1) years, $p<0.001$), more often male (37.3% vs 35.8%, $p=0.015$), more often smoker (8.3% vs 9.5%, $p=0.002$) and had a marginally lower BMI (29.6 (+/-4.8) vs. 29.8 (+/-5.1) kg/m^2, $p=0.001$). Full characteristics of both groups can be found in supplementary file 1.

The 7071 patients that were available for development of the prediction model were randomly divided into two groups. A training cohort (n=4951) and a cohort for internal validation of the prediction models (n=2120). Patient characteristics of these groups did not show important differences (Table 2). The incidence of residual complaints at 1-year follow-up is presented in Table 3. Of the outcome variables under study, residual complaints on kneeling (female 72% /male 59%) and squatting (female 71% / male 56%) were reported most frequently. Least residual complaints were scored for walking (female16% / male 12%) and pain at rest (female18% / male 14%).

The significant factors identified in the univariate analyses were all entered into the multivariable regression models. The majority of the individual PROM questions and PROM total scores were included in the different models. Other patient characteristics such as Charnley score, side affected, previous surgery to the affected joint and smoking status showed to be significant univariable predictors in only a few models. In almost all models the factor that showed the highest predictive value was the pre-operative PROM question corresponding with the factor that was predicted. For example, in the model predicting residual complaints when rising from a chair for female patients, KOOS-PS question 6 (the question addressing problems with sit-to-stand movement) showed to be the strongest predictor with an odds ratio of 1.61 (95% CI 1.374 to 1.882).

The full prediction models are presented as clinical calculators in supplementary file 2 for female patients and supplementary file 3 for male patients (available online only). The calculator presents an individual patients' chance of residual symptoms concerning pain at rest and during activity, sit-to-stand movement, stair negotiation, walking, performance of activities of daily living, kneeling and squatting and the chance of dissatisfaction and no overall treatment success. An example of the output provided by the prediction model is presented in figure 2.

The performance of these models in terms of discrimination are presented in Table 4 for both the training and validation cohort. The predictive algorithms for residual symptoms concerning rising from a chair, stair negotiation, walking, activities of daily living and treatment success showed acceptable discriminative values (AUC 0.68 – 0.74) and explained fraction of variance (Nagelkerke R^2 0.13-0.21). The prediction models for residual complaints regarding kneeling, squatting, pain and satisfaction showed the least favorable results on discrimination (AUC 0.58 – 0.64) and explained variance (Nagelkerke R^2 0.04-0.11). The calibration of the models in the validation cohort is presented using calibration curves. Figure 3 shows the performance of the models for women and Figure 4 for men.

Discussion

We have developed and internally validated prediction models for residual symptoms 12 months after TKA. The models predict the chance of residual symptoms on a set of outcome

parameters that are considered most important in pre-operative expectation management. The output of the models presents an individual patients' chance of residual complaints (figure 2). These models generally showed fair discrimination and good calibration.

The present study shows that a considerable proportion of patients has residual complaints after TKA. Residual complaints on kneeling and squatting were reported most frequently, while the least residual complaints were reported for walking and pain at rest. Previously, TKA patients have been reported to still experience substantial functional impairment compared with their age-matched peers, especially in biomechanically demanding activities.[9,37] The results of the present study emphasize the fact that residual complaints are common and this should be considered in pre-operative decision making and expectation management.

In the study design we chose to develop separate prediction models for male and female patients. Residual complaints showed to be considerably more frequent in women over the whole range of outcome parameters that were analyzed. This confirms the conclusion of previous reports in this regard,[17] and supports the choice to develop separate prediction models for male and female patients in the present study. The predictive performance of the models did not show important differences between sexes.

The discriminative performance of most prediction models developed in the present study can be considered acceptable, with AUC values in the internal validation cohort of around 0.70.[38] The prediction models for residual complaints regarding kneeling, squatting, pain and dissatisfaction with treatment result showed the least favorable results. Because of the study design, prognostic variables included were limited to the variables registered in the LROI. As a result, not all pre-operative factors that are known to influence outcome after TKA, such as pre-operative expectations and radiological OA severity, could be included as predictors for the model.[10,39] Therefore, their influence on outcome and patient satisfaction could not be incorporated by the models, which might have limited performance of the models. At the same time, by including only predictors that patients currently already provide, we made sure the models can easily be integrated in clinical practice.

For most outcome variables adequate calibration curves are reported. For predicted probabilities of residual complaints exceeding 0.5, the loess lines tend to deviate from the ideal line. This is probably due to the low numbers of patients in this category. Therefore, predicted probabilities of residual complaint exceeding 50% have to be interpreted with some caution. Considering residual complaints on kneeling and squatting the distribution of predicted probabilities was reversed in comparison to all other categories. This is in line with the relatively high prevalence of residual complaints; 57% for male and 72% for female patients. For kneeling and squatting prediction for probabilities of 40% and higher seem adequately calibrated.

To the authors' knowledge, this is the first time such a comprehensive set of prediction models on treatment result after TKA is reported. Previous efforts have been made to predict the treatment result of TKA patients.[11,13–15] These studies mainly aimed to predict overall response or treatment success in general.[11,13–15] For this purpose, Dowsey et al. created a prognostic nomogram predicting probabilities of nonresponse to TKA.[10] A limited set of predictors was derived, and internal validation showed acceptable calibration and discrimination (c-statistic 0.74) quite similar to the performance of our models on treatment success.[10] Unfortunately, these favorable results were only partly supported on external validation of this predictive nomogram by Riddle et al.[13] Especially calibration showed poor agreement between actual vs predicted probabilities of nonresponse.[13] Therefore, the model developed by Riddle et al. does not seem applicable beyond the population in which it was developed.[10,13] We intend to use the models presented in the present study in a very comparable population as in which they were developed. Therefore, we do not expect similar calibration problems.

For effective expectation management, a prediction tool should ideally provide specific information on the most likely outcome on pain and function for an individual patient.[6] Pua et al. constructed a model on risk for walking limitations after TKA.[11] In contrast to the models described in the previous paragraph that predict outcome in general, the model by Pua et al. is the only predictive model for a specific functional outcome parameter available in literature. The model is based on a predictor set that has limited overlap with the predictors used in the present study, and used data on postoperative recovery in addition to pre-operative measurements. The predictive performance of their nomogram seems to be quite comparable to our model on walking limitations, with a reported c-statistic of 0.71.[11]

The candidate predictors for the prediction models constructed in the present study are part of routinely collected data. Since this information is always obtained for patients in the Dutch system, it was not deemed necessary to reduce the number of predictors to obtain the smallest predictor set possible.[36] Further reduction would only lead to loss of predictive power without an increase in usability or reduction of the burden for patients. The data entered in routinely obtained questionnaires can directly be transformed into an individualized prediction for probability of residual complaints (example output in figure 2). These predictions are a good basis for improved pre-operative expectation management.

A limitation of the present study is that only 19% of the complete cohort of TKA patients had completed PROMs at baseline and 12 months follow up. Therefore, only this subset of the total cohort could be included for development of the models. This is probably partly caused by the knee PROM follow-up program was only introduced in 2015 in the Netherlands and implementation has just recently started in some clinics. To avoid selection bias in hospitals that included only a very few patients, hospitals with <25 patients registered were excluded

from the analysis. The group included for modeling statistics and the group that could not be included showed significant differences for some patient characteristics, but these differences were very small and therefore probably not clinically relevant. For this reason, the results of the present study can be considered generalizable to the whole Dutch population of patients undergoing primary TKA for knee OA. Patient characteristics, methods of measurement and healthcare systems might differ across populations, and can potentially influence predictive performance.[13] These models are specific for the Dutch healthcare setting. Generalization beyond this population would therefore warrant external validation and possibly recalibration. Nevertheless, the concept of outcome prediction based on routinely gathered patient characteristics and PROMs is likely to be applicable beyond the Dutch healthcare setting.

Another limitation is that although we showed that a useful prediction can be made for the probability of residual symptoms after TKA, the predictions cannot be used as explicit indication criteria for TKA. A cut off value for appropriateness of TKA is not provided and the level of discrimination would not justify such strong conclusions. Furthermore, the indication for TKA has a subjective nature where each patient has to consider the risks and benefits against their own values.[16] In the authors' opinion, the prediction results are useful for identifying patients at risk for residual complaints, and individualizing expectation management in this regard.

Conclusion

A considerable proportion of patients have residual complaints after TKA. The present study showed that demographic and PROMs data collected in the LROI can be used to predict the probability of residual symptoms after TKA. The models developed in the present study predict the chance of residual symptoms for an individual patient on 10 specific items concerning functional outcome and pain relief. This prediction can be useful for individualized expectation management in patients planned for TKA.

Funding:
This study was supported by a grant from the van Rens Funds, and conducted with data of the Dutch Arthroplasty Registry.

References

1. Noble PC, Conditt M a, Cook KF, Mathis KB. The John Insall Award: Patient expectations affect satisfaction with total knee arthroplasty. *Clin Orthop Relat Res*. 2006;452(452):35-43. doi:10.1097/01.blo.0000238825.63648.1e
2. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ. Patient satisfaction after total knee arthroplasty: Who is satisfied and who is not? *Clin Orthop Relat Res*. 2010;468(1):57-63. doi:10.1007/s11999-009-1119-9
3. Hamilton DF, Lane J V, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. *BMJ Open*. 2013;3(4):1-8. doi:10.1136/bmjopen-2012-002525
4. Tolk JJ, Janssen RPA, Haanstra TM, Bierma-zeinstra SMA, Reijman M. The EKSPECT study : the influence of Expectation modification in Knee arthroplasty on Satisfaction of PatiEnts : study protocol for a randomized Controlled Trial. *Trials*. 2018;19(437).
5. Mancuso C, Graziano S, Briskie L, et al. Randomized trials to modify patients' preoperative expectations of hip and knee arthroplasties. *Clin Orthop Relat Res*. 2008;466(2):424-431. doi:10.1007/s11999-007-0052-z
6. Judge A, Arden NK, Cooper C, et al. Predictors of outcomes of total knee replacement surgery. *Rheumatology (Oxford)*. 2012;51(10):1804-1813. doi:10.1093/rheumatology/kes075
7. Mancuso CA, Sculco TP, Wickiewicz TL, Jones EC, Robbins L, Warren RF. Patients' expectations of knee surgery. *J Bone Jt Surg Am*. 2001;83-A(7):1005-1012.
8. Lange T, Schmitt J, Kopkow C, Rataj E, Günther K-P, Lützner J. What Do Patients Expect From Total Knee Arthroplasty? A Delphi Consensus Study on Patient Treatment Goals. *J Arthroplasty*. 2017:1-8. doi:10.1016/j.arth.2017.01.053
9. Tilbury C, Haanstra TM, Leichtenberg CS, et al. Unfulfilled Expectations After Total Hip and Knee Arthroplasty Surgery: There Is a Need for Better Preoperative Patient Information and Education. *J Arthroplasty*. 2016;31(10):2139-2145. doi:10.1016/j.arth.2016.02.061
10. Dowsey MM, Spelman T, Choong PFM. Development of a Prognostic Nomogram for Predicting the Probability of Nonresponse to Total Knee Arthroplasty 1 Year After Surgery. *J Arthroplasty*. 2015;31:1654-1660. doi:10.1016/j.arth.2016.02.003
11. Pua YH, Seah FJT, Clark RA, Poon CLL, Tan JWM, Chong HC. Development of a prediction model to estimate the risk of walking limitations in patients with total knee arthroplasty. *J Rheumatol*. 2016;43(2):419-426. doi:10.3899/jrheum.150724
12. Van Onsem S, Van Der Straeten C, Arnout N, Deprez P, Van Damme G, Victor J. A New Prediction Model for Patient Satisfaction After Total Knee Arthroplasty. *J Arthroplasty*. 2016;31:1-9. doi:10.1016/j.arth.2016.06.004
13. Riddle DL, Golladay GJ, Jiranek WA, Perera RA. External Validation of a Prognostic Model for Predicting Nonresponse Following Knee Arthroplasty. *J Arthroplasty*. 2017;32(4):1153-1158. doi:10.1016/j.arth.2016.11.007
14. Barlow T, Dunbar M, Sprowson A, Parsons N, Griffin D. Development of an outcome prediction tool for patients considering a total knee replacement--the Knee Outcome Prediction Study (KOPS). *BMC Musculoskelet Disord*. 2014;15(1):451. doi:10.1186/1471-2474-15-451
15. Lungu E, Desmeules F, Dionne CE, Belzile EL, Vendittoli P-A. Prediction of poor outcomes six months following total knee arthroplasty in patients awaiting surgery. *BMC Musculoskelet Disord*. 2014;15:299. doi:10.1186/1471-2474-15-299
16. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can Preoperative Patient-reported Outcome Measures Be Used to Predict Meaningful Improvement in Function After TKA? *Clin Orthop Relat Res*. 2016;475(1):1-9. doi:10.1007/s11999-016-4770-y
17. Cherian J, O'Connor M, Robinson K, Jauregui J, Adleberg J, Mont M. A Prospective, Longitudinal Study

of Outcomes Following Total Knee Arthroplasty Stratified by Gender. *J Arthroplasty*. 2015;30:1372-1377. doi:10.1055/s-0034-1388658
18. Volkmann ER, FitzGerald JD. Reducing gender disparities in post-total knee arthroplasty expectations through a decision aid. *BMC Musculoskelet Disord*. 2015;16:16. doi:10.1186/s12891-015-0473-x
19. Dutch Arthroplasty Register (LROI). Report on data quality coverage and completeness. http://www.lroi-rapportage.nl/data-quality-coverage-and-completeness. Published 2016.
20. Peters RM, van Beers LWAH, van Steenbergen LN, et al. Similar Superior Patient-Reported Outcome Measures for Anterior and Posterolateral Approaches After Total Hip Arthroplasty: Postoperative Patient-Reported Outcome Measure Improvement After 3 months in 12,774 Primary Total Hip Arthroplasties Using the Ante. *J Arthroplasty*. 2018;33(6):1786-1793. doi:10.1016/j.arth.2018.01.055
21. Fitz-Henry J. The ASA classification and peri-operative risk. *Ann R Coll Surg Engl*. 2011;93(3):185–187.
22. Charnley J. The long-term results of low-friction arthroplasty of the hip performed as a primary intervention. *J Bone Jt Surgery, Br …*. 1972;54(1):61-76.
23. Perruccio A V, Lohmander LS, Canizares M, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. *Osteoarthr Cart*. 2008;16:542-550. doi:10.1016/j.joca.2007.12.014
24. Ruyssen-Witrand A, Fernandez-Lopez CJ, Gossec L, Anract P, Courpied JP, Dougados M. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. *Clin Exp Rheumatol*. 2011;29(2):231-237.
25. de Groot IB, Favejee MM, Reijman M, Verhaar J a N, Terwee CB. The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: a validation study. *Health Qual Life Outcomes*. 2008;6:16. doi:10.1186/1477-7525-6-16
26. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Outcome. *Arthritis Care Res (Hoboken)*. 2011;63 Suppl 1(November):S208-28. doi:10.1002/acr.20632
27. Haverkamp D, Breugem SJM, Sierevelt IN, Blankevoort L, van Dijk CN. Translation and validation of the Dutch version of the Oxford 12-item knee questionnaire for knee arthroplasty. *Acta Orthop*. 2005;76(3):347-352. doi:10.1080/00016470510030814
28. Hamilton DF, Loth FL, MacDonald DJ, et al. Treatment Success Following Joint Arthroplasty: Defining Thresholds for the Oxford Hip and Knee Scores. *J Arthroplasty*. 2018;33(8):2392-2397. doi:10.1016/j.arth.2018.03.062
29. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337-343.
30. Lamers L, Stalmeier P, McDonnell J, Krabbe P, van Busschbach J. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. *Ned Tijdschr Geneeskd*. 2005;149(28):1574-1578.
31. Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Qual Life Res*. 2015;24(7):1775-1784. doi:10.1007/s11136-014-0910-6
32. Pinto PR, McIntyre T, Ferrero R, Almeida A, Araújo-Soares V. Risk Factors for Moderate and Severe Persistent Pain in Patients Undergoing Total Knee and Hip Arthroplasty: A Prospective Predictive Study. *PLoS One*. 2013;8(9). doi:10.1371/journal.pone.0073917
33. Smith EJ, Soon V-L, Boyd A, McAllister J, Deakin A, Sarungi M. What do Scottish Patients Expect of Their Total Knee Arthroplasty? *J Arthroplasty*. 2016;31(4):786-792. doi:10.1016/j.arth.2015.10.021
34. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17(1):1-10. doi:10.1186/s12874-017-0442-1
35. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33:517-535. doi:10.1002/sim.5941

36. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med*. 2015;162(1):W1. doi:10.7326/M14-0698
37. Noble PC, Gordon MJ, Weiss JM, Reddix RN, Conditt M a, Mathis KB. Does total knee replacement restore normal knee function? *Clin Orthop Relat Res*. 2005;(431):157-165. doi:10.1097/01.blo.0000150130.03519.fb
38. Gortmaker S, Hosmer D, Lemeshow S. Applied logistic regression. *Contemp Sociol*. 1994;23:159.
39. Jain D, Nguyen L-CL, Bendich I, et al. Higher patient expectations predict higher patient-reported outcomes, but not satisfaction, in total knee arthroplasty patients: a prospective multicenter study. *J Arthroplasty*. 2017;32(9S):S166-S170. doi:10.1016/j.arth.2017.01.008

Table 1. Thresholds for residual symptoms		
Factor to be predicted	**Corresponding question**	**Cut-off value for non-response**
1. Residual pain at rest	NRS pain rest	>3
2. Residual pain during activity	NRS pain activity	>3
3. Rising from a chair	KOOS-PS question 3 *'Indicate the degree of difficulty you have experienced in the last week due to your knee problem: Rising from sitting.'*	Moderate or higher (score 2-4)
4. Stair negotiation	OKS question 12 ***'Could you walk down a flight of stairs?'***	With moderate difficulty or worse (score 1-3)
5. Walking	OKS question 4 ***For how long are you able to walk before the pain in your knee becomes severe? (With or without a stick)***	5 - 15 minutes or less (score 1-3)
6. Activities of daily living	OKS question 9 ***How much has pain from your knee interfered with your usual work? (including housework)***	Moderately or more (score 1-3)
7. Kneeling	KOOS-PS question 6 *'Indicate the degree of difficulty you have experienced in the last week due to your knee problem: Kneeling'*	Moderate or higher (score 2-4)
8. Squatting	KOOS-PS question 7 *'Indicate the degree of difficulty you have experienced in the last week due to your knee problem: squatting'*	Moderate or higher (score 2-4)
9. Dissatisfaction	NRS satisfaction	<8
10. No treatment success	OKS total score	<32.5
NRS; Numerical Rating Scale. KOOS-PS; Knee injury and Osteoarthritis Outcome Score - Physical Function Short Form. OKS; Oxford Knee Score.		

Table 2. Patient characteristics and baseline scores of training and test cohort.				
	Training cohort		Test cohort	
	Female (n=3120)	Male (n=1831)	Female (n=1315)	Male (n=805)
Age, years	68.8 (8.6)	68.1 (8.1)	68.8 (8.6)	67.8 (8.2)
ASA classification, n (%)				
I	380 (12.2)	285 (15.6)	160 (12.2)	132 (16.4)
II	2213 (70.9)	1231 (67.2)	934 (71.0)	531 (66.0)
III-IV	527 (16.9)	315 (17.2)	221 (16.8)	142 (17.6)
Smoking, yes n (%)	219 (7.0)	187 (10.2)	91 (6.9)	87 (10.8)
Charnley score, n (%)				
A	1328 (42.6)	878 (48.0)	551 (41.9)	395 (49.1)
B1	1024 (32.8)	585 (31.9)	435 (33.1)	267 (33.2)
B2	663 (21.3)	320 (17.5)	284 (21.6)	125 (15.5)
C	105 (3.4)	48 (2.6)	45 (0.3)	17 (0.2)
BMI, kg/m^2	30.0 (5.2)	28.9 (4.0)	30.1 (5.4)	28.9 (4.1)
Side affected, right n (%)	1677 (53.8)	948 (51.8)	700 (53.2)	426 (52.9)
Previous surgery on affected joint, yes n (%)	852 (27.3)	787 (43.0)	369 (28.1)	351 (43.6)
EQ-VAS	67.1 (19.0)	72.0 (18.7)	67.5 (19.2)	71.5 (18.0)
EQ-index	0.57 (0.27)	0.66 (0.23)	0.58 (0.27)	0.65 (0.23)
KOOS-PS total	52.9 (15.1)	47.5 (14.4)	52.8 (14.7)	47.8 (13.8)
OKS total	22.3 (6.8)	25.4 (6.7)	22.3 (7.1)	25.2 (7.1)
NRS Pain activity	7.4 (1.9)	7.0 (2.0)	7.4 (1.8)	7.0 (2.1)
NRS Pain rest	5.3 (2.6)	4.6 (2.6)	5.3 (2.5)	4.7 (2.6)
Data presented as mean (standard deviation) or otherwise as mentioned. ASA classification; American Society of Anesthesiologists (ASA) classification system. BMI; Body Mass Index. EQ; EuroQol. VAS; visual analogue scale. KOOS-PS; Knee injury and Osteoarthritis Outcome Score - Physical Function Short Form. OKS; Oxford Knee Score. NRS; Numerical Rating Scale.				

Table 3. Frequency of residual complaints.			
Dependent variable	Sex	Training cohort (♀n=3120, ♂n=1831)	Validation cohort (♀n=1315, ♂n=805)
Residual pain rest (n [%])	♀	568 (18.2)	230 (17.5)
	♂	443 (14.2)	115 (14.3)
Residual pain activity (n [%])	♀	886 (28.4)	368 (28.0)
	♂	434 (23.7)	188 (23.4)
Rising from a chair (n [%])	♀	721 (23.1)	301 (22.9)
	♂	361 (19.7)	158 (19.6)
Stair negotiation (n [%])	♀	876 (27.8)	359 (27.3)
	♂	297 (16.2)	140 (17.4)
Walking (n [%])	♀	493 (15.8)	199 (15.1)
	♂	220 (12.0)	93 (11.6)
Activities of daily living (n [%])	♀	761 (24.4)	330 (25.1)
	♂	342 (18.7)	161 (20.0)
Kneeling (n [%])	♀	2237 (71.7)	940 (71.5)
	♂	1086 (59.3)	483 (60.0)
Squatting (n [%])	♀	2228 (71.4)	948 (72.1)
	♂	1038 (56.7)	459 (57.0)
Dissatisfaction (n [%])	♀	933 (29.9)	396 (30.1)
	♂	450 (24.6)	197 (24.5)
No treatment success (n [%])	♀	658 (21.1)	284 (21.6)
	♂	258 (14.1)	116 (14.4)

Table 4. Performance of predictive models – discrimination.							
		Training data set				Validation data set	
Dependent variable	Sex	AUC (95%CI)		Nagelkerke R2		AUC (95%CI)	
Pain rest	♀	0.69	(0.66-0.72)	0.11		0.65	(0.61-0.69)
	♂	0.68	(0.64-0.72)	0.09		0.62	(0.57-0.67)
Pain activity	♀	0.65	(0.63-0.67)	0.08		0.63	(0.60-0.66)
	♂	0.68	(0.65-0.71)	0.11		0.64	(0.62-0.66)
Rising from a chair	♀	0.76	(0.74-0.78)	0.21		0.73	(0.70-0.76)
	♂	0.73	(0.71-0.75)	0.16		0.70	(0.66-0.74)
Stair negotiation	♀	0.72	(0.70-0.74)	0.17		0.71	(0.69-0.73)
	♂	0.72	(0.69-0.75)	0.13		0.69	(0.66-0.72)
Walking	♀	0.74	(0.72-0.76)	0.16		0.72	(0.69-0.75)
	♂	0.71	(0.69-0.73)	0.13		0.69	(0.67-0.71)
Activities of daily living	♀	0.70	(0.67-0.73)	0.13		0.67	(0.64-0.70)
	♂	0.72	(0.70-0.74)	0.14		0.67	(0.64-0.70)
Kneeling	♀	0.65	(0.63-0.67)	0.08		0.64	(0.60-0.68)
	♂	0.65	(0.62-0.68)	0.09		0.62	(0.58-0.66)
Squatting	♀	0.64	(0.62-0.66)	0.07		0.63	(0.60-0.66)
	♂	0.63	(0.61-0.65)	0.08		0.60	(0.57-0.63)
Dissatisfaction	♀	0.61	(0.59-0.64)	0.04		0.58	(0.55-0.61)
	♂	0.63	(0.60-0.66)	0.06		0.59	(0.55-0.63)
No treatment success	♀	0.73	(0.70-0.76)	0.16		0.70	(0.67-0.73)
	♂	0.73	(0.71-0.75)	0.15		0.69	(0.67-0.71)
AUC; Area Under the Curve. CI; confidence interval.							

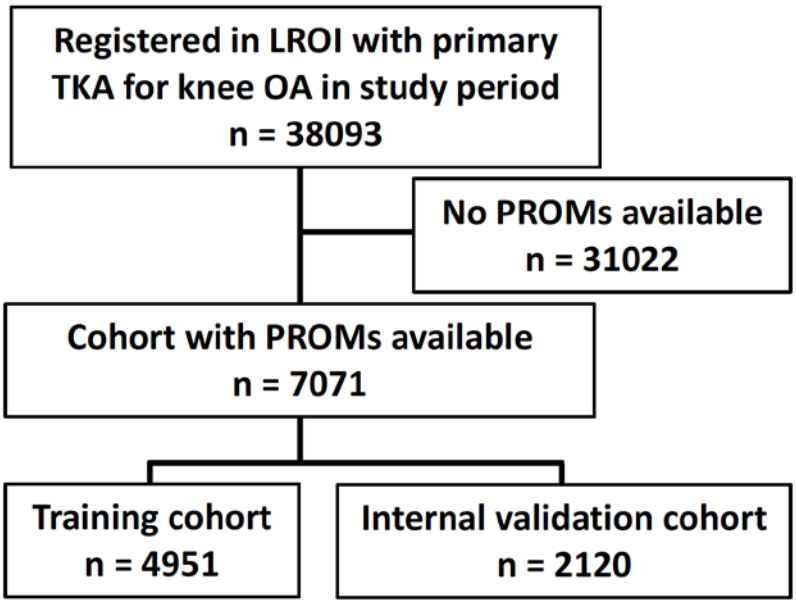

Figure 1. Flowchart for inclusion and exclusion of patients.

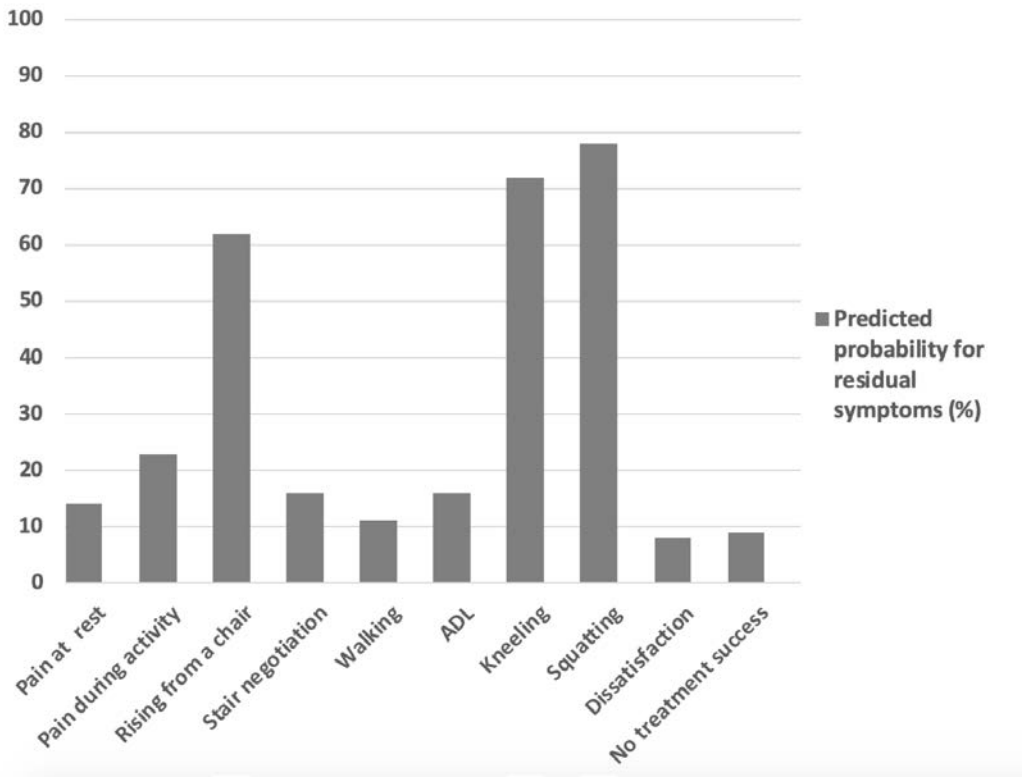

Figure 2. Example of output provided by the predictive model. The individual chances of residual symptoms for a patient on each of the 10 items presented are calculated based on patient characteristics, baseline EQ-5D scores, Oxford knee score and Knee injury and Osteoarthritis Outcome Score - Physical Function Short Form scores.

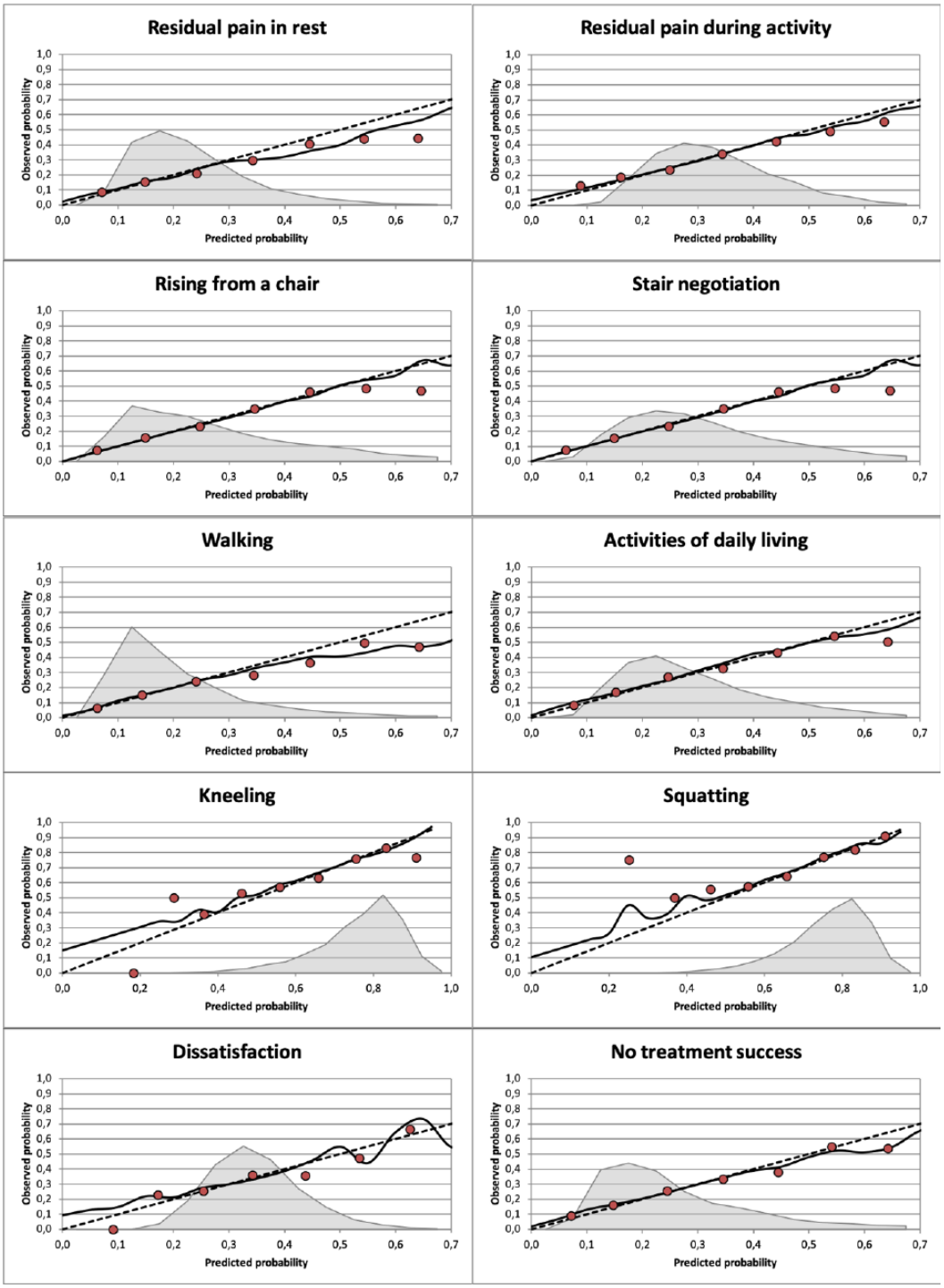

Figure 3. *Calibration curves for female patients. Predicted probability of residual symptoms on each specific outcome parameter is given on the x-axis, and the observed probability of residual symptoms is given on the y-axis. The dashed diagonal line represents perfect agreement between the predicted and actual probability of nonresponse. Tenths of the predicted risk are presented as reds dots, and augmented by a smoothed (loess) line over the entire predicted probability range (continuous line). In gray above the X-axis the distribution of the predicted probabilities.*

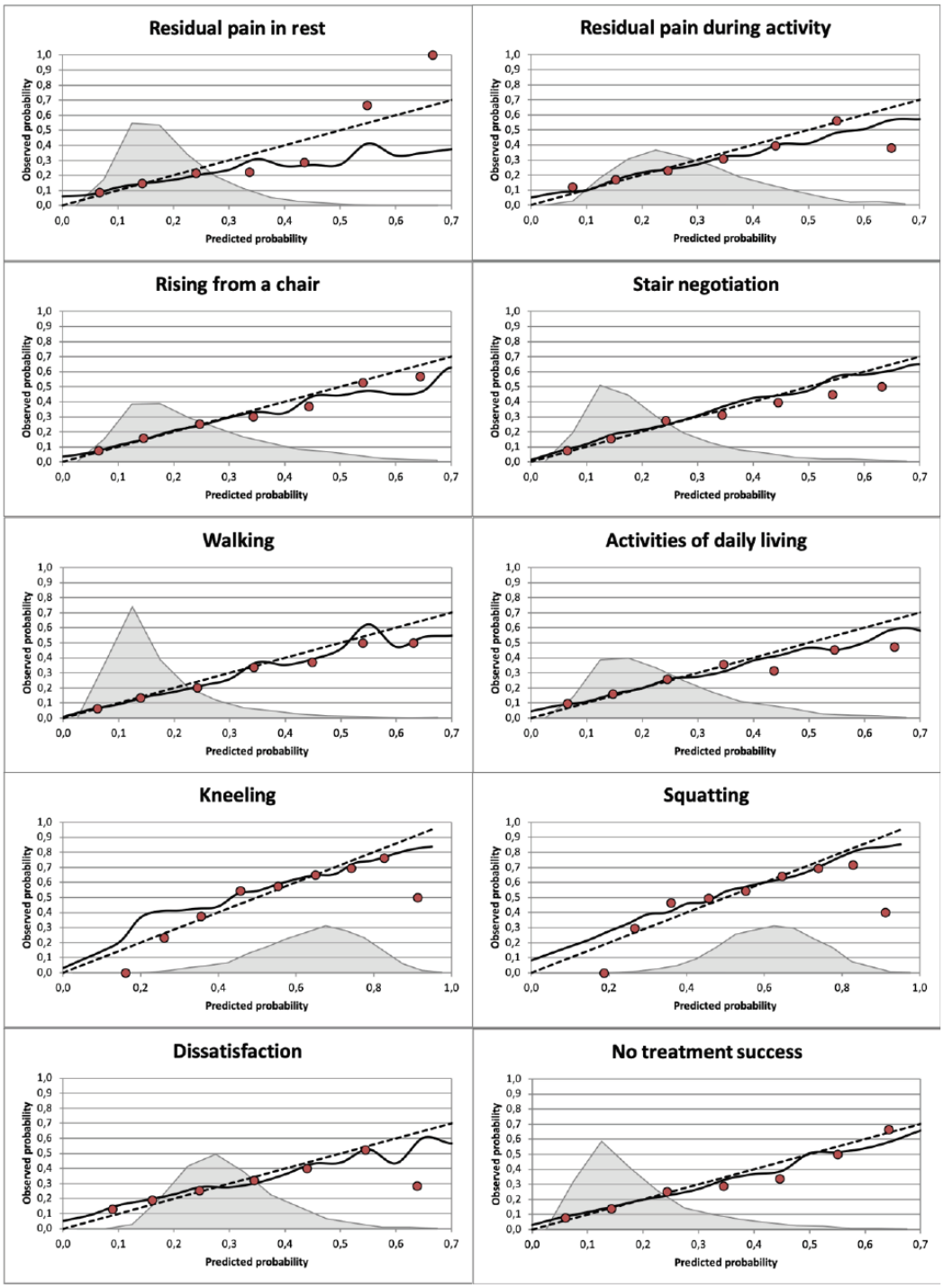

Figure 4. *Calibration curves for male patients. Predicted probability of residual symptoms on each specific outcome parameter is given on the x-axis, and the observed probability of residual symptoms is given on the y-axis. The dashed diagonal line represents perfect agreement between the predicted and actual probability of nonresponse. Tenths of the predicted risk are presented as reds dots, and augmented by a smoothed (loess) line over the entire predicted probability range (continuous line). In gray above the X-axis the distribution of the predicted probabilities.*

Supplementary file 1 – patients with PROM data vs patient without PROM data registered in LROI.			
	Cohort with PROMs available (n=7071)	Cohort with no PROMs available (n=31022)	p-value*
Sex female, n (%)	4435 (62.7)	212516(64.2)	p=0.015
Age	68.4 (8.5)	68.8 (9.1)	p<0.001
ASA classification, n (%) I II III-IV	 962 (13.6) 4886 (69.1) 1223 (17.3)	 4219 (13.6) 21591 (69.6) 5212 (16.8)	p=0.615
Smoking, yes n (%)	587 (8.3)	2947 (9.5)	p=0.002
Charnley score, n (%) A B1 B2 C	 3161 (44.7) 2305 (32.6) 1393 (19.7) 212 (3.0)	 13246 (42.7) 10547 (34.0) 6356 (21.1) 838 (2.7)	p<0.001
BMI, kg/m^2	29.6 (4.8)	29.8 (5.1)	p=0.001
Side affected, right n (%)	3733 (52.8)	16256 (52.4)	p=0.809
Previous surgery on affected joint, yes n (%)	2376 (33.6)	10389 (33.5)	p=0.562
Independent-Samples T-test or Chi-Square depending on the type of variable.			