Prediction of 90-day mortality after Total Hip Arthroplasty: a simplified and externally validated model based on observational registry data from Sweden, England and Wales

Anne Garland, PhD\*1,2,3, Erik Bulow, MSc\*2,4, Erik Lenguerrand, MSc5, Prof Ashley Blom, PhD5,6, Prof J. Mark Wilkinson, PhD7, Adrian Sayers, MSc5, Prof Ola Rolfson, PhD2,4, Prof Nils P. Hailer, MD1,2

2020-05-13

\* Authors contributed equally.

## Affiliations

1. Department of Orthopaedics, Institute of Surgical Sciences, Uppsala University Hospital, Uppsala, Sweden
2. The Swedish Hip Arthroplasty Register, Gothenburg, Sweden
3. Department of Orthopaedics, Visby Hospital, Visby, Sweden
4. Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
5. Translational Health Sciences, Bristol Medical School, University of Bristol, UK.
6. The National Institute of Health Research Biomedical Research Centre Bristol
7. Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK.

## Addresses

* AG: Andersrogatan 5, SE-621 45 Visby, SWEDEN
* EB/OR: Svenska Hoftprotesregistret, Registercentrum Vastra Gotaland, SE-413 45 Gothenbutg, SWEDEN
* EL/AB/AS: Musculoskeletal Research Unit, University of Bristol, Beacon House, Queens Road, Bristol, BS8 1QU, UK
* JMW: Department of Oncology and Metabolism, The Medical School, Beech Hill Road, Sheffield, S10 2RX, UK
* NPH: Akademiska sjukhuset, 751 85 Uppsala, SWEDEN

## Correspondence

Erik Bulow. + 46 70 08 234 28. [erik.bulow@vgregion.se](mailto:erik.bulow@vgregion.se). Postal: Svenska Hoftprotesregistret, Registercentrum Vastra Gotaland, SE-413 45 Gothenburg, SWEDEN

##### PAGE BREAK

# Abstract

**Background:**  Shared individualized decision making prior to total hip arthroplasty (THA) includes discussions of the risk of mortality. However, diagnosis code-based instruments such as the Charlson co-morbidity index are impossible to use in clinical practice, and the American Society for Anesthesiologists (ASA) class is imprecise. We thus developed a simple model to predict 90-day mortality after elective THA.

**Methods:** We performed a prospective multi-national population-based registry study. We included 53,099 patients with cemented THA due to osteoarthritis from the Swedish Hip Arthroplasty Register for model derivation and internal validation, as well as 125,428 patients from the National Joint Register from England and Wales for external model validation. A model was developed using a bootstrap ranking procedure with a least absolute shrinkage and selection operator (LASSO) logistic regression model combined with piecewise linear regression. Discriminative ability was evaluated by the area under the receiver operating characteristics curve (AUC). Calibration belt plots were used for model calibration.

**Findings:**  We propose a parsimonious main effects model combining age, sex, American Society for Anesthesiologists (ASA) class, the presence of cancer, diseases of the central nervous system, kidney disease, and obesity. The model had good discrimination, both internally (AUC = 0·78, 95 % CI 0·75 - 0·81) and externally (AUC = 0·75, 95 % CI 0·73 - 0·76). This model was superior to traditional models based on the Charlson (AUC = 0·66, 95 % CI 0·62 - 0·70) and Elixhauser (AUC = 0·64, 95 % CI 0·59 - 0·68) co-morbidity indices. The model was well calibrated for predicted probabilities up to 5%.

**Interpretation:**  We developed a parsimonious model to predict 90-day mortality after elective cemented THA based on easily available demographic data and the presence of four defined co-morbidities. This may facilitate individualized risk assessment prior to one of the most common surgical interventions. We published a web-calculator to aid clinical decision-making (<https://shpr.registercentrum.se/om-registret-1/forskning/prediktion-av-90-dagarsmortalitet/p/SkyeTsTFB>).

**Funding:**  We had no external funding.

##### PAGE BREAK

# Research in context

## Evidence before this study

We performed an initial ad hoc literature review based on Google Scholar, personal recommendations and relevant references. We then used 2Dsearch (<https://2dsearch.com> accessed 2020-05-11) to build a search query for a more systematic review: (femur OR hip OR bone OR bones OR osteoarthritis OR rheumatoid OR rheumatism) AND (prognostic OR prediction OR forecast OR forecasting OR forecasts OR predict OR predicted OR predicting OR predictions OR predicts OR projections) AND (death OR mortality OR deaths OR survival OR survive) AND (replacement OR replaced OR replacements OR arthroplasty OR "Arthroplasty, Replacement" OR prosthesis OR implant OR prostheses OR prosthetic OR prosthetics) AND (calibration OR calibrate OR calibrated OR validation OR validate OR validated OR validating OR validity OR precision OR accuracy OR accurate OR precise OR "external validation" OR evaluation OR verification OR internal) AND (build OR building OR construct OR create OR develop OR establish OR train OR derive OR derivation) AND (models OR modelling OR "statistical model" OR modeling) AND (comorbidity OR comorbid OR co-morbid OR co-occurring OR diagnoses OR co-morbidity OR Elixhauser OR Charlson OR "comorbidity index" OR "comorbidity score" OR "Rx Risk V" OR ASA). We used no time or language constraints and used this query in PubMed (<https://pubmed.gov> accessed 2020-05-11), and identified 143 articles, whereof one was a duplicate. We performed an article screening with Rayyan (<https://rayyan.qcri.org/> accessed 2020-05-12). 131 articles were excluded based on titles, and an additional 6 after reading the abstracts. 5 articles remained, whereof 3 were already considered during the ad hoc review process. Research on cadavers, canines, and patients with diagnoses other than elective hip disorders (including hip fractures) were excluded during the screening process, as were prediction models with outcomes other than mortality. In addition, we used Open Grey (<https://opengrey.eu> accessed 2020-05-12) with the query above, as well as a simpler query (hip AND prediction). We found 9 items, whereof none was considered relevant. The same queries were used with Epistemonikos (<https://epistemonikos.org> accessed 2020-05-12). No items were found using the longer query, but 33 systematic reviews based on the shorter. 30 where excluded after title screening, and the remaining 3 after screening of abstracts.

In summary, several prediction models have been developed for short-term mortality in patients operated with total hip arthroplasty (THA). Most of them are based on small sample sizes, make use of too elaborate predictors based on data that are unavailable in a clinical setting, only consider in-hospital deaths, are inadequately internally or externally validated or not validated at all, or have no national population coverage. Only few model descriptions are accompanied by a practical tool for use in clinical practice (often only nomograms or a mathematical formula are supplied). Some relevant models are further described in the main text. Meta-analysis is most relevant for pooling individual effect estimates or measures of predictive power for a model validated on different cohorts. Hence, no relevant meta-analysis was applicable to our study.

## Added value of this study

Our study is the first to develop and externally validate a parsimonious prediction model for early mortality after THA based on two European cohorts, and to simultaneously provide a simple web-calculator for use in clinical practice.

## Implications of all the available evidence

Prediction models must consider all aspects of the relevant cohorts, including different health care systems, population demographics, and treatment traditions. Our tool was derived on a Swedish cohort and then validated on a similar cohort from England and Wales. We believe that the model might have broader clinical relevance and is suitable for clinical usage.

##### PAGE BREAK

# Introduction

Shared decision making has evolved into an integral part of patient-physician interactions prior to surgical interventions, and the weighing of risks against benefits is central to this process. In terms of gained quality of life and cost-utility, total hip arthroplasty (THA) is an enormously successful procedure. 90-day mortality after THA performed due to osteoarthritis is low, ranging from 0·2 % to 0·6 % in large joint registry studies.1 The risk of short-term mortality is nevertheless an essential part of preoperative discussions between patients and surgeons.

Co-morbidity is associated with a shorter remaining life span, but the Charlson and Elixhauser co-morbidity indices poorly predict mortality after THA.1,2,2–5 Additionally, these complex co-morbidity instruments are based on information from extensive in- or out-patient databases, each measure being defined by 1,178 and 1,516 individual International Classification of Diseases (ICD)-10 codes, respectively. Each co-morbidity index also exists in numerous versions, making interpretation and comparisons difficult.6 The use of these indices is therefore limited to research settings but they are impossible to use under clinical circumstances.

Pre-operative co-morbidity data have been used to develop clinically applicable prediction models for early mortality and adverse events such as surgical site infection in the context of THA surgery.7–13 The first prediction model based on a European cohort was the patient decision support tool for joint replacement in the UK (<https://jointcalc.shef.ac.uk/>). However, it remains unclear to what level the depth of co-morbidity characterization influences the accuracy of mortality prediction models.

Here we aimed to develop a parsimonious prediction model of 90-day postoperative mortality after THA with internal and external validation of discrimination and calibration, and to compare this to the accuracy of prediction models based on existing co-morbidity measures.14

# Participants and Methods

We used data from the Swedish Hip Arthroplasty Register (SHAR) for model derivation and internal validation of discrimination and calibration. The best model was then validated externally on patients from England and Wales recorded in the National Joint Registry for England, Wales, Northern Ireland, the Isle of Man and the States of Guernsey (NJR). We focused on patients with cemented THA due to osteoarthritis. Cementation is the most common fixation technique used in Sweden, thus used as inclusion criteria to decrease heterogeneity. Only the last operated hip was accounted for in patients with bilateral THA, since those are more likely to resemble patients with unilateral THA (Figure 1).15

## Derivation cohort (Sweden)

The patient cohort was identified from the SHAR. The inclusion period started 2008 since the American Society for Anesthesiologists (ASA) class and Body Mass Index (BMI) were systematically recorded from then on. The observation period ended 2015, since we had access to co-morbidity from the national patient register up to this point. Deterministic data linkage was performed by 10-digit identity numbers, assigned to all Swedish residents at birth or immigration. We used data from a variety of sources, as previously described.16 Age, sex, BMI, ASA class, type of hospital (university/county/rural/private) and year of surgery were collected from the SHAR, with a completeness of 96-98 %.17 Data on educational level (low = up to 9 years/middle = 10-12 years/high = at least 12 years) was recorded for more than 98 % of the population with 85 % accuracy in the longitudinal integration database for health insurance and labor market studies (LISA) from Statistics Sweden.18 Civil status (married/un-married/divorced/widow[er]), were also collected from LISA. The Swedish national patient register (NPR) was used to assess co-morbidity during the year preceding index surgery. This register contains all relevant diagnoses coded by ICD-10, as well as dates of admission and discharge for in- and outpatient episodes in all private and public hospitals. Completeness for NPR is above 99 % and 85-95 % of all diagnostic codes are valid.19 Death dates were linked from the national population register.

Patients with missing information on BMI or a measurement above 50 were excluded, as were patients with missing information on ASA class or class IV and above, as well as patients with unknown educational level or type of hospital.

## External validation cohort (England and Wales)

Similar inclusion criteria were applied to the validation cohort from England and Wales. We excluded no patients based on missing information on BMI, educational level, or civil status however, since those variables were not used in the final model. The validation cohort was created by linking NJR data to the secondary uses service database from the National Health Service (NHS). This database includes co-morbidity recorded in the admitted patient care module of the hospital episode statistics (HES) database in England, and the patient episode database for Wales.20 Out-patient deaths were linked from the office of national statistics. Data linkage used NHS-numbers, as well as patient name, age, sex and address.21

## Defining co-morbidity

Co-morbidity was defined by individual ICD-10 codes grouped into 17 categories according to Charlson and 31 categories according to Elixhauser.22 Some co-morbidities were identified in both indices, and some distinct co-morbidities were closely related (such as hypertension with and without complications, or abuse of either drugs or alcohol). We combined individual diagnostic groups to establish 21 broader categories of co-morbidity (Table 1). Conditions were merged according to clinical relevance as to be recognized in a patient-physician meeting without access to external register data.

## Statistical analysis for model development

We used the Kaplan-Meier estimator to assess unadjusted mortality after cemented THA. Further analysis was based on logistic regression, since no loss to follow up occurred in the Swedish cohort within the first 90 days, and we assumed the same was true for patients from England and Wales. We used a modelling procedure with bootstrap ranking and a logistic least absolute shrinkage and selection operator (LASSO).23 Numeric variables (age and BMI) were normalized before modelling to have mean = 0 and standard deviation = 1. Thus, the magnitude (absolute values) of their estimated coefficients would indicate variable importance on the same scale as the categorical variables. Co-morbidities recorded for at least one patient who died within 90 days, and one who did not, were included in the modelling process. 1,000 bootstrap samples were drawn from the observed data set. We used 10-fold cross validation for every bootstrap sample with a broad range of potential penalty values (:s) in a logistic LASSO model. We then only kept :s minimizing the mean cross-validated deviances in each sample. Those :s were used to estimate model coefficients for each potential predictor. The magnitude (absolute values) of those estimates were used as a measure of variable importance. Piece-wise linear regression was used to detect a breaking point where a significant drop in variable importance was observed. Potential predictors with variable importance above this breaking point were considered important and kept as model candidates. The whole process was repeated 100 times. Covariates that were selected at least once out of the 100 times were used in a main effects model of multivariable logistic regression without penalty, and without pre-normalization of numeric variables (main model). A reduced model with variables chosen at least 33 out of the 100 times was used as a simpler alternative for comparison. This model, without cancer as a predictor, was also evaluated separately, considering that medical indications for THA surgery may be different for patients with cancer compared to patients without malignancies. Univariable models with the ASA class, Charlson or Elixhauser co-morbidity indices were used for benchmarking, as well as a multivariable model including only age and sex. Each model including age was fitted three times, once with age as a main effect and twice with restricted cubic splines, either by two or three knots. Odds ratios for the final model were estimated with 95 % confidence intervals.

## Statistical analysis for model validation

Each model was used to predict the probability of death within 90 days for patients from the SHAR (internal validation). Sensitivity and specificity were estimated to form receiver operating characteristic (ROC) curves and the area under those curves (AUC) were used as a measure of discriminative ability. Models with a lower 95 % confidence limit above 0·7, were considered good. Those intervals were based on percentiles from 2,000 non-parametric bootstrap samples. We used the bias-corrected Somers’ rank correlation based on 100 resamples to adjust for optimism.24 Calibration bands were made to graphically assess model calibration, comparing predicted probabilities and observed proportions.25 The reduced model was then evaluated externally. An AUC with 95 % confidence interval was calculated for the model as-is. Re-calibration of the model intercept was then performed to account for different mortality rates in Sweden compared to England and Wales. An updated over-all slope was also calculated to account for country-specific treatment differences.26 Calibration for this re-calibrated model was illustrated in the same calibration belt plot as for the internal calibration.

## Statistical tools

We built an online web calculator available at <https://shpr.registercentrum.se/om-registret-1/forskning/prediktion-av-90-dagarsmortalitet/p/SkyeTsTFB> to be used in clinical practice.

We used R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) with significant packages tidyverse, tidymodels, furrr, pROC and shiny. All R-scripts and necessary configurations (but no personal data) is available at <http://doi.org/10.5281/zenodo.3732852>. A linked Binder environment is also available for interactive usage.

## Ethical approval

Ethical approval for this study was obtained from the Regional Ethical Review Board in Gothenburg (360-13) and by the NJR in the UK (RSC2017/21). Informed consent was not mandatory according to the Swedish patient data law from 2009, and the UK law for pseudonymised data.

# Indirect Patient and Public Involvement

There is a patient representative at SHAR to ensure that the patients’ interests are respected in planned research.

# Results

## Study participants

53,099 patients (age 35 - 99, 61 % females) were included in the derivation cohort from SHAR (Figure 1 left panel). 175 (0·33 %) of those patients died within 90 days and no one was censored before that. The unadjusted risk of 90-day mortality was therefore 0·33 % (95 % CI 0·28 - 0·38). Further characteristics of the study population are presented in Table 2. 26 % and 48 % of the Swedish patients had at least one pre-surgery co-morbidity according to Charlson and Elixhauser. The proportion of patients with ASA class III was 18 %.

In addition, 125,428 patients were included for the external validation cohort from England and Wales (Figure 1 right panel). Their unadjusted risk of 90 day mortality was 0·52 % (95 % CI 0·49 - 0·56), which was higher compared to the derivation cohort from Sweden.

## Model development and internal validation

There were five co-morbidities that were not recorded for any patient who died in the Swedish derivation cohort: acquired immunodeficiency syndrome by the human immunodeficiency virus (AIDS/HIV), coagulopathy, fluid electrolyte disorders, liver disease, and weight loss. Those variables were therefore excluded as potential predictors prior to any statistical model derivation. The derived main model included age, sex, ASA class, the presence of cancer, diseases of the central nervous system, kidney disease, diagnosed obesity, heart condition, anemia, and myocardial infarction. The reduced model, with covariates included at least 33 out of 100 times, was restricted to age, sex, ASA class, the presence of cancer, diseases of the central nervous system, kidney disease, and diagnosed obesity (Table 3).

There were no differences between models including age as a main effect, compared to those were age was modeled by restricted cubic splines with either two or three knots. We therefore focused on the more parsimonious models with age as a main effect. Similarly, the correction for optimism only affected the third decimals of the AUC confidence intervals and was therefore omitted.

The main and reduced models were no different regarding discriminative power (AUC = 0·79, 95 % CI 0·75 - 0·82 versus AUC = 0·78, 95 % CI 0·75 - 0·81). We therefore considered the reduced model as superior due to its simplicity. Traditional models performed poorly with 95 % confidence intervals not above 0.7: The Charlson co-morbitiy index had an AUC of (AUC = 0·66, 95 % CI 0·62 - 0·70) and the Elixhauser co-morbidity an AUC of (AUC = 0·64, 95 % CI 0·59 - 0·68; Figure 2 left panel and Figure 3).

The ability of the reduced model to estimate probabilities of death within 90 days is further illustrated in Figure 4. Model calibration was good for estimated probabilities up to 3 % and acceptable up to 5 %, although with predicted probabilities usually higher than observed proportions (Figure 5).

Estimated model coefficients and corresponding odds ratios for the reduced model are presented in Table 4.

Omitting cancer from the reduced model did not affect the AUC or calibration for estimated probabilities below 3 %, but calibration outside this range deteriorated, and we thus retained cancer as an important predictor.

## External validation

The discriminative ability for the reduced model was not statistically significantly different when applied to the external validation cohort (AUC = 0·75, 95 % CI 0·73 - 0·76) compared to the internal validation (AUC = 0·78, 95 % CI 0·75 - 0·81) (Figure 2 right panel and Figure 3). Calibration of the re-calibrated model was slightly inferior compared to the internal calibration. Predicted probabilities between 0·5 % and 1·5 % were lower than observed proportions and the estimated 95 % confidence bands were wider. Over-all, calibration was good for predicted probabilities below 3 % and acceptable below 5 % (Figure 5).

# Discussion

## Principal findings

We found that a multivariable main effects logistic regression model with age, sex, ASA class, the presence of cancer, diseases of the central nervous system, kidney disease, and diagnosed obesity was considerably better at predicting death within 90 days after THA than traditional models based on the Charlson or Elixhauser co-morbidity indices.

The resulting model predicted the probability of death within 90 days as: where is the vector of estimated coefficients presented in table 4. This formula is considered valid for patients aged 35 - 99 years and for predicted probabilities up to 5 %.

## Clinical usage

Our model could be used in clinical practice, either by the formula above, or by accessing a simple web calculator online (<https://shpr.registercentrum.se/om-registret-1/forskning/prediktion-av-90-dagarsmortalitet/p/SkyeTsTFB>). For example, a 35-year-old woman with ASA class I and none of the important co-morbidities would have a 0·028 ‰ risk to die within 90 days of surgery. A woman, 67 years old, i.e. the first age quantile, would have an elevated risk of 0·37 ‰. A 99-year-old man, i.e. the maximum observed age, with ASA class III and cancer, would have a risk of 8·8 %. Note, however, that covariate patters with observed probabilities above 5 % were rare (0·13 %, n = 70), and estimated mortality risks above 5 % are therefore subject to extrapolation. The observed proportions of deaths for patients with similar characteristics is likely lower (as indicated by Figure 5). Some risk calculators ignore this problem,7,13 but we think this should be acknowledged.

Variables in our model were chosen based on statistical properties and not due to clinical relevance. Variables and estimated coefficients should therefore not be assigned any exact epidemiological and/or causal meaning per se.

Age and sex are well-known predictors of remaining life span, but it is less obvious that this relation must be linear. We used restricted cubic splines to allow a more flexible relation, but found no difference in AUC compared to simpler main effect models. ASA class III has the largest estimated coefficient among all predictors, indicating large relative importance. This is clinically reasonable since a label of “severe systemic disease” should be based on a relevant patient assessment prior to surgery. ASA class is nevertheless known to have a high degree of internal variability.27 It has been compared to the Charlson co-morbidity index, but not with respect to mortality after THA. Patients with ASA class IV-VI were excluded since those categories describe severe disease, moribund and brain-dead individuals. We suspect that patients with such conditions might have been misclassified.

Obesity was not statistically significant by itself ( 0·05) but was still relevant as a predictor due to unobserved heterogeneity. We noted that the proportion of patients registered in the SHAR with BMI above 30, the WHO definition of obesity, was much higher (23 %) than the proportion of patients with a diagnosis code for obesity (ICD-10 = E66) recorded in the Swedish national patient register (1·9 %). We believe this might be due to the fact that BMI is systematically recorded in the SHAR, but the diagnosis of obesity might not have been made during previous contacts with the healthcare system.

## Strengths and limitations

A strength of this study is its nationwide design with data from both Sweden, England and Wales. The Swedish registers are valid with low proportions of missing data.17,19 Some concerns have been raised regarding validity of the hospital episodes statistics (HES) database from England and Wales however. A systematic review found that the overall median diagnostic accuracy (comparing ICD-codes from HES to individual case notes) was 80 %.28 Both in- and outpatient co-morbidity data were available for Sweden, but only in-patient data for England and Wales. We did not have access to any data from primary care or nursing homes in any of the counties.

The linkage procedure between NJR and HES has been previously described,29 where it was noted that privately funded patients were not included in HES. 17 % of the patients had missing personal data or did not allow linkage, and 6 % were not found in HES although their data were available from the NJR.

In addition to data validity, the indications for choosing between different fixation techniques differ in Sweden compared to England and Wales. Cementation is the most commonly used fixation principle in Sweden but not so in England and Wales, where younger patients are more likely to be operated with cementless implants. To only include cemented THA therefore implies an over-representation of older and frailer patients from England and Wales compared to Sweden. Those patients might be more likely to die within 90 days, and this assumption would be in agreement with the calibration plot where estimated probabilities below 1·5 % underestimated the observed proportion of deaths in the external validation cohort from the NJR.

We investigated whether the effect of age on mortality was non-linear using restricted cubic splines. A similar approach might be relevant for BMI, which we only modelled as a linear main effect. A U-shaped association between mortality and BMI categorized as underweight (below 18·5 kg/m2), normal (8·5 to 25 kg/m2), or obese (above 25 kg/m2) has been found for example in England and Wales.30 Our recordings of BMI might be imprecise however. Some hospitals report actual measurements but it is unknown to what extent those data are self-estimated by the patients or by health care personnel. Therefore, we aimed to avoid overfitting BMI based on too elaborate transformations.

It should be noted that the risk model does not study THA as an observed intervention. We merely followed the cohort who already had THA. Hence, deaths within 90 days might occur for those patients regardless if THA is inserted or not. The proximity in time however, the maximum of 90 days from THA to death, is an indication that the operation might be the main cause of death for the majority of deceased patients.

## Conclusion

Our results indicate that the risk of early postoperative mortality after cemented THA can be pre-operatively assessed by a parsimonious prediction model. We hope that this model, with its accompanying web calculator, will facilitate shared decision-making between physicians and their patients in need of THA.

##### PAGE BREAK

# Contribution of authors

AG and NPH initiated the study and managed the ethical review board application in Sweden. AB and JMW conceptualized the external validation and managed the ethical review board application in the UK. EB developed the statistical model. EL and AS performed external validation with data from NJR. AG and EB drafted, and all authors edited and finalized the manuscript.

# Acknowledgement

We thank the patients and staff of all the hospitals in England, Wales and Northern Ireland who have contributed data to the National Joint Registry. We are grateful to the Healthcare Quality Improvement Partnership (HQIP), the NJR Research Committee and staff at the NJR Centre for facilitating this work. We also thank all contributors to the Swedish Hip Arthroplasty Register (SHAR), especially Szilard Nemes, previous senior statistician, for his involvement in the study. The authors have conformed to the NJR’s standard protocol for data access and publication. The views expressed represent those of the authors and do not necessarily reflect those of the National Joint Registry Steering Committee or the Healthcare Quality Improvement Partnership (HQIP) who do not vouch for how the information is presented.

##### PAGE BREAK

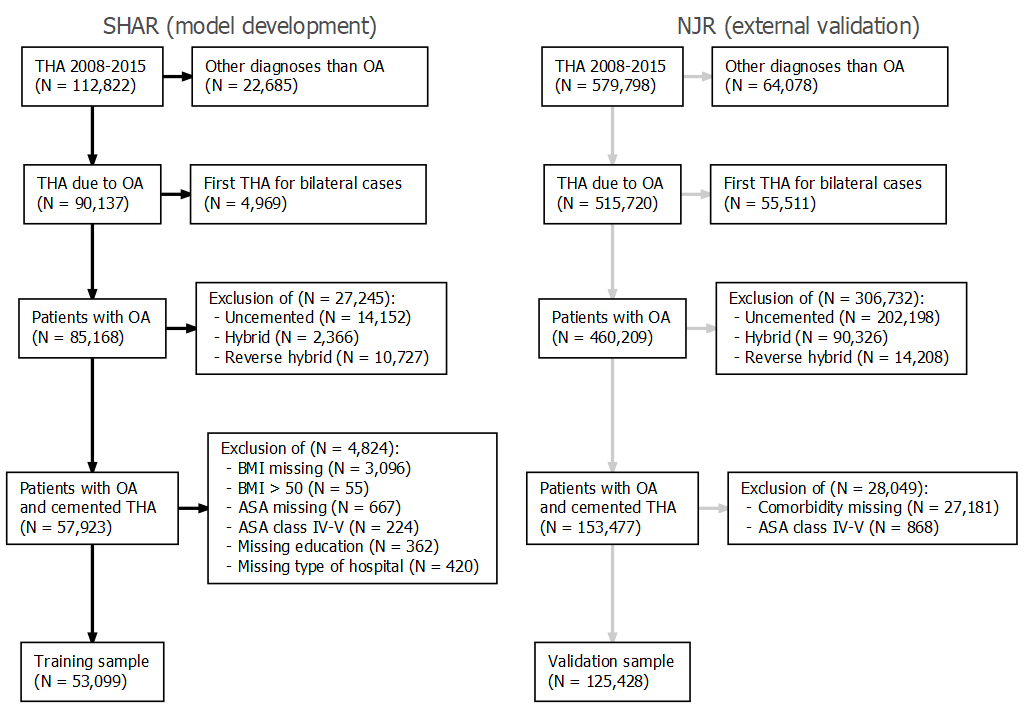


Figure 1: Flowchart depicting inclusion criteria and number of patients. Data from the Swedish Hip Arthroplasty Register were used for model derivation and internal validation (left). Data from the National Joint Registry in England and Wales were used for external validation (right).

##### PAGE BREAK

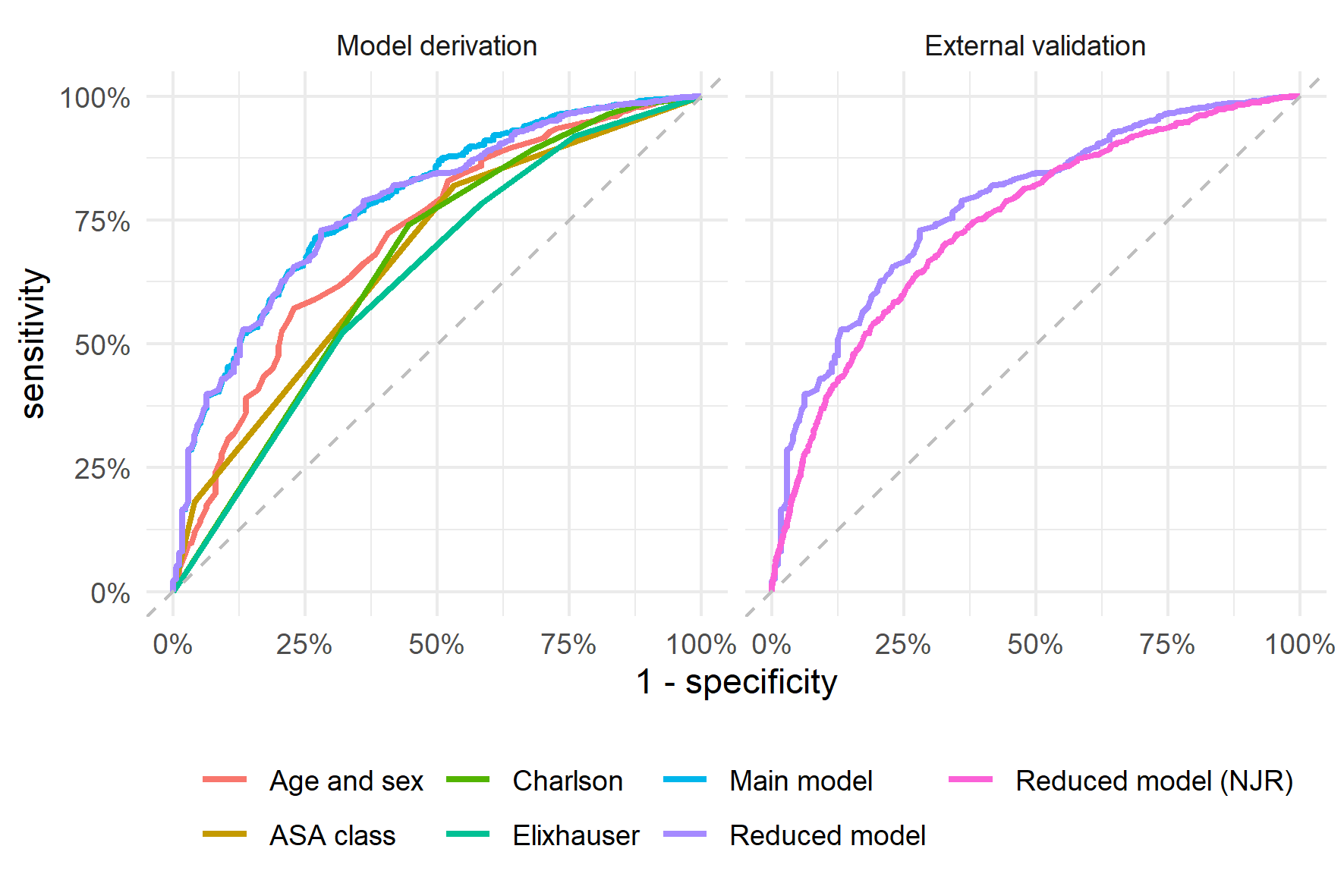


Figure 2: Receiver Operation Characteristics (ROC) curves. The area under the ROC curve (AUC) for the reduced model derived on data from the SHAR was very similar to the main model (left panel). External validation of this model, on data from the National Joint Registry for England and Wales, produced a similar curve as well (right panel).

##### PAGE BREAK

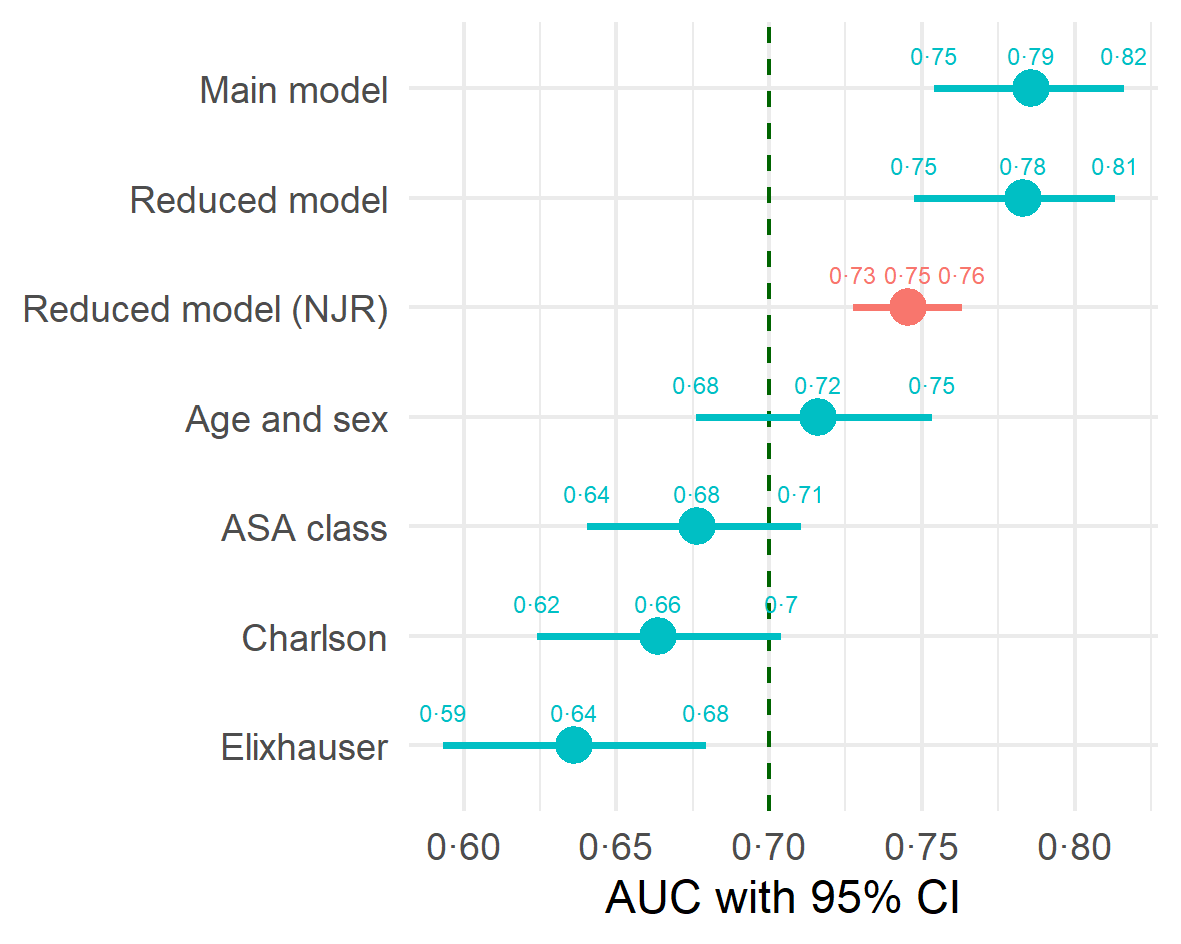


Figure 3: Area Under the Receiver Operation Characteristics Curve (AUC) as a measure of predictive discriminative ability with 95 % bootstrap confidence intervals. AUC above 0·7 were considered good. The reduced model was similar to the main model, and not statistically significantly inferior when used with external data from the NJR.

##### PAGE BREAK

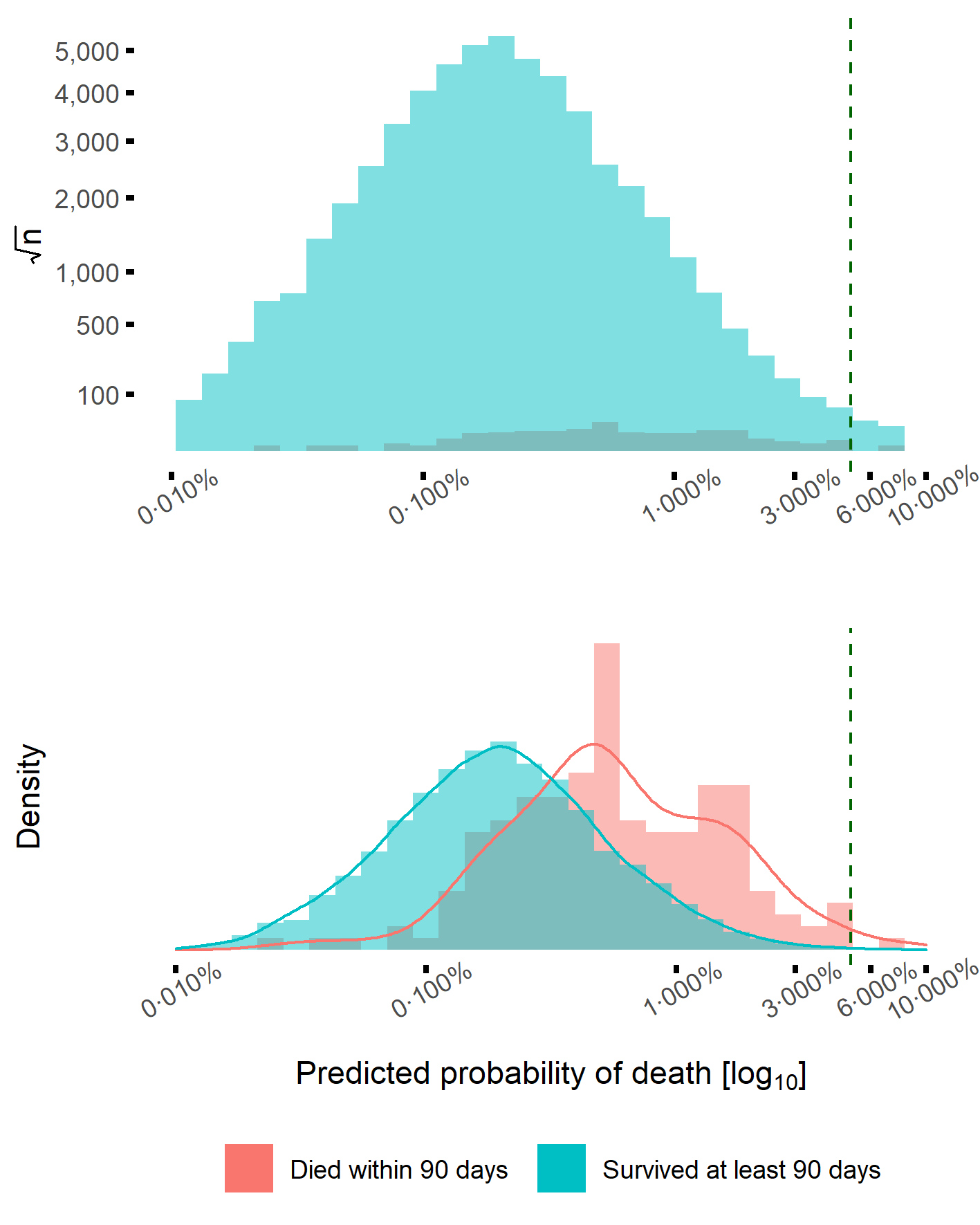


Figure 4: The vast majority of patients survived more than 90 days after THA, and therefore the blue bars dominate the histogram (upper panel; note the scales). A normalized density plot however reveals that patients who died within 90 days were, on average, estimated to have a higher probability to do so (lower panel). There were very few observations with covariate patterns resulting in death probabilities higher than 5 % (green line). Estimated probabilities above this limit are therefore subject to extrapolation.

##### PAGE BREAK

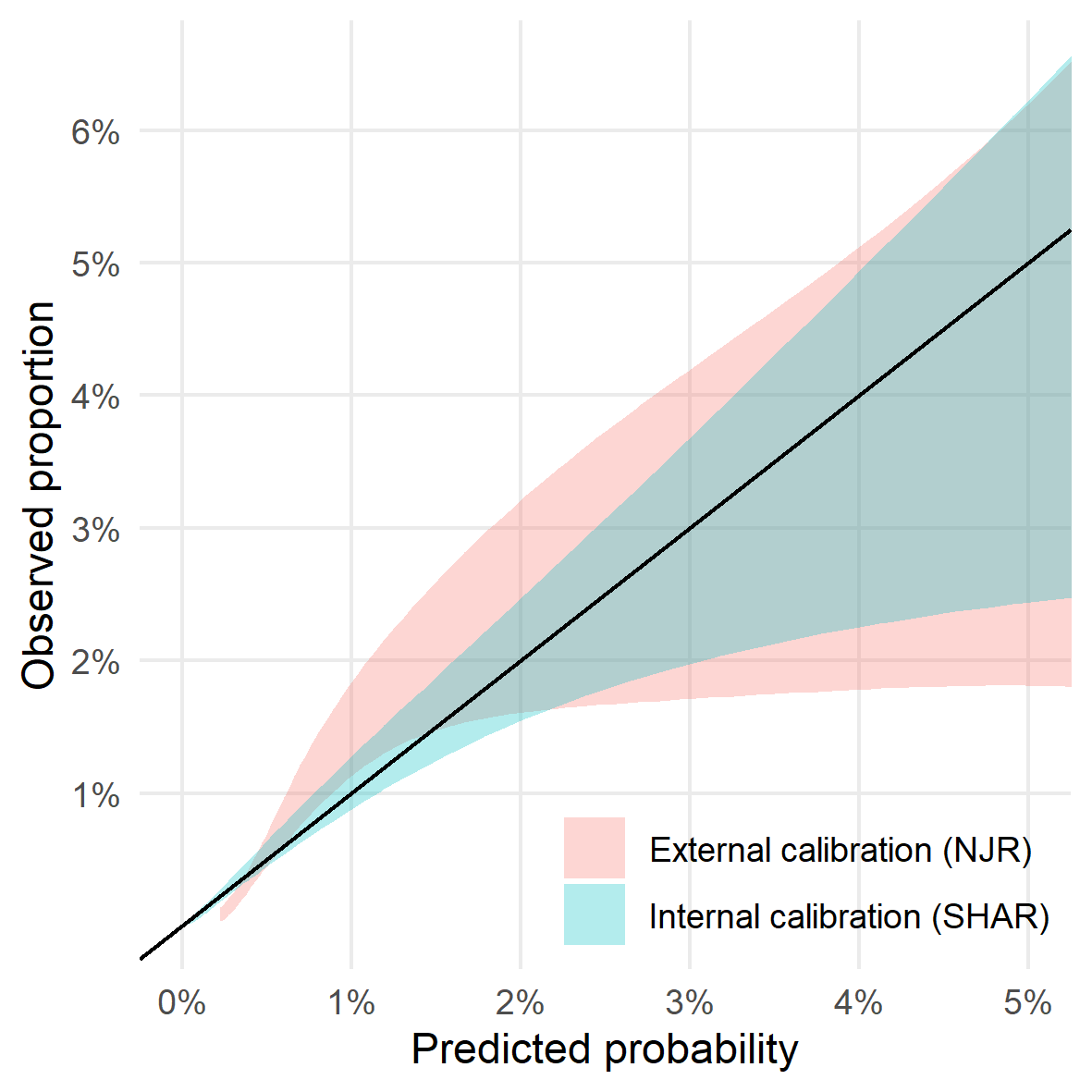


Figure 5: This figure illustrates calibration between observed proportions and predicted probabilities with 95 % confidence intervals. Deviations above the diagonal line indicated that the model under-estimated the probability of death within 90 days. Such deviation was statistically significant for the external validation for predictions below 1·5 %. In contrast, with higher predicted probabilities, the model over-estimated the observed proportions of deaths.

##### PAGE BREAK

Table 1: Categorization of co-morbidities based on diagnostic groups according to Charlson and Elixhauser. (CNS = central nervous system.)

|  |  |  |
| --- | --- | --- |
| Comorbidities by groups | Charlson | Elixhauser |
| AIDS/HIV | AIDS/HIV | AIDS/HIV |
| Anemia |  | Blood loss anemia, Deficiency anemia |
| Arrhythmia |  | Cardiac arrhythmias |
| Arterial hypertension |  | Hypertension uncomplicated, Hypertension complicated |
| Cancer | Malignancy, Metastatic solid tumor | Lymphoma, Metastatic cancer, Solid tumor |
| CNS disease | Dementia, Hemiplegia or paraplegia | Depression, Paralysis, Other neurological disorders, Psychoses |
| Coagulopathy |  | Coagulopathy |
| Diabetes | Diabetes without complication, Diabetes complication | Diabetes uncomplicated, Diabetes complicated |
| Drug alcohol abuse |  | Alcohol abuse, Drug abuse |
| Fluid electrolyte disorders |  | Fluid electrolyte disorders |
| Heart condition | Congestive heart failure | Congestive heart failure, Valvular disease |
| Myocardial infarction | Myocardial infarction |  |
| Hypothyroidism |  | Hypothyroidism |
| Kidney disease | Renal disease | Renal failure |
| Liver disease | Mild liver disease, Moderate or severe liver disease | Liver disease |
| Lung airways disease | Chronic pulmonary disease | Chronic pulmonary disease, Pulmonary circulation disorder |
| Obesity |  | Obesity |
| Peptic ulcer | Peptic ulcer disease | Peptic ulcer disease |
| Rheumatic disease | Rheumatic disease | Rheumatoid arthritis |
| Vascular disease | Peripheral vascular disease, Cerebrovascular disease | Peripheral vascular disorder |
| Weight loss |  | Weight loss |

##### PAGE BREAK

Table 2: Characteristics of the study population in the model derivation cohort (SHAR) and external validation cohort (NJR). Educational levels were classified as low (up to 9 years), middle (10-12 years) and high (at least 12 years). BMI = Body mass index. ASA = American Society for Anesthesiologists classification. CNS = central nervous system.

|  |  |  |  |
| --- | --- | --- | --- |
| what | level | SHAR | NJR |
| n |  | 53,099 | 125,428 |
| Age (mean (SD)) |  | 72·68 (7·76) | 73·51 (8·67) |
| Sex = Female (%) |  | 32,440 (61·1) | 82,247 (65·6) |
| BMI (mean (SD)) |  | 27·19 (4·39) |  |
| ASA class (%) |  |  |  |
|  | I | 9,589 (18·1) | 10,890 (8·7) |
|  | II | 33,881 (63·8) | 89,388 (71·3) |
|  | III | 9,629 (18·1) | 25,150 (20·05) |
| Hospital (%) |  |  |  |
|  | University | 24,460 (46·1) |  |
|  | County | 16,507 (31·1) |  |
|  | Rural | 9,940 (18·7) |  |
|  | Private | 2,192 ( 4·1) |  |
| Education (%) |  |  |  |
|  | Low | 11,645 (21·9) |  |
|  | Middle | 20,604 (38·8) |  |
|  | High | 20,850 (39·3) |  |
| Civil status (%) |  |  |  |
|  | Married | 29,436 (55·4) |  |
|  | Single | 12,888 (24·3) |  |
|  | Widow/widower | 10,775 (20·3) |  |
| Charlson (%) |  |  |  |
|  | 0 | 39,256 (73·9) | 80,236 (64·0) |
|  | 1 | 8,117 (15·3) | 33,447 (26·7) |
|  | 2 | 3,762 ( 7·1) | 9,049 (7·2) |
|  | 3 | 1,176 ( 2·2) | 2,116 (1·7) |
|  | 4+ | 788 ( 1·5) | 580 (0·5) |
| Elixhauser (%) |  |  |  |
|  | 0 | 27,773 (52·3) | 33,546 (26·8) |
|  | 1 | 13,766 (25·9) | 42,467 (33·9) |
|  | 2 | 7,239 (13·6) | 28,188 (22·5) |
|  | 3+ | 4,321 ( 8·1) | 21,227 (16·9) |
| AIDS/HIV (%) |  | 5 ( 0·0) | 0 |
| Anemia (%) |  | 416 ( 0·8) | 2,306 (1·8) |
| Arrhythmia (%) |  | 4,505 ( 8·5) | 13,040 (10·4) |
| Arterial hypertension (%) |  | 16,677 (31·4) | 66,837 (53·3) |
| Cancer (%) |  | 2,715 ( 5·1) | 3,055 (2·4) |
| CNS disease (%) |  | 1,682 ( 3·2) | 7,075 (5·6) |
| Coagulopathy (%) |  | 192 ( 0·4) | 583 (0·5) |
| Diabetes (%) |  | 4,077 ( 7·7) | 13,874 (11·1) |
| Drug alcohol abuse (%) |  | 223 ( 0·4) | 2,057 (1·6) |
| Fluid electrolyte disorders (%) |  | 304 ( 0·6) | 2,774 (2·2) |
| Heart condition (%) |  | 2,639 ( 5·0) | 5,746 (4·6) |
| Myocardial infarction (%) |  | 2,186 ( 4·1) | 3,774 (3·0) |
| Hypothyroidism (%) |  | 1,791 ( 3·4) | 10,453 (8·3) |
| Kidney disease (%) |  | 551 ( 1·0) | 7,370 (5·9) |
| Liver disease (%) |  | 207 ( 0·4) | 630 (0·5) |
| Lung airways disease (%) |  | 2,878 ( 5·4) | 18,337 (14·6) |
| Obesity (%) |  | 1,000 ( 1·9) | 10,380 (8·3) |
| Peptic ulcer (%) |  | 341 ( 0·6) | 685 (0·6) |
| Rheumatic disease (%) |  | 1,922 ( 3·6) | 5,956 (4·8) |
| Vascular disease (%) |  | 1,686 ( 3·2) | 3,656 (2·9) |
| Weight loss (%) |  | 35 ( 0·1) | 509 (0·4) |

##### PAGE BREAK

Table 3: Variables selected by the bootstrap ranking procedure, and therefore kept in the main model. Variables chosen at least 33 times out of 100 were kept in the reduced model as well. ASA class II and III were kept in both models, since those are usually distinguished anyway. To simplify the models further would be possible by lumping class I and II together. Variables chosen more frequently are likely more important predictors. This is not necessarily true however since one of several strongly correlated variables might be chosen spuriously. (CNS = central nervous system.)

|  |  |
| --- | --- |
| variable | n |
| Cancer | 100 |
| CNS disease | 100 |
| Kidney disease | 100 |
| ASA class III | 100 |
| Age | 50 |
| Obesity | 49 |
| Male sex | 38 |
| Heart condition | 20 |
| Anemia | 9 |
| ASA class II | 8 |
| Myocardial infarction | 6 |

##### PAGE BREAK

Table 4: Estimated coefficients (beta) and odds ratios (OR) with 95 % confidence intervals for the reduced model.

|  |  |  |  |
| --- | --- | --- | --- |
| term | beta | OR 95 % CI | p |
| (intercept) | -13·28 |  |  |
| Cancer | 0·71 | 2·04 (1·30-3·08) | <0·01 |
| CNS disease | 0·9 | 2·45 (1·35-4·11) | <0·01 |
| Kidney disease | 1·33 | 3·79 (2·04-6·52) | <0·01 |
| ASA class II | 0·88 | 2·42 (1·20-5·78) | 0·03 |
| ASA class III | 1·66 | 5·27 (2·56-12·77) | <0·01 |
| Age | 0·08 | 1·08 (1·06-1·11) | <0·01 |
| Obesity | 0·77 | 2·17 (0·91-4·36) | 0·05 |
| Male sex | 0·61 | 1·85 (1·36-2·52) | <0·01 |

##### PAGE BREAK

# Bibliography

1 Garland A, Gordon M, Garellick G, Kärrholm J, Sköldenberg O, Hailer NP. Risk of early mortality after cemented compared with cementless total hip arthroplasty. *The Bone & Joint Journal* 2017; **99-B**: 37–43.

2 Inacio MCS, Pratt NL, Roughead EE, Graves SE. Using medications for prediction of revision after total joint arthroplasty. *The Journal of arthroplasty* 2015; **30**: 2061–70.

3 Gordon M, Stark A, Sköldenberg OG, Kärrholm J, Garellick G. The influence of comorbidity scores on re-operations following primary total hip replacement: Comparison and validation of three comorbidity measures. *The bone & joint journal* 2013; **95-B**: 1184–91.

4 Hofstede SN, Gademan MGJ, Vlieland TPMV, Nelissen RGHH, de Mheen PJM-v. Preoperative predictors for outcomes after total hip replacement in patients with osteoarthritis: A systematic review. *BMC musculoskeletal disorders* 2016; **17**: 212.

5 Bülow E, Rolfson O, Cnudde P, Rogmark C, Garellick G, Nemes S. Comorbidity does not predict long-term mortality after total hip arthroplasty. *Acta Orthopaedica* 2017; **88**: 1–6.

6 Brusselaers N, Lagergren J. The charlson comorbidity index in registry-based research. *Methods of Information in Medicine* 2017; **56**: 401–6.

7 Bozic KJ, Ong K, Lau E *et al.* Estimating risk in medicare patients with THA: An electronic risk calculator for periprosthetic joint infection and mortality. *Clinical Orthopaedics and Related Research®* 2013; **471**: 574–83.

8 Bozic KJ. Orthopaedic healthcare worldwide: Shared medical decision making in orthopaedics. *Clinical Orthopaedics and Related Research* 2013; **471**: 1412–4.

9 Harris AHS, Kuo AC, Bozic KJ *et al.* American joint replacement registry risk calculator does not predict 90-day mortality in veterans undergoing total joint replacement. *Clinical Orthopaedics and Related Research* 2018; **476**: 1869–75.

10 Harris AH, Kuo AC, Bowe T, Gupta S, Nordin D, Giori NJ. Prediction models for 30-Day mortality and complications after total knee and hip arthroplasties for veteran health administration patients with osteoarthritis. *The Journal of Arthroplasty* 2018; **33**: 1539–45.

11 Inacio MCCS, Pratt NLL, Roughead EEE, Graves SEE. Evaluation of three co-morbidity measures to predict mortality in patients undergoing total joint arthroplasty. *Osteoarthritis and cartilage* 2016; **24**: 1718–26.

12 Price A, Smith J, Dakin H *et al.* The Arthroplasty Candidacy Help Engine tool to select candidates for hip and knee replacement surgery: Development and economic modelling. *Health Technology Assessment* 2019; **23**: 1–216.

13 Harris AHS, Kuo AC, Weng Y, Trickey AW, Bowe T, Giori NJ. Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? *Clinical Orthopaedics and Related Research* 2019; **477**: 452–60.

14 Manning DW, Edelstein AI, Alvi HM. Risk prediction tools for hip and knee arthroplasty. *The Journal of the American Academy of Orthopaedic Surgeons* 2016; **24**: 19–27.

15 Bülow E, Nemes S, Rolfson O. Are the first or the second hips of staged bilateral THAs more similar to unilateral procedures? A study from the swedish hip arthroplasty register. *Clinical Orthopaedics and Related Research* 2020;: 1.

16 Cnudde P, Rolfson O, Nemes S *et al.* Linking Swedish health data registers to establish a research database and a shared decision-making tool in hip replacement. *BMC Musculoskeletal Disorders* 2016; **17**: 414.

17 Kärrholm J, Mohaddes M, Odin D, Vinblad J, Rogmark C, Rolfson O. Svenska höftprotesregistret årsrapport 2017. 2018 <https://registercentrum.blob.core.windows.net/shpr/r/-rsrapport-2016-SJirXXUsb.pdf>.

18 Ludvigsson JF, Svedberg P, Olén O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *European Journal of Epidemiology* 2019; **34**: 423–37.

19 Ludvigsson JF, Andersson E, Ekbom A *et al.* External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011; **11**: 450.

20 Thorn JC, Turner E, Hounsome L *et al.* Validation of the hospital episode statistics outpatient dataset in england. *PharmacoEconomics* 2016; **34**: 161–8.

21 Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. *International Journal of Epidemiology* 2016; **45**: 954–64.

22 Quan H, Sundararajan V, Halfon P *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical care* 2005; **43**: 1130–9.

23 Guo P, Zeng F, Hu X *et al.* Improved variable selection algorithm using a LASSO-Type penalty, with an application to assessing hepatitis b infection relevant factors in community residents. *PLOS ONE* 2015; **10**: e0134151.

24 Miller ME, Hui SL, Tierney WM. Validation techniques for logistic regression models. *Statistics in Medicine* 1991; **10**: 1213–26.

25 Nattino G, Finazzi S, Bertolini G. A new test and graphical tool to assess the goodness of fit of logistic regression models. *Statistics in Medicine* 2016; **35**: 709–20.

26 Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: A study on sample size and shrinkage. *Statistics in Medicine* 2004; **23**: 2567–86.

27 Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995; **50**: 195–9.

28 Burns EM, Rigby E, Mamidanna R *et al.* Systematic review of discharge coding accuracy. *Journal of Public Health* 2012; **34**: 138–48.

29 Smith AJ, Dieppe P, Porter M, Blom AW. Risk of cancer in first seven years after metal-on-metal hip replacement compared with other bearings and general population: Linkage study between the National Joint Registry of England and Wales and hospital episode statistics. *BMJ (Online)* 2012; **344**. DOI:[10.1136/bmj.e2383](https://doi.org/10.1136/bmj.e2383).

30 Mouchti S, Whitehouse MR, Sayers A, Hunt LP, MacGregor A, Blom AW. The association of body mass index with risk of long-term revision and 90-Day mortality following primary total hip replacement. *The Journal of Bone and Joint Surgery* 2018; **100**: 2140–52.