European Journal of Cardio-Thoracic Surgery 2023, 64(3), ezad301
https://doi.org/10.1093/ejcts/ezad301 Advance Access publication 6 September 2023

ORIGINAL ARTICLE

GENERAL ADULT CARDIAC

Cite this article as: van Dijk WB, Leeuwenberg AM, Grobbee DE, Siregar S, Houterman S, Daeter EJ *et al.* Dynamics in cardiac surgery: trends in population characteristics and the performance of the EuroSCORE II over time. Eur J Cardiothorac Surg 2023; doi:10.1093/ejcts/ezad301.

Dynamics in cardiac surgery: trends in population characteristics and the performance of the EuroSCORE II over time

Wouter B. van Dijk [a], Artuur M. Leeuwenberg [a], Diederick E. Grobbee [a], Sabrina Siregar [b], Saskia Houterman [c], Edgar J. Daeter [c,d], Martine C. de Vries [e], Rolf H. H. Groenwold [f,g], Ewoud Schuit [a,*] and on behalf of the Cardiothoracic Surgery Registration Committee of the Netherlands Heart Registration[†]

[a] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
[b] Department of Cardiothoracic Surgery, Erasmus Medical Center, Erasmus University, Rotterdam, Netherlands
[c] Netherlands Heart Registration, Utrecht, Netherlands
[d] Department of Cardiothoracic Surgery, St. Antonius Hospital, Nieuwegein, Netherlands
[e] Department of Medical Ethics and Health Law, Leiden University Medical Center, Leiden University, Leiden, Netherlands
[f] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden University, Leiden, Netherlands
[g] Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden University, Leiden, Netherlands

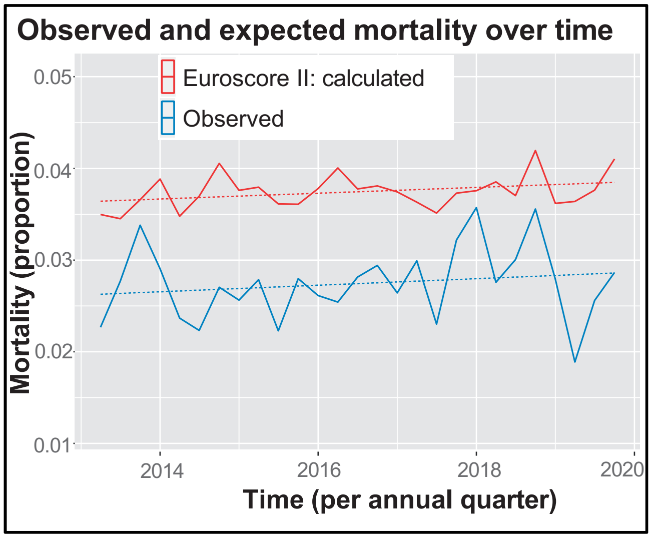
* Corresponding author. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. Tel: +31887569110; e-mail: e.schuit@umcutrecht.nl (E. Schuit).

Received 27 July 2022; received in revised form 21 June 2023; accepted 5 September 2023

Trends in the performance ofthe EuroSCORE II over time

Summary

The EuroSCORE II consistently overestimates mortality risks of all types of major cardiothoracic surgical procedure in adult patients who underwent cardiothoracic surgery in the Netherlands.



Observed and expected mortality over time

Legend: Upper graph, dashed lines: average of observed/EuroSCORE II calculated mortality; dotted lines: trend line (LM) of observed/EuroSCORE II calculated mortality.

[†]See Supplementary Material for Cardiothoracic Surgery Registration Committee members of the Netherlands Heart Registration

Abstract

OBJECTIVES: The aim of this study was to investigate the performance of the EuroSCORE II over time and dynamics in values of predictors included in the model.

METHODS: A cohort study was performed using data from the Netherlands Heart Registration. All cardiothoracic surgical procedures performed between 1 January 2013 and 31 December 2019 were included for analysis. Performance of the EuroSCORE II was assessed across 3-month intervals in terms of calibration and discrimination. For subgroups of major surgical procedures, performance of the EuroSCORE II was assessed across 12-month time intervals. Changes in values of individual EuroSCORE II predictors over time were assessed graphically.

RESULTS: A total of 103 404 cardiothoracic surgical procedures were included. Observed mortality risk ranged between 1.9% [95% confidence interval (CI) 1.6–2.4] and 3.6% (95% CI 2.6–4.4) across 3-month intervals, while the mean predicted mortality risk ranged between 3.4% (95% CI 3.3–3.6) and 4.2% (95% CI 3.9–4.6). The corresponding observed:expected ratios ranged from 0.50 (95% CI 0.46–0.61) to 0.95 (95% CI 0.74–1.16). Discriminative performance in terms of the *c*-statistic ranged between 0.82 (95% CI 0.78–0.89) and 0.89 (95% CI 0.87–0.93). The EuroSCORE II consistently overestimated mortality compared to observed mortality. This finding was consistent across all major cardiothoracic surgical procedures. Distributions of values of individual predictors varied broadly across predictors over time. Most notable trends were a decrease in elective surgery from 75% to 54% and a rise in patients with no or New York Heart Association I class heart failure from 27% to 33%.

CONCLUSIONS: The EuroSCORE II shows good discriminative performance, but consistently overestimates mortality risks of all types of major cardiothoracic surgical procedures in the Netherlands.

Keywords: Prediction models • Population dynamics • Cardiothoracic surgery

ABBREVIATIONS

AUC	Area under the receiver-operating characteristic curve
CIs	Confidence intervals
CITL	Calibration in the large
NHR	Netherlands Heart Registration
NYHA	New York Heart Association
O:E	Observed:expected
STS	Society of Thoracic Surgery

INTRODUCTION

The EuroSCORE II model aims to support clinicians and their patients to determine whether benefits of cardiac surgery outweigh mortality risks associated with these procedures [1]. After implementation in clinical practice prediction models, like the EuroSCORE II, are generally used for many years without any reassessment of their performance. Consequently, prediction models are at risk of showing decreased performance over time, a phenomenon also known as concept drift [2]. Mortality predictions of the original EuroSCORE model, for instance, were found to increasingly overestimate the mortality risk over time [3–5]. Underlying mechanisms were found in changing patient risk profiles, improved outcomes and the introduction of novel interventions [5]. This drift of the original EuroSCORE was also the main reason to introduce the EuroSCORE II, as in use today [1].

As prediction models become more frequently in use in clinical practice it is increasingly important to regularly check their performance and update them when needed. The increased interest in registries, real-world data, and the concept of learning healthcare systems will likely facilitate more frequent or even continuous assessment of concept drift [6]. As the EuroSCORE II almost celebrates its 10-year anniversary it is time to assess its performance and dynamics in characteristics of the patients for whom the model is potentially applied.

Therefore, the aim of our study was to investigate the performance of the EuroSCORE II and dynamics in underlying values of predictors over time.

METHODS

Ethical statement

An exemption from the Dutch law on medical research on human beings was obtained from the Medical Ethical Review Committee Utrecht prior to this study (protocol number 20-698/C).

Study design

This study was a cohort study assessing performance trends and population dynamics of the EuroSCORE II over time, using data from the Netherlands Heart Registration (NHR). The NHR comprises prospectively collected data on almost 1.5 million adults with cardiovascular disease and >200 000 patients who underwent cardiac surgery in one of the 16 Dutch cardiothoracic medical centres since 2007 [7]. Data on the NHR comprise a wide range of patient characteristics (including the original EuroSCORE and EuroSCORE II variables), procedural variables and outcome measures. Data received from participating medical centres was collected, reviewed and audited in accordance with the NHR manual (https://www.nhr.nl).

Study population

Since the EuroSCORE II was published in 2012, data for all cardiothoracic surgery procedures performed in adult patients in the Netherlands were extracted from the NHR from 1 January 2013, onwards, up to 31 December 2019. No distinction was made between first interventions and reinterventions as reinterventions are included as a predictor in the EuroSCORE II.

Outcome measures

The primary outcome for this study was the performance of the EuroSCORE II model in predicting in-hospital mortality over time [1]. In addition, this study aimed to investigate dynamics of the values of predictors of the EuroSCORE II model.

EuroSCORE II

In short, the EuroSCORE II model was developed to predict in-hospital mortality of patients after cardiac surgery [1]. The model coefficient values are provided in Supplementary Material, Table S2. The dataset substantiating model development comprised 22,381 cardiothoracic patients from 154 hospitals in 43 countries. Ten-fold cross-validation was conducted during model development, dividing the original dataset into 10 equally sized random samples of which 90% was used for fitting the model and 10% for validation. Performance of the EuroSCORE II comprised an observed:expected (O:E) ratio of 1.06 (observed: 4.18%; expected: 3.95%), and a c-statistic of 0.81 in the validation data. For more details on the EuroSCORE II model we refer to the original publication [1].

Data analysis

The main analysis included all patients consecutively included in the NHR within the previously defined timeframe. Missing data on individual predictors and the outcome was assessed by calculating overall proportions of missing data, the proportion of missing data across the 3-month intervals, and the overall proportion of missing data per individual predictor. A table was constructed to compare patients with information on all predictors and the outcome available (i.e. without missing values) to patients with 1 or more missing values on any of the predictors or the outcome. To test for differences between these groups, chi-square tests and *t*-tests were performed for categorical and continuous predictors, respectively. Missing data were imputed using multiple imputation chained equations. The imputation model included all predictor variables and the outcome. Data were imputed 10 times in each 3-month period to retain potential heterogeneity in model performance or population characteristics over time. Confidence intervals (CIs) were pooled from 200 bootstrap samples evenly distributed over all imputation sets, following Schomaker and Heumann [8].

Model performance was assessed in terms of calibration and discrimination across 3-month intervals. These intervals were chosen such that at least 100 patients with the outcome, i.e., mortality, were included in the interval to allow reliable external validation of the EuroSCORE II [9]. Calibration refers to agreement between the predicted mortality risk and the observed proportions of cardiac surgery patients who died in-hospital. Calibration was assessed graphically in calibration plots and numerically in terms of the O:E ratio, the calibration in the large (CITL) and the calibration slope. The O:E ratio was calculated and plotted by dividing the incidence of mortality by the average mortality risk as calculated by the EuroSCORE II in the same group of patients. An O:E ratio of 1 indicates good overall agreement between observed and expected in-hospital mortality. The CITL and calibration slope were derived from a calibration plot. The calibration line in the plot was described with the calibration slope (*b*) and with an intercept (*a*), given that the calibration slope is set to 1 (*a*|*b* = 1, CITL), as proposed by Cox [10]. With perfect calibration, the CITL would be zero, and the calibration slope would equal one. Discrimination refers to the ability of the model to distinguish between patients who died and patients who survived and was assessed with the area under the receiver-operating characteristic curve (AUC) [11]. Trends over time for all model performance measures were examined in plots drawn with non-parametric locally weighted smoothing.

Subsequent similar analyses were performed for subgroups stratified for 4 major surgical procedures (isolated coronary artery bypass grafting, aortic valve surgery, mitral valve surgery, aorta surgery) across 12-month intervals. For these latter analyses, 12-month intervals were chosen to increase the number of events per sampling interval.

Similar plots were constructed to assess dynamics in values of individual model predictors. To assess effects of individual predictors on the overall model performance, a plot comprising all individual predictors centred around their initial 3 months means was produced.

All analyses were performed using R software, version 4.0.3.18 [12].

RESULTS

Population

In total, 103 404 procedures of adult patients who received cardiothoracic surgery were recorded in the NHR between 1 January 2013 and 31 December 2019. Overall, 7.90% of data were missing, with the highest percentage of missing values for New York Heart Association (NYHA) class (30.2%), poor mobility (20.2%) and Canadian Cardiovascular Society IV angina pectoris score (18.3%) (see Supplementary Material, Table S1 and Supplementary Material, Fig. S1 for missing data trends per predictor over time). For 36 947 patients (35.7%), at least 1 predictor or outcome value was missing. Notably more data were missing in the first 2 years of the dataset compared to later years. After imputation of missing data, all 103 404 patients were included for further analysis. Three-month segments comprised an average of 3693 patients, of whom an average of 103 patients died.

Overall, patients were 65.8 years of age at the time of surgery, with 27.5% being female, and an average mortality of 2.8% (Table 1). Characteristics of our study population were largely comparable with the population that was used for the development of the EuroSCORE II (Table 1), with the exception that we saw less dialysis (0.4% vs 1.1%), more urgent and emergency procedures (43.8% vs 22.8%), and more thoracic aorta surgeries (9.7% vs 7.3%). Despite the seemingly higher-risk population, the overall mortality rate was lower in our population compared to the EuroSCORE II population (2.8% vs 3.9%).

Overall EuroSCORE II performance

Across 3-month time intervals, observed mortality of cardiothoracic surgery ranged between 1.9% (95% CI 1.6–2.4) and 3.6% (95% CI 2.6–4.4) over time (Fig. 1**)**. Expected mortality, calculated by the EuroSCORE II, ranged between 3.4% (95% CI 3.3–3.6) and 4.2% (95% CI 3.9–4.6) over time. Plots showed the observed

Table 1: Overall baseline characteristics of EuroSCORE II predictors

Characteristic	*n* (%) or mean (SD)	
	Validation population (current study) *N* = 103 404	Development population (EuroSCORE II study) [1] *N* = 22 381
Outcome		
In-hospital mortality, *n* (%)	2678 (2.8)	873 (3.9)
Predictor		
Age, years (SD)	65.8 (11.0)	64.6 (12.5)
Female, *n* (%)	27 988 (27.5)	6919 (30.9)
Creatinine concentration, mmol/l (SD)	93.2 (47.8)	96.4 (57.1)
Dialysis, *n* (%)	412 (0.40)	244 (1.1)
Insulin-dependent diabetes mellitus, *n* (%)	8169 (7.9)	1705 (7.6)
NYHA, *n* (%)		
Class I	34 139 (33.0)	NR
Class II	36 563 (35.4)	NR
Class III	26 419 (25.6)	NR
Class IV	6193 (6.0)	NR
LV function, % (SD)	51.4 (9.49)	NR
CCS4, *n* (%)	6204 (6.0)	NR
Active endocarditis, *n* (%)	2698 (2.6)	497 (2.2)
Extracardiac arteriopathy, *n* (%)	11 788 (11.3)	NR
Chronic pulmonary disease, *n* (%)	10 960 (10,6)	NR
Pulmonary artery systolic pressure, mmHg, (SD)	26.6 (6.57)	NR
Recent myocardial infarction, *n* (%)	22 335 (21.6)	NR
N/M mobility, *n* (%)	2554 (2.5)	NR
Urgency, *n* (%)		
Elective	57 033 (55.2)	17 165 (76.7)
Urgent	38 673 (37.4)	4135 (18.5)
Emergency	6659 (6.4)	972 (4.3)
Salvage	1039 (1.0)	109 (0.5)
Critical preoperative state, *n* (%)	4022 (3.9)	924 (4.1)
Weight of procedure, *n* (%)		
Isolated CABG	53 432 (51.7%)	10 448 (46.7)
1 procedure (non-CABG)	24 827 (24.4)	NR
2 procedures	19 202 (18.6%)	NR
3 or more procedures	5943 (5.7%)	NR
Thoracic aorta surgery, *n* (%)	10 065 (9.7)	1636 (7.3)
Previous cardiac surgery (redo), *n* (%)	7052 (6.8)	NR

CABG: coronary artery bypass grafting; CCS4: CCS class 4 angina; LV: left ventricular; N/M mobility: neurological or musculoskeletal dysfunction severely affecting mobility; NR: not reported in the original EuroSCORE II publication [1]; NYHA: New York Heart Association; SD: standard deviation.

mortality to be consistently overestimated by the model. O:E ratios ranged from 0.50 (95% CI 0.46–0.61) to 0.95 (95% CI 0.74–1.16).

CITL for the EuroSCORE II ranged from -0.81 (95% CI -1.03 to 0.60) to -0.02 (95% CI -0.12 to 0.08) (Supplementary Material, Figs S2 and S3). Slopes for the EuroSCORE II centred around 1, ranging from 0.85 (95% CI 0.70 to 1.00) to 1.21 (95% CI 1.05–1.38).

AUCs of the EuroSCORE II ranged between 0.82 (95% CI 0.78–0.89) and 0.89 (95% CI 0.87–0.93) (Fig. 2). The mean AUC over time was 0.86 (95% CI 0.83–0.89).

EuroSCORE II performance per major procedure type

For isolated coronary artery bypass grafting observed mortality ranged from 1.0% (95% CI 1.0–1.2) to 1.1% (95% CI 1.0–1.2), and expected mortality from 1.5% (95% CI 1.5–1.6) to 1.6% (95% CI 1.6–1.6) (Fig. 3). For aortic valve surgery observed mortality ranged from 2.9% (95% CI 2.2–3.7) to 3.5% (95% CI 2.6–4.2), and expected mortality from 4.6% (95% CI 4.3–5.1) to 5.1% (95% CI 4.7–5.4). For mitral valve surgery observed mortality ranged from 4.4% (95% CI 3.4–5.7) to 5.0% (95% CI 3.3–6.6), and expected mortality from 5.1% (95% CI 4.6–5.6) to 6.5% (95% CI 5.5–7.7). For major aortic surgery observed mortality ranged from 7.0% (95% CI 4.9–8.8) to 8.4% (95% CI 6.4–10.8), and expected mortality from 8.9% (95% CI 8.4–10.1) to 11.4% (95% CI 10.4–12.3).

Dynamics of predictor values

Trends in values of individual model predictors varied broadly between predictors (Fig. 4). Most notable trends were the decrease in elective surgery from 75% to 54%, the rise in patients with no or NYHA I class heart failure from 27% to 33%, the rise in patients with a recent myocardial infarction from 19% to 23% and the rise in aorta surgery from 7% to 11%.

DISCUSSION

This study examined the model performance of the EuroSCORE II over time since its introduction. Our results show that the

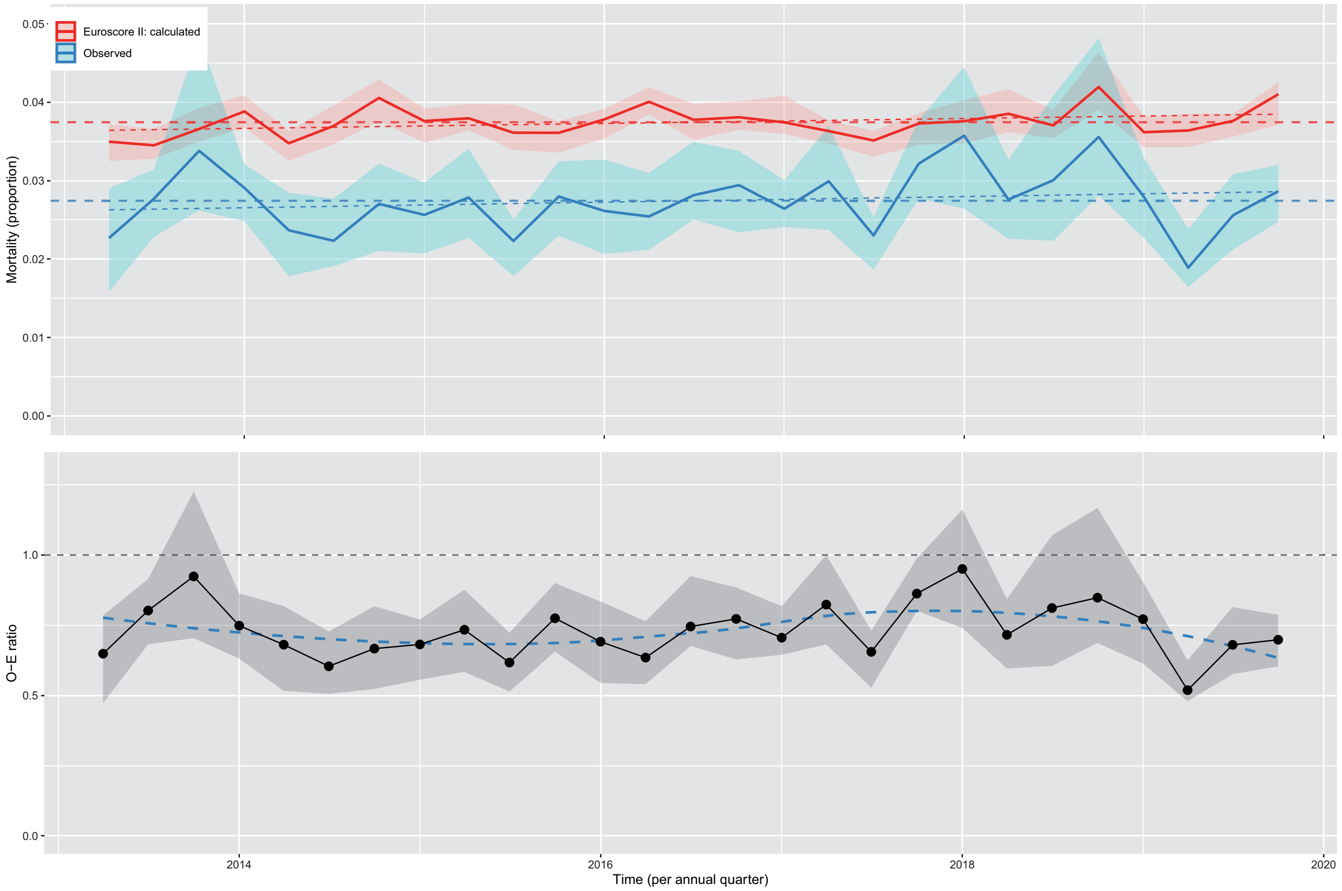


Figure 1: Observed and expected mortality of the EuroSCORE II over time. Upper graph, solid lines: average of observed/EuroSCORE II calculated mortality at specific time points; dahsed lines: overall average observed/EuroSCORE II calculated mortality; dotted lines: trend line (LM) of observed/EuroSCORE II calculated mortality, lower graph, solid line: observed:expected ratios; dashed line: trend line (locally weighted smoothing) of the observed:expected ratio.

EuroSCORE II persistently overestimates in-hospital mortality after cardiothoracic surgery in the Netherlands. Dynamics of values of individual model predictors showed an increase in urgency of procedures with subsequent decrease in elective procedures, and a decrease in heart failure severity. These changes did not result in a substantial change of the EuroSCORE II's model performance over time.

Since its introduction the performance of the EuroSCORE II has been assessed multiple times for specific populations or in comparison to other risk scores. In general, authors found mixed results for calibration, which may be partly explained by design flaws in the development of the score [13], and satisfactory results for discrimination (i.e. an AUC of >0.7) of the EuroSCORE II [14–17]. For instance, in patients undergoing aortic valve surgery the model was found to perform satisfactory [17], while in patients undergoing minimally invasive cardiothoracic surgery calibration was found to be accurate for low-risk patients only [15]. In a recent study on risks of cardiothoracic surgery for octogenarians, the EuroSCORE II was found to steadily underestimate mortality risks [16]. When compared to other risk prediction models like the original EuroSCORE, the logistic version of the original EuroSCORE and Society of Thoracic Surgery (STS) short-term risk score, the EuroSCORE II is typically found to perform better than the first 2 models and equal to the latter [14, 18]. Notably, a single-centre study validating the EuroSCORE II in a Dutch medical centre found the model to systematically underestimate in-hospital mortality, while our study found overestimation of mortality for all major cardiothoracic procedures [19]. A possible explanation for these different research findings is that the single-centre study was a subset of the data used in this study.

Noticeable also was the high number of missing values for certain predictors like NYHA, and poor mobility. In studies on the performance of the EuroSCORE I with NHR data in the years before the introduction of the EuroSCORE II substantially less predictor values were missing [4, 20]. The EuroSCORE II seems to encompass predictors that are less frequently reported by clinicians or that are not applicable for many patients, even though they are relevant for the model. For example, the default value of the NYHA predictor assumes patients to be scored as NYHA I at minimum. However, if patients are not dyspneic it is not possible for clinicians to assign patients an NYHA class at all, which means no NYHA class is reported for that patient.

Future perspectives

Dynamics in populations underlying clinical prediction models are increasingly recognized and acknowledged in research [21]. Still, existing prediction models are only updated infrequently. As previously mentioned, the original EuroSCORE was used for a decade

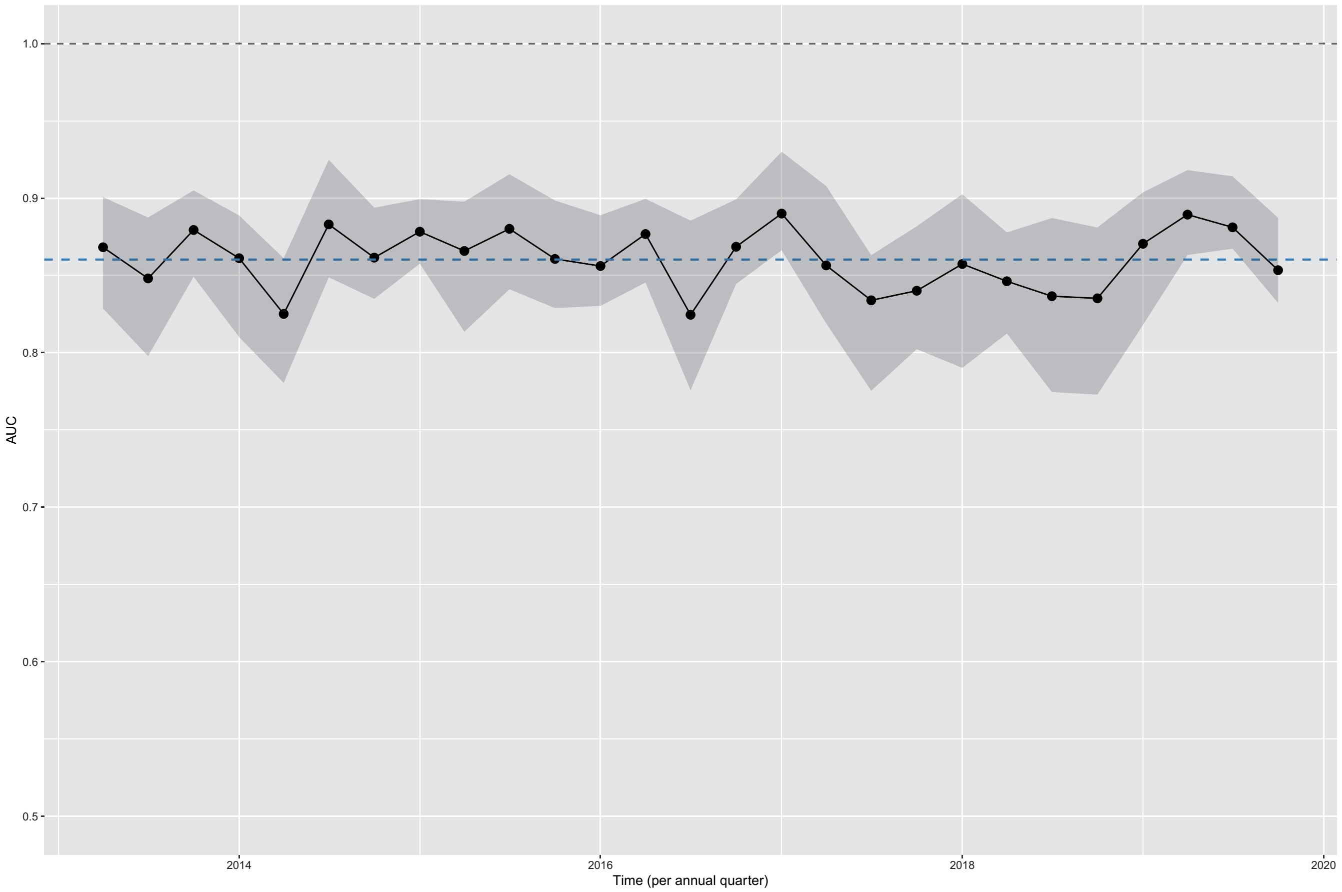


Figure 2: Area under the receiver-operating characteristic curve with 95% confidence interval of the EuroSCORE II over time. Dots: average interval AUC; Wider area: 95% confidence intervals of interval AUC; dashed line: overall average AUC. AUC: area under the receiver-operating characteristic curve.

before receiving its much-needed update to the EuroSCORE II. In comparison, introduced in 2008 and updated in 2018, the STS short-term risk score was updated after a decade of use also [22]. Similar to the EuroSCORE, the STS short-term risk score was updated to adjust for changing patient characteristics [22].

As this study shows, it is difficult to determine when the time is right to update existing clinical prediction models. Ten years after the original EuroSCORE was reported to be no longer suitable for use due to decreasing performance the model was updated [5]. This study shows that despite changes in the underlying population, at least in the Netherlands, the EuroSCORE II consistently overestimates mortality after cardiac surgery, but its performance does not deteriorate. Other authors rightly point out that prediction models should not be deemed axiomatic [22], yet leaving the question open when to update prediction models. According to the EuroSCORE website, the model (EuroSCORE II) needs recalibration, 'as the importance of data in informing clinical decisions is increasing' [23], meaning a EuroSCORE III will be introduced in the future. When developing EuroSCORE III it will be important to consider the criticism received on the EuroSCORE II [13].

To assess prediction model deviations, other industries use the notion of concept drift [2, 23]. Introduced to monitor sensor deviations in hardware industries, concept drift monitors how and when prediction models depart from their original outcomes (the 'concept'), or when 1 or more of their underlying predictors diverge [24].

After detecting model drift, the models must be updated. Several authors have explored model updating of prediction models, in particular in cardiac surgery too [20, 25]. To date, however, these dynamic modelling methods have not found their way to the models applied in clinical practice yet. Most likely, connections between clinical and methodological researchers need to be improved for that. Future research might focus on time limits for clinical prediction models to be (re-)assessed, and when needed, updated. Particularly when these models are greatly relied upon in clinical practice such as the EuroSCORE II.

Limitations

Several limitations apply to this study. First, this study was performed using data from the NHR comprising a subset of Dutch cases. Regardless, the NHR is a very large and comprehensive registry. Similar analysis in other countries could therefore yield different results due to differences in populations and intervention criteria. Second, some predictors comprised up to 30% missing values, with less data missing as time progressed. Moreover, missing values were found to be not missing completely at random. Especially in such circumstances a complete case analysis, which excludes patients with 1 or more missing predictor or outcome values from the analysis, is known to lead to biased results [26]. Therefore, we feel confident that our approach was the best possible way to deal with the encountered problem of missing data.

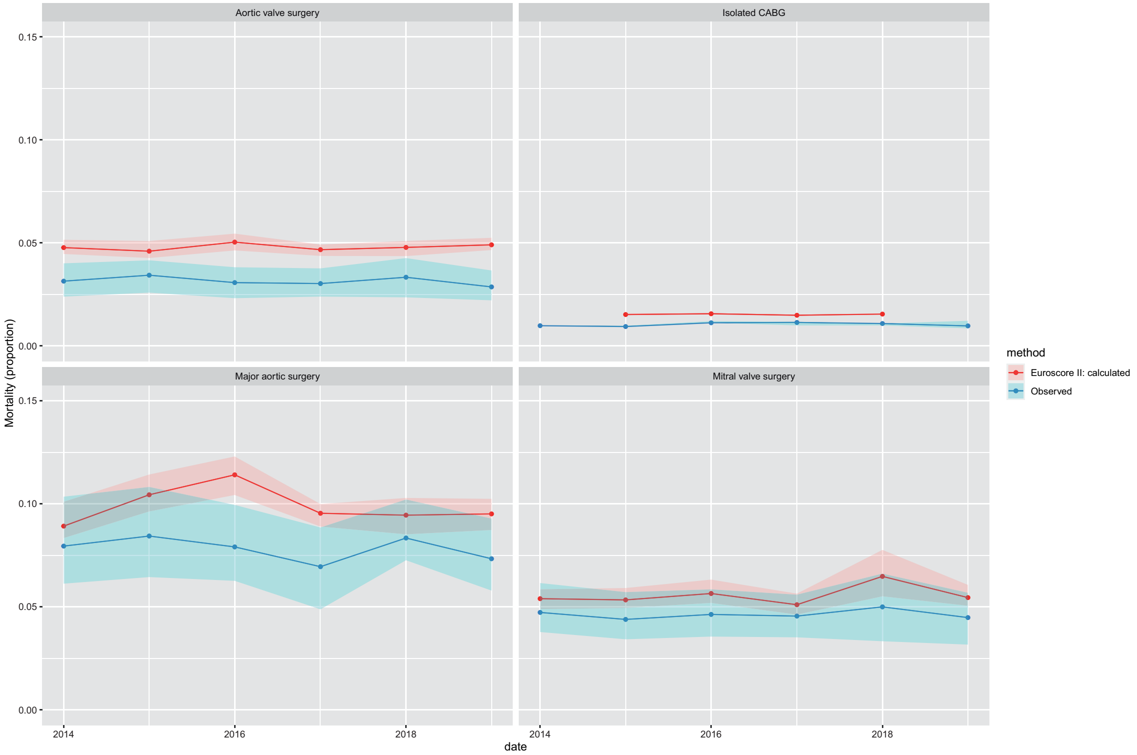


Figure 3: EuroSCORE II performance per major procedure type. Upper lines: average of EuroSCORE II calculated mortality; Lower lines: observed mortality for the 4 major procedure types: aortic valve surgery (top left), isolated coronary artery bypass graft (coronary artery bypass grafting; top right), major aortic surgery (bottom left), and mitral valve surgery (bottom right).

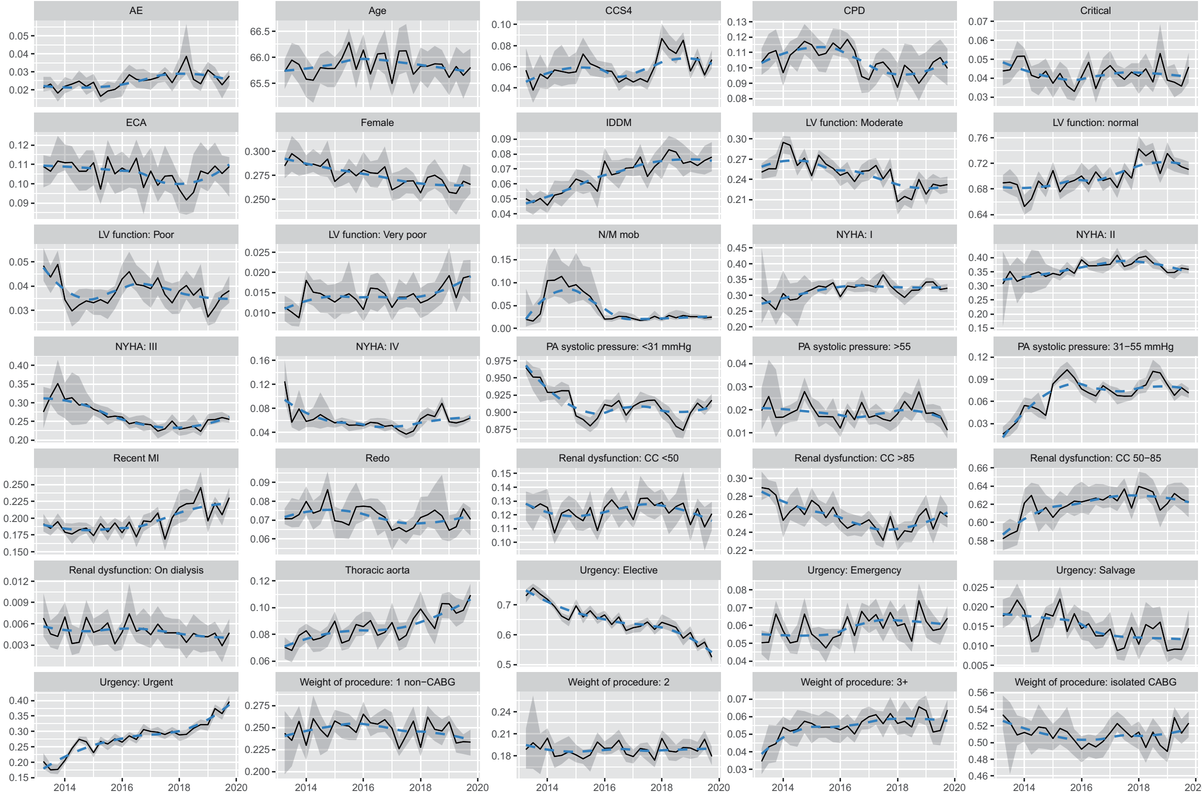


Figure 4: Trends in values of individual EuroSCORE II model predictors over time. Solid lines: proportions of individual predictors; dashed lines: trend line (locally weighted smoothing) of predictor dynamics. AE: active endocarditis; CABG: coronary artery bypass grafting; CCS4: CCS class 4 angina; CPD: chronic pulmonary disease; Critical: critical preoperative state; ECA: extracardiac arteriopathy; IDDM: insulin-dependent diabetes mellitus; LV function: left ventricular function; N/M mob: neurological or musculoskeletal dysfunction severely affecting mobility; NYHA: New York Heart Association; PA systolic: pulmonary artery systolic pressure; Recent MI: recent myocardial infarction; Redo: previous cardiac surgery. Weight of procedure '1 non-CABG': single major cardiac procedure which is not isolated CABG; 2: 2 major cardiac procedures; 3+: 3 or more major cardiac procedures.

CONCLUSIONS

Even though the population characteristics of cardiac surgery patients in the Netherlands have changed over time, the EuroSCORE II shows good discriminative performance but consistently overestimates the in-hospital mortality risk after cardiothoracic surgery in the Netherlands over time for all major procedures. Still, more attention is needed for dynamic trends and modelling in clinical prediction models.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *EJCTS* online.

ACKNOWLEDGEMENTS

The authors would like to mention their gratitude to the Netherlands Heart Registration for making available the data substantiating this work.

Funding

This work was supported by the Netherlands Organisation for Health Research and Development (ZonMW) (grant number 91217027).

Conflict of interest: None of the authors reported any conflicts of interest.

DATA AVAILABILITY

The data underlying this article were provided by the NHR by permission. Data will be shared on request to the corresponding author with permission of the NHR.

Author contributions

Wouter B. van Dijk: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing—original draft; Writing—review & editing. **Artuur M. Leeuwenberg:** Data curation; Formal analysis; Investigation. **Diederick E. Grobbee:** Funding acquisition; Supervision; Writing—review & editing. **Sabrina Siregar:** Supervision; Writing—review & editing. **Saskia Houterman:** Data curation; Supervision; Writing—review & editing. **Edgar J. Daeter:** Supervision; Writing—review & editing. **Martine C. de Vries:** Funding acquisition; Supervision; Writing—review & editing. **Rolf H. H. Groenwold:** Conceptualization; Methodology; Supervision; Writing—review & editing. **Ewoud Schuit:** Conceptualization; Formal analysis; Investigation; Methodology; Supervision; Writing—review & editing.

Reviewer information

European Journal of Cardio-Thoracic Surgery thanks Gary L. Grunkemeier, Vito Domenico Bruno, Dileep C Unnikrishnan and the other, anonymous reviewer(s) for their contribution to the peer review process of this article.

REFERENCES

[1] Nashef SAM, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR et al Euroscore II. Eur J Cardiothorac Surg 2012;41:734–45.

[2] Kukar M. Drifting concepts as hidden factors in clinical studies. Lect Notes Comput Sci 2003;2780:355–64.

[3] Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16:9–13.

[4] Siregar S, Groenwold RHH, de Heer F, Bots ML, van der Graaf Y, van Herwerden LA. Performance of the original EuroSCORE. Eur J Cardiothorac Surg 2012;41:746–54.

[5] Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K et al Dynamic trends in cardiac surgery: why the logistic euroscore is no longer suitable for contemporary cardiac surgery and implications for future risk models. Eur J Cardiothorac Surg 2013;43:1146–52.

[6] Nakatsugawa M, Cheng Z, Kiess A, Choflet A, Bowers M, Utsunomiya K *et al.* The needs and benefits of continuous model updates on the accuracy of RT-induced toxicity prediction models within a learning health system. Int J Radiat Oncol Biol Phys 2019;103:460–7.

[7] van Veghel D, Marteijn M, de Mol B; Measurably Better Study Group (The Netherlands) and Advisory Board. First results of a national initiative to enable quality improvement of cardiovascular care by transparently reporting on patient-relevant outcomes. Eur J Cardiothorac Surg 2016;49:1660–9.

[8] Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. Stat Med 2018;37:2252–66.

[9] Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 2005;58:475–83.

[10] Cox DR. Two further applications of a model for binary regression. Biometrika 1958;45:562–5.

[11] Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing, 2019.

[12] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[13] Collins G, Altman D. Design flaws in EuroSCORE II. Eur J Cardiothorac Surg 2013;43:871.

[14] Velicki L, Cemerlic-Adjic N, Pavlovic K, Mihajlovic BB, Bankovic D, Mihajlovic B *et al.* Clinical performance of the EuroSCORE II compared with the previous EuroSCORE Iterations. Thorac Cardiovasc Surg 2014;62:288–97.

[15] Margaryan R, Moscarelli M, Gasbarri T, Bianchi G, Kallushi E, Cerillo AG *et al.* EuroSCORE performance in minimally invasive cardiac surgery discrimination ability and external calibration. Innovations (Phila) 2017;12:282–286.

[16] Kuplay H, Bayer Erdoğan S, Baştopçu M, Karpuzoğlu E, Er H. Performance of the EuroSCORE II and the STS score for cardiac surgery in octogenarians. Turk Gogus Kalp Damar Cerrahisi Derg 2021;29:174–82.

[17] Duchnowski P, Hryniewiecki T, Kuśmierczyk M, Szymanski P. Performance of the EuroSCORE II and the society of thoracic surgeons score in patients undergoing aortic valve replacement for aortic stenosis. J Thorac Dis 2019;11:2076–81.

[18] Sullivan PG, Wallach JD, Ioannidis JPA. Meta-analysis comparing established risk prediction models (EuroSCORE II, STS score, and ACEF score) for perioperative mortality during cardiac surgery. Am J Cardiol 2016; 118:1574–82.

[19] Hogervorst EK, Rosseel PMJ, van de Watering LMG, Brand A, Bentala M, van der Meer BJM *et al.* Prospective validation of the EuroSCORE II risk model in a single Dutch cardiac surgery centre. Neth Heart J 2018;26:540–51.

[20] Siregar S, Nieboer D, Vergouwe Y, Versteegh MIM, Noyez L, Vonk ABA et al Improved prediction by dynamic modeling: an exploratory study in the adult cardiac surgery database of the Netherlands Association for Cardio-Thoracic Surgery. Circ Cardiovasc Qual Outcomes 2016;9:171–81.

[21] Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW, Bossuyt P *et al.*; On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:1–7.

[22] Shahian DM, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC *et al.* The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: part 1—background, design considerations, and model development. Ann Thorac Surg 2018;105:1411–1418.

[23] EuroSCORE 3. https://www.euroscore.org/index.php?id=38 (28 2022 November, date last accessed).

[24] Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. ACM Comput Surv 2014;46:1–37.

[25] Vergouwe Y, Nieboer D, Oostenbrink R, Debray TPA, Murray GD, Kattan MW *et al.* A closed testing procedure to select an appropriate method for updating prediction models. Stat Med 2017;36:4529–39.

[26] Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. J Clin Epidemiol 2006;59:1092–101.