

Temporal External Validation of Successive RCRI (Revised Cardiac Risk Index) Updates across Observational Healthcare Databases: a Multinational Retrospective Cohort Study

Alexander Saelmans

a.saelmans@erasmusmc.nl

Erasmus MC

Jenna M. Reps

Erasmus MC

Evan Minty

Johns Hopkins University

Angela Leis Machin

Hospital del Mar Research Institute

Miguel-Angel Mayer

Hospital del Mar Research Institute

Juan Manuel Ramirez-Anguila

Hospital del Mar Research Institute

Farnoosh Haji-Sheikhi

Stanford University

Priya Desai

Stanford University

Hannah Morgan-Cooper

Stanford University

Thomas Falconer

Columbia University

Andrey Soares

University of Colorado Anschutz Medical Campus

Alvaro Martinez Perez

Instituto de Investigación Sanitaria La Fe

Carlos Lopez Gomez

Instituto de Investigación Sanitaria La Fe

Mateus de Lima Freitas

Hospital Israelita Albert Einstein

Cesar Augusto Madid Truys

Hospital Israelita Albert Einstein

Peter R. Rijnbeek

Erasmus MC

Ross D. Williams

Erasmus MC

Research Article

Keywords: Revised Cardiac Risk Index, RCRI, Updating, Recalibration, Clinical Prediction Models

Posted Date: September 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10626993/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: Competing interest reported. A.S., J.R., P.R. and R.W. work for a research group at Erasmus University Medical Center that receives/received an unconditional grant for methodological research by Johnson & Johnson. The grant is for the institute. Dr Reps is an employee and shareholder of Johnson & Johnson.

Abstract

Background

We aim to evaluate whether the Revised Cardiac Risk Index (RCRI) recalibration methods worked.

Methods

We performed temporal external validation and external validation of different recalibrated versions of the RCRI, an implemented prediction model, on observational health databases. The three versions included the original RCRI model (oRCRI), as well as the Duceppe and MDCalc recalibrations. (I) Temporal external validation was performed in 6-month intervals from 2010 to 2025, while (II) external validation overall covered the entire period from 2010 to 2025.

Results

We performed temporal external validation on eight databases. The Area Under the Receiver Operating Characteristic Curve showed a statistically significant ($p = 0.0099$) decrease. Calibration metrics and Net-Benefit were relatively consistent. Based on calibration metrics and Net-Benefit values, the RCRI versions can be ranked from best performing to worst performing as follows: MDCalc, oRCRI and Duceppe. In addition, we performed external validation on nine (including the eight used for temporal external validation) databases with an AUROC range of 0.62–0.86 and calibration plots that indicate an overestimation of risk. For an RCRI score of 1, the Net-Benefit range was the most substantial for the MDCalc version (-0.0013 – 0.022), followed by the oRCRI version (-0.00056 – 0.023) and the Duceppe version (-0.025 – (-0.012)).

Conclusions

While the RCRI recalibrations show consistent calibration and Net-Benefit, the significant decline in AUROC highlights challenges that extend beyond the RCRI itself. As patient populations and care patterns evolve, even well-established clinical prediction models require ongoing reassessment, illustrating both the value and limitations of recalibration alone.

Introduction

Clinical prediction models estimate an individual's future risk of a specific outcome based on multiple predictors derived from historical data. However, when these models are implemented, shifts in healthcare data and practices over time can lead to decreased model performance, necessitating regular updates.^{1–4} There are various types of updating strategies to maintain a model's performance over time: recalibration or revision, model extension, meta-modeling and dynamic updating. Recalibration or revision, e.g. deletion, adjusts a model for novel data settings. The optimal approach for updating remains unclear.⁵

An example⁶ of a clinically used binary prediction model that has been updated by means of recalibration is the Revised Cardiac Risk Index (RCRI) model for major postoperative cardiac complications. This original model (oRCRI) was developed in 1999 and internally validated with an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.759 ± 0.032 .⁷ A systematic review by *Ford et al.*⁸ summarized external validations of this model with a median AUROC of 0.75 (95% CI 0.72–0.79). The RCRI model was updated by means of recalibration.^{9,10}

RCRI has been incorporated in medical guidelines of the United States of America (USA), Canada and Europe.^{9,11,12} Physicians use RCRI to identify high-risk patients. These patients can then be monitored more intensively. For example, a patient with a history of ischemic heart disease and cerebrovascular disease would then be monitored regularly with Troponin measurements the days following noncardiac surgery. This facilitates timely intervention should myocardial ischemia occur. Conversely, a change in the guidelines might have impacted the Net-Benefit of the model. A 2014 guideline advised an RCRI threshold of greater than or equal to a score of two.¹³ In contrast, a 2024 guideline advised an RCRI threshold of greater than a score of one.¹¹ The updates to these guidelines could thus impact the Net-Benefit of the model.

Whilst the RCRI prediction model has been incorporated in clinical guidelines worldwide, due to the model's age, it may not work as well as it did due to model and data 'drift'.^{14,15} A temporal external validation of the model would inform researchers better on the performance trend of a binary recalibrated model. There is currently a lack of research into the optimal prognostic model updating strategy. The RCRI model is a rare example of a model that has been clinically implemented and temporally updated via recalibration. Our study aims to gain insight into how successful model recalibration was at maintaining predictive performance over time for RCRI. A multi-database temporal external validation of the RCRI model and its updated versions allows for the assessment of the updating strategies themselves, and on the impact this has on the Net-Benefit. This study aims to temporally and externally validate different recalibrated versions of the RCRI prediction model on large observational healthcare databases to evaluate the impact of the recalibration methods.

Methods

Data sources

We performed external validation and temporal external validation on claims, general practitioner (GP) and electronic health record (EHR) databases from four different countries. The included databases are depicted in Table 1. These databases are mapped to the Observational Medical Outcome Partnership

Table 1
Included databases (n = 9) for external validation of the Revised Cardiac Risk Index (RCRI)

Database	Acronym	Country	Data provenance	Dates covered YYYY/MM	Patients (n)	%Female
Merative MarketScan® Medicare Supplemental Database	MDCR	USA	Claims	2000/01 2025/04	11,454,804	55.7
Merative MarketScan® Medicaid Supplemental Database	MDCD	USA	Claims	2006/01 2024/12	33,778,687	55.8
Merative MarketScan® Commercial Claims and Encounters	CCAE	USA	Claims	2000/01 2025/04	179,646,560	51.5
Japan Medical Data Center	JMDC	Japan	Claims	2005/01 2024/12	22,647,725	49.3
Integrated Primary Care Information	IPCI	NL	GP	2006/01 2025/06	3,059,100	51.2
Optum® De-identified Electronic Health Record Dataset	Optum® EHR	USA	EHR	2007/01 2025/04	171,915,506	53.5
STANford medicine Research data Repository-OMOP	STARR	USA	EHR	2008 2025	4,330,208	53.5
Columbia University Irving Medical Center	CUIMC	USA	EHR	2005/01 2025/12	7,374,357	55.5
Parc de Salut Mar Barcelona Information System	IMASIS	Spain	EHR	1990/01 2025/01	1,771,000	48.3
University of Colorado Anschutz Medical Campus-Health Data Compass*	CU_AMC	USA	EHR	1937 2025	43,035	57.2
IIS La Fe*	HULAFE	Spain	EHR	2012/01 2025/05	2,455,704	52.5
Hospital Israelita Albert Einstein*	HIAE	Brazil	EHR	2017/01 - 2025/12	4,118,200	55.5
EHR = Electronic health record, GP = General Practitioner, NL = The Netherlands, USA = United States of America						
*Because of insufficient counts for the outcome cohorts, these databases did not yield any results						

Revised cardiac risk index task

The RCRI prediction task we implemented in this paper is depicted in Fig. 1. The population that the model is applied to (the target cohort) consists of patients undergoing elective major non-cardiac surgery, e.g. a splenectomy. These patients were included if they had a minimum observation period of 365 days prior to index (surgery date). Myocardial infarction, cardiac arrest and death are the outcomes that are investigated during a time-at-risk window of 30 days.⁷

Three versions of RCRI for validation

The equally weighted predictors for the RCRI model are elevated-risk surgery (suprainguinal vascular, intrathoracic, or intraperitoneal procedures), history of ischemic heart disease, pre-operative treatment with insulin, history of congestive heart failure, pre-operative serum creatinine >2mg/dL (> 176.8 μmol/L) and history of cerebrovascular disease. For external validation, we used three versions of the RCRI model:

- 1. oRCRI: the original version by Lee *et al.* (1999)⁷
- 2. Duceppe: the recalibrated version by Duceppe *et al.* (2017)⁹
- 3. MDCalc: the recalibrated *MDCalc* version. (available in 2025)¹⁰

The score-to-risk probability mappings have been summarized in Table 2. For example, for the oRCRI version a patient with only a history of ischemic heart disease, would be assigned an RCRI-score of 1 and thus an accompanying risk probability of 0.9%.

Table 2
Original (oRCRI) and recalibrated versions of the Revised Cardiac Risk Index (RCRI)

RCRI score	Risk percentage		
	oRCRI	Duceppe	MDCalc
0	0.4	3.9	0.5
1	0.9	6.0	1.1
2	7	10.1	5
≥ 3	11	15	10

Target & outcome cohort differences across RCRI versions

The target and outcome cohorts differ slightly between the oRCRI and Duceppe recalibration. The target cohort includes non-elective emergency surgeries in the *Duceppe et al* study, but excludes these in the *Lee et al* study. The myocardial infarction in the outcome cohort was determined by means of serial cardiac markers:

- *Duceppe et al*: more recent Troponin values
- *Lee et al*: Creatinine kinase-MB (CK-MB) values

MDCalc does not specify whether emergency surgeries should be included in the target cohort.

Target, outcome and covariate cohorts used for this study

We adopted a cautious approach when defining the target cohort for elective surgery, focusing on procedures that are predominantly performed in an elective setting. For sensitivity, we did not exclude cases with a prior emergency visit, which means the cohort still contains some emergency surgeries (more in line with Duceppe). For the outcome cohort, nowadays Troponin is used more than CK-MB for the diagnosis of myocardial ischemia (more in line with Duceppe). We did not rely solely on CK-MB or Troponin cutoff values to identify (asymptomatic) myocardial ischemia, as these measurements are insufficiently reported in the used databases. Instead, we assumed that a diagnosis of myocardial ischemia would be recorded when such thresholds were exceeded. As IPCI is a GP database, surgical procedures are not captured. Therefore, we adjusted the target cohort for this specific database. Instead of a collection of non-cardiac surgical procedures it consists of patients with a visit to the hospital longer than 1 day for non-cardiac surgical specialisms, e.g. general surgery. When it comes to the covariates, we expanded on the original definition of pre-operative serum creatinine >2mg/dL (> 176.8 μmol/L). Next to this renal function value cutoff, we added renal impairment as creatinine is not always recorded in observational healthcare databases.

Phenotyping of the cohort definitions

We phenotyped the target, predictor and outcome cohorts by means of CohortDiagnostics.¹⁷ The incidence rates and gender ratios extracted from Cohort Diagnostics were compared to the ones collected from the literature.^{18,19} We observed a stable stratification per calendar year. For the selection of databases, we looked into whether required procedures were recorded within the database on the EHDEN (European Health Data & Evidence Network) portal (www.ehden.eu). The EHDEN portal includes most of the OMOP databases in Europe. Also, the EHDEN portal allows you to look for certain procedures within the database.^{20,21} The phenotyping process is described in the *Supplementary Table 1* and *Supplementary Fig. 1*.

External Validation

To compare overall predictive ability, we externally validated the three versions of the RCRI model in each database for the patients undergoing elective major non-cardiac surgery between 2010 to 2025 in each database. If a person had multiple surgeries, the model was only used to predict their risk after the first surgery that satisfied inclusion criteria (see section Revised cardiac risk index task). Performance was evaluated by comparing the RCRI predicted risk for a patient against the truth (whether the patient experienced myocardial infarction, cardiac arrest or death within 30 days after the surgery).

Temporal External Validation

To investigate how the models' performances change over time we evaluated all versions of the RCRI for patients undergoing their first elective major non-cardiac surgery (that satisfied inclusion criteria) during a fixed 6-month period in each database. The 6-month periods were [January 2010-June 2010], [July 2010-December 2010],..., [January 2025 -June 2025]. To investigate temporal trends, we conducted a mixed effects model analysis (with fixed effect of time and random database effect) to analyze whether the performance change was statistically significant over time. We excluded IPCI from the mixed effects model analysis, because of a different implementation of the target population. A random intercept was estimated for each database separately. We considered P values < 0.05 statistically significant.

Performance evaluation in external and temporal external validation

We evaluated discrimination, calibration performance and Net-Benefit over time and overall. Discrimination entails distinguishing well between participants who will and will not have the outcome of interest. We used the area under the receiver operating characteristic curve (AUROC) and Area Under the Precision-Recall Curve (AUPRC) to measure discrimination. AUROC is the plot of sensitivity vs. 1-specificity as the value of the cut-off point moves from 0 to 1. AUPRC summarizes the trade-off between precision and recall and provides an insightful perspective when dealing with imbalanced datasets.⁵

Calibration is defined as the agreement between predicted and observed risks. There are different forms of calibration. We visualized calibration numerically by means of difference between calibration in the large mean prediction and calibration in the large observed risk, Eavg, E90 and scaled Brier. Moreover, for overall external validation we visualized calibration with calibration plots.⁵

Decision curve analysis allows for the assessment of the clinical strategy by evaluating its Net-Benefit for a set threshold. The possible thresholds are represented by the x-axis and the Net-Benefit by the y-axis. Therefore, it depicts the trade-off between true- and false-positive predictions for a set threshold. For the Net-Benefit analysis, we extracted the values for the thresholds corresponding to an RCRI score of 1 and 2. We selected these thresholds as a recent guideline recommends physicians to apply a cut-off value of RCRI score > 1 and an older guideline recommends a cut-off value of RCRI score ≥ 2.^{11,13}

We performed all statistical analyses in R 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) with HADES²² packages “Strategus”, “PatientLevelPrediction”, “CohortGenerator”, “CohortDiagnostics” and “CohortExplorer”. In addition, we used “lme4” and “lmerTest” packages for the mixed effects model analysis. We created a package “RCRValidation” for this study to facilitate performing the analysis on various databases. (<https://github.com/ohdsi-studies/RCRValidation>)

Results

External validation

Baseline characteristics

We performed external validation on nine databases. The outcome percentages range from 0.34–3.1%. The database’s counts of the target and outcome cohorts and outcome percentages overall are shown in *Supplementary Table 2*.

Area Under the Receiver Operating Characteristic Curve (AUROC) and Area Under the Precision-Recall Curve (AUPRC)

The AUROC range is moderate to good with 0.62–0.86. For AUPRC, the range is 0.014–0.081. The AUROC and AUPRC for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases overall is depicted in Table 3.

Table 3

Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), difference between calibration in the large scaled Brier and net-benefit (RCRI score of 1 & RCRI score of 2) for original, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across o D = Duceppe recalibration, M = MD Calc version.

Database	AUROC*	AUPRC*	Calibration in the large difference			Eavg			E90			Scaled Brier		
			O	D	M	O	D	M	O	D	M	O	D	M
OptumEHR	0.75	0.072	0.010	0.050	0.0051	0.015	0.050	0.0086	0.034	0.068	0.018	0.26	0.57	0.16
MDCD	0.77	0.044	0.018	0.058	0.014	0.018	0.058	0.014	0.058	0.098	0.048	0.50	0.72	0.45
MDCR	0.77	0.060	0.0099	0.050	0.0058	0.013	0.049	0.0078	0.036	0.077	0.026	0.31	0.65	0.22
CCAE	0.86	0.043	0.0095	0.049	0.0084	0.0094	0.049	0.0084	0.047	0.078	0.029	0.59	0.83	0.57
JMDC	0.79	0.015	0.011	0.050	0.010	0.011	0.050	0.0010	0.057	0.089	0.038	0.72	0.88	0.72
IPCI	0.62	0.014	-0.00024	0.036	0.00057	0.0014	0.036	0.0015	0.0071	0.044	0.0051	-0.044	0.83	0.093
IMASIS	0.83	0.052	0.0080	0.047	0.0066	0.0082	0.047	0.0067	0.034	0.077	0.023	0.46	0.79	0.42
STARR	0.81	0.081	0.0040	0.043	0.0013	0.0073	0.043	0.0048	0.031	0.060	0.014	0.19	0.64	0.095
CUIMC	0.79	0.044	0.018	0.058	0.014	0.018	0.058	0.014	0.058	0.099	0.048	0.53	0.73	0.48
CCAE = Merative MarketScan® Commercial Claims and Encounters, MDCD = Merative MarketScan® Medicaid Supplemental Database, MDCR = Merative Medical Data Center, STARR = STAnford medicine Research data Repository-OMOP, CUIMC = Columbia University Irving Medical Center, IPCI = Integrated Primary Electronic Health Record Dataset *As recalibration does not impact the AUROC and AUPRC values, for these values one column was sufficient for the oRCRI														
~Net-Benefit values were made green when positive														

Calibration metrics

Calibration plots are depicted in Fig. 2 and indicate overestimation of risk. This overestimation is most pronounced for the Duceppe model and least pronounced for the MDCalc model. The range for the difference between calibration in the large mean prediction and calibration in the large observed risk; oRCRI - 0.00024-0.018, Duceppe 0.036-0.058, MDCalc 0.00057-0.014. For the Eavg the range is; oRCRI 0.0014-0.018, Duceppe 0.036-0.058, MDCalc 0.0010-0.014. E90's range is; oRCRI 0.0071-0.058, Duceppe 0.044-0.099, MDCalc 0.0051-0.048. The range for the Scaled Brier is; oRCRI - 0.044-0.72, Duceppe 0.57-0.88, MDCalc 0.093-0.72. In summary, other calibration metrics also showed a similar pattern of overestimation. This overestimation is most pronounced for the Duceppe model and least pronounced for the MDCalc model. The difference between calibration in the large mean prediction and calibration in the large observed risk, Eavg, E90 and scaled Brier for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases overall is shown in Table 3.

Net-Benefit

For an RCRI score of 1, the Net-Benefit range was most substantial for the MDCalc version (-0.0013-0.022), followed by the oRCRI version (-0.00056-0.023) and the Duceppe version that never reached a positive value for any of the included databases (-0.025 - (-0.012)). For an RCRI score of 2, the Net-Benefit range was only partially positive for the MDCalc version (-0.0050-0.0041). The oRCRI and Duceppe versions did not show a positive Net-Benefit for any of the included databases (oRCRI - 0.012 - (-0.00055), Duceppe - 0.025 - (-0.0011)). Table 3 depicts the Net-Benefit (RCRI score = 1 & RCRI score = 2) for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases overall.

Temporal External Validation

Baseline characteristics

We performed temporal external validation on eight databases. The database's counts of the target and outcome cohorts and outcome percentages per 6-month period can be seen in *Supplementary Table 3*. Furthermore, the outcome percentages per time window have been plotted in *Supplementary Fig. 2*. OptumEHR and STARR's outcome percentages rise substantially from 2023 onwards. Apart from this substantial rise, the outcome percentages are relatively consistent.

Trend of Area Under the Receiver Operating Characteristic Curve (AUROC) and Area Under the Precision-Recall Curve (AUPRC)

The trend of the AUROC and AUPRC for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases over 6-month time periods is depicted in Fig. 3AB. The figure shows a relatively stable discrimination over time, except for the IPCI database. However, in mixed effects model analysis, AUROC (excluding IPCI) shows a statistically significant ($p = 0.0099$) decrease of 0.0015. This corresponds to an AUROC decline of approximately 0.044 over a 15-year period. *Supplementary Fig. 3* shows the AUROC trend, including the standard deviations

Trend of calibration metrics

The trend of the difference between calibration in the large mean prediction and calibration in the large observed risk for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases over 6-month time periods is shown in Fig. 3C. Also, *Figs. 3DEF* depict the trend of Eavg, E90 and scaled Brier. These calibration metrics all show a relatively consistent trend, except for the scaled Brier in the IPCI database.

Trend of Net-Benefit

Figures 3GH depict the trend of the Net-Benefits (RCRI score = 1 & RCRI score = 2) for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases over 6-month time periods. These figures show a relatively stable trend of the Net-Benefit.

Discussion

We performed external validation and temporal external validation of different recalibrated versions of the Revised Cardiac Risk Index (RCRI) to evaluate whether the recalibration methods worked. The RCRI recalibrations have shown relatively consistent, yet poor, performance over time. The Duceppe recalibration's performance may be inferior to the oRCRI version. However, the MDCalc recalibration appears to be slightly superior to the oRCRI version.

Interestingly, our temporal external validation results show relatively stable trends when it comes to calibration and Net-Benefit. This is in contrast with what one would expect as the literature describes that shifts in healthcare data and practices over time can lead to decreased model performance.^{1-4,14} The mixed effects model analysis does, however, show a statistically significant decrease over time for AUROC of 0.044 over a 15-year period. Moreover, the MDCalc recalibration shows better calibration and Net-Benefit compared to the Duceppe recalibration. Firstly, the Duceppe model overestimates more, based on both the trend of calibration metrics in temporal external validation and the calibration plots and metrics in external validation. Secondly, the Duceppe recalibration never reaches a positive Net-Benefit value. In contrast, the MDCalc recalibration shows the most positive Net-Benefit values. The differences between the target and outcome cohort definitions between the Duceppe recalibration and oRCRI configuration cannot explain this as the cohort definitions used, favored the Duceppe recalibration cohorts.

Calibration remains relatively stable but consistently poor across recalibration methods. This might have to do with the changes across the surgical fields over time. For example, the transition from open to minimally invasive, e.g. laparoscopy, surgeries²³ or the increased use of intraoperative imaging and navigation systems.²⁴ It may also be related to variations in how hospitals record data or to changes in ICD coding over time. Apart from these possible explanations for RCRI's poor performance, more advanced updating, e.g. addition or deletion of predictors, could have improved the calibration.²⁵ In addition,

the performance of the RCRI model versions across databases is relatively consistent, except for IPCI across the entire temporal validation and for Optum® EHR and STARR databases after 2023. We can explain the fluctuations observed for the IPCI results by the different target cohort used and it being a General Practitioner Database, not covering surgeries. The STARR observed increase appears to reflect expanded documentation and repeated coding in the source data, rather than a true rise in incident Myocardial Infarction events. The STARR team is actively working with their hospital's Epic team to understand what workflow or regulatory changes contributed to this shift. We are unsure as to what might have caused Optum® EHR's fluctuations.

Based on our results, caution should be exercised when formulating or revising guideline recommendations for the use of prediction models. In 2014, the Guideline for Perioperative Cardiovascular Management for Noncardiac Surgery¹³ recommended to use a cut-off value of an RCRI score of more than 2. Conversely, our results show that the ranges of the three RCRI versions never (except for the MDCalc recalibration, partially) reach positive Net-Benefit values. The 2024 Guideline for Perioperative Cardiovascular Management for Noncardiac Surgery¹¹ recommendation's, however, cut-off value of an RCRI score of more than 1 does show positive ranges of the Net-Benefit. In addition, the difference in score-to-risk probability mappings impacts the Net-Benefit. In contrast, patients that identify as high risk are identical for each recalibrated version, as the mappings do not impact the RCRI score. This is potentially problematic for the calculation of the Net-Benefit for these score-to-risk probability mappings. The added value of the score-to-risk probability mappings may therefore decrease because of these limitations.

Strengths and limitations

We assessed RCRI performance by means of discrimination, calibration and Net-Benefit. This allows for a complete overview of the performance and is a strength of this study. Another strength is the number of included databases across four countries and the number of patients they represent. Conversely, temporal external validation requires a relatively high number of patients per time window. In one out of nine databases it was not possible to perform temporal external validation in time periods of 6 months, because of an insufficient number of patients per time window. Additionally, in three out of eight temporal external validation databases not all time windows allowed for temporal external validation. For three databases (CU_AMC, HULAFE & HIAE), no results could be included because of insufficient counts for the outcome cohorts. Another limitation is the target and outcome cohort variation between the oRCRI and Duceppe versions of the RCRI model.

Future work

Conducting a temporal external validation of models that use various updating strategies, such as removing or adding predictors, could provide valuable understanding of how these different approaches impact model performance.

Conclusion

While the RCRI recalibrations show consistent calibration and Net-Benefit, the significant drop in discrimination highlights challenges that extend beyond the RCRI itself. As underlying populations and care patterns change, even well-known clinical prediction models must be revisited and refined. These results highlight the value and limitations of updating by means of recalibration of clinical prediction models.

Declarations

Ethics approval and consent to participate

All patient data included in this study were deidentified. Participating databases obtained Institutional Review Board (IRB) approval based on a study protocol provided by A.S. Due to the large size of the study population, obtaining individual informed consent from all participants was not feasible and was therefore waived, where applicable, in accordance with local regulations. The New England Institutional Review Board determined that studies conducted in Optum/IBM data are exempt from study-specific IRB review, as these studies do not qualify as human subjects research. No experiments were conducted on humans in this study. The research methods were conducted in accordance with appropriate guidelines. The human data was collected in accordance with Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

This multinational retrospective cohort study was performed with aggregated results. Therefore, the database managers only shared results with us. However, interested parties could contact these database managers for further inquiries or data requests. The Optum, Merative CCAE, Merative MDCR, and Merative MDCD data that support the findings of this study are available from Merative MarketScan Research Databases and Optum but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Funding

A.S., J.R., P.R. and R.W. work for a research group at Erasmus University Medical Center that receives/received an unconditional grant for methodological research by Johnson & Johnson. The grant is for the institute

References

1. E S. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer 2009.
2. Steyerberg EMK, van der Windt D. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*. 2013;10(2):e1001381.
3. Vaid A, Suarez-Farinas ASM. Implications of the use of artificial intelligence predictive models in health care settings: a simulation study. *Ann Intern Med*. 2023;176(10):1358–69.
4. Castillo D. What is Covariate Shift. <https://www.seldon.io/what-is-covariate-shift>. Published 2021. Updated 08-07-2021. Accessed 2025.
5. Binuya MEE, Schats W, Schmidt M, Steyerberg E. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol* 2022;22.
6. Alexander Saelmans TS, Victor Pera A, Markus E, Fridgeirsson LJ, Lieke Schiphof-Godart, Peter Rijnbeek, Jenna Reps, Ross Williams Implementation and Updating of Clinical Prediction Models: A Systematic Review. *Mayo Clin Proc Digit Health*. 2025;3:100228.
7. Lee THME, Mangione CM, Thomas EJ, Polanczyk CA. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation*. 1999;100:1043–9.
8. Ford MKSBW, Wijeyesundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann Intern Med*. 2010;152:26–35.
9. Duceppe EPJ, MacDonald P, Lyons K, McMullen M. Canadian Cardiovascular Society Guidelines on Perioperative Cardiac Risk Assessment and Management for Patients Who Undergo Noncardiac Surgery. *Can J Cardiol* 2017;33.
10. Chan J. Revised Cardiac Risk Index for Pre-Operative Risk. *MDCalc*. <https://www.mdcalc.com/calc/1739/revised-cardiac-risk-index-pre-operative-risk>. Accessed.
11. Thompson AFK, Smilowitz NR, de las Fuentes L, Mukherjee D, AHA/ACC/ACS/. ASNC/HRS/ SCA/SCCT/SCMR/SVM Guideline for Perioperative Cardiovascular Management for Noncardiac Surgery. *JACC*. 2024;84(19).
12. Halvorsen SMJ, Cassese S, Hall TS, Abdelhamid M. 2022 ESC Guidelines on cardiovascular assessment and management of patients undergoing non-cardiac surgery. *European Heart Journal*. 2022;43:3826–3924.
13. Lee A, Fleisher KEF, Andrew D, Auerbach SAB, Joshua A, Beckman BB. 2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery. *Circulation* 2014;130.
14. Sahiner B, Samala WCRK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol* 2023;96.
15. Wong A. Understanding Model Drift and Its Impact on Health Care Policy. *JAMA Health Forum*. 2025;6:e252724.
16. Voss EAMR, Matcho A, Ma Q, Knoll C, Schuemie M. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inf Assoc*. 2015;22:553–64.
17. Gilbert JRG, Schuemie M, Ryan P, Weaver J, CohortDiagnostics. Diagnostics for OHDSI Cohorts. R package version 3.3.0. <https://github.com/OHDSI/CohortDiagnostics>. Published 2024. Accessed 2025.
18. Martin SS, Almarzooq AAZI, Anderson CAM. P Arora 2024 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association. *Circulation*. 2024;149.
19. NA. National Center for Health Statistics. Percentage of coronary heart disease for adults aged 18 and over, United States, 2019–2023. National Health Interview Survey. Generated interactively. https://wwwn.cdc.gov/NHISDataQueryTool/SHS_adult/index.html. Published 2025. Accessed 2025.
20. NA. EHDEN. <https://www.ehden.eu/>. Accessed 2025.
21. Blacketer C, Moinat MSM, Voss EA, Camprubi M, Rijnbeek PR, Ryan PB. Advancing Real-World Evidence Through a Federated Health Data Network (EHDEN): Descriptive Study. *J Med Internet Res* 2025(27).
22. Schuemie MRJ, Black A, Defalco F, Evans L. Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research. *IOS Press*. 2023;310:966–70.
23. Ace St John IC, Natalia S, Kubicki SM, Kavic. The Rise of Minimally Invasive Surgery: 16 Year Analysis of the Progressive Replacement of Open Surgery with Laparoscopy. *JSLs* 2020;24(4).
24. Andrea Teatini EP, Davit L, Aghayan RP, Kumar R, Palomar. Åsmund Avdem Fretland, Bjørn Edwin, Ole Jakob Elle. The effect of intraoperative imaging on surgical navigation for laparoscopic liver resection surgery. *Sci Rep*. 2019;9(1):18687.
25. Vergouwe Y, Oostenbrink DNR, Debray TPA, Murray GD, Kattan MW, Koffijberg H. A closed testing procedure to select an appropriate method for updating prediction models. *Statist Med*. 2017;36:4529–39.

Figures

**Index date:
Major elective
noncardiac surgery**

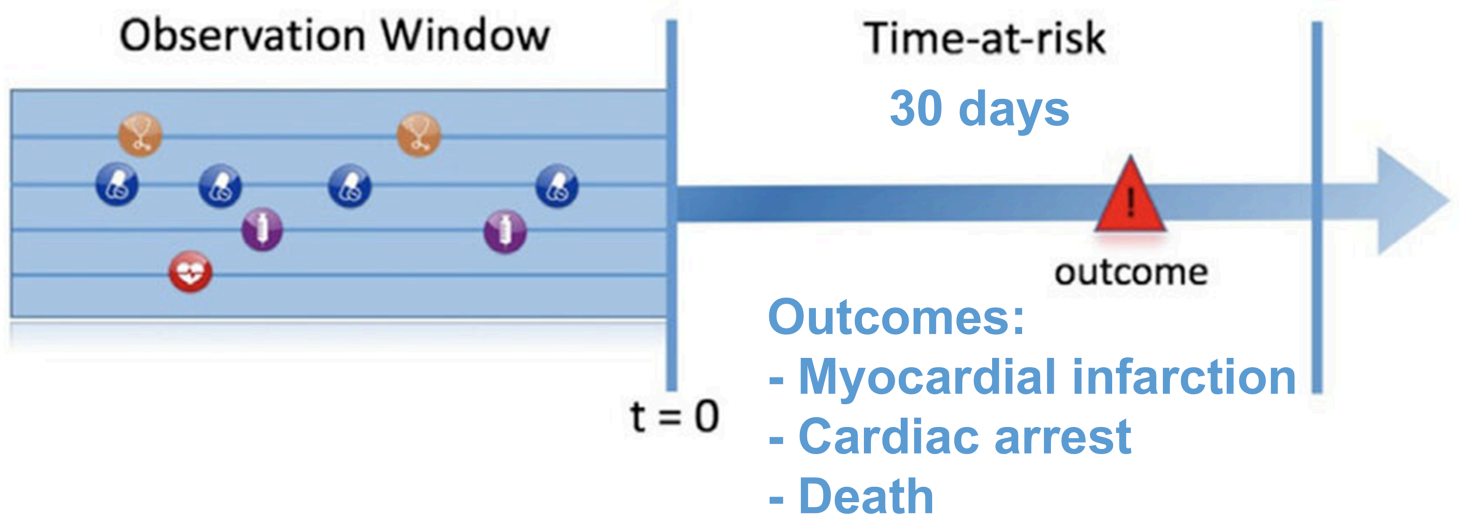


Figure 1

Original RCRI model (oRCRI) design

**Index date:
Major elective
noncardiac surgery**

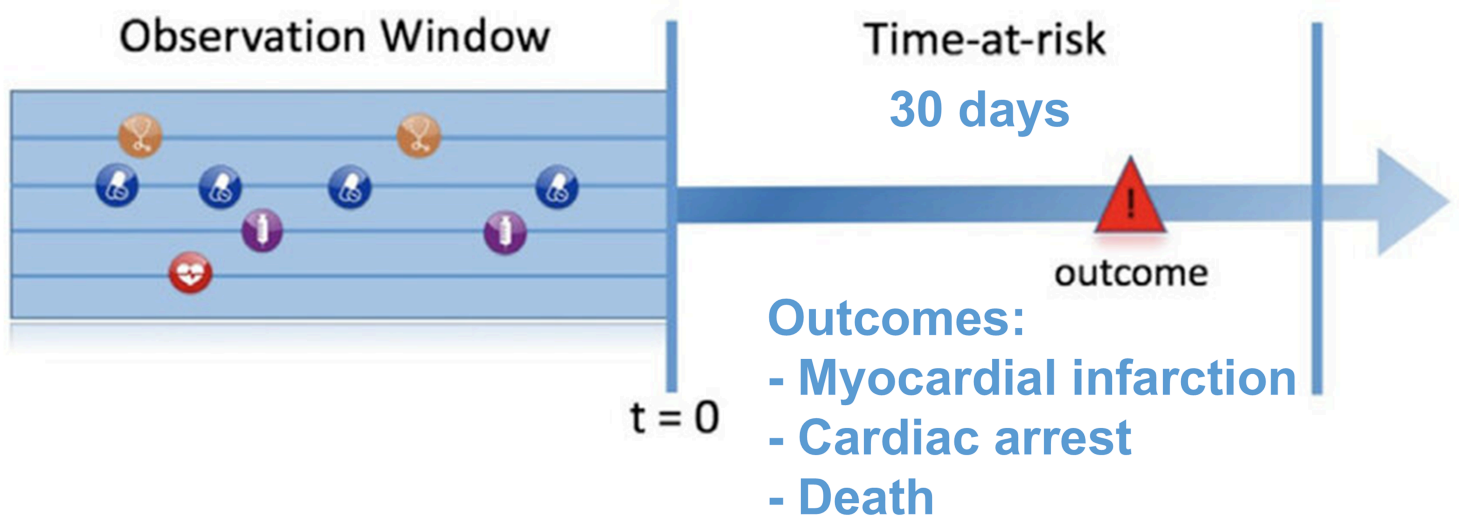


Figure 1

Original RCRI model (oRCRI) design

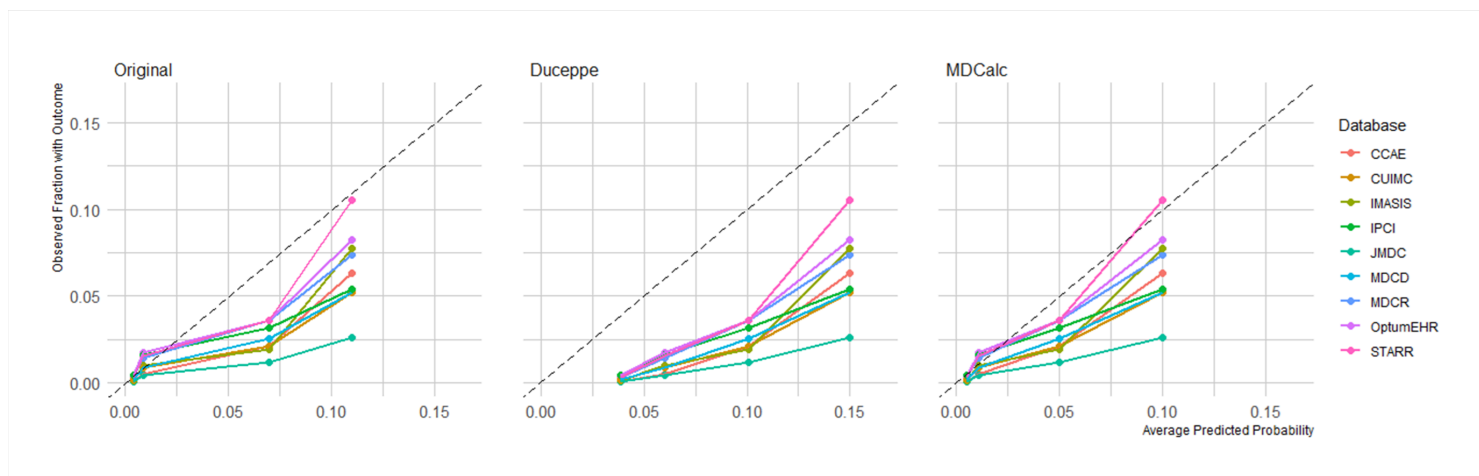


Figure 2

Calibration plots for external validation of the oRCRI, Dupeppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases

CCAE = Merative MarketScan® Commercial Claims and Encounters, MDCCD = Merative MarketScan® Medicaid Supplemental Database, MDCR = Merative MarketScan® Medicare Supplemental Database, JMDC = Japan Medical Data Center, STARR = STANford medicine Research data Repository-OMOP, CUIMC = Columbia University Irving Medical Center, IPCI = Integrated Primary Care Information, Optum® EHR = Optum® De-identified Electronic Health Record Dataset IMASIS = Parc de Salut Mar Barcelona Information System

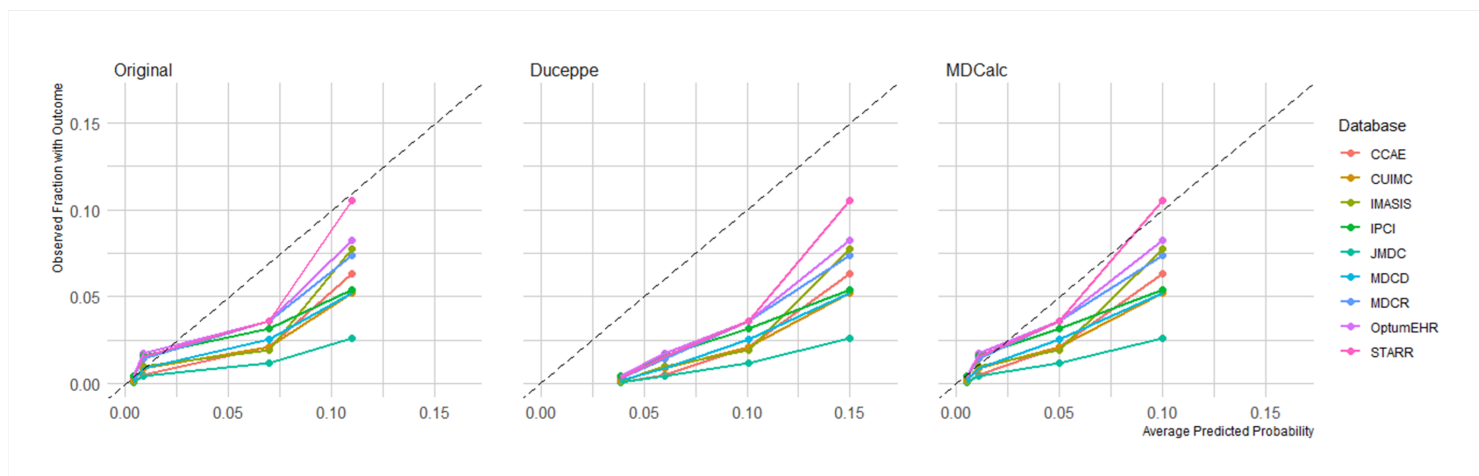


Figure 2

Calibration plots for external validation of the oRCRI, Dupeppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases

CCAE = Merative MarketScan® Commercial Claims and Encounters, MDCCD = Merative MarketScan® Medicaid Supplemental Database, MDCR = Merative MarketScan® Medicare Supplemental Database, JMDC = Japan Medical Data Center, STARR = STANford medicine Research data Repository-OMOP, CUIMC = Columbia University Irving Medical Center, IPCI = Integrated Primary Care Information, Optum® EHR = Optum® De-identified Electronic Health Record Dataset IMASIS = Parc de Salut Mar Barcelona Information System

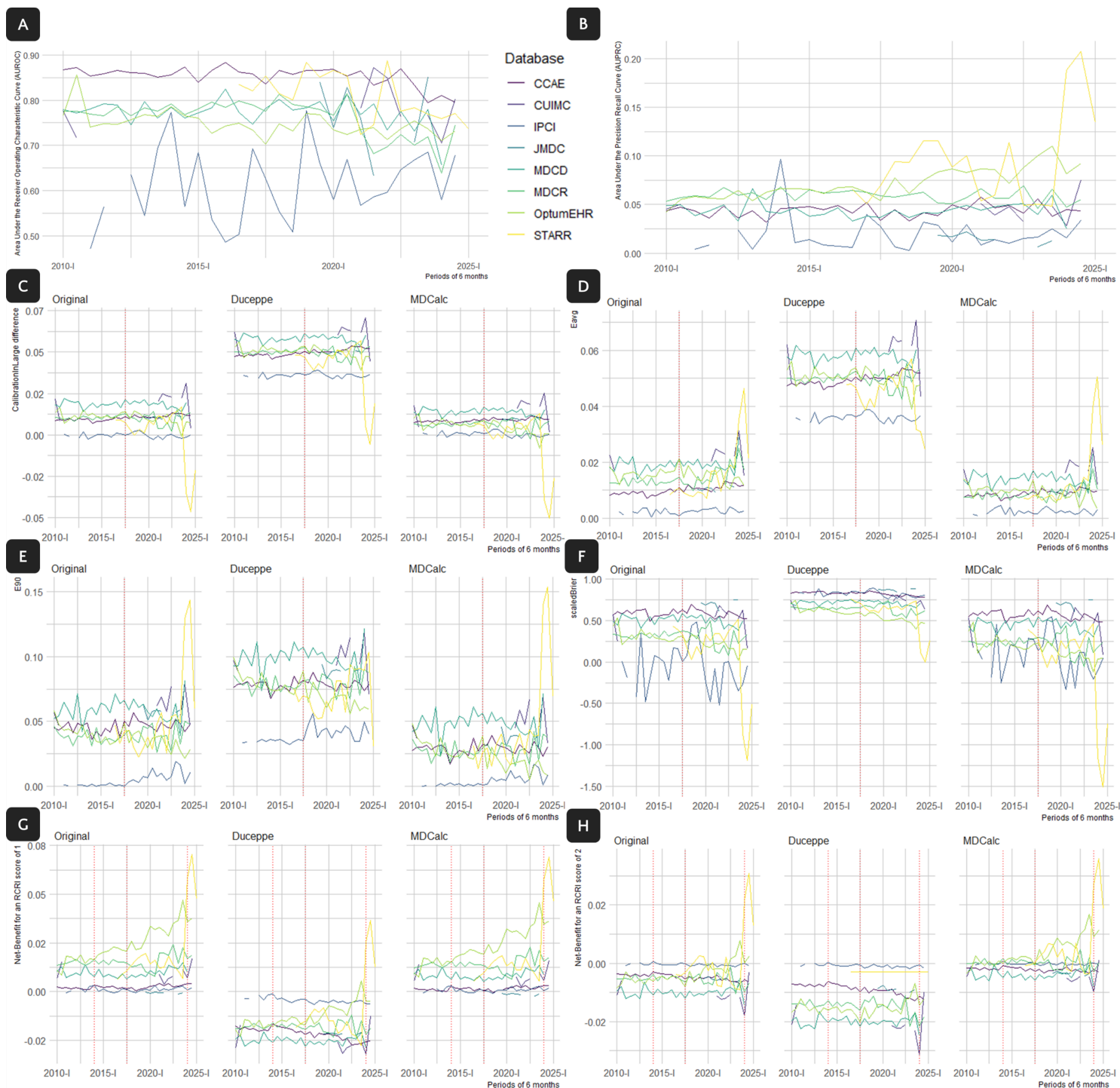


Figure 3

Trend of Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), difference between calibration in the large mean prediction and calibration in the large observed risk, Eavg, E90, scaled Brier and net-benefit (RCRI-score of 1) for oRCRI, Dupeppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases over 6-month time periods. **A** Area Under the Receiver Operating Characteristic Curve (AUROC)* **B** Area Under the Precision-Recall Curve (AUPRC)* **C** Difference between calibration in the large mean prediction and calibration in the large observed risk **D** Eavg **E** E90 **F** scaled Brier **G** Net-Benefit (RCRI-score of 1) **H** Net-Benefit (RCRI-score of 2). The dotted lines represent the 2014 guideline (cutoff of RCRI-score of 2), 2017 (Dupeppe recalibration) and 2024 guideline (cutoff of RCRI-score of 1). CCAE = Merative MarketScan® Commercial Claims and Encounters, MDCCD = Merative MarketScan® Medicaid Supplemental Database, MDCR = Merative MarketScan® Medicare Supplemental Database, JMDC = Japan Medical Data Center, STARR = STAnford medicine Research data Repository-OMOP, CUIMC = Columbia University Irving Medical Center, IPCI = Integrated Primary Care Information, Optum® EHR = Optum® De-identified Electronic Health Record Dataset

*As recalibration does not impact the AUROC and AUPRC values, for these values one graph was sufficient for the oRCRI, Dupeppe and MDCalc versions.

~Year-I (e.g. 2010-I) indicates the first 6 months of a year

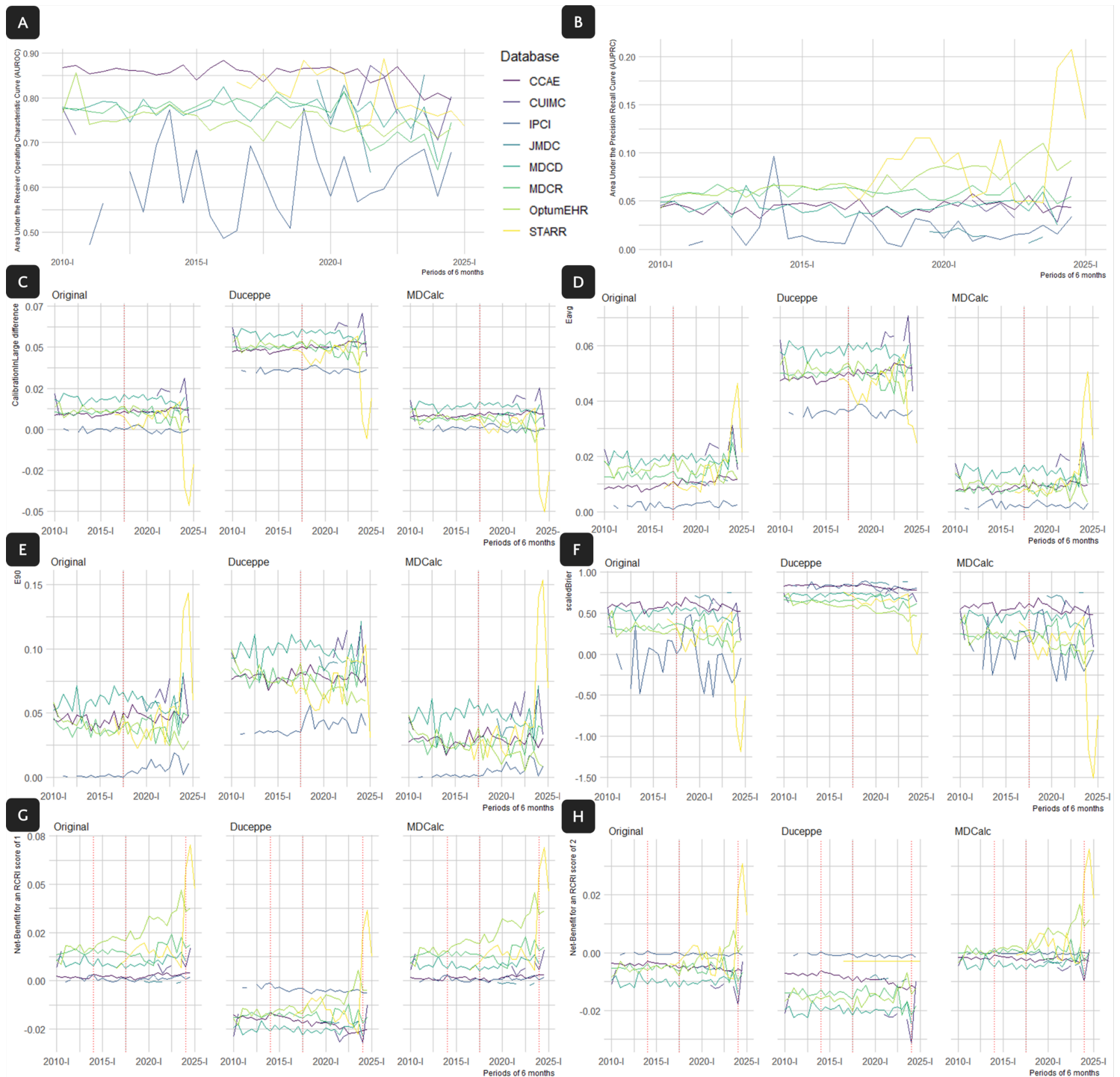


Figure 3

Trend of Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), difference between calibration in the large mean prediction and calibration in the large observed risk, Eavg, E90, scaled Brier and net-benefit (RCRI-score of 1) for oRCRI, Duceppe and MDCalc versions of the Revised Cardiac Risk Index (RCRI) across observational health databases over 6-month time periods. **A** Area Under the Receiver Operating Characteristic Curve (AUROC) **B** Area Under the Precision-Recall Curve (AUPRC) **C** Difference between calibration in the large mean prediction and calibration in the large observed risk **D** Eavg **E** E90 **F** scaled Brier **G** Net-Benefit (RCRI-score of 1) **H** Net-Benefit (RCRI-score of 2). The dotted lines represent the 2014 guideline (cutoff of RCRI-score of 2), 2017 (Duceppe recalibration) and 2024 guideline (cutoff of RCRI-score of 1). CCAE = Merative MarketScan® Commercial Claims and Encounters, MDCC = Merative MarketScan® Medicaid Supplemental Database, MDCR = Merative MarketScan® Medicare Supplemental Database, JMDC = Japan Medical Data Center, STARR = STAnford medicine Research data Repository-OMOP, CUIMC = Columbia University Irving Medical Center, IPCI = Integrated Primary Care Information, Optum® EHR = Optum® De-identified Electronic Health Record Dataset

*As recalibration does not impact the AUROC and AUPRC values, for these values one graph was sufficient for the oRCRI, Duceppe and MDCalc versions.

~Year-I (e.g. 2010-I) indicates the first 6 months of a year

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytable1IncidenceRatesPhenotyping.docx](#)
- [Supplementarytable1IncidenceRatesPhenotyping.docx](#)
- [Supplementarytable2Baselineexternalvalidation.docx](#)
- [Supplementarytable2Baselineexternalvalidation.docx](#)
- [Supplementarytable3Baselinetemporalvalidation.docx](#)
- [Supplementarytable3Baselinetemporalvalidation.docx](#)
- [SupplementaryFiguresRCRI.docx](#)
- [SupplementaryFiguresRCRI.docx](#)