

Temporal External Validation of Successive QRISK Updates across Observational Healthcare Databases: a Multinational Retrospective Cohort Study

Alexander Saelmans¹, Jenna M Reps^{1,2}, David Stables³, Peter R Rijnbeek¹, Ross D Williams¹

1 Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands

2 Johnson & Johnson, Titusville, NJ, USA

3 Endeavour Predict, Leeds, United Kingdom

Background

Clinical prediction models estimate an individual's future risk of an outcome based on multiple predictors derived from historical data. However, changes in healthcare data and practices over time can degrade model performance, making regular updates essential.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.⁵ Recalibration or revision, e.g. deletion, adjusts a model for novel data settings.^{5,6} Also, model extension entails incorporation of new predictors to the model. The optimal approach for updating, however, remains unclear.⁷

To partially bridge this knowledge gap, we investigate QRISK. QRISK is a cardiovascular disease prediction model that has been clinically implemented and updated in the United Kingdom (UK). The survival model predicts either coronary heart disease (CHD), stroke or transient ischemic attack (TIA) within ten years. Four versions⁸⁻¹¹ of the QRISK model are available, because the model has been updated over time. In the UK, QRISK is used to decide which treatment a patient should be offered to reduce their cardiovascular risk. The decision to prescribe a statin or not is aided by the QRISK score. For this decision, a cut-off risk percentage of ten percent is used for QRISK II, III and IV.^{12,13}

Model degradation over time is a known issue with clinically implemented models.^{14,15} A recent systematic review of model implementation showed few models have updating strategies.⁵ The optimal updating strategy for clinical prediction models is not well researched. A study investigating the simple recalibration updating strategy of RCRI found a significant decrease in discrimination.⁶ QRISK is an example of a clinically implemented model that has been updated via model extension. In this study we aim to quantify how much the model extension improved the model's performance over time on large observational healthcare databases.

Methods

We performed temporal external validation and external validation of QRISK versions 1-4 stratified by gender on ten observational health databases, including CPRD (Clinical Practice Research Datalink), a UK general practitioner database. The aim of this study is to evaluate whether the addition and deletion of predictors from QRISK, worked. (I) Temporal external validation was performed in 6-month intervals from 2007 to 2024, while (II) external validation overall covered the original QRISK version development periods.

We adjusted the target cohort with the index being the first outpatient visit per 6-month time window. The outpatient visit is only included as index after meeting original index requirements (birthday, registration with the practice and installation of the practice computer system). We believe this target cohort design will more closely resemble how the model is applied in the real-world. Figure 1 depicts this target design difference with patient-level examples for both original as temporal external validation.

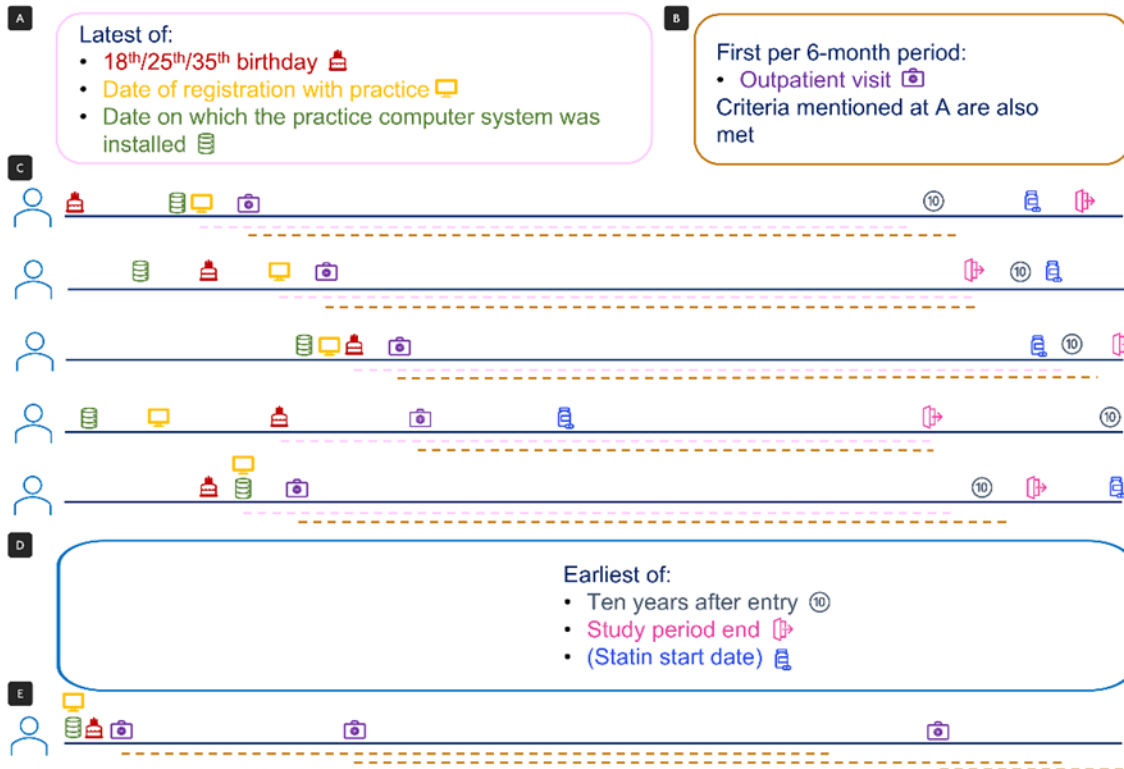


Figure 1. Depiction of target cohort design differences at a patient level

A Original target cohort inclusion criteria **B** Adjusted target cohort inclusion criteria for temporal external validation **C** Five example patients and accompanying timelines per target cohort group – Overall External Validation **D** Outcome cohort inclusion criteria **E** One example patient and accompanying timeline – Temporal External Validation

Apart from this adjustment, we replicated the target cohort exclusions in line with the original target cohorts. However, in the data we used, Townsend scores are not available. Therefore, we did not exclude patients if no Townsend score was available. We performed temporal external validation for ten models, consisting of five QRISK versions (I, IIa, IIb, III and IV) by gender. These ten adjusted target cohorts were specific to their corresponding QRISK version. In addition to these model-specific cohorts, we conducted further temporal validation using:

- QRISK 1 target cohorts (first version of QRISK, year: 2007, age range: 35-74, exclusion of diabetes mellitus patients)
- QR4 target cohorts (fourth version of QRISK, year: 2024, age range: 18-84, exclusion of patients who have been prescribed statins at cohort entry)

This enabled us to assess how broadening the age range influenced QRISK model performance.

We performed all statistical analyses in R 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) with HADES 20 packages “PatientLevelPrediction”, “CohortGenerator”, “CohortDiagnostics” and “CohortExplorer”. We also used the “survival” package for calculating the “C-statistic” as this is not included in “PatientLevelPrediction”. We created a package “QRISKvalidation” for this study to facilitate performing the analysis on various databases. (<https://github.com/ohdsi-studies/QRISKvalidation>)

Results

We performed temporal external validation on ten databases. Evaluating the different QRISK versions in the same study population indicated negligible impact of model updating on discrimination performance. QRISK model versions were developed using different study populations and incremental improvement were observed in discrimination across populations, as depicted in Figure 2. Also, calibration metrics showed overestimation of risk and the Net-Benefit was negative across all databases and versions. In addition, we performed external validation on ten databases with c-statistics approximating the development c-statistics in CPRD, calibration metrics that indicate an overestimation of risk and a negative Net-Benefit.

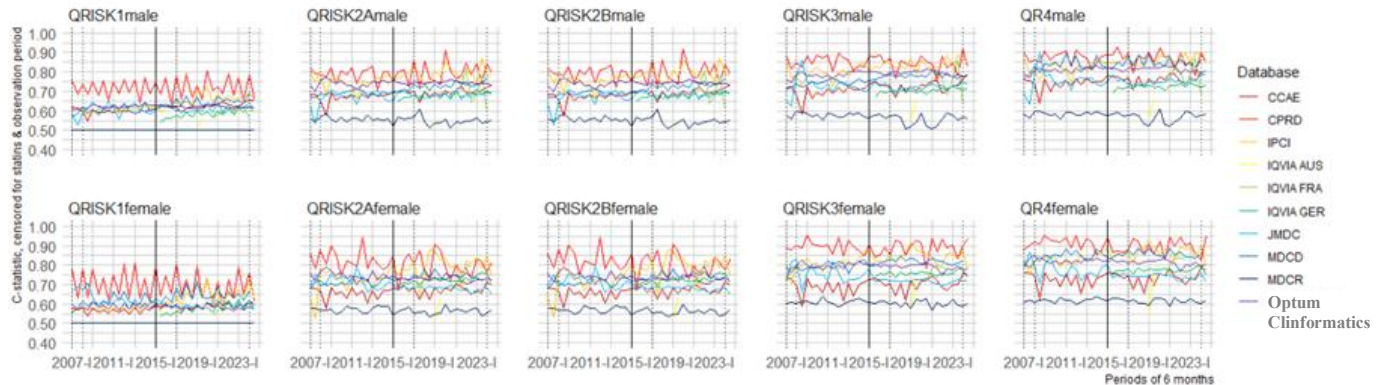


Figure 2. The trend of C-statistic, censored for both statins and observation period end, for the QRISK versions, QRISK I-IV, by gender for a time-at-risk of ten years across observational healthcare databases over 6-month time periods. **A** QRISK I ♂ **B** QRISK IIA ♂ **C** QRISK IIB ♂ **D** QRISK III ♂ **E** QR IV ♂ **F** QRISK I ♀ **G** QRISK IIA ♀ **H** QRISK IIB ♀ **I** QRISK III ♀ **J** QR IV ♀ The black line represents the last time period with a full 10 year time-at-risk window. The dotted lines represent the publication of the different QRISK versions (I: 2007, II: 2008, III: 2015 and IV: 2024)

CPRD = Clinical Practice Research Datalink, IPCI = Integrated Primary Care Information, IQVIA GER = IQVIA Disease Analyzer Germany, IQVIA FRA = IQVIA Disease Analyzer France, IQVIA AUS = IQVIA Disease Analyzer Australia, Optum Clinformatics® = Optum's de-identified Clinformatics® Data Mart Database, CCAE = Merative MarketScan® Commercial Claims and Encounters, MDCC = Merative MarketScan® Medicaid Supplemental Database, MDCR = Merative MarketScan® Medicare Supplemental Database, JMDC = Japan Medical Data Center

~Year-I (e.g. 2007-I) indicates the first 6 months of a year

Conclusion

QRISK is an example of a model that has been updated via extension. However, the results suggest these updates were due to expanding the target population rather than a need to adjust for model degradation. Interestingly, the models appear to be stable in performance within target populations.

References

1. E S. Clinical prediction models: a practical approach to development, validation, and updating. Springer2009.
2. Steyerberg E MK, van der Windt D. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med. 2013;10(2):e1001381.
3. Vaid A SA, Suarez-Farinas M. Implications of the use of artificial intelligence predictive models in health care settings: a simulation study. Ann Intern Med. 2023;176(10):1358-1369.
4. D C. What is Covariate Shift. <https://www.seldon.io/what-is-covariate-shift>. Published 2021. Updated 08-07-2021. Accessed.

5. Alexander Saelmans TS, Victor Pera, Aniek Markus, Egill Fridgeirsson, Luis John, Lieke Schiphof-Godart, Peter Rijnbeek, Jenna Reys, Ross Williams Implementation and Updating of Clinical Prediction Models: A Systematic Review. *Mayo Clin Proc Digit Health*. 2025;3:100228.
6. Saelmans A RJ, Minty E, Leis Machin A, Mayer MA. Temporal External Validation and External Validation of an Implemented and Updated Binary Prediction Model, the Revised Cardiac Risk Index (RCRI), on Observational Health Databases: a Multinational Retrospective Cohort Study. 2026.
7. Binuya M EE, Schats W, Schmidt M, Steyerberg E. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Medical Research Methodology*. 2022;22.
8. Hippisley-Cox J CC, Vinogradova Y, Robson J, May M. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ*. 2007.
9. Hippisley-Cox J CC, Vinogradova Y, Robson J, Minhas R. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*. 2008;336.
10. Hippisley-Cox J CC, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ*. 2017;357.
11. Hippisley-Cox J CC, Bafadhel M, Russell REK, Sheikh A. Development and validation of a new algorithm for improved cardiovascular risk prediction. *Nature Medicine*. 2024:1440-1447.
12. Summary of national guidance for lipid management for primary and secondary prevention of cardiovascular disease. <https://www.england.nhs.uk/aac/publication/summary-of-national-guidance-for-lipid-management/>. Published 2023. Accessed.
13. Cardiovascular Risk Score (QRISK2) Patient Information Leaflet. <https://www.winchmorehillpractice.nhs.uk/2021/09/08/cardiovascular-risk-score-qrisk2-patient-information-leaflet/>. Published 2021. Accessed.
14. Sahiner B CW, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol*. 2023;96.
15. Wong A SJ. Understanding Model Drift and Its Impact on Health Care Policy. *JAMA Health Forum*. 2025;6:e252724.