Drug Safety (2022) 45:563–570
https://doi.org/10.1007/s40264-022-01161-8

ORIGINAL RESEARCH ARTICLE

Using Iterative Pairwise External Validation to Contextualize Prediction Model Performance: A Use Case Predicting 1-Year Heart Failure Risk in Patients with Diabetes Across Five Data Sources

Ross D. Williams[1] · Jenna M. Reps[2] · Jan A. Kors[1] · Patrick B. Ryan[2] · Ewout Steyerberg[3] · Katia M. Verhamme[1] · Peter R. Rijnbeek[1]

Accepted: 9 February 2022

Abstract

Introduction External validation of prediction models is increasingly being seen as a minimum requirement for acceptance in clinical practice. However, the lack of interoperability of healthcare databases has been the biggest barrier to this occurring on a large scale. Recent improvements in database interoperability enable a standardized analytical framework for model development and external validation. External validation of a model in a new database lacks context, whereby the external validation can be compared with a benchmark in this database. Iterative pairwise external validation (IPEV) is a framework that uses a rotating model development and validation approach to contextualize the assessment of performance across a network of databases. As a use case, we predicted 1-year risk of heart failure in patients with type 2 diabetes mellitus.

Methods The method follows a two-step process involving (1) development of baseline and data-driven models in each database according to best practices and (2) validation of these models across the remaining databases. We introduce a heatmap visualization that supports the assessment of the internal and external model performance in all available databases. As a use case, we developed and validated models to predict 1-year risk of heart failure in patients initializing a second pharmacological intervention for type 2 diabetes mellitus. We leveraged the power of the Observational Medical Outcomes Partnership common data model to create an open-source software package to increase the consistency, speed, and transparency of this process.

Results A total of 403,187 patients from five databases were included in the study. We developed five models that, when assessed internally, had a discriminative performance ranging from 0.73 to 0.81 area under the receiver operating characteristic curve with acceptable calibration. When we externally validated these models in a new database, three models achieved consistent performance and in context often performed similarly to models developed in the database itself. The visualization of IPEV provided valuable insights. From this, we identified the model developed in the Commercial Claims and Encounters (CCAE) database as the best performing model overall.

Conclusion Using IPEV lends weight to the model development process. The rotation of development through multiple databases provides context to model assessment, leading to improved understanding of transportability and generalizability. The inclusion of a baseline model in all modelling steps provides further context to the performance gains of increasing model complexity. The CCAE model was identified as a candidate for clinical use. The use case demonstrates that IPEV provides a huge opportunity in a new era of standardised data and analytics to improve insight into and trust in prediction models at an unprecedented scale.

✉ Ross D. Williams
r.williams@erasmusmc.nl

[1] Department of Medical Informatics, Erasmus MC, University Medical Center Rotterdam, Doctor Molewaterplein 40, 3015 GD Rotterdam, The Netherlands

[2] Janssen Research and Development, Titusville, NJ, USA

[3] Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

Key Points

External validation lacks context, which inhibits understanding of model performance.

Iterative pairwise external validation provides contextualised model performance across databases and across model complexity.

1 Introduction

External validation has been identified as an essential aspect of the development of clinical prediction models and a key part of the evidence-gathering process needed to create impactful models that are adopted in the clinic [1]. Currently, the majority of prediction models are not externally validated; where they are, they are poorly reported [2].

A major issue preventing the external validation of models is the lack of interoperability of healthcare databases [3]. There are two main problems to solve: databases use different coding systems (e.g. *International Classification of Diseases, Tenth Revision* and Systemized Nomenclature of Medicine—Clinical Terms) and have different structures [4]. A solution to this is to (1) convert each database into a common format to improve syntactic interoperability and (2) standardize to common vocabularies to improve semantic interoperability.

Standardization of the format and vocabulary of these databases allows for the development of standardized tools and a framework for conducting prediction research [5, 6]. Using these standard tools and conducting research according to open science principles [7] removes many difficulties associated with externally validating prediction models. Some challenges remain, including the interpretation of results in the context of the new database. Furthermore, important privacy concerns often need to be respected in the development process [8]. For example, many data owners are unable to share patient-level data, so any development process must be able to cater for this [9].

1.1 Performance Contextualization

Traditionally, a prediction model is trained on one database using predictors selected by domain experts, and this model is then validated on other databases [10, 11]. These models often consist of a limited number of predictors [12]. Recently, data-driven approaches have been used to leverage all the information in electronic health records (EHRs), which can result in models with many predictors. The question is, how do we decide whether the model works well in other databases? For this, the standard approach is to compare the discriminative performance and model calibration with the performance obtained on the training data [13–15]. Any performance drop could be because the model was too tuned to the training data to properly transport to unseen data, i.e. the model was overfit or needs recalibration. However, the performance achieved could also be similar to that of a model that is trained on that same database. In other words, the model performs as well as possible in the context of the available data in that database. We need a model development approach that provides this context. Furthermore, simpler models are preferred as they are easier to implement, and—as such—understanding the performance gain compared with the baseline of using only age and sex is valuable to contextualize the performance of the more complex model [16, 17].

In this article, we introduce iterative pairwise external validation (IPEV), a framework to better contextualise the performance of prediction models, and demonstrate its value when developing and validating a prediction model in a network of databases. The use case for this model is the prediction of the 1-year risk of heart failure (HF) following the initialisation of a secondary drug to treat type 2 diabetes mellitus (T2DM). As described in detail in a literature review [18], the pathophysiological connection between diseases and their frequent adverse interactions should affect treatment choice [19]. The 2019 American Diabetes Association guidelines [20] recommend that patient treatments should be stratified according to an established or high risk of HF. Specifically, the guidelines state that thiazolidinediones should be avoided in patients with HF and that sodium–glucose co-transporter-2 inhibitors are preferred in patients at high risk of HF. The guidelines appear to be trending towards a more personalized treatment strategy [21, 22]. As such, there is an opportunity to use risk prediction to further personalize treatment in the intermediate steps before treatment with insulin. This use case presents the opportunity to both evaluate IPEV and simultaneously create a potentially clinically impactful model.

2 Methods

2.1 Analysis Methods

2.1.1 Iterative Pairwise External Validation

IPEV is a new model development and validation procedure that involves a two-step procedure. The first step is to create two models per database: a model with only age and sex as covariates, which serves as a baseline for what a simple model can achieve, and a more complex data-driven model that assesses the maximum achievable performance. The second step is then validating these models, both internally and externally, in the other databases. A diagram of this process can be seen in Fig. 1.

2.1.2 Candidate Covariates

Two sets of covariates are used to develop models. One set consists of only age and sex and is used to create a baseline model. The other set is used to build a more complex data-driven model and consists of age, sex, and binary variables

indicating the presence or absence of comorbidity (based on presence of disease codes) any time prior to index and of procedures and drugs that occurred in the year prior to index date. The binary variables constructed are for any condition, procedure, or drug in the patient's history. For example, if a diagnosis of liver failure is recorded in a patient's medical records prior to the index date, then we create a candidate binary variable named 'liver failure any time prior' that has a value of 1 for patients with a record of liver failure in their history and 0 otherwise.

The use of these two sets of covariates shows the achievable performance for a simple set of covariates that can then be used to assess any added value of a more complex model. This gives a context to the performance gains relative to the increased model complexity.

2.1.3 Evaluation Analysis

For performance analysis, we consider the area under the receiver operating characteristic curve (AUC) as a measure of discrimination. An AUC of 0.5 corresponds to a model randomly assigning risk, and an AUC of 1 corresponds to a model that can perfectly rank patients in terms of risk (assigns higher risk to patients who will develop the outcome compared with those who will not). For calibration assessment, we use calibration graphs and visually assess whether the calibration is deemed sufficient.

2.2 Proof of Concept

Predicting the 1-year risk of developing HF following initiation of a second pharmaceutical treatment for T2DM was selected as a proof of concept. This case study could help inform treatment decisions by comparing an individual patient's risk of HF with the known safety profiles of the different medications.

2.2.1 Data Sources

The analyses were performed across a network of five observational healthcare databases. All databases contained either claims or EHR data from the USA and were transformed into the Observational Medical Outcomes Partnership common data model (OMOP CDM), version 5 [23].

Table 1 describes the databases included in this study. The complete specification for the OMOP CDM, version 5, is available at https://ohdsi.github.io/CommonDataModel/cdm531.html.

2.2.2 Cohort Definitions

2.2.2.1 Target Cohort The target population consisted of patients with T2DM who were treated with metformin and who became new adult users of one of sulfonylureas, thiazolidinediones, dipeptidyl peptidase-4 inhibitors, glucagon-like peptide-1 receptor agonists, or sodium-glucose co-transporter-2 inhibitors. The index date was the first prescription of one of these secondary treatments. We required all subjects to have a T2DM diagnosis based on the presence of a disease code and use of metformin prior to the index date. Patients with HF or patients treated with insulin on or prior to the index date were excluded from the analysis. Patients were required to have been enrolled for at least 365 days before cohort entry.

2.2.3 Outcome Definitions

The outcome was defined using the presence of a diagnosis code of HF occurring for the first time in the patient's history between 1 and 365 days post index.

The cohort definition is available at https://github.com/ohdsi-studies/PredictingHFinT2DM/tree/main/validation/inst/cohorts.

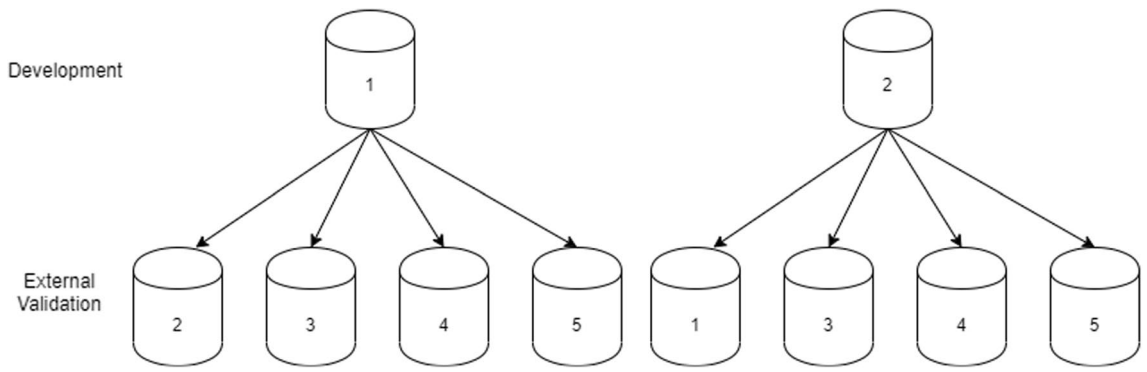

Fig. 1 Rotation of databases for model development and external validation in the iterative pairwise external validation method

Table 1 Database characteristics

Database	Acronym	Country	Data type	Time period	Database size (million patients)
Optum® de-identified EHR dataset	Optum EHR	USA	EHR	2006–2018	87
IBM MarketScan® commercial database	CCAE	USA	Claims	2000–2018	155
IBM MarketScan® multi-state Medicaid database	MDCD	USA	Claims	2006–2017	30
IBM MarketScan® Medicare supplemental database	MDCR	USA	Claims	2000–2018	10
Optum® de-identified Clinformatics® data mart database	Optum Clinformatics	USA	Claims	2000–2018	98

EHR electronic health record

The study period contained data from 2000 to 2018. The exact period varies between the databases and is available in Table 1.

2.2.4 Covariates

In total, we derived around 39,000 candidate covariates. These included more than 26,000 conditions, 13,000 procedures and drugs, and demographic information.

2.2.5 Statistical Analysis

Model development followed the framework for the creation and validation of patient-level prediction models presented in Reps et al. [5]. We used a 'train–test split' method to perform internal validation. In each target population cohort, a random sample of 75% of the patients ('training sample') was used to develop the prediction model, and the remaining 25% of the patients ('test sample') was used to internally validate the prediction model developed.

We used regularized logistic regression risk models, also known as least absolute shrinkage and selection operator. Regularisation is a process to limit overfitting in model development. This process works by assigning a 'cost' to the inclusion of a variable, and the variable must contribute more to the model performance than this cost in order to be included. If this condition is not met, the coefficient of the covariate becomes 0, which therefore eliminates the covariate from the model, providing an in-built feature selection [24].

2.2.6 Open-Source Software

We used the PatientLevelPrediction R-package (version 4.0.1) and R (v4.0.2) to perform all analyses. All development analysis code and cohort definitions are available at https://github.com/ohdsi-studies/PredictingHFinT2DM. The validation package is available at https://github.com/ohdsi-studies/PredictingHFinT2DM/tree/main/validation.

3 Results

Across all databases, we selected 403,187 patients with T2DM initiating second-line treatment. Of these, 12,173 developed HF during the 1-year follow-up. Next, we performed patient-level prediction of HF. The number of patients and the AUCs are given in Table 2.

The AUC results, as shown in Fig. 2, show reasonable performance. The main diagonal of the heatmaps indicates the internal validation. All other results are from external validation. The mean AUCs across internal and external validation were 0.78 (Commercial Claims and Encounters [CCAE]), 0.76 (IBM MarketScan® multi-state Medicaid database), 0.76 (IBM MarketScan® Medicare supplemental database [MDCR]), 0.78 (Optum Clinformatics), and 0.78 (Optum EHR). The best performing models in terms of discrimination were developed in CCAE, Optum Clinformatics, and Optum EHR and appeared to be the most consistent across the external validations. When comparing the baseline model, consisting of only age and sex, with the full model, the performances dropped. For example, for CCAE, the data-driven model achieved 0.78 compared with the baseline model of 0.64 for CCAE and 0.80 (data driven) and 0.69 (baseline) for Optum Clinformatics.

Of note, models externally validated in the MDCR dataset consistently outperformed the model that was developed there. This occurred for the data-driven model (internal 0.73), with the external validation of CCAE, Optum Clinformatics, and Optum EHR achieving 0.75, 0.76, and 0.74, respectively.

We assessed the calibration of the three models with the best discrimination (CCAE, Optum Clinformatics, and Optum EHR). The calibration results from these three models across the external validations are shown in Fig. 3. The models generally appear to be well calibrated.

Concerning the best model produced, the CCAE and Optum Clinformatics developed models had the best discrimination

Table 2 Number of patients and internal validation performance per database

Database	Patients with T2DM (*n*)	Patients with HF (*n*)	Incidence (%)	Age, years mean ± SD	Female (%)	Full model AUC	Age Sex AUC
CCAE	112,989	1843	1.6	53 ± 8	46	0.78	0.64
MDCD	15,860	650	4.1	50 ± 12	64	0.77	0.65
MDCR	22,433	1658	7.4	73 ± 6	48	0.73	0.64
Optum Clinformatics	92,272	4332	4.7	63 ± 13	48	0.80	0.69
Optum EHR	159,633	3690	2.3	58 ± 12	49	0.81	0.71

AUC area under the concentration–time curve, *CCAE* Commercial Claims and Encounters, *EHR* electronic health record, *HF* heart failure, *MDCD* Medicaid, *MDCR* Medicare, *SD* standard deviation, *T2DM* type 2 diabetes mellitus

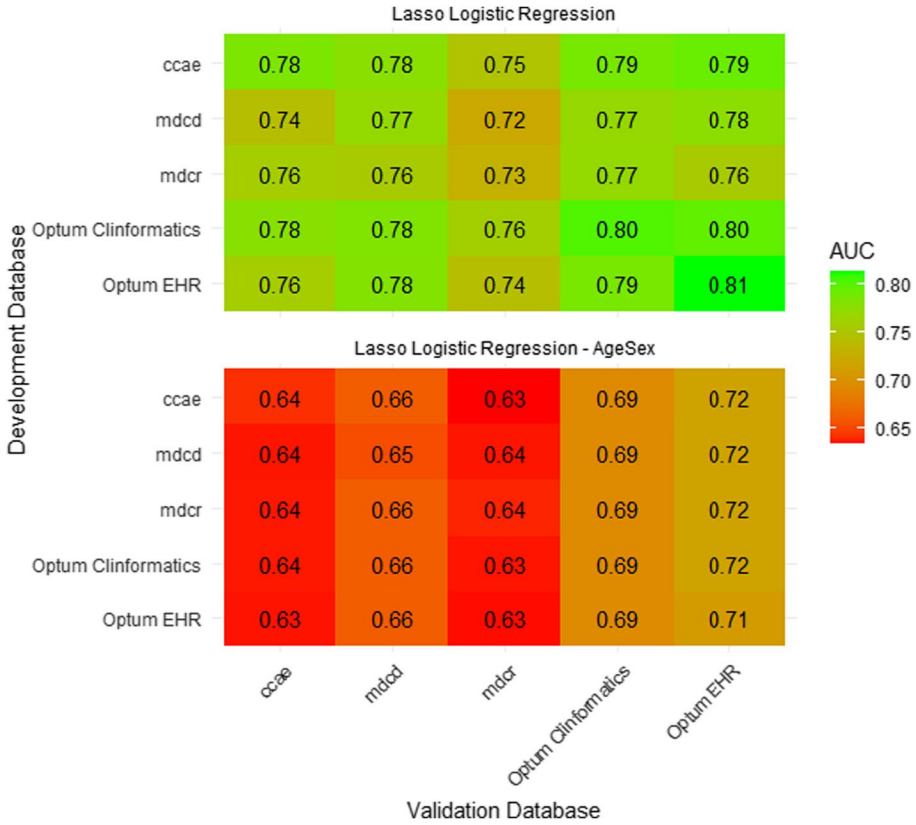

Fig. 2 A heatmap of the area under the concentration–time curve values across internal validation (values on the lead diagonal) and external validations of the developed prediction models. The colour scale runs from red (low discriminative ability) to green (high discriminative ability). The upper section details the performances for the data-driven model. The lower half details the same but then for the age and sex model. *AUC* area under the concentration–time curve, *ccae* Commercial Claims and Encounters, *EHR* electronic health records, *mdcd* Medicaid, *mdcr* Medicare

performance. The CCAE model contained 195 covariates, compared with 413 for Optum Clinformatics, so it is preferred. The names and coefficients of the covariates in the CCAE model are available in Appendix 1 in the electronic supplementary material (ESM).

For the CCAE-developed model, demographic plots are provided in the ESM. These plots show the calibration of the model stratified by sex across age groups.

All results are available in a study application located at https://data.ohdsi.org/PredictingHFinT2DM/.

4 Discussion

This study demonstrates the use of IPEV for model development and external validation. External validation of a prediction model has traditionally lacked any contextual information on what the expected performance in the database should be. By including a baseline and data-driven model developed in each database, context can be added to the performance of a model externally validated in this database.

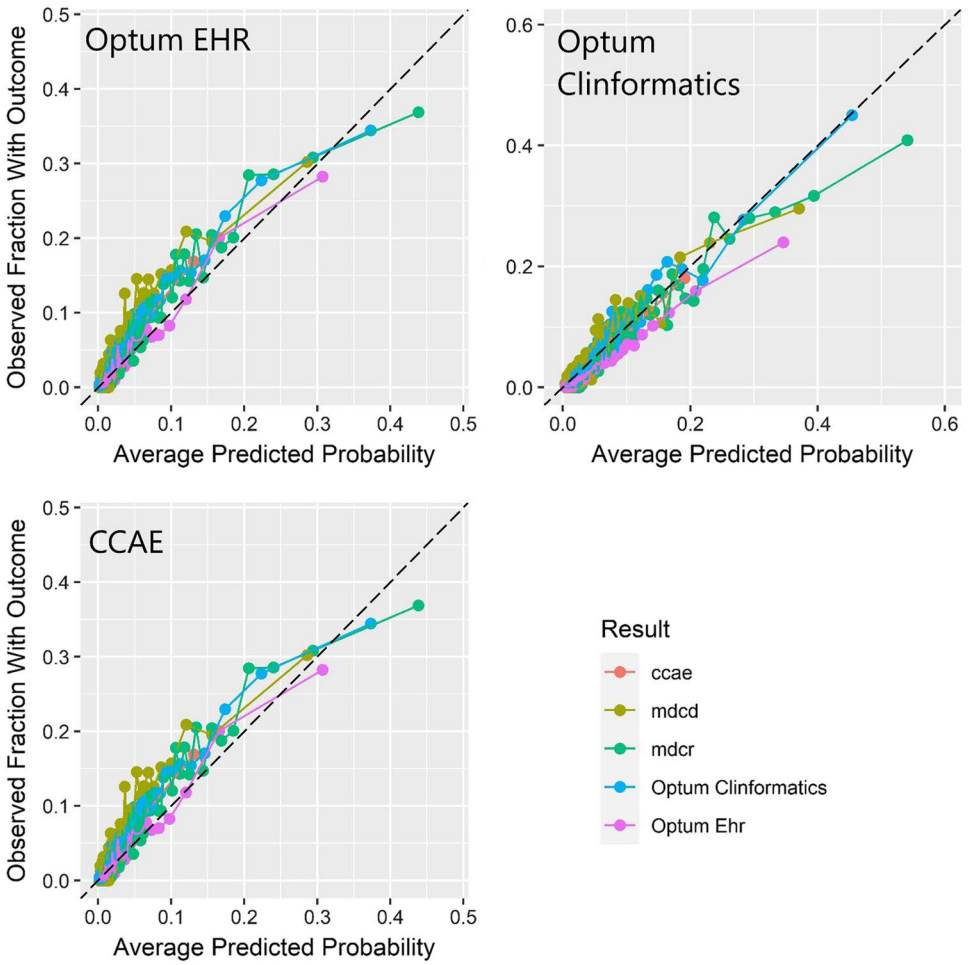

Fig. 3 Internal and external calibration of the Optum EHR, Optum Clinformatics, and CCAE trained models. *CCAE* Commercial Claims and Encounters, *EHR* electronic health records, *mdcd* Medicaid, *mdcr* Medicare

Recent improvements in database interoperability and standardisation of tools made it possible to utilise IPEV to develop and contextually validate models for predicting HF in T2DM. This contextual validation provides a more rigorous approach to model assessment. For example, where a model's performance drops from training to external validation but achieves performance consistent with expectations in the external validation database, this then raises the question of what the difference is between the two databases. Similarly, if a model achieves a lower performance than expected in a new database, this can be interpreted as overfitting to training data.

The inclusion of a baseline model (using only age and sex covariates) in each training step provides context to the performance gain from increasing model complexity. By comparing the more complex model with this baseline model, a better assessment of complexity–performance trade-off can be made to analyse the potential for clinical implementation. If a large disparity in performance between these two models is observed, a parsimonious model (of around ten variables) could be created to attempt to bridge the gap between the performance of the complex model and the ease of implementation of the baseline model. The interpretation of the results is aided by the inclusion of a heatmap. This allows for easy visual inspection of performance across external validations. Once differences in performance across external validation have been demonstrated, it would be interesting to investigate the case mix of the cohorts in the database as well as the prevalence of the predictors to better understand these performance differences [25].

Considering the specific use case, the performance of the CCAE model developed in this paper suggests it could be used in treatment planning. This model had good discriminative performance that was consistent across external validations (AUC internal 0.78, external 0.75–0.79). There was a minor loss in discrimination for some of the external validations, for example MDCR had the lowest AUC (0.75). This lower performance was in line with the internal validation of the database, and MDCR had the worst performance across all the external validations, suggesting it is a more problematic dataset in which to make predictions. Possible explanations for this are that the underlying case mix of patients could make discrimination more difficult. For

example, patients in this database are generally older, so it could become more difficult to separate them; there is also little to no overlap in the ages of patients between CCAE and MDCR. Another reason could be that the lower numbers of patients might mean data are insufficient to provide a reliable estimate or to develop the optimal model. Specifics of performance in different demographics are available in the Shiny R package. The model showed reasonable calibration across internal and external validations, with some overestimation of risk for patients at higher risk. The Optum EHR external validation showed a larger miscalibration and could benefit from some recalibration before implementation. When we compared the data-driven and baseline models, the performance of the latter across all the validations for all models was only moderate and often produced a drop of 0.1–0.2 AUC, demonstrating that the increase in complexity provided significant performance gains. Age and sex alone were insufficient to accurately predict future HF, and more complex models are needed.

Calibration is important when using a model for clinical decision making, and this result highlights that our model likely requires recalibration when applied to case mixes that differ from the development database.

The model could be implemented at either the treatment facility or the health authority level. Using the previously discussed American Diabetes Association treatment guidelines, the use of a risk model to stratify patients can be impactful, and the evidence generated in this paper suggests that the CCAE-developed model could be a candidate for clinical use. If patients can be assessed on their risk of HF, their treatment can be personalised, helping to prevent medication switching or the addition of new medicines to treat HF when there are diabetes treatments with known beneficial HF effects. To our knowledge, this is the only model available in open source that can be used for this specific prediction problem.

This method is scalable and can be expanded to use more databases as they become available. An example is through the European Health Data and Evidence Network (EHDEN) project, which is currently standardizing 100 databases to the OMOP CDM. This network could be leveraged to provide context to the external validation of prediction models at an unprecedented scale. This would lead to improved models, stronger evidence, and a bigger clinical impact. When considering the case of a federated data network such as EHDEN, IPEV is particularly suitable. As privacy concerns prevent the sharing of patient-level data, a development and validation process that does not require this is necessary. IPEV incorporates 'privacy by design', whereby research can be performed by separate researchers at separate locations without sharing patient data. This is a major advantage as it maintains the ability to produce excellent and clinically impactful research without introducing any new privacy or security concerns. This means that the method can be used under the standard procedures of obtaining institutional review board approval, while maintaining data security and improving the quality of research without significantly burdening researchers.

A limitation of this method is that it does not use the full data available for training. There is evidence to suggest that combining data can improve internal validation. However, this requires researchers to share data and violates data privacy concerns. Further, methods such as federated learning are compatible with IPEV. If a researcher is particularly concerned with improving the performance of the developed model, they could combine $n-1$ databases and test in the nth, then rotate through development using IPEV, leaving out one database at a time, simultaneously increasing the data available for training and maintaining external validity.

5 Conclusion

Using IPEV lends weight to the model development process. The rotation of development through multiple databases provides context, allowing for thorough analysis of performance. The inclusion of a baseline model in all modelling steps provides further context to the performance gains of increasing model complexity. IPEV provides a huge opportunity in a new era of standardised data and analytics to improve insights into and trust in prediction models on an unprecedented scale.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40264-022-01161-8.

Declarations

Funding This project received support from the EHDEN project. EHDEN received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement no. 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations.

Conflicts of interest Ross D. Williams, Jan A. Kors, and Ewout Steyerberg have no conflicts of interest that are directly relevant to the content of this article. Katia M. Verhamme and Peter R. Rijnbeek work for a research group that has received unconditional research grants from Boehringer-Ingelheim, GSK, Janssen Research & Development, Novartis, Pfizer, Yamanouchi, and Servier. Jenna M. Reps and Patrick B. Ryan are employees and shareholders of Janssen Research & Development and shareholders of Johnson & Johnson.

Availability of data and material Ross D. Williams is responsible for the data. Privacy concerns mean the patient-level data for the study are not available. All results are available at https://data.ohdsi.org/PredictingHFinT2DM/.

Code availability All code used in the study is provided open source at https://github.com/ohdsi-studies/PredictingHFinT2DM.

Author contributions RDW had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. PBR and KMV contributed significantly to the development of the cohort definitions. All authors contributed substantially to the study design, data analysis and interpretation, and the writing of the manuscript. All authors read and approved the final version.

Ethics approval The use of IBM and Optum databases was reviewed by the New England Institutional Review Board (IRB) and were determined to be exempt from IRB approval.

Consent to participate The manuscript uses secondary data. As such, no human participants were involved in the study, and informed consent for participation was not necessary at any site.

Consent for publication Not applicable.

References

1. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245–7.
2. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;19(14):40.
3. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med. 2019;2:79.
4. Kent S, Burn E, Dawoud D, Jonsson P, Ostby JT, Hughes N, et al. Common problems, common data model solutions: evidence generation for health technology assessment. Pharmacoeconomics. 2021;39(3):275–85.
5. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018;25:969–75.
6. Reps JM, Williams RD, You SC, Falconer T, Minty E, Callahan A, et al. Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. BMC Med Res Methodol. 2020;20(1):102.
7. Woelfle M, Olliaro P, Todd MH. Open science is a research accelerator. Nat Chem. 2011;3(10):745–8.
8. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health. 2018;1(39):95–112.
9. Mandl KD, Kohane IS. Federalist principles for healthcare data networks. Nat Biotechnol. 2015;33(4):360–3.
10. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. Heart. 2012;98(9):683–90.
11. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg. 2015;102(3):148–58.
12. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;16(353):i2416.
13. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691–8.
14. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.
15. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2020;14(1):49–58.
16. Helgeson C, Srikrishnan V, Keller K, Tuana N. Why simpler computer simulation models can be epistemically better for informing decisions. Philos Sci. 2021;88(2):213–33.
17. Zhang J, Wang Y, Molino P, Li L, Ebert DS. Manifold: a model-agnostic framework for interpretation and diagnosis of machine learning models. IEEE Trans Vis Comput Graph. 2019;25(1):364–73.
18. Tousoulis D, Oikonomou E, Siasos G, Stefanadis C. Diabetes mellitus and heart failure. Eur Cardiol Rev. 2014;9(1):37–42.
19. Nichols GA, Hillier TA, Erbey JR, Brown JB. Congestive heart failure in type 2 diabetes: prevalence, incidence, and risk factors. Diabetes Care. 2001;24(9):1614–9.
20. Care F. Standards of medical care in diabetes 2019. Diabetes Care. 2019;42(Suppl 1):S124–38.
21. Association AD. Updates to the standards of medical care in diabetes-2018. Diabetes Care. 2018;41(9):2045–7.
22. Marathe PH, Gao HX, Close KL. American diabetes association standards of medical care in diabetes 2017. J Diabetes. 2017;9(4):320–4.
23. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19(1):54–60.
24. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc B Metorol. 1996;58(1):267–88.
25. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68(3):279–89.