External Validation of Existing Dementia Prediction Models

Henrik John[1], Jan Kors[1], Egill Fridgeirsson[1], Jenna Reps[2], Peter Rijnbeek[1]

(1) Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands (2) Janssen Research and Development, Raritan, NJ, United States

Introduction

Dementia is an umbrella term to describe various illnesses that affect cognition and may lead to mental degradation. Early diagnosis of individuals at high risk of dementia allows for improved care and risk-factor targeted intervention. In recent years models have increasingly been developed on observational health data. These routinely collected data from administrative claims and electronic health records are considered to enhance a model's applicability at the point of care.

However, the systematic reviews of Hou et al. and Goerdten et al. conclude that although many dementia risk prediction models have been developed, only a handful of them have been externally validated [1, 2]. External validation assesses a model's reliability for clinical use in external data sources that have not been used for model development. A lack of external validation can lead to a plethora of proposed models with little evidence about which are reliable and under what circumstances.

In this study, we aim to externally validate existing dementia prediction models. To that end, we define replicability criteria, review published models, and externally validate three selected models using routinely collected health data from administrative claims and electronic health records.

Materials and Methods

The replicability criteria that a study must report are presented in the following table and were directly inferred from the prediction approach in OHDSI, where among a population at risk, we predict which patients at a defined moment in time (the index) will experience some outcome during a time-at-risk.

Category	Replicability criteria	Description
Population settings	Target population definition	Definition or description of the population for which predictions are made.
	Index date	Date at which a patient qualifies for inclusion in the target population.
	Time-at-risk	Time window in which a model's predictions are valid relative to the index date.
	Outcome definition	Definition or description of the outcome to be predicted during the time-at-risk.
Statistical analysis settings	Prediction method	Prediction methods in this study are limited to logistic regression and Cox proportional hazard for predicting a binary outcome.
	Predictor definitions	Predictor descriptions or definitions in terms of data source codes.
	Predictor time window	Time window in which the predictor is assessed.
	Model specifications	The prediction model, e.g., parameters to construct the model given a prediction method. We also distinguish here between fully and partially specified models.

Included dementia prediction studies were reviewed for these criteria to obtain the current state of reporting in the literature. Moreover, we selected three well reported models for replication and external validation in a network of observational databases, with the aim to investigate factors beyond our criteria that may impact successful external validation. These three models will in the remainder of this poster be referred to based on their first author names Walters, Mehta, and Nori, respectively [3-5].

The following databases were selected for external validations of these models as they contain an adequate number of elderly patients.

Database	Acronym	No. of patients (million)	Country	Data type
IBM MarketScan Medicare Suppl.	MDCR	10	US	Claims
Iqvia Germany Disease Analyzer	IQGER	30	DE	GP, EHR
Optum Socioeconomic Status	OPSES	85	US	Claims
Optum Electronic Health Records	OPEHR	94	US	EHR
Clinical Practice Research Datalink	CPRD	13	UK	GP
Integrated Primary Care Information	IPCI	2.5	NL	GP
Iqvia Medical Research Database	IMRD	18	UK	GP

Results and Discussion

The inclusion criteria of our literature search were met by 35 studies, which described a total of 59 prediction models. The following table summarizes the reporting of our replicability criteria in the included articles.

Category	Replicability criteria	Reported by no. of models (%)
Population settings	Target population definition	59 (100)
	Index date	23 (39)
	Time-at-risk	39 (66)
	Outcome definition	59 (100)
Statistical analysis settings	Prediction method	59 (100)
	Predictor definitions	46 (78)
	Predictor time window	21 (36)
	Model specifications: Full model	8 (14)
	Model specifications: Partial model	19 (32)

Our results showed that while reporting was complete for some criteria such as target and outcome definitions, reporting of statistical analysis criteria are mostly insufficient to fully replicate the dementia prediction models.

Moreover, our external validation of three selected models (Walters, Mehta, and Nori) showed that even if reporting was sufficient for replication, it did not guarantee that replication and external validation becomes non-trivial, because predictors had to be present, and inclusion and exclusion criteria of target and outcome had to be generalizable to other data sources. Specific problems that we encountered where the following:

- Walters: Uses a "social deprivation score", which ranges from 1 to 5 indicating social deprivation. The information in this variable has been established through a linkage, which is no longer available, or unlikely to exist in other databases across the world.
- Mehta: Does not report a time-at-risk, which was estimated to be 5 years. Also does not provide the baseline hazard so that only a risk stratification model could be replicated rather than the original Cox proportional hazard mode.
- Nori: Does not report a time-at-risk, which was estimated to be 5 years.

Performance across external data sources showed substantial differences in discrimination performance measured as the area under the receiver operating characteristic curve (AUROC) and 95% CI as presented in the following table.

Model	Internal	MDCR	IQGER	OPSES	OPEHR	CPRD	IPCI	IMRD
Walters	0.84 THIN	0.69 (0.69 – 0.69)*	0.75 (0.75 – 0.75)*	0.74 (0.74 – 0.74)*	0.73 (0.73 – 0.73)*	0.67 (0.66 – 0.67)*	0.76 (0.75 – 0.77)*	0.68 (0.68 – 0.69)*
Mehta	0.81 CPRD	0.69 (0.69 – 0.70)	0.72 (0.71 – 0.72)	0.71 (0.70 – 0.71)	0.73 (0.73 – 0.73)	0.79 (0.78 – 0.80)	0.78 (0.76 – 0.80)	0.79 (0.78 – 0.80)
Nori	0.69 Optum	0.66 (0.66 – 0.67)	0.67 (0.66 – 0.68)	0.67 (0.66 – 0.68)	0.62 (0.62 – 0.63)	0.68 (0.67 – 0.69)	0.64 (0.62 – 0.67)	0.68 (0.68 – 0.69)

We believe that the lack of external validation in dementia prediction literature can to some extend be attributed to the insufficient reporting of models. Models should be developed with external validation in mind. This could for example mean to report all aspects of the model explicitly. Such transparency is best achieved programmatically through code lists and underlying logic rather than literal descriptions, for example by providing a full description of the model (development) in code, ideally against a common data model. This approach will likely eliminate ambiguity as a source of error.

Development choices should not rely on properties unique to the development database, e.g., the Walters model contained criteria to define the target population and predictors that did not exist in the external data sources, for example the cohort entry event "one year following new registration with a THIN practice".

In general, authors should avoid uncommon predictors during model development to guarantee replicability, if the model is meant to be applied in external healthcare settings. Instead of building a single model with multiple, complex cohort entry events, it can be beneficial to build a model for each entry event, which may be easier to interpret and replicate. The Nori model suffered from this problem as it had a complex target population definition with multiple entry events. Defining the time-at-risk window is crucial to indicate in which time window a model's predictions are valid. Using the full follow-up of a population is not a valid approach, as follow-up can vary per person.

Conclusion

We reviewed 35 studies that proposed a total of 59 dementia risk models. We observed that reporting is mostly insufficient to fully replicate and externally validate published dementia prediction models, and therefore, it is uncertain how well these models would work in other clinical settings. In addition, we replicated and externally validated three existing dementia prediction models and encountered difficulties beyond our replicability criteria, such as ambiguous cohort or predictor definitions. We recommend that reporting should be more explicit and have external validation in mind if the model is meant to be applied in different settings.

References:

1. Hou XH et al. Models for predicting risk of dementia: a systematic review. Journal of Neurology, Neurosurgery and Psychiatry. 2019;90(4):373-9.

2. Goerdten J et al. Statistical methods for dementia risk prediction and recommendations for future work: A systematic review. Alzheimer's and Dementia: Translational Research and Clinical Interventions. 2019;5:563-9.

3. Walters K, et al. Predicting dementia risk in primary care: development and validation of the Dementia Risk Score using routinely collected data. BMC Med. 2016;14:6.

4. Mehta HB et al. Development and validation of the RxDx-Dementia risk index to predict dementia in patients with type 2 diabetes and hypertension. Journal of Alzheimer's Disease. 2016;49(2):423-32.

5. Nori VS et al. Identifying incident dementia by applying machine learning to a very large administrative claims dataset. PLoS ONE. 2019;14(7).

Contact: l.john@erasmusmc.nl

This project has 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 EFPIA.