PROTOCOL: External validation of existing dementia prediction models on observational health data

Luis H. John, Jan A. Kors, Jenna M. Reps, Peter R. Rijnbeek

03/03/2021

Version 2

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.

Version: 2

Authors:

Luis H. John[1], Jan A. Kors[1], Jenna M. Reps[2], Peter R. Rijnbeek[1]

[1] Erasmus University Medical Center, [2] Janssen Research and Development

Date: 3 March 2021

Acknowledgment: The analysis is based in part on work from the Observational Health Sciences and Informatics (OHDSI) collaborative. OHDSI (https://ohdsi.org/) is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics.

Disclosures: JMR is an employee of Janssen Research and Development.

Funding statement: 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.

Table of Contents

1 LIST OF ABBREVIATIONS

ADRD	Alzheimer's disease and related dementias
CDM	Common Data Model
CI	Confidence Interval
CPM	Clinical Prediction Model
CYCLOPS	Cyclic coordinate descent for logistic, Poisson and survival analysis
LASSO	Least absolute shrinkage and selection operator
MEHTA	Mehta's RxDx-Dementia Risk Index[1]
NORIS	Nori's dementia prediction model[2]
O	Outcome cohort
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PLP	Patient-level prediction R-package framework
SNOMED	Systematized Nomenclature of Medicine
T	Target cohort
WALTERS	Walters' Dementia Risk Score[3]

2 EXECUTIVE SUMMARY

This study will externally validate three existing dementia prediction models that are proposed in clinical prediction model literature. We will assess the models' performance in real-world settings and discuss the quality of model reporting.

The three clinical prediction models (CPMs) that will be examined are Walters' Dementia Risk Score[3] (WALTERS), Mehta's RxDx-Dementia Risk Index[1] (MEHTA) and Nori's ADRD dementia prediction model[2] (NORI). WALTERS predicts 5-year risk of first recorded dementia diagnosis among patients aged 60–79 using a cox proportional hazard model. MEHTA predicts risk of incident dementia among patients diagnosed with type 2 diabetes mellitus and hypertension using a cox proportional hazard model. NORI predicts Alzheimer's disease and related dementias (ADRD) among patients aged 45 and older using a logistic regression model.

We will conduct the study using data from observational databases across the European Health Data & Evidence Network (EHDEN) and the Observational Health Data Sciences and Informatics (OHDSI) network and perform analyses using the OHDSI Patient Level Prediction (PLP) package framework.

3 AMENDMENTS AND UPDATES

Version	Date	Author	Description
1	1 March 2021	L.H. John	First draft
2	3 March 2021	L.H. John	Release version

4 MILESTONES

Milestone	Planned / Estimated Date
Start of analysis	1 October 2020
End of initial analysis	31 March 2021
End of final analysis	
Submission of manuscript	

5 RATIONAL & BACKGROUND

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.[4] Various prediction models for identifying individuals who are at high risk of dementia have been developed. For example, the cardiovascular Risk Factors, Aging and Dementia (CAIDE) risk score predicts late-life dementia risk based on midlife vascular risk factors.[5]

In recent years dementia prediction models have been increasingly developed on observational health data. Because this kind of data is routinely collected from administrative claims and electronic health records, it is considered to enhance model's applicability at the healthcare system level as it reflects the data normally available to a clinician in practice.[3]

While observational health data generally does not include dedicated dementia tests, genetic or imaging data, and commonly used variables such as education, which have previously been shown to be predictive, various studies have shown good internal validation performance when developing models. Notable examples are Walters et al. who developed dementia prediction models using electronic health record data from the THIN database and Albrecht et al. who developed mid-life and late-life predictive models for earlier diagnosis of Alzheimer's Disease and Related Dementias (ADRD) using administrative claims data from the OptumLabs Data Warehouse.[3,6] Also Nori et al. uses very large claims datasets to develop dementia prediction models.[2]

However, 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.[7,8] External validation assesses a model's reliability for clinical use in external data sources that have not been used for model development. Riley et al. concluded that most prediction model research focuses on model development rather than external validation leading to a plethora of proposed model with little evidence about which are reliable and under what circumstances.[9] Whether a model can be externally validated also demonstrates whether it could be applied in clinical practice in the first place. Moreover, assessing a replicated model's discrimination and calibration performance can help inform of its clinical usefulness.

In this study we aim to externally validate three clinical prediction models: (1) Walters' Dementia Risk Score[3] (WALTERS), (2) Mehta's RxDx-Dementia Risk Index[1] (MEHTA) and (3) Nori's ADRD dementia prediction model[2] (NORI). WALTERS predicts 5-year risk of first recorded dementia diagnosis among patients aged 60–79 using a cox proportional hazard model. MEHTA predicts risk of incident dementia among patients diagnosed with type 2 diabetes mellitus and hypertension using a cox proportional hazard model. NORI predicts Alzheimer's disease and related dementias (ADRD) among patients aged 45 and older using a logistic regression model.

We will also investigate how well these models are reported to allow replication and whether the chosen predictors are commonly found in observational health databases.

6 OBJECTIVES

Aim: Externally validate existing prognostic models that predict late-life dementia in the general population using routinely collected health data from administrative claims and electronic health records.

Approach: Evaluate the performance of the Walters' Dementia Risk Score[3], Mehta's RxDx-Dementia Risk Index[1] and Nori's ADRD dementia prediction model[2] across a diverse range of observational data from administrative claims, electronic health records, and general practitioners. These evaluations will assess model discrimination (c-statistic or AUROC) and calibration (Harrell's E) before and after

simple (intercept and slope) recalibration. Moreover, we will discuss the quality of reporting in the research articles about the three models given the external validation results.

7 METHODS

7.1 Model selection

Erasmus University Medical Center conducted a systematic review of scientific articles between 2009 and 2020 that propose clinical prediction models for dementia and its various subtypes.

Articles were included if they met the following criteria: (1) the sample was population-based; (2) the article provided a risk model to predict dementia in non-demented individuals, prospectively; (3) the article included measurements of discrimination (AUROC or c-statistic). A total of 500 articles were found, and 65 were selected based on our criteria. We learned that the most popular prediction algorithms are logistic regression and cox proportional hazard models, which made up 60% of the studies reviewed.

In addition, a list of required information was compiled that a scientific article needs to report in order to reliably replicate a model for external validation (Table below).

Required model information	Description
Development database	The data source information gives insight into the origin of the data (administrative claims, general practitioner, hospital, etc.) as well as the source data model (ICD9, ICD10, Read, etc.).
Algorithm	The prediction algorithm used. We are limiting ourselves to the most popular linear algorithms: logistic regression and cox proportional hazard for a binary outcome
Coefficients	A numeric value associating the predictor to the outcome risk.
Predictor definitions	A description of the predictor or, better, a list of data source codes.
Predictor time window	The time window in which the predictor was assessed.
Baseline hazard/Intercept	The baseline hazard estimate in a cox proportional hazard model or the intercept term in a logistic regression model. Importantly, it is possible to assess a model's discrimination performance (C-statistic/AUROC) without the baseline hazard/intercept.
Index date	The date at which a patient qualifies for the target cohort
Target population definition	A definition of the population for which predictions are made
Outcome definition	A definition of the outcome to be predicted.
Time-at-risk	A time window in which a model's predictions are valid.

Scientific articles that could provide a large number of items in this table (about 80%) were pre-selected. Final models were selected on the basis of their feasibility (projected presence of predictors and sufficiently large populations for target and outcome) of external validation in observational health databases mapped to the OMOP CDM.

7.2 Study design

This study will follow a retrospective, observational, patient-level prediction design. We define 'retrospective' to mean the study will be conducted using data already collected prior to the start of the study. We define 'observational' to mean there is no intervention or treatment assignment imposed by the study. We define 'patient-level prediction' as a modelling process wherein an outcome is predicted within a time at risk relative to the target cohort start and/or end date.

Prediction is performed using a set of covariates derived using data prior to the start of the target cohort.

Figure 1 illustrates the prediction problem we will address. Among a population at risk, we aim to predict which patients at a defined moment in time (t = 0) will experience some outcome during a time-at-risk. Prediction is done using only information about the patients in an observation window prior to that moment in time.

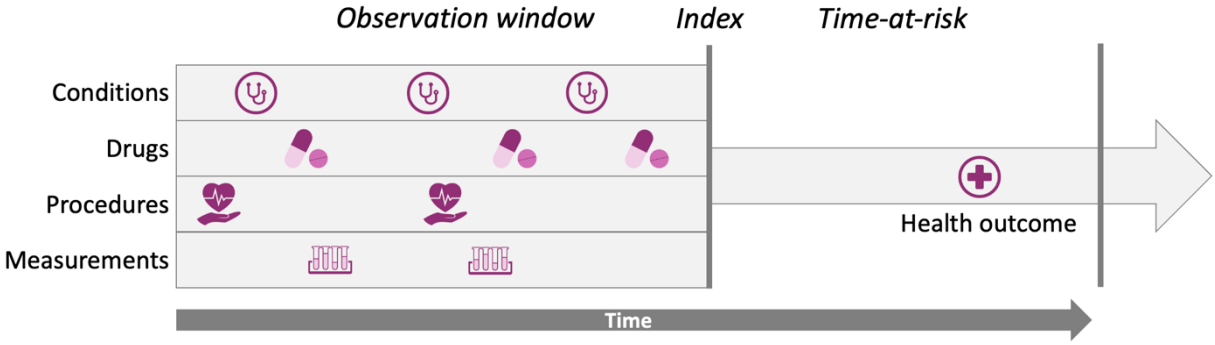

Figure 1. Patient-level prediction time windows and index date.

We follow the TRIPOD guidance for transparent reporting of the model results.[10]

7.3 Data sources

The analyses will be performed across the EHDEN and OHDSI network in multiple observational databases. These databases have been converted to the OMOP Common Data Model (CDM), version 5.0 or higher. The complete specification for the OMOP CDM is available at https://github.com/OHDSI/CommonDataModel.

Data Source Name	Contributor	Description
Optum De-Identified Clinformatics Data Mart Database (OptumInsight, Eden Prairie, MN)	Janssen	Adjudicated administrative health claims database for members with private health insurance, who are fully insured in commercial plans or in administrative services only (ASOs), Legacy Medicare Choice Lives (prior to January 2006), and Medicare Advantage (Medicare Advantage Prescription Drug coverage starting January 2006).
Optum De-identified Electronic Health Record Dataset	Janssen	A medical records database representing Humedica's Electronic Health Record data.
IBM MarketScan® Medicare Supplemental Database (MDCR)	Janssen	Represents health services of retirees (aged 65 or older) in the United States with primary or Medicare supplemental coverage through privately insured fee-for-service, point-of-service, or capitated health plans. These data include adjudicated health insurance claims (e.g., inpatient, outpatient, and outpatient pharmacy). Additionally, it captures laboratory tests for a subset of the covered lives.
Iqvia Disease Analyzer (DA) Germany	Janssen	Data collected from physician practices and medical centres for all ages. Mostly primary care physician data, however, some data from specialty practices (where practices are electronically

		connected to each other) and some lab data is included.
Integrated Primary Care Information (IPCI) database	Erasmus MC	Dutch database containing the complete medical record of more than 3 million patients provided by more than 450 GPs geographically spread over the Netherlands.
Clinical Practice Research Datalink (CPRD)	Janssen	Governmental, not-for-profit research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA), a part of the Department of Health, United Kingdom (UK). CPRD consists of data collected from UK primary care for all ages.
Iqvia Medical Research Data UK (IMRD)	Iqvia	Real world data, incorporating data from THIN, a Cegedim Database, which contains longitudinal non-identified primary care electronic medical records (EMR) from more than 18 million patients in the UK.

7.4 Study populations

Populations are defined as computable target and outcome cohorts as previously described[11,12].

7.4.1 Target cohorts [T]

Target component	WALTERS	MEHTA	NORI
Cohort name	Walters Target 0 Prior 0 Post	Mehta Target 0 Prior 0 Post [V2]	Nori Target 0 Prior 0 Post [v1]
Description	Patients with an outpatient visit occurrences, who are between 60 and 79 years old	Patients with a condition occurrence of T2DM, who are >= 60 years old	Patients with an outpatient visit occurrence, who are >= 45 years old
Inclusion criteria	• No prior dementia, memory loss, confusion, cognitive decline • No prior drug-induced (and alcoholinduced) dementia • No prior dementia in other conditions NOS • No HIV positive • No huntington's • No lewy body disease • No parkinson's	• Hypertension • No prior dementia	• No prior alzheimer's disease • No prior alzheimer's medication

	• No pick's • No cruetzfeldt-jacob disease		

7.4.2 Outcome cohorts [O]

Outcome component	WALTERS	MEHTA	NORI
Cohort name	Persons with dementia	Mehta Outcome	Nori Outcome
Description	Patients with a condition occurrence of dementia	Patiets with a condition occurrence of dementia; or a drug exposure of a dementia prescription medication	Patients with a condition occurrence of ADRD; or a drug exposure for ADRD; or a drug exposure of memantine with a condition ocurrence of ADRD; or an inpatient visist occurrence with a condition occurrence of ADRD
Inclusion criteria	-	-	-

7.4.3 Population settings

Population settings	WALTERS	MEHTA	NORI
binary	TRUE	TRUE	TRUE
includeAllOutcomes	TRUE	TRUE	TRUE
firstExposureOnly	FALSE	TRUE	FALSE
washoutPeriod	365	365	1825
removeSubjectsWithPriorOutcome	FALSE	FALSE	FALSE
priorOutcomeLookback	9999	9999	9999
requireTimeAtRisk	TRUE	TRUE	TRUE
minTimeAtRisk	365	365	1094
riskWindowStart	1	1	1
startAnchor	'cohort start'	'cohort start'	'cohort start'
endAnchor	'cohort start'	'cohort start'	'cohort start'
riskWindowEnd	1825	1825	1095
verbosity	'INFO'	'INFO'	'INFO'

7.4.4 Covariate populations

Custom covariates that do not use the standard covariate settings of the FeatureExtraction R-package are defined as executable ATLAS cohorts. A complete list of these custom covariates can be found in 7.4.5 under cohort type "Covariate".

7.4.5 Executable cohort definitions

Cohorts are defined in ATLAS. These include definitions of the target cohort, outcome cohort as well as definitions of covariates. The following table is a complete list of ATLAS cohorts that this study utilizes.

Cohort ID	Model	Cohort type	Cohort name	URL
1	WALTERS	Target	Walters Target 0 Prior 0 Post	
2	WALTERS	Outcome	Persons with dementia	
3	WALTERS	Covariate	BMI	
4	WALTERS	Covariate	BMI squared	
5	WALTERS	Covariate	Smoking status never	
6	WALTERS	Covariate	Smoking status past	
7	WALTERS	Covariate	Smoking status current	
8	WALTERS	Covariate	History of alcohol problems	
9	WALTERS	Covariate	History of diabetes	
10	WALTERS	Covariate	Depression	
11	WALTERS	Covariate	Stroke	
12	WALTERS	Covariate	Atrial fibrillation	
13	WALTERS	Covariate	Current aspirin use	
14	WALTERS	Covariate	Social deprivation	
15	MEHTA	Target	Mehta Target 0 Prior 0 Post [V2]	
16	MEHTA	Outcome	Mehta Outcome	
17	MEHTA	Covariate	Myocardial infarction	
18	MEHTA	Covariate	Congestive heart failure	
19	MEHTA	Covariate	Coronary and peripheral vascular disease	
20	MEHTA	Covariate	Cerebrovascular disease	
21	MEHTA	Covariate	Chronic pulmonary disease	
22	MEHTA	Covariate	Rheumatologic disease	
23	MEHTA	Covariate	Peptic ulcer disease	
24	MEHTA	Covariate	Renal disease and end stage renal disease	
25	MEHTA	Covariate	Mild liver disease and moderate or severe liver disease	
26	MEHTA	Covariate	Any malignancy	
27	MEHTA	Covariate	Epilepsy	
28	MEHTA	Covariate	Hyperlipidemia	
29	MEHTA	Covariate	Parkinsons disease	

30	MEHTA	Covariate	Cardiac disease ASCVD	
31	MEHTA	Covariate	Glaucoma	
32	MEHTA	Covariate	Transplantation	
33	MEHTA	Covariate	Thyroid disorder	
34	MEHTA	Covariate	Gout	
35	MEHTA	Covariate	Crohns and ulcerative disease	
36	MEHTA	Covariate	Pain and inflammation and pain	
37	MEHTA	Covariate	Depression	
38	MEHTA	Covariate	Psychotic illness	
39	MEHTA	Covariate	Bipolar disorders	
40	MEHTA	Covariate	Anxiety and tension	
41	NORI	Target	Nori Target 0 Prior 0 Post [v1]	
42	NORI	Outcome	Nori Outcome	
43	NORI	Covariate	MEMORY LOSS	
44	NORI	Covariate	PARALYSIS AGITANS	
45	NORI	Covariate	MILD COGNITIVE IMPAIRMENT SO STATED	
46	NORI	Covariate	BIPOLAR DISORDER UNSPECIFIED	
47	NORI	Covariate	UNSPECIFIED PSYCHOSIS	
48	NORI	Covariate	LOSS OF WEIGHT	
49	NORI	Covariate	DEPRESSIVE DISORDER NOT ELSEWHERE CLASSIFIED	
50	NORI	Covariate	ALTERED MENTAL STATUS	
51	NORI	Covariate	PERSONAL HISTORY OF FALL	
52	NORI	Covariate	OTHER CONVULSIONS	
53	NORI	Covariate	UNSPECIFIED FALL	
54	NORI	Covariate	OTHER CHRONIC PAIN	
55	NORI	Covariate	ACUTE BUT ILLDEFINED CEREBROVASCULAR DISEASE	
56	NORI	Covariate	URGE INCONTINENCE	
57	NORI	Covariate	OTHER ALTERATION OF CONSCIOUSNESS	

58	NORI	Covariate	UNSPECIFIED CONSTIPATION	
59	NORI	Covariate	UNSPECIFIED URINARY INCONTINENCE	
60	NORI	Covariate	ENCOUNTER FOR LONGTERM CURRENT USE OF OTHER MEDICATIONS	
61	NORI	Covariate	LACK OF COORDINATION	
62	NORI	Covariate	OTHER MALAISE AND FATIGUE	
63	NORI	Covariate	DIABETES MELLITUS	
64	NORI	Covariate	ABNORMALITY OF GAIT	
65	NORI	Covariate	DIZZINESS AND GIDDINESS	
66	NORI	Covariate	UNSPECIFIED CEREBRAL ARTERY OCCLUSION WITH CEREBRAL INFARCTION	
67	NORI	Covariate	DIABETES MELLITUS 2	
68	NORI	Covariate	EDEMA	
69	NORI	Covariate	MUSCLE WEAKNESS GENERALIZED	
70	NORI	Covariate	URINARY TRACT INFECTION SITE NOT SPECIFIED	
71	NORI	Covariate	SCREENING MAMMOGRAPHY COMPUTERAIDED DETECTION	
72	NORI	Covariate	COMPUTED TOMOGRAPHY HEAD OR BRAIN	
73	NORI	Covariate	RADIOLOGIC EXAMINATION CHEST SINGLE VIEW FRONTAL	
74	NORI	Covariate	VENLAFAXINE HCL	
75	NORI	Covariate	DULOXETINE HCL	
76	NORI	Covariate	TOLTERODINE TARTRATE	
77	NORI	Covariate	SERTRALINE HCL	
78	NORI	Covariate	CITALOPRAM HYDROBROMIDE	

79	NORI	Covariate	POTASSIUM CHLORIDE	
80	NORI	Covariate	OXYBUTYNIN CHLORIDE	
81	NORI	Covariate	HYDROCODONE BIT ACETAMINOPHEN	
82	NORI	Covariate	PROPOXYPHENE ACETAMINOPHEN	
83	NORI	Covariate	SULFAMETHOXAZOLE TRIMETHOPRIM	
84	NORI	Covariate	METFORMIN HCL	
85	NORI	Covariate	BLOOD SUGAR DIAGNOSTIC	
86	NORI	Covariate	LISINOPRIL	
87	NORI	Covariate	CEPHALEXIN MONOHYDRATE	
88	NORI	Covariate	SIMVASTATIN	
89	NORI	Covariate	CLOPIDOGREL BISULFATE	
90	NORI	Covariate	TRAMADOL HCL	
91	NORI	Covariate	GABAPENTIN	
92	NORI	Covariate	FUROSEMIDE	

8 DATA ANALYSIS PLAN

8.1 Algorithm settings

Algorithm setting	WALTERS	MEHTA	NORI
Prediction type	survival	survival	binary
Seed	NULL	NULL	NULL
Variance	0.01	0.01	0.01
Mapping	$1 - 0.9969^{e^X}$	X (no baseline hazard provided)	$\frac{1}{1 + e^{1.96 - X}}$

8.2 Standard covariate settings

The following standard covariate settings were used for the three models. Note that a majority of covariates used in the three models are not standard covariates, but rather covariates defined through executable cohort definitions in ATLAS. These have already been listed in 7.4.5.

Standard covariates	WALTERS	MEHTA	NORI
VisitCountMediumTerm	FALSE	FALSE	FALSE
ObservationShortTerm	FALSE	FALSE	FALSE
shortTermStartDays	-30	-30	-30
MeasurementRangeGroupShortTerm	FALSE	FALSE	FALSE
ConditionOccurrenceLongTerm	FALSE	FALSE	FALSE
DrugEraStartLongTerm	FALSE	FALSE	FALSE
VisitCountShortTerm	FALSE	FALSE	FALSE

Chads2Vasc	FALSE	FALSE	FALSE
ConditionGroupEraStartLongTerm	FALSE	FALSE	FALSE
ConditionEraShortTerm	FALSE	FALSE	FALSE
Dcsi	FALSE	FALSE	FALSE
DrugGroupEraLongTerm	FALSE	FALSE	FALSE
DrugGroupEraShortTerm	FALSE	FALSE	FALSE
ConditionEraStartLongTerm	FALSE	FALSE	FALSE
temporal	FALSE	FALSE	FALSE
DemographicsIndexMonth	FALSE	FALSE	FALSE
ConditionOccurrencePrimaryInpatientLongTerm	FALSE	FALSE	FALSE
ConditionEraAnyTimePrior	FALSE	FALSE	FALSE
addDescendantsToInclude	FALSE	FALSE	FALSE
ConditionGroupEraStartMediumTerm	FALSE	FALSE	FALSE
ProcedureOccurrenceLongTerm	FALSE	FALSE	FALSE
DrugExposureLongTerm	FALSE	FALSE	FALSE
DrugEraStartShortTerm	FALSE	FALSE	FALSE
DistinctIngredientCountMediumTerm	FALSE	FALSE	FALSE
DistinctMeasurementCountShortTerm	FALSE	FALSE	FALSE
MeasurementRangeGroupLongTerm	FALSE	FALSE	FALSE
ConditionGroupEraOverlapping	FALSE	FALSE	FALSE
MeasurementRangeGroupMediumTerm	FALSE	FALSE	FALSE
DrugGroupEraStartMediumTerm	FALSE	FALSE	FALSE
MeasurementAnyTimePrior	FALSE	FALSE	FALSE
MeasurementMediumTerm	FALSE	FALSE	FALSE
includedCovariateIds			
ConditionOccurrenceAnyTimePrior	FALSE	FALSE	FALSE
DistinctConditionCountLongTerm	FALSE	FALSE	FALSE
MeasurementValueLongTerm	FALSE	FALSE	FALSE
DrugEraShortTerm	FALSE	FALSE	FALSE
DrugGroupEraAnyTimePrior	FALSE	FALSE	FALSE
DrugEraOverlapping	FALSE	FALSE	FALSE
ConditionOccurrencePrimaryInpatientAnyTimePrior	FALSE	FALSE	FALSE
ConditionEraMediumTerm	FALSE	FALSE	FALSE
ConditionEraOverlapping	FALSE	FALSE	FALSE
ConditionEraStartShortTerm	FALSE	FALSE	FALSE
ObservationAnyTimePrior	FALSE	FALSE	FALSE
VisitConceptCountShortTerm	FALSE	FALSE	FALSE
DemographicsEthnicity	FALSE	FALSE	FALSE
DistinctIngredientCountLongTerm	FALSE	FALSE	FALSE
ConditionOccurrencePrimaryInpatientShortTerm	FALSE	FALSE	FALSE
DemographicsAgeGroup	FALSE	FALSE	FALSE
DistinctProcedureCountShortTerm	FALSE	FALSE	FALSE
DistinctObservationCountMediumTerm	FALSE	FALSE	FALSE
includedCovariateConceptIds	FALSE	FALSE	FALSE
DrugGroupEraStartShortTerm	FALSE	FALSE	FALSE
addDescendantsToExclude	FALSE	FALSE	FALSE
DrugEraLongTerm	FALSE	FALSE	FALSE
DistinctConditionCountShortTerm	FALSE	FALSE	FALSE
ConditionGroupEraShortTerm	FALSE	FALSE	FALSE

ConditionEraStartMediumTerm	FALSE	FALSE	FALSE
VisitCountLongTerm	FALSE	FALSE	FALSE
DemographicsRace	FALSE	FALSE	FALSE
ProcedureOccurrenceAnyTimePrior	FALSE	FALSE	FALSE
DistinctObservationCountLongTerm	FALSE	FALSE	FALSE
ProcedureOccurrenceMediumTerm	FALSE	FALSE	FALSE
CharlsonIndex	FALSE	FALSE	FALSE
DemographicsPriorObservationTime	FALSE	FALSE	FALSE
MeasurementShortTerm	FALSE	FALSE	FALSE
DistinctProcedureCountMediumTerm	FALSE	FALSE	FALSE
ConditionEraLongTerm	FALSE	FALSE	FALSE
DrugGroupEraStartLongTerm	FALSE	FALSE	FALSE
DemographicsGender	TRUE	TRUE	FALSE
DeviceExposureAnyTimePrior	FALSE	FALSE	FALSE
ObservationLongTerm	FALSE	FALSE	FALSE
DemographicsIndexYearMonth	FALSE	FALSE	FALSE
ConditionOccurrenceMediumTerm	FALSE	FALSE	FALSE
longTermStartDays	-365	-365	-365
DemographicsAge	TRUE	TRUE	FALSE
DrugGroupEraOverlapping	FALSE	FALSE	FALSE
DistinctMeasurementCountLongTerm	FALSE	FALSE	FALSE
MeasurementRangeGroupAnyTimePrior	FALSE	FALSE	FALSE
DistinctConditionCountMediumTerm	FALSE	FALSE	FALSE
DrugGroupEraMediumTerm	FALSE	FALSE	FALSE
ProcedureOccurrenceShortTerm	FALSE	FALSE	FALSE
ObservationMediumTerm	FALSE	FALSE	FALSE
ConditionGroupEraAnyTimePrior	FALSE	FALSE	FALSE
Chads2	FALSE	FALSE	FALSE
DrugExposureAnyTimePrior	FALSE	FALSE	FALSE
DeviceExposureLongTerm	FALSE	FALSE	FALSE
DemographicsTimeInCohort	FALSE	FALSE	FALSE
DistinctMeasurementCountMediumTerm	FALSE	FALSE	FALSE
MeasurementValueShortTerm	FALSE	FALSE	FALSE
DeviceExposureMediumTerm	FALSE	FALSE	FALSE
ConditionGroupEraStartShortTerm	FALSE	FALSE	FALSE
ConditionOccurrencePrimaryInpatientMediumTerm	FALSE	FALSE	FALSE
MeasurementLongTerm	FALSE	FALSE	FALSE
DemographicsIndexYear	FALSE	FALSE	FALSE
MeasurementValueMediumTerm	FALSE	FALSE	FALSE
DrugEraStartMediumTerm	FALSE	FALSE	FALSE
MeasurementValueAnyTimePrior	FALSE	FALSE	FALSE
DistinctObservationCountShortTerm	FALSE	FALSE	FALSE
DrugEraMediumTerm	FALSE	FALSE	FALSE
ConditionGroupEraLongTerm	FALSE	FALSE	FALSE
DrugExposureShortTerm	FALSE	FALSE	FALSE
DistinctIngredientCountShortTerm	FALSE	FALSE	FALSE
DeviceExposureShortTerm	FALSE	FALSE	FALSE
mediumTermStartDays	-180	-180	-180
DemographicsPostObservationTime	FALSE	FALSE	FALSE

VisitConceptCountLongTerm	FALSE	FALSE	FALSE
VisitConceptCountMediumTerm	FALSE	FALSE	FALSE
excludedCovariateConceptIds			
ConditionGroupEraMediumTerm	FALSE	FALSE	FALSE
DrugExposureMediumTerm	FALSE	FALSE	FALSE
DistinctProcedureCountLongTerm	FALSE	FALSE	FALSE
DrugEraAnyTimePrior	FALSE	FALSE	FALSE
endDays	0	0	0
ConditionOccurrenceShortTerm	FALSE	FALSE	FALSE

8.3 Model recalibration

The WALTERS and the MEHTA models will be recalibrated during analysis. Therefore, each study is effectively executed two times, once without recalibration and once with recalibration. All population and model settings remain constant.

8.4 Model evaluation

Predictions will be made for individual patients meeting the inclusion criteria for the target populations for each CPM. For the WALTERS and NORI models, two sets of evaluation metrics will be ascertained: 1) for the original model; 2) for the model with a recalibrated intercept and slope.

8.4.1 Discrimination

We will use the area under the receiver operating characteristic curve (AUROC) to evaluate the discriminative performance of the logistic regression model and the concordance statistic (c-statistic) to evaluate the discrimination performance of the cox proportional hazard models.

Across study sites we will describe the distribution (i.e., median and inter-quartile range) of calculated AUROC/c-statistic for each model to evaluate the discriminative ability of the prediction model in an external database. The chosen recalibration techniques (slope and intercept) do not alter the discrimination of the original models.

8.4.2 Calibration

Calibration-in-the-large, a measure of global fit, as well as calibration slope and Harrell's E (E_{AVG}) standardized to the outcome rate, will be assessed. E_{AVG} computes the average absolute calibration error (difference between the observed outcome rate and the estimated probabilities, where the observed rate is estimated using a non-parametric locally weighted scatterplot smoothing). E_{90}, representing the 90th percentile of the absolute calibration error, will also be calculated. For each model evaluation we will describe the distribution (i.e., median and inter-quartile range) of calibration measures across the study sites. A plot of the predicted risk against the observed risk is used to visualize the calibration. Lastly, we compute Harrell's E_{MAX} index, which is the maximum absolute difference between a smooth calibration curve and the diagonal line of perfect calibration.

We will describe the impact of recalibration techniques (slope and intercept) on these calibration metrics.

8.5 Analysis execution settings

For the WALTERS prediction model there is 1 target cohort evaluated for 1 outcome over 1 model over 1 covariate setting and over 1 population setting. Additionally, 12 custom covariates are created which are based on 12 executable cohort definitions. The WALTERS model is recalibrated, for which reason the analysis is executed twice, once without recalibration and once with recalibration.

For the MEHTA prediction model there is 1 target cohort evaluated for 1 outcome over 1 model over 1 covariate settings and over 1 population setting. Additionally, 24 custom covariates are created which are based on 24 executable cohort definitions.

For the NORI prediction model there is 1 target cohort evaluated for 1 outcome over 1 model over 1 covariate setting and over 1 population setting. Additionally, 50 custom covariates are created which are based on 50 executable cohort definitions. The Nori model is recalibrated for which reason the analysis is executed twice, once without recalibration and once with recalibration.

9 STRENGTHS AND LIMITATIONS

9.1 Strengths

- Observational health data that are routinely collected from administrative claims, electronic health records, etc., is considered to enhance a model's applicability at the healthcare system level as it reflects the data normally available to a clinician in practice.[3] If dementia models transport well, they may be interesting candidates for future research into clinical usefulness.

9.2 Limitations

- The external validation datasets may not have a sufficient number of (some of) the different outcomes to be used in the analysis.
- Although the OMOP CDM standardises the vocabularies of the datasets, the concept recording distributions are likely to differ between databases and it is unknown how much this will limit model transportability.

10 PROTECTION OF HUMAN SUBJECTS

For this study, personal data which are collected in national/regional electronic health record databases from various countries, will be processed. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

In agreement with these regulations, rather than combining person level data and performing a central analysis, only local analyses will be run, which generate nonidentifiable aggregate summary results. All the databases used in this study have a well-developed mechanism to ensure that regulations dealing with ethical use of the data and adequate privacy control are adhered to. If required, the protocol has been reviewed by the Institutional Review Boards of the respective databases.

10.1 Management and reporting of adverse events and adverse reactions

This study uses coded data that already exist in an electronic database. In this type of database, it is not possible to link (i.e., identify a potential causal association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (i.e., identifiable patient, identifiable reporter, a suspect product, and event) are not available and adverse events are not reportable as individual adverse events reports. The study results will be assessed for medically important results.

11 PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

Dissemination activities to be undertaken will have mainly, although not exclusively, a scientific nature (articles, presentations at conferences, etc.). The study results will be posted on the OHDSI

website and EHDEN website after completion of the study. One paper describing the study and its results will be written and submitted for publication to a peer-reviewed scientific journal.

12 Study Repository

The latest executable files of this study can be found on the Medical Informatics, Erasmus MC, GitHub repository.

- Study files: https://github.com/mi-erasmusmc/EmcDementiaModelValidation

To execute the study, please follow the instructions in the readme.md file of above mentioned repository.

13 References

1. Mehta HB, Mehta V, Tsai C-L, Chen H, Aparasu RR, Johnson ML. 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):423-432.
2. Nori VS, Hane CA, Martin DC, Kravetz AD, Sanghavi DM. Identifying incident dementia by applying machine learning to a very large administrative claims dataset. *PLoS ONE.* 2019;14(7).
3. Walters K, Hardoon S, Petersen I, 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. Stephan BC, Kurth T, Matthews FE, Brayne C, Dufouil C. Dementia risk prediction in the population: are screening models accurate? *Nat Rev Neurol.* 2010;6(6):318-326.
5. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *The Lancet Neurology.* 2006;5(9):735-741.
6. Albrecht JS, Hanna M, Kim D, Perfetto EM. Predicting diagnosis of Alzheimer's Disease and related dementias using administrative claims. *Journal of Managed Care and Specialty Pharmacy.* 2018;24(11):1138-1145.
7. Hou XH, Feng L, Zhang C, Cao XP, Tan L, Yu JT. Models for predicting risk of dementia: a systematic review. *Journal of Neurology, Neurosurgery and Psychiatry.* 2019;90(4):373-379.
8. Goerdten J, Čukić I, Danso SO, Carrière I, Muniz-Terrera G. Statistical methods for dementia risk prediction and recommendations for future work: A systematic review. *Alzheimer's and Dementia: Translational Research and Clinical Interventions.* 2019;5:563-569.
9. Riley RD, Ensor J, Snell KI, 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.
10. 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):148-158.
11. 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 *Journal of the American Medical Informatics Association.* 2018;25(8):969-975.

12. Observational Health Data Sciences and Informatics. *The Book of OHDSI.* Observational Health Data Sciences and Informatics,; 2019.