Study Protocol

P4-C1-021

DARWIN EU® - Alzheimer's Disease: Incidence, Prevalence, and Individual's Characteristics

08/12/2025

Version 3.0

Authors: Marzyeh Amini, Rana Jajou, Melissa Leung, Talita Duarte-Salles, Adam Black, Ger Inberg, Natasha Yefimenko

Public

CONTENTS

Study title	DARWIN EU® - Alzheimer's Disease: Incidence, Prevalence, and Individual's Characteristics
Protocol version	V3.0
Date	08/12/2025
EUPAS number	EUPAS1000000826
Active Substance	n/a
Medicinal Product	n/a
Research question and objectives	• What are the overall incidence rates and prevalence of Alzheimer's disease in the general adult population, stratified by calendar year, sex, and age categories? • What are the demographic characteristics, diagnostic procedures, and clinical profile of individuals with newly diagnosed Alzheimer's disease? The specific objectives are: 1. To estimate overall incidence and prevalence of Alzheimer's disease in the general adult population, stratified by calendar year, sex, and age. 2. To describe the demographic characteristics, diagnostic procedures, and clinical profile of individuals with newly diagnosed Alzheimer's disease.
Countries of study	Croatia, Denmark, Germany, The Netherlands, United Kingdom
Authors	Marzyeh Amini, m.amini@darwin-eu.org Rana Jajou, r.jajou@darwin-eu.org Melissa Leung, m.leung@darwin-eu.org Talita Duarte Salles, t.duarte@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
Aβ	Amyloid beta
ACE	Angiotensin-converting enzyme
AD	Alzheimer's disease
ADL	Activity of daily living
ApoE ε4	Apolipoprotein E epsilon 4
ARB	Angiotensin II receptor blockers
ATC	Anatomical Therapeutic Chemical
CC	Coordinating centre
CDM	Common Data Model
CHMP	Committee for Medicinal Products for Human Use (EMA)
CI	Confidence Interval
CIPH	Croatian Institute of Public Health
CPRD	Clinical Practice Research Datalink
DAC	Data Analytics Centre
DAR	Cause of Death Registry
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DKMA	Danish Medicines Agency
DOAC	Direct oral anticoagulants
DOI	Declaration of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
EBM	Einheitlicher Bewertungsmaßstab
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GmBh	Gesellschaft mit beschränkter Haftung
GDPR	General Data Protection Regulation
GLP-1	Glucagon-like peptide-1
GP	General Practitioner
HMA	Heads of Medicines Agencies
HRI	Health Risk Institute
ICD-10	International Classification of Diseases, Tenth Revision
ICPC-1	International Classification of Primary Care, first version

Acronyms/terms	Description
ICU	Intensive Care Unit
IgG1	Immunoglobulin G1
InGef RDB	InGef Research Database (Germany)
IP	Inpatient
IPCI	Integrated Primary Care Information (Netherlands)
IQR	Interquartile Range
IQVIA DA	IQVIA Disease Analyzer
IRB	Institutional Review Board
LOINC	Logical Observation, Identifiers, Names and Codes
MCI	Mild Cognitive Impairment
MRI	Magnetic Resonance Imaging
NAJS	National Public Health Information System (Croatia)
NDORMS	Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
NFC	Near Field Communication
NHG	Nederland Huisartsen Genootschap
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
ONS	Office for National Statistics
OP	Outpatient
OPS	Operationen- und Prozedurenschlüssel
PCSK9	Proprotein convertase subtilisin/kexin type 9
PET-F18	Positron Emission Tomography with Fluorine-18
PY	Person-years
PZN	Pharmazentralzummer -pharmaceutical reference number
RRE	Remote Research Environment
RWD	Real-World Data
RWE	Real-World Evidence
RxNorm	Medical prescription normalized
SD	Standard Deviation
SGLT2	Sodium-glucose cotransporter-2
SHI	Statutory Health Insurances
SNOMED	Systematized Nomenclature of Medicine
UK	United Kingdom
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Alzheimer's Disease: Incidence, Prevalence, and Individual's Characteristics

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigators	Rana Jajou Marzyeh Amini Melissa Leung	Erasmus MC
Co-Principal Investigator	Talita Duarte-Salles	Erasmus MC
Data Scientist	Adam Black Ger Inberg Maarten van Kessel Ionna Nika Cesar Barbosa Ross Williams	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
NAJS	Antea Jezidžić Marko Čavlina Karlo Pintaric Anamaria Jurcevic Jakov Vukovic	Croatian Institute of Public Health
DK-DHR	Elvira Bräuner Susanne Bruun	Danish Medicines Agency
InGef RDB	Josephine Jacob Raeleesha Norris Alexander Harms Annika Vivirito	Institut für angewandte Gesundheitsforschung Berlin GmbH
IQVIA DA Germany	Dina Vojinovic Akram Sharim Mendez Rangel Ellen Gerritsen Gargi Jadhav Hugo Vernooij Isabella Kaczmarczyk	IQVIA
IPCI	Katia Verhamme	Erasmus MC
CPRD GOLD	Antonella Delmestri	University of Oxford

*Data partners do not have an investigator role. Data partners execute code at their data source, review and approve their results.

3. ABSTRACT

Title

DARWIN EU® - Alzheimer's Disease: Incidence, Prevalence, and Individual's Characteristics

Rationale and background

Alzheimer's disease (AD) is the most common cause of dementia, characterised by progressive cognitive decline and loss of independence. Its burden is rising globally due to population ageing, with substantial implications for healthcare systems and society. Recent therapeutic advances, such as the approval of anti-amyloid monoclonal antibodies, are reshaping AD management but also highlight the need for comprehensive and updated real-world evidence on the epidemiology and characteristics of affected populations.

This DARWIN EU® study will provide contextual information through a comprehensive overview of the incidence and prevalence of AD in the general adult population in European countries, as well as demographic and clinical characteristics of individuals with AD.

Research question and objectives

Research questions

What are the overall incidence rates and prevalence of AD? What are the demographic characteristics, diagnostic procedures, and clinical profiles of individuals with newly diagnosed AD?

Objectives

1. To estimate overall incidence and prevalence of AD in the general adult population, stratified by calendar year, sex, and age.
2. To describe the demographic characteristics, diagnostic procedures, and clinical profile of individuals with newly diagnosed AD.

Methods

Study design

- Population level cohort study (objective 1)
- Patient level cohort study (objective 2)

Population

Objective 1: All individuals aged 18 years or older present in the data sources during the study period from 01/01/2014 to 31/12/2024, or earliest date of loss to follow-up, death, end of the observation period or data availability, or the earliest date of AD diagnosis with at least 365 days of database history.

Objective 2: All individuals with newly (incident) diagnosed AD aged 18 years or older, diagnosed with AD during the study period from 01/01/2014 to 31/12/2024, with at least 365 days of database history.

Variables

Exposure:

Not applicable.

Outcome:

Alzheimer's disease.

Relevant covariates:

At the diagnosis date, covariates will include sex, age categories (18–55; 56–65; 66–75; 76–85; 86+), and calendar year.

Within 365 days prior to diagnosis, covariates will be diagnostic procedures, such as recording of brain MRI and recording of brain PET-F18, predefined clinical features, including neuropsychiatric symptoms, alterations in activities of daily living (ADLs), and caregiver support, as well as medication use. The latter will include AD drugs and medications for comorbidities, such as antiplatelets, anticoagulants/antithrombotic, glucose-lowering therapies (insulin and oral agents), antihypertensives, antiarrhythmics/rhythm control drugs, lipid-lowering drugs.

At any time prior and up to diagnosis, covariates will include predefined comorbidities of interest, such as Down's syndrome, stroke, atrial fibrillation, myocardial infarction, heart failure, hypertension, diabetes, hypercholesterolemia, and hypertriglyceridemia. In addition, the occurrence of a prior diagnosis of Mild Cognitive Impairment (MCI) will be described.

Data sources

1. Croatia: Croatian National Public Health Information System (NAJS)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Germany: InGef Research Database (InGef RDB)
4. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
5. Netherlands: Integrated Primary Care Information (IPCI)
6. United Kingdom, Clinical Practice Research Datalink GOLD (CPRD GOLD)

Study size

No sample size has been calculated, as this is an exploratory study which will not test a specific hypothesis. Based on a preliminary feasibility assessment, the expected number of persons counts for AD in the data sources included in this study will range from 13,400 (IPCI) to 109,700 (DK-DHR).

Statistical analysis

Objective 1: Yearly incidence rates per 100,000 person-years and prevalence of AD will be estimated, overall and stratified by calendar year, sex, and age categories. Incidence rates and prevalence will be reported with 95% confidence intervals.

Objective 2: Demographic characteristics, pre-specified comorbidities, concomitant medications, diagnostic procedures, and prior cognitive diagnosis will also be described and will be reported as counts and proportions. For MCI, time from MCI recording to AD diagnosis will be described as well.

A minimum cell count of 5 will be used when reporting results, with any smaller count reported as "<5" and zero counts as "0".

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	12 November 2025
Creation of Analytical code	24 October 2025
Execution of Analytical Code on the data	6 November 2025
Draft Study Report	24 November 2025
Final Study Report	To be confirmed by EMA

*Planned dates are dependent on obtaining approvals from the internal review boards of the data sources.

6. RATIONALE AND BACKGROUND

Alzheimer's disease (AD) is a disorder that causes degeneration of the cells in the brain, and it is the main cause of dementia, which is characterised by a decline in thinking and independence in personal daily activities.(1) AD is considered a multifactorial disease associated with several risk factors, such as increasing age, genetic factors, head injuries, vascular diseases, infections, and environmental factors.(1)

The burden of AD is increasing globally, with significant implications for public health and the economy.(2-4) Demographic analyses suggest that these patterns are driven by decreases in fertility coupled with increases in life expectancy, which together lead to large changes in the age structure of the population (i.e., larger number of people at the oldest ages). These changes have led to increases in the number of people affected by dementia, including AD, over time.(2, 4) In 2021, global mortality from AD and other dementias among individuals aged 60 years and older reached approximately 1,922,970 cases (95% confidence interval [CI]: 480,348 to 5,104,315), and the prevalence was 52,560,253 cases (95% CI: 41,399,948 to 65,633,448). Projections suggest a near fourfold increase in AD cases by 2050, driven by population growth and aging, with females disproportionately affected.(5)

This DARWIN EU® study is proposed to provide contextual information through a comprehensive overview of the prevalence and incidence of AD, as well as individual's demographics and clinical characteristics in European countries in the last decade.

7. RESEARCH QUESTION AND OBJECTIVES

Research questions

What are the overall incidence rates and prevalence of AD in the general adult population, stratified by calendar year, sex, and age categories?

What are the demographic characteristics, diagnostic procedures, and clinical profiles of individuals with newly diagnosed AD?

Research objectives

This study aims to provide contextual information on the incidence and prevalence of AD in the general population and demographic- and clinical characteristics of individuals with newly diagnosed AD.

The specific objectives of this study are:

1. To estimate overall incidence and prevalence of AD in the general adult population, stratified by calendar year, sex, and age.

2. To describe the demographic characteristics, diagnostic procedures, and clinical profile of individuals with newly diagnosed AD.

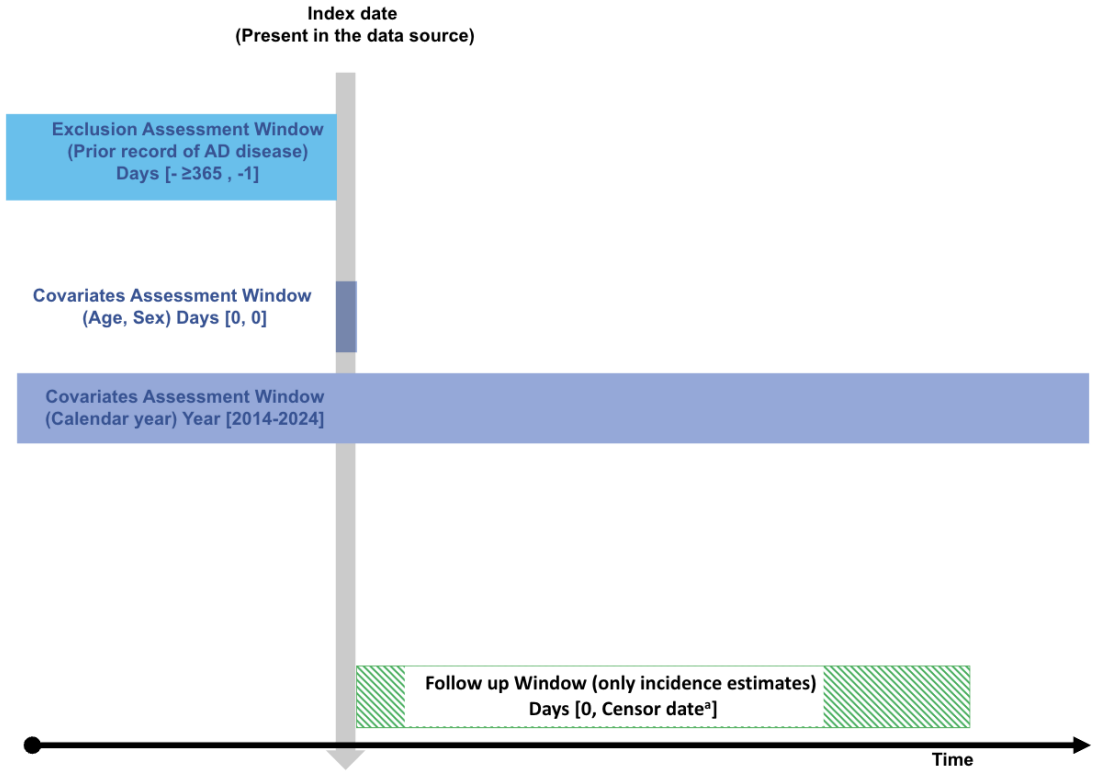
8. RESEARCH METHODS

8.1. Study design

Population level and patient level cohort studies will be conducted using routinely collected health data from six data sources across five European countries.

- A population level characterisation study will be conducted to address objective 1.
- A patient level characterisation study will be conducted to address objective 2.

A graphical description of the study design for each objective is shown in **Figure 1. Graphical depiction of the population level study design (Objective 1).** and **Figure 2. Graphical depiction of the patient level study design (Objective 2).**



AD= Alzheimer disease
a. Earliest date of: AD diagnosis, loss to follow-up, death, end of the observation period or data availability

Figure 1. Graphical depiction of the population level study design (Objective 1).

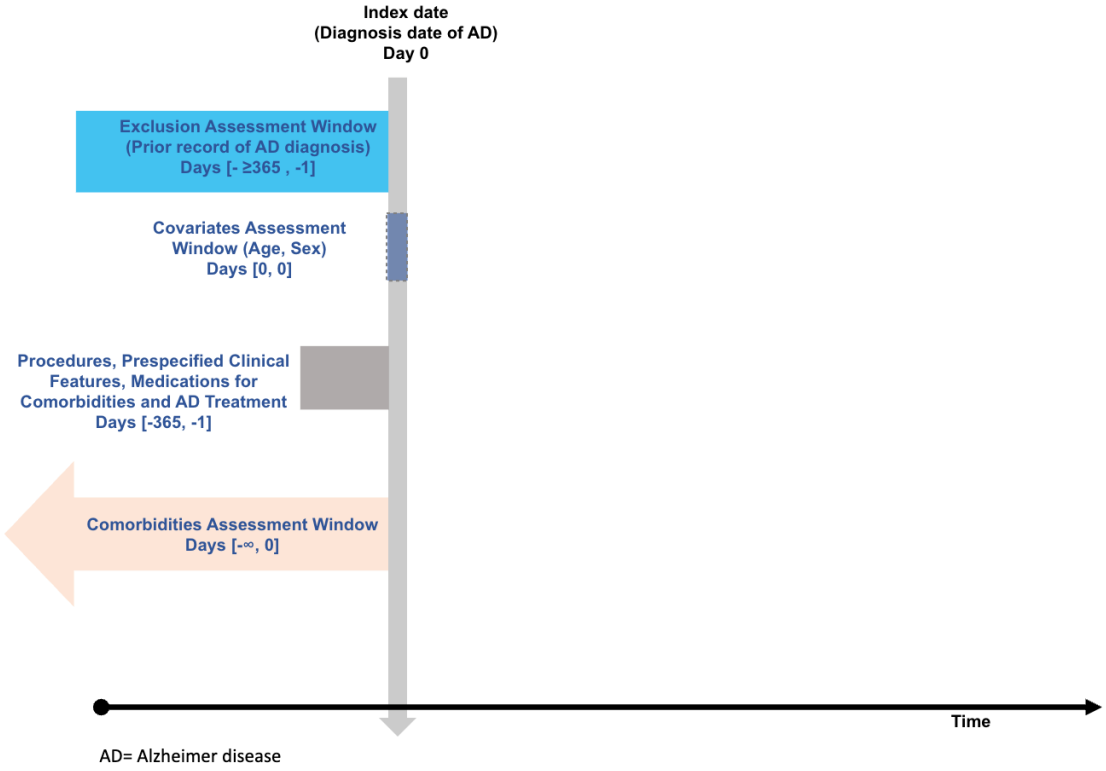


Figure 2. Graphical depiction of the patient level study design (Objective 2).

8.2. Follow-up

The follow-up will be defined as follows for each cohort:

For the cohort of the general adult population (Objective 1):

The index date will be the first date during the study period, on which individuals are minimum 18 years of age and have sufficient prior data availability (minimum 365 days).

Follow-up will start on the study start date (01/01/2014) or the date on which the requirement of prior history is fulfilled.

Follow-up will end on the earliest date of AD diagnosis, loss to follow-up, death, end of observation period (the latest available data), or study end date of 31/12/2024.

For the cohort of individuals with newly diagnosed AD (Objective 2):

The index date will be the date of first (incident) AD diagnosis, on which individuals are minimum 18 years of age and have sufficient prior data availability (minimum 365 days). From the index date, retrospective data will be assessed on comorbidities, diagnostic procedures, prespecified clinical features, and medications for comorbidities and AD treatment.

An example of entry and exit into the denominator population is shown in **ANNEX III. Included observation time for the denominator population.**

8.3. Study population with inclusion and exclusion criteria

The study population will be defined as follows for each cohort:

For the cohort of the general adult population (Objective 1):

Inclusion criteria

- All individuals present during the study period
- Minimum of 365 days of available data history
- Individuals are aged ≥18 at index date

Exclusion criteria

- For incidence estimates: individuals with any prior record of AD diagnosis before index date.

For the cohort of individuals with newly diagnosed AD (Objective 2):

Inclusion criteria

- Individuals with newly diagnosed (incident) AD present during the study period
- Minimum of 365 days of continuous prior observation before the AD index date
- Individuals are aged ≥18 at index date

Exclusion criteria

- Individuals with any prior record of AD diagnosis before index date.
- Individuals with unspecified dementia.

8.4. Study setting and data sources

This study will be conducted using routinely collected data from six primary/secondary care data sources in the DARWIN EU® network of data partners from five European countries. All data were a priori mapped to the OMOP CDM. Information on data sources included and a rationale for their choice in terms of ability to capture the relevant data is described in **Table 1. Description of the selected data sources**. and **ANNEX I. Description of data sources** and **ANNEX II. Fitness for use assessment.**

Table 1. Description of the selected data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
Croatia	National Public Health Information System (NAJS)	Primary care: GPs, secondary care: specialists (ambulatory or hospital outpatient care), hospital inpatient care	Registry	4.3M	2017–2024	Objectives 1 and 2
Denmark	Danish Data Health Registries (DK-DHR)	Community pharmacists, secondary care: specialists (ambulatory or hospital outpatient care), hospital inpatient care	Registry	5.98M	1995–2024	Objectives1 and 2
Germany	InGef Research Database (InGef RDB)	Primary care: GPs, community pharmacists, primary care specialists (e.g., paediatricians), secondary care: specialists (ambulatory or hospital outpatient care), hospital inpatient care, claims data	Claims	7.67M	2015–2024	Objectives 1 and 2
Germany	IQVIA Disease Analyzer Germany (IQVIA DA Germany)	Primary care: GPs, primary care specialists (e.g., paediatricians)	Outpatient General Practitioner Care	4.48M	1992–2024	Objectives 1 and 2
Netherlands	Integrated Primary Care Information (IPCI)	Primary care: GPs	Outpatient General Practitioner Care	1.33M	2006–2024	Objectives 1 and 2
United Kingdom	Clinical Practice Research Datalink (CPRD) GOLD	Primary care	EHR	2.83M	1988–2024	Objectives 1 and 2

NAJS=National Public Health Information System, DK-DHR=Danish Data Health Registries, InGef RDB=InGef Research Database, IQVIA DA=IQVIA Disease Analyzer, IPCI=Integrated Primary Care Information, CPRD=Clinical Practice Research Datalink
GP=General Practitioner, M=millions, EHR=Electronic Health Records

Data sources selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness for this study while covering different regions of Europe (**ANNEX I. Description of data sources** and **ANNEX II. Fitness for use assessment**).

8.5. Study period

The study period covers from 01/01/2014 to 31/12/2024 or the most recent data available for each contributing data source. It should be noted that in the NAJS data source, accurate data are available starting from 01/01 2017 (see **ANNEX I. Description of data sources**), while in InGef RDB, data are available starting from 01/01/2015. Therefore, to meet the 365 days prior history requirement, the study period for NAJS will start from 01/01/2018 to 31/12/2024; and from 01/01/2016 to 31/12/2024 for InGef RDB.

8.6. Variables

8.6.1. Exposure

None.

8.6.2. Outcome

For objective 1, the outcome will be defined as the occurrence of AD recorded in the data sources.

The definition of Alzheimer's disease will be based on SNOMED codes, diagnostic codes explicitly referring to AD or primary degenerative dementia of the Alzheimer type (including presenile/senile onset, uncomplicated, or with depression, delirium, delusions, or behavioural disturbance) will be included. The preliminary concept sets used for the identification of individuals with AD are described in **ANNEX V. List of stand-alone documents.** These codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes,(6) which involve the review of code lists by clinical experts, and the review of phenotypes after their execution in the participating data sources.

For objective 2, AD occurrence is part of the inclusion criteria. There is no outcome for objective 2.

8.6.3. Covariates, including confounders, effect modifiers, and other variables

All objectives:

The covariates are as follows:

- Sex
    - Female/male
- Age at index date will be calculated using January 1st of the year of birth as proxy for the actual birthday. Age groups will be categorised considering that early-onset AD is usually identified in individuals aged 65 years and younger, and late-onset AD in individuals aged (older than 65 years), namely:
    - 18–55 years
    - 56–65 years
    - 66–75 years
    - 76–85 years
    - 86+ years

Objective 1:

- Calendar year: Calendar time will be based on the calendar year of AD incidence and will be from 2014 to 2024.

Objective 2:

The diagnostic procedures and clinical profile of individuals with newly diagnosed AD will be assessed at different time windows of interest as follows:

- Within 365 days prior to diagnosis:
  - Recording of diagnostic procedures:
    - Brain MRI (date of test)
    - Brain PET-F18 (date of test)

    For these procedures, only information on whether the test was recorded/performed (yes/no) derived from the date of the test will be described. Of note, information on test results is not consistently captured across data sources.
  - Clinical features:
    - Neuropsychiatric symptoms: depression, delusions, behavioural disturbances, and delirium, as captured by diagnostic codes.
    - Alterations in activities of daily living (ADLs): any changes, limitations, or dependencies related to an individual's ability to perform basic or instrumental self-care tasks. This category encompasses findings such as alteration in ADL, need for assistance with ADLs, difficulty performing personal hygiene activities, functional dependency, physical functional dependency, unfitness for activity, and limitations in instrumental activities of daily living.
    - Caregiver support: based on recordings of clinical findings related to caregiver support: problems with life management, an individual's dependence on a care provider, or the need for personal care assistance.
    - Alzheimer's disease drugs: include memantine, donepezil, rivastigmine, galantamine. Code list with a preliminary list of products is shown in **ANNEX V. List of stand-alone documents, Table S2**.
    - Medications for predefined comorbidities including antiplatelets, anticoagulants/antithrombotic, glucose-lowering therapies (insulin and oral agents), antihypertensives, antiarrhythmics/rhythm control drugs, lipid-lowering drugs.

    If, during study execution, the counts for ADLs or caregiver support are found to be too low to yield reliable estimates, these variables will be excluded from the final analysis and reported as deviation for the protocol with the corresponding justification.
- At any time prior and up to diagnosis:
  - Predefined comorbidities of interest:
    - Down's syndrome
    - Stroke
    - Atrial fibrillation
    - Myocardial infarction
    - Heart failure

- Hypertension
- Diabetes
- Hypercholesterolemia
- Hypertriglyceridemia

- Prior cognitive diagnosis:
  - Mild Cognitive Impairment (MCI) recorded before AD diagnosis

The preliminary concept sets used for the identification of covariates are described in **ANNEX V. List of stand-alone documents**. All codes will be refined during the study execution following the DARWIN EU® phenotyping standard processes,(6) which involve the review of code lists by clinical experts, and the review of phenotypes after their execution in the participating data sources.

8.7. Study size

No sample size has been calculated, as this is a descriptive disease epidemiology study which will not test a specific hypothesis. In addition, the study will be based on secondary use of data (i.e., data already collected for other purposes than research) to estimate incidence and prevalence of AD in the general adult population. Thus, the sample size is driven by the availability of data for individuals with AD. Based on a preliminary feasibility assessment, the expected number of person-counts for AD diagnosis in the data sources included in this study will range from 13,400 (IPCI) to 109,700 (DK-DHR). These numbers are based on the overall number of conditions or observations registries in each data source with no filter by study period or inclusion and exclusion criteria.

8.8. Analysis

8.8.1. Federated network analyses

All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing individual's data.

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed (see **ANNEX IV. Operational and reporting considerations**), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

8.8.2. Data privacy protection

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in a secure online repository (Data Transfer Zone). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources are checked, after which they are made available to the team, and the Study Dissemination Phase can start. All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing individual-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

8.8.3. Statistical model specification and assumptions of the analytical approach considered

All objectives

We will use the R packages *IncidencePrevalence*, *CohortCharacteristics*, and *PatientProfiles* developed by DARWIN EU®(7) for the population level characterisation, individual-level characterisation, and large-scale characterisation, based on OMOP CDM mapped data.

Objective 1 (Incidence and prevalence of AD)

Incidence calculations

Yearly incidence rates of AD will be calculated as the number of newly diagnosed cases of AD divided by the total person-time at risk, expressed per 100,000 person-years (PY), for each calendar year. Eligible participants will contribute person-time from cohort entry until the earliest of the following: first AD diagnosis during the study period, death, end of observation, or end of the study period (31 December 2024). Participants without a diagnosis will contribute time at risk until censoring, as described in **8.2. Follow-up** section. 95% confidence intervals will be based on a Poisson distribution.

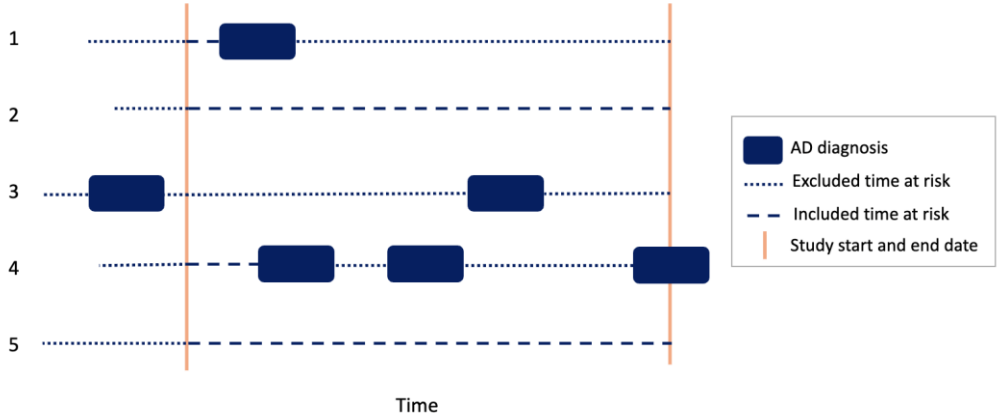


Figure 3. Incidence examples for Alzheimer's Disease (AD).

An illustration of the calculation of incidence of AD is shown above in **Figure 3. Incidence examples for Alzheimer's Disease (AD).** Person IDs 1 and 4 contribute person-time at risk from the study start date until the occurrence of their first recorded AD diagnosis (incident case). Person ID 3 is excluded from the analysis due to a previous record of AD disease. Person IDs 2 and 5 contribute person-time at risk from the study start date until the study end date, as no AD diagnosis is observed during the study period or before study entry. Periods of observation prior to eligibility are excluded from the time at risk.

Prevalence calculations

Prevalence will be calculated as annual period prevalence which summarises the number of individuals with a recorded diagnosis of AD who are alive and actively observed at a given calendar year. Therefore, period prevalence gives the proportion of individuals exposed at any time during a specified interval. Binomial 95% confidence intervals will be calculated.

An illustration of the calculation of period prevalence is shown below in **Figure 4. Example of Period Prevalence Calculation for Alzheimer's Disease (AD).** Between time *t+2* and *t+3*, two of the five study participants have a recorded AD diagnosis, giving a prevalence of 40%. For the period *t* to *t+1*, all five participants contribute observation time, with one of the five having an AD diagnosis, giving a prevalence of 20%.

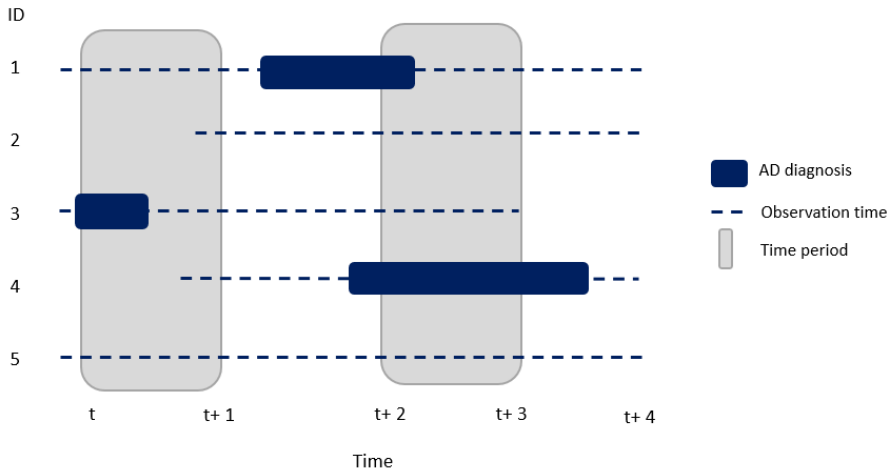


Figure 4. Example of Period Prevalence Calculation for Alzheimer's Disease (AD).

Objective 2 (Diagnostic procedures, individual-level, and clinical profile characterisation)

Descriptive statistics will be used to summarise demographic and clinical characteristics at different time windows as described in the section **8.6.3. Covariates, including confounders, effect modifiers, and other variables.**

Categorical variables (e.g., sex, comorbidities, medication use) will be described using counts and percentages.

Continuous variables (e.g., age at diagnosis) will be described using means, standard deviations, medians, and interquartile ranges.

For individuals with a recorded diagnosis of MCI prior to their first AD diagnosis, the absolute/relative frequency as well as the time elapsed between MCI occurrence and AD diagnosis will be calculated. The duration will be defined as the time in days between the date of the earliest recorded MCI diagnosis and the index date of AD diagnosis. If multiple MCI records are present, the first available record will be used. Time elapsed between recording of MCI and AD diagnoses will be summarised as descriptive statistics (median, interquartile range, mean, standard deviation) and presented both overall and stratified by sex and age categories.

8.8.4. Output

Output will include a PDF report including an executive summary, and the following tables and figures:

- Figure 1. Yearly incidence rates per 100,000 PYs (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD, 2014–2024.
- Figure 2. Yearly incidence rates per 100,000 PYs (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD stratified by sex: (a) Females and (b) Males.
- Figure 3. Yearly incidence rates per 100,000 PYs (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD stratified by age categories.
- Figure 4. Overall prevalence (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD,2014–2024.
- Figure 5. Prevalence (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD stratified by sex.
- Figure 6. Prevalence (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD stratified by age categories.

- Table 1. Distribution of baseline characteristics among participants (number and %, median and IQR, mean and SD) per cohorts of interest by data source.
- Table 2. Distribution of clinical profile characteristics within 365 days prior to diagnosis among individuals with Alzheimer's disease (number and %), by data source.
- Table 3. Distribution of comorbidities at any time prior and up to diagnosis among individuals with Alzheimer's disease (number and %), by data source.
- Table 4. Distribution of prior cognitive diagnosis at any time prior and up to diagnosis among individuals with Alzheimer's disease (number and %), by data source.
- Table S1. Study attrition of individuals included in each cohort during the study period within each data source.
- Table S2 (one for each data sources). Incidence rates per 100,000 PYs (95% confidence intervals) of AD in the general adult population in <data source name> stratified by calendar year, sex, and age categories.
- Table S3 (one for each data sources). Prevalence (95% confidence intervals) of AD in the general adult population in <data source name> stratified by calendar year, sex, and age categories.
- Figure S1. (One for each data sources) Yearly incidence rates per 100,000 PYs of AD in data source (data source name) stratified by sex and age categories.
- Figure S2. (One for each data sources) Prevalence of AD in data source (data source name) stratified by sex and age categories.

An interactive dashboard will be generated by incorporating all the results (tables and figures) included in the PDF report mentioned above.

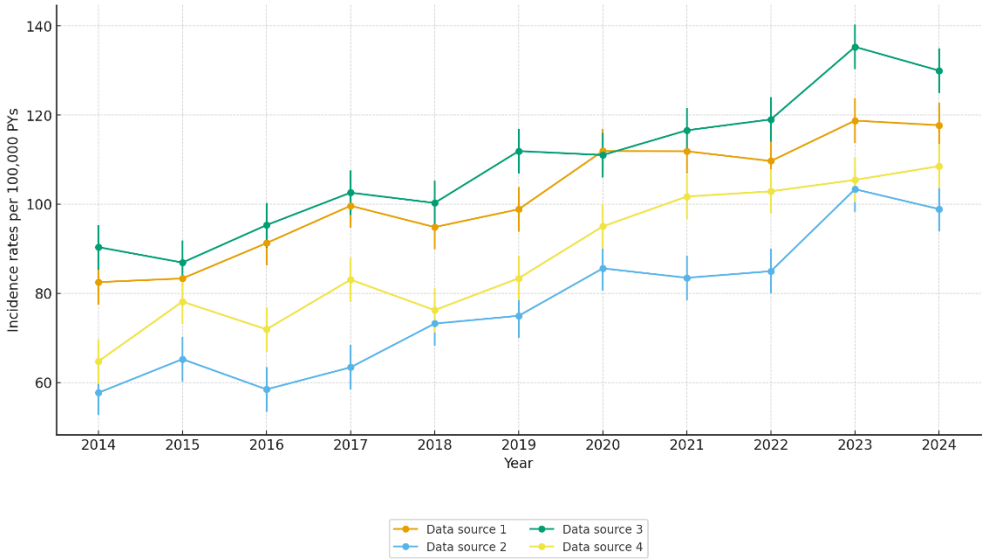


Figure 1. Yearly incidence rates per 100,000 PYs (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD, 2014–2024.

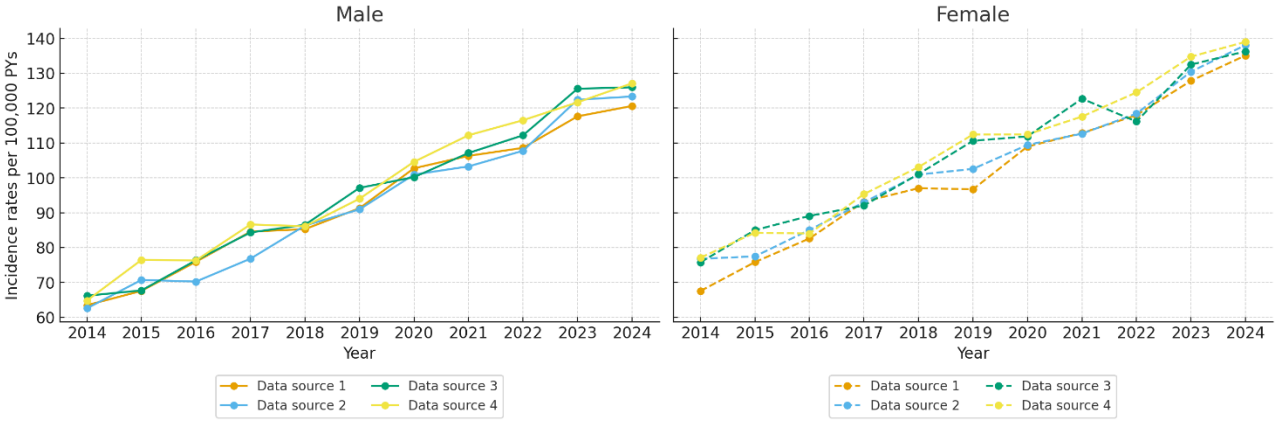


Figure 2. Yearly incidence rates per 100,000 PYs of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD stratified by sex, 2014–2024.

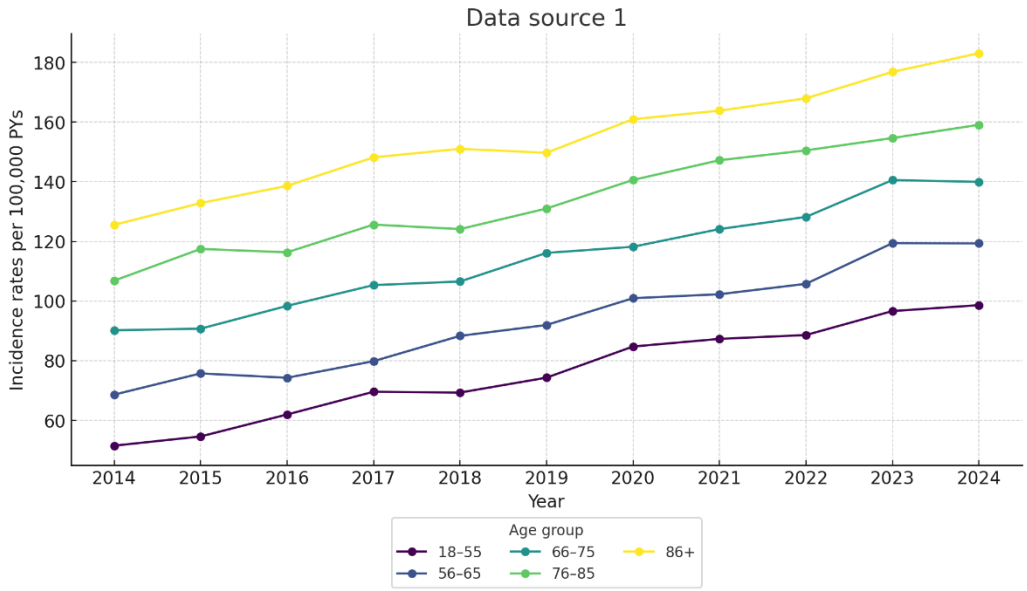


Figure 3. (One for each data source) Yearly incidence rates per 100,000 PYs of AD in data source (data source name) stratified by age categories, 2014–2024.

For all above depicted graphs, details on yearly incidence rates of AD, including the number of participants, follow-up PY, number of individuals with AD, and incidence rate (95% CIs), will be presented in the Appendix table below and via an interactive ShinyApp.

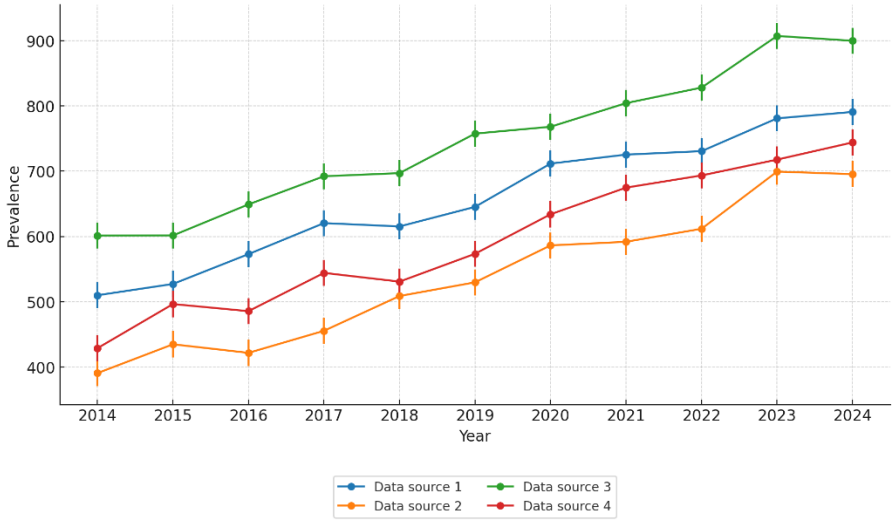


Figure 4. Overall prevalence (95% CIs) of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD, 2014–2024.

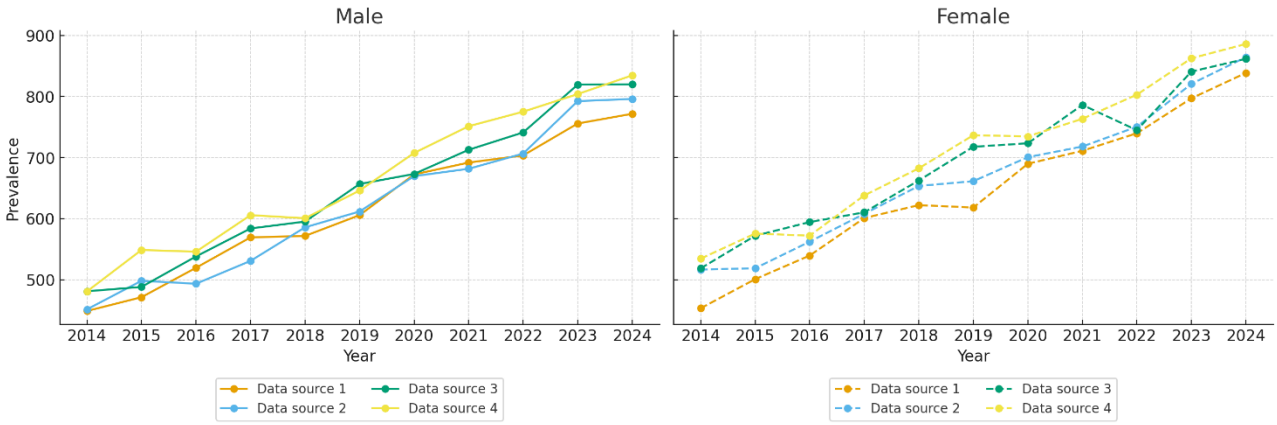


Figure 5. Prevalence of AD in data sources NAJS, DK-DHR, InGef RDB, IQVIA DA Germany, IPCI, CPRD GOLD stratified by sex, 2014–2024.

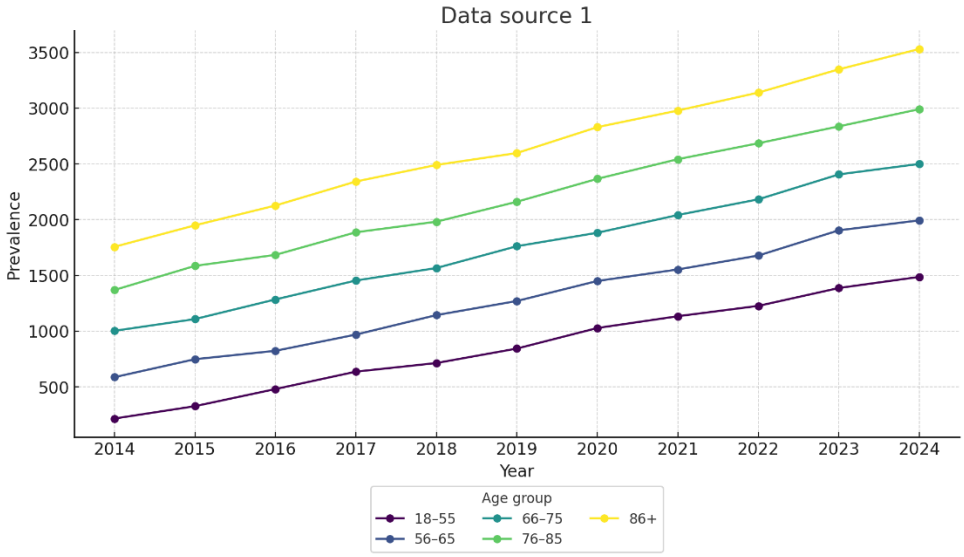


Figure 6. (One for each data source) Prevalence of AD in data source (data source name) stratified by age categories, 2014–2024.

For all above depicted graphs, details on prevalence of AD in the general population, including the number of participants, number of individuals with AD, and prevalence (95% CIs), will be presented in the Annex table below and via an interactive ShinyApp.

Table 1. Distribution of baseline characteristics among participants (number and %, median and IQR, mean and SD) per cohorts of interest by data source.

	Characteristics	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
General adult population	Overall, N						
	Median age (IQR)						
	Mean age (SD)						
	Age groups, in year N (%)						
	18–55						
	56–65						
	66–75						
	76–85						
	86+						
	Sex, N (%)						
	Male						
	Female						
Individuals with Alzheimer disease	Overall, N						
	Median age (IQR) at index date						
	Mean age (SD) at index date						

	Characteristics	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
	Age groups, in year N (%)						
	18–55						
	56–65						
	66–75						
	76–85						
	86+						
	Median index year (IQR)						
	Sex, N (%)						
	Male						
	Female						

IQR=Interquartile Range, SD=Standard Deviation

Table 2. Distribution of clinical profile characteristics within 365 days prior to diagnosis among individuals with Alzheimer's disease (number and %), by data source.

	Characteristics	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
Diagnosis procedures	Brain MRI, N (%)						
	Brain PET-F18, N (%)						
Prespecified clinical features	Neuropsychiatric symptoms, N (%)						
	Alterations in activities of daily living, N (%)						
	Caregiver support, N (%)						
Alzheimer's disease drugs	Memantine, N (%)						
	Donepezil, N (%)						
	Rivastigmine, N (%)						
	Galantamin, N (%)						
Medications for comorbidities	Antiplatelets, N (%)						
	Anticoagulants/antithrombotics, N (%)						
	Oral glucose-lowering drugs, N (%)						
	Insulin, N (%)						
	Antihypertensives, N (%)						

	Characteristics	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
	Antiarrhythmics/rhythm control drugs, N (%)						
	Lipid lowering drugs, N (%)						

MRI=Magnetic Resonance Imaging, PET-F18=Positron Emission Tomography with Fluorine-18

Table 3. Distribution of comorbidities at any time prior and up to diagnosis among individuals with Alzheimer's disease (number and %), by data source.

Characteristics	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
Down's syndrome, N (%)						
Stroke, N (%)						
Atrial fibrillation, N (%)						
Myocardial infarction, N (%)						
Heart failure, N (%)						
Hypertension, N (%)						
Diabetes, N (%)						
Hypercholesterolemia, N (%)						
Hypertriglyceridemia, N (%)						
MCI before AD diagnosis						

MCI=Mild Cognitive Impairment

Table 4. Distribution of prior cognitive diagnosis at any time prior and up to diagnosis among individuals with Alzheimer's disease (number and %), by data source.

Characteristics	Data source 1	Data source 2	Data source 3	Data source 4	Data source 5	Data source 6
MCI, N (%)						
Time from MCI occurrence to AD diagnosis (min-max)						
Time from MCI occurrence to AD diagnosis (median, p25-p75)						
Time from MCI occurrence to AD diagnosis (mean, SD)						

MCI=Mild Cognitive Impairment, AD=Alzheimer's disease, SD=Standard Deviation

ANNEX RESULTS

Table S1. Study attrition of individuals included in each cohort during the study period within each data source.

General	Starting population												
General	Missing year of birth												
General	Missing sex												
General	No observation time available during study period												
General	Doesn't satisfy age criteria during the study period												
General	Prior history requirement not fulfilled during study period												
General	No observation time available after applying												

General	Starting population												
	age and prior history criteria												
Incidence	Starting analysis population												
Incidence	Excluded due to prior event (do not pass outcome washout during study period)												
Incidence	Not observed during the complete data source interval												

Table S2. (one for each data sources) Incidence rates per 100,000 PYs (95% confidence intervals) of AD in the general adult population in <data source name> stratified by calendar year, sex, and age categories.

	Data source name			
	Number of participants	**Follow-up (person-years)**	**Number of individuals with AD**	**Incidence Rates/100,000 PYs (95% CI)**
Calendar year				
2014				
2015				
2016				
2017				
2018				
2019				
2020				
2021				
2022				
2023				
2024				
Sex				
Male				
Female				
Age categories				
18–55				
56–65				
66–75				
76–85				
86+				

AD=Alzheimer's disease, PY=Person-years, CI=Confidence Interval

Table S3. (one for each data sources) Prevalence (95% confidence intervals) of AD in the general adult population in <data source name> stratified by calendar year, sex, and age categories.

	Data source name			
	Number of individuals with AD	**Number of participants**	**Prevalence**	**95% CI**
Calendar year				
2014				
2015				
2016				
2017				
2018				
2019				
2020				
2021				
2022				
2023				
2024				
Sex				
Male				
Female				
Age groups				
18–55				
56–65				
66–75				
76–85				
86+				

AD=Alzheimer's disease, CI=Confidence Interval

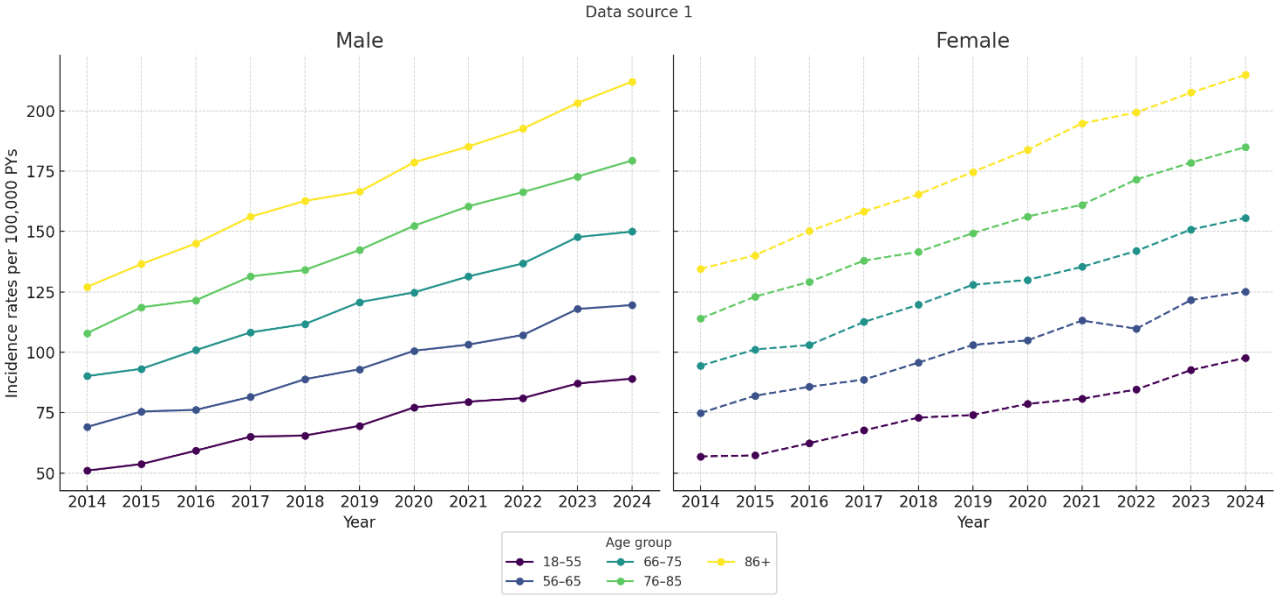


Figure S1. (One for each data sources) Yearly incidence rates per 100,000 PYs of AD in data source (data source name) stratified by sex and age categories.

Figure S2. (One for each data sources) Prevalence of AD in data source (data source name) stratified by sex and age categories.

An interactive dashboard in the Shiny application will be generated by incorporating all the results (tables and figures) included in the pdf report mentioned above.

The results of the large-scale characterisation will be made available exclusively through the interactive Shiny application.

8.9. Evidence synthesis

Results from analyses described in **Section 8.8.3 Statistical model specification and assumptions of the analytical approach considered** section will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

The study will be informed by routinely collected health care data and so, data quality issues must be considered when interpreting the results. In particular, data on genetics (e.g., APOE4, APP, PSEN1, PSEN2), AD-specific biomarkers (amyloid-β, tau), disease staging, and functional or lifestyle factors are either unavailable or inconsistently captured across the selected data sources, restricting the depth of characterisation of individuals. Information on behavioural and psychological symptoms, activities of daily living, caregiver support, and quality of life is also sparse and not uniformly recorded. Variability across databases in coding practices, availability of radiology and neuropsychological assessments, and completeness of comorbidity information may introduce heterogeneity and potential misclassification, particularly when distinguishing AD from other dementias. This study will be based on a phenotype definition which will only include AD disease (code ID: 378419) and MCI before AD diagnosis (code ID: 37398899). Therefore, we may miss early AD cases coded as "unspecified dementia". Finally, due to feasibility and timelines, only a limited number of data sources can be included, which may limit representativeness. Follow-up studies, which will include additional data sources to account for this, will be performed. Moreover, the documentation of comorbidities, necessary for characterisation on the individual-level, may vary across data sources.

Additionally, the calculated estimates will only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purpose rather than primarily for research use. Consequently, using six primary care/ secondary care data sources from Croatia, Denmark, Germany, The Netherlands, and the United Kingdom limits generalisability to those countries.

While OMOP provides mapping to established vocabularies such as SNOMED, inaccuracies or gaps in these mappings can occur, impacting the accuracy and completeness of data in different data sources. Outcome misclassification may also occur due to coding limitations. Certain data sources, such as InGef RDB, and IPCI, include fewer specific codes for certain outcomes of interest, increasing the risk of misclassification.

Another limitation of the InGef RDB claims data source is the potential misclassification of the dates of diagnosis for conditions diagnosed outside of a hospital setting. Unlike inpatient diagnoses, which are associated with an exact diagnosis date and mapped directly to standard concepts, outpatient diagnoses are recorded on a quarterly basis without a specific date. All outpatient diagnoses within a given quarter are documented as occurring on the last day of that period. For example, if an individual is first diagnosed with AD during an outpatient visit on 12 May, the diagnosis will be recorded as occurring on 30 June, the last day of the second quarter. This misclassification can lead to a time lag of up to three months between the date of actual diagnosis and the recorded date in the data source. This can, for example, impact results where the time is calculated from MCI occurrence to AD diagnosis (**Table 4. Distribution of prior cognitive diagnosis at any time prior and up to diagnosis among individuals with Alzheimer's disease (number and %), by data source.**), in which the time would not be accurately calculated**.**

10. REFERENCES

1. Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules [Internet]. 2020; 25(24).

2. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. The Lancet Public Health. 2022;7(2):e105-e25.

3. Liu W, Deng W, Gong X, Ou J, Yu S, Chen S. Global burden of Alzheimer's disease and other dementias in adults aged 65 years and over, and health inequality related to SDI, 1990–2021: analysis of data from GBD 2021. BMC Public Health. 2025;25(1):1256.

4. Wolters FJ, Chibnik LB, Waziry R, Anderson R, Berr C, Beiser A, et al. Twenty-seven-year time trends in dementia incidence in Europe and the United States: The Alzheimer Cohorts Consortium. Neurology. 2020;95(5):e519-e31.

5. Yu DT, Li RX, Sun JR, Rong XW, Guo XG, Zhu GD. Global mortality, prevalence and disability-adjusted life years of Alzheimer's disease and other dementias in adults aged 60 years or older, and the impact of the COVID-19 pandemic: a comprehensive analysis for the global burden of disease 2021. BMC Psychiatry. 2025;25(1):503.

6. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, et al. Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). Pharmacoepidemiol Drug Saf. 2024;33:e70042. https://doi.org/10.1002/pds.70042

7. Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. Pharmacoepidemiology and Drug Safety. 2024;33(1):e5717.

11. ANNEXES

ANNEX I. Description of data sources

Croatia, National Public Health Information System (NAJS)

#	Section	Description
1	Database Identification and country	NAJS (Croatian National Public Health Information System) Croatia
2	Data partner information section	Croatian Institute of Public Health, Department of Data Science and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. Geographic coverage covers whole Croatia, with various levels of resolution for different registries. Current estimates for the population in Croatia will be available at: https://podaci.dzs.hr/hr/podaci/stanovnistvo/procjena-stanovnistva/ for each year.
4	Healthcare setting / type of data	Primary care – GPs, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. For both inpatient and outpatient setting diagnoses, medication, procedures, and measurements are captured. The year of availability of information depends on the setting: • 2014 for lab tests • 2015 for general practitioners • 2016 for secondary conciliatory care • 2017 for hospital records • 2020 for vaccination records
5	Data collection process	Inpatient hospital billing systems, and Other. Data is entered by clinicians at healthcare contact, then combined by CIPH into the NAJS database.
6	General representativeness	The data is collected from public health records, as the majority of health care in Croatia is public. Personal details are collected to a better extent for insured individuals compared to uninsured patients.
7	Data content /source coding	Medication prescriptions are recorded with ATC codes, and diagnoses with ICD10 codes.
8	Data Harmonization	Complete Records from 2017 include insured patients with reliable IDs. Uninsured patients do not have reliable IDs. For example, if a patient changed her status from insured to uninsured, or vice versa, she could be counted several times, as could tracking records from before 2017 and after. By using the unique personal identifier for Croatian citizens, it can be checked and verified.
9	Quality control (database specific)	There is a network of registry personnel (leaders, administrators, coders, sources) working on data coverage and other quality dimensions. An analytical team routinely checks for erroneous entries in hospital records, removing double entries, false dates, and overlapping stays. Entries without enough data or with obviously erroneous dates from primary care analysis are being excluded.
10	Linkage	The national death registry is updated monthly and primary care is updated weekly. Specific registries are included in NAJS (e.g. diabetes registry), where inclusion criteria vary across these registries.
11	Vital status	NAJS is linked to the national death registry.
12	Limitations	Hospital data is available from 2017 onwards. This is often used as start of data collection, while laboratory and GP data is captured before that (since 2014 and 2015 respectively). The total and active person count in the NAJS data is larger than the current population of Croatia. This explained by a) the person table included deceased and all previously insured people and b) there is no information about insurance ending. It is known that a lot of people migrated (300k-400k) and weren't included in the last population census but still are in the NAJS database. In-hospital administrations are managed via paper drug charts and hospital discharge summaries are currently not captured into NAJS.
13	Main references	No main reference provided

#	Section	Description
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111155 Website: https://www.hzjz.hr/nacionalni-javnozdravstveni-informacijski-sustav-najs/

Denmark, Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Database Identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA), Data Analytics Centre (DAC)
3	Coverage and timespan	Data collection since: 1995 Extent: Nation-wide. The data is representative of the entire Danish population.
4	Healthcare setting / type of data	Community pharmacists, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnosis (including rare diseases and pregnancy data), hospital admissions, discharge and ICU data, Cause of death, Drug prescriptions, dispensing, vaccination and contraception, Procedures, Devices, and Sociodemographic information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonization	Complete No.
9	Quality control (database specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g., non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority does running and comprehensive checks of the registry table data validity of the variables, breaks in data, changes in variable coding, missingness, etc. We perform checks of missingness/completeness in relation to requested variables. In essence, we are receiving a dump of a mirror of the data that is controlled by the SDS. The documentation performed by SDS is available online, in Danish primarily https://www.esundhed.dk/Dokumentation (all variables), but also in English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
10	Linkage	There is no linkage in this data source.
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	There are no clinical measurements in the data.

#	Section	Description
13	Main references	Schmidt M, Schmidt SAJ,Adelborg K,Sundbøll J,Laugesen K,Ehrenstein V,Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

Germany, InGef Research Database (InGef RDB)

#	Section	Description
1	Database Identification and country	InGef RDB (InGef Research Database) Germany
2	Data partner information section	Institut für angewandte Gesundheitsforschung Berlin GmbH
3	Coverage and timespan	Data collection since: 2014 Extent: Nation-wide. The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs. The data in the database depicts all health care use which has been reimbursed by the SHI.
4	Healthcare setting / type of data	Primary care – GPs, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The following data elements are collected: pregnancy data, hospital admission and/or discharge also with ICU admission. Prescription, dispensing drugs and Advanced therapy medicinal products. Contraception, medical devices, vaccinations, procedures, diagnoses and demographic information.
5	Data collection process	Insurance/administrative claims. The data in the database depicts all health care use which has been reimbursed by the SHI (statutory health insurances).
6	General representativeness	The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs.
7	Data content /source coding	The ATC and OPS (Operationen- und Prozedurenschlüssel) are used for prescription and dispensing drugs. For Procedures the EBM (Einheitlicher Bewertungsmaßstab - doctor's fee scale) and for ambulatory procedures; OPS (Operationen- und Prozedurenschlüssel) for operations conducted at the hospital are used. Medical events are coded in ICD-10-GM and another vocabulary used is PZN (Pharmazentralzummer -pharmaceutical reference number).
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No. In the German statutory health system, a person can only be enrolled in one health insurance at a time. However, if a person changes from one contributing insurer to another, a new ID number will be generated.
9	Quality control (database specific)	Before entering the InGef database, the data elements are checked with respect to data format, completeness, and plausibility. After each data update, data are compared with the previous data update in regard to number of records, number of data providers, etc.

#	Section	Description
		Due to the anonymized nature of the database, no direct validation of the data (e.g. using medical charts as the gold standard) is possible. Data delivery by health care providers is generally based upon standardized data requirements and formats provided by the National Association of Statutory Health Insurance Funds (compare: https://www.gkv-datenaustausch.de/leistungserbringer/leistungserbringer.jsp)
10	Linkage	No
11	Vital status	The Cause of Death is not captured, the date of death is captured.
12	Limitations	The cause of death is not captured and there is no linkage with other data sources. Approx. 10.5 Million insurees are included in the database, 7.8 Million of these actively insured in 2024. This corresponds to 7% of the total German population. Data are longitudinally linked over a period of currently ten years.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): 26530279
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

Germany, IQVIA Disease Analyser (DA)

#	Section	Description
1	Database Identification and country	IQVIA DA Germany (IQVIA Disease Analyzer Germany) Germany
2	Data partner information section	IQVIA
3	Coverage and timespan	Data collection since: 1989 Extent: Nation-wide. GP and specialists in Germany using specific patient management software.
4	Healthcare setting / type of data	Primary care – GPs, and primary care specialists (e.g., paediatricians). Diagnoses, medication, and procedures from an ambulatory setting. Medications are recorded as prescriptions of marketed products.
5	Data collection process	Outpatient electronic health records. By clinicians at healthcare contact.
6	General representativeness	No specific details on general representativeness given.
7	Data content /source coding	Prescription is on product code level (German PZN), ICD10, NFC, Local lab coding.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. There can be patients registered under different ID numbers, because there is no linkage between different GPs.
9	Quality control (database specific)	Data is quality checked on plausibility.
10	Linkage	No.
11	Vital status	Death information is derived from medical events.

#	Section	Description
12	Limitations	No database-specific limitations documented. General limitations of data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104282 Website: https://www.iqvia.com/

The Netherlands, The Integrated Primary Care Information (IPCI)

#	Section	Description
1	Database Identification and country	IPCI (Integrated Primary Care Information) Netherlands
2	Data partner information section	Erasmus University Medical Center Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – GPs. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. Data is aggregated by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons are mostly uniquely identified, with the exception of when persons change GP practice (when the same individual can receive several different identifiers).
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs. However, in the Netherlands, patients typically have one GP and changing practice is uncommon.
9	Quality control (database specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical

#	Section	Description
		data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M,de Ben C,Leyba AR,Mosseveld BMT,Verhamme KMC,van der Lei J,Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182143
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

Clinical Practice Research Datalink GOLD, United Kingdom (University of Oxford)

#	Section	Description
1	Database Identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – GPs, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records.

#	Section	Description
		Data is entered by clinicians into the EHR. Data is processed by CPRD and provides data releases for research.
6	General representativeness	CPRD GOLD has been assessed and found to be broadly representative of the UK general population in terms of age, gender, and ethnicity. In CPRD GOLD in January 2025 there were 2,730,707 current acceptable patients (i.e. registered at currently contributing practices that use Vision software, excluding transferred out, deceased patients, and those flagged by CPRD as not acceptable for clinical research for data quality issues). This equals to 4.07%, based on the UK population estimates of 67,026,300 from the Office of National Statistics (mid-2023). Current patients are only from Scotland, Wales, and Northern Ireland. Historically, GOLD does contain data from England as well.
7	Data content /source coding	Gemscript, Read, dm+d
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (database specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (uts). Each practice is associated to an uts date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this uts date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice uts date. CPRD also check data quality standards at the patient level and associate each patient to a flag, reporting if its data is acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at Practice Level
11	Vital status	Vital status is retrieved from the GP records. Population registry (ONS) data can be requested on a study-by-study basis and linked. This data only covers England and is planned to be mapped to OMOP in the future. The cause of death is not captured.
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL,Dedman D,Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://cprd.com

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

Croatia, National Public Health Information System (NAJS)

NAJS includes primary care, outpatient specialist care, and inpatient care registries providing information on AD occurrence, radiology, activities of daily living, caregiver support, medications, and relevant comorbidities in the general adult population (≥18 years). The CDM population comprises all publicly insured persons residing in Croatia starting in 2017. The inclusion of NAJS enhances the geographical diversity of data sources, with adequate data coverage over the study period. A preliminary feasibility assessment estimates approximately 28,900 individuals with AD in NAJS.

Data availability and follow-up are sufficient, with records available from 01/01/1995 (accurate data are available starting from the year 2017, see **ANNEX I. Description of data sources**) and the most recent data extraction on 02/08/2025, fully aligned with the study period. No study-specific limitations have been identified for this data source. Furthermore, IRB approval for NAJS is obtained within approximately one month, supporting the feasibility of study execution within the study timelines.

Denmark, Danish Data Health Registries (DK-DHR)

DK-DHR includes nationwide registry data, comprising inpatient, outpatient, and emergency care data from hospitals, providing information on AD occurrence, radiology, medications, and relevant comorbidities in the general adult population (≥18 years). However, caregiver support and activities of daily living are not available. The inclusion of DK-DHR enhances the geographical diversity of data sources, with adequate data coverage over the study period. A preliminary feasibility assessment estimates approximately 107,900 individuals with AD in DK-DHR. The absence of primary care data means that chronic comorbidities will not be detected during this study if they are treated only in primary care. This is also true for behavioural and neuropsychiatric symptoms only registered in primary care. Moreover, no data is available on cognitive test results and functional status. Brain MRI is included in MRI of head, but it was not mapped on a more granular level.

Data availability and follow-up are sufficient, with records available from 01/01/1995 and the most recent data extraction on 10/04/2025, fully aligned with the study period. No study-specific limitations have been identified for this data source. Furthermore, DK-DHR operates under blanket IRB approval, ensuring feasibility of study execution within the current timelines.

Germany, InGef Research Database (InGef RDB)

InGef RDB includes primary care, hospital inpatient care, and secondary outpatient care claims data, providing information on AD occurrence, radiology, medications, and relevant comorbidities in the general adult population (≥18 years). However, activities of daily living, caregiver support, and review of mild cognitive impairment are not available. The CDM population comprises individuals from all regions in Germany who are covered by one of the approximately 50 contributing statutory health insurance companies, ~15% of the German population starting in 2015. The inclusion of InGef RDB enhances the geographical diversity of data sources, with adequate data coverage over the study period. A preliminary feasibility assessment estimates approximately 27,500 individuals with AD in InGef RDB.

Data availability and follow-up are sufficient, with records available from 01/01/2015 and the most recent data extraction on 18/04/2025, fully aligned with the study period. No study-specific limitations have been identified for this data source. Furthermore, InGef RDB operates under blanket IRB approval, ensuring feasibility of study execution within the current timelines.

Germany, IQVIA Disease Analyser (DA)

IQVIA DA Germany includes primary and outpatient secondary care electronic health record data, providing information on AD occurrence, radiology, activities of daily living, caregiver support, medications, and

relevant comorbidities in the general adult population (≥18 years). The inclusion of IQVIA DA Germany enhances the geographical diversity of data sources, with adequate data coverage over the study period. A preliminary feasibility assessment estimates approximately 79,300 individuals with AD in IQVIA DA Germany.

Data availability and follow-up are sufficient, with records available from 01/01/1992 and the most recent data extraction on 10/04/2025, fully aligned with the study period. No study-specific limitations have been identified for this data source. Furthermore, IQVIA DA Germany operates under blanket IRB approval, ensuring feasibility of study execution within the current timelines.

The Netherlands, The Integrated Primary Care Information (IPCI)

IPCI includes primary care electronic health records data, providing information on AD occurrence, radiology, activities of daily living, caregiver support, medications, and relevant comorbidities in the general adult population (≥18 years). The inclusion of IPCI enhances the geographical diversity of data sources, with adequate data coverage over the study period. A preliminary feasibility assessment estimates approximately 13,400 individuals with AD in IPCI.

Data availability and follow-up are sufficient, with records available from 01/01/2006 and the most recent data extraction on 16/04/2025, fully aligned with the study period. No study-specific limitations have been identified for this data source. Furthermore, IRB approval for IPCI is obtained within approximately one month, supporting the feasibility of study execution within the study timelines.

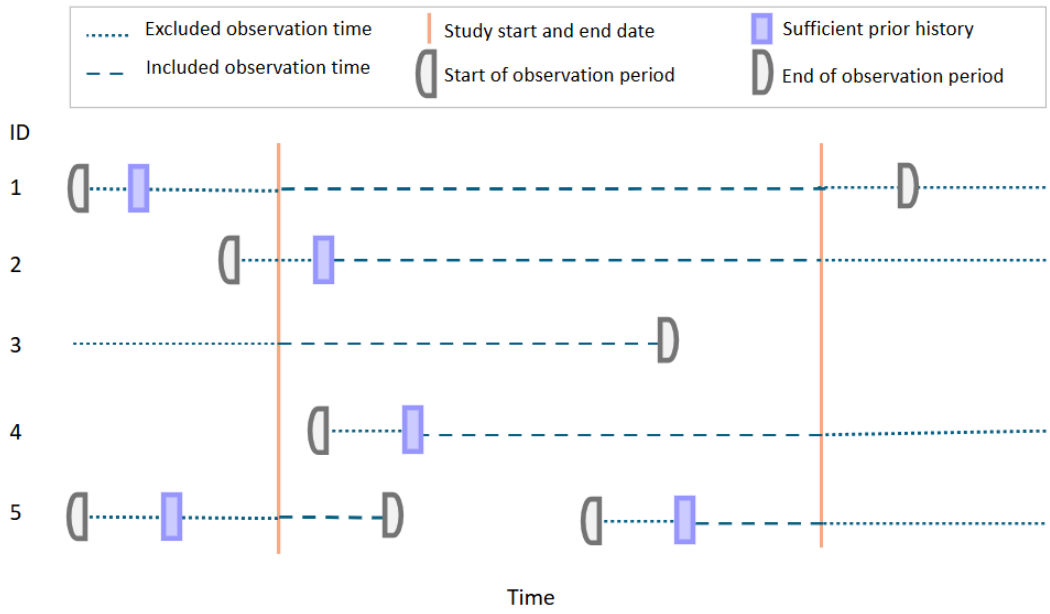
Clinical Practice Research Datalink GOLD, United Kingdom (University of Oxford)

CPRD GOLD includes primary care-GP, primary specialist care, secondary specialist care of outpatient electronic health records, providing information on AD occurrence, radiology, activities of daily living, caregiver support, medications, and relevant comorbidities in the general adult population (≥18 years). The CDM population comprises all persons residing in United Kingdom starting in 2015. The inclusion of CPRD GOLD enhances the geographical diversity of data sources, with adequate data coverage over the study period. A preliminary feasibility assessment estimates approximately 87,500 individuals with AD in CPRD GOLD.

Data availability and follow-up are sufficient, with records available from 01/01/1988 and the most recent data extraction on 15/03/2025, fully aligned with the study period. No study-specific limitations have been identified for this data source. Furthermore, CPRD GOLD operates under blanket IRB approval, ensuring feasibility of study execution within the current timelines.

ANNEX III. Included observation time for the denominator population

In this example, person ID 1 already has sufficient prior history before the start date and the observation period ends after the study end date, so this person will contribute during the complete study period. Person IDs 2 and 4 enter the study only when they have sufficient prior history. Person ID 3 leaves when exiting the data source (the end of the observation period). Lastly, person ID 5 has two observation periods in the data source. The first period contributes time from the study start until the end of the observation period, the second starts contributing time again once sufficient prior history is reached and exits at study end date.



ANNEX IV. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enables the use of standardised analytics and using DARWIN EU® tools across the network, since the structure of the data and the terminology system is harmonised. The OMOP CDM was developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org.

The analytic code for this study will be written in R and will use standardized analytics wherever possible. Each data partner will execute the study code against their data source containing patient-level data and then return the results (csv files), which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for the study report.

Data storage and protection

For this study, participants from various EU member states will process personal data from individuals that is collected in national/regional electronic health record data sources. 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.

All data sources used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses will be run, which generate non-identifiable aggregate summary results.

The output files are stored in the DARWIN EU® Remote Research Environment (RRE). These output files do not contain any data that allow identification of subjects included in the study. The RRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

QUALITY CONTROL

Data source quality control

When defining drug cohorts, non-systemic products will be excluded from the list of included codes summarised on the ingredient level.

When defining cohorts for indications, a systematic search of possible codes for inclusion will be identified using the *CodelistGenerator* R package (https://github.com/darwin-eu/CodelistGenerator). This package allows the user to define a search strategy and will use this to query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* (https://github.com/OHDSI/CohortDiagnostics) and *DrugExposureDiagnostics* (https://cran.r-project.org/web/packages/DrugExposureDiagnostics/index.html) R packages will be run, if needed, to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records.

The study code will be based on DARWIN EU® R packages: *IncidencePrevalence* to estimate Incidence and Prevalence, and *CohortCharacteristics* and *PatientProfiles* for the population level characterisation, individual-level characterisation, and large-scale characterisation. These packages will include numerous

automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

A PDF report including an executive summary, and the specified tables and/or figures will be submitted to EMA by the DARWIN EU® CC upon completion of the study.

An interactive dashboard incorporating all the results (tables and figures) will be provided alongside the PDF report. The full set of underlying aggregated data used in the dashboard will also be made available, if requested.

ANNEX V. List of stand-alone documents

Table S1. Preliminary list of conditions definitions.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID
Alzheimer's disease	Alzheimer's disease	378419	-
	Descendant concepts of Alzheimer's disease: *Altered behavior in Alzheimer's disease, Alzheimer disease with psychosis, Alzheimer's disease co-occurrent with delirium, Alzheimers continuum, Autosomal dominant Alzheimer disease due to mutation of amyloid precursor protein, Autosomal dominant Alzheimer disease due to mutation of presenilin 1, Autosomal dominant Alzheimer disease due to mutation of presenilin 2, Behavioral disturbance co-occurrent and due to late onset Alzheimer dementia, Delusions in Alzheimer's disease, Dementia of the Alzheimer type with behavioral disturbance, Depressed mood in Alzheimer's disease, Early Alzheimers disease, Early onset Alzheimer's disease with behavioral disturbance, Familial Alzheimer's disease of early onset, Familial Alzheimer's disease of late onset, Focal Alzheimer's disease, Frontal variant non-amnestic Alzheimer disease, High level Alzheimers neuropathology changes, Intermediate level Alzheimers neuropathology changes, Logopenic non-amnestic Alzheimer disease, Low level Alzheimers neuropathology changes, Mixed dementia, Non-amnestic Alzheimer disease, Non-familial Alzheimer's disease of early onset, Non-familial Alzheimer's disease of late onset, Primary degenerative dementia of the Alzheimer type, presenile onset, Primary degenerative dementia of the Alzheimer type, presenile onset in remission, Primary degenerative dementia of the Alzheimer type, presenile onset, uncomplicated, Primary degenerative dementia of the Alzheimer type, presenile onset, with delirium, Primary degenerative dementia of the Alzheimer type, presenile onset, with delusions, Primary degenerative dementia of the Alzheimer type, presenile onset, with depression, Primary degenerative dementia of the Alzheimer type, senile onset, Primary degenerative dementia of the Alzheimer type, senile onset in remission, Primary degenerative dementia of the Alzheimer type, senile onset, uncomplicated, Primary degenerative dementia of the Alzheimer type, senile onset, with behavioral*	*Descendant concept IDs:* *44784643, 37166126, 37395572, 3170377, 608051, 603149, 608060, 37117145, 44782726, 43530664, 44782727, 3179057, 44782432, 4043241, 4043243, 4043377, 37167060, 3184947, 3176418, 37167043, 3177168, 43021816, 36716558, 4043242, 4043244, 4218017, 44782941, 4277444, 4277746, 4182539, 4019705, 4220313, 44782940, 4278830, 762578, 4167839, 4204688, 4097384, 1340510, 4043379.*	

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID
	disturbance, Primary degenerative dementia of the Alzheimer type, senile onset, with delirium, Primary degenerative dementia of the Alzheimer type, senile onset, with delusions, Primary degenerative dementia of the Alzheimer type, senile onset, with depression, Progression of Alzheimer's disease, Progressive aphasia in Alzheimer's disease.		
Down's syndrome	Complete trisomy 21 syndrome, Partial trisomy 21 in Down's syndrome,	439125, 4110269	-
Stroke	Normal cardiac stroke volume, Decreased cardiac stroke volume, Stroke test finding, Ischemic stroke, Increased cardiac stroke volume, Haemorrhagic stroke, Autosomal recessive leukoencephalopathy, ischemic stroke, retinitis pigmentosa syndrome, Acute stroke, Late effects of cerebral ischemic stroke, Cerebellar stroke, Stroke co-occurrent with migraine, Epilepsy due to perinatal stroke, Decreased cardiac stroke volume index, Cerebral ischemic stroke due to subarachnoid hemorrhage, Cerebral ischemic stroke due to hypercoagulable state, Cerebral ischemic stroke due to global hypoperfusion with watershed infarct, Cerebral ischemic stroke due to dissection of artery, Seizure disorder as sequela of stroke, Sequela of lacunar stroke, Sequela of cardioembolic stroke, Weakness of face muscles as sequela of stroke, Sequela of thrombotic stroke, Weakness of extremities as sequela of stroke, Occlusion of cerebral artery with stroke, Spasticity as sequela of stroke, Alteration of sensation as late effect of stroke, Weakness as a late effect of stroke, Hemiplegia and/or hemiparesis following stroke	372654, 440426, 761790, 4006295, 4023571, 4045755, 4046363, 4079008, 4088120, 4099974, 4111710, 4153352, 4159140, 4159152, 4168056, 4181404, 4219010, 4236498, 4243337, 4280420, 4310996, 4338810, 35609033, 36675148, 36684840, 36716860, 36716999, 37110241, 37110521, 37118679, 37312013, 37312014, 37312015, 37312017, 43530665, 43530732, 43530736, 43530744, 43531592, 43531595, 43531605, 43531610, 43531617, 44782753, 44782781	3663227, 37110521, 36684840
Atrial fibrillation	Atrial fibrillation	313217	-
Myocardial infarction	Myocardial infarction	4329847	-
Heart failure	Heart failure	316139	-
Hypertension	Hypertensive disorder, Hypertension secondary to endocrine disorder, Hypertension secondary to kidney transplant, Rebound hypertension, Secondary diastolic hypertension, Secondary hypertension, Transient hypertension, Transient hypertension of pregnancy, Transient hypertension of pregnancy - not delivered, Hypertensive complication, Hypertensive renal disease, Hypertensive heart failure, Hypertensive heart AND renal disease	316866, 4110948, 4178312, 4221991, 4253928, 319826, 4199306, 441922, 136760, 42709887, 201313, 444101, 195556	4071202, 762994, 43021830, 137940, 141639, 4062906, 42599748

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID
Diabetes	Diabetes mellitus, Complication due to diabetes mellitus	201820, 442793	-
Hypercholesterolemia	Pure hypercholesterolemia, Hypercholesterolemia	437827, 4029305	-
Hypertriglyceridemia	Hypertriglyceridemia	4120314	-
Brain MRI	Brain MRI measuring method, MRI for measurement of brain volume, MRI of brain, MRI of brain and brain stem without contrast, MRI of brain and cervical spinal cord without contrast, MRI of brain and facial bones without contrast, MRI of brain and internal auditory canal without contrast, MRI of brain without contrast, MRI of cerebrospinal fluid flow without contrast, MRI of head, Magnetic Resonance Imaging (MRI) of Brain using Other Contrast, Magnetic Resonance Imaging (MRI) of Brain using Other Contrast, Unenhanced and Enhanced, Magnetic Resonance Imaging (MRI) of Sella Turcica/Pituitary Gland using Other Contrast, Magnetic Resonance Imaging (MRI) of Sella Turcica/Pituitary Gland using Other Contrast, Unenhanced and Enhanced, Magnetic resonance imaging for measurement of brain volume with contrast, Magnetic resonance spectroscopy of brain without contrast	35810878, 44784284, 37311324, 36713175, 36713041, 36713233, 36713228, 36713262, 36713055, 4082979, 2789360, 2789359, 2789363, 2789362, 44784285, 36713050	
Brain PET-F18	PET CT of brain, PET CT of brain using 18F-FDOPA (fluorodopa 18-F), PET+CT Brain for amyloidosis, PET+CT Brain for tau protein, Positron Emission Tomographic (PET) Imaging of Brain using Fluorine 18 (F-18), Positron emission tomography of brain using fluorodeoxyglucose (18-F), Positron emission tomography with computed tomography of brain using fluorodeoxyglucose (18-F), Positron emission tomography with computed tomography of brain using flutemetamol (18-F)	37109091, 37310737, 36304731, 37021253, 2793369, 35622623, 35608074, 36713668	
Neuropsychiatric symptoms	Altered behavior in Alzheimer's disease, Alzheimer's disease co-occurrent with delirium, Behavioral disturbance co-occurrent and due to late onset Alzheimer dementia, Delusions in Alzheimer's disease, Dementia of the Alzheimer type with behavioral disturbance, Depressed mood in Alzheimer's disease, Disinhibited behavior due to dementia, Early onset Alzheimer's disease with behavioral disturbance, Hallucinations co-occurrent and due to late onset dementia	44784643, 37395572, 37117145, 44782726, 43530664, 44782727, 37311665, 44782432, 37109222	-

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID
Alterations in activities of daily living	Activity of daily living (ADL) alteration, Assisting with activity of daily living, Finding related to ability to perform personal hygiene activity, Functionally dependent, Instrumental activity of daily living, Physical functional dependency, Unfit for activity	4032520, 4128088, 4274102, 36713755, 4044726, 4030753, 44811145	-
Caregiver support	Caregiver support, Problem related to life management difficulty, Patient dependence on care provider, Need for personal care assistance	4303295, 43020485, 4022076, 42535090	-
MCI before AD diagnosis	Mild cognitive impairment review	37398899	-

MRI=Magnetic Resonance Imaging, PET-F18=Positron Emission Tomography with Fluorine-18, CT=Computed tomography, ADL=Activity of daily living, MCI=Mild Cognitive Impairment, AD=Alzheimer's disease

Table S2. Preliminary list of medicines definitions.

Substance Name	Concept name	Ingredient Concept ID (including descendants)	Exclude concept ID
Alzheimer disease drugs			
Memantine	Memantine	701322	-
Donepezil	Donepezil	715997	-
Rivastigmine	Rivastigmine	733523	-
Galantamine	Galantamine	757627	-
Hypertension drugs			
Antihypertensives:			
ACE inhibitors (lisinopril, enalapril)	lisinopril, enalapril	1308216, 1341927	-
ARBs (losartan, valsartan)	losartan, valsartan	1367500, 1308842	-
Beta-blockers (metoprolol, atenolol)	metoprolol, atenolol	1307046, 1314002	-
Calcium channel blockers (amlodipine, diltiazem)	amlodipine, diltiazem	1332418, 1328165	-
Diuretics (hydrochlorothiazide, furosemide, spironolactone)	hydrochlorothiazide, furosemide, spironolactone	974166, 956874, 970250	-
Stroke / Myocardial infarction / Atrial fibrillation / Heart failure drugs			
Antiplatelets:			
aspirin	aspirin	1112807	-
clopidogrel	clopidogrel	1322184	-
ticagrelor	ticagrelor	1116632	-
prasugrel	prasugrel	40163718	-
Anticoagulants/antithrombotics:			-
warfarin	warfarin	1310149	-
heparin	heparin	1367571	-
DOACs (apixaban, rivaroxaban, dabigatran, edoxaban)	apixaban, rivaroxaban, dabigatran etexilate, edoxaban	43013024, 40241331, 40228152, 45892847	-
fondaparinux	fondaparinux	1315865	-
bivalirudin	bivalirudin	19084670	-
Rhythm control drugs:			
amiodarone	amiodarone	1309944	-
flecainide	flecainide	1354860	-
propafenone	propafenone	1353256	-
sotalol	sotalol	1370109	-
dronedarone	dronedarone	40163615	-
dofetilide	dofetilide	1362979	-
Heart failure treatments:			-
ACE inhibitors (lisinopril, enalapril, ramipril, captopril, fosinopril, perindopril, quinapril,	lisinopril, enalapril, ramipril, captopril, fosinopril, perindopril, quinapril,	1308216, 1341927, 1334456, 1340128, 1363749, 1373225,	-

Substance Name	Concept name	Ingredient Concept ID (including descendants)	Exclude concept ID
trandolapril, benazepril, moexipril, cilazapril)	trandolapril, benazepril, moexipril, cilazapril	1331235, 1342439, 1335471, 1310756, 19050216	
ARBs (losartan, valsartan, candesartan, telmisartan)	losartan, valsartan, candesartan, telmisartan	1367500, 1308842, 1351557, 1317640	-
beta-blockers (metoprolol, bisoprolol, carvedilol, nebivolol)	metoprolol, bisoprolol, carvedilol, nebivolol	1307046, 1338005, 1346823, 1314577	-
diuretics (hydrochlorothiazide, furosemide, spironolactone, bumetanide, torsemide, metolazone, chlorothiazide, amiloride, chlorthalidone, indapamide, triamterene, eplerenone)	hydrochlorothiazide, furosemide, spironolactone, bumetanide, torsemide, metolazone, chlorothiazide, amiloride, chlorthalidone, indapamide, triamterene, eplerenone	974166, 956874, 970250, 932745, 942350, 907013, 992590, 991382, 1395058, 978555, 904542, 1309799	-
mineralocorticoid receptor antagonists (spironolactone, eplerenone)	spironolactone, eplerenone	970250, 1309799	-
SGLT2 inhibitors (dapagliflozin, empagliflozin)	dapagliflozin, empagliflozin	44785829, 45774751	-
sacubitril/valsartan	sacubitril / valsartan Oral Tablet	46275724	46275719, 1308842
ivabradine	ivabradine	46234437	-
vericiguat	vericiguat	739665	-
digoxin	digoxin	1326303	-
hydralazine/ isosorbide dinitrate	hydralazine/ isosorbide dinitrate Oral Tablet	40140795	-
Diabetes drugs			
Insulin	Insulins and analogues	21600713	-
Oral glucose-lowering drugs:			
Metformin (biguanide)	biguanide	1593986	-
Sulfonylureas (glipizide, gliclazide)	glipizide, gliclazide	1560171, 19059796	-
DPP-4 inhibitors (sitagliptin, linagliptin)	sitagliptin, linagliptin	1580747, 40239216	-
SGLT2 inhibitors (dapagliflozin, empagliflozin)	dapagliflozin, empagliflozin	44785829, 45774751	-
GLP-1 receptor agonists (liraglutide, semaglutide)	liraglutide, semaglutide	40170911, 793143	-
Hypercholesterolemia / Hypertriglyceridemia drugs			
Lipid-lowering drugs:			
Statins (atorvastatin, simvastatin, rosuvastatin)	atorvastatin, simvastatin, rosuvastatin	1545958, 1539403, 1510813	-
Fibrates (fenofibrate, gemfibrozil)	fenofibrate, gemfibrozil	1551803, 1558242	-
Ezetimibe	Ezetimibe	1526475	-
PCSK9 inhibitors (alirocumab, evolocumab)	alirocumab, evolocumab	46275447, 46287466	-

Substance Name	Concept name	Ingredient Concept ID (including descendants)	Exclude concept ID
Omega-3 fatty acids (for triglyceride lowering)	Omega-3 fatty acids	19106973	-
Down's syndrome drugs			
No disease-specific chronic medications	-	-	-

ACE=Angiotensin-converting enzyme, ARB=Angiotensin II receptor blockers, DOAC=Direct oral anticoagulants, SGLT2=Sodium-glucose cotransporter-2, GLP-1=Glucagon-like peptide-1, PCSK9=Proprotein convertase subtilisin/kexin type 9

ANNEX VI. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: Alzheimer's Disease: Incidence, Prevalence, and Individual's Characteristics

EU PAS Register® number: EUPAS1000000826 **Study reference number:** P4-C1-021

Section 1: Milestones		**Yes**	**No**	**N/A**	**Section Number**
1.1	Does the protocol specify timelines for				
	1.1.1 Start of data collection[1]	☒	☐	☐	8.5
	1.1.2 End of data collection[2]	☒	☐	☐	8.5
	1.1.3 Progress report(s)	☐	☐	☒	
	1.1.4 Interim report(s)	☐	☐	☒	
	1.1.5 Registration in the EU PAS Register®	☐	☒	☐	
	1.1.6 Final report of study results.	☐	☒	☐	

Comments:

Section 2: Research question		**Yes**	**No**	**N/A**	**Section Number**
2.1	Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	
	2.1.1 Why is the study conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	6
	2.1.2 The objective(s) of the study?	☒	☐	☐	7
	2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	8.3
	2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
	2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design		**Yes**	**No**	**N/A**	**Section Number**
3.1	Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	8.8

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

3.4	Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	8.2
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	☒	☐	☐	8.5
	4.2.2 Age and sex	☒	☐	☐	8.3
	4.2.3 Country of origin	☒	☐	☐	8.4
	4.2.4 Disease/indication	☒	☐	☐	8.6
	4.2.5 Duration of follow-up	☒	☐	☐	8.2
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.3

Comments:

Section 5: Exposure definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
5.1	Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☒	
5.2	Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3	Is exposure categorised according to time windows?	☐	☐	☒	
5.4	Is intensity of exposure addressed? (e.g. dose, duration)	☐	☐	☒	
5.5	Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6	Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
6.1	Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	8.6.2

Section 6: Outcome definition and measurement		**Yes**	**No**	**N/A**	**Section Number**
6.2	Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8.6.2
6.3	Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☒	
6.4	Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias		**Yes**	**No**	**N/A**	**Section Number**
7.1	Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	☒	
7.2	Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☒	
7.3	Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☐	☐	☒	

Comments:

Section 8: Effect measure modification		**Yes**	**No**	**N/A**	**Section Number**
8.1	Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☐	☐	☒	

Comments:

Section 9: Data sources		**Yes**	**No**	**N/A**	**Section Number**
9.1	Does the protocol describe the data source(s) used in the study for the ascertainment of:				
	9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	☒	
	9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	8.4
	9.1.3 Covariates and other characteristics?	☒	☐	☐	8.4
9.2	Does the protocol describe the information available from the data source(s) on:				
	9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☒	
	9.2.2 Outcomes? (e.g. date of occurrence, multiple events, severity measures related to event)	☒	☐	☐	8.4

Section 9: Data sources		**Yes**	**No**	**N/A**	**Section Number**
	9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.4, ANNEX I
9.3	Is a coding system described for:				
	9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☒	
	9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	8.4, ANNEX I
	9.3.3 Covariates and other characteristics?	☒	☐	☐	8.4, ANNEX I
9.4	Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	ANNEX I

Comments:

Section 10: Analysis plan		**Yes**	**No**	**N/A**	**Section Number**
10.1	Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8
10.2	Is study size and/or statistical precision estimated?	☒	☐	☐	8.7
10.3	Are descriptive analyses included?	☒	☐	☐	8.8.3
10.4	Are stratified analyses included?	☒	☐	☐	8.8
10.5	Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6	Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	
10.7	Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8	Are relevant sensitivity analyses described?	☐	☐	☒	

Comments:

Section 11: Data management and quality control		**Yes**	**No**	**N/A**	**Section Number**
11.1	Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	ANNEX IV
11.2	Are methods of quality assurance described?	☒	☐	☐	ANNEX IV
11.3	Is there a system in place for independent review of study results?	☐	☐	☒	

Comments:

Section 12: Limitations		**Yes**	**No**	**N/A**	**Section Number**
12.1	Does the protocol discuss the impact on the study results of: 12.1.1 Selection bias?	☒	☐	☐	

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1.2 Information bias? 12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☒ ☐	☐ ☐	☐ ☒	9
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.7, ANNEX II

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have the requirements of the Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	ANNEX II
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	ANNEX IV

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	4

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	ANNEX IV
15.2 Are plans described for disseminating study results externally, including publication?	☐	☐	☒	

Comments:

ANNEX VII. Glossary

Additional definitions are available in the EMA Glossary of terms https://www.ema.europa.eu/en/about-us/glossaries.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g., side effects).

Common Data Model (CDM)

A standardized data structure that enables data from multiple sources to be harmonized, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Centre in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A data source or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g., its content, quality, and structure), essential for identifying relevant data sources in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care data sources, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.

Name of the main author of the protocol: Marzyeh Amini

Date: 12/11/2025

Signature: *M. Amini*