Real4Reg

Unlocking Real-World Data with AI

AARHUS UNIVERSITY

# Use cases in preauthorisation & evaluation
# Use case 1
# Expert workshop 27 January 2026

Lead: Aarhus University (Professor Vera Ehrenstein, Associate professor Erzsébet Horváth-Puhó, statistician István Bakos)

Contributors: DKMA, UEF, Fraunhofer SCAI, BfArM, Infarmed, CSC

# Real4Reg

- Development, optimisation and implementation of **artificial intelligence (AI) methods for RWD analyses** in regulatory decision-making and HTA along the product lifecycle.

- **Duration:** 4 years (January 2023 – December 2026)

- **Consortium**: 10 partners from 6 European countries (Leader BfArM)

Real4Reg

Unlocking Real-World Data with AI

DAC DATA ANALYTICS CENTRE | LÆGEMIDDELSTYRELSEN DANISH MEDICINES AGENCY | AARHUS UNIVERSITY | UNIVERSITY OF EASTERN FINLAND

Federal Institute for Drugs and Medical Devices | HDL | CSC

DZNE

*Partner until December 2025

EIWH

Fraunhofer SCAI | Infarmed Autoridade Nacional do Medicamento e Produtos de Saúde, I.P. | EUPALS

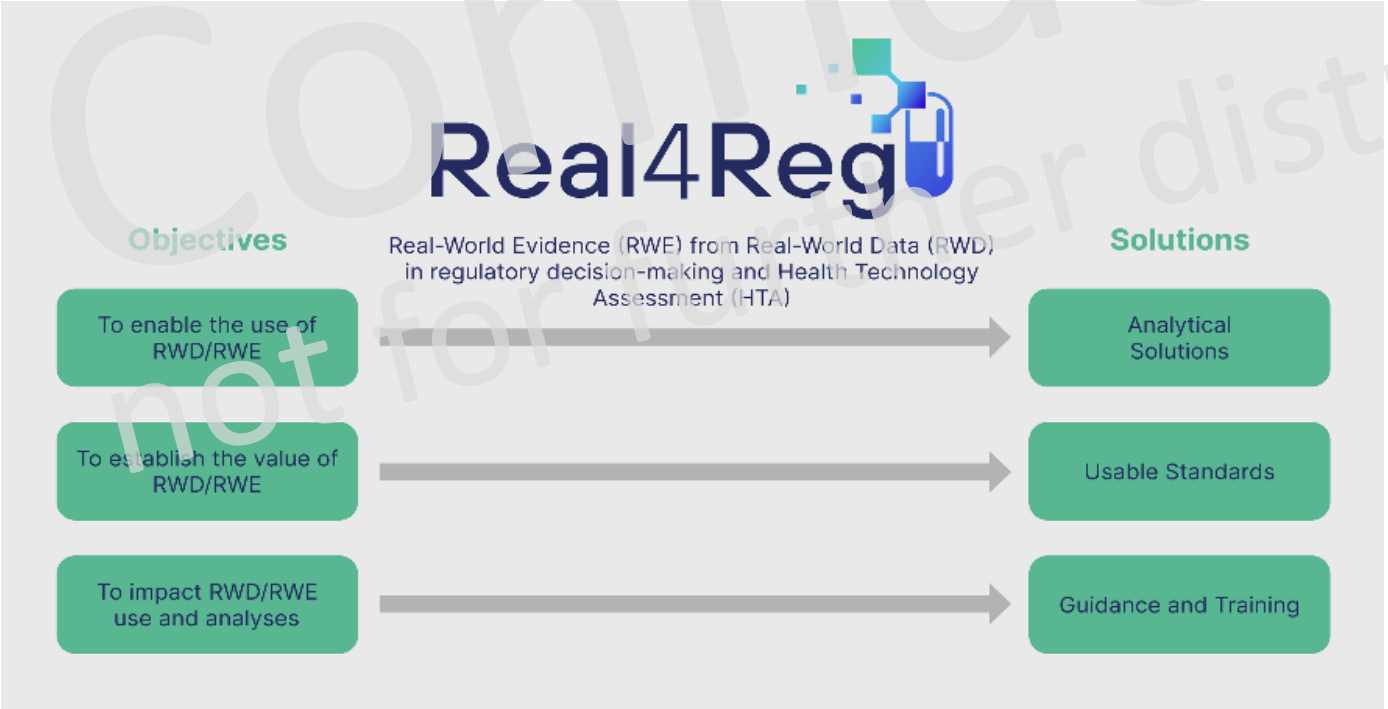
# Real4Reg - Databases

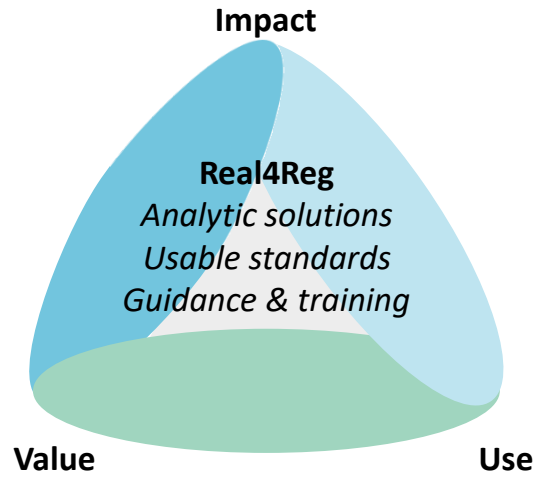
- Real4Reg partners from four different European countries process **individual's personal data** accruing in national/regional electronic **health records databases**

| Country | Size of source population | Data provenance/type |
|---|---|---|
| **Denmark** | 5.9 Million | • Demographics (age, sex)<br>• Procedures<br>• Diagnoses (hospital/specialist/cancer registry)<br>• Dispensings<br>• Healthcare utilisation |
| **Finland** | 5.6 Million | • Demographics (age, sex)<br>• Procedures<br>• Diagnoses (hospital/specialist/cancer registry)<br>• Dispensings<br>• Healthcare utilisation<br>• **Socioeconomic data** |
| **Germany** | 72.8 Million | • Demographics (age, sex)<br>• Claims - diagnoses<br>• Claims - prescriptions<br>• **Socioeconomic data** |
| **Portugal** | 10.3 Million | • Demographics (age, sex)<br>• Procedures<br>• Diagnoses<br>• Dispensings<br>• Healthcare utilisation |

# Real4Reg - Objectives

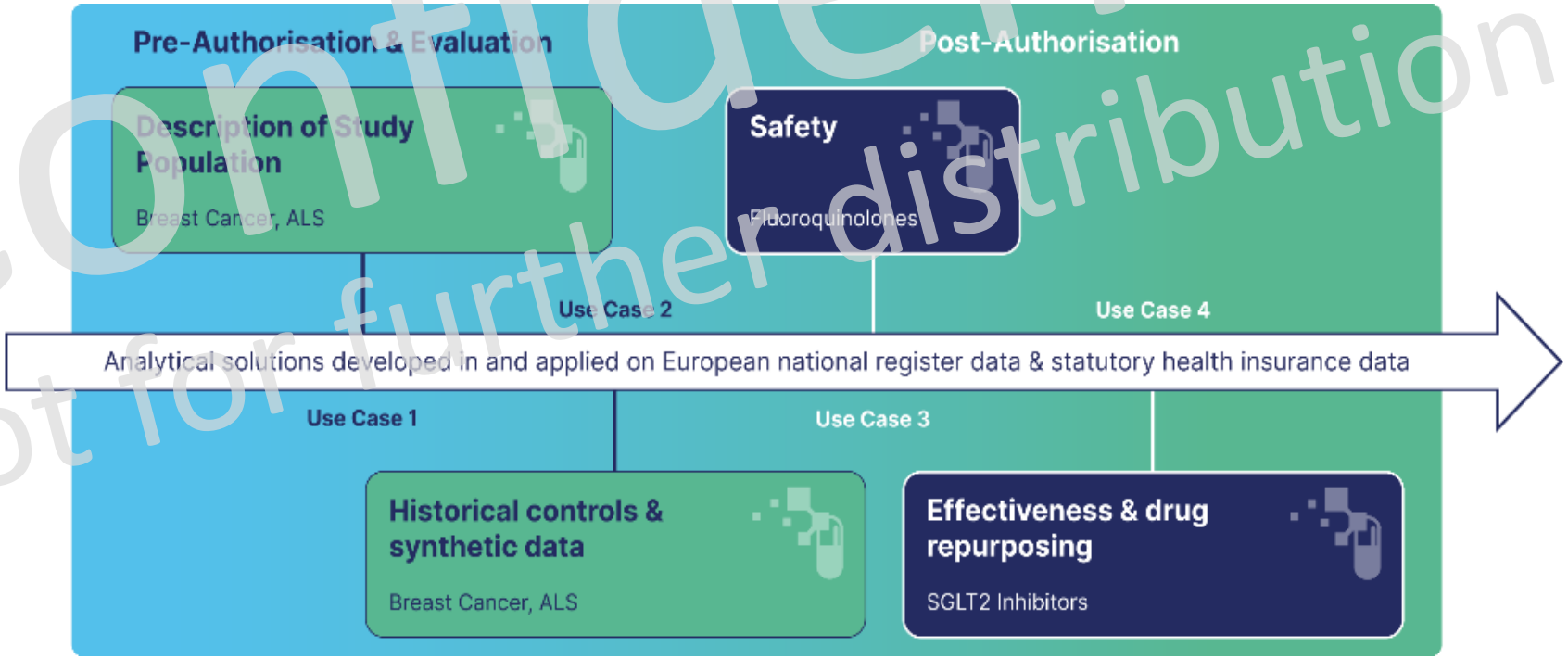
- Development of **new and optimised methods for RWD analysis**, addressing the challenges and opportunities across different health care systems. Ultimately, this will support **better decision-making** about medicines, benefiting patients' health.



# Real4Reg – Use Cases

- To address the need to **demonstrate the potential of RWE in regulatory decision-making and HTA**, Real4Reg is employing different highly relevant use cases from regulatory practice and across the product lifecycle.

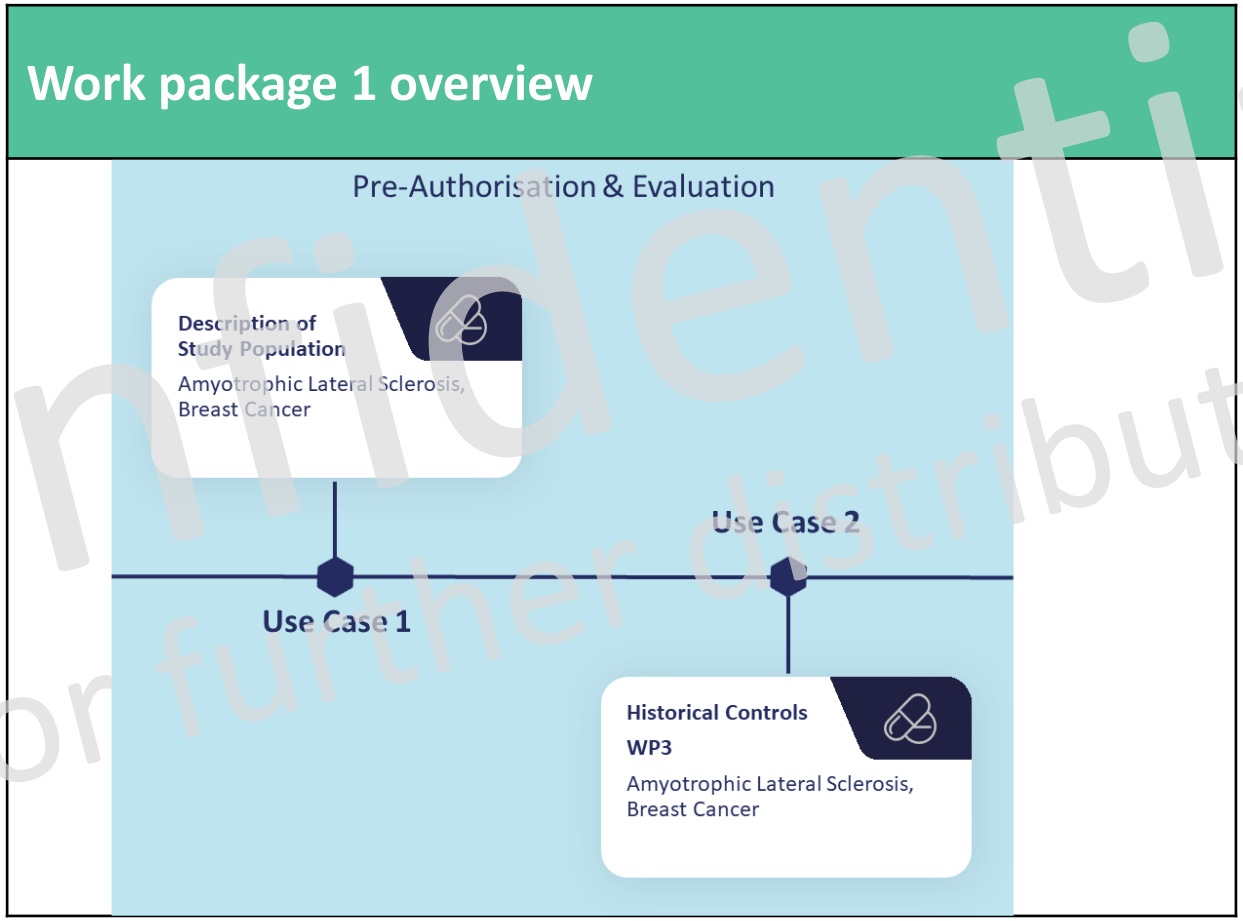


# Background and Rationale: Use Case 1

- **Randomised controlled trials (RCTs)** remain the primary requirement for the initial authorisation of treatments based on their efficacy and safety
- **Real-world evidence (RWE)** – evidence generated from real-world data (RWD) – has been traditionally used for assessing postauthorisation safety and effectiveness monitoring
- Regulators, including the European Medicines Agency (EMA), recognise the potential of **RWE to complement RCTs in regulatory decision-making**, including the preauthorisation phases, especially where RCTs have limitations (e.g., rare diseases, severe conditions, vulnerable subgroups, lack of equipoise)

# Barriers to RWE use in the Preauthorisation & Evaluation Phase

- Lack of awareness about the available data sources
- Stakeholders' skepticism/concern about data quality
- Slow data access for regulatory and HTA purposes






ALS: amyotrophic lateral sclerosis

# Rationale for Selecting ALS and Breast Cancer

- The two diseases represent two **diverse** preauthorisation settings in which RWE use may be relevant
- Breast cancer the most **common** and a well studied malignancy in women, with well-developed treatment protocols. Prognosis differs by patient characteristics. Men with breast cancer represent an **understudied** small **subgroup** with less optimal prognosis than women
- ALS is a **rare** motor neuron disorder, characterised by high fatality and absence of cure. Pharmacotherapy with riluzole is the only approved option used for symptom relief

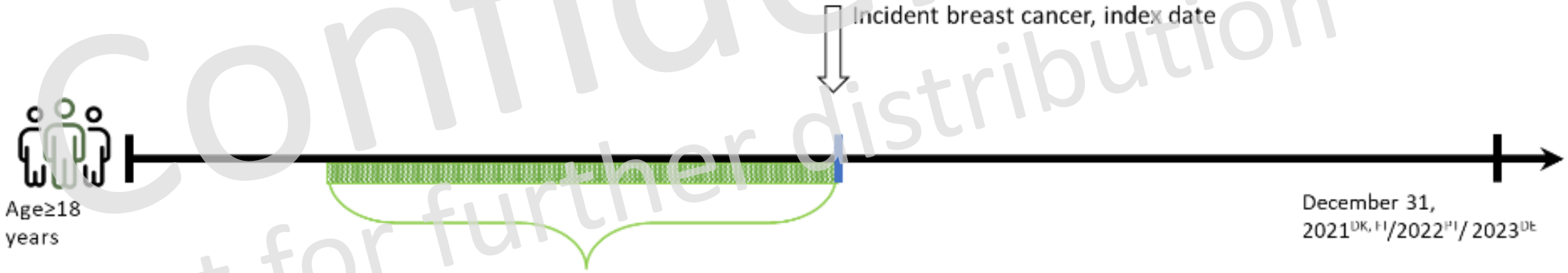
# Rationale for Selecting ALS and Breast Cancer

Because evidence from RCTs in rare breast cancer subgroups and in ALS patients is difficult to obtain due to small number of affected patients and strict eligibility criteria required by RCTs, these patient populations may stand to benefit from use of RWE in the preauthorisation phase

# Use Case 1: Specific Objectives

- Obtain **data access** and carry out data pre-processing tasks including quality control and conversion of the native data in each data source to the OMOP common data model (CDM)
- Provide a **descriptive overview** of the breast cancer and ALS populations, including premorbid characteristics, disease characteristics, treatment, and prognosis – as a proof of concept
- Examine sources of **heterogeneity** of the results across the country-specific patient populations and across data provenance, coding practices, accessibility, representativeness, completeness, temporal variation in variables' availability, and validity

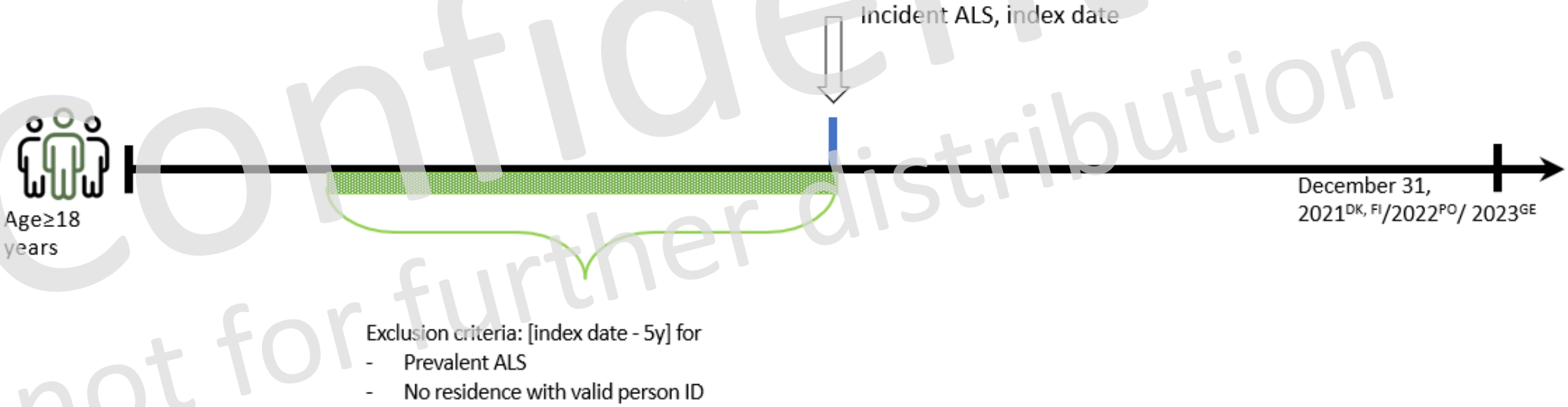
# Methods – Study Population and Design Breast Cancer




Exclusion criteria: [index date-5 y] for

- Prevalent malignancy (except nonmelanoma skin cancer)
- No residence with valid person ID

# Methods – Study Population and Design ALS



# Data Access and Management

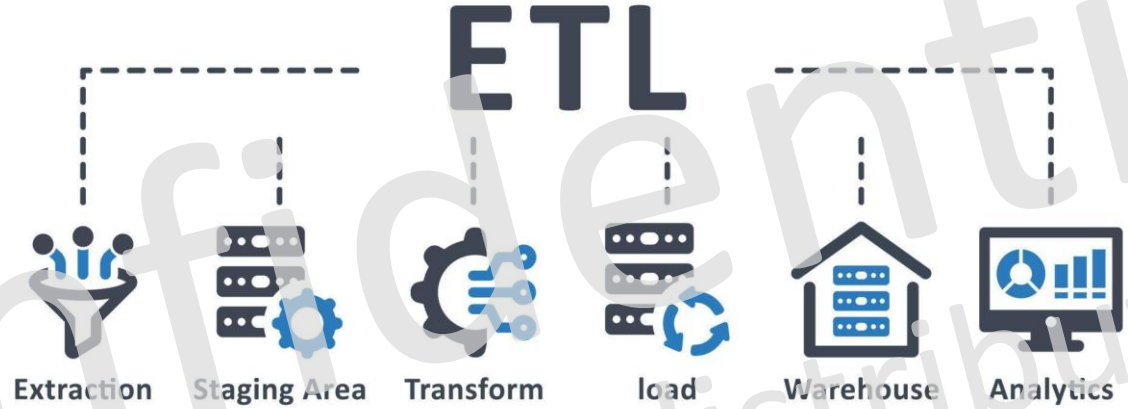
1. Ethical permissions and other legal requirements fulfilled
2. **Protocol-based data application at each country's data custodian**
3. **Upon access: local data checking, cleaning, plausibility checks**
4. Data coding according to the common protocol with local "flavours"



Protocol available from the HMA-EMA Catalogues of real-world data sources and studies EUPAS105544

# OMOP Implementation

# Extract Transform Load Flow




Medical databases from Finland, Denmark, Portugal, ...

Cleaning and normalizing

OMOP Common Data Model



https://www.vecteezy.com/vector-art/17094538-etl-icon-vector-illustration-etl-extract-transform-load-extraction-staging-area-data-warehouse-analytics-infographic-template-presentation-concept-banner-icon-set-icons

# ETL Process Insights

| Initial Design | Adapted Design |
| --- | --- |
| • Based on available metadata | • After access to data<br>• Different variable names and table structure<br>• Switching to in-process database due to memory issues |

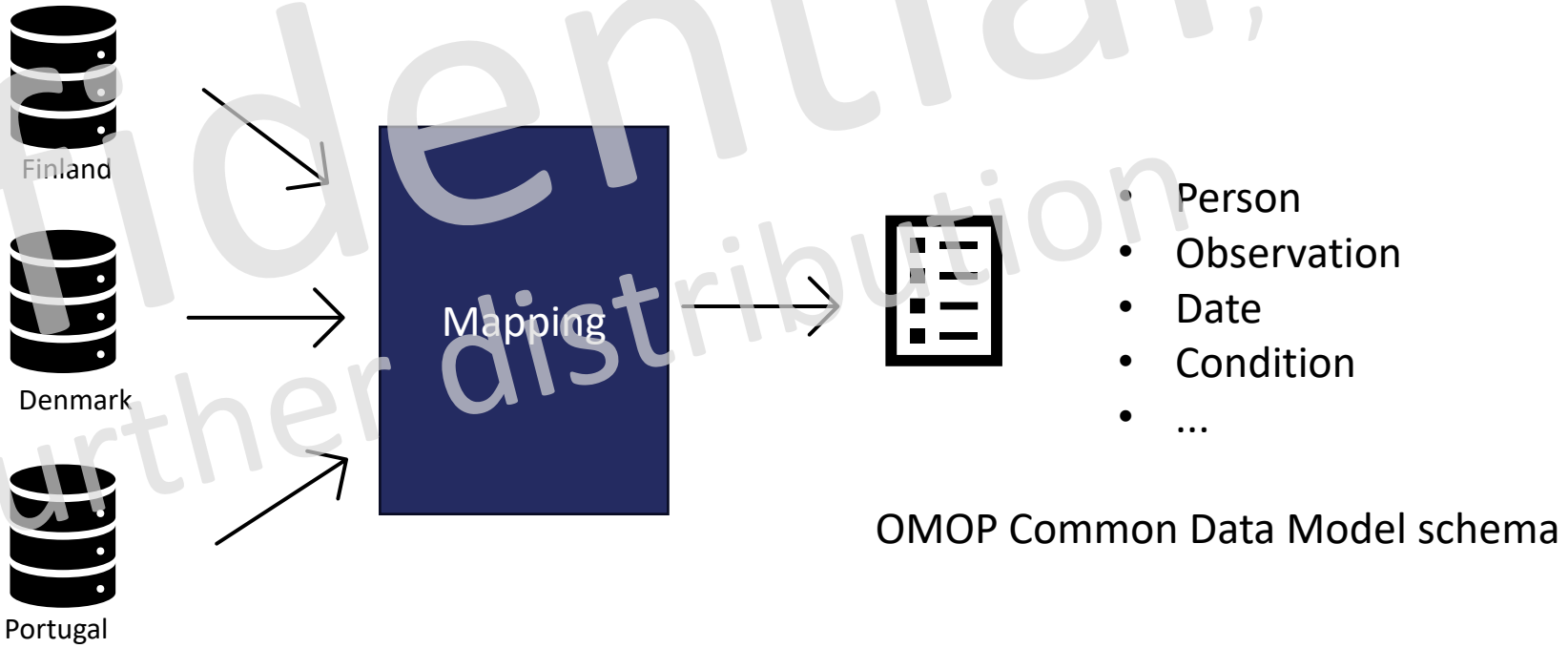
Adaptation of scripts was required

**Learnings**

Importance of early availability of data and of accurate design of metadata

# Target Schema Mapping

- Effective mapping of country-specific codes to a standardized vocabulary is essential for data integration and analysis in health research.
- This mapping ensures that disparate data sources can be aligned and analyzed cohesively within the **Observational Medical Outcomes Partnership** (OMOP) framework.

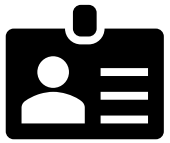


# Vocabulary Mapping through Athena

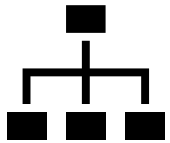
- Athena, a searchable database maintained by the **Observational Health Data Sciences and Informatics** (OHDSI), facilitates this mapping process.

- It assists users in identifying local medical codes and matching them to their corresponding OMOP equivalents by providing crucial information during the mapping process.

Domain ID
Categorises relevant health information



Concept IDs
Unique identifiers for related medical concepts



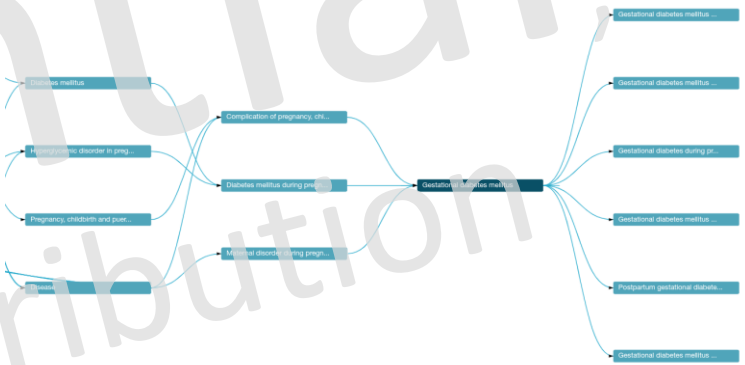
Hierarchical relationships
Parent and child terms

# Vocabulary Mapping through Athena: Example

Example: **Gestational diabetes mellitus**

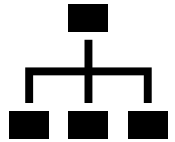
Condition

40395652
Vocabulary: SNOMED
Concept Class ID: Disorder



Domain ID
Categorises relevant health information

Concept IDs
Unique identifiers for related medical concepts



Hierarchical relationships
Parent and child terms

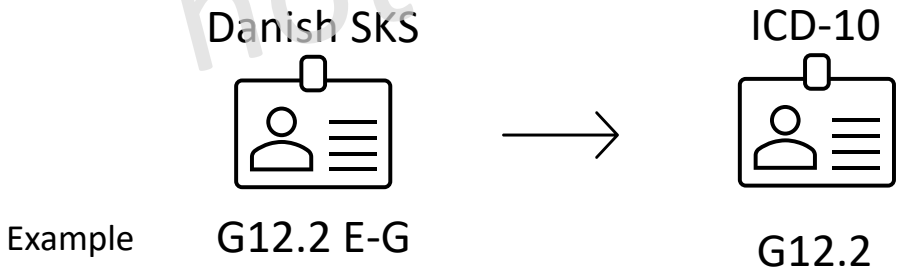
# Challenges with Athena

## Missing specific codes

- Country-specific codes for, e.g., special reimbursement or socioeconomic codes are not present

## Missing entire vocabularies

- Includes e.g. ICD10-GM or CM
- Misses Danish healthcare classification system SKS




Example: certain chemotherapy treatment regimens are non-standard or country-specific; Usage of HemOnc vocabulary for mapping non-standard vocabulary procedures

| | | | | | |
|---|---|---|---|---|---|
| BWHA135 | VCR | 0.76 | APPROVED | 35807043 | Vincristine monotherapy |
| BWHA429 | ponatinib | 0.58 | APPROVED | 35804084 | Ponatinib monotherapy |
| BWHC31 | goserelin | 0.73 | APPROVED | 35804276 | Goserelin monotherapy |
| BWHA253 | gemcitabin+oxaliplatin+capecitabin | 0.57 | FLAGGED | 35804755 | CapeOx |
| BWHA447 | binimetinib | 0.63 | APPROVED | 35806140 | Binimetinib monotherapy |
| BWHA176 | fludarabinfosfat+cyclophosphamid+mitoxantron (FMC) | 0.53 | APPROVED | 35804577 | FCM |

**Learnings:**

- Creation of manual mappings necessary
- Simplification of extended Danish ICD codes which were more specific than international ICD-10 version
- Mapping of unique ICD codes was unfeasible

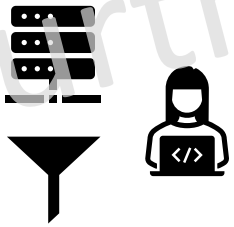
# Missing Information In Source Data

- OMOP CDM is very detailed and assumes much information to be recorded in the source

In reality …

Mixtures of structured data and free text

E.g. number of dispensed drugs in Finland



Manual editing required for extracting information

Only possible for a selected subset

Missing details

E.g. drug_exposure_end_date and days_supply not reported in all cases

Learnings:
Importance of standard input formats and necessity of complete information for drug administration

# Ambiguity of Standard Mappings: ATC

- **Introducing standard vocabularies:**
  - **ATC (Anatomical Therapeutic Chemical Classification)**: A classification system that categorises drugs based on their anatomical and therapeutic properties, assigning a unique code to each drug or group of drugs.

Example: Metformin

| | |
|---|---|
| **A** | Alimentary tract and metabolism<br>(1st level, anatomical main group) |
| **A10** | Drugs used in diabetes<br>(2nd level, therapeutic subgroup) |
| **A10B** | Blood glucose lowering drugs, excl. insulins<br>(3rd level, pharmacological subgroup) |
| **A10BA** | Biguanides<br>(4th level, chemical subgroup) |
| **A10BA02** | Metformin<br>(5th level, chemical substance) |

https://www.who.int/tools/atc-ddd-toolkit/atc-classification

**ATC 1st level**

The system has fourteen main anatomical or pharmacological groups (1st level). The ATC 1st levels are shown in the figure.

**ATC 2nd level**

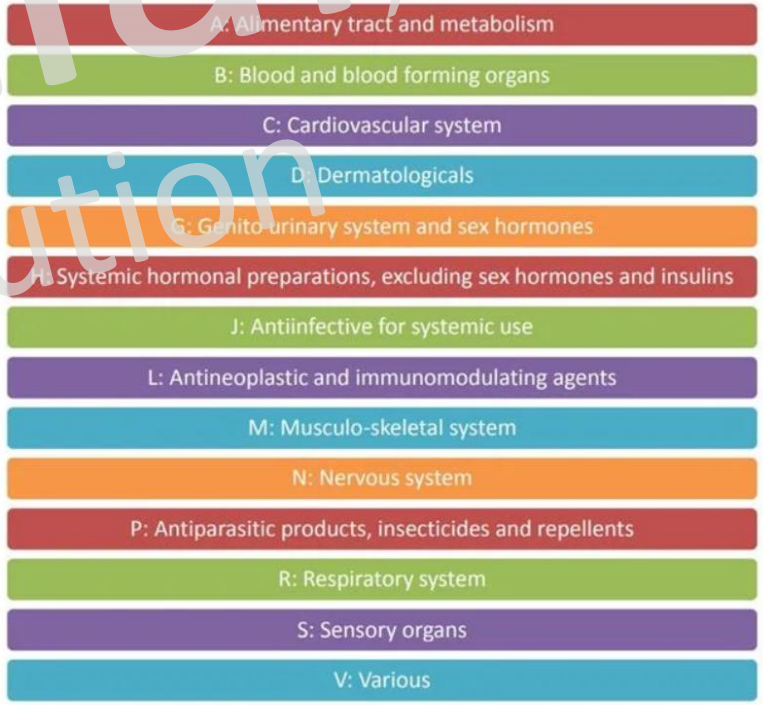
Pharmacological or Therapeutic subgroup

**ATC 3rd& 4th levels**

Chemical, Pharmacological or Therapeutic subgroup

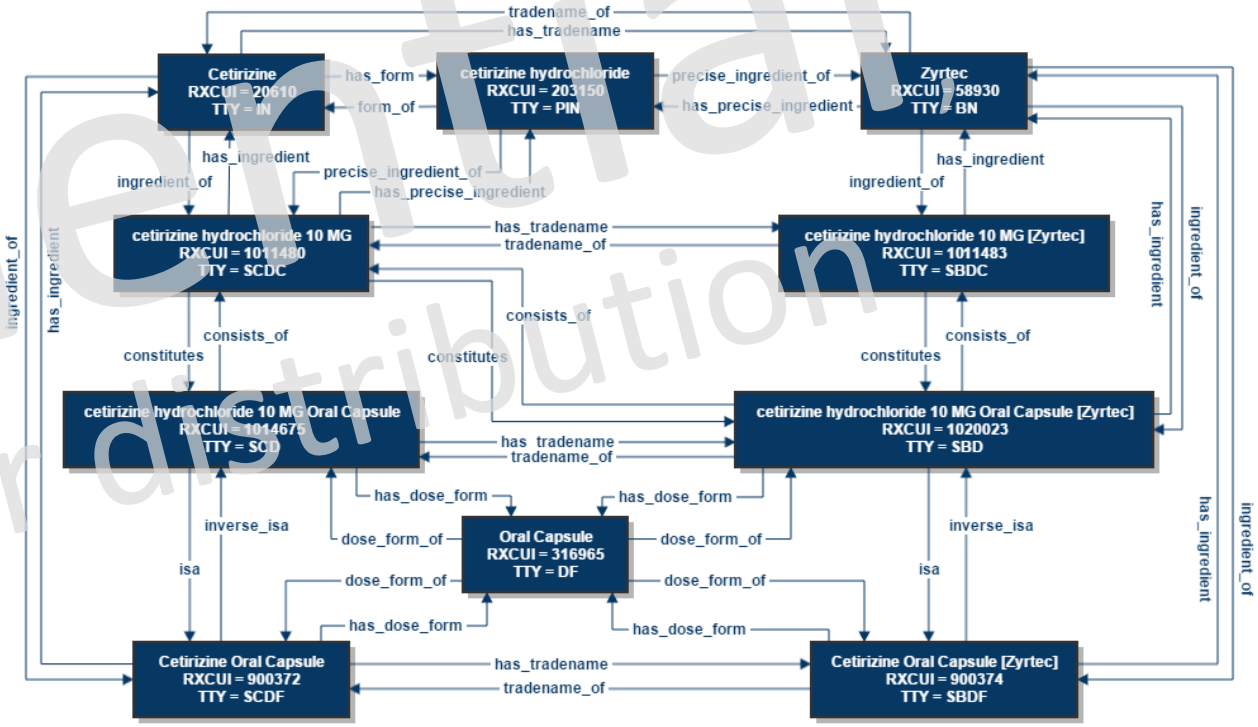
**ATC 5th level**

Chemical substance



# Ambiguity of Standard Mappings: RxNorm

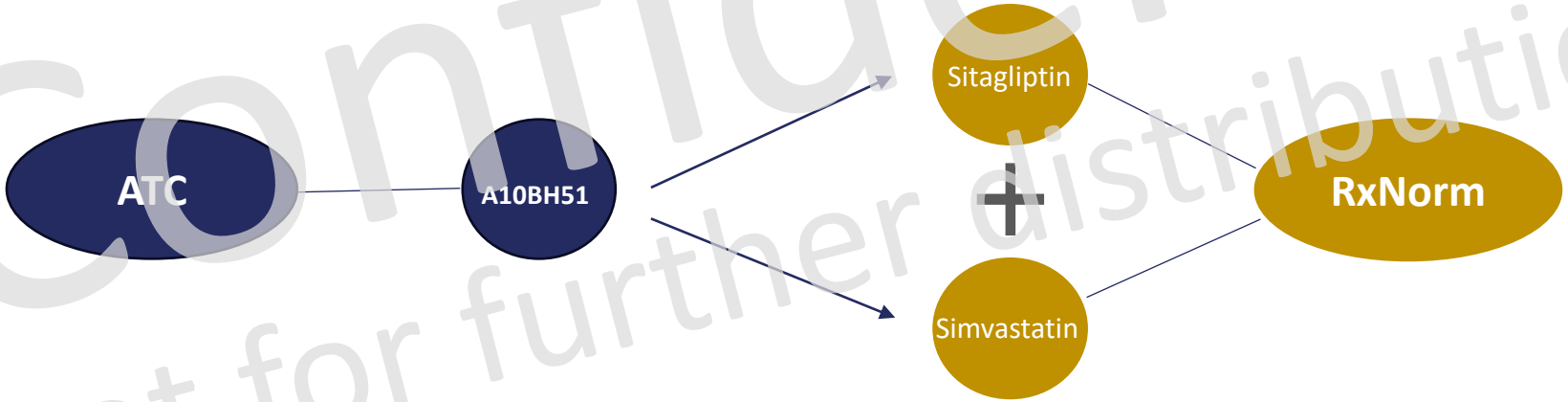
- **Introducing standard vocabularies:**
  - **RxNorm**: A standardised nomenclature for clinical drugs that provides normalised names and identifiers for medications and links them to various drug classification systems
  - **Linking country specific codes to standard vocabularies: FinOMOP added VNR (Nordic Article Number) which links to standard concepts from RxNorm (**A six-digit identification code for a specific medicine with marketing authorisation in the Nordic countries)




https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

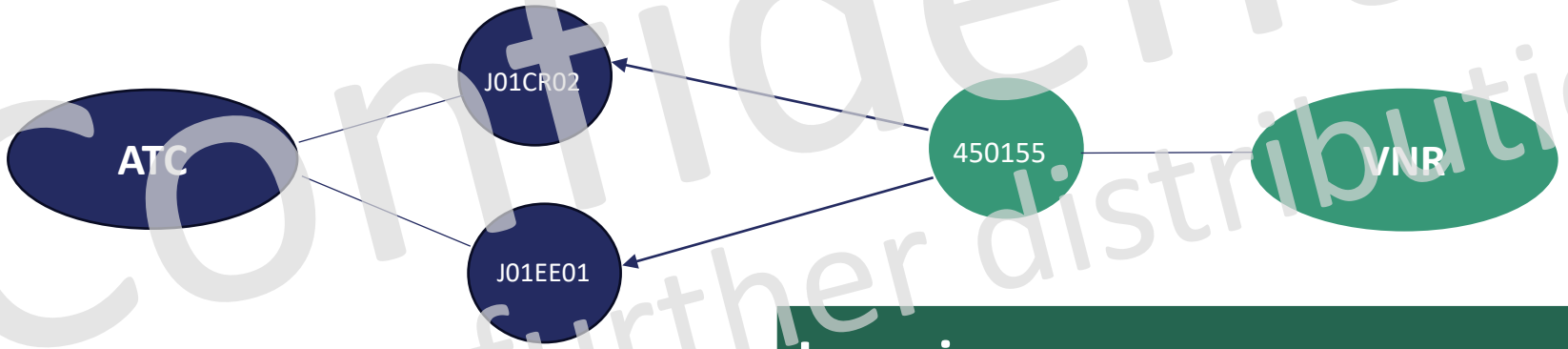
# Ambiguity of Standard Mappings: ATC and RxNorm

- **Challenges of One-to-Many Mappings:**
  - One-to-many mappings occur when ATC codes represent products consisting of multiple components, leading to mapping to distinct RxNorm concepts.



# Ambiguity of Standard Mappings: ATC and VNR

- **Challenges of One-to-Many Mappings:**
  - VNRs should be unique, but exceptions were found where they represent the same ATC codes due to **multiple indications** or **formulation variations**




**Learnings:**

Relying solely on standard RxNorm concepts can lead to an increase in drug counts, making manual verification of mapped codes necessary or the direct use of source concept IDs (from ATC or VNR) advisable.

# Ambiguity of Standard Mappings: Impact

- **Problem:** Data inflation results when processing a single row of native data generates multiple rows in the OMOP format.

-> The standard OMOP format lacks a mechanism for preserving the linkage between these related rows, complicating the identification of unique health events, especially since each condition_occurrence_ID is distinct.

| Reference ID | Vocabulary Code | Date | Person ID |
|---|---|---|---|
| | | | |

health event

| Condition_occurrence_id | Concept ID | Condition Start Date | Condition End Date | Person ID |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |

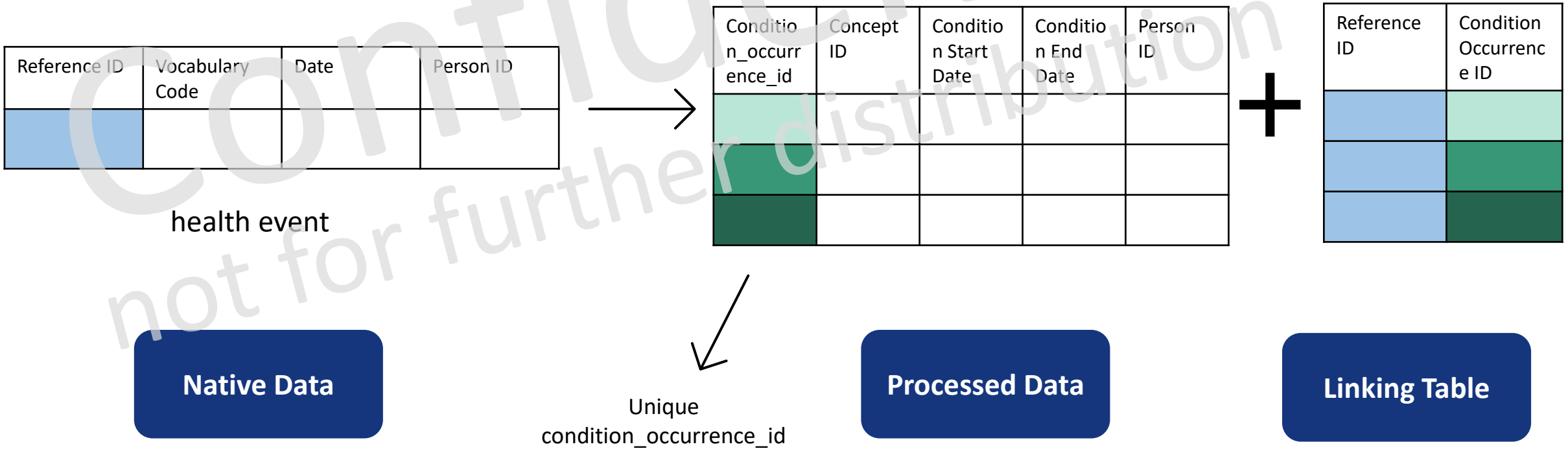
Unique condition_occurrence_id

**Native Data**

**Processed Data**

# Solving Ambiguity of Standard Mappings

- **Solution:** An additional identifier for the source drug product UUID linking to the occurrence_id in OMOP is needed to accurately track affected rows during analysis.



# Solving Ambiguity of Standard Mappings

| Reference ID | Vocabulary Code | Date | Person ID |
|---|---|---|---|
| | | | |

| Condition_occurrence_id | Concept ID | Condition Start Date | Condition End Date | Person ID |
|---|---|---|---|---|
| | | | | |
| | | | | |

+

| Reference ID | Condition Occurrence ID |
|---|---|
| | |
| | |

**Learnings:**

- Establishing unique identifiers for linking occurrence IDs to unique source events is critical to improve the accuracy and clarity of health data interpretations.

# Computational Demands

- Source data can be very large depending on population size and observation period.
- Secure environments can come with limited hardware and/or software limitations.
  - We encountered data tables (e.g., 65 GB) that exceeded available RAM (64 GB).
  - Environments that are not set up for multiuser usage or machine learning workflows.
  - Secure access makes usage more difficult as even remote -> secure environment clipboard access is deactivated.
  - Software cannot always be easily installed.

**Learnings:**

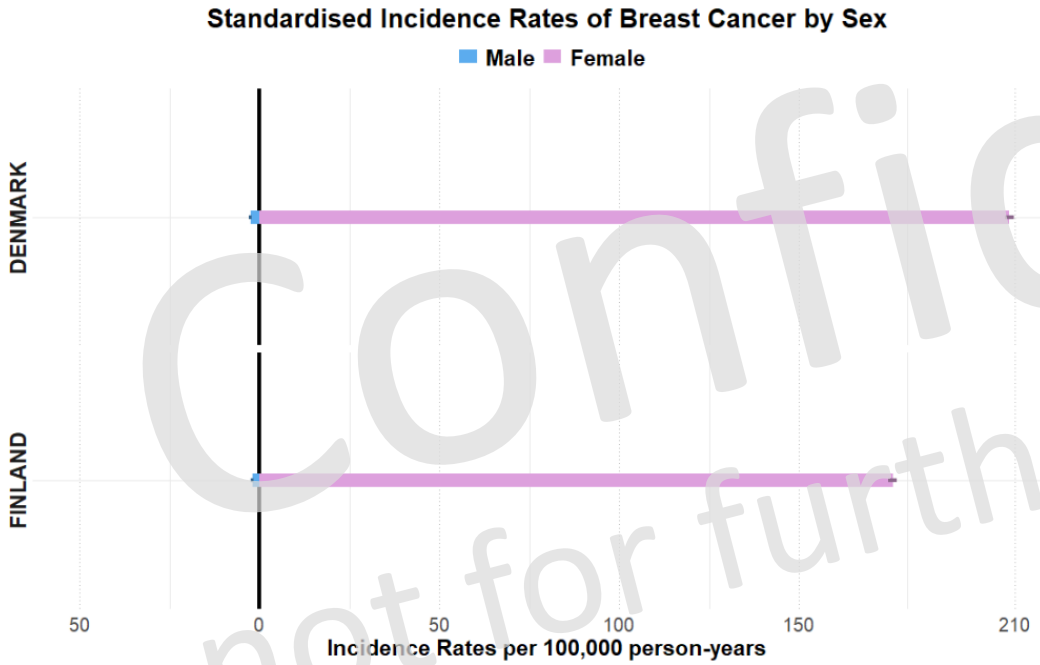
- Employing incremental file processing strategies and leveraging high-performance database systems are crucial for efficient data handling.
- Effective import and export protocols necessary for data flow and operational efficiency
- Adequate access to necessary software and stable internet connectivity is essential to avoid disruptions.

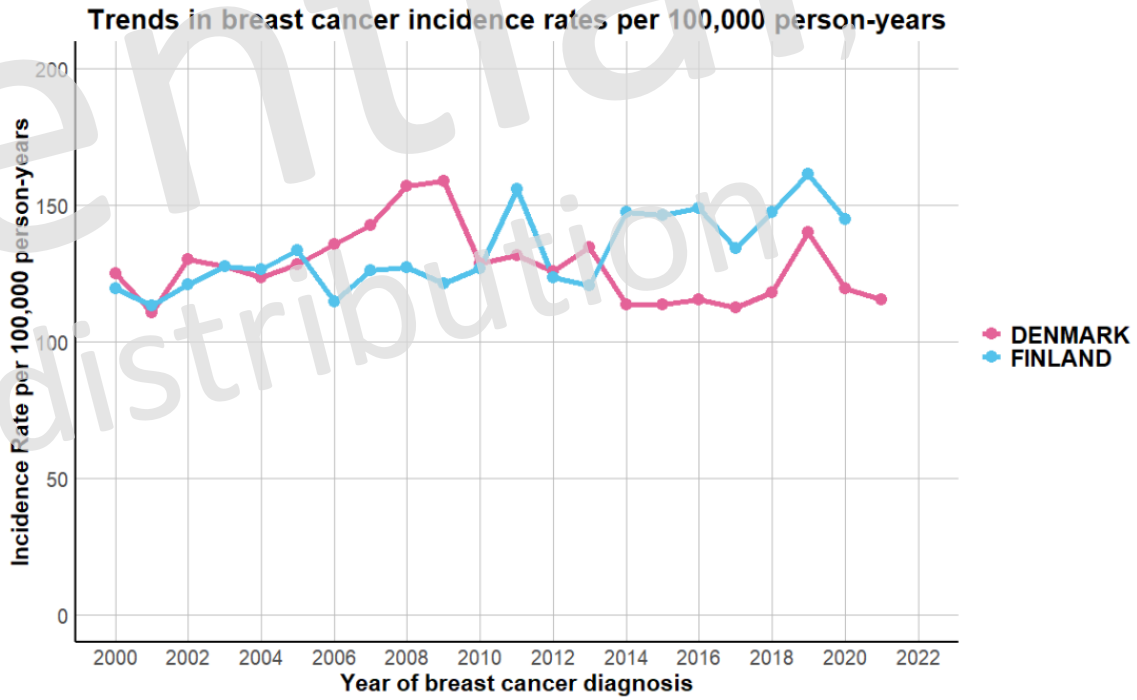
# Use Case 1 Results

# Data Access and Analysis

- **Planned**: single analyst writes analytic script for OMOP-converted data in all countries
- **Workarounds**: due to delays in data access and need to change data custodians to enable relevant data access, each country conducted analyses of data in native (pre-OMOP) format ahead of the analyses in the OMOP-converted data.
- Descriptive statistics of the breast cancer and ALS populations:
  - Incidence rates
  - Demographics, comorbidities, treatment
  - Disease progression, including survival

# Results – Breast Cancer: Incidence Rates (Native)






Results on incidence rates of breast cancer available from Denmark and Finland

# Results – Breast cancer, Women: Demographics

| Characteristic | Denmark | Finland | Portugal |
|---|---|---|---|
| | Total=97550 | Total=87840 | Total=97549 |
| **Age at diagnosis, years** | | | |
| 18 - 34 | 1525 (2%) | 1005 (1%) | 2445 (3%) |
| 35 - 44 | 7240 (7%) | 5205 (6%) | 11989 (12%) |
| 45 - 54 | 18745 (19%) | 18215 (21%) | 22747 (23%) |
| 55 - 64 | 25330 (26%) | 25160 (29%) | 22588 (23%) |
| 65 - 74 | 23490 (24%) | 20315 (23%) | 20133 (21%) |
| >=75 | 21220 (22%) | 17945 (20%) | 17647 (18%) |
| **Employment** | | | |
| Self-employed | NA | 3180 (4%) | NA |
| Upper-level employee | NA | 8860 (10%) | NA |
| Lower-level employee | NA | 18265 (21%) | NA |
| Manual worker | NA | 6535 (7%) | NA |
| Pensioner | NA | 45260 (52%) | NA |
| Other | NA | 5720 (7%) | NA |
| NA | 100% | 0 | 100% |

# Results – Breast Cancer, Women: Comorbidities

| | Denmark | Finland | Portugal |
|---|---|---|---|
| | Total=97550 | Total=87840 | Total=97549 |
| **Comorbidities** | | | |
| Psychiatric disorders | 740 (1%) | 3260 (4%) | 327 (0%) |
| Dementia and Alzheimer's | 1210 (1%) | 2620 (3%) | 640 (1%) |
| Cardiovascular diseases | 7840 (8%) | 22935 (26%) | 8009 (8%) |

# Results – Breast Cancer, Women: Treatment (12 Months Post-diagnosis)

| | Denmark | Finland | Portugal |
|---|---|---|---|
| | Total=97550 | Total=87840 | Total=97549 |
| Radiotherapy | 56755 (58%) | 21485 (24%) | 29019 (30%) |
| Chemotherapy | 36370 (37%) | 21745 (25%) | 57047 (58%) |
| Total or partial lymphadenectomy | 57665 (59%) | 33500 (38%) | 28752 (29%) |
| Mastectomy | 33155 (34%) | 37030 (42%) | 34732 (36%) |
| Breast conservation surgery | 47265 (48%) | 42150 (48%) | 39678 (41%) |
| Hormonal therapy for breast cancer | 36420 (37%) | 58225 (66%) | 29944 (31%) |
| HER directed treatments | 9400 (10%) | 4750 (5%) | 6234 (6%) |

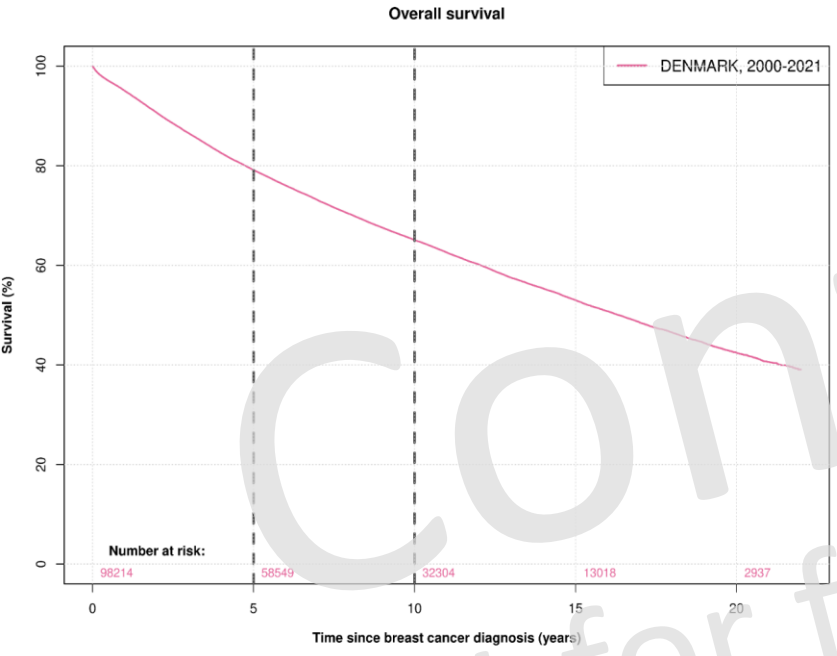
# Results – Breast Cancer, Women: Entire Follow-up

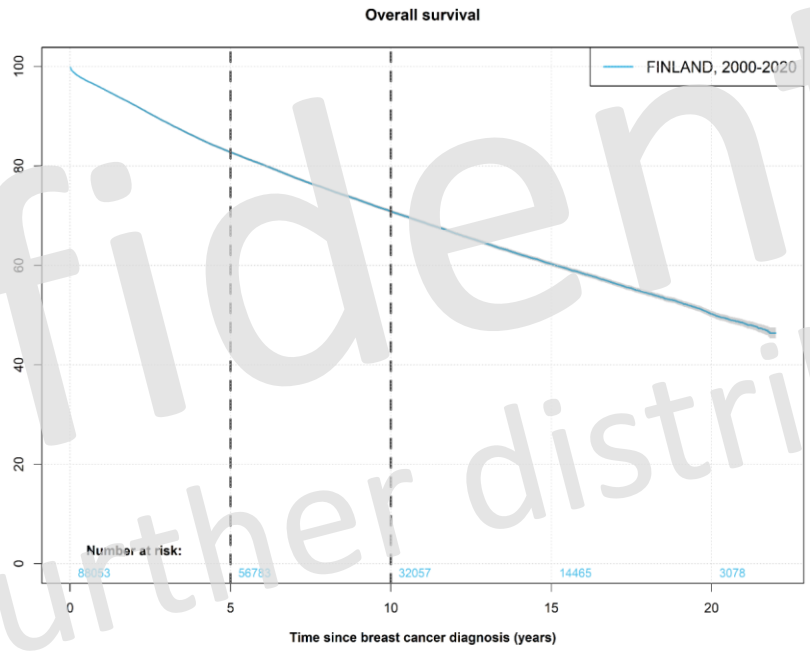
| Characteristic | Denmark | Finland | Portugal |
|---|---|---|---|
| | Total=97550 | Total=87840 | Total=97549 |
| **Signals for disease progression** | | | |
| **Stage at diagnosis** | | | |
| TNM Stage: 1 | 42845 (44%) | NA | 13795 (14%) |
| TNM Stage: 2 | 23320 (24%) | NA | 6062 (6%) |
| TNM Stage: 3 | 2875 (3%) | NA | 401 (0%) |
| TNM Stage: NA | 28515 (29%) | 100% | 77842 (79%) |
| Finland Stage: 1 localised | NA | 40290 (46%) | NA |
| Finland Stage: 2 non-localised | NA | 29625 (34%) | NA |
| Finland Stage: NA | 100% | 17930 (20%) | 100% |
| **Primary tumour size** | | | |
| No growth beyond primary cells | 5795 (6%) | 100 (0%) | 4978 (5%) |
| Tumour growth beyond primary cells | 76330 (78%) | 20150 (95%) | 21549 (22%) |
| Undefined growth | 15425 (16%) | 320 (1%) | 71022 (73%) |
| **Metastases at diagnosis** | | | |
| Metastasis - spread to lymph nodes | 30155 (31%) | 8545 (40%) | 8059 (8%) |
| Metastasis - spread to other organs | 3475 (4%) | 300 (1%) | 496 (1%) |
| **Other signals of progression** | | | |
| New primary malignancy | 8050 (8%) | 5995 (7%) | 9385 (10%) |
| Breast cancer specific mortality | 16535 (17%) | 12000 (14%) | 8699 (16%) |
| All-cause mortality | 32890 (34%) | 25875 (29%) | 24746 (25%) |

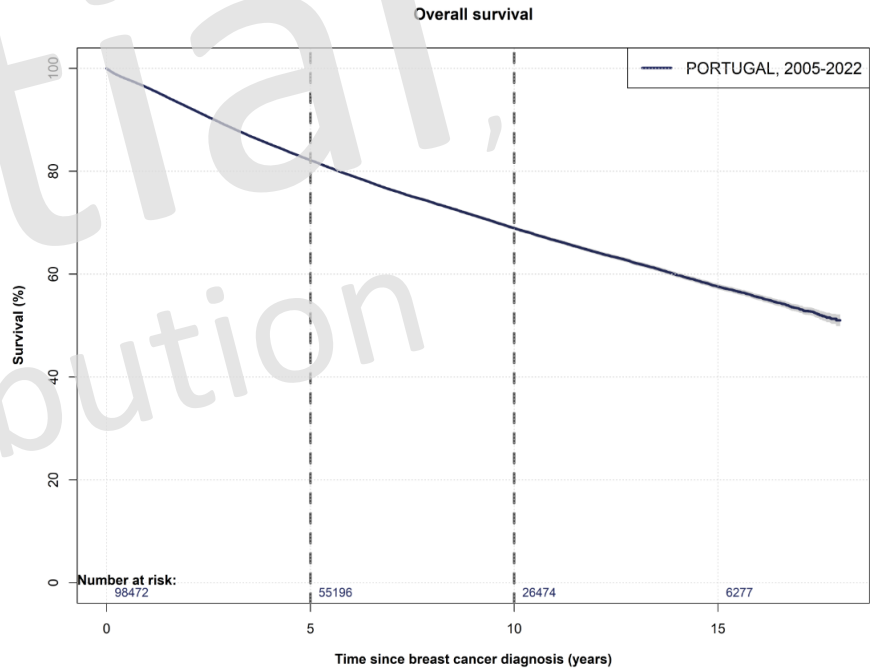
# Results – Breast Cancer, Women: Entire Follow-up

| Characteristic | Denmark | Finland | Portugal |
|---|---|---|---|
| | Total=97550 | Total=87840 | Total=97549 |
| **Signals for disease progression** | | | |
| **Stage at diagnosis** | | | |
| TNM Stage: 1 | 42845 (44%) | NA | 13795 (14%) |
| TNM Stage: 2 | 23320 (24%) | NA | 6062 (6%) |
| TNM Stage: 3 | 2875 (3%) | NA | 401 (0%) |
| TNM Stage: NA | 28515 (29%) | 100% | 77842 (79%) |
| Finland Stage: 1 localised | NA | 40290 (46%) | NA |
| Finland Stage: 2 non-localised | NA | 29625 (34%) | NA |
| Finland Stage: NA | 100% | 17930 (20%) | 100% |
| **Primary tumour size** | | | |
| No growth beyond primary cells | 5795 (6%) | 100 (0%) | 4978 (5%) |
| Tumour growth beyond primary cells | 76330 (78%) | 20150 (95%) | 21549 (22%) |
| Undefined growth | 15425 (16%) | 320 (1%) | 71022 (73%) |
| **Metastases at diagnosis** | | | |
| Metastasis - spread to lymph nodes | 30155 (31%) | 8545 (40%) | 8059 (8%) |
| Metastasis - spread to other organs | 3475 (4%) | 300 (1%) | 496 (1%) |
| **Other signals of progression** | | | |
| New primary malignancy | 8050 (8%) | 5995 (7%) | 9385 (10%) |
| Breast cancer specific mortality | 16535 (17%) | 12000 (14%) | 8699 (16%) |
| All-cause mortality | 32890 (34%) | 25875 (29%) | 24746 (25%) |

# Results – Breast Cancer, All: Overall Survival



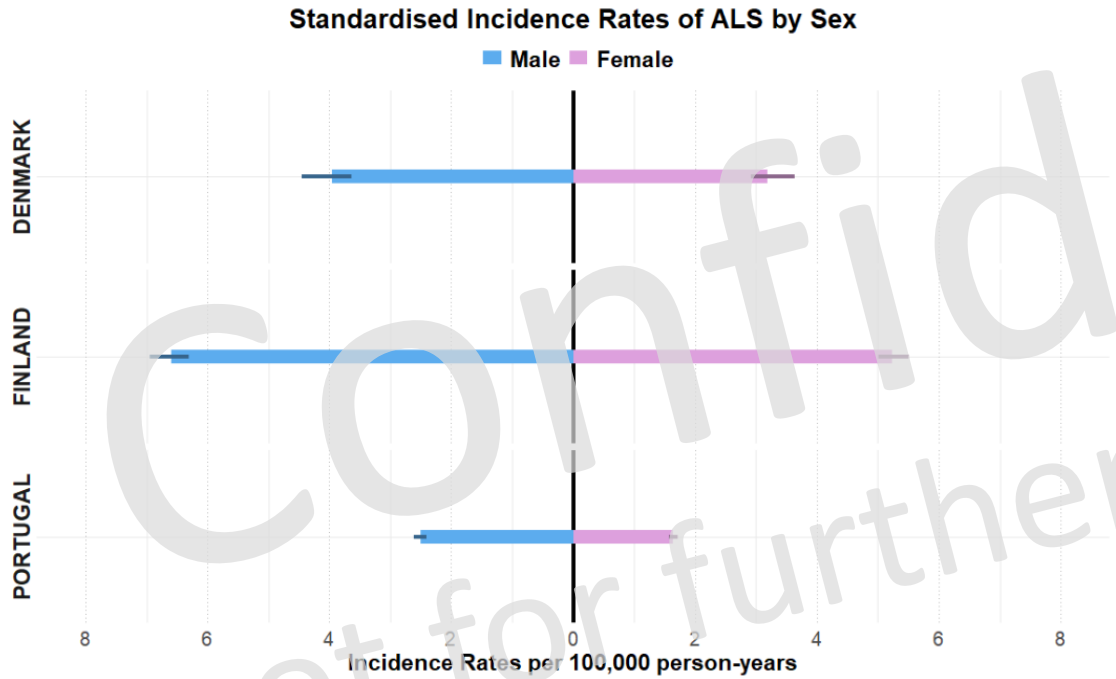



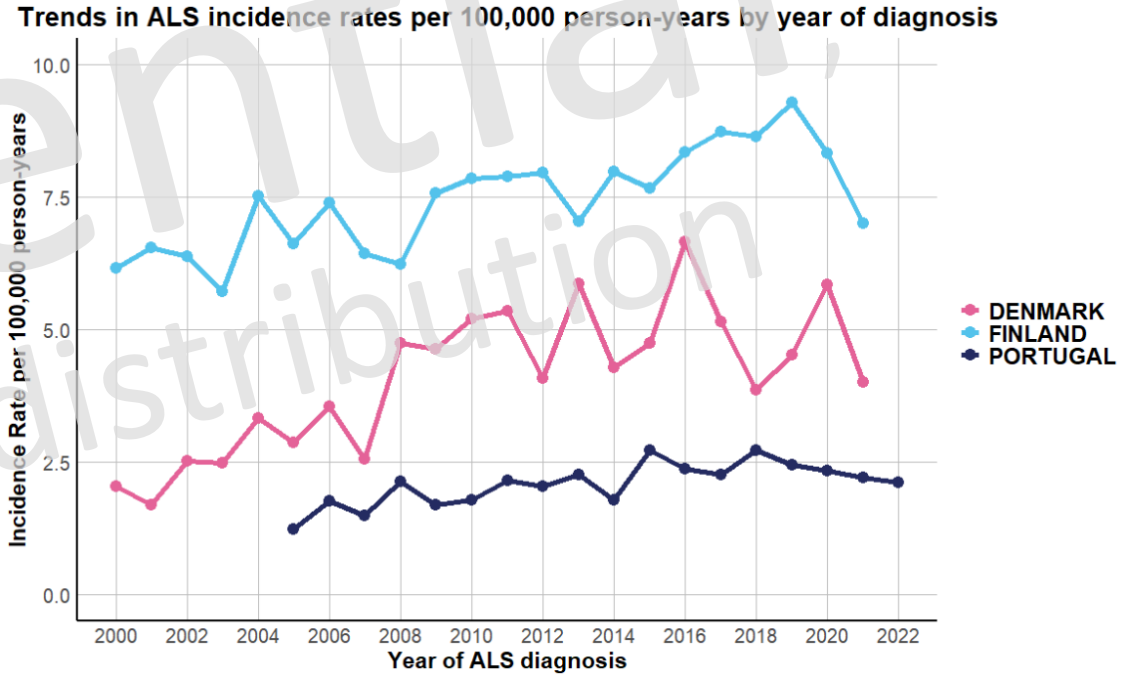



# Results – Breast Cancer, Women/Men: Age

| Age at diagnosis, years | WOMEN | | | MEN | | |
|---|---|---|---|---|---|---|
| | Denmark | Finland | Portugal | Denmark | Finland | Portugal |
| | Total=97550 | Total=87840 | Total=97549 | Total=665 | Total=420 | Total=923 |
| 18 - 34 | 1525 (2%) | 1005 (1%) | 2445 (3%) | 5 (1%) | 5 (1%) | <10 |
| 35 - 44 | 7240 (7%) | 5205 (6%) | 11989 (12%) | 20 (3%) | 10 (2%) | <40 |
| 45 - 54 | 18745 (19%) | 18215 (21%) | 22747 (23%) | 60 (9%) | 50 (13%) | 100 (11%) |
| 55 - 64 | 25330 (26%) | 25160 (29%) | 22588 (23%) | 135 (20%) | 105 (24%) | 199 (22%) |
| 65 - 74 | 23490 (24%) | 20315 (23%) | 20133 (21%) | 215 (33%) | 125 (30%) | 300 (33%) |
| >=75 | 21220 (22%) | 17945 (20%) | 17647 (18%) | 225 (34%) | 130 (31%) | 285 (31%) |

# Results – ALS: Incidence Rate (Native)



# Results – ALS: Demographics

| Characteristic | Denmark | Finland | Portugal |
|---|---|---|---|
| | N=2810 | N=5400 | N=4061 |
| **Sex** | | | |
| Female | 1265 (45%) | 2615 (48%) | 1825 (45%) |
| Male | 1545 (55%) | 2785 (52%) | 2236 (55%) |
| **Age, years** | | | |
| 18 - 34 | 25 (1%) | 60 (1%) | 58 (1%) |
| 35 - 44 | 95 (3%) | 145 (3%) | 154 (4%) |
| 45 - 54 | 285 (10%) | 605 (11%) | 385 (9%) |
| 55 - 64 | 680 (24%) | 1440 (27%) | 907 (22%) |
| 65 - 74 | 1060 (38%) | 1835 (34%) | 1435 (35%) |
| >=75 | 655 (23%) | 1315 (24%) | 1122 (28%) |
| Median (quartiles) | 68 (60-75) | 68 (59-75) | 69 (61-76) |
| **Employment** | | | |
| Self-employed | NA | 160 (3%) | NA |
| Upper-level employee | NA | 305 (6%) | NA |
| Lower-level employee | NA | 445 (8%) | NA |
| Manual worker | NA | 420 (8%) | NA |
| Pensioner | NA | 3725 (69%) | NA |
| Other | NA | 345 (6%) | NA |
| NA | 100% | 0 | 100% |

# Results – ALS: Clinical Characteristics

| Characteristic | Denmark | Finland | Portugal |
|---|---|---|---|
| | N=2810 | N=5400 | N=4061 |
| **Comorbidities** | | | |
| Psychiatric disorders | 15 (1%) | 175 (3%) | 24 (1%) |
| Dementia and Alzheimer's | 115 (4%) | 325 (6%) | 249 (6%) |
| Cardiovascular diseases | 455 (16%) | 1920 (36%) | 1290 (32%) |
| Respiratory diseases | 205 (7%) | 655 (12%) | 363 (9%) |
| Other muscular atrophy types | 20 (1%) | NA | 17 (0%) |
| **Characteristics in 5 years before ALS diagnosis** | | | |
| Fasciculations, cramps, muscle twitching | 45 (2%) | 225 (4%) | 16 (0%) |
| Dysarthris | 95 (3%) | 700 (13%) | 49 (1%) |
| Fronto-temporal dementia | 10 (0%) | 40 (1%) | 13 (0%) |
| Mood disorders (incl. depression, bipolar) | 85 (3%) | 300 (6%) | 169 (4%) |
| Other mild motor impairment | 505 (18%) | 1350 (25%) | 33 (1%) |
| Anxiety | 20 (1%) | 95 (2%) | 43 (1%) |

# Results – ALS: Treatment (12 Months Post-diagnosis)

| | Denmark | Finland | Portugal |
|---|---|---|---|
| | N=2810 | N=5400 | N=4061 |
| Intensive care unit service admission with mechanical ventilation | 255 (9%) | 75 (1%) | 76 (2%) |
| Non-invasive ventilation | 560 (20%) | NA | 880 (22%) |
| Riluzole | 1635 (58%) | 1600 (30%) | NA |
| Antiepileptics [incl. gabapentinoids, levetiracetam] | 105 (4%) | 340 (6%) | 226 (12%) |
| Skeletal muscle relaxants | 485 (17%) | 480 (9%) | 274 (14%) |
| Benzodiazepines and related drugs | 1180 (42%) | 1440 (27%) | 634 (33%) |
| Antidepressants/mood disorder medication | 1200 (43%) | 1820 (34%) | 1018 (52%) |

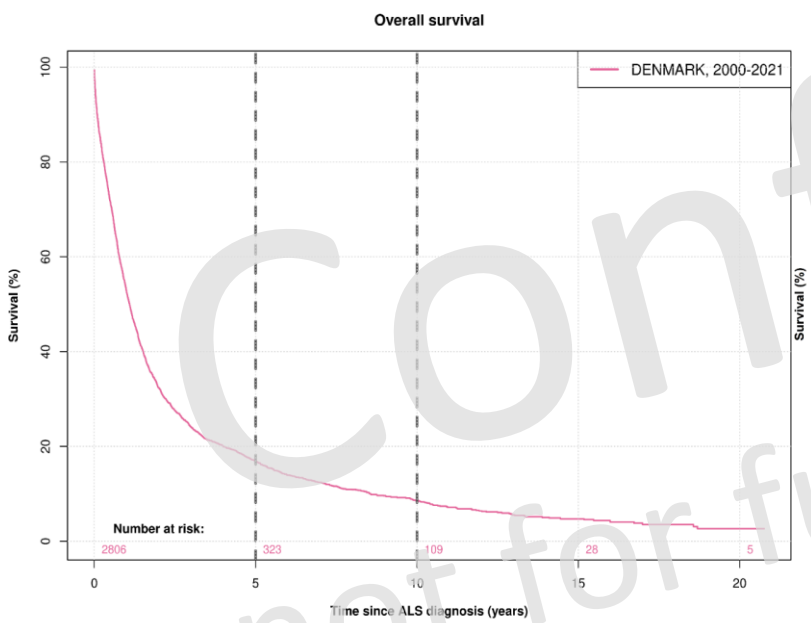
# Results – ALS: Treatment (12 Months Post-diagnosis)

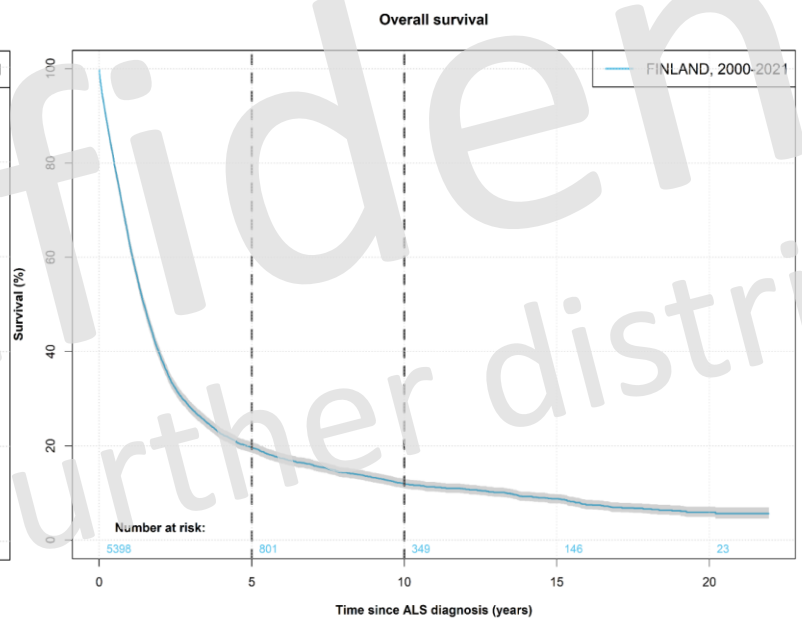
| | Denmark | Finland | Portugal |
|---|---|---|---|
| | N=2810 | N=5400 | N=4061 |
| | | | |
| Intensive care unit service admission with mechanical ventilation | 255 (9%) | 75 (1%) | 76 (2%) |
| Non-invasive ventilation | 560 (20%) | NA | 880 (22%) |
| Riluzole | 1635 (58%) | 1600 (30%) | NA |
| Antiepileptics [incl. gabapentinoids, levetiracetam] | 105 (4%) | 340 (6%) | 226 (12%) |
| Skeletal muscle relaxants | 485 (17%) | 480 (9%) | 274 (14%) |
| Benzodiazepines and related drugs | 1180 (42%) | 1440 (27%) | 634 (33%) |
| Antidepressants/mood disorder medication | 1200 (43%) | 1820 (34%) | 1018 (52%) |

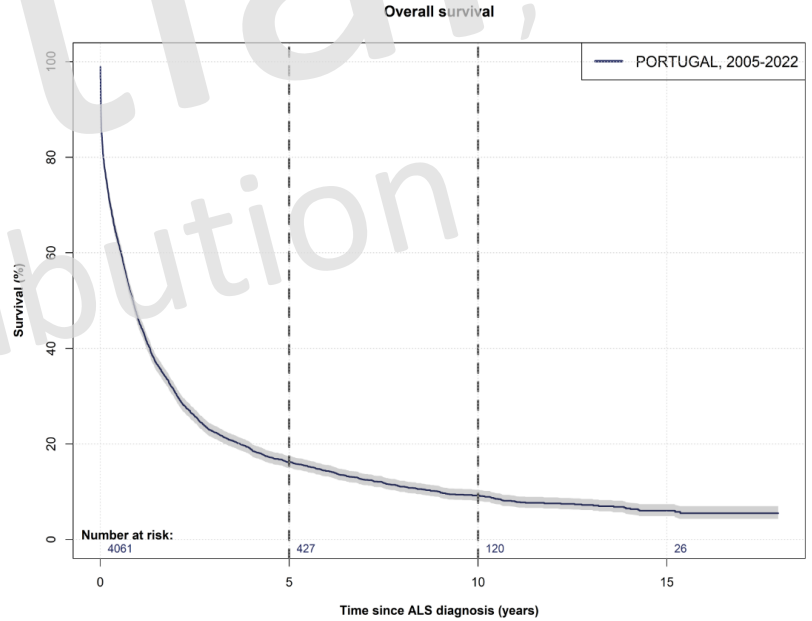
# Results – ALS: Prognosis (Entire Follow-up)

| | Denmark | Finland | Portugal |
|---|---|---|---|
| | N=2810 | N=5400 | N=4061 |
| Respiratory infection [pneumonia] | 750 (27%) | 1470 (27%) | 1232 (30%) |
| Intensive care unit service admission with mechanical ventilation | 440 (16%) | 150 (3%) | 125 (3%) |
| Non-invasive ventilation | 805 (29%) | NA | 1229 (30%) |
| PEGs (percutaneous gastrostomy) | 380 (13%) | 90 (2%) | 547 (13%) |
| Cachexia (wasting syndrome) | 80 (3%) | 10 (0%) | 162 (4%) |
| Acute respiratory failure | 180 (6%) | 145 (3%) | 682 (17%) |
| ALS specific mortality | 2010 (72%) | 3810 (71%) | 1397 (65%) |
| All-cause mortality | 2350 (84%) | 4420 (82%) | 3368 (83%) |

# Results – ALS: Overall Survival



# Results – ALS: Native vs OMOP Analyses, Selected Characteristics

| | Denmark | | Portugal | |
|---|---|---|---|---|
| | NATIVE | OMOP | NATIVE | OMOP |
| | N=2810 | N=2991 | N=4061 | N=4071 |
| **Sex** | | | | |
| Female | 1265 (45%) | 1340 (45%) | 1825 (45%) | 1827 (45%) |
| Male | 1545 (55%) | 1651 (55%) | 2236 (55%) | 2244 (55%) |
| **Age, years** | | | | |
| 18 - 34 | 25 (1%) | 34 (1%) | 58 (1%) | 58 (1%) |
| 35 - 44 | 95 (3%) | 102 (3%) | 154 (4%) | 154 (4%) |
| 45 - 54 | 285 (10%) | 309 (10%) | 385 (9%) | 387 (10%) |
| 55 - 64 | 680 (24%) | 713 (24%) | 907 (22%) | 908 (22%) |
| 65 - 74 | 1060 (38%) | 1131 (38%) | 1435 (35%) | 1438 (35%) |
| >=75 | 655 (23%) | 702 (23%) | 1122 (28%) | 1126 (28%) |
| Median (quartiles) | 68 (60-75) | 68 (59-74) | 69 (61-76) | 68 (60-75) |
| **Comorbidities** | | | | |
| Psychiatric disorders | 15 (1%) | 26 (1%) | 24 (1%) | 24 (1%) |
| Dementia and Alzheimer's disease | 115 (4%) | 157 (5%) | 249 (6%) | 249 (6%) |
| Cardiovascular diseases | 455 (16%) | 456 (15%) | 1290 (32%) | 1292 (32%) |
| Respiratory diseases | 205 (7%) | 203 (7%) | 363 (9%) | 364 (9%) |
| Other muscular atrophy types | 20 (1%) | 53 (2%) | 17 (0%) | 17 (0%) |

# Main Findings/Lessons Learned

- Data access delays is an important **barrier** to timely RWE generation
- Correctness of data delivery cannot be automatically assumed and needs to be a standard **quality control** item for researcher before data use
- **Heterogeneity** of the results across the country-specific patient populations stems from the true differences, and artefacts of data flow, completeness of recording and reimbursement practices, and data provenance
- The heterogeneity, including type, validity, and completeness should be used for assessing a given data source **fitness for purpose** in relation to specific research/regulatory questions

# Conclusions

- Use of real-world-data in regulatory preauthorisation/evaluation decisions requires expert knowledge to assess the impact of data limitations
- Harmonisation of data to the same structure does not remove the underlying heterogeneity of data quality, flow and completeness
- Results for breast cancer and ALS are generated for illustration and should not be taken as indicative of disease epidemiology, however, they are broadly consistent with the published evidence

# Thank you!

Prof. Vera Ehrenstein
**Aarhus University**
Department of Clinical Epidemiology
E-Mail: ve@clin.au.dk

PhD Candidate Mohamed Aborageh
**Fraunhofer SCAI**
Department of AI & Data Science
E-Mail: mohamed.aborageh@scai.fraunhofer.de



**Real4Reg administration contacts:**

**Real4Reg management**
Federal Institute for Drugs and Medical Devices
Division 5 Research

real4reg@bfarm.de

**Real4Reg communication**
INFARMED, IP.
Information and Strategic Planning Division

real4reg@infarmed.pt

**www.real4reg.eu**

Funded by the European Union
Horizon Europe
Project No 101095353