Pharmacoepidemiology and Drug Safety

WILEY

ORIGINAL ARTICLE

EpiCore—A Common Data Model for Pharmacoepidemiological Studies in Denmark, Norway, and Sweden

Peter Bjødstrup Jensen[1] | Jacob H. Andersen[1] | Martin Thomsen Ernst[1] | Morten Olesen[1] | Øystein Karlstad[2] | Kari Furu[2] | Julia Eriksson[3] | Karin Gembert[3] | Anton Pottegård[1]

[1]Clinical Pharmacology, Pharmacy, and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark | [2]Department of Chronic Diseases, Norwegian Institute of Public Health, Oslo, Norway | [3]Center for Pharmacoepidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Correspondence: Anton Pottegård (apottegaard@health.sdu.dk)

Received: 4 March 2025 | **Revised:** 23 September 2025 | **Accepted:** 8 October 2025

Funding: The authors received no specific funding for this work.

Keywords: CDM | pharmacoepidemiology | research infrastructure

ABSTRACT

Purpose: The use of common data models (CDMs) is increasing; however, the complexity of many CDM frameworks constitutes a barrier for their use. For many local and collaborative use cases, simpler CDMs can suffice. Here, we propose the EpiCore CDM, a simple CDM framework for use in Scandinavian pharmacoepidemiological studies.

Methods: The EpiCore CDM was developed based on a set of guiding principles. It should (i) accommodate the most common elements of typical data sources in the field and region, (ii) be accessible to users without needing advanced technical expertise or database infrastructure, (iii) prioritize structural and syntactic harmonization of data and defer clinical concept mapping to the analytical phase, (iv) be usable in both collaborative and single site settings, and (v) include support for quality control procedures.

Results: The EpiCore CDM comprises two mandatory administrative tables (person and observation), six optional event tables (diagnosis, procedure, encounter, drug, primcare, and cancer) and three optional lookup tables (drug_info, organisation_info, and prescriber_info). Each table, along with its columns and constraints is specified according to an EpiCore CDM specification template. This provides easy documentation and integrates with an R-package called EpiCoreAssistant, which provides quality control tools for testing the compliance of a CDM instance with the EpiCore specification. In the event that a project requires customization of the CDM, this is easily implemented in the template and testing. A step-by-step description is presented, demonstrating the steps involved in a typical CDM-based collaborative pharmacoepidemiological study using the EpiCore CDM.

Conclusions: We present the EpiCore CDM, a specification template and an R package that offers a simple framework for improved workflows, standardizations and collaboration, focused on Scandinavian pharmacoepidemiological studies and with relevance for a broad palette of register-based health care researchers.

Summary

- The EpiCore CDM provides a streamlined framework tailored for Scandinavian pharmacoepidemiological studies, with lower complexity compared to existing CDMs.
- The EpiCore CDM is designed for users without advanced technical expertise and accommodates common data elements while allowing customization for specific research needs.
- The EpiCore CDM prioritizes syntactic and structural data harmonization, deferring clinical concept mapping to the analytical phase.
- The EpiCore CDM is accompanied by the EpiCoreAssistant R package, to support implementation and compliance with the CDM specification and to enhance data quality control.

1 | Introduction

Nationwide health registries can provide high validity data with long follow-up for entire populations [1]. Nevertheless, exposures or outcomes may be so rare that data from a single country do not provide sufficient statistical precision [2–4]. In other situations, one may want to test a result by independently replicating it on data from another data source [5]. Finally, some studies on, for example, disease patterns or drug utilization benefit from being implemented across countries to document variation [6, 7]. A common solution in all these cases, is to run the same analysis in multiple countries and combine or compare the results.

Two main strategies are commonly used to accomplish this [8]. The "common protocol" strategy lets each site independently implement the same analysis based on an agreed-upon study protocol. The "common program" strategy instead lets each site run the same analytical code. The latter strategy requires a common data model (CDM), an agreed representation of data that specifies the naming of tables and columns, their content constraints, and the relationships between them.

CDM-based infrastructure provides several additional benefits in the local setting. For example, it encourages standardization of data cleaning processes, commonly referred to as extract, transform, load (ETL), sharing and streamlining of code and enables better code review.

The choice of CDM depends on the nature and scale of the project or collaboration, as well as factors such as technical expertise, available infrastructure, CDM experience, existing workflows and rules for data access. General CDM frameworks [9] such as the Observational Medical Outcomes Partnership (OMOP) [10] and Sentinel [11], offer comprehensive data models for all kinds of health data and a rich ecosystem of tools, but require a substantial investment of resources and expertise to implement and maintain. Study-specific CDM frameworks [9], such as ConcepTION [12] are tailored to specific research domains (e.g., medicines in pregnancy) and are typically populated on a project-by-project basis. While general CDMs like OMOP are well suited for large-scale international collaborations they can be unnecessarily complex for smaller domain-specific and project-based collaborations.

In this paper, we present the Epidemiological Core CDM (EpiCore), a pragmatic CDM designed for register-based pharmacoepidemiological research in Scandinavia. EpiCore formalizes and extends data models originally developed for local studies and data management in Denmark, later refined and applied in published and ongoing multi-country collaborations involving Denmark, Norway, and Sweden [3, 5, 6]. The framework includes the EpiCoreAssistant [13] R package with quality control (QC) functionality, including an excel template with the specification of the EpiCore CDM, where modifications are easily made. We aim for EpiCore to provide an accessible framework that reduces the need for repetitive CDM design, supports the benefits outlined above, and facilitates streamlined, high-quality collaborations. In this paper, we (i) outline the considerations underpinning EpiCore's design, (ii) describe its structure, (iii) introduce the accompanying R package, (iv) illustrate a collaborative CDM-based workflow, and (v) discuss perspectives and future developments.

2 | Considerations (i)

In this section, we present the specific considerations that were used to guide the process of designing and presenting the EpiCore CDM framework.

2.1 | One Size Fits Most

We deliberately prioritized simplicity over comprehensiveness. Accordingly, the EpiCore CDM consists of a core set of tables intended to accommodate key information, from the most relevant data sources for register-based pharmacoepidemiological research in the Scandinavian region. This focus is justified by the contextual similarities in these countries and by our experience that most local and collaborative studies are fully supported by this core set.

To accommodate future development and project-specific needs, the framework allows flexible extension of the model with new tables for currently unsupported data domains, such as, for example, perinatal data. The flexibility extends to things like adding columns and codebook values to existing tables in those rare use cases where this is required.

2.2 | Simple Technical Requirements

We wanted the EpiCore CDM to offer an accessible, pragmatic framework, with technical and administrative requirements that align with those of a typical researcher or research group in our field and region. In our experience, that means a file-based framework with few dependencies, that can easily be set up both locally and within protected analytical environments where the researchers often have limited administrative privileges. Although database systems (as required e.g., by OMOP) generally offer better performance over files, many researchers

still have little or no experience using them, let alone installing and administrating them. In Denmark databases are not even available on the national analysis environments.

2.3 | No Concept Mapping

We believe that a simple CDM should focus on the technical aspects of data harmonization. That is, an agreed representation of tables, columns and their relational constraints (structural harmonization) and codebooks, formats and datatypes (syntactic harmonization). EpiCore does not require mapping of native vocabularies to a single standard vocabulary of clinical concepts. Instead, it stores the original codes and coding systems—a design like the approach taken by ConcePTION CDM.

This shifts the standardization and documentation of meaning (semantic harmonization) to the analytical level rather than the CDM level. This allows (and requires) each site to specify study variables in terms of the local codes and contexts they know best, typically based on a draft in the project protocol. This approach can accommodate the inherent heterogeneity between sites and avoids the false sense of comparability that a shared vocabulary can sometimes create.

Naturally, thorough documentation of this process is required, but we do not consider it part of the EpiCore CDM itself. Instead, resources such as the HARPER protocol are recommended to guide this documentation at the project level [14].

2.4 | Supporting Internal Workflows

The research groups behind this paper have found great value in adopting CDM-based infrastructures in our internal workflows. These simplify complex source data into clean, research-ready formats accessible to non-experts, promote code reuse, and encourage well-documented, tested functions for common tasks. They also support internal code review and quality control. We envision EpiCore CDM as a practical choice for local research infrastructures, enabling easier scaling from local studies to multinational collaborations and external validations.

2.5 | Supporting Quality Control

EpiCore should include standardized quality control tools to verify that a CDM instance resulting from ETL complies with the specification. This allows problems to be caught early—before shared analysis scripts fail or produce incorrect results, saving time and avoiding frustration. The QC tools do not dictate the ETL process itself; responsibility for implementing, testing and versioning the ETL remains with the site.

3 | CDM Design (ii)

In this section we describe the elements of the EpiCore CDM and their relations, which are graphically illustrated in Table 1 and Figure 1. For the full specification, including a number of details and specific scenarios, please refer to the CDM specification document provided as Supporting Information S1

3.1 | Instance

We use the word "instance" to describe any realization of the EpiCore CDM. That is, an instance is the product of running an ETL process, for example for a specific project. As the EpiCore CDM is file-based, an instance consists of one file per CDM table, stored together in a single folder.

In principle, any file format may be used if the content complies with the specification. However, the current EpiCoreAssistant package supports .csv and .fst (from the R fst package [15]). The .fst format offers substantially faster read and write performance than .csv, which can greatly improve downstream analyses.

3.2 | Tables

We distinguish between three types of tables in the EpiCore CDM: (i) administrative tables, (ii) event tables, and (iii) look-up tables.

3.2.1 | Administrative Tables

The EpiCore CDM consists of two required administrative tables, the person table and the observation table. Records in the persons table capture the key demographic data of the CDM population and assign each person a unique identifier, the id. The observation table defines the periods of time a person has been eligible for data capture by the data sources that supply data for the CDM. In the Scandinavian countries, data capture is mandatory and nationwide across all source registers, so there is typically no need to register observation periods by source. It is enough to capture the non-overlapping periods of residency for each person. Observation periods are truncated by death, emigration or data censoring. Any person must have at least one observation period.

3.2.2 | Event Tables

Event tables capture discrete facts about persons in the CDM such as a diagnosis code, a prescribed drug or other medical events. Each event table contains additional columns of attributes that provide context to the fact, such as the date, the health care setting, type of fact, or classification. EpiCore is designed to accommodate clinical data from the following health care settings: (i) Primary care, comprising, for example, general practice, is captured in the CDM table primcare, (ii) secondary healthcare, comprising in and outpatient encounters, is captured in the CDM tables diagnosis, procedure and encounter, and (iii) prescription fills are captured in the CDM table drug. Finally, (iv) the cancer table is specifically intended for records of cancers coming from dedicated cancer registries. All event tables are in principle optional.

TABLE 1 | Description of tables, columns, and relationships of the EpiCore CDM. See Supporting Information S1 for all details. Categorical columns are underlined. See Supporting Information S1 for codebook of allowed key-label pairs.

Type	Table	Description	Columns and relationships between columns
Administrative (Required)	person	One record for each person in the CDM. The id from person acts as foreign key in all other data tables of the CDM	**PK:** id **FK:** **Other:** birthdate, sex, deathdate
	observation	One record for each non-overlapping period of time a person is eligible for data capture by all the register sources that contribute data. In DK, NO and SE, this corresponds to periods of national recidency. Each person must have at least one record in observation	**PK:** id, obs_start **FK:** id **Other:** obs_end, end_reason
Event (Optional)	encounter	One record for each health care encounter associated to a person in the CDM. Encounters are characterized by a setting (e.g. hospital) and a type (e.g. inpatient). Encounters can span one or multiple days. Org_id allows linking additional attributes of the health care provider to the encounter	**PK:** enc_id **FK:** id org_id **Other:** enc start, enc end, encsetting, enctype
	diagnosis	One record for each diagnosis associated to a person. Multiple diagnosis records may be linked to the same encounter id in the encounter table (if that table is included). Org_id allows direct linking of additional attributes of the health care provider to the diagnosis	**PK:** id, diag, dxdate, enc_id **FK:** id enc_id org_id **Other:** dxtype, classification
	procedure	One record for each procedure associated to a person. Multiple procedure records may be linked to the same encounter id in the encounter table (if that table is included). Org_id allows direct linking of additional attributes of the health care provider to the procedure	**PK:** id, proc, pxdate, enc_id **FK:** id enc_id org_id **Other:** pxtype, classification
	drug	One record for each drug event. Most commonly a record corresponds to a drug redeemed at a pharmacy, but can be any relevant prescription, dispensing or other exposure to a drug. presc_id allows linking addition attributes of the prescriber. drug_id (vnr in the Scandinavian setting) allows linking additional attributes of the specific drug	**PK:** No primary key **FK:** id presc_id drug_id **Other:** atc, rxdate, ddd, amout, drug_setting, ddd, amount, indication, indi_classification
	primcare	One record for each primary care diagnosis or service associated to a person. The definition of primary care may differ in different countries and may e.g. include services provided by private specialists (Denmark)	**PK:** id, pmc_date, pmc_code **FK:** id org_id **Other:** classification, nrdups
	cancer	One record for each new registered cancer associated to a person. Intended for cancer registrations coming from dedicated (and typically validated) cancer registers. It is possible for the same information to be both in the diagnosis table and cancer table	**PK:** id, ccode, cdate, morpho **FK:** id **Other:** classification
Lookup (Optional)	drug_info	One record for each distinct drug id used in the drugs CDM table. Allows joining of elevant attributes to drug records.	**PK:** drug_id **FK:** **Other:** strength, dose form, package_amount, adm_route, ..
	organisation_info	One record for each unique org_id referenced in the CDM event tables. An org_id can e.g. specify a hospital department or GP clinic. Allows joining additional attributes of the org_id to those event records.	**PK:** org_id **FK:** **Other:** org_name, org_setting, org_specialty, org_locality, ...
	prescriber_info	One record for each unique presc_id referenced by a record in the drug table. A presc_id can e.g. specify a unique prescriber person or a clinic or department responsible for the prescription	**PK:** presc_id **FK:** **Other:** presc_type, locality, presc_specialty ...,

Note: Relationships between columns in different tables are represented as lines with crow's foot notation: ●—○<: one to zero or more. ●—<: one to one or more. PK: Primary key. No duplicate rows allowed across identical value(s) of primary key column(s). FK: Foreign key. A foreign key column must reference a primary key in another table.

3.2.3 | Look-Up Tables

Look-up tables do not contain individual level data. They provide a way to link facts or attributes in administrative or event tables to additional (less frequently used) attributes. This reduces complexity and table size of the administrative and event tables. The drug_info table links distinct drug_id's (Nordic article numbers ["varenummer"] in the Scandinavian setting) in the drug table, to additional attributes, such as strength, administration route, name, etc. The organisation_info look-up table, links, for example, information about a department or provider (e.g., their medical specialty) to a record in an event table. Finally, the prescriber_info lookup table links a prescriber in the drug table to additional attributes about the prescriber.

3.3 | Constraints

Constraints act as assumptions about data, that users can expect to hold. In addition to the constraint given by the table and column names, the EpiCore CDM specification template, imposes additional constraints at the (i) table, (ii) column, and (iii) relation level.

3.3.1 | Table Level Constraints

The main table level constraint is the primary key, which describes the set of columns in the table, across which every record must be unique. In the persons table, this is the id alone, but in

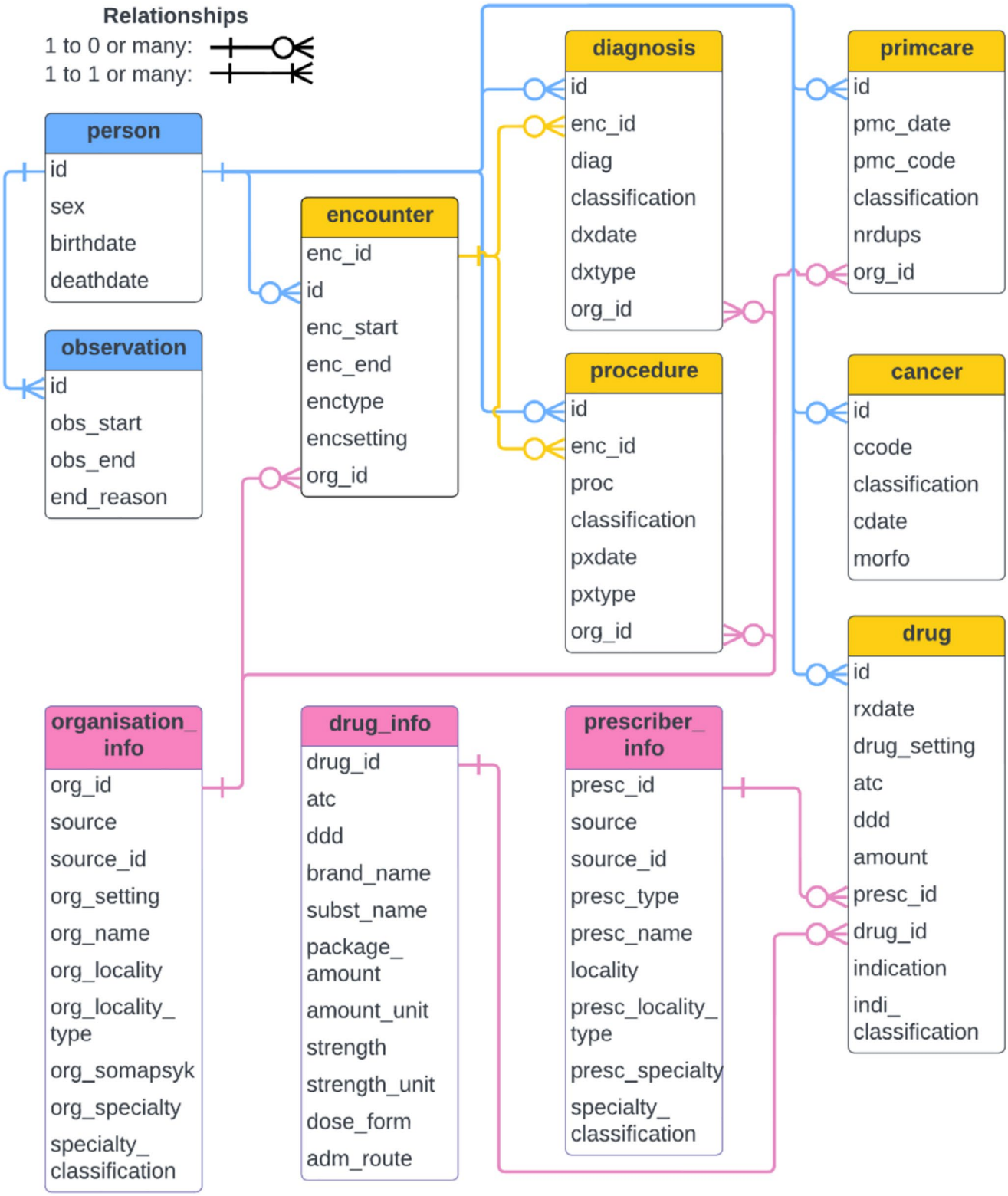

FIGURE 1 | Relationship diagram for tables and columns in the EpiCore CDM. Blue tables: Administrative mandatory tables. Yellow tables: Optional event tables. Purple tables: Optional lookup tables. Relationships between columns in different tables are represented as lines with crow's foot notation: ⟝ one to zero or more. ⟞ one to one or more. For example, any record in person must have at least one record in observation, but not necessarily any records in encounter.

other tables it may involve multiple columns. In the diagnoses table for example, the same person cannot have two records on the same date, with the same diagnoses code linked to the same encounter_id.

3.3.2 | Column Level Constraints

The values of any column in any table must comply with one of the following types: character (e.g., diagnosis codes), integer

(e.g., id columns), numeric (numerical columns not compatible with integer), date (e.g., diagnosis date), datetime (e.g., encounter start), or categorical (integer column with a finite set of allowed values that are further specified in a codebook). A column can be designated as required, meaning it must exist in the table, and it can also be characterized by whether missing values are allowed for individual records. For more details, see Supporting Information S1.

3.3.3 | Relational Constraints

Relationships describe the constraints imposed by links between columns in different tables, known as foreign key constraints. For example, all event tables have an id column with a many-to-one foreign key relationship to the id column of the person table. This constraint dictates, that any record in these tables must reference an id in the persons table, and that any id in the person table may have zero or more records in the linked table. A similar many-to-one relationship exists between records in the event tables diagnosis and procedure with the event table encounter and between columns linked to look-up tables, like the drug_id column in the drug table and the drug_info look-up table.

4 | R Package and Quality Control (iii)

It is essential to have a standardized way to assert that the EpiCore CDM instance resulting from an ETL procedure, conforms with the specification. We have built an R package "EpiCoreAssistant [13]" that includes a quality control module to accomplish this. It uses the CDM specification document (Supporting Information S1) as the standardized template for the test and outputs QC results in an easy-to-read html report.

Tests for conformance are applied stepwise at different levels: (level 1) Compares available tables in the CDM instance to expected tables in the specification and tests if they are readable, (level 2) Compliance tests within each table for constraints such as data types, categorical values, formats, missingness and primary keys, (level 3) Tests between tables of relational constraints. The EpiCoreAssistant is described and documented in the R package documentation.

5 | CDM Based Workflow (iv)

In this section we describe what an EpiCore-based workflow in a collaborative multisite project might look like. In Supporting Information S2, we provide a similar description of a local (single site) EpiCore-based workflow. Once the current and long-term aim and scope of the collaboration are agreed, the project follows these steps.

5.1 | Data Source Clarification

Which data elements from available data sources in each site will be required to answer study questions.

5.2 | CDM Specification and Dissemination

Starting from the EpiCore CDM specification, determine if modifications are needed to accommodate the required data. This may include adding new tables for additional data domains (e.g., laboratory results or perinatal data) or leaving out tables not relevant to the project (e.g., the procedure table). Column-specific modifications are also possible such as adding additional allowed categorical values. The agreed model is documented in the CDM specification template and shared within the project group.

5.3 | Data Transformation

Each site develops an ETL package, to transform source data into an instance of the agreed CDM representation. The ETL program should be versioned, with a clear indication of the EpiCore CDM version that is the target of the ETL. The instance is tested for compliance using the EpiCoreAssistant package. Any incompliance is resolved by adjusting the ETL program, until the compliance test is passed.

5.4 | Analysis Program

Based on the protocol, the analysis scripts are developed and versioned by one of the study sites and then shared among sites. Each site contributes site-targeted study variable specifications in a structured format that acts as inputs to the analysis scripts, which are otherwise ideally identical for all sites. Each site runs the analysis and shares aggregate results and estimates with the coordinating site.

6 | Future Developments (v)

We consider the EpiCore CDM proposed in this paper a stable release; however we recognize that some data domains are not well accommodated. This includes, for example, perinatal data, laboratory results, causes of death and sociodemographic information. We plan to release separate table templates for these domains, to act as extensions to the original EpiCore data model. We will draw inspiration from existing CDM work in the Scandinavian setting such as, for example, the NorPreSS CDM [16] for perinatal data.

The key feature of the EpiCoreAssistant R package is the compliance checking. We hope to improve this beyond testing conformance (structure and syntax), to include a small selection of tests for completeness and plausibility [17] It is also planned to add support for the .parquet file format [18], as this format has great potential for optimizing performance and cross-platform analytics in a file-based CDM setup like EpiCore.

7 | Discussion (vi)

We have presented EpiCore, a pharmacoepidemiological CDM designed for the types of projects that constitute a large part of

both collaborative and internal pharmacoepidemiological research in the Scandinavian countries. The design is intentionally simple, which we view as both feasible and appropriate given the many similarities across Scandinavian healthcare systems in terms of national coverage, underlying structures, and data source availability [1].

To facilitate adoption, we have developed an accompanying R package (EpiCoreAssistant) together with a standardized approach for documenting and adapting the EpiCore CDM specification. Together, these tools constitute an accessible framework and suggested workflow for register-based research. The technical expertise and infrastructure required to implement this framework are within the reach of any research unit or individual researcher with basic experience in handling register data. We acknowledge that a file-based implementation may, in some settings, pose technical performance challenges compared to database systems, particularly for high-concurrency or transactional use cases. However, we believe these are outweighed by the advantages in accessibility and ease of use. Moreover, modern columnar file formats such as .fst and .parquet offer competitive performance in terms of access speed, compression, and memory efficiency, thereby mitigating many of the limitations traditionally associated with file-based designs.

Although we like pragmatic solutions whenever possible, we also recognize the value of more complex CDM initiatives. For large-scale international collaborations and infrastructure projects, more sophisticated CDMs are often required. For example, the Data Analysis and Real World Interrogation Network (DARWIN EU [19]) project, which aims to conduct over 140 annual studies on drug use, safety, and effectiveness across the European Union, relies on the OMOP CDM to harmonize healthcare data across participating countries. However, adopting the OMOP CDM requires technical skills, infrastructure and resources that may not be realistic in many research groups, particularly those with limited experience in CDM-based workflows.

To realize the benefits of CDM-based research, transforming source data into the CDM must be as streamlined as possible. ETL development is complex and requires detailed knowledge of source data, creating a strong incentive to standardize ETL procedures. In Denmark, we plan to release an ETL package for selected source data, with similar efforts in Norway and Sweden to follow and be shared as they become available.

Our hope is that the EpiCore CDM will gain traction, as a starting point for many collaborative CDM-based studies in the Scandinavian setting. Moreover, we believe that it can serve as a valuable training platform for broader CDM skill development, thereby equipping research groups with the expertise needed to adopt more complex frameworks in the future.

7.1 | Plain Language Summary

High-quality research in healthcare requires contributions from many individuals and, ideally, multiple countries or settings. Since data are generally not allowed to be combined across sites, CDMs are used to ensure data are represented consistently, enabling the same analyses to run smoothly everywhere.

In many cases, simple CDMs are sufficient, particularly in comparable settings such as the Scandinavian countries. However, many existing CDMs demand significant resources, infrastructure, and expertise to implement. The EpiCore CDM presented in this paper is a pragmatic and user-friendly model specifically designed for pharmacoepidemiological research using Scandinavian register data.

EpiCore supports the key data sources most commonly used in this research area, including prescribed drugs, hospital encounters, and associated diagnoses and procedures. It is designed to be flexible, allowing users to adjust the model when needed, for example, to incorporate additional data sources. We also believe that EpiCore can enhance the way researchers manage and analyze data in their local settings.

To complement the CDM, we have developed an automated tool, EpiCoreAssistant, which helps researchers verify that their source data have been correctly transformed into the EpiCore CDM structure.

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. K. Laugesen, J. F. Ludvigsson, M. Schmidt, et al., "Nordic Health Registry-Based Research: A Review of Health Care Systems and Key Registries," *Clinical Epidemiology* 13 (2021): 533–554, https://doi.org/10.2147/CLEP.S314959.

2. H. Kieler, M. Artama, A. Engeland, et al., "Selective Serotonin Reuptake Inhibitors During Pregnancy and Risk of Persistent Pulmonary Hypertension in the Newborn: Population Based Cohort Study From the Five Nordic Countries," *BMJ (Clinical Research Ed.)* 344 (2011): d8012, https://doi.org/10.1136/bmj.d8012.

3. S. M. Baxter, L. C. Lund, J. H. Andersen, et al., "Glucagon-Like Peptide 1 Receptor Agonists and Risk of Thyroid Cancer: An International Multisite Cohort Study," *Thyroid* 35, no. 1 (2025): 69–78, https://doi.org/10.1089/thy.2024.0387.

4. A. Pottegård, L. C. Lund, Ø. Karlstad, et al., "Arterial Events, Venous Thromboembolism, Thrombocytopenia, and Bleeding After Vaccination With Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: Population Based Cohort Study," *BMJ (Clinical Research Ed.)* 373 (2021): n1114, https://doi.org/10.1136/bmj.n1114.

5. E. Simonsen, L. C. Lund, M. T. Ernst, et al., "Use of Semaglutide and Risk of Non-Arteritic Anterior Ischemic Optic Neuropathy: A Danish–Norwegian Cohort Study," preprint, December 11, 2024, https://doi.org/10.1101/2024.12.09.24318574.

6. R. Wesselhoeft, P. B. Jensen, A. Talati, et al., "Trends in Antidepressant Use Among Children and Adolescents: A Scandinavian Drug Utilization Study," *Acta Psychiatrica Scandinavica* 141, no. 1 (2020): 34–42, https://doi.org/10.1111/acps.13116.

7. S. R. Raman, K. K. C. Man, S. Bahmanyar, et al., "Trends in Attention-Deficit Hyperactivity Disorder Medication Use: A Retrospective Observational Study Using Population-Based Databases," *Lancet Psychiatry* 5, no. 10 (2018): 824–835, https://doi.org/10.1016/S2215-0366(18)30293-1.

8. R. Gini, M. C. J. Sturkenboom, J. Sultana, et al., "Different Strategies to Execute Multi-Database Studies for Medicines Surveillance in Real-World Setting: A Reflection on the European Model," *Clinical Pharmacology & Therapeutics* 108, no. 2 (2020): 228–235, https://doi.org/10.1002/cpt.1833.

9. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP), "Guide on Methodological Standards in Pharmacoepidemiology (Revision 11). EMA/95098/2010," http://www.encepp.eu/standards_and_guidance.

10. S. J. Reisinger, P. B. Ryan, D. J. O'Hara, et al., "Development and Evaluation of a Common Data Model Enabling Active Drug Safety Surveillance Using Disparate Healthcare Databases," *Journal of the American Medical Informatics Association* 17, no. 6 (2010): 652–662, https://doi.org/10.1136/jamia.2009.002477.

11. R. J. Desai, K. Marsolo, J. Smith, et al., "The FDA Sentinel Real World Evidence Data Enterprise (RWE-DE)," *Pharmacoepidemiology and Drug* 33, no. 10 (2024): e70028, https://doi.org/10.1002/pds.70028.

12. N. H. Thurin, R. Pajouheshnia, G. Roberto, et al., "From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding," *Clinical Pharmacology and Therapeutics* 111, no. 1 (2022): 321–331, https://doi.org/10.1002/cpt.2476.

13. P. B. Jensen, "Epicoreassistant. Epicoreassistant Gitlab Repository," 2025, https://gitlab.sdu.dk/pharmacoepi/epicoreassistant.

14. S. V. Wang, A. Pottegård, W. Crown, et al., "HARmonized Protocol Template to Enhance Reproducibility of Hypothesis Evaluating Real-World Evidence Studies on Treatment Effects: A Good Practices Report of a Joint ISPE/ISPOR Task Force," *Value in Health* 25, no. 10 (2022): 1663–1672, https://doi.org/10.1016/j.jval.2022.09.001.

15. M. Klik, "fst: Lightning Fast Serialization of Data Frames," 2022, http://www.fstpackage.org.

16. J. M. Cohen, C. E. Cesta, L. Kjerpeseth, et al., "A Common Data Model for Harmonization in the Nordic Pregnancy Drug Safety Studies (NorPreSS)," *Norsk Epidemiologi* 29, no. 1–2 (2021): 117–123.

17. M. G. Kahn, T. J. Callahan, J. Barnard, et al., "A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data," *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)* 4, no. 1 (2016): 18, https://doi.org/10.13063/2327-9214.1244.

18. "Apache Parquet," https://parquet.apache.org/docs/.

19. European Medicines Agency (EMA), "Data Analysis and Real World Interrogation Network (DARWIN EU)," accessed August 21, 2024, https://www.ema.europa.eu/en/about-us/how-we-work/big-data/real-world-evidence/data-analysis-real-world-interrogation-network-darwin-eu.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** pds70241-sup-0001-SupinfoS1.xlsx. **Data S2:** pds70241-sup-0002-SupinfoS2.docx.