

Review

# Methods Used in the Development of Common Data Models for Health Data: Scoping Review

Najia Ahmadi<sup>1</sup>, MSc; Michele Zoch<sup>1</sup>, Dipl-Wi-Inf; Patricia Kelbert<sup>2</sup>, MSc; Richard Noll<sup>3</sup>, MSc; Jannik Schaaf<sup>3</sup>, Dr rer med; Markus Wolfien<sup>1,4</sup>, Dr-Ing; Martin Sedlmayr<sup>1</sup>, Prof Dr

<sup>1</sup>Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

<sup>2</sup>Fraunhofer Institute for Experimental Software Engineering IESE, Kaiserslautern, Germany

<sup>3</sup>Institute of Medical Informatics, Goethe University Frankfurt, University Hospital, Frankfurt, Germany

<sup>4</sup>Center for Scalable Data Analytics and Artificial Intelligence, Dresden/Leipzig, Germany

**Corresponding Author:**

Najia Ahmadi, MSc

Institute for Medical Informatics and Biometry

Carl Gustav Carus Faculty of Medicine

Technische Universität Dresden

Fetscherstr 74

Dresden, 01307

Germany

Phone: 49 351458 87 7704

Email: [najia.ahmadi@tu-dresden.de](mailto:najia.ahmadi@tu-dresden.de)

## Abstract

**Background:** Common data models (CDMs) are essential tools for data harmonization, which can lead to significant improvements in the health domain. CDMs unite data from disparate sources and ease collaborations across institutions, resulting in the generation of large standardized data repositories across different entities. An overview of existing CDMs and methods used to develop these data sets may assist in the development process of future models for the health domain, such as for decision support systems.

**Objective:** This scoping review investigates methods used in the development of CDMs for health data. We aim to provide a broad overview of approaches and guidelines that are used in the development of CDMs (ie, common data elements or common data sets) for different health domains on an international level.

**Methods:** This scoping review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. We conducted the literature search in prominent databases, namely, PubMed, Web of Science, Science Direct, and Scopus, starting from January 2000 until March 2022. We identified and screened 1309 articles. The included articles were evaluated based on the type of adopted method, which was used in the conception, users' needs collection, implementation, and evaluation phases of CDMs, and whether stakeholders (such as medical experts, patients' representatives, and IT staff) were involved during the process. Moreover, the models were grouped into iterative or linear types based on the imperativeness of the stages during development.

**Results:** We finally identified 59 articles that fit our eligibility criteria. Of these articles, 45 specifically focused on common medical conditions, 10 focused on rare medical conditions, and the remaining 4 focused on both conditions. The development process usually involved stakeholders but in different ways (eg, working group meetings, Delphi approaches, interviews, and questionnaires). Twenty-two models followed an iterative process.

**Conclusions:** The included articles showed the diversity of methods used to develop a CDM in different domains of health. We highlight the need for more specialized CDM development methods in the health domain and propose a suggestive development process that might ease the development of CDMs in the health domain in the future.

(*JMIR Med Inform* 2023;11:e45116) doi: [10.2196/45116](https://doi.org/10.2196/45116)

**KEYWORDS**

common data model; common data elements; health data; electronic health record; Observational Medical Outcomes Partnership; stakeholder involvement; Data harmonisation; Interoperability; Standardized Data Repositories; Suggestive Development Process; Healthcare; Medical Informatics;

## Introduction

### Rationale

Integration of heterogeneous data is a ubiquitous topic in modern medicine. The arising large variety of data has the potential to provide in-depth insights about different aspects of clinical care and can lead to improvements in health care [1,2]. Yet, challenges, such as the identification and access of relevant data, the association between different data sources, and the assurance of data quality given the structural variations among data sources, still pose major barriers [3,4]. Common data models (CDMs) provide the possibility of harmonizing data from disparate sources, storing information in a standard structure by defining the syntax and semantics of data, and enabling operations on data using standard analysis methods [5]. In particular, a CDM contains a unified set of metadata, allowing data and its information content to be shared across applications and institutional borders, and thus enabling harmonized data integration and analysis on an international scale [6].

In the health domain, there are different types of CDMs (eg, CDMs for harmonization and storage of electronic health record-based patient data). An example is the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) developed by the Observational Health Data Science and Informatics (OHDSI) community, which ensures homogeneous storage of observational health care data across different databases with similar formats and terminologies [7]. There are also further CDMs for clinical data, like Sentinel CDM, Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM), and National Patient-Centered Clinical Research Network (PCORnet) [8], and data warehouse models, like Informatics for Integrating Biology and the Bedside (i2b2) [9]. Moreover, some CDMs define the data from patient cohorts and describe a medical specialty or a group of diseases. For example, there are specific CDMs for the domain of rare diseases [10,11] or radiology [12]. Overall, there is a large variety of CDMs in the literature for common, rare, and context-specific medical examinations, and each of them follows a more self-defined development process.

As described by Melles et al [13], a practical design meets the users' needs. While designing a CDM in the health domain, in addition to the developers (ie, IT staff and computer scientists), the primary stakeholders (ie, patients and clinicians) are particularly interested in the outcome. It is therefore recommended to include them in the design process as early as possible [13,14]. In addition to the stakeholders, the medical context is also quite complex and requires extensive medical and technical expertise to ensure the usefulness of the model after its development. This is why the development process of a CDM is critical and a comprehensive development method or guideline is necessary.

Studies, such as those by Gericke and Blessing [15] and Bobbe et al [16], have already tried to determine the commonalities and differences in development processes across disciplines. Bobbe et al [16] performed a comparison of design models from academic theory and professional practice, and discussed 8 types

of design processes. In particular, the basic design cycle, V design process, human-centered design, hypercyclic design, Munich procedural model, double diamond model, frog model, and IDEO model were presented. Additionally, Melles et al [13] introduced categories for models, namely, whether a model is activity-based or stage-based, solution-oriented or problem-oriented, and design-focused or project-focused.

However, given the complexity of the health domain and the importance of many stakeholders taking part in the process, it might be difficult to transfer models from other disciplines. This is why we aim to derive such a process and review the available CDM instances in the domain. Exemplarily, the results of this scoping review will be integrated into the design and development of a CDM for the SATURN ("Smarter Arztportal für Betroffene mit unklarer Erkrankung" ["Smart physicians' platform for patients with unclear diseases"]) Project in the future [17]. This project aims to develop an artificial intelligence-based diagnosis support tool for primary care physicians. With the help of user-centered design, the requirements of a decision support tool, especially for noncharacteristic symptoms, will be studied. The medical focus is on the diagnosis of unclear and rare medical conditions. This is why, in this review, we focus on the similarities between the CDM development methods in rare medical conditions and common medical conditions in order to determine whether the methods for common medical conditions can be adopted for rare medical conditions as well. On a technical level, rule-based systems, machine learning, and case-based reasoning will be implemented. As part of this project, CDMs for 3 groups of rare diseases, namely, endocrinology, gastroenterology, and pneumology, will be developed.

Our review contributes to the analysis of CDM development methods in the health domain on an international scale and aims to explore the actual involvement of stakeholders, especially medical experts, in the development process. To the best of our knowledge, this is the first scoping review focusing on CDM development methods in the health domain.

### Objectives and Research Questions

This scoping review has been conducted to provide an overview of the methods used for the initial and further development of CDMs in the health domain. We divided the overall development process into conception, users' needs collection (eg, collection of evidence, review of the literature, and guidelines), and implementation, as well as individual evaluations within the phases. We consider the conception phase as an initial step, where the CDM is theoretically designed along with stakeholders. Subsequently, the essential elements previously identified are gathered in the "users' needs collection" phase. The finalized process, in which the conceptualized model is implemented and ready-to-use, is termed the implementation phase.

According to the rationale and objective explained above, this scoping review examines the following questions:

1. How are CDMs methodically developed in the health domain? What requirement analysis methods, design processes, and validation methods were used?

- How or when do stakeholders, especially medical experts, get involved in the development process?
- How can the CDM development methods be classified based on their requirement analysis methods, design processes, validation methods, and model type?

## Methods

### Protocol and Registration

To ensure methodological quality, this scoping review has followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist [18]. According to this checklist, we published and registered the review protocol [19]. Out of the 22 items of the PRISMA checklist, 20 have been considered in this review (Multimedia Appendix 1).

### Search Strategy

To achieve a comprehensive query, an initial search was performed in PubMed with the term “common data model.” Six randomly chosen articles matching the topic were analyzed [10-12,20-22]. The keywords associated with the articles listed in Table 1 were considered and subsequently tested in the query.

The combination of terms that delivered the highest number of matching articles was included in our final search string.

Some studies used the term *data set* [11], and others defined alternative data elements that can be part of a data set or data model [10]; thus, to avoid the exclusion of certain studies, we jointly used the following terms in our search string: *common data model*, *common data element*, and *common data sets*. We also added the short forms of these terms in our search string and analyzed the relevance of the results by simply looking into the resulting literature. Additionally, we added the following terms in our search string to ensure that the included CDMs were developed within the health domain: medical, medicine, health, healthcare, health care, electronic health, clinical, and disease. The search string used in PubMed is presented in Table 2. It was developed as a combination of the mentioned terms, their possible variations, and where applicable, Medical Subject Headings (MeSH) [23]. The search strings used in the other 3 databases have been provided in Multimedia Appendix 2.

The query was designed and tested by the author NA and was approved by all coauthors. The resulting articles were added to Rayyan (Rayyan Systems Inc) [24] for further screening and annotation.

**Table 1.** Six randomly chosen articles for the construction of the search string and their keywords.

Article title	Keywords
The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration [10]	Registries, common data elements, European platform, rare diseases, patient registration, and EPIRARE
A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research [11]	Common data elements, interoperability, metadata, minimum data set, national health program, and rare diseases
Development and validation of the Radiology Common Data Model (R-CDM) for the international standardization of medical imaging data [12]	Metadata, standardization, and radiology information system
Common data model for natural language processing based on two existing standard information models: CDA+GrAF [20]	Natural language processing, medical informatics, data model, information model, HL7 clinical document architecture, and ISO graph annotation format
Genomic common data model for biomedical data in clinical practice [21]	High-throughput nucleotide sequencing, data analysis, and observational study
Towards a newborn screening common data model: The Utah Newborn Screening Data Model [22]	Newborn screening, newborn screening laboratory information management system, common data model, interoperability, electronic data exchange, NBS, LIMS, and standards

**Table 2.** Search strings used to identify articles from PubMed.

Search aspects	Variations	Search string <sup>a</sup>
Common data model	Common data model (CDM), common data element (CDE), and common data sets (CDS)	(“common data model” AND CDM) OR (“common data element*” AND CDE) OR “Common Data Elements”[Mesh] OR “common dataset*” OR “common data set*”
Health care	Medical, medicine, health, healthcare, health care, electronic health, and disease	medical OR medicine OR “Medicine”[Mesh] OR health OR “Health”[Mesh] OR healthcare OR “health care” OR “electronic health” OR clinical OR disease OR “Disease”[Mesh]

<sup>a</sup>The common data model and health care search terms were combined with “AND.”

In particular, literature from 2000 to 2022 was considered, which is an extension of the previously published study protocol [19]. It is also noteworthy that the MeSH terms were only available in PubMed. The language of the articles was limited to English. Using the Boolean operators “AND” and “OR,” the systematic

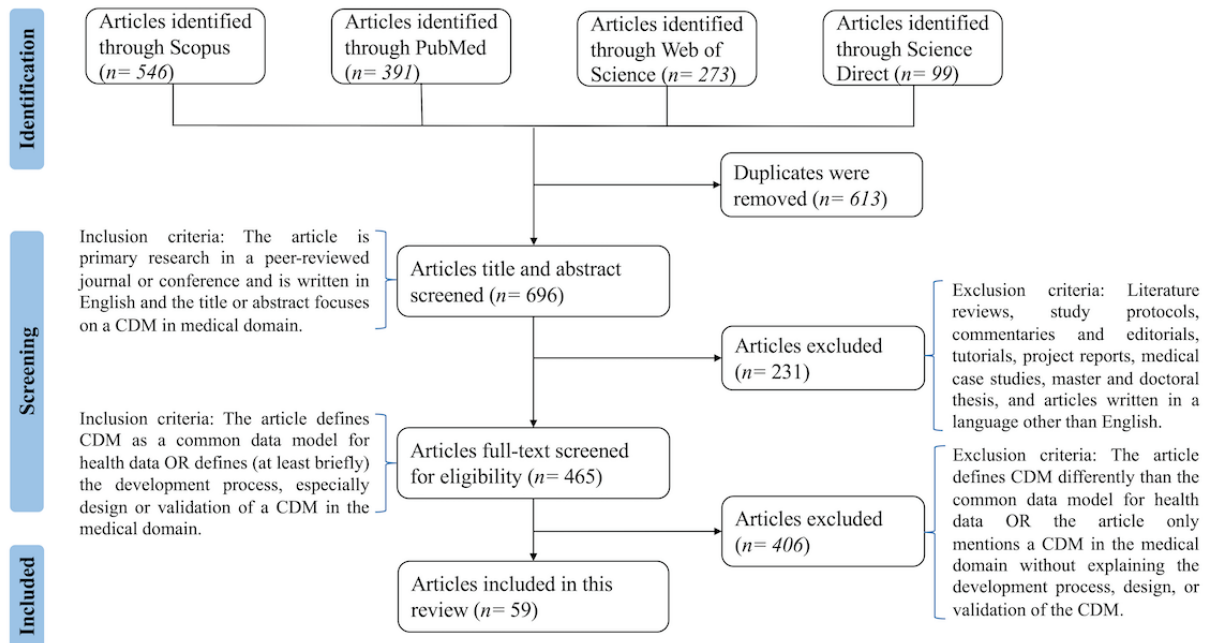
search was carried out in the following electronic databases: PubMed, Web of Science, Science Direct, and Scopus. The search was performed in March 2022. The publication date tag in PubMed and Web of Science was set to January 1, 2000, to March 15, 2022, and that in Science Direct and Scopus was set

to 2000 to 2022 (it is not possible to specify the month and day in Science Direct and Scopus).

### Inclusion and Exclusion Criteria

The inclusion and exclusion criteria are summarized in [Multimedia Appendix 3](#) and are visualized along with the number of outcome articles in [Figure 1](#).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing the paper selection process and the inclusion and exclusion criteria. CDM: common data model.



### Selection and Review of Articles

Duplicates were removed using the built-in function in Rayyan [24]. The process of deletion was monitored by the author NA. After eliminating duplicates, the selection of studies was performed in 2 steps. The title and abstract screening steps were performed by the authors in groups of two. The articles were tagged as “include,” “exclude,” or “maybe.” Tagged articles were decided upon based on the tags described in [Table 3](#).

Disagreements were resolved by a third author. This process was initially carried out on 10% of the articles to confirm the accuracy of our inclusion and exclusion criteria, and clarify ambiguities. After the title and abstract screening, the full text of the included articles was screened by the authors, again in groups of two. The selected articles were included in the data extraction step.

**Table 3.** Description of tags used by the authors in the article screening process.

Author 1	Author 2	Decision
Include	Include	Included
Include	Exclude	Discuss and decide together
Include	Maybe	Include
Exclude	Maybe	Exclude
Maybe	Maybe	Discuss and decide together

### Data Charting and Extraction Process

A data charting table was developed and refined throughout the study, with several iterations. This table contained a list of items that were extracted from all included publications. All authors examined 10% of the articles for the defined data items and refined the data charting table, if necessary. The data charting table, including the extracted information from articles, is included in [Multimedia Appendix 4](#).

For each article, we focused on 4 major aspects: (1) the meta information, such as DOI, authors, year, country, and project

name, if applicable; (2) the medical condition for which the CDM was built, whether the condition is rare or common, the organ affected by the condition, and whether the condition is long term (longer than a year) or short term; (3) methodological information, such as requirement analysis, design, and validation process; whether the design process was linear or iterative; and advantages and disadvantages of the method, as stated in the respective article; and (4) information about stakeholder involvement. The extracted data elements, their categories, and their definitions are shown in [Table 4](#).

**Table 4.** Data extraction sheet with specified elements, categories, and subcategories, including their definitions.

Category and subcategory	Definition
<b>Meta information</b>	
DOI	A link to the article
Author	First author's name
Publication year	Year of the publication date of the article
Country of study	Country of the leading author's affiliation
Project name	If applicable; when the CDM <sup>a</sup> study was part of a project/consortium
<b>Medical background</b>	
Medical condition	Name of the medical condition for which the CDM was built
Organ function	Organ affected by the medical condition
Short-term/long-term condition	Short term: less than a year; long term: longer than a year
Is the condition rare or common?	Is the medical condition considered rare or common based on its occurrence? Available answers: common medical condition, rare medical condition, and conditions that can be rare and common.
<b>Requirement analysis method</b>	
Literature analysis	It includes searching in a variety of literature, such as extraction of frequent CDEs <sup>b</sup> from real-world data, data harmonization across studies, multicenter longitudinal and observational studies, consensus documents and guidelines, primary outcome data of trials, review of instruments, and forms like report forms, users' needs collection forms, etc.
Interview/questionnaire	It includes expert interviews, focus group meetings, working group meetings, consensus meetings, workshops and discussions, and online surveys.
Delphi	Delphi or modified Delphi was used. Delphi techniques involve experts evaluating complex issues iteratively, where knowledge is incomplete or uncertain. Typically, the response from the previous questionnaire is appended to the next questionnaire [25].
Review of existing CDEs	When an existing CDE was validated/reviewed.
<b>Design</b>	
Creation of new CDEs	If there were no CDEs in the domain and the experts tried to come up with some CDEs using literature in the field.
Modification of existing CDEs	If existing CDEs in a disease domain were modified.
Reuse of existing CDEs (without modification)	If existing CDEs in the domain were used without any modification.
<b>Validation</b>	
External experts	It includes only external validation of any sort, such as public reviews on a website from experts or nonexperts in the field. Excluded are experts that were part of the conception process of the model.
Others	Any other type of validation, such as internal reviews, working group consensus, etc.
<b>Model type</b>	
Iterative	When at least one iterative process was performed during development of the CDM.
Linear	When there was no iteration in the process.
<b>Stakeholder information</b>	
Were stakeholders involved in the design process?	Yes/no
Which stakeholders were involved?	Patients' representatives, clinicians, domain experts, computer scientists, IT personnel, and registry staff
When did they get involved in the process?	In users' needs collection (when experts were involved in the preanalysis step, eg, collection of evidence, review of literature, guidelines, etc), in conception (when experts were involved in conception of the CDEs), in evaluation (when the model was evaluated via experts), and in implementation (when experts were involved in the implementation of the model).
What was the nature of stakeholder involvement?	Through expert workshops, semistructured interviews, questionnaires, etc
<b>Pros and cons of methods as mentioned in the article</b>	

Category and subcategory	Definition
Pros	Advantages of the method as stated in the article
Cons	Disadvantages of the method as stated in the article

<sup>a</sup>CDM: common data model.

<sup>b</sup>CDE: common data element.

## Visualization and Summarization of Results

At the end of the data extraction, the data items collected in [Table 4](#) were summarized and visualized. A flowchart according to the PRISMA-ScR guidelines was designed to show the article processing approach ([Figure 1](#)). Tables, timeline plots, histogram charts, pie plots, and scatter histograms were used to display the extracted data items. The graphics and the required analysis were performed using Python version 3.9.12 (Python Software Foundation), with matplotlib, pandas, and NumPy packages. The script used for the plots is publicly available [[19,26](#)].

First, we aimed for a broad overview of available CDMs and whether original CDMs were developed or existing CDMs were modified, as well as whether they focused on common or rare diseases and addressed a specific organ function. Second, to answer our first research question, we documented the medical domain of each article, whether the medical condition was considered as long term (more than a year) or short term (less than a year), and the affected organ as stated in the respective original article. To classify the development process of the CDMs, we documented 4 categories of data information for each article: requirement analysis, design, validation, and model type ([Table 4](#)). We categorized the methodology that was used for the requirement analysis (ie, why a CDM was needed), as well as the context to design a set of common data elements (CDEs). For validation, we distinguished between external evaluation and any other type of evaluation. The “other” category included the evaluations performed by the same clinical experts who were involved in the conception process, such as working group consensus, user evaluations, reviews performed via the members of the project, statistical tests, and pilot tests conducted within the project. Additionally, we investigated

stakeholder involvement in the development stages in those studies and whether the studies followed an iterative or linear method of development. We used the advantages and disadvantages of the methods as stated in the articles ([Table 4](#)) and formulated them into a list of constraints in the area of CDM development to further highlight the need for streamlined methods. Finally, after analyzing the included CDMs, we summarized the most frequent methods used in the included literature in a suggestive development process that could be a reasonable basis to start with when developing a novel model.

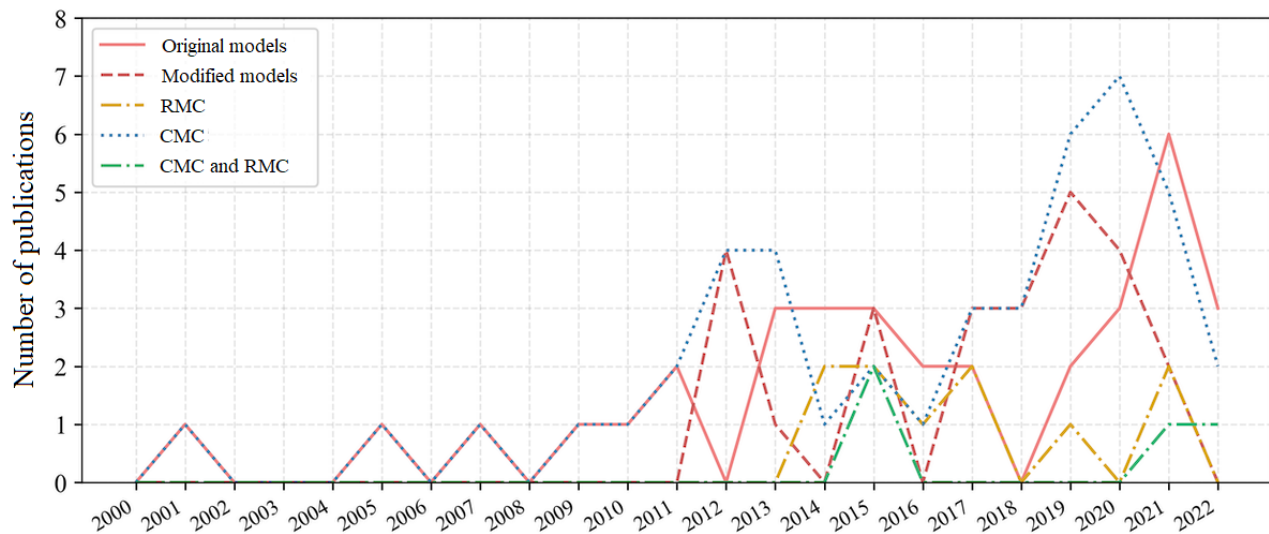
## Results

### Selection of Articles

In total, we identified 1309 articles from PubMed, Web of Science, Science Direct, and Scopus search engines. From the identified articles, after duplicate removal, 695 articles were included in the title and abstract screening. Finally, 465 articles underwent full-text screening, and of these, 59 matched the full-text screening criteria of this review and were finally included. We excluded articles that did not describe the development or evaluation of a CDM in the health domain. Additionally, articles that were not publicly available and those in a language other than English were excluded. The article identification process along with the inclusion and exclusion criteria are shown in [Figure 1](#).

The selected articles defined CDMs, common data sets, or CDEs for common or rare medical conditions. All included articles were published between 2000 and 2022. As shown in [Figure 2](#), the number of articles that focused on CDM development increased after 2011 and continued to increase in the last years.

**Figure 2.** The number of publications focusing on common data model (CDM) development per year from 2000 to 2022. The line chart compares the number of articles developing original CDMs (original models) with the number of articles developing CDMs via modification of existing models (modified models), and compares the number of articles developing CDMs for rare medical conditions (RMCs), the number of articles developing CDMs for common medical condition (CMCs), and the number of articles developing CDMs for both kinds of conditions (CMCs and RMCs). In addition to the increase in the number of articles from 2011 in general, we can see that CDMs for rare diseases were only developed starting from 2014.



### Country of Publication

We categorized the articles into countries based on the affiliation of the first author. Among the 59 articles, 26 (44%) were published in the United States, 8 (14%) were published in Canada, and 6 (10%) were published in Germany. The number of articles according to country is as follows: Belgium, 2 [27,28]; Canada, 8 [29-36]; China, 1 [37]; Denmark, 2 [38,39]; France, 2 [11,40]; Germany, 6 [41-46]; Italy, 1 [10]; Spain, 1 [47]; Republic of Korea, 1 [48]; Norway, 3 [49-51]; Switzerland, 1 [52]; Taiwan, 1 [53]; the Netherlands, 1 [54]; United Kingdom, 3 [55-57]; and United States, 26 [58-83].

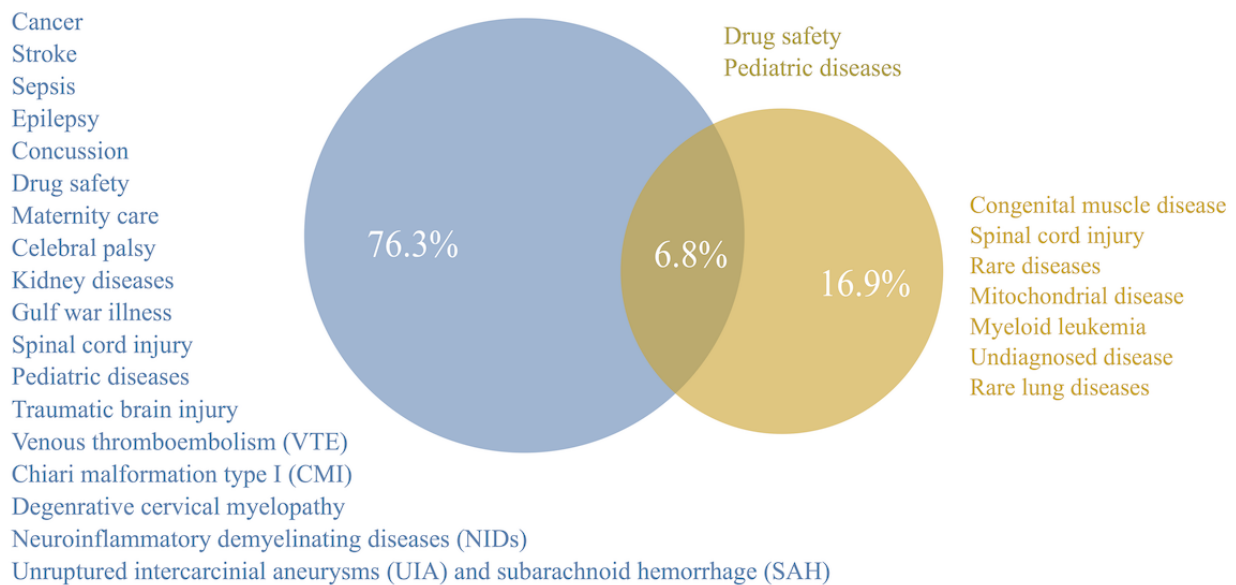
### Medical Conditions and Their Domains

According to our research, CDMs were developed for a variety of medical domains in the past 22 years; however, we divided them into 3 categories, namely, rare, common, and rare and common (both). An aggregated list of the medical conditions and their domains is shown in Figure 3. A full list of the medical conditions extracted during this scoping review is shown in Multimedia Appendix 4. An organ function overview and the long- and short-term conditions are shown in Multimedia

Appendix 5. Among these, 10 (17%) CDMs were designed for rare medical conditions, such as myeloid leukemia and rare lung diseases, and mitochondrial diseases [41,44-46,59]. Moreover, 1 CDM, namely, the CDM in the study by Berger et al [44], was designed for undiagnosed diseases in general.

Among the 59 articles, 45 involved the development of a CDM for common medical conditions. These included traumatic brain injury [27,28,30], spinal cord injury in children and youth [67], dental caries [68], sport-related concussion [65], cerebral palsy [29], degenerative cervical myelopathy [55], unruptured intracranial aneurysms and subarachnoid hemorrhage [32,42,55,60], Chiari malformation type I [63], breast implant [43], stroke [37], venous thromboembolism [33], pediatric epilepsy [61], pediatric critical illness [62], pregnancy drugs and treatments [49], sepsis [31], medication use in pregnancy and breastfeeding [40], degenerative cervical myelopathy [55], Gulf War illness [58], neuroinflammatory demyelinating disease [43], traumatic brain injury [27], and neurologic disorder and stroke [69]. Wandner et al [66] focused on clinical pain management, and Jaboyedoff et al [52] focused on pediatric diseases in general.

**Figure 3.** Characteristics of the included studies. A Venn diagram showing the proportions of identified common data models (CDMs) for common medical conditions (76.3%; blue), rare medical conditions (16.9%; golden yellow), and medical conditions that could fit into both categories (6.8%). Additionally, an aggregated list of medical conditions that CDMs were developed for in the studies is shown in 3 different colors according to their categories.



### Stakeholder Involvement

To investigate the involvement of stakeholders, we summarized at which particular stage they were involved in the CDM development process. Out of the 59 included articles, 54 (92%) mentioned at least one stakeholder in the design process. Additionally, we were interested in the different types of stakeholders that were involved, how they were involved, and at what stage of the process they typically got involved. As shown in Figure S1 in [Multimedia Appendix 6](#), stakeholders were mostly involved in the initial stage, namely, the conception phase. Domain experts and clinicians were the most common stakeholders involved in the studies (Figure S2 in [Multimedia Appendix 6](#)). Additionally, while many different methods were used to involve the stakeholders, such as expert groups, surveys, consensus meetings, interviews, teleconferences, questionnaires,

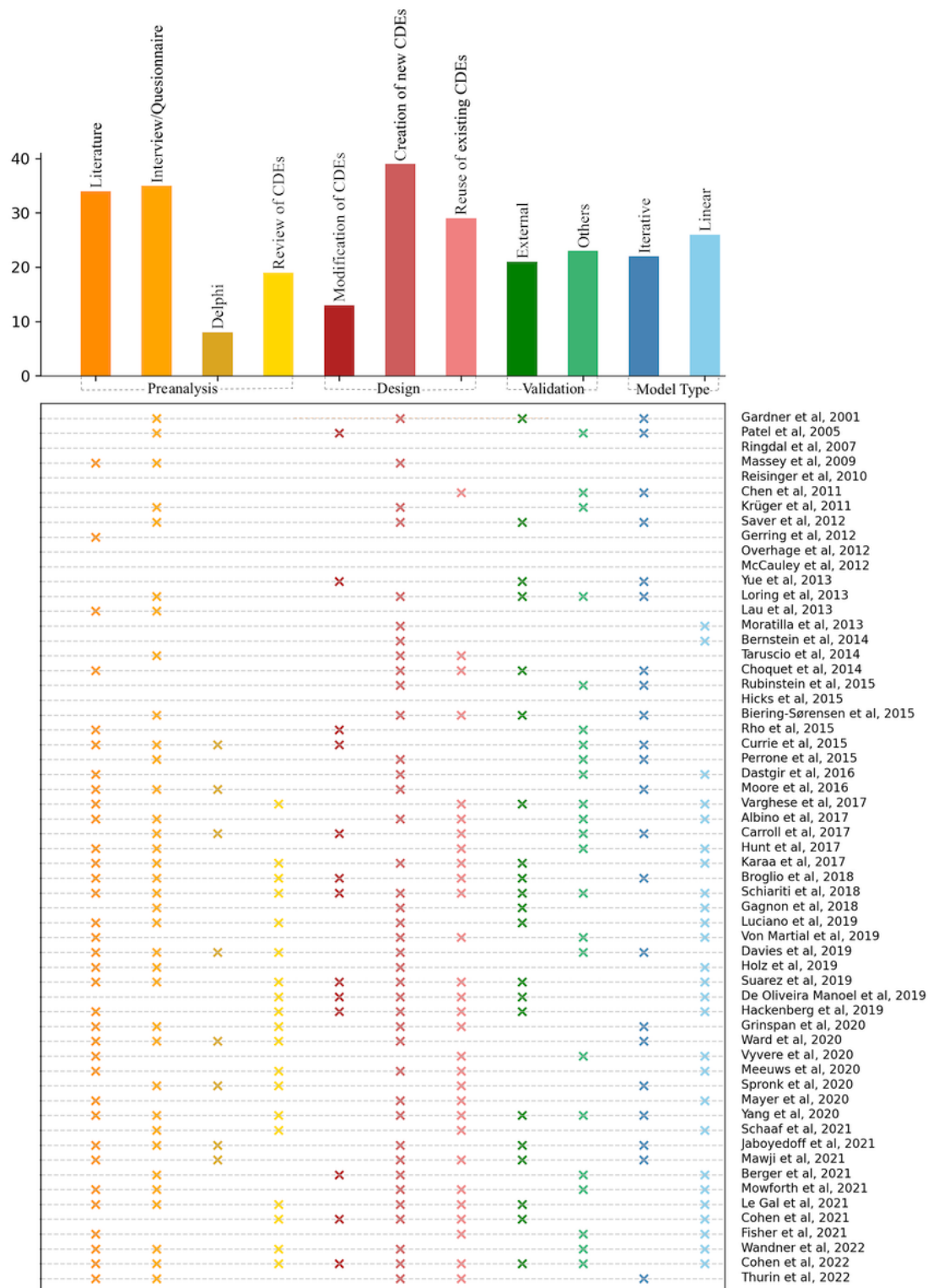
and workshops, “working group” was the most frequent method used (Figure S3 in [Multimedia Appendix 6](#)).

### Design Process

The methods used in the articles for designing a CDM were literature analysis, interview, Delphi, and review of existing CDEs. From our extraction table ([Multimedia Appendix 4](#)), we noted that 39 articles involved the definition of an original model/set of CDEs, 13 involved the modification of an existing set of CDEs, and 29 involved the use of an existing set of CDEs without any modifications. The external evaluation included web-based feedback, public review and comments, and feedback in a conference, among others. Finally, we found that 26 articles involved a rather linear design method and 22 others involved an iterative process. The list of articles that involved the use of each of these categories is shown in [Figure 4](#). Detailed information is presented in [Multimedia Appendix 4](#).



**Figure 4.** Methodological information on the articles [10,11,27-83]. The y-axis shows the list of articles by publication year. The x-axis shows the methodological categories. The scatter plot includes a cross mark when the Boolean is true for a specific article, for example, if the authors have used literature analysis as a preanalysis method, a cross (x) is added. The sum of cross marks in each column contributes to the bar size of the bar plot positioned on the x-axis. To improve visibility, each subcategory is shown with a different color. The subcategories of the same category are grouped via the same family of colors. CDE: common data element.



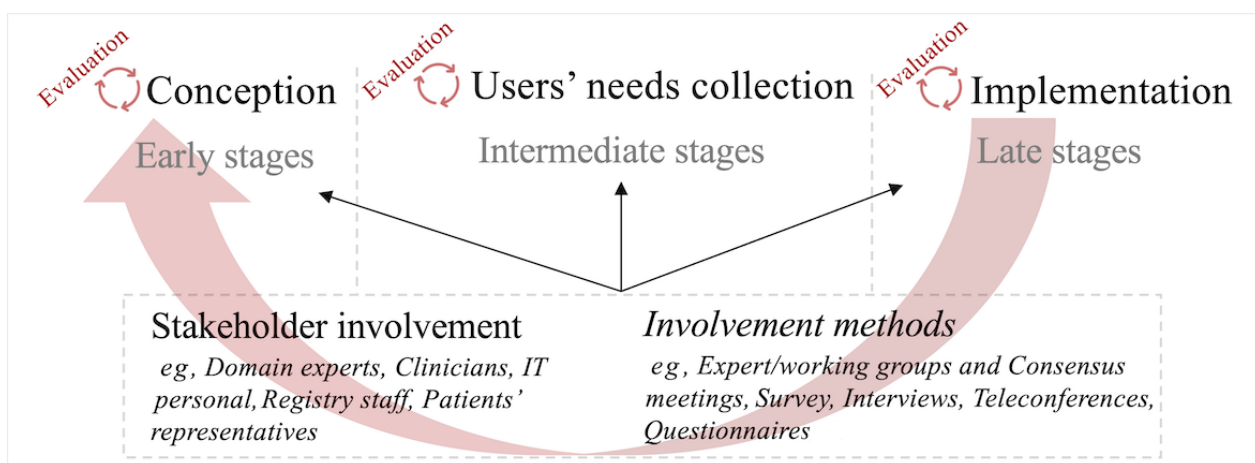
**Methodological Constraints Highlighted in Previous Studies**

The included articles presented a range of constraints in the development process from the methods used in the different stages of the process to the applicability of the outcome

elements. For example, Thurin et al [40] performed interviews with a single data access provider per data source and mentioned that other data access providers might conceptualize the data source differently. Additionally, they tested the applicability of the developed model only on the included data sources in the project. The model might require modification to use it with

other data sources. The limited sample size used to test the developed model is a common problem in rare conditions [44] given the rarity of the disease. One of the limitations mentioned by Broglio et al [65] is that some of their developed CDEs require special expertise that might not be implementable in certain settings. Grinspan et al [61] mentioned that some subcategories of epilepsy syndrome were merged at a level higher into a single category, which might have led to reduced data resolution, although uphill mapping is often used, especially in the OMOP context [5]. Additionally, the elements considered do not cover every possible influencing element, and the source was limited to only US-based patients, which means the elements can differ once an international data level is considered. They also included CDEs that were documented as free text, and processing of such elements might require natural language processing applications. The authors also highlighted the possible bias caused by the methodology used for consensus and discussion, and the Delphi approach, focus groups, and interviews might have also influenced the outcome of the study.

**Figure 5.** Summary of a basic common data model development process.



## Discussion

### Overview

One of the major challenges faced by CDM developers in the health domain is the lack of a comprehensive methodology or workflow to follow, which is also reflected in this review. The general models from industrial design and even academia (eg, the model introduced by Bobbe et al [16]) do not generally translate one-to-one to the health domain. The medical context is usually complex, and the involvement of stakeholders, such as clinicians, patients' representatives, and IT staff, is of utmost importance to ensure the applicability of a to-be-developed CDM. In addition, user-friendly, adaptable, and straightforward models are preferred in health care as one can start working with them without requiring a substantial amount of time [84].

This scoping review provides a summary of the development methods for CDMs and categorizes them based on the requirement analysis method, design process, validation approach, and model type. A variety of methods were used in the requirement analysis step in the articles, starting from searching in different types of literature and medical guidelines

### Essence of the CDM Development Process

Our outcomes showed that a heterogeneous variety of methods or processes were used in CDM development in the included articles, which highlights the need for a more streamlined field-specific development method. Therefore, we summarized our analysis outcomes into a suggestive development process (Figure 5), considering the 3 development steps that have been identified from the included models in this study, namely, conception, users' needs collection, and implementation. We suggest that evaluation and validation should be integrated into every stage of development, which gives the stages an iterative nature, and feedback should be integrated into the process as much as possible. We also emphasize the involvement of stakeholders in the process as early as possible and propose continuous involvement until the end of the development process because in every phase, questions might arise that need to be answered from different perspectives.

[43,44] to interviews [29], the Delphi approach [31], and a review of existing CDEs. A full list of these articles is shown in Figure 4 and Multimedia Appendix 4.

The majority of the developed CDMs have been designed for common medical conditions, and only 10 articles involved the design of a particular CDM for rare diseases. However, we did not find a significant difference in the development process of a CDM for rare and common conditions. Interestingly, based on our analysis, we can conclude that common medical conditions were the focus of CDM studies from early 2000, whereas the first CDM for rare conditions was developed in 2014. Despite methodological similarities, every article usually mentioned following a more individualistic method of development. This may arise because rare conditions occur rarely and the number of patients included in studies is limited [44]. Moreover, finding an expert for each rare or unclear disease is a challenging task. Additionally, most of the information crucial in the diagnosis of such diseases (like symptoms or phenotypes and genotypes) is currently stored in unstructured forms (eg, clinical notes). Extraction of such information requires a lot of time and effort from technical and clinical stakeholders [41].

Thus, given the variety of studies, the methods used for common conditions might be adaptable for rare conditions. Considering that a CDM is an essential part of data harmonization (a necessity in the health domain), we see highly emphasized development models as essential. Therefore, after analyzing the included CDMs, we summarized a suggestive development process that is shown in [Figure 5](#), which could be the starting point for conceptualizing and implementing novel CDMs.

### Limitations

The findings of our study are subject to certain limitations. First, our analysis is restricted to the selected databases, namely, PubMed, Web of Science, Science Direct, and Scopus. Additionally, the scope of our investigation is confined to articles published within a specific time frame and written in English. Moreover, we did not conduct any assessment of the quality of the included articles. In addition, it may also be worth noting that the authors of this review have varying interdisciplinary backgrounds, expertise levels, and experiences in the CDM field. However, to optimize the screening and analyzing processes, we performed them in pairs and first tested the method on a subset of 10% of the articles, resulting in a minimal number of conflicts.

### Conclusion

We considered 4 steps in the development of a CDM: conception, users' needs collection, implementation, and evaluation. We could identify 4 groups of methods that were most often used in the articles as part of the requirement analysis of the CDM development process. These were literature analysis, interviews, Delphi approaches, and review of existing CDEs. The articles considered in this review either developed a new CDE or made use of an existing set of CDEs with or without modification.

Most of the articles involved at least one stakeholder from among domain experts, clinicians, IT staff, registry staff, and patients' representatives, and mostly from the initial step, which was conception. The methods used to involve the stakeholders were expert groups, surveys, consensus meetings, interviews, working groups, teleconferences, questionnaires, and workshops, and among these, working groups were most often used.

We conclude that the methods used in the development of CDMs in the health domain are heterogeneous and this field is lacking solid guidelines that may ease up this process, especially in terms of the reusability and adaptability of a CDM. This is why the proposed outline ([Figure 5](#)) could be a reasonable basis to start with. In our future work, we plan to test and improve the proposed outline for developing a CDM.

### Acknowledgments

This work was accomplished as part of the SATURN (Smarter Arztportal für Betroffene mit unklarer Erkrankung; Smart physicians' platform for patients with unclear diseases) Project funded by the German Federal Ministry of Health as part of the research focus "Digital Innovation," Module 3: "Smart Algorithms and Expert Systems" (funding codes: 2520DAT02C, 2520DAT02B, and 2520DAT02D).

### Data Availability

The script used for analysis and visualization in this review is available at GitHub [26], and the study protocol can be accessed on the Open Science Framework (OSF) [19].

### Authors' Contributions

NA, MZ, PK, RN, MW, JS, and MS contributed to conceptualization and methodology. NA contributed to data acquisition. NA, MZ, PK, RN, and MW contributed to the literature screening. NA contributed to data analysis and interpretation. NA contributed to writing and preparing the original draft. NA, MZ, PK, RN, MW, and JS contributed to reviewing and editing the manuscript. NA and MW contributed to the visualization. MS contributed to resources. All authors take responsibility for the scientific integrity of the work. All authors have read and agreed to the published version of the manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist. [\[PDF File \(Adobe PDF File\), 940 KB-Multimedia Appendix 1\]](#)

### Multimedia Appendix 2

Search strings used in the PubMed, Web of Science, Science Direct, and Scopus databases to search for articles. [\[DOCX File , 17 KB-Multimedia Appendix 2\]](#)

### Multimedia Appendix 3

Inclusion and exclusion criteria for the title and abstract screening, and the full-text screening.

[\[DOCX File, 14 KB-Multimedia Appendix 3\]](#)

### Multimedia Appendix 4

Additional study information.

[\[XLSX File \(Microsoft Excel File\), 64 KB-Multimedia Appendix 4\]](#)

### Multimedia Appendix 5

Characteristics of the included studies.

[\[PDF File \(Adobe PDF File\), 99 KB-Multimedia Appendix 5\]](#)

### Multimedia Appendix 6

Summary of derived stakeholder information.

[\[PDF File \(Adobe PDF File\), 198 KB-Multimedia Appendix 6\]](#)

## References

1. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data* 2019 Jun 19;6(1):54 [doi: [10.1186/s40537-019-0217-0](#)]
2. Razzak MI, Imran M, Xu G. Big data analytics for preventive medicine. *Neural Comput Appl* 2020 Mar 16;32(9):4417-4451 [FREE Full text] [doi: [10.1007/s00521-019-04095-y](#)] [Medline: [32205918](#)]
3. Asche CV, Seal B, Kahler KH, Oehrlein EM, Baumgartner MG. Evaluation of healthcare interventions and big data: review of associated data issues. *Pharmacoeconomics* 2017 Aug;35(8):759-765 [doi: [10.1007/s40273-017-0513-5](#)] [Medline: [28474299](#)]
4. Kent S, Burn E, Dawoud D, Jonsson P, Østby JT, Hughes N, et al. Common problems, common data model solutions: evidence generation for health technology assessment. *Pharmacoeconomics* 2021 Mar;39(3):275-285 [FREE Full text] [doi: [10.1007/s40273-020-00981-9](#)] [Medline: [33336320](#)]
5. Maier C, Lang L, Storf H, Vormstein P, Bieber R, Bernarding J, et al. Towards implementation of OMOP in a German university hospital consortium. *Appl Clin Inform* 2018 Jan;9(1):54-61 [FREE Full text] [doi: [10.1055/s-0037-1617452](#)] [Medline: [29365340](#)]
6. Simko LC, Chen L, Amtmann D, Gibran N, Herndon D, Kowalske K, et al. Challenges to the standardization of trauma data collection in burn, traumatic brain injury, spinal cord injury, and other trauma populations: a call for common data elements for acute and longitudinal trauma databases. *Arch Phys Med Rehabil* 2019 May;100(5):891-898 [FREE Full text] [doi: [10.1016/j.apmr.2018.10.004](#)] [Medline: [31030731](#)]
7. Hripcsak G, Duke J, Shah N, Reich C, Huser V, Schuemie M, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574-578 [FREE Full text] [Medline: [26262116](#)]
8. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform* 2016 Dec;64:333-341 [FREE Full text] [doi: [10.1016/j.jbi.2016.10.016](#)] [Medline: [27989817](#)]
9. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010 Feb 26;17(2):124-130 [FREE Full text] [doi: [10.1136/jamia.2009.000893](#)] [Medline: [20190053](#)]
10. Taruscio D, Mollo E, Gainotti S, Posada de la Paz M, Bianchi F, Vittozzi L. The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration. *Arch Public Health* 2014 Oct 13;72(1):35 [FREE Full text] [doi: [10.1186/2049-3258-72-35](#)] [Medline: [25352985](#)]
11. Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. *J Am Med Inform Assoc* 2015 Jan;22(1):76-85 [FREE Full text] [doi: [10.1136/amiajnl-2014-002794](#)] [Medline: [25038198](#)]
12. Park C, You SC, Jeon H, Jeong CW, Choi JW, Park RW. Development and Validation of the Radiology Common Data Model (R-CDM) for the International Standardization of Medical Imaging Data. *Yonsei Med J* 2022;63(Suppl):S74 [doi: [10.3349/ymj.2022.63.s74](#)]
13. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. *Int J Qual Health Care* 2021 Jan 12;33(Supplement\_1):37-44 [FREE Full text] [doi: [10.1093/intqhc/mzaa127](#)] [Medline: [33068104](#)]
14. Sacristan JA, Aguaron A, Avendaño C, Garrido P, Carrion J, Gutierrez A, et al. Patient involvement in clinical research: why, when, and how. *Patient Preference and Adherence* 2016 Apr;6:31-640 [doi: [10.2147/ppa.s104259](#)]

15. Gericke K, Blessing L. An analysis of design process models across disciplines. In: DS 70: Proceedings of DESIGN 2012, the 12th International Design Conference. 2012 Presented at: 12th International Design Conference; May 21-24, 2012; Dubrovnik, Croatia
16. Bobbe T, Krzywinski J, Woelfel C. A comparison of design process models from academic theory and professional practice. In: DS 84: Proceedings of the DESIGN 2016 14th International Design Conference. 2016 Presented at: 14th International Design Conference; May 16-19, 2016; Dubrovnik, Croatia
17. SATURN Projekt. URL: <https://www.saturn-projekt.de/> [accessed 2023-07-06]
18. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
19. Ahmadi N, Zoch M, Kelbert P, Noll R, Schaaf J, Wolfien M, et al. Methods used in the development of Common Data Models for health data – A Scoping Review Protocol. OSF. URL: <https://osf.io/rza84/> [accessed 2023-07-06]
20. Meystre SM, Lee S, Jung CY, Chevrier RD. Common data model for natural language processing based on two existing standard information models: CDA+GrAF. *J Biomed Inform* 2012 Aug;45(4):703-710 [FREE Full text] [doi: [10.1016/j.jbi.2011.11.018](https://doi.org/10.1016/j.jbi.2011.11.018)] [Medline: [22197801](https://pubmed.ncbi.nlm.nih.gov/22197801/)]
21. Shin S, You S, Roh J, Park Y, Park R. Genomic Common Data Model for Biomedical Data in Clinical Practice. *Stud Health Technol Inform* 2019 Aug 21;264:1843-1844 [doi: [10.3233/SHTI190676](https://doi.org/10.3233/SHTI190676)] [Medline: [31438371](https://pubmed.ncbi.nlm.nih.gov/31438371/)]
22. Jones D, Shao J, Wallis H, Johansen C, Hart K, Pasquali M, et al. Towards a newborn screening common data model: the Utah Newborn Screening Data Model. *Int J Neonatal Screen* 2021 Oct 27;7(4):70 [FREE Full text] [doi: [10.3390/ijns7040070](https://doi.org/10.3390/ijns7040070)] [Medline: [34842615](https://pubmed.ncbi.nlm.nih.gov/34842615/)]
23. Chapman D. Advanced search features of PubMed. *J Can Acad Child Adolesc Psychiatry* 2009 Feb;18(1):58-59 [FREE Full text] [Medline: [19270851](https://pubmed.ncbi.nlm.nih.gov/19270851/)]
24. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016 Dec 05;5(1):210 [FREE Full text] [doi: [10.1186/s13643-016-0384-4](https://doi.org/10.1186/s13643-016-0384-4)] [Medline: [27919275](https://pubmed.ncbi.nlm.nih.gov/27919275/)]
25. Niederberger M, Spranger J. Delphi technique in health sciences: a map. *Front Public Health* 2020;8:457 [FREE Full text] [doi: [10.3389/fpubh.2020.00457](https://doi.org/10.3389/fpubh.2020.00457)] [Medline: [33072683](https://pubmed.ncbi.nlm.nih.gov/33072683/)]
26. Ahmadi N. VisualisationWithPython. GitHub. URL: <https://github.com/NajiaAhmadi/VisualisationWithPython> [accessed 2023-07-06]
27. Meeuws S, Yue JK, Huijben JA, Nair N, Lingsma HF, Bell MJ, et al. Common Data Elements: Critical Assessment of Harmonization between Current Multi-Center Traumatic Brain Injury Studies. *J Neurotrauma* 2020 Jun 01;37(11):1283-1290 [FREE Full text] [doi: [10.1089/neu.2019.6867](https://doi.org/10.1089/neu.2019.6867)] [Medline: [32000562](https://pubmed.ncbi.nlm.nih.gov/32000562/)]
28. Vande Vyvere T, De La Rosa E, Wilms G, Nieboer D, Steyerberg E, Maas A, CENTER-TBI Participants Investigators. Prognostic validation of the NINDS common data elements for the radiologic reporting of acute traumatic brain injuries: A CENTER-TBI Study. *J Neurotrauma* 2020 Jun 01;37(11):1269-1282 [FREE Full text] [doi: [10.1089/neu.2019.6710](https://doi.org/10.1089/neu.2019.6710)] [Medline: [31813313](https://pubmed.ncbi.nlm.nih.gov/31813313/)]
29. Schiari V, Fowler E, Brandenburg JE, Levey E, McIntyre S, Sukal-Moulton T, et al. A common data language for clinical research studies: the National Institute of Neurological Disorders and Stroke and American Academy for Cerebral Palsy and Developmental Medicine Cerebral Palsy Common Data Elements Version 1.0 recommendations. *Dev Med Child Neurol* 2018 Oct 15;60(10):976-986 [FREE Full text] [doi: [10.1111/dmcn.13723](https://doi.org/10.1111/dmcn.13723)] [Medline: [29542813](https://pubmed.ncbi.nlm.nih.gov/29542813/)]
30. Hunt C, Michalak A, Ouchterlony D, Marshall S, Masanic C, Vaidyanath C, et al. Common Data Elements for Concussion in Tertiary Care: Phase One in Ontario. *Can J Neurol Sci* 2017 Nov 30;44(6):676-683 [doi: [10.1017/cjn.2017.222](https://doi.org/10.1017/cjn.2017.222)] [Medline: [29391082](https://pubmed.ncbi.nlm.nih.gov/29391082/)]
31. Mawji A, Li E, Chandna A, Kortz T, Akech S, Wiens MO, et al. Common data elements for predictors of pediatric sepsis: a framework to standardize data collection. *PLoS One* 2021;16(6):e0253051 [FREE Full text] [doi: [10.1371/journal.pone.0253051](https://doi.org/10.1371/journal.pone.0253051)] [Medline: [34111209](https://pubmed.ncbi.nlm.nih.gov/34111209/)]
32. de Oliveira Manoel AL, van der Jagt M, Amin-Hanjani S, Bambakidis NC, Brophy GM, Bulsara K, Unruptured AneurysmsSAH – CDE Project Investigators. Common data elements for unruptured intracranial aneurysms and aneurysmal subarachnoid hemorrhage: recommendations from the Working Group on Hospital Course and Acute Therapies-Proposal of a Multidisciplinary Research Group. *Neurocrit Care* 2019 Jun;30(Suppl 1):36-45 [doi: [10.1007/s12028-019-00726-3](https://doi.org/10.1007/s12028-019-00726-3)] [Medline: [31119687](https://pubmed.ncbi.nlm.nih.gov/31119687/)]
33. Le Gal G, Carrier M, Castellucci LA, Cuker A, Hansen J, Klok FA, ISTH CDE Task Force. Development and implementation of common data elements for venous thromboembolism research: on behalf of SSC Subcommittee on official Communication from the SSC of the ISTH. *J Thromb Haemost* 2021 Jan;19(1):297-303 [FREE Full text] [doi: [10.1111/jth.15138](https://doi.org/10.1111/jth.15138)] [Medline: [33405381](https://pubmed.ncbi.nlm.nih.gov/33405381/)]
34. Gagnon I, Friedman D, Beauchamp MH, Christie B, DeMatteo C, Macartney G, et al. The Canadian Pediatric Mild Traumatic Brain Injury Common Data Elements Project: Harmonizing Outcomes to Increase Understanding of Pediatric Concussion. *J Neurotrauma* 2018 Aug 15;35(16):1849-1857 [doi: [10.1089/neu.2018.5887](https://doi.org/10.1089/neu.2018.5887)] [Medline: [30074870](https://pubmed.ncbi.nlm.nih.gov/30074870/)]
35. Massey KA, Magee LA, Dale S, Claydon J, Morris TJ, von Dadelszen P, et al. A current landscape of provincial perinatal data collection in Canada. *J Obstet and Gynaecol Canada* 2009 Mar;31(3):236-246 [doi: [10.1016/s1701-2163\(16\)34122-6](https://doi.org/10.1016/s1701-2163(16)34122-6)]

36. Lau F, Downing M, Tayler C, Fassbender K, Lesperance M, Barnett J. Toward a population-based approach to end-of-life care surveillance in Canada: initial efforts and lessons. *J Palliat Care* 2018 Dec 19;29(1):13-21 [doi: [10.1177/082585971302900103](https://doi.org/10.1177/082585971302900103)]
37. Yang Y, Xu H, Qi B, Niu X, Li M, Zhao D. Stroke screening data modeling based on openEHR and NINDS Stroke CDE. 2020 Presented at: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 16-19, 2020; Seoul, Korea [doi: [10.1109/BIBM49941.2020.9313127](https://doi.org/10.1109/BIBM49941.2020.9313127)]
38. Biering-Sørensen F, Alai S, Anderson K, Charlifue S, Chen Y, DeVivo M, et al. Common data elements for spinal cord injury clinical research: a National Institute for Neurological Disorders and Stroke project. *Spinal Cord* 2015 Apr 10;53(4):265-277 [FREE Full text] [doi: [10.1038/sc.2014.246](https://doi.org/10.1038/sc.2014.246)] [Medline: [25665542](https://pubmed.ncbi.nlm.nih.gov/25665542/)]
39. Bernstein K. Reporting of drug allergies for use in a national decision support system. *Stud Health Technol Inform* 2014;205:68-72 [Medline: [25160147](https://pubmed.ncbi.nlm.nih.gov/25160147/)]
40. Thurin NH, Pajouheshnia R, Roberto G, Dodd C, Hyeraci G, Bartolini C, et al. From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding. *Clin Pharmacol Ther* 2022 Jan 26;111(1):321-331 [FREE Full text] [doi: [10.1002/cpt.2476](https://doi.org/10.1002/cpt.2476)] [Medline: [34826340](https://pubmed.ncbi.nlm.nih.gov/34826340/)]
41. Holz C, Kessler T, Dugas M, Varghese J. Core Data Elements in Acute Myeloid Leukemia: A Unified Medical Language System-Based Semantic Analysis and Experts' Review. *JMIR Med Inform* 2019 Aug 12;7(3):e13554 [FREE Full text] [doi: [10.2196/13554](https://doi.org/10.2196/13554)] [Medline: [31407666](https://pubmed.ncbi.nlm.nih.gov/31407666/)]
42. Hackenberg KAM, Algra A, Al-Shahi Salman R, Frösen J, Hasan D, Juvela S, Unruptured AneurysmsSAH CDE Project Investigators. Definition and prioritization of data elements for cohort studies and clinical trials on patients with unruptured intracranial aneurysms: proposal of a multidisciplinary research group. *Neurocrit Care* 2019 Jun;30(Suppl 1):87-101 [doi: [10.1007/s12028-019-00729-0](https://doi.org/10.1007/s12028-019-00729-0)] [Medline: [31102238](https://pubmed.ncbi.nlm.nih.gov/31102238/)]
43. von Martial S, Brix TJ, Klotz L, Neuhaus P, Berger K, Warnke C, et al. EMR-integrated minimal core dataset for routine health care and multiple research settings: a case study for neuroinflammatory demyelinating diseases. *PLoS One* 2019;14(10):e0223886 [FREE Full text] [doi: [10.1371/journal.pone.0223886](https://doi.org/10.1371/journal.pone.0223886)] [Medline: [31613917](https://pubmed.ncbi.nlm.nih.gov/31613917/)]
44. Berger A, Rustemeier A, Göbel J, Kadioglu D, Britz V, Schubert K, et al. How to design a registry for undiagnosed patients in the framework of rare disease diagnosis: suggestions on software, data set and coding system. *Orphanet J Rare Dis* 2021 May 01;16(1):198 [FREE Full text] [doi: [10.1186/s13023-021-01831-3](https://doi.org/10.1186/s13023-021-01831-3)] [Medline: [33933089](https://pubmed.ncbi.nlm.nih.gov/33933089/)]
45. Varghese J, Holz C, Neuhaus P, Bernardi M, Boehm A, Ganser A, et al. Key Data Elements in Myeloid Leukemia. *Stud Health Technol Inform* 2016;228:282-286 [Medline: [27577388](https://pubmed.ncbi.nlm.nih.gov/27577388/)]
46. Schaaf J, Chalmers J, Omran H, Pennekamp P, Sitbon O, Wagner T, et al. The Registry Data Warehouse in the European Reference Network for Rare Respiratory Diseases - background, conception and implementation. *Stud Health Technol Inform* 2021 May 24;278:41-48 [doi: [10.3233/SHTI210049](https://doi.org/10.3233/SHTI210049)] [Medline: [34042874](https://pubmed.ncbi.nlm.nih.gov/34042874/)]
47. Moratilla JM, Alonso-Calvo R, Molina-Vaquero G, Paraiso-Medina S, Perez-Rey D, Maojo V. A data model based on semantically enhanced HL7 RIM for sharing patient data of breast cancer clinical trials. *Stud Health Technol Inform* 2013;192:971 [Medline: [23920745](https://pubmed.ncbi.nlm.nih.gov/23920745/)]
48. Rho MJ, Kim SR, Park SH, Jang KS, Park BJ, Hong JY, et al. Common data model for decision support system of adverse drug reaction to extract knowledge from multi-center database. *Inf Technol Manag* 2015 Jul 3;17(1):57-66 [doi: [10.1007/s10799-015-0240-6](https://doi.org/10.1007/s10799-015-0240-6)]
49. Cohen JM, Cesta CE, Kjerpeseth L, Leinonen MK, Hálfðánarson Ó, Karlstad Ø, et al. A common data model for harmonization in the Nordic Pregnancy Drug Safety Studies (NorPreSS). *Nor J Epidemiol* 2021 Aug 16;29(1-2):117-123 [doi: [10.5324/nje.v29i1-2.4053](https://doi.org/10.5324/nje.v29i1-2.4053)]
50. Krüger A, Lockey D, Kurolo J, Di Bartolomeo S, Castrén M, Mikkelsen S, et al. A consensus-based template for documenting and reporting in physician-staffed pre-hospital services. *Scand J Trauma Resusc Emerg Med* 2011 Nov 23;19(1):71 [FREE Full text] [doi: [10.1186/1757-7241-19-71](https://doi.org/10.1186/1757-7241-19-71)] [Medline: [22107787](https://pubmed.ncbi.nlm.nih.gov/22107787/)]
51. Ringdal KG, Lossius HM, SCANTEM ad hoc group on Scandinavian MTOSTrauma Registry. Feasibility of comparing core data from existing trauma registries in scandinavia. Reaching for a Scandinavian major trauma outcome study (MTOS). *Scand J Surg* 2007 Jun 24;96(4):325-331 [doi: [10.1177/145749690709600412](https://doi.org/10.1177/145749690709600412)] [Medline: [18265862](https://pubmed.ncbi.nlm.nih.gov/18265862/)]
52. Jaboyedoff M, Rakic M, Bachmann S, Berger C, Diezi M, Fuchs O, et al. SwissPedData: standardising hospital records for the benefit of paediatric research. *Swiss Med Wkly* 2021 Dec 20;151:w30069 [FREE Full text] [doi: [10.4414/smw.2021.w30069](https://doi.org/10.4414/smw.2021.w30069)] [Medline: [34964587](https://pubmed.ncbi.nlm.nih.gov/34964587/)]
53. Chen S, Hsu C, Huang C. Annotating Taiwan Cancer Registry to caDSR for International Interoperability. In: Zhang Y, editor. *Future Communication, Computing, Control and Management. Lecture Notes in Electrical Engineering*, vol 141. Berlin, Heidelberg: Springer; 2012:257-263
54. Spronk P, Begum H, Vishwanath S, Crosbie A, Earnest A, Elder E, et al. Toward International Harmonization of Breast Implant Registries: International Collaboration of Breast Registry Activities Global Common Data Set. *Plast Reconstr Surg* 2020 Aug;146(2):255-267 [doi: [10.1097/PRS.0000000000006969](https://doi.org/10.1097/PRS.0000000000006969)] [Medline: [32740572](https://pubmed.ncbi.nlm.nih.gov/32740572/)]
55. Mowforth O, Khan D, Wong M, Pickering G, Dean L, Magee J, AO Spine RECODE-DCM Steering CommitteeAO Spine RECODE-DCM Consortium. Gathering Global Perspectives to Establish the Research Priorities and Minimum Data Sets for Degenerative Cervical Myelopathy: Sampling Strategy of the First Round Consensus Surveys of AO Spine

- RECODE-DCM. *Global Spine J* 2022 Feb;12(1\_suppl):8S-18S [FREE Full text] [doi: [10.1177/21925682211047546](https://doi.org/10.1177/21925682211047546)] [Medline: [34879754](https://pubmed.ncbi.nlm.nih.gov/34879754/)]
56. Currie AC, Cahill R, Delaney CP, Faiz OD, Kennedy RH. International expert consensus on endpoints for full-thickness laparoendoscopic colonic excision. *Surg Endosc* 2016 Apr 27;30(4):1497-1502 [doi: [10.1007/s00464-015-4362-z](https://doi.org/10.1007/s00464-015-4362-z)] [Medline: [26123345](https://pubmed.ncbi.nlm.nih.gov/26123345/)]
57. Davies BM, Khan DZ, Mowforth OD, McNair AGK, Gronlund T, Koliass AG, et al. RE-CODE DCM (REsearch Objectives and Common Data Elements for Degenerative Cervical Myelopathy): A Consensus Process to Improve Research Efficiency in DCM, Through Establishment of a Standardized Dataset for Clinical Research and the Definition of the Research Priorities. *Global Spine J* 2019 May 08;9(1 Suppl):65S-76S [FREE Full text] [doi: [10.1177/2192568219832855](https://doi.org/10.1177/2192568219832855)] [Medline: [31157148](https://pubmed.ncbi.nlm.nih.gov/31157148/)]
58. Cohen D, Sullivan K, McNeil R, Gulf War Illness Common Data Elements Working Group., Symptoms Assessment Working Group., McNeil R, Systems Assessment Working Group., et al. A common language for Gulf War Illness (GWI) research studies: GWI common data elements. *Life Sci* 2022 Feb 01;290:119818 [FREE Full text] [doi: [10.1016/j.lfs.2021.119818](https://doi.org/10.1016/j.lfs.2021.119818)] [Medline: [34352259](https://pubmed.ncbi.nlm.nih.gov/34352259/)]
59. Karaa A, Rahman S, Lombès A, Yu-Wai-Man P, Sheikh MK, Alai-Hansen S, Mito Working Group Member Participants. Common data elements for clinical research in mitochondrial disease: a National Institute for Neurological Disorders and Stroke project. *J Inher Metab Dis* 2017 May 16;40(3):403-414 [FREE Full text] [doi: [10.1007/s10545-017-0035-5](https://doi.org/10.1007/s10545-017-0035-5)] [Medline: [28303425](https://pubmed.ncbi.nlm.nih.gov/28303425/)]
60. Suarez JI, Sheikh MK, Macdonald RL, Amin-Hanjani S, Brown RD, de Oliveira Manoel AL, Unruptured Intracranial AneurysmsSAH CDE Project Investigators. Common Data Elements for Unruptured Intracranial Aneurysms and Subarachnoid Hemorrhage Clinical Research: A National Institute for Neurological Disorders and Stroke and National Library of Medicine Project. *Neurocrit Care* 2019 Jun 13;30(Suppl 1):4-19 [doi: [10.1007/s12028-019-00723-6](https://doi.org/10.1007/s12028-019-00723-6)] [Medline: [31087257](https://pubmed.ncbi.nlm.nih.gov/31087257/)]
61. Grinspan ZM, Patel AD, Shellhaas RA, Berg AT, Axeen ET, Bolton J, Pediatric Epilepsy Learning Healthcare System. Design and implementation of electronic health record common data elements for pediatric epilepsy: Foundations for a learning health care system. *Epilepsia* 2021 Jan 24;62(1):198-216 [FREE Full text] [doi: [10.1111/epi.16733](https://doi.org/10.1111/epi.16733)] [Medline: [33368200](https://pubmed.ncbi.nlm.nih.gov/33368200/)]
62. Ward S, Flori H, Bennett T, Sapru A, Mourani P, Thomas N, et al. Design and Rationale for Common Data Elements for Clinical Research in Pediatric Critical Care Medicine. *Pediatr Crit Care Med* 2020 Nov;21(11):e1038-e1041 [FREE Full text] [doi: [10.1097/PCC.0000000000002455](https://doi.org/10.1097/PCC.0000000000002455)] [Medline: [32639472](https://pubmed.ncbi.nlm.nih.gov/32639472/)]
63. Luciano MG, Batzdorf U, Kula RW, Rocque BG, Maher CO, Heiss J, Chiari I Malformation Common Data Element Working Group. Development of common data elements for use in chiari malformation type I clinical research: an NIH/NINDS project. *Neurosurgery* 2019 Dec 01;85(6):854-860 [FREE Full text] [doi: [10.1093/neuros/nyy475](https://doi.org/10.1093/neuros/nyy475)] [Medline: [30690581](https://pubmed.ncbi.nlm.nih.gov/30690581/)]
64. Mayer CS, Williams N, Huser V. Identification of common data elements from pivotal FDA trials. *AMIA Annu Symp Proc* 2020;2020:813-822 [FREE Full text] [Medline: [33936456](https://pubmed.ncbi.nlm.nih.gov/33936456/)]
65. Broglio SP, Kontos AP, Levin H, Schneider K, Wilde EA, Cantu RC, et al. National Institute of Neurological Disorders and Stroke and Department of Defense Sport-Related Concussion Common Data Elements Version 1.0 recommendations. *J Neurotrauma* 2018 Dec 01;35(23):2776-2783 [FREE Full text] [doi: [10.1089/neu.2018.5643](https://doi.org/10.1089/neu.2018.5643)] [Medline: [29717643](https://pubmed.ncbi.nlm.nih.gov/29717643/)]
66. Wandner L, Domenichiello A, Beierlein J, Pogorzala L, Aquino G, Siddons A, NIH Pain Consortium Institute and Center Representatives. NIH's Helping to End Addiction Long-term Initiative (NIH HEAL Initiative) clinical pain management common data element program. *J Pain* 2022 Mar;23(3):370-378 [FREE Full text] [doi: [10.1016/j.jpain.2021.08.005](https://doi.org/10.1016/j.jpain.2021.08.005)] [Medline: [34508905](https://pubmed.ncbi.nlm.nih.gov/34508905/)]
67. Carroll A, Vogel LC, Zebracki K, Noonan VK, Biering-Sørensen F, Mulcahey MJ. Relevance of the international spinal cord injury basic data sets to youth: an Inter-Professional review with recommendations. *Spinal Cord* 2017 Sep 28;55(9):875-881 [doi: [10.1038/sc.2017.14](https://doi.org/10.1038/sc.2017.14)] [Medline: [28244501](https://pubmed.ncbi.nlm.nih.gov/28244501/)]
68. Albino J, Tiwari T, Gansky S, Henshaw M, Barker J, Brega A, Early Childhood Caries Collaborating Centers. The basic research factors questionnaire for studying early childhood caries. *BMC Oral Health* 2017 May 19;17(1):83 [FREE Full text] [doi: [10.1186/s12903-017-0374-5](https://doi.org/10.1186/s12903-017-0374-5)] [Medline: [28526003](https://pubmed.ncbi.nlm.nih.gov/28526003/)]
69. Fisher J, Krisa L, Middleton D, Leiby B, Harrop J, Shah L, et al. Validation of the National Institute of Neurological Disorders and Stroke Spinal Cord Injury MRI Common Data Elements Instrument. *AJNR Am J Neuroradiol* 2021 Feb 11;42(4):787-793 [doi: [10.3174/ajnr.a7000](https://doi.org/10.3174/ajnr.a7000)]
70. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012 Jan 01;19(1):54-60 [FREE Full text] [doi: [10.1136/amiajnl-2011-000376](https://doi.org/10.1136/amiajnl-2011-000376)] [Medline: [22037893](https://pubmed.ncbi.nlm.nih.gov/22037893/)]
71. Gardner D, Knuth KH, Abato M, Erde SM, White T, DeBellis R, et al. Common data model for neuroscience data and data model exchange. *J Am Med Inform Assoc* 2001 Jan 01;8(1):17-33 [FREE Full text] [doi: [10.1136/jamia.2001.0080017](https://doi.org/10.1136/jamia.2001.0080017)] [Medline: [11141510](https://pubmed.ncbi.nlm.nih.gov/11141510/)]
72. Patel AA, Kajdacsy-Balla A, Berman JJ, Bosland M, Datta MW, Dhir R, et al. The development of common data elements for a multi-institute prostate cancer tissue bank: the Cooperative Prostate Cancer Tissue Resource (CPCTR) experience. *BMC Cancer* 2005 Aug 21;5(1):108 [FREE Full text] [doi: [10.1186/1471-2407-5-108](https://doi.org/10.1186/1471-2407-5-108)] [Medline: [16111498](https://pubmed.ncbi.nlm.nih.gov/16111498/)]

73. Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. *J Am Med Inform Assoc* 2010 Nov 01;17(6):652-662 [[FREE Full text](#)] [doi: [10.1136/jamia.2009.002477](https://doi.org/10.1136/jamia.2009.002477)] [Medline: [20962127](https://pubmed.ncbi.nlm.nih.gov/20962127/)]
74. Loring D, Lowenstein D, Barbaro N, Fureman B, Odenkirchen J, Jacobs M, et al. Common data elements in epilepsy research: development and implementation of the NINDS epilepsy CDE project. *Epilepsia* 2011 Jun;52(6):1186-1191 [[FREE Full text](#)] [doi: [10.1111/j.1528-1167.2011.03018.x](https://doi.org/10.1111/j.1528-1167.2011.03018.x)] [Medline: [21426327](https://pubmed.ncbi.nlm.nih.gov/21426327/)]
75. McCauley SR, Wilde EA, Anderson VA, Bedell G, Beers SR, Campbell TF, Pediatric TBI Outcomes Workgroup. Recommendations for the use of common outcome measures in pediatric traumatic brain injury research. *J Neurotrauma* 2012 Mar 01;29(4):678-705 [[FREE Full text](#)] [doi: [10.1089/neu.2011.1838](https://doi.org/10.1089/neu.2011.1838)] [Medline: [21644810](https://pubmed.ncbi.nlm.nih.gov/21644810/)]
76. Gerring JP, Wade S. The essential role of psychosocial risk and protective factors in pediatric traumatic brain injury research. *J Neurotrauma* 2012 Mar 01;29(4):621-628 [[FREE Full text](#)] [doi: [10.1089/neu.2011.2234](https://doi.org/10.1089/neu.2011.2234)] [Medline: [22091875](https://pubmed.ncbi.nlm.nih.gov/22091875/)]
77. Saver JL, Warach S, Janis S, Odenkirchen J, Becker K, Benavente O, et al. Standardizing the Structure of Stroke Clinical and Epidemiologic Research Data. *Stroke* 2012 Apr;43(4):967-973 [doi: [10.1161/strokeaha.111.634352](https://doi.org/10.1161/strokeaha.111.634352)]
78. Dastgir J, Rutkowski A, Alvarez R, Cossette S, Yan K, Hoffmann R, et al. Common Data Elements for Muscle Biopsy Reporting. *Arch Pathol Lab Med* 2016 Jan;140(1):51-65 [[FREE Full text](#)] [doi: [10.5858/arpa.2014-0453-OA](https://doi.org/10.5858/arpa.2014-0453-OA)] [Medline: [26132600](https://pubmed.ncbi.nlm.nih.gov/26132600/)]
79. Perrone RD, Neville J, Chapman AB, Gitomer BY, Miskulin DC, Torres VE, et al. Therapeutic area data standards for autosomal dominant polycystic kidney disease: a report from the Polycystic Kidney Disease Outcomes Consortium (PKDOC). *Am J Kidney Dis* 2015 Oct;66(4):583-590 [doi: [10.1053/j.ajkd.2015.04.044](https://doi.org/10.1053/j.ajkd.2015.04.044)] [Medline: [26088508](https://pubmed.ncbi.nlm.nih.gov/26088508/)]
80. Yue JK, Vassar MJ, Lingsma HF, Cooper SR, Okonkwo DO, Valadka AB, TRACK-TBI Investigators. Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. *J Neurotrauma* 2013 Nov 15;30(22):1831-1844 [[FREE Full text](#)] [doi: [10.1089/neu.2013.2970](https://doi.org/10.1089/neu.2013.2970)] [Medline: [23815563](https://pubmed.ncbi.nlm.nih.gov/23815563/)]
81. Hicks K, Tchong J, Bozkurt B, Chaitman B, Cutlip D, Farb A, American College of Cardiology, American Heart Association. 2014 ACC/AHA Key Data Elements and Definitions for Cardiovascular Endpoint Events in Clinical Trials: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Cardiovascular Endpoints Data Standards). *Circulation* 2015 Jul 28;132(4):302-361 [[FREE Full text](#)] [doi: [10.1161/CIR.000000000000156](https://doi.org/10.1161/CIR.000000000000156)] [Medline: [25547519](https://pubmed.ncbi.nlm.nih.gov/25547519/)]
82. Moore SM, Schiffman R, Waldrop-Valverde D, Redeker NS, McCloskey DJ, Kim MT, et al. Recommendations of common data elements to advance the science of self-management of chronic conditions. *J Nurs Scholarsh* 2016 Sep;48(5):437-447 [[FREE Full text](#)] [doi: [10.1111/jnu.12233](https://doi.org/10.1111/jnu.12233)] [Medline: [27486851](https://pubmed.ncbi.nlm.nih.gov/27486851/)]
83. Rubinstein YR, McInnes P. NIH/NCATS/GRDR® Common Data Elements: a leading force for standardized data collection. *Contemp Clin Trials* 2015 May;42:78-80 [[FREE Full text](#)] [doi: [10.1016/j.cct.2015.03.003](https://doi.org/10.1016/j.cct.2015.03.003)] [Medline: [25797358](https://pubmed.ncbi.nlm.nih.gov/25797358/)]
84. De Vito Dabbs A, Myers B, Mc Curry K, Dunbar-Jacob J, Hawkins R, Begey A, et al. User-centered design and interactive health technologies for patients. *Comput Inform Nurs* 2009;27(3):175-183 [[FREE Full text](#)] [doi: [10.1097/NCN.0b013e31819f7c7c](https://doi.org/10.1097/NCN.0b013e31819f7c7c)] [Medline: [19411947](https://pubmed.ncbi.nlm.nih.gov/19411947/)]

## Abbreviations

**CDE:** common data element

**CDM:** common data model

**MeSH:** Medical Subject Headings

**OMOP:** Observational Medical Outcomes Partnership

**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

**SATURN:** Smartes Arztportal für Betroffene mit unklarer Erkrankung (Smart physicians' platform for patients with unclear diseases)

*Edited by C Lovis; submitted 16.12.22; peer-reviewed by A Lamer, X Ma; comments to author 31.01.23; revised version received 09.03.23; accepted 08.06.23; published 03.08.23*

*Please cite as:*

Ahmadi N, Zoch M, Kelbert P, Noll R, Schaaf J, Wolfien M, Sedlmayr M

*Methods Used in the Development of Common Data Models for Health Data: Scoping Review*

*JMIR Med Inform* 2023;11:e45116

URL: <https://medinform.jmir.org/2023/1/e45116>

doi: [10.2196/45116](https://doi.org/10.2196/45116)

PMID:



©Najia Ahmadi, Michele Zoch, Patricia Kelbert, Richard Noll, Jannik Schaaf, Markus Wolfien, Martin Sedlmayr. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 03.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.