

Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?

A Validation Study Based on Prescribing Codeine in Children

Gianmario Candore^{1,*}, Karin Hedenmalm¹, Jim Slattery², Alison Cave², Xavier Kurz² and Peter Arlett²

Exploring and combining results from more than one real-world data (RWD) source might be necessary in order to explore variability and demonstrate generalizability of the results or for regulatory requirements. However, the heterogeneous nature of RWD poses challenges when working with more than one source, some of which can be solved by analyzing databases converted into a common data model (CDM). The main objective of the study was to evaluate the implementation of the Observational Medical Outcome Partnership (OMOP) CDM on IQVIA Medical Research Data (IMRD)-UK data. A drug utilization study describing the prescribing of codeine for pain in children was used as a case study to be replicated in IMRD-UK and its corresponding OMOP CDM transformation. Differences between IMRD-UK source and OMOP CDM were identified and investigated. In IMRD-UK updated to May 2017, results were similar between source and transformed data with few discrepancies. These were the result of different conventions applied during the transformation regarding the date of birth for children younger than 15 years and the start of the observation period, and of a misclassification of two drug treatments. After the initial analysis and feedback provided, a rerun of the analysis in IMRD-UK updated to September 2018 showed almost identical results for all the measures analyzed. For this study, the conversion to OMOP CDM was adequate. Although some decisions and mapping could be improved, these impacted on the absolute results but not on the study inferences. This validation study supports six recommendations for good practice in transforming to CDMs.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Studies have been conducted on aspects of conversion of databases into an Observational Medical Outcome Partnership (OMOP) common data model (CDM) that supported, in general, the usefulness of the model. However, different factors have been identified that can influence results, and these varied between databases.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ The study explores loss of information and inaccuracy resulting from the conversion of the IQVIA Medical Research Data (IMRD)-UK data into OMOP CDM by replicating on both the source and the transformed data a study of codeine prescribing in children.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ For this study, the conversion of IMRD-UK to OMOP CDM was adequate. Although some decisions and mapping could be improved, these impacted on the absolute study results, but not on the study inferences. This study supports six recommendations for good practice in transforming to CDMs.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ The proposed recommendations will support a more routine use of a CDM as a basis for regulatory decision making and enable a more efficient and uniform conduct of multidatabase studies.

¹Business Data Department, European Medicines Agency, Amsterdam, The Netherlands; ²Pharmacovigilance and Epidemiology Department, European Medicines Agency, Amsterdam, The Netherlands. *Correspondence: Gianmario Candore (gianmario.candore@ema.europa.eu)

Received October 17, 2019; accepted December 17, 2019. doi:10.1002/cpt.1785

Real-world data (RWD) is being used for evidence generation to support regulatory decision making. Although current use is most frequent in the postauthorization setting, opportunities can arise across the entire product life cycle.¹

One challenge of working with RWD is that the quantity of information related to any product or disease is not within the control of the researcher. For some studies a single data source can resolve a question completely, for others it may be necessary to combine several different sources. Even if one source seems to provide an accurate answer to a clinical question, concerns may remain that the number of patients exposed is not sufficient or that results are not generalizable due to substantive differences in clinical practice or variation in other factors across health systems. Thus, even with apparently straightforward questions, it may be wise to explore more than one data source, preferably arising from diverse healthcare systems.² These or similar considerations are often reflected in regulatory requirements, for example, regarding the determination of orphan status for a drug^{3,4} or the setting of postauthorization studies. This is also demonstrated in the way the European Medicines Agency (EMA) supports medicines evaluation by having in-house access to different primary healthcare databases and, when needed, considering additional data sources.

Naturally, working with more than one dataset with different data structures, variables, and coding practices is a challenge and this is particularly relevant in the European landscape, rich with RWD but of heterogeneous nature.^{1,5,6} Moreover, addressing the same research question in separate studies on different datasets will encounter many of the same challenges, resulting in duplication of efforts to address them with consequences on time and resources needed. It also raises concerns that different researchers may adopt different practices in addressing them making it challenging to determine if heterogeneity in the results is intrinsic to the population studied or methodological. This has led researchers to propose that some alignment of the datasets prior to and independent of any particular study would be highly desirable and would enable a more efficient and uniform implementation of a single study protocol across diverse databases.⁵⁻⁸

The fundamental prerequisite to be able to use any common data model (CDM) is that it faithfully represents the source data. Two main risks need to be assessed and monitored over time:

- **Completeness:** Is there any information loss that can be caused by (i) the CDM data structure not being able to accommodate all the different types of variables present in the source database and needed for the analysis; (ii) the existence of terms in the source dictionary that have no counterparts and cannot be mapped to the CDM standard dictionaries, if these are used; or (iii) variables or free text fields containing relevant information in the source database that are not included in the transformation to the CDM.
- **Accuracy:** Is there any difference in the representation of the data in the CDM vs. the source data that can be caused by (i) errors during the extraction, transformation, and load (ETL) or in the mapping to the CDM dictionaries; (ii) rules used in the ETL that do not reflect the conventions or methodologies

usually applied by the data owner or analysts; or (iii) a mapping to a term in the CDM standard dictionary that is less granular.

The Observational Medical Outcome Partnership (OMOP) was a public-private partnership project to develop new methods for observational research.⁹ The OMOP CDM was developed as part of that project and has since been adopted and maintained by the international collaborative Observational Health Data Sciences and Informatics (OHDSI).^{9,10} What distinguishes the OMOP CDM from other CDMs is that converted databases not only share a common data structure¹¹ but also a common terminology thanks to the mapping of medical constructs in the source database to common dictionaries.^{7,12}

Multiple studies have been conducted on aspects of conversion of databases into OMOP CDM in the United States, Europe, and Asia.^{5,6,13-21} European databases included the Austrian health claims data,^{15,16} the German University hospital data,¹⁸ and the Clinical Practice Research Datalink (CPRD) and IQVIA Medical Research Data incorporating data from THIN, A Cegedim Database (IMRD-UK, formerly known as THIN) primary care databases in the United Kingdom.^{5,13} The studies have supported, in general, the usefulness of the model. However, incomplete mapping of codes, codes that may have to be mapped on a less detailed level than in the original database, as well as differences in underlying data models have been identified as factors that can influence results.²² For IMRD-UK, in particular, mapping of laboratory, physical examination, and lifestyle data were also not available in the Zhou *et al.* study,⁵ leading to the conclusions that the OMOP CDM was of limited use for quality epidemiological analyses.

It has been suggested that conversion of a database into OMOP CDM should be considered as an iterative process as improvements in mapping and CDM versions occur over time.²² Hence, continued assessment has been recommended and, considering also that the previous study in IMRD-UK was done a few years ago and used an older specification of the OMOP CDM (version 2), a new study can provide updated evidence regarding whether limitations previously encountered have subsequently been overcome.

In June 2013, risk minimization measures (RMMs) for codeine were introduced in the European Union for the treatment of pain in children.²³⁻²⁵ Following the introduction of the above measures, a collaborative study²⁶ of the impact of the RMM on the prescribing of codeine in children was conducted on multiple databases: IMRD-France and IMRD-Germany,^{26,27} BIFAP (Spain),²⁸ and CPRD GOLD (UK).²⁹

This current study aimed to replicate the Hedenmalm *et al.*²⁶ study but with the main objective to compare the prescribing of codeine for the treatment of pain in children in the OMOP CDM converted database against results in the original IMRD-UK source database. The specific objective was to explore any potential loss of information and inaccuracy resulting from the conversion and, if so, whether these changes would have impacted on the interpretation of study results and lead to a different conclusion. The wider aims also included:

- To provide examples of discrepancies that can focus attention on potential systematic errors in the data transformation procedure, which then may suggest routine quality assurance checks.
- To test the iterative improvement of the data transformation process by studying two different versions of the OMOP CDM converted IMRD-UK database.
- To provide recommendations for good practice in the transformation process.
- To assess the potential utility of data transformed into OMOP CDM for regulatory decision making.

METHODS

Setting

The IMRD-UK database contains electronic primary care medical records extracted from over 700 general practices across the United Kingdom covering ~ 6% of the UK population. Data are representative of the UK population in terms of age, deprivation, and geographic distribution,³⁰ and patients are linked via an anonymous patient ID number allowing them to be followed longitudinally over time. In the United Kingdom, patients are required to register with a general practitioner (GP) for their primary health care.

Data sources

IQVIA provided both the original IMRD-UK source data and the OMOP CDM converted database. The May 2017 version (1705) was used for the initial comparison between OMOP CDM and source data. After the initial analysis and feedback about the results, the analysis was rerun on the September 2018 version (1809).

The OMOP CDM version used in both transformations was 5.2.

Terminologies

In IMRD-UK, diagnoses, symptoms, procedures, and other relevant health information are recorded using the Read Code clinical classification system, drug prescribing using Gemscript codes.

OMOP CDM uses SNOMED as the standard dictionary for medical conditions, and RxNorm/RxNorm Extension as the standard dictionary for medicines.^{14,31}

Study design and population

This was a retrospective study of pediatric patients from up-to-standard practices. The study population included patients with an age < 18 years. Each patient's observation period began at the latest of the patient's registration date, the acceptable mortality recording date of the practice, the Vision date (the date when the practice started using the vision practice management software to record consultations) or January 1, 2010; and ended at the earliest of the date of transfer out of the practice, the 18th birthday, the date of last practice data collection, or December 31, 2016/June 30, 2018, according to the IMRD-UK version used. Children 1 year or older were required to have been registered at the practice for 1 year before study entry. Children below the age of 1 year were required to have been registered for at least the number of days between their birth date and the registration date. For example, a child who was registered at an age of 3 months could be included at the earliest at an age of 6 months.

Variables

Products were identified that were considered to capture the use of codeine for the treatment of pain and grouped in the following categories: (i) liquid oral formulation of plain (single substance) codeine, (ii) solid oral formulation of plain codeine, (iii) codeine in combination with analgesics/nonsteroidal anti-inflammatory drugs, and (iv) codeine in combination with an analgesic and an antihistamine.

For a description of the procedures for selecting Gemscript codes and RxNorm concept names for codeine-containing products and use of other analgesics, please see the **Supplementary Text**. The methodology and resulting codes used to identify the Read and SNOMED terms for tonsillectomy or adenoidectomy (TA) and obstructive sleep apnea (OSA) are provided in the **Tables S1** and **S2**.

Measurements

Prevalence was calculated as the proportion between the number of children with a prescription for a selected codeine product during the time period (year or 6 months) and the number of children in the study population at the middle of the corresponding time period. In children with a first record of TA, the proportion with a codeine prescription within a time period of up to ± 30 days of the TA, in the presence or not of OSA, was calculated. Finally, duration of treatment and the proportion of children with a prescription of another analgesic within 90 days prior to the date of the codeine prescription were calculated.

Duration of treatment

In source IMRD-UK, "prescription duration" is not a required field, and only 4.5% of drugs prescribed were recorded with a duration value. Therefore, the main method recommended by the data provider to calculate duration is to divide the drug quantity prescribed (99.4% of drugs prescribed had a valid quantity value) by a daily dosage calculated from free text (55.6% of drugs prescribed had a value classified as valid; invalid values are primarily due to prescriptions with instructions to take "as needed" or "as directed").³² When it was not possible to calculate the duration following the approach recommended, no imputation was performed, and drug duration was considered as missing.

In IMRD-UK OMOP CDM, in the table "drug_exposure," the duration is provided using the information in the field "prescription duration" of the source data, not following the recommended approach by the data provider. However, both the fields reporting the drug quantity prescribed and the free text with the daily dosage are reported in the OMOP CDM; this allowed calculating the duration as in IMRD-UK source.

Analysis and statistical methods

All of the analyses were planned and performed by researchers independently of IQVIA. Results were analyzed descriptively. Differences between OMOP CDM and IMRD-UK source data were identified and, if they were found to have appreciable impact on the study results, further investigated.

Analyses were performed using SAS Enterprise Guide version 7.13 statistical software. The same analyst wrote the programming code used in both source and transformed data, using a structure as similar as possible, and implementing the same conventions when interpreting the protocol.

The relevant medical condition codes and drug codes were identified by the same researcher separately in the source and OMOP CDM dictionaries in order to be able to compare the entire study conduct.

As in the original codeine collaborative study, joinpoint regression analysis with log-linear model^{33–35} was used to evaluate statistically significant changes in prescribing trend.

RESULTS

IMRD-UK version 1705

Study population. Using OMOP CDM, slightly fewer children were included in the study, 1,725,353 vs. 1,783,223, a difference of 3.2% compared with the IMRD-UK source data (**Figure 1**). This difference was mainly due to how the date of birth was assigned in OMOP CDM (2.9%), where patients younger than



Figure 1 Study population of children 0–17 years in IQVIA Medical Research Data (IMRD)-UK source vs. IMRD-UK Observational Medical Outcome Partnership common data model (CDM; database version 1705). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

15 years at the time of the last data collection for the database were assigned a birth date in January instead of their actual month of birth. This meant that part of the population was miscoded as older and, in patients younger than 1 year, this increased the length of the required observation period before entering the cohort. In the event that children younger than 1 year were only registered in the database for a few months, this increase resulted in them being unable to satisfy the condition of entry for the CDM version.

A second reason was related to the methodology used to calculate the 1-year observation period before entering the cohort (0.3% of the difference; **Figure S1**). In the OMOP CDM, created to accommodate different databases with different rules to calculate the starting date, there is only one value for the start date of the patient observation period. When transforming IMRD-UK, this is the result of choosing the latest of the registration dates, the acceptable mortality recording date of the practice, and the Vision date. The 1-year observation period could then only be calculated from the latest of these dates. In the source database, however, all the different dates are available and, as in our normal practice and used in other studies,³⁶ the 1-year observation period was calculated from the registration date.

Codeine prescriptions. The number of children with a codeine prescription during the study period was smaller when using OMOP CDM compared with IMRD-UK source data, 55,480 vs. 64,226, respectively, a difference of 13.6% (**Figure 2**). A small proportion (2.1%) of this difference was a result of having fewer patients in the study population, whereas the remainder (11.5%) was a result of two codeine treatments being miscoded as devices.

The discrepancies noted were reflected in the prevalence calculations with the OMOP CDM showing a lower prevalence throughout the whole period in both age groups (**Figure 3**). However, results from both databases confirmed a slight decrease in overall prescribing of codeine between the start and end of the study period, a decrease that started before the introduction of the regulatory action. The decrease was more pronounced in children under the age of 12 years, with little evidence of a decrease in children 12–17 years.

Results by formulation (**Figure S2**) showed that the prevalence estimates almost overlapped between OMOP CDM and source data in the groups unaffected by the miscoded codeine treatment. In contrast, for the combinations with the other analgesics and antihistamines group, which contained the miscoded codeine treatments, the impact on the prevalence was marked.



Figure 2 Exposure to codeine for pain in children 0–17 years in IQVIA Medical Research Data (IMRD)-UK source vs. IMRD-UK Observational Medical Outcome Partnership common data model (CDM; database version 1705). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

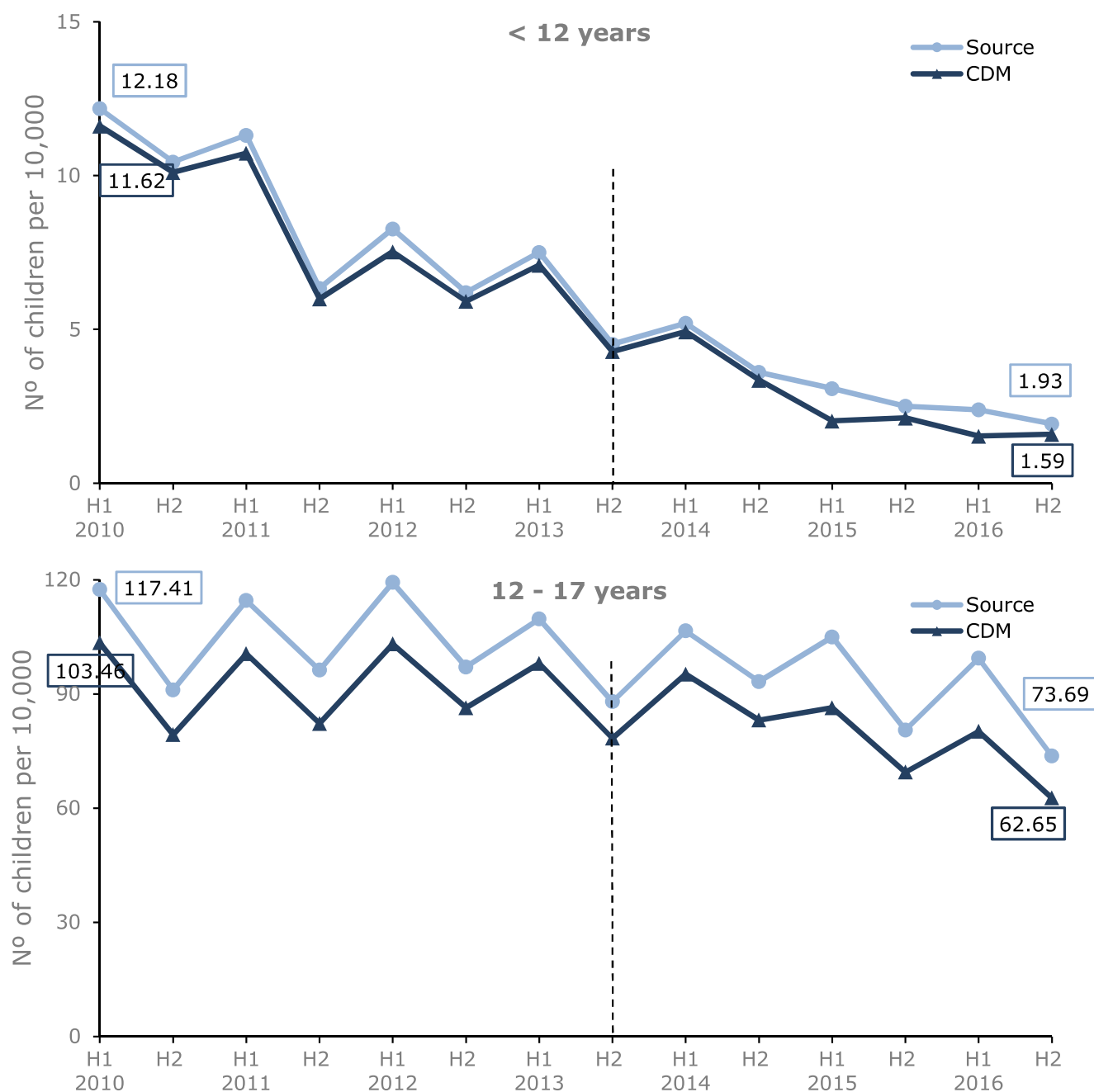


Figure 3 Six-monthly prevalence (per 10,000) of codeine prescribing for pain in children 0–17 years by age group in IQVIA Medical Research Data (IMRD)-UK source vs. IMRD-UK Observational Medical Outcome Partnership common data model (CDM; database version 1705). The black vertical line represents the date of introducing the risk minimization measures. The x-axis shows the half year (H1 = January 1 to June 30, H2 = July 1 to December 31). [Colour figure can be viewed at wileyonlinelibrary.com]

Children undergoing TA. The proportion of children with a prescription for codeine within 30 days of a TA is shown in **Figure 4**. No miscoded products were used in children undergoing TA, and the results are superimposable. Both OMOP CDM and source data show a pronounced drop in prescribing of codeine starting in the first half of 2013, when the RMM was still under discussion. Only a small proportion of children undergoing TA had a diagnosis of OSA recorded prior to the TA (~ 5%); in this subgroup of children, the drop

coincided with the RMM and prescribing of codeine decreased to almost zero after the RMM.

Duration of treatment and prior use of other analgesics. Finally, results related to median duration and the proportion of children who had received a prescription for another analgesic within 90 days prior to the codeine prescription showed no significant change during the study period and had almost identical patterns across the two databases (**Figures S3 and S4**).

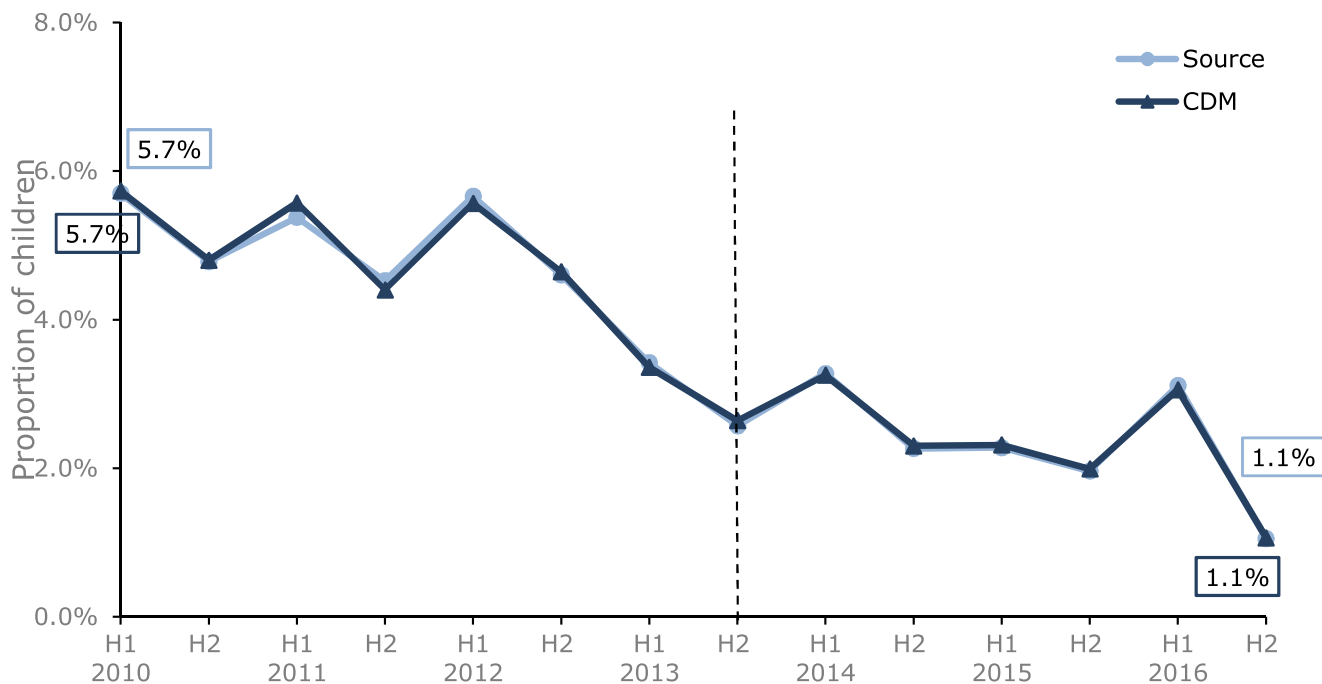


Figure 4 Proportion of children 0–17 years with codeine treatment within 30 days of undergoing tonsillectomy or adenoidectomy by six monthly periods in IQVIA Medical Research Data (IMRD)-UK source vs. IMRD-UK Observational Medical Outcome Partnership common data model (CDM; database version 1705). The black vertical line represents the date of introducing the risk minimization measures. The x-axis shows the half year (H1 = January 1 to June 30, H2 = July 1 to December 31). [Colour figure can be viewed at wileyonlinelibrary.com]

Joinpoint analysis of prescribing trend. Results of the joinpoint analysis in children 0–11 years confirmed a decreasing trend during the whole study period and did not identify any change in prescribing trends in neither source nor OMOP CDM databases (Figure S5).

IMRD-UK version 1809

Using IMRD-UK version 1809 (and an extended study period ending in June 2018), the results were almost identical. In the OMOP CDM, 1,953,668 children were included in the study population, 9,556 (0.5%) less than in the source database. The discrepancy is due to the different methodology adopted in the calculation of the 1-year observation period, as was also seen in the previous analysis.

The number of children with a codeine prescription during the study period was also very similar: 71,293 in OMOP CDM vs. 72,830 in IMRD-UK source, a difference of 2.1%.

No difference in overall prevalence of prescribing of codeine by age group between OMOP CDM and IMRD-UK source was observed (Figure 5). All of the other elements of the RMM studied also showed a complete alignment between the two databases.

DISCUSSION

The main objective of the study was to evaluate the implementation of OMOP CDM on IMRD-UK data running the same study on both and investigating unexpected differences. This facilitated the exploration of any potential loss of information and inaccuracy and informed on whether the results and conclusions of the study on the transformed dataset were as reliable and valid as on the source dataset.

For this particular study, besides the few differences noted, the transformation could be considered adequate confirming previous studies that suggest a CDM is suitable for use in observational research and providing new evidence that the transformation can also be applied successfully on the IMRD-UK database. The OMOP CDM structure successfully accommodated all the variables needed, and all the analyses in the source data could be performed on the converted data, although the study was not overly complicated. The content was also faithfully captured except for the miscoded codeine treatments.

Interestingly, the identified differences had different causes and from each of them a recommendation is proposed (Table 1).

The different conventions adopted regarding the date of birth for children younger than 15 years old and the value used for calculating the exposure duration raised two considerations. First, inaccuracy might come not only from the mapping to the standard dictionaries, considered as the main “weakness” of the OMOP CDM, but from every aspect of the transformation, potentially affecting also those CDMs that focus only on the data structure. Second, it is critical to perform the transformation in close collaboration with both those knowledgeable about the source data, to ensure that conventions in the source are applied correctly, and with experts in the destination model, knowledgeable about the transformation process.

The fact that the CDM allowed less discretion in selecting the start of the observation period may be seen as both a strength and a weakness. On one hand, it enforces a common definition between databases avoiding nuances derived by small differences in the interpretation of the same protocol across different studies. On the other hand, there may be genuine reasons to prefer different options in different studies. From our point of view, it forced us

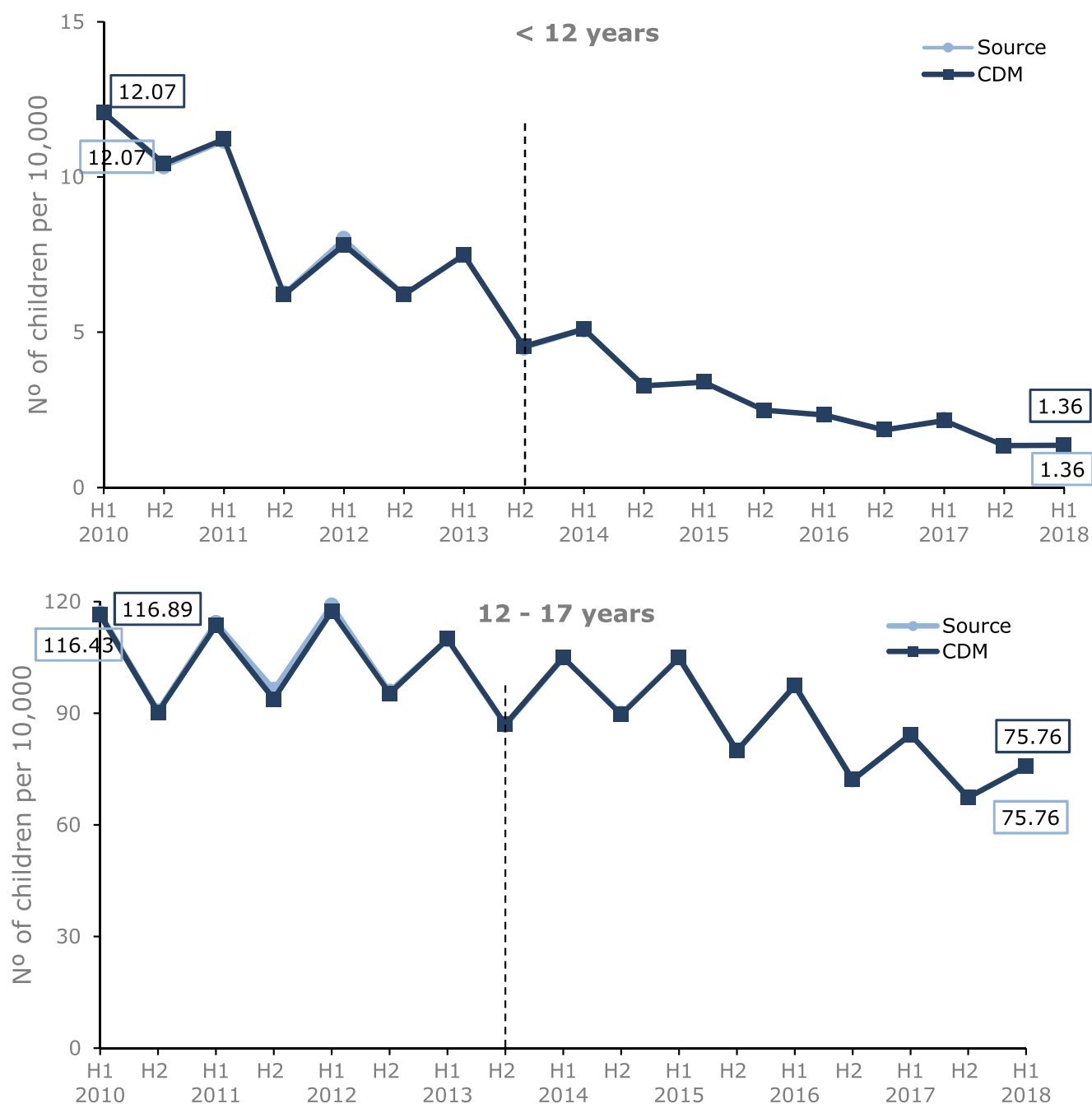


Figure 5 Six-monthly prevalence (per 10,000) of codeine prescribing for pain in children 0–17 years by age group in IQVIA Medical Research Data (IMRD)-UK source vs. IMRD-UK Observational Medical Outcome Partnership common data model (CDM; database version 1809). The black vertical line represents the date of introducing the risk minimization measures. The x-axis shows the half year (H1 = January 1 to June 30, H2 = July 1 to December 31). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.com)]

to recognize that options were available in the source data and to explicitly consider whether our current methodology is more suitable or somewhat arbitrary. In other words, the discipline of using a CDM can prompt an explicit evaluation of what constitutes good scientific practice.

Finally, the misclassification of two drug treatments to the device domain showed the importance of details. Previous studies have shown good results of checks regarding the proportion of

mapped terms and regarding individual reviews of the top most frequently occurring mapped and unmapped conditions, procedures, and drugs: These are all necessary and useful checks. However, the more the OMOP CDM is used, the more important it is to perform detailed checks on more individual terms as they might reveal systematic errors or discrepancies relevant in studies with less frequent exposures or outcomes (e.g., focusing on particular formulations or specific conditions). Moreover, it

Table 1 Recommendations when transforming a database into an OMOP CDM

Recommendations	Detailed recommendations	Rationale
Work in close collaboration with a broad range of expertise during the transformation	<ul style="list-style-type: none"> Involve experts in <ul style="list-style-type: none"> The source data The destination model Terminologies 	<ul style="list-style-type: none"> Ensure conventions used in the source data are applied correctly Ensure the transformation process requirements are carefully applied Wide stakeholder engagement will increase adoption and support the sustainability of the model
Operationalize reliability and validity by building clear and consistent rules for the transformation and applying a data quality framework	<ul style="list-style-type: none"> Ensure the data quality framework has been verified by an independent party Widen checks between source and transformed data to most of the individual terms Ensure the data quality framework includes running of validation studies and investigate small differences that arise 	<ul style="list-style-type: none"> Increase reliability of the transformed data for regulatory decision making Harmonize quality standards and checks between different organization transforming the data to CDMs Widen the routine use of CDMs to studies with less frequent exposures or outcomes
Avoid the same database being transformed by multiple organizations	<ul style="list-style-type: none"> Agree a framework where the same database is transformed by only one organization (preferably the data owner for expertise and data protection considerations) 	<ul style="list-style-type: none"> Drive consistency, avoiding the potential of different conventions and inaccuracies More efficient use of resources
Provide clear and nontechnical documentation to increase transparency	<ul style="list-style-type: none"> Document clearly how each variable in the CDM version has been defined from the source data Incorporate and document any validation done Retain the source data in the transformed dataset 	<ul style="list-style-type: none"> Help users to understand conventions used during the transformation Allow transparency and reproducibility of data and tools to facilitate credible and robust evidence
Use implementation of international standards for dictionary where possible	<ul style="list-style-type: none"> Adopt the ISO standard for IMPD 	<ul style="list-style-type: none"> Link with other health data using the same international standard More complete and granular dictionary that will improve the quality of the mapping
Promote an open communication between users and organizations performing the data transformation	<ul style="list-style-type: none"> Create an iterative process for improvements that include feedback from users Ensure that CDM is dynamic, extendable and learn from experience 	<ul style="list-style-type: none"> Users can highlight nuances and influence future developments to widen the cases where a CDM can be used

CDM, common data model; IMPD, Identification of Medicinal Products; ISO, International Organization for Standardization; OMOP, Observational Medical Outcome Partnership.

is also recommended to include the running of similar studies to the current one as a final step in the quality checks after the initial transformation and to investigate even small differences that arise.

Reaching this level of details in the checks requires a significant effort of resources. As there is the possibility, which has already happened in the past, that the same database could be transformed into a CDM by different organizations, it is strongly recommended that one database is transformed only by one organization to have a more efficient use of the resources. This will also drive consistency avoiding creating multiple CDMs of the same database with the potential of different conventions used and different inaccuracies. Moreover, a data quality framework, including a set of standards and detailed quality checks agreed, harmonized, and verified by an independent party, could help increasing the quality of the checks in an efficient way. International standardization and harmonization, including the new International Organization for Standardization standard for medicinal products,³⁷ may also contribute to improve the quality of mapping in the future with the added benefit to facilitate links with other health data sources.

It is important to consider that this was the first time the data provider, IQVIA, transformed IMRD-UK into OMOP CDM; both the provider of the data transformation and the analysts were on a learning curve and small mistakes or differences in interpreting the data are part of this process. This was reflected in the rerun of the analysis with the 1809 version of the database, after adjusting the ETL process based on experiences from the first analysis, where the results were almost identical between the two databases. This highlights the role of the users and analysts of the transformed data and of the data transformation organizations in creating an iterative process for improvements in the transformation to a CDM. The former, with their feedback, can highlight small nuances and influence future developments to widen the use cases; the latter, being open and reactive, can help in creating a virtuous cycle.

Despite the discrepancies found, it is important to highlight that they did not affect the conclusions and interpretation of the study. Results in both the OMOP CDM and the source database suggest that: (i) prescribing of codeine for treatment of pain in children, especially below the age of 12 years, decreased over time starting before the introduction of the RMM; (ii)

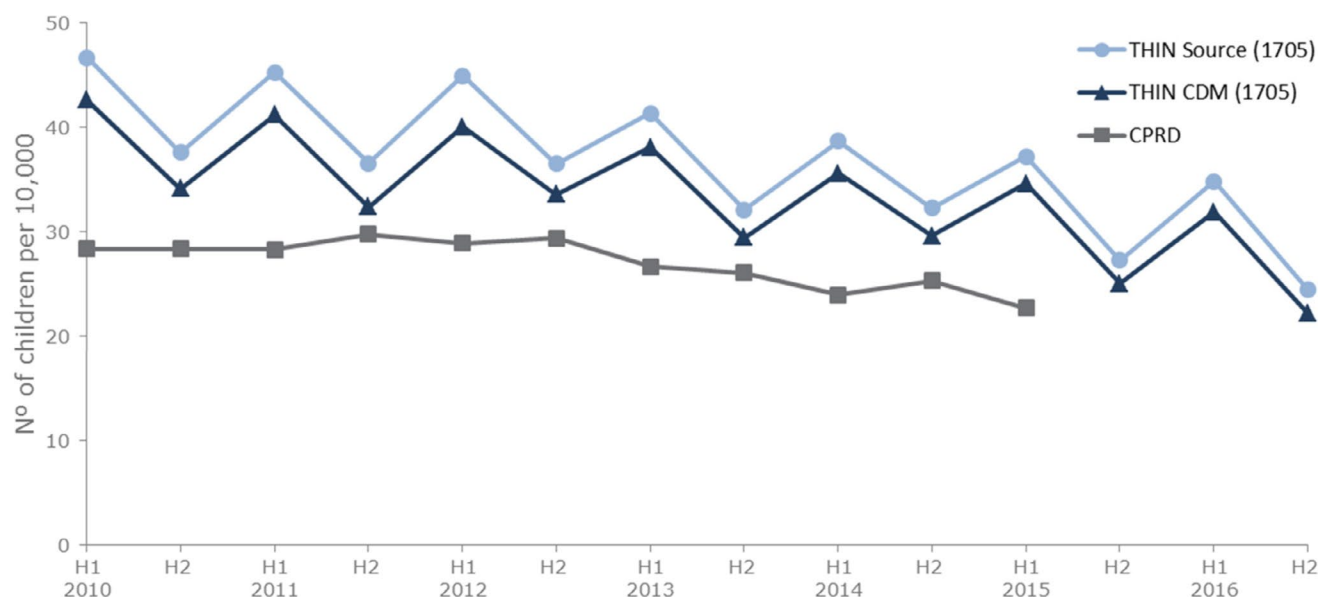


Figure 6 Six-monthly prevalence (per 10,000) of codeine prescribing for pain in children 0–17 years by age group in IQVIA Medical Research Data (IMRD)-UK source, IMRD-UK Observational Medical Outcome Partnership common data model (CDM; database version 1705), and Clinical Practice Research Datalink (CPRD). [Colour figure can be viewed at wileyonlinelibrary.com]

prescribing of codeine within 30 days of undergoing TA decreased considerably while the RMM was under discussion (and went to almost zero if a diagnosis of OSA was recorded); and (iii) there was no observed change in duration of use and prescription of other analgesics.

The question about generalizability of these results to other use cases and to other data sources is valid and important but cannot be answered by this study. We are currently creating a more general validation plan that will specify further studies and attempt to answer the question of how much and what type of evidence is required to support the more routine use of a CDM for regulatory decision making. At the same time, we are keen to see further work by others that may address different databases. The more use cases are tested, the more guidance can be provided on which studies the OMOP CDM is more suited for, as a proper understanding of the limitations of the underlying data is always required to ensure that appropriate conclusions are inferred, the extent of which depends on the specific study and research question.

Being able to run a multiple database study using a CDM has several advantages: Preparing one programming code for data management and analysis that is able to run on all the transformed databases has significant implications on the speed of the analysis and resources required. Moreover, given that the data model, the terminologies, and the analysis method are the same, differences in results across databases are more likely to be due to genuine heterogeneity derived by the underlying patient populations and the data captured about them, influenced by the national health systems, rather than due to divergent approaches taken to analyze the data.

For this particular study, the same analyst wrote the programming codes for both source and transformed data; this assured that minor decisions in the interpretation of the protocol were aligned, and, where this was not possible, differences were

highlighted (as for the calculation of the 1-year observation period). The consequence was that the discrepancies seen were due to shortcomings in the conversion of the underlying data and not due to differences in interpreting and executing the study protocols. It is recognized that the latter can cause important variance in the results,³⁸ and this has been seen in our experience while running studies on multiple databases to which the EMA has access.

This research has focused on challenges associated with the implementation of a CDM. Such challenges must always be viewed in the light of the problems inherent in doing multidatabase studies without a CDM. IMRD-UK and CPRD GOLD (from practices with Vision software) are two population-based electronic health record databases from general practitioners in the United Kingdom that are similar in structure and content with a considerable overlap of patients.³⁹ CPRD was one of the databases used in the Hedenmalm *et al.* study.²⁶ The results of the CPRD study, compared with those found in this study on IMRD-UK, not only revealed a lower prevalence in the prescribing of codeine but, more remarkably, an absence of seasonality (Figure 6). The presence or absence of seasonality seemed to be caused by differences in how the month of birth, not available for children older than 15 years, was handled (for instance, whether it was imputed in January or in July). Such difficulty to explain effects in untransformed databases from essentially similar populations provide a cogent motivation for pursuing the idea of a CDM.

CONCLUSION

To optimize the regulatory decision-making process, regulators need timely access to data that are of high quality, relevant for benefit-risk assessment, which supports multiple use cases, representative of the whole of Europe, and generated through

a transparent methodology.^{40,41} These requirements translate into being able to efficiently access and analyze multiple databases, and, therefore, the use of a CDM is a core component.

For this study, the conversion to OMOP CDM was adequate. Although some decisions and mapping could be improved, these impacted on the absolute study results, but not on the study inferences. This validation study supports six recommendations for good practice in transforming to CDMs (Table 1) and we encourage researchers to conduct validation studies comparing source and CDM transformed for different datasets and different research questions. By sharing our learnings, we will drive up the reliability and utility of studies on observational data as evidence for decision making.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

Supplementary Figures.

Supplementary Tables.

Supplementary Text.

ACKNOWLEDGMENTS

The authors thank IQVIA for providing the IQVIA Medical Research Data incorporating data from THIN, A Cegedim Database, transformed in OMOP CDM.

FUNDING

No funding was received for this work.

CONFLICT OF INTERESTS

All authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

All authors wrote the manuscript. G.C., J.S., X.K., and P.A. designed the research. G.C. and K.H. performed the research. G.C. and K.H. analyzed the data.

DISCLAIMER

The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.

© 2020 The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Cave, A., Kurz, X. & Arlett, P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin. Pharmacol. Ther.* **106**, 36–39 (2019).
2. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Chapter 4: approaches to data collection. In ENCePP Guide on Methodological Standards in Pharmacoepidemiology. European Medicines Agency <http://www.encepp.eu/standards_and_guidances/methodologicalGuide4.shtml> Accessed August 26, 2019.
3. European Commission. Guideline on the format and content of applications for designation as orphan medicinal products and on the transfer of designations from one sponsor to another, 27.03.2014. Belgium: European Commission Health and

Consumers Directorate-General <https://ec.europa.eu/health/sites/health/files/files/orphanmp/2014-03_guideline_rev4_final.pdf> (2014).

4. European Medicines Agency. Orphan designation: Overview <<https://www.ema.europa.eu/en/human-regulatory/overview/orphan-designation-overview>>.
5. Zhou, X. *et al.* An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance. *Drug Saf.* **36**, 119–134 (2013).
6. Overhage, J.M., Ryan, P.B., Reich, C.G., Hartzema, A.G. & Stang, P.E. Validation of a common data model for active safety surveillance research. *J. Am. Med. Inform. Assoc.* **19**, 54–60 (2012).
7. Xu, Y. *et al.* A comparative assessment of observational medical outcomes partnership and mini-sentinel common data models and analytics: implications for active drug safety surveillance. *Drug Saf.* **38**, 749–765 (2015).
8. Platt, R.W., Platt, R., Brown, J.S., Henry, D.A., Klungel, O.H. & Suissa, S. How pharmacoepidemiology networks can manage distributed analyses to improve replicability and transparency and minimize bias. *Pharmacoepidemiol. Drug Saf.* **29**, 3–7 (2019).
9. Hripcsak, G. *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies Health Technol. Inform.* **216**, 574–548 (2015).
10. Observational Health Data Sciences and Informatics (OHDSI). Observational health data sciences and informatics <<https://www.ohdsi.org/>>.
11. Ball, R., Robb, M., Anderson, S.A. & Dal Pan, G. The FDA's sentinel initiative—a comprehensive approach to medical product surveillance. *Clin. Pharmacol. Ther.* **99**, 265–268 (2016).
12. Schneeweiss, S., Brown, J.S., Bate, A., Trifiro, G. & Bartels, D.B. Choosing among common data models for real-world data analyses fit for making decisions about the effectiveness of medical products. *Clin. Pharmacol. Ther.* **107**, 827–833 (2020).
13. Matcho, A., Ryan, P., Fife, D. & Reich, C. Fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model. *Drug Saf.* **37**, 945–959 (2014).
14. Voss, E.A. *et al.* Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J. Am. Med. Inform. Assoc.* **22**, 553–564 (2015).
15. Haberson, A., Rinner, C. & Gall, W. Standardizing Austrian claims data using the OMOP common data model: a feasibility study. *Stud. Health Technol. Inform.* **258**, 151–152 (2019).
16. Haberson, A., Rinner, C., Schoberl, A. & Gall, W. Feasibility of mapping Austrian health claims data to the OMOP common data model. *J. Med. Syst.* **43**, 314 (2019).
17. Lima, D.M., Rodrigues, J.F. Jr, Traina, A.J.M., Pires, F.A. & Gutierrez, M.A. Transforming two decades of ePR data to OMOP CDM for clinical research. *Stud. Health Technol. Inform.* **264**, 233–237 (2019).
18. Maier, C. *et al.* Towards implementation of OMOP in a German University Hospital Consortium. *Appl. Clin. Inform.* **9**, 54–61 (2018).
19. Makadia, R. & Ryan, P.B. Transforming the premier perspective hospital database into the Observational Medical Outcomes Partnership (OMOP) common data model. *EGEMS* **2**, 1110 (2014).
20. Yoon, D. *et al.* Conversion and data quality assessment of electronic health record data at a Korean tertiary teaching hospital to a common data model for distributed network research. *Healthcare Inform. Res.* **22**, 54–58 (2016).
21. You, S.C. *et al.* Conversion of National Health Insurance Service-National Sample Cohort (NHIS-NSC) Database into Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM). *Stud. Health Technol. Inform.* **245**, 467–470 (2017).
22. Rijnbeek, P.R. Converting to a common data model: what is lost in translation?: Commentary on "fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model". *Drug Saf.* **37**, 893–896 (2014).

23. European Medicines Agency (EMA)/350259/2013. PRAC recommends restricting the use of codeine when used for pain relief in children London, United Kingdom <http://www.ema.europa.eu/docs/en_GB/document_library/Referrals_document/Codeine_containing_medicinal_products/Recommendation_provided_by_Pharmacovigilance_Risk_Assessment_Committee/WC500144445.pdf> (2013).
24. European Medicines Agency. Restrictions on use of codeine for pain relief in children – CMDh endorses PRAC recommendation London, United Kingdom <http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2013/06/news_detail_001829.jsp&mxml:id=WC0b01ac058004d5c1> (2013).
25. European Medicines Agency (EMA)/H/A-31/1342. Assessment report for codeine-containing medicinal products indicated in the management of pain in children 2013 [updated June 24, 2013. 1–18] <http://www.ema.europa.eu/docs/en_GB/document_library/Referrals_document/Codeine_containing_medicinal_products/Recommendation_provided_by_Pharmacovigilance_Risk_Assessment_Committee/WC500147065.pdf>.
26. Hedenmalm, K. *et al.* A European multicentre drug utilisation study of the impact of regulatory measures on prescribing of codeine for pain in children. *Pharmacoepidemiol. Drug Saf.* **28**, 1086–1096 (2019).
27. Becher, H., Kostev, K. & Schroder-Bernhardi, D. Validity and representativeness of the "Disease Analyzer" patient database for use in pharmacoepidemiological and pharmacoeconomic studies. *Int. J. Clin. Pharmacol. Ther.* **47**, 617–626 (2009).
28. Salvador Rosa, A., Moreno Perez, J.C., Sonogo, D., Garcia Rodriguez, L.A. & de Abajo Iglesias, F.J. The BIFAP project: database for pharmaco-epidemiological research in primary care. *Aten. Primaria* **30**, 655–661 (2002).
29. Herrett, E. *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int. J. Epidemiol.* **44**, 827–836 (2015).
30. Blak, B.T., Thompson, M., Dattani, H. & Bourke, A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Informatics Primary Care* **19**, 251–255 (2011).
31. OHDSI/CommonDataModel <<https://github.com/OHDSI/CommonDataModel/wiki>>.
32. IQVIA. *THIN Research Format Data Guide*. Version 1809 (IQVIA, London, UK, 2018).
33. Kim, H.J., Fay, M.P., Feuer, E.J. & Midthune, D.N. Permutation tests for joinpoint regression with applications to cancer rates. *Stat. Med.* **19**, 335–351 (2000).
34. National Cancer Institute. Joinpoint Regression Program, version 4.5.0.1 – June 2017 <<https://surveillance.cancer.gov/joinpoint/>> (2017).
35. Fay, M.P., Tiwari, R.C., Feuer, E.J. & Zou, Z. Estimating average annual percent change for disease rates without assuming constant change. *Biometrics* **62**, 847–854 (2006).
36. Morales, D.R., Slattery, J., Pacurariu, A., Pinheiro, L., McGettigan, P. & Kurz, X. Relative and absolute risk of tendon rupture with fluoroquinolone and concomitant fluoroquinolone/corticosteroid therapy: population-based nested case-control study. *Clin. Drug Investig.* **39**, 205–213 (2018).
37. Tranchard, S. Revised IDMP standards to improve description of medicinal products worldwide <<https://www.iso.org/news/ref2234.html>> (2017).
38. Schneeweiss, S. & Glynn, R.J. Real-world data analytics fit for regulatory decision-making. *Am. J. Law Med.* **44**, 197–217 (2018).
39. Carbonari, D.M. *et al.* Use of demographic and pharmacy data to identify patients included within both the Clinical Practice Research Datalink (CPRD) and The Health Improvement Network (THIN). *Pharmacoepidemiol. Drug Saf.* **24**, 999–1003 (2015).
40. Cave, A. Opportunities for 'big data' in medicines development and regulatory science. 34th International Conference of Pharmacoepidemiology & Therapeutic Risk Management, Prague, Czech Republic, August 24, 2018.
41. European Medicines Agency (EMA)/614680/2018. A common data model for Europe? Why? Which? How? <https://www.ema.europa.eu/en/documents/report/common-data-model-europe-why-which-how-workshop-report_en.pdf> (2018).