Quality registry data represented in the OMOP CDM

Rowdy de Groot, MSc, rowdy.degroot@amsterdamumc.nl[1], Ronald Cornet, PhD, r.cornet@amsterdamumc.nl[1], Nicolette de Keizer, PhD, n.f.keizer@amsterdamumc.nl[1] , Nirupama Benis, PhD, n.benis@amsterdamumc.nl[1], Ferishta Raiez, PhD, f.raiez@amsterdamumc.nl[1]

[1] Department of Medical Informatics, Amsterdam UMC, Location AMC, Amsterdam Public Health research institute, Amsterdam, The Netherlands

Abstract

NICE (National Intensive Care Evaluation) is a quality registry that collects observational data on Intensive Care Unit (ICU) admissions from all Dutch ICUs extracted from their EHRs. In other countries, ICU registries similar to NICE exist and there is a strong intention to conduct joint research. However, this is hampered by the fact that every registry has its own dataset and data definitions. Using a common information model could solve this. Therefore, our aim was to transform quality-registry data from a proprietary format to OMOP CDM.

The process for the implementation was guided by the Extract, Transform and Load course from the EHDEN academy. The book of OHDSI and tools were used to execute the process of designing the Extract, Transform, Load (ETL) based on OMOP CDM.

The implementation of the NICE minimal data set, containing 204 items, to OMOP CDM was completed. All NICE data items could be represented in OMOP CDM and 89.9% of the data values could be mapped in Usagi. The ETL was implemented as an SQL database in PostgreSQL.

OMOP CDM proved to be a viable information model to represent data from an ICU quality registry. The tools OHDSI provides proved to be useful in the process of designing the ETL process and to transform the NICE data to OMOP CDM compliant data source.

Research Category (please highlight or circle which category best describes your research)

Methodological research

Introduction/background

With the increasing need to improve the infrastructure supporting the reuse of data, more attention is given to making data FAIR (Findable, Accessible, Interoperable, Reusable)[1]. OHDSI (Observational Health Data Sciences and Informatics) is an initiative that supports data sources to become partners in a research network to make their data sources FAIR and enable the possibility of answering large multisite research and policy questions by reusing data[2].

The NICE (National Intensive Care Evaluation) registry is such a data source including data on Intensive Care Unit (ICU) admissions from all ICUs in the Netherlands[3]. ICUs extract a predefined dataset from their Electronic Patient Record (EHR) and upload these to the central NICE database on a monthly basis. NICE provides ICUs with audit and feedback to monitor clinical performance over time and to benchmark with national averages or best practices in order to improve quality of care. Data from the NICE registry is also used for clinical research[4]. In other countries, ICU registries similar to NICE exist and there is a strong intention to conduct joint research.

Although NICE has a rich dataset, re-use of data for joint research with other data sources or (inter-) national research groups is hampered by a lack of standardization in data model and data definitions.

For the NICE registry to become FAIR and improve data reuse, data should be represented using a standardized information model. The information model is an important component in the process of making data FAIR. It is a representation of concepts, relationships, constraints, rules and operations that are used to specify data semantics in a certain domain[5]. Information models clearly define the data items in the database and the relationships between them. Therefore, information models are useful for interoperability (the 'I' of FAIR) between databases.

The OHDSI initiative makes use of OMOP CDM, a person-centric relational model mostly suitable for observational data[2,6]. However, it is not clear whether OMOP CDM could also be used to represent data from a quality registry such as NICE. In contrast to raw EHR data, NICE contains many aggregated data

items in a certain context, such as the highest blood pressure in first 24 hours of ICU admission, that could be hard to represent in a common data model. The primary aim of this study was to represent the NICE quality registry data in OMOP CDM and to investigate whether OMOP CDM is suitable for an observational quality registry. The secondary aim was to evaluate the OHDSI tools to facilitate the whole ETL process of the NICE quality registry.

Methods

For this study the Minimum Data Set (MDS) was used, which includes demographic information, laboratory and physiological outcomes in the first 24 hours of admission, reason for ICU admission, and IC and hospital mortality and length of stay[3]. The MDS consists of 204 variables.

The book of OHDSI was used to perform the process of designing the Extract Transform, Load (ETL) for OMOP CDM[7]. Figure 1 shows the process of designing and implementing the ETL and the tools that were used[7]. The process for the implementation was guided by the Extract, Transform and Load course from the EHDEN academy[10]. ACHILLES was used in RStudio and connected with the database in PostgreSQL to prove that it can characterize and assess the data quality.

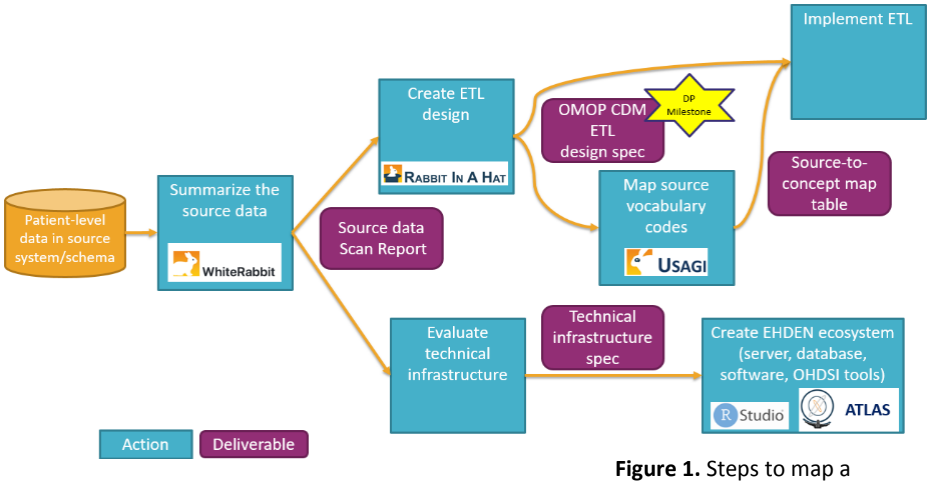

Figure 1. Steps to map a dataset to the OMOP CDM.

Results

In Rabbit-in-a-Hat all source data items could be mapped to CDM fields. In Usagi 89.9% of the source data values were mapped to an OMOP concept id. Data values for items containing contextual information such as "*blood pressure in the first 24 hours after ICU admission*" could not be fully mapped. The whole ETL design process took about 35 hours to complete. The ETL designed in Rabbit-in-a-Hat was used as a blueprint for the implementation of the NICE MDS data in the OMOP CDM. Usagi provided the SOURCE_TO_CONCEPT table. This table searches for the corresponding concept id of the source code. However, source codes are not used in the NICE database. Therefore, an extra table CODES was created which links source values in the dataset to the corresponding source code given in Usagi. With the CODES table, it was possible to fill the OMOP CDM tables with the correct concept ids. The ACHILLES tests were completed by which we could characterize and assess the data quality.

Discussion/Conclusion

OMOP CDM proved to be a viable information model to represent data from an ICU quality registry such as NICE. Only the aggregated data items in a certain context were hard or sometimes impossible to map in Usagi. Future research is needed on how to optimize the representation of these data items in OMOP CDM. The tools OHDSI provides proved to be useful in the process of designing the ETL process to transform the NICE data to OMOP CDM compliant data.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3.
2. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. Studies in health technology and informatics. 2015;216:574
3. van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data resource profile: the Dutch National Intensive Care Evaluation (NICE) registry of admissions to adult intensive care units. International journal of epidemiology. 2015;44(6):1850-h.
4. NICE. Wat we doen 2020 [cited 2020 10th of January]. Available from: https://www.stichting-nice.nl/watwedoen.jsp
5. Lee YT, editor Information modeling: From design to implementation. Proceedings of the second world manufacturing congress; 1999: International Computer Science Conventions Canada/Switzerland.
6. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. Journal of the American Medical Informatics Association. 2011;19(1):54-60.
7. OHDSI. The Book of OHDSI 2019 [cited 2020 10th of January]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/
8. OHDSI. WhiteRabbit 2020 [cited 2020 27th of January]. Available from: https://github.com/OHDSI/WhiteRabbit
9. OHDSI. Usagi 2020 [cited 2020 27th of January]. Available from: https://github.com/OHDSI/Usagi
10. EHDEN. EHDEN academy 20020 [cited 2020 27th May]. Available from: https://academy.ehden.eu/.