

Research and Applications

Using the Observational Medical Outcomes Partnership Common Data Model for a multi-registry intensive care unit benchmarking federated analysis: lessons learned

Aasiyah Rashan ^{*}, Daniel P. Püttmann ¹, MSc^{2,3,4}, Nicolette F. de Keizer, PhD^{2,3,4}, Dave A. Dongelmans, MD, PhD^{3,5}, Ronald Cornet, PhD^{2,6}, Otavio Ranzani, PhD^{7,8}, Wangari Waweru-Siika, MMed⁹, Matthew Smith, PhD¹⁰, Steve Harris ¹, PhD¹, Abi Beane, PhD¹¹, Ferishta Bakhshi-Raiez, PhD^{2,3,4}, for the Collaboration for Research, Implementation and Training in Critical Care—Asia and Africa Investigators, and the Dutch National Intensive Care Registry

¹Institute of Health Informatics, University College London, London WC1E 6BT, United Kingdom, ²Department of Medical Informatics, Amsterdam Public Health Institute, Amsterdam UMC, University of Amsterdam, Amsterdam 1105 AZ, The Netherlands, ³National Intensive Care Evaluation (NICE) Foundation, Amsterdam 1105 AZ, The Netherlands, ⁴Quality of Care, Amsterdam Public Health Institute, Amsterdam 1105 AZ, The Netherlands, ⁵Department of Intensive Care, Amsterdam University Medical Centers Location Academic Medical Center, Amsterdam 1105 AZ, The Netherlands, ⁶Digital Health, Amsterdam Public Health Institute, Amsterdam 1105 AZ, The Netherlands, ⁷Barcelona for Global Health, ISGlobal, 08036 Barcelona, Spain, ⁸DataHealth Lab, Institut de Recerca Sant Pau (IR SANT PAU), 08041 Barcelona, Spain, ⁹Department of Anaesthesia, Aga Khan University, Nairobi 30270-00100, Kenya, ¹⁰Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom, ¹¹Pandemic Science Hub, Institute of Regeneration and Repair, University of Edinburgh, Edinburgh EH16 4UU, United Kingdom,

*Corresponding author: Aasiyah Rashan, MRes, Institute of Health Informatics, University College London, Gower Street, London WC1E 6BT, United Kingdom (aasiyah@nicslk.com)

A. Rashan and D.P. Püttmann are joint first authors of this work.

Abstract

Objective: Federated analysis is a method that allows data analysis to be performed on similar datasets without exchanging any data, thus facilitating international research collaboration while adhering to strict privacy laws. This study aimed to evaluate the feasibility of using federated analysis to benchmark mortality in 2 critical care quality registry databases converted to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), describing challenges to and recommendations for performing federated analysis on data transformed to OMOP CDM.

Materials and Methods: To identify as many challenges as possible and to be able to complete the benchmarking phase, a 2-step approach was taken during implementation. The first step was a naive implementation to allow challenges to surface naturally; the second step was developing solutions for the encountered challenges. Expected patient mortality risk was calculated by applying the Acute Physiology and Chronic Health Evaluation II (APACHE II) model to data from OMOP CDM databases containing adult ICU encounters between July 1, 2019 and December 31, 2022. An analysis script was developed to calculate comparable, registry level standardized mortality ratios. Challenges were recorded and categorized into predefined categories: “data preparation,” “data analysis plan,” and “data interpretation.” Challenges specific to the OMOP CDM were further categorized using published steps from an existing generic harmonization process.

Results: A total of 7 challenges were identified, 4 of which were related to data preparation, 1 to data analysis, and 1 to data interpretation. Out of all 7 challenges, 4 stemmed from decisions made during the implementation of OMOP CDM. Several recommended solutions were distilled from the naive approach.

Discussion: Federated analysis facilitated by a CDM is a feasible option for critical care quality registries. However, future analysis is influenced by decisions made during the CDM implementation process. Thus, prior publication of data dictionaries and the use of metadata to communicate data handling and data source classification during CDM implementation will improve the efficiency and accuracy of subsequent analysis.

Lay Summary

Datasets on critical care admissions are increasingly available for international research, but they are often stored in different formats and use different clinical coding systems. This makes it difficult to perform the same analysis across multiple datasets, as the data need to be moved to a central location and converted into a shared format before it can be analyzed. An alternative approach, known as “bringing the code to the data” rather than “bringing the data to the code,” has been suggested. In this method, each dataset is first converted into a common data model. Then, the same analysis code is run on each dataset without moving the data from their original locations. This study tested whether

Received: March 5, 2025; Revised: May 1, 2025; Editorial Decision: May 27, 2025; Accepted: June 2, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of the American Medical Informatics Association.

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 reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

this method could be used to compare risk-adjusted mortality rates across different critical care datasets. The study found that this approach is feasible, but it comes with challenges, including differences in data availability, data collection, and how data are interpreted across datasets. These challenges were addressed by researchers sharing detailed information about each dataset before agreeing on the analysis plan. This study recommends that this type of detailed information sharing should be a standard step in any shared analysis using common data models.

Key words: federated analysis; global health; data standardization; OMOP CDM; APACHE II; registries; critical care.

Objective

International research collaboration is a significant opportunity for critical care quality registries (henceforth referred to as quality registries for brevity). These registries aim to facilitate the improvement of critical care services through benchmarking, facilitating registry embedded research, and providing data for audit and feedback. Benchmarking is the process of comparing risk adjusted outcomes over a specific period of time, across different peers or sites, and among patient populations to identify opportunities for improvement, learn from best performers, and enhance quality of care.¹⁻⁴ The recent COVID-19 pandemic prompted efforts to expand existing quality registry databases internationally. In response, quality registries demonstrated their potential to generate data for infectious disease-specific surveillance, service forecasting, and clinical research.^{5,6}

The COVID-19 pandemic also highlighted limitations in the ability of quality registries to collaborate for research internationally.⁷⁻⁹ Quality registry datasets are rarely aligned with each other, as they are typically designed to meet the specific needs and capabilities of individual health systems. For example, a universal feature of quality registries is the inclusion of a specific severity of illness score calculated with prognostic models. These models are developed to enable case-mix adjustment, allowing for equitable comparison of health outcomes across a diverse patient population within these registries. However, the implementation of prognostic models to calculate severity of illness scores may differ between quality registries and geographic regions. Until recently, efforts to standardize datasets and prognostic score calculation internationally have been hampered by concerns about data availability and heterogeneity of the global critical care population and care processes.¹⁰

Quality registry datasets require substantial transformation, if the datasets are to be harmonized and made suitable for international collaborative research. Traditionally, this involved transferring data to a central location for cleaning and transformation prior to its analysis. However, such processes are inefficient and increasingly prohibitive to the international community. This is particularly true for countries where cooperation is impeded by strict data sharing regulations.¹¹⁻¹³ An increasingly proposed solution is federated analysis, which involves conducting analysis on multiple datasets using shared analysis scripts, without the need to physically relocate the data to a central repository. Casaletto et al. call it “bringing the code to the data,” rather than the reverse.¹⁴

For federated analysis to be successful, datasets need to align in data models, data elements, and terminology systems. Common data models (CDMs) have sought to provide such standardization and enable data to be analyzed without inspecting source data. Multiple CDMs have been proposed based on their suitability for heterogeneous patient populations and their applicability to clinical documentation. Examples of CDMs include those which standardize trial data¹⁵ and longitudinal data for primary care.¹⁶ Such CDMs further

utilize existing terminology systems like Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and Logical Observation Identifiers Names and Codes (LOINC) to reduce ambiguity over data elements and their values.¹⁷⁻²¹

Quality registries are increasingly adopting the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) as a framework for standardizing data.²²⁻²⁴ Developed by the Observational Health Data Sciences and Informatics (OHDSI) collaborative, this CDM offers a standardized table structure for a wide variety of observational data and additional open access tools for transforming the data and conducting federated analyses. However, to our knowledge, no studies have been published on the feasibility of doing federated analysis on critical care registry data. Therefore, this study aims to evaluate the feasibility of federated analysis to benchmark mortality outcomes on 2 quality registry databases converted to the OMOP CDM and to describe challenges and recommendations for performing a federated analysis on data transformed to the OMOP CDM.

Materials and methods

Data sources

This study used the OMOP CDM databases of 2 quality registries: the National Intensive Care Evaluation (NICE),¹ and the Collaboration for Research, Implementation and Training in Critical Care Asia and Africa (CCAA).² NICE includes data from Dutch Intensive Care Units (ICUs), and CCAA is a network of 17 quality registries operational in countries in Africa and Asia. These quality registries are members of the LOGIC international critical care benchmarking consortium³ and have both independently applied Extract, Transform, and Load (ETL) processes²⁵ to generate their OMOP CDM databases.

The OMOP CDM has a relational table structure centered around a “person,” for example, an ICU patient. This person has a visit occurrence and visit detail, which map to a hospital encounter and ICU encounter respectively. Linked to the person and their visits are measurements, condition occurrences, and procedure occurrences, examples being blood pressure, diabetes, and dialysis, respectively. Any variable that does not fit into the previous 3 categories is mapped to observation, such as a length of stay.²⁶

The NICE registry was established in 1996 and receives data from all Dutch adult ICUs since 2016.¹ Data are gathered from existing electronic health records (EHRs), where it is manually and automatically validated by trained intensivists, ICU-nurses, or data managers, and uploaded monthly to the NICE database. Data include demographics, comorbidities, reasons for admission, highest and lowest laboratory results and physiology in the first 24 hours of ICU admission, organ support information, and patient-level ICU and hospital outcomes such as length of stay and mortality.²⁷ Consequently, the following OMOP tables were filled in the NICE OMOP database: person, visit occurrence, visit detail,

condition occurrence, procedure occurrence, measurement, observation, observation period, and death in ICU.²⁶ NICE was standardized using OMOP CDM version 5.3, and vocabulary version v20220510.²² The OHDSI Achilles tool and the OHDSI Data Quality Dashboard, which run data quality checks on OMOP databases using a harmonized framework, was implemented and the database passed the checks.^{22,28,29}

CCAA was formally established in 2020, with quality registries collecting data prospectively from 2019.² Data are manually gathered daily from paper or electronic records by trained data collectors overseen by trained intensivists and received through a common data platform. The dataset includes information on patient demographics, comorbidities,³⁰ reasons for admission (entered using SNOMED CT terms),³¹ highest or lowest laboratory and point of care tests, and physiology and organ support in the first 24 hours of ICU admission along with ICU and hospital outcome.² The CCAA OMOP database contains all OMOP CDM tables used in the NICE OMOP database described above, with the addition of the care site and drug exposure tables.²⁶ CCAA was standardized using OMOP CDM version 5.4, and vocabulary version v20220602. The OHDSI Data Quality Dashboard was implemented, and the database passed the checks.²⁹

Federated analysis for multi-registry benchmarking Severity of illness score

The standardized mortality ratio (SMR) is a commonly used measure for benchmarking case-mix adjusted outcomes in patient populations. It is the ratio of observed deaths to expected deaths over a time frame and is an indicator of the quality of care. A ratio of one means the quality of care is as expected, a higher ratio indicates a possible deterioration, and a lower ratio indicates a possible improvement.

To calculate the SMR, the observed mortality at ICU discharge in each ICU was divided by its expected mortality, which is the mean of predicted mortality risks in all patients multiplied by the number of patients. The risk of mortality was calculated using the Acute Physiology and Chronic Health Evaluation II (APACHE II) model.³² Although APACHE II was developed to predict hospital mortality, ICU mortality was chosen as the outcome since hospital mortality is not routinely available in the CCAA dataset and external validations have suggested that APACHE II can perform sufficiently on this alternative outcome.³³ Although superseded by APACHE IV, APACHE II is validated internationally and is still widely used to describe population severity of illness and outcomes for critical care research. Moreover, all its component variables are routinely collected by both NICE and CCAA.^{33–35}

Population selection

For NICE, all data were included. For CCAA, 12 collaborating quality registries which had previously agreed on transferring their data to the OMOP CDM format were included. From both sources, all patients admitted to adult ICUs between July 1, 2019 and December 31, 2022 were included. Following APACHE II exclusion criteria, patients were excluded if they were less than 17 years of age or had a diagnosis of burns.³² Patients without an APACHE II reason for admission recorded were also excluded. If patients had

multiple ICU admissions during a hospital admission, only data from their first admission were used.³²

Analysis

For each data source, demographic information, the APACHE II score, and patient ICU outcomes were reported as N (%) or Median (IQR). The availability and distribution of each component variable for APACHE II was reported for both databases. SMRs were calculated per ICU. To visualize the variation in SMRs among different ICUs, so-called funnel plots were used. In such a plot, each ICU is represented as a dot. The value on the x-axis indicates the expected number of deaths for the respective ICU, while the value on the y-axis represents the SMR. The horizontal line represents the average SMR, which has a value of 1. The curved lines represent the corresponding 95% control limits (CL), emphasizing that there is no single “normal” SMR value. Instead, a range of values is considered normal depending on the expected number of deaths.³⁶

In the extraction scripts, Structured Query Language (SQL) was used to select patients meeting eligibility and inclusion criteria. The OHDSI R package SQLRender version 1.16.1 was used to translate SQL queries between SQL dialects.³⁷ R software version 4.1 was used for data analyses.³⁸ The federated analysis was performed by running the same analysis scripts separately on both data sources. Analysis scripts were developed collaboratively and stored on GitHub and Figshare, see Data availability section.

Two-step approach

To identify as many challenges as possible, and to be able to perform the federated analyses, a 2-step approach was taken during implementation. For the naive approach, the study group agreed on the analysis plan above. Author A.R. then developed an analysis script for the CCAA database, and shared it with author D.P., who attempted to apply it on the NICE database. Challenges identified during this first step were recorded and categorized, see details in the section below. In the second step, the study group met to discuss challenges encountered during the first step. They agreed on revisions to the initial analysis plan and identified solutions, which A.R. and D.P. implemented to complete the federated analysis. The solutions are described in more detail in the “Results” section.

Categorization of challenges

Challenges identified during step 1 were recorded, and discussed by the authors A.R., D.P., N.K., D.D., R.C., A.B., and F.B.R. These challenges were displayed in a table under the headings “data preparation,” “analysis,” and “data interpretation.” Those specific to the OMOP CDM were further categorized using a generic harmonization process published by Henke et al., who proposed 9 steps to undertake harmonization of data to a common data model.³⁹ The causes of challenges not related to the OMOP CDM were also described. The solutions used to enable federated analyses are described, along with recommendations for those performing federated analyses on critical care registry data.

Results

Participants and benchmarking

The data sources included 243 346 admissions in 74 ICUs from the NICE data source and 147 495 admissions in 186 ICUs from the CCAA data source, together representing 13 quality registries over the 3.5 years, see Table 1. Figure 1 describes the dataflow of the study. Availability and distribution of the APACHE II variables are described in Table S1 for CCAA and Table S2 for NICE. APACHE II SMRs ranged from 0.7 to 2.4 for all quality registries, see Figure 2. Analysis scripts were shared between the databases by which the federated analysis was successfully performed. The difference in OMOP CDM versions did not affect the variables used in this analysis, as the only difference between versions involved renaming variables which were not utilized in this study.

Table 1. Demographics and outcomes from the OMOP datasets.

Variable	CCAA	NICE
<i>Demographics</i>		
Number of ICUs	186	74
Number of patients	147 495	243 346
Age Median (IQR)	55 (36-67)	66 (55-74)
Male, N (%)	82 996 (56.3)	151 755 (62.4)
<i>Severity of illness</i>		
APACHE II score, Median (IQR)	14 (9.4-19.7)	15 (11.0-20.0)
<i>Outcomes</i>		
ICU mortality, N (%)	30 636 (20.8)	25 424 (10.5)
ICU length of stay (Days), Median (IQR)	2.3 (1.2-4.7)	1.1 (0.8-3.0)

Abbreviations: APACHE, Acute Physiology and Chronic Health Evaluation; CCAA, Critical Care Asia-Africa; ICU, Intensive Care Unit; NICE, National Intensive Care Evaluation; OMOP, Observational Medical Outcomes Partnership.

Challenges identified during the federated analysis

Seven challenges were identified during step 1, 4 of which related to data preparation, 2 to analysis plan, and 1 to data interpretation. Table 2 outlines these challenges, along with their impacts, the solutions implemented in step 2, and recommendations for future analyses. Of the 7 challenges, 4 related specifically to CDM transformation. These challenges arose from decisions made during the ETL process prior to this study regarding semantic mapping, structural mapping, and coverage of vocabularies.

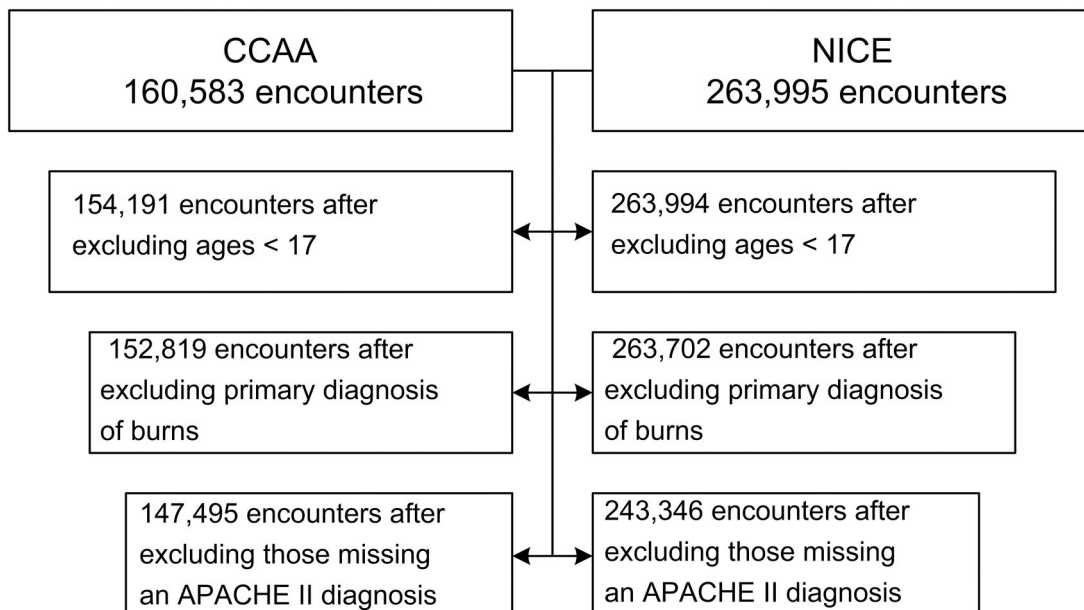
Solutions implemented for each challenge

Technical issues (C1)

NICE and CCAA stored their data on separate servers, requiring different dialects for querying. NICE uses Transact-SQL (T-SQL) queries for their Microsoft SQL Server, while CCAA uses Procedural Language/PostgreSQL (PL/pgSQL) queries for their PostgreSQL server.^{40,41} However, OHDSI's R package SQLRender only converts from T-SQL to other dialects and cannot perform the reverse conversion. Therefore, although author A.R. initially wrote the extraction query in PL/pgSQL, author D.P. manually rewrote the query to T-SQL. The analysis script then used SQLRender to automatically convert the query to PL/pgSQL.

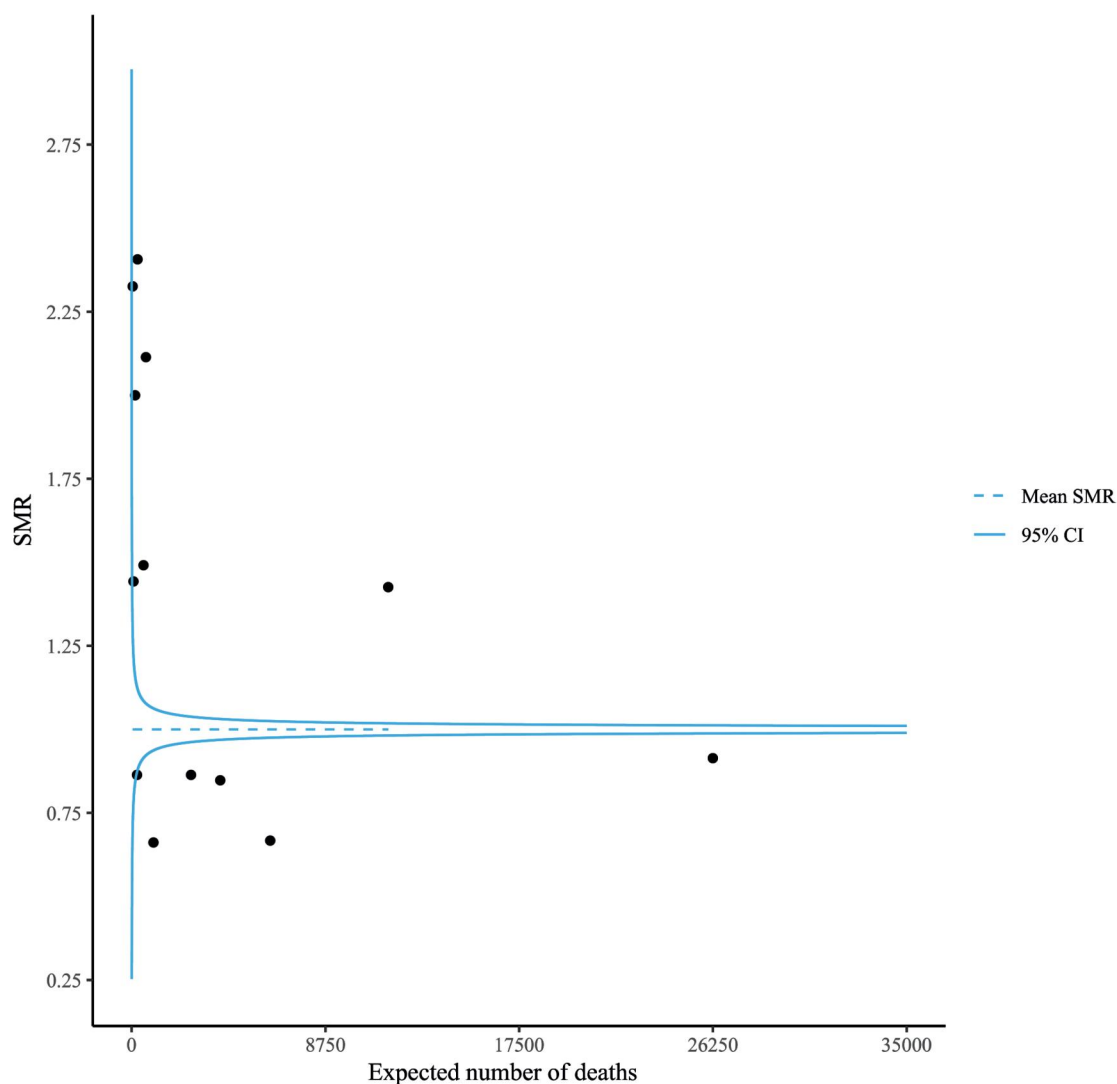
Variations in standardized concept codes and tables (C2 and C3)

The OMOP Standardized Vocabularies allow the same clinical concept to be stored using different concept codes. This affected the physiology variables used for APACHE II. For example, leukocyte count could be stored as either "Leukocytes [#volume] in Blood by Automated count" (3000905) or "Leukocytes [#volume] in Blood by Manual count" (3003282). Thus, the data extraction analysis script



Abbreviations: CCAA, Critical Care Asia-Africa; NICE, National Intensive Care Evaluation; SMR, Standardised Mortality Ratio; ICU, Intensive Care Unit; APACHE, Acute Physiology and Chronic Health Evaluation.

Figure 1. Flowchart of data inclusion and attrition. APACHE, Acute Physiology and Chronic Health Evaluation; CCAA, Critical Care Asia-Africa; ICU, Intensive Care Unit; NICE, National Intensive Care Evaluation; SMR, Standardised Mortality Ratio.



Abbreviations: CCAA, Critical Care Asia-Africa; NICE, National Intensive Care Evaluation; SMR, Standardised Mortality Ratio; ICU, Intensive Care Unit; APACHE, Acute Physiology and Chronic Health Evaluation.

Figure 2. Funnel plot of APACHE II SMRs calculated over the period from July 2019 to December 2022. Each dot represents a registry (12 from CCAA and 1 from NICE). APACHE, Acute Physiology and Chronic Health Evaluation; CCAA, Critical Care Asia-Africa; ICU, Intensive Care Unit; NICE, National Intensive Care Evaluation; SMR, Standardised Mortality Ratio.

was adapted to build SQL queries based on a customizable configuration file which specified a variable's concept code. The concept codes could be based on either standard OMOP vocabularies or on custom codes stored in the data source only. The OMOP CDM allows concepts such as comorbidities and past medical history to be stored in various OMOP tables like Condition versus Observation, depending on interpretation. Therefore, the data extraction analysis script was adapted to allow different tables to be specified in the customizable configuration file.

Variations in collection and mapping of reasons for ICU admission (C4 and C6)

Calculation of the APACHE II mortality probability requires reasons for ICU admission to be coded as APACHE II diagnostic categories. Since 2018, NICE has derived these from the APACHE IV classifications. CCAA collects reasons for ICU

admission as SNOMED CT codes, which are also mapped to APACHE IV classifications prior to mapping to APACHE II classifications. APACHE IV reasons for admission cannot be mapped to standardized OMOP vocabularies and were stored as unmapped data in the OMOP databases. Therefore, the data extraction script was adapted to extract reasons for admission separately for each data source. The data analysis script was also adapted to separately map reasons for admission from each data source into the APACHE II diagnostic categories. The calculation of the APACHE II probability of mortality was then performed using a shared data analysis script.

Privacy concerns (C5)

The original intention described in the naive analysis plan was to report SMRs per ICU. However, NICE did not store identifying ICU information in their OMOP database due to privacy concerns and legal regulations with participating

Table 2. Challenges, their impacts, solutions, and recommendations for future analysis.

Challenge (Category)	Description	Impact	Solution	Recommendation
C1 (Data Preparation): Different SQL dialects between OMOP CDM databases	<p>The 2 data sources are stored in different types of SQL Server (MS SQL & PostgreSQL) and therefore required different SQL dialects to query the server.</p> <p>This was an OMOP CDM related challenge that related to the ETL process in generic harmonization process.</p>	Data from the 2 OMOP CDM databases could not be extracted using a shared SQL query.	The OHDSI SQLRender package was used to convert T-SQL code to PL/pgSQL.	Use the SQLRender package to translate queries where necessary. When using the package, ensure that queries are written in T-SQL, regardless of the servers used, as the package only translates T-SQL queries.
C2 (Data Preparation): Variation in base vocabularies used for concept codes	<p>The 2 data sources occasionally use different base vocabularies to derive their concept codes within the OMOP CDM structure.</p> <p>This was an OMOP CDM related challenge that related to semantic mapping in the generic harmonization process.</p>	Standardized concept codes could not be directly written into the shared data extraction script.	The data extraction script was adapted to build SQL queries based on a customizable configuration file in which the concept code for each variable was specified for each data source.	Proactively account for variations in concept codes by using a customizable configuration file for data extraction, rather than hardcoding concept codes into an SQL query.
C3 (Data Preparation): Variation in OMOP tables used to store concept codes	<p>Specific concepts such as comorbidities and past medical history can be stored in multiple OMOP tables depending on interpretation.</p> <p>This was an OMOP CDM related challenge that related to structural mapping in the generic harmonization process.</p>	Clinical concepts could not be directly linked to the OMOP tables in the SQL queries.	The data extraction queries were adapted to allow different tables to be specified in the customizable configuration file.	Anticipate that concept codes may be stored in different OMOP tables. Ensure the customizable configuration file and data extraction script can accommodate querying different tables based on user specifications.
C4 (Data Preparation): Lack of standardized vocabulary mappings for reasons for admission	<p>The vocabularies used for reasons for admission did not have mappings to standardized OMOP vocabularies.</p> <p>This was an OMOP CDM related challenge that related to coverage analysis of vocabularies in the generic harmonization process.</p>	The data extraction scripts for reasons for admission could not be shared between data sources.	The data extraction script was adapted so that reasons for admission were extracted separately for each data source.	Recognize when standardized extraction or federated analysis is not possible and use separate data extraction scripts per source. Standardize data formats as quickly as possible to resume federated analysis with shared scripts
C5 (Analysis plan): Lack of ICU identifiers in one data source	<p>The 2 data sources applied different measures to enhance patient anonymity. Hospital identification was removed in NICE when loading data to the OMOP database.</p> <p>This challenge was a result of data storage decisions and was not related to the OMOP CDM.</p>	SMR reporting at ICU level was not possible.	Data were aggregated to provide SMRs at the quality registry level.	Communicate data requirements between data sources before drafting the analysis plan. Confirm data availability early, using shared scripts if needed.

(continued)

Table 2. (continued)

Challenge (Category)	Description	Impact	Solution	Recommendation
C6 (Analysis plan): Variation in vocabularies used to collect and store reason for ICU admission in the OMOP CDM	NICE stores reasons for admission as APACHE IV diagnosis codes. CCAA stores reasons for admission using SNOMED CT terms. This challenge was caused by variation in data collection methods and was not directly related to the OMOP CDM.	A shared analysis script could not be used to calculate APACHE II probability of mortality.	Separately from the federated analysis, an additional data source-specific mapping script was developed to map the reasons for admission to the APACHE II diagnostic codes. The APACHE II probability of mortality was then calculated using a shared script.	Recognize when standardized extraction or federated analysis is not possible and use separate data extraction or analysis scripts per source. Standardize data formats as quickly as possible to resume federated analysis with shared scripts.
C7 (Data interpretation): Different causes of missingness	The causes of missing data were different between the 2 data sources. This challenge was a result of variations in clinical practice, and resource availability between the NICE and CCAA sites, and was not related to the OMOP CDM.	The ability of APACHE II to predict mortality was impacted by degree and pattern of missingness.	The analysis was adapted to be context-sensitive regarding the meaning of missing data. In NICE data, missing data were handled as being within the normal range, as described in the APACHE II paper, while multiple imputation was used for CCAA data.	If data interpretation differs between sources, apply appropriate methods for each rather than enforcing a single approach.

Abbreviations: APACHE, Acute Physiology and Chronic Health Evaluation; C1-7, Challenge 1-7; CCAA, Critical Care Asia-Africa; ICU, Intensive Care Unit; NICE, National Intensive Care Evaluation; OHDSI, Observational Health Data Sciences and Informatics; OMOP CDM, Observational Medical Outcomes Partnership Common Data Model; PL/pgSQL, Procedural Language/PostgreSQL (used by PostgreSQL); SMR, Standardized Mortality Ratio; SQL, Structured Query Language; T-SQL, Transact-SQL (used by Microsoft SQL Server).

hospitals. Therefore, the analysis was adapted to report SMRs per quality registry instead of per ICU.

Data interpretation differences (C7)

Missing data were interpreted differently for each data source, due to differences in data collection context. For NICE, missing values were presumed to have been considered normal by clinicians and were thus handled as being within the normal clinical range for that specific variable, as described in the original APACHE II publication. This resulted in that component of the APACHE II score having a weight of “0,” indicating that the patient’s value for that variable did not contribute to an increased risk of death. For CCAA data, missing values could stem from resource constraints rather than clinician assumptions of normality. Therefore, multiple imputation with chained equations was used to impute missing values in the CCAA dataset.^{33,42} Multiple imputation for CCAA was performed using Predictive Mean Matching.^{43–45} Predictors were quality registry, ICU survival status, ICU length of stay, and the APACHE II component variables. Multiple imputation was performed 30 times, with 100 iterations per imputation, to produce 30 complete datasets in which missing values for each APACHE II component variable were imputed. The APACHE II probability of mortality was calculated separately per patient on each imputed dataset and the means over the imputed values were reported as median (IQR). For the SMRs, the expected number of deaths per quality registry was calculated separately on each imputed dataset. These were then combined using Rubin’s rules, and separate funnel plots were created using the point estimate and the upper and lower 95% confidence intervals, see [Figure 2](#).⁴³

Discussion

This study demonstrates the feasibility of using quality registry data to perform a federated data analysis to benchmark clinical outcomes internationally. It describes the steps undertaken to perform federated analysis and explores the challenges associated with both the federated analysis and the interoperability of datasets following implementation of the OMOP CDM.

Challenges C2, C3, and C6 in [Table 2](#) are directly linked to the flexibility of the OMOP CDM and its underlying terminology systems, which aim to harmonize data from diverse sources for research.¹⁷ OHDSI offers guidance, but when data do not exactly fit the model, OMOP CDM allows users flexibility in implementation. For example, OMOP CDM users are allowed to categorize comorbidities as observations, rather than conflating them with other medical conditions. The analysis script needs to accommodate these possible variations. As CDMs are increasingly used internationally, collaboration among researchers developing quality registries could enhance standardization in CDM implementation and the application of terminologies. Although achieving a common minimal dataset may be impractical, improving data findability and interpretability through the publication of data dictionaries and mapping frameworks could enhance future analyses.⁴⁶ Guidelines such as the FAIRs metrics^{46,47} and initiatives like the LOGIC consortium³ already advocate for greater transparency in reporting data structures, publicly available analysis scripts, and metadata.

Pre-empting OMOP CDMs challenges of variation in implementation, OHDSI developed the web application “ATLAS,” enabling users to group concept codes into a single clinical entity.⁴⁸ Our study group sought to use this tool in this analysis, but ATLAS currently supports only clinical characterization, population-level estimation, and patient-level prediction, not other study designs such as the benchmarking used in this analysis.⁴⁹ Potential ATLAS users should be aware that installation requires a substantial level of technical expertise and possibly support from engineers familiar with the software. While collaboration between healthcare, research, and information technology sectors is growing, expertise in these areas remains limited, particularly in low and lower-middle-income healthcare research teams. This study group consisted of clinicians and data scientists, with some support from software engineers. This study reinforces the need for investment in these disciplines.

Challenge C1 was a consequence of OMOP CDM’s support for different types of SQL servers to host the database. The OHDSI community recognize that OMOP CDM’s utility is improved by its ability to support existing server structures. As many quality registries and EHRs are already operational, this flexibility allows curators to apply the CDM to existing datasets. However, this required development of analysis scripts to support different SQL dialects for federated analysis. While OHDSI offers a tool to translate queries from T-SQL dialect to other dialects, translation between dialects without using T-SQL is not supported. Researchers conducting federated analysis should be aware of these current limitations during analysis script development.

Missing data were handled differently for each data source, due to differences in presumed mechanisms by which the data were generated within the respective clinical contexts. For NICE, missing values for laboratory and physiological measures were presumed to have been considered normal by clinicians and were thus imputed using normal value imputation. For CCAA, it is understood that missing values may stem from resource constraints influencing access and use of tests, and therefore, multiple imputation was used. This challenge was not related to the OMOP CDM, although the variation in solutions highlights the necessity of considering data generation contexts and availability of metadata before developing the study protocol and analysis plan.

OHDSI recommends performing network studies by creating an analysis script using a single OMOP database and then running it on other datasets within the OHDSI data network.⁵⁰ However, this study found this method, corresponding to the naive analysis initially attempted, to be infeasible, as there were discrepancies in data generation processes and availability among the data sources. An alternative approach, where authors first share a protocol and list of required metadata with collaborators, then edit the protocol, analysis plan, and analysis script based on metadata received, may mitigate some of these challenges. The “metadata” shared for each data source should include data dictionaries, concept mappings, results of OHDSI database characterization and data quality checks, and proportions of missingness for variables, along with any study or protocol specific information required. Other studies seeking to perform federated analysis emphasized the importance of assessing data availability, coverage, and quality before planning the analysis.^{7,51} The studies additionally recommended that collaborators validate

OMOP concept code lists before analysis to mitigate the challenge of varied concepts codes.⁵²

The use of real-world data for research is becoming increasingly important to gain insights into patient care and to gain knowledge related to the diagnosis, treatment, and prevention of diseases. Critical care services are increasingly at the forefront of response to disease, environmental, and political emergencies. Quality registry data have an important role in improving services, and standardization of quality registry data is making international analysis possible. This study is one of the first to use the OMOP CDM for federated analysis between international quality registries. Federated analysis can be a valuable tool for enabling research studies using existing accessible healthcare databases and can provide a mechanism for increasing interoperability. OMOP CDM can contribute to this under the condition that required metadata are discussed among network collaborators prior to the analysis. Future research should focus on how to identify challenges before development of the analysis, and how to mitigate them.

Acknowledgments

We thank all hospitals participating in the NICE registry for their efforts to provide complete and high-quality data.

CCAA is a collaborative of registries, of which this work is only possible by the work of the clinical and research teams working in each registry's partner hospitals.

Collaborators from Collaboration for Research, Implementation and Training in Critical Care—Asia are listed by registry in alphabetical order (National and site leads are in **bold**):

AFGHANISTAN (Registry for Intensive Care in Afghanistan—RICA): Roya Afzali (Malalai Maternity Hospital, Kabul), Noorullah Ahmadzai (Nangarhar Regional Specialization Hospital, Nangarhar), Mirwais Azizi (Mohammad Ali Jennah Hospital, Kabul), Nasibullah Barukzai (Indira Gandhi Institute of Child Health, Kabul), Maryam Barukzay (Indira Gandhi Institute of Child Health, Kabul), Naqibullah Danish (Nangarhar Regional Specialization Hospital, Nangarhar), Maliha Farooq (Herat Regional Hospital, Herat), **Maryam Shamal Ghalib** (Wazir Akbar Khan Hospital, Kabul), Owais Urhman Ghalib (Jamhoriyat Hospital, Kabul), Bibi Fazila Habibi, (Shefajo Hospital, Kabul), Hussain Hussaini (Ariana medical Complex, Kabul), Bibi Zainab Haidari (Rabia Balkhi Hospital, Kabul), Rahim Mazloomiyar (Ariana Medical Complex, Kabul), Shoaib Mirzada (Specialized & Tertiary Hospital 102 Beds, Khaira Khane, Kabul), Obaid Urahman Mohammadi (Ibn-e-Sina Ajal Hospital, Kabul), Meher Negar (Wazir Mohammad Akbar Khan National Hospital, Kabul), Bahar Nadim (Malalai Maternity Hospital, Kabul), Bibi Nasa Niazi (Isteqlal Hospital, Kabul), Ahmad Seyar Quraishi (Indira Gandhi Chlid Hospital, Kabul), Abdul Majid Rahimi (Afghan-Japan National Specialty & Tertiary Communicable Diseases Hospital, Kabul), Muhammad Dawood Safi (Indira Gandhi Institute of Child Health, Kabul), Muhammad Hamid Rahimi Safi (Shaikh Zayed Hospital, Kabul), Guldad Khan Saifi (Nangarhar Regional Specialization Hospital, Nangarhar), Asia Shamal (Malalai Maternity Hospital, Kabul), Ahmad Zakariya Shinwary, and

Jafar Khan Suliman Khil (Nangarhar Teaching Hospital, Kabul).

BANGLADESH: Hiranmoy Dutta (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Enshad Ekramullah (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), **Aniruddha Ghose** (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Md Hassanuzzaman (Department of Neurology, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Muna Islam (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Mahabubul Alam Khondokar (Department of Neurology, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Md Abdur Rahim (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Md Harun Or Rashid (Department of Anesthesia & ICU, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Md Abdus Sattar (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Abdullah Abu Sayeed (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Sarkar Shoman (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Md. Rezaul Hoque Tipu (Department of Anesthesia & ICU, Chittagong Medical College & Hospital, Chattogram, Bangladesh), **Rabiul Alam Md Erfan Uddin** (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), Mohammed Jashim Uddin (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh), and ASM Zahed (Department of Medicine, Chittagong Medical College & Hospital, Chattogram, Bangladesh).

ETHIOPIA: **Menbeu Sultan** (St Paul's hospital Millenium Medical College, Addis Ababa) and **Ayalew Zewdie** (St Paul's hospital Millenium Medical College, Addis Ababa).

GHANA: **John Amuasi** (Global Health and Infectious Disease Research Group, Kumasi Centre for Collaborative Research in Tropical Medicine; Global Health Department, School of Public Health, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; Department of Implementation Research, Bernhard Nocht Institute of Tropical Medicine, Hamburg, Germany), **Joe Bonney** (Global Health and Infectious Disease Research Group, Kumasi Centre for Collaborative Research in Tropical Medicine; Emergency Medicine, Komfo Anokye Teaching Hospital, Kumasi, Ghana), **Moses Siaw Frimpong** (Global Health and Infectious Disease Research Group, Kumasi Centre for Collaborative Research in Tropical Medicine; Anaesthesia and Intensive Care, Komfo Anokye Teaching Hospital, Kumasi, Ghana), Ibrahim Kwaku Duah (Global Health and Infectious Disease Research Group, Kumasi Centre for Collaborative Research in Tropical Medicine, Kumasi, Ghana).

INDIA (Indian Registry of Intensive care—IRIS): All investigators from the Indian Registry of Intensive care—IRIS.

MALAYSIA: **Mohd Shahnaz Hasan** (Universiti Malaya, Kuala Lumpur), **Mohd Basri Mat Nor** (International Islamic University Malaysia, Kuantan, Pahang), and **Mohd Zulfakar**

Mazlan (School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan).

NEPAL (Nepal Intensive Care Registry Foundation—NICRF): **Sushila Paudel** (Nepal Intensive Care Research Foundation, Kathmandu), **Subekshya Luitel** (Nepal Intensive Care Research Foundation, Kathmandu), **Isha Amatya** (Nepal Intensive Care Research Foundation, Kathmandu), **Diptesh Aryal** (Nepal Intensive Care Research Foundation, Kathmandu), **Basanta Gauli** (Chitwan Medical College, Chitwan), **Praveen Giri** (Kirtipur Hospital, Kathmandu), **Kishor Khanal** (Nepal Medcity Hospital, Kathmandu), **Sushil Khanal** (Grande International Hospital, Kathmandu), **Sabin Koirala** (HAMS Hospital, Kathmandu), **Sanjay Lakhey** (B&B Hospital, Kathmandu), **Hem Raj Paneru** (Tribhuvan University Teaching Hospital, Kathmandu), **Sushila Paudel**, **Lalit Rajbanshi** (Birat Medical College and Teaching Hospital/Birat Nursing Home, Biratnagar), **Sangina Ranjit** (Dhulikhel Hospital, Dhulikhel, Kathmandu), **Yam Roka** (Neuro Cardio and Multispeciality Hospital, Biratnagar), **Pramesh Sundar Shrestha** (Om Hospital, Kathmandu), **Raju Shrestha** (B&C Hospital, Jhapa), **Pradeep Tiwari** (Civil Service Hospital, Kathmandu), **Shital Adhikari** (Chitwan Medical College, Chitwan), **Subhash Prasad Acharya** (Tribhuvan University Teaching Hospital, Kathmandu).

KENYA: **Wangari Waweru-Siika** (Aga Khan University, Nairobi).

PAKISTAN: (Pakistan Registry of Intensive Care—PRICE): **Ali Abbas** (Ziauddin Group of Hospitals, Karachi), **Aftab Ahmed** (Abbasi Shaheed Hospital, Karachi), **Afzal Ahmed** (GMMM Teaching Hospital, Sukkur), **Tanvir Alam** (Civil Hospital, Karachi), **Basit Ali** (Services Hospital, Lahore), **Basit Ali** (Jinnah Hospital, Lahore), **Liaquat Ali** (Jinnah Hospital, Lahore), **Mohsin Ali** (Liaquat University Hospital, Hyderabad), **Samad Ali** (Ziauddin Group of Hospitals, Karachi), **Sayed Muneeb Ali** (Pakistan Institute of Medical Sciences, Islamabad), **Junaid Anwar** (National Hospital & Medical Center, Lahore), **Mustehsan Bashir** (Mayo Hospital Lahore, Lahore), **Mobin Chaudhary** (Pakistan Kidney and Liver Institute, Lahore), **Kamran Cheema** (Services Hospital, Lahore), **Ahmed Farooq** (Doctors Hospital, Lahore), **Saeeda Haider** (The Indus Hospital, Karachi), **Fakhir Raza Haidri** (SIUT, Karachi), **Madiha Hashmi** (Ziauddin University, Karachi), **Noor Hassan** (Civil Hospital, Karachi), **Noor Hassan** (South City Hospital, Karachi), **Noor Hassan** (Ziauddin Group of Hospitals, Karachi), **Muhammad Hayat** (North West General Hospital, Peshawar), **Imran ul Haq** (Khyber Teaching Hospital, Peshawar), **Fivzia Herekar** (The Indus Hospital, Karachi), **Iqbal Hussain** (Pakistan Kidney and Liver Institute, Lahore), **Muhammad Ibrahim** (Peoples Medical University, Nawabshah), **Shahbaz Ikram** (Doctors Hospital, Lahore), **Muhammad Imran** (National Institute of Cardiovascular Diseases, Karachi), **Arther John** (The Indus Hospital, Karachi), **Saima Kamal** (Dow International Hospital, Karachi), **Muhammad Kamran** (North West General Hospital, Peshawar), **Osama Khalid** (Ziauddin Group of Hospitals, Karachi), **Amir Khan** (Abbasi Shaheed Hospital, Karachi), **Amir Khan** (Patel Hospital, Karachi), **Farhan Khan** (Patel Hospital, Karachi), **Farman Ali Khan** (Khyber Teaching Hospital, Peshawar), **Farman Ali Khan** (North West General Hospital, Peshawar), **Imran Khan** (Lady Reading Hospital, Peshawar), **Imran Khan** (North West General Hospital,

Peshawar), **Rashid Nasim Khan** (Darul Sehat Hospital, Karachi), **Shereen Khan** (Fatima Jinnah Chest Hospital, Quetta), **Quratul Ain Khan** (Ziauddin Group of Hospitals, Karachi), **Saleh Khaskheli** (Peoples Medical University, Nawabshah), **Zafar Iqbal Khatak** (Lady Reading Hospital, Peshawar), **Amin Khawaja** (National Institute of Cardiovascular Diseases, Karachi), **Muhammad Nasir Khoso** (South City Hospital, Karachi), **Aneela Altaf Kidwai** (Abbasi Shaheed Hospital, Karachi), **Ashok Kumar** (Ziauddin Group of Hospitals, Karachi), **Mukesh Kumar** (Ziauddin Group of Hospitals, Karachi), **Vinod Kumar** (Jinnah Post-Graduate Medical Center, Karachi), **Vinod Kumar** (National Institute of Cardiovascular Diseases, Karachi), **Irfan Malik** (Lahore General Hospital, Lahore), **Sobia Masood** (National Institute of Cardiovascular Diseases, Karachi), **Shahryar Maqsood** (Jinnah Hospital, Lahore), **Shahryar Maqsood** (Lahore General Hospital, Lahore), **Maqsood Meher** (GMMM Teaching Hospital, Sukkur), **Kashif Memon** (Liaquat University Hospital, Hyderabad), **Hafeez Muhammad** (Aziz Bhatti Shaheed Hospital, Gujrat), **Imran Muhammad** (Rehman Medical Institute, Peshawar), **Kamran Muhammad** (North West School of Medicine, Peshawar), **Nadeem Muneer** (Jinnah Post-Graduate Medical Center, Karachi), **Mazhar Ali Naqvi** (Services Hospital, Lahore), **Rehan Niazi** (Lahore General Hospital, Lahore), **Sajjad Orakzai** (Lady Reading Hospital, Peshawar), **Junaid Patel** (The Indus Hospital, Karachi), **Sumaira Qabulio** (Ziauddin Hospital Keemari, Karachi), **Muddasir Qadir** (Pakistan Kidney and Liver Institute, Lahore), **Arslan Raha-tullah** (North West General Hospital, Peshawar), **Asim Rana** (Bahria International Hospital, Lahore), **Ahmed Ranjha** (Jinnah Hospital, Lahore), **Ali Raza** (Ziauddin Group of Hospitals, Karachi), **Attaur Rehman** (Patel Hospital, Karachi), **Fawadur Rehman** (SIUT, Karachi), **Yasir Rehman** (The Indus Hospital, Karachi), **Nawal Salahuddin** (National Institute of Cardiovascular Diseases, Karachi), **Sairah Sadaf** (Sheikh Zayed Medical College, Rahim Yar Khan), **Anjum Saleem** (Sheikh Zayed Medical College, Rahim Yar Khan), **Jodat Saleem** (Lahore General Hospital, Lahore), **Rana Imran Sikandar** (Pakistan Institute of Medical Sciences, Islamabad), **Imtiaz Ali Shah** (Pakistan Institute of Medical Sciences, Islamabad), **Naseem Ali Shah** (Hameed Latif Hospital, Lahore), **Jhonsan Shahzad** (Darul Sehat Hospital, Karachi), **Muhammad Sheharyar** (Lady Reading Hospital, Peshawar), **Ilyas Shehzad** (Ziauddin Group of Hospitals, Karachi), **Arshad Taqi** (National Hospital & Medical Center, Lahore), **Moazzam Tarar** (Jinnah Hospital, Lahore), **Akash Thakrani** (Ziauddin Group of Hospitals, Karachi), **Ahmed Zia** (Patel Hospital, Karachi), **Muhammad Ashraf Zia** (Jinnah Hospital, Lahore).

SIERRA LEONE: **Eva Hanciles** (Connaught Hospital, University of Sierra Leone Teaching Hospitals Complex, Freetown), and **Luigi Pisani** (Mahidol Oxford Tropical Research Unit, Bangkok).

SOUTH AFRICA: **Yumna Beukes** (Division of Critical Care, Groote Schuur Hospital, University of Cape Town, Cape Town), **Esihle Gili** (Division of Critical Care, Groote Schuur Hospital, University of Cape Town, Cape Town), **Kieyaam Johnson** (Groote Schuur Hospital, University of Cape Town, Cape Town), **Malcolm Miller** (Division of Critical Care, Groote Schuur Hospital, University of Cape Town, Cape Town), **David Thomson** (Division of Critical Care, Groote Schuur Hospital, University of Cape Town, Cape Town).

UGANDA: **Martha Alupo** (Mulago National Referral Hospital, Kampala), **Adam Hewitt Smith** (Busitema University Faculty of Health Sciences, Mbale, Uganda and Queen Mary University of London, London, United Kingdom), **Dennis Kakaire** (Lubaga Hospital, Kampala), **Herbert Kiwalya** (Busitema University Faculty of Health Sciences, Mbale), **Joseph Kiwanuka** (Mbarara University of Science and Technology, Mbarara), **Arthur Kwizera** (Makerere University, Kampala), **Joshua Muhanguzi** (Mbarara University of Science and Technology, Mbarara), and **Cornelius Sendagire** (Makerere University, Kampala).

Author contributions

Aasiyah Rashan (Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Daniel P. Püttmann (Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Validation, Visualization, Writing—original draft, Writing—review & editing), Nicolette F. de Keizer (Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review & editing), Dave A. Dongelmans (Conceptualization, Methodology, Supervision, Writing—review & editing), Ronald Cornet (Conceptualization, Methodology, Supervision, Writing—review & editing), Otavio Ranzani (Methodology, Supervision, Writing—review & editing), Wangari Waweru-Siika (Data curation, Writing—review & editing), Matthew Smith (Supervision, Writing—review & editing), Steve Harris (Funding acquisition, Supervision, Writing—review & editing), Abi Beane (Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing—original draft, Writing—review & editing), and Ferishta Bakhshi-Raiez (Conceptualization, Data curation, Methodology, Project administration, Supervision, Validation, Writing—original draft, Writing—review & editing)

Supplementary material

[Supplementary material](#) is available at *JAMIA Open* online.

Funding

This work was supported in part by the Wellcome Trust [220211/Z/20/Z]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. A.R. is partially supported by a University College London UK Research and Innovation Centre for Doctoral Training in AI-enabled Healthcare Studentship [Grant number EP/S021612/1]. A.R., S.H., and A.B. are also supported by Wellcome Collaboration for Research, Implementation and Training in Critical Care Asia-Africa (CCAA) [Grant number WT215522/Z/19/Z]. OR is funded by the Ramón y Cajal program (RYC2023-002923-C) awarded by the Spanish Ministry of Science, Innovation and Universities (MICIU/AEI/10.13039/501100011033) and by the European Social Fund Plus (ESF+).

Conflicts of interest

The authors have no competing interests to declare.

Data availability

Access to the data used in the analysis can be granted through the submission of an extraction form, which will be reviewed by a committee before access is granted. Extraction forms can be submitted here:

- NICE (website): <https://www.stichting-nice.nl/extractie-verzoeken.jsp>
- CCAA (email): DAC@nicsslk.com

A list of R packages used in the analysis is available in [Table S3](#). Data dictionaries, analysis scripts, and APACHE IV to APACHE II mapping tables are available on the figshare website for replication purposes. The Digital object identifier (DOI) for each item is listed below:

Data dictionaries:

<https://doi.org/10.6084/m9.figshare.26268322.v1>

Analysis scripts:

<https://doi.org/10.6084/m9.figshare.25837042.v2>
<https://github.com/aasiyahrashan/benchmarking-OMOP/releases/tag/v1.0.0>
<https://github.com/aasiyahrashan/SeverityScoresOMOP/releases/tag/v1.0>

NICE mapping table:

<https://doi.org/10.6084/m9.figshare.25837876.v1>

CCAA mapping table:

<https://doi.org/10.6084/m9.figshare.26268361.v1>

Ethics approval

The institutional research board (IRB) of the Amsterdam University Medical Centre reviewed the research proposal and waived the need for informed consent on September 6, 2023 (IRB number METC-number 2023.0259). The Oxford Tropical Research Ethics Committee (OxTREC) reviewed the research proposal and waived the need for informed consent on February 12, 2024.

References

1. National Intensive Care Evaluation. Dutch National Intensive Care Evaluation (NICE) registry [Internet]. Accessed August 15, 2024. <https://www.stichting-nice.nl>
2. Pisani L, Rashan T, Shamal M, Collaboration for Research, Implementation and Training in Critical Care—Asia Investigators, et al. Performance evaluation of a multinational data platform for critical care in Asia. *Wellcome Open Res.* 2021;6:251.
3. Dongelmans DA, Pilcher D, Beane A, et al. Linking of global intensive care (LOGIC): an international benchmarking in critical care initiative. *J Crit Care.* 2020;60:305-310.
4. Salluh JIF, Quintairos A, Dongelmans DA, Linking of Global Intensive Care (LOGIC) and Japanese Intensive care Patient

- Database (JIPAD) Working Group, et al. National ICU Registries as enablers of clinical research and quality improvement. *Crit Care Med.* 2024;52:125-135.
5. Wortel SA, Bakhshi-Raiez F, Sjoie WGTS, van der Zwan EPA, de Keizer NF, Dongelmans DA. The role of the Dutch National Intensive Care Evaluation registry during the COVID-19 pandemic. *Neth J Crit Care.* 2022;30:152-155.
 6. Garcia-Gallo E, Merson L, Kennon K, ISARIC Clinical Characterization Group, et al. ISARIC-COVID-19 dataset: a prospective, standardized, global dataset of patients hospitalized with COVID-19. *Sci Data.* 2022;9:454.
 7. Kroes JA, Bansal AT, Berret E, et al. Blueprint for harmonising unstandardised disease registries to allow federated data analysis: prepare for the future. *ERJ Open Res.* 2022;8:00168. 2022.
 8. Tacconelli E, Gorska A, Carrara E, et al. Challenges of data sharing in European Covid-19 projects: a learning opportunity for advancing pandemic preparedness and response. *Lancet Reg.* 2022;21:100467.
 9. Freeman EE, McMahon DE, Hruza GJ, et al. International collaboration and rapid harmonization across dermatologic COVID-19 registries. *J Am Acad Dermatol.* 2020;83:e261-e266.
 10. Quintairos A, Pilcher D, Salluh JIF. ICU scoring systems. *Intensive Care Med.* 2023;49:223-225. <https://doi.org/10.1007/s00134-022-06914-8>.
 11. Lenert L, McSwain BY. Balancing health privacy, health information exchange, and research in the context of the COVID-19 pandemic. *J Am Med Inform Assoc.* 2020;27:963-966.
 12. Zwitter A, Gstrein OJ. *Big Data, Privacy and COVID-19: Learning from Humanitarian Expertise in Data Protection.* Springer; 2020:1-7.
 13. Bradford L, Aboy M, Liddell K. COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes. *J Law Biosci.* 2020;7:lsaa034.
 14. Casaletto J, Bernier A, McDougall R, Cline MS. Federated analysis for privacy-preserving data sharing: a technical and legal primer. *Annu Rev Genomics Hum Genet.* 2023;24:347-368.
 15. Kuchinke W, Aerts J, Semler SC, Ohmann C. CDISC standard-based electronic archiving of clinical trials. *Methods Inf Med.* 2009;48:408-413.
 16. THIN. The Health Improvement Network [Internet]. Accessed August 15, 2024. <https://www.the-health-improvement-network.com/>
 17. 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;19:54-60.
 18. Sentinel Initiative. Sentinel common data model [Internet]. Accessed August 15, 2024. <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model>
 19. Roblin DW, Rubenstein KB, Tavel HM, et al. Development of a common data model for a multisite and multiyear study of virtual visit implementation: a case study. *Med Care* 2023;61:S54-S61.
 20. Weber GM, Klann J, Mendis M, Murphy S, Potenzzone R, Rice P. i2b2 Common data model documentation v1.0. [Internet]. 2021. Accessed August 15, 2024. <https://community.i2b2.org/wiki/display/BUN/i2b2+Common+Data+Model+Documentation>
 21. Health Level 7. HL7 FHIR [Internet]. Accessed August 15, 2024. <https://www.hl7.org/fhir/>
 22. Puttmann D, De Keizer N, Cornet R, Van Der Zwan E, Bakhshi-Raiez F. FAIRifying a Quality Registry using OMOP CDM: challenges and solutions. *Stud Health Technol Inform.* 2022;294:367-371.
 23. Biedermann P, Ong R, Davydov A, et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. *BMC Med Res Methodol.* 2021;21:238.
 24. Yu Y, Zong N, Wen A, et al. Developing an ETL tool for converting the PCORnet CDM into the OMOP CDM to facilitate the COVID-19 data integration. *J Biomed Inform.* 2022;127:104002.
 25. Blacketer C, Voss E. The Book of OHDSI: Observational Health Data Sciences and Informatics. Extract transform load [Internet]. 2019. Accessed August 15, 2024. <https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html>
 26. OHDSI. Common Data Model GitHub Repository [Internet]. Accessed August 15, 2024. <https://github.com/OHDSI/CommonDataModel>
 27. National Intensive Care Evaluation. NICE data dictionary [Internet]. Accessed August 15, 2024. <https://www.stichting-nice.nl/dd>
 28. DeFalco F, Ryan P, Schuemie M, et al. *Achilles: Achilles Data Source Characterization.* R package version 1.7.2., 2023.
 29. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. *J Am Med Inform Assoc.* 2021;28:2251-2257. <https://doi.org/10.1093/jamia/ocab132>.
 30. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373-383.
 31. SNOMED International. SNOMED CT National Library of Medicine [Internet]. Accessed August 15, 2024. <https://www.nlm.nih.gov/healthit/snomedct/index.html>
 32. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II. *Crit Care Med.* 1985;13:818-829.
 33. Haniffa R, Pubudu De Silva A, Weerathunga P, et al. Applicability of the APACHE II model to a lower middle income country. *J Crit Care.* 2017;42:178-183.
 34. Czajka S, Ziębińska K, Marczenko K, Posmyk B, Szczepańska AJ, Krzych ŁJ. Validation of APACHE II, APACHE III and SAPS II scores in in-hospital and one year mortality prediction in a mixed intensive care unit in Poland: a cohort study. *BMC Anesthesiol.* 2020;20:296-298.
 35. Sungono V, Hariyanto H, Soesilo TEB, et al. Cohort study of the APACHE II score and mortality for different types of intensive care unit patients. *Postgrad Med J.* 2022;98:914-918.
 36. Verburg IW, Holman R, Peek N, Abu-Hanna A, de Keizer NF. Guidelines on constructing funnel plots for quality indicators: a case study on mortality in intensive care unit patients. *Stat Methods Med Res.* 2018;27:3350-3366.
 37. Schuemie M, Suchard M. SqlRender: rendering parameterized SQL and translation to dialects. [Internet]. Accessed August 15, 2024. <https://github.com/OHDSI/SqlRender>
 38. R Foundation for Statistical Computing: CRAN Team. R: a language and environment for statistical computing. R version 4.1.2 (2021-11-01) ed [Internet]. Accessed August 15, 2024. Index of/ bin/windows/base/old/4.1.2 (r-project.org)
 39. Henke E, Zoch M, Peng Y, Reinecke I, Sedlmayr M, Bathelt F. Conceptual design of a generic data harmonization process for OMOP common data model. *BMC Med Inform Decis Mak.* 2024;24:58.
 40. The PostgreSQL Global Development Group. PostgreSQL: the world's most advanced open source relational database [Internet]. Accessed August 15, 2024. <https://www.postgresql.org/>
 41. Microsoft. Transact-SQL reference (Database Engine) [Internet]. Accessed August 15, 2024. <https://learn.microsoft.com/en-us/sql/t-sql>
 42. Gamage D, Beane A, Venkataraman R, et al. Validation of a simplified risk prediction model using a cloud based critical care registry in a lower-middle income country. Network for Improving Critical Care Systems and Training Aasiyah Rashan Network for Improving Critical care Systems and Training; 2020.
 43. Buuren S. Flexible imputation of missing data. Multiple imputation in a nutshell [Internet]. Bookdown. 2018. Accessed August 15, 2024. <https://stefvanbuuren.name/fimd/sec-nutshell.html>.
 44. Allison P. Imputation by predictive mean matching: promise & peril [Internet]. 2015. Accessed August 15, 2024. <https://statisticalhorizons.com/predictive-mean-matching/>

45. Buuren S. Flexible imputation of missing data. Predictive mean matching [Internet]. Bookdown. 2018. Accessed August 15, 2024. <https://stefvanbuuren.name/fimd/sec-pmm.html>
46. Devaraju A, Huber R, Mokrane M, et al. FAIRsFAIR data object assesment metrics. Zenodo; 2022. <https://doi.org/10.5281/zenodo.6461229>
47. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.
48. OHDSI. ATLAS GitHub repository [Internet]. Accessed August 15, 2024. <https://github.com/OHDSI/Atlas>
49. Madigan D. The Book of OHDSI: Observational Health Data Sciences and Informatics. Data analytics use cases [Internet]. 2019. Accessed August 15, 2024. <https://ohdsi.github.io/TheBookOfOhdsi/DataAnalyticsUseCases.html>
50. Falconer T, Chen R, Hripcsak G. Running an OHDSI Network Study [Internet]. 2016. Accessed August 15, 2024. https://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=research:nci-ohdsi_instruction_manual2.pdf
51. Kiwuwa-Muyingo S, Todd J, Bhattacharjee T, Taylor A, Greenfield J. Enabling data sharing and utilization for African population health data using OHDSI tools with an OMOP-common data model. *Front Public Health*. 2023;11:1116682.
52. Abedtash H. Using OHDSI Data Network for capturing real-world evidence: our experience with a multi-country study on an obese and overweight cohort. In: *Observational Health Data Sciences and Informatics Symposium*. Washington, DC: Observational Health Data Sciences and Informatics; 2019.