International Journal of Population Data Science

Journal Website: www.ijpds.org

IMECCHI-DATANETWORK: empowering knowledge generation through international data network

Le Pogam, Marie Annick¹, Metcalf, Amy², Johnsen, Søren Paaske³, Hider, Phil⁴, Patel, Alka⁵, Kim, Hongsoo⁶, Carlini, Emanuele⁷, Perego, Raffaele⁷, Quan, Hude⁸, and Gini, Rosa⁹

¹Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital (CHUV), Lausanne

²Alberta Children's Hospital Research Institute, University of Calgary, Calgary

³Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus

⁴Department of Population Health, University of Otago, Christchurch

⁵O'Brien Institute for Public Health and Alberta Health Services, Calgary

⁶Graduate School of Public Health, Seoul National University, Seoul

⁷Istituto di Scienza e Tecnologie dell'Informazione (ISTI), National Research Council (NRC), Pisa

⁸Department of Community Health Sciences, University of Calgary and O'Brien Institute for Public Health, Calgary

⁹Agenzia regionale di sanità della Toscana, Florence

Objectives

The International Methodology Consortium for Coded Health Information (IMECCHI), an international collaboration of health services researchers, launched the IMECCHI-DATANETWORK initiative in October 2015. Its main objective is to enable replication of observational studies across countries through a distributed data infrastructure.

Approach

In a distributed data infrastructure, individual raw data are not shared. Instead, data are converted locally using a common data model (CDM) and loaded into a common software for data processing and analysis. Whenever a study protocol is agreed upon and ethically approved, an ad hoc procedure is programmed - using that software - including the data processing steps needed to create the analytical dataset from the CDM: record linkage, case selection, sampling, matching, etc. The procedure is then shared and locally run by each partner to generate an analytical dataset of integrated data. Analytical datasets may then be shared and pooled for statistical analyses.

Results

Six partners of the IMECCHI collaboration, located in countries across 4 continents (Canada, Denmark, Italy, New Zealand,

South Korea, and Switzerland), currently participate in the initiative. They first conducted a survey to describe the origin, content, completeness and main attributes of each table in their original databases. Based on the results of the survey, a CDM was created, encompassing 4 tables of coded or structured data to be linked at the individual level using a common personal identifier: (1) characteristics of the subjects with dates of birth and death; (2) hospital discharge summaries with diagnosis and procedure codes, and admission, discharge and procedure dates; (3) drug dispensing information with date of dispensing, drug name, duration of the amount of active principle according to the Defined Daily Dose of the World Health Organization; (4) causes of death. In each table, additional attributes describe the coding systems in which the other attributes are coded. Using such specific attributes facilitates interoperability across multiple coding systems. The open source Java-based software, The-Matrix, which operates on flat csv files using a domain-specific programming language, was chosen to embed the ad hoc procedure.

Conclusion

Within the IMECCHI-DATANETWORK initiative, databases from various countries will be locally converted in a CDM which will facilitate study replication in a distributed fashion while granting interoperability across coding systems. Through such international data networks, data are empowered for creating re-

*Corresponding Author:

Email Address: marie-annick.le-pogam@chuv.ch (M. Le Pogam)

sults which are generalizable to multiple countries. Cross-border data sharing and international comparisons are also facilitated.