International Journal of Population Data Science

Journal Website: www.ijpds.org

EUROlinkCAT: Common Data Model

Loane, M^{1*}, Densem, J², Morris, J³, and Tan, J³

¹Ulster University

²Bio-Medical Computing Ltd

³Queen Mary University London

Background

Over 130,000 children are born in Europe every year with congenital anomalies which are a major cause of infant mortality, childhood morbidity and long-term disability. A European data linkage study (EUROlinkCAT) aims to investigate the health and educational outcomes of children up to 10 years of age with congenital anomalies, born between 1995 and 2014. While congenital anomaly data including information on potential risk factors are standardised across the EUROCAT network, information on mortality, morbidity and educational outcomes are not.

Objective

To create a common data model that transforms key variables in local databases to standardised formats enabling data on health and educational outcomes to be pooled and analysed across multiple registries.

Method

Twenty-two EUROCAT registries in fourteen countries are participating in the study. Each registry records uniformly coded data on cases of congenital anomaly registered in their local population using the EUROCAT Data Management Program. The registries will link their congenital anomaly data to their local mortality, hospital discharge, prescriptions and educational data. The linked individual case data cannot leave the local institution or “safe haven” environment, therefore verification and validation of all derived variables, data transformations and proxy variables must be performed locally.

Findings

Creating a common data model is challenging as there are diverse coding classification systems, languages, healthcare and educational systems in Europe. As with many administrative datasets, the common data model is based on coded data rather than the often richer “free text” information.

Conclusion

The use of administrative datasets across Europe enables pooling of data on rare outcomes and allows hypotheses on the health and education of children to be investigated. However, a common data model must be applied to ensure that data from multiple sites conform to a standard format.

*Corresponding Author:

Email Address: ma.loane@ulster.ac.uk (M Loane)