

Integrating electronic health records from different sources across the UK: lessons from a record linkage study

Bandyopadhyay, A^{1*}, Tingay, K¹, Akbari, A¹, Griffiths, L², Cortina-Borja, M², Bedford, H², Walton, S², Dezateux, C³, and Lyons, RA¹

¹Swansea University Medical School, Swansea, UK

²Great Ormond Street Institute of Child Health, University College London, London, UK

³Centre for Primary Care and Public Health, Queen Mary University London

Background

Harmonisation of different data sources from various electronic health records (EHRs) across systems enhances the potential scope and granularity of data available to health data research.

Objective

To describe data harmonisation of routine electronic health-care records in Wales and Scotland linked to a UK longitudinal birth cohort, the Millennium Cohort Study (MCS).

Methods

Comparable secondary care data was linked, with parental consent, to MCS information for 1838 and 1431 children participating in MCS and residing in Wales and Scotland, by assigning, respectively, unique Anonymised Linkage Fields to person-based records in the privacy protecting Secure Anonymised Information Linkage (SAIL) databank at Swansea University, and by the National Health Service (NHS) Information Standards Division. Survey and non-response weights were created to account for the clustered sample, sample attrition and consent to linkage. Heterogeneous variables from the Patient Episode Dataset for Wales, Emergency Department Data Set for Wales, Scottish Medical Record 01 and Accident and Emergency dataset for Scotland were harmonised enabling data to be pooled and standardised for research.

Findings

Overall linkage to harmonised health care data was achieved for 98.9% (99.9% for Wales and 97.6% for Scotland) of consented MCS participants. 66% of children experienced at least one hospital admission (total 5747 hospital admissions) up to their 14th birthday, while 60% attended A&E departments at least once (total 5221 attendances) between their 9th and 14th birthday. We managed date granularity by generating random dates of birth, standardising periods of data collection, identifying inconsistencies and then mapping and bridging differences in definitions of periods of care across countries and datasets.

Conclusions

Combining and harmonising data from multiple sources and linking them to information from a longitudinal cohort create useful resources for population health research. These methods are reproducible and can be utilised by other researchers and projects.

*Corresponding Author:

Email Address: a.bandyopadhyay@swansea.ac.uk (A Bandyopadhyay)

