Cronfa - Swansea University Open Access Repository

_____

This is an author produced version of a paper published in:
International Journal of Population Data Science

_____

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa44392

_____

Paper:

_____

Challenges of identifying children with Cystic Fibrosis to explore inequalities in social and health outcomes, using multiply linked data sources.

Griffiths, R[1], Akbari, A[2], Schlueter, D[3], Taylor-Robinson, D[4], and Tucker, D[5]

[1]Farr Institute, Swansea University

[2]Health Data Research UK - Wales and Northern Ireland, Swansea University Medical School

[3]Lancaster University

[4]Liverpool University

[5]Public Health Wales

Introduction

Cystic fibrosis (CF) is the most common life-limiting inherited disease in white populations, with most patients dying prematurely from respiratory failure. As it is rare, it is important to reduce misclassification. We therefore aimed to assess how well CF children could be identified across routine data in Wales.

Objectives and Approach

Data from the Secure Anonymised Information Linkage (SAIL) databank, identified children with CF from 1998 to 2016, within hospital, General Practice (GP), and the Welsh Congenital Abnormality Register (CARIS), which uses new born screening to identify congenital abnormalities, including CF. The International Classification of Diseases (ICD10) E84 was used to identify CF children in both hospital and CARIS data, with READ codes used in the GP data (approximately 80% coverage of Wales). The data was linked using anonymised linking fields and matching rates analysed, as unmatched records in linked data can reduce the utility of the data for epidemiological studies.

Results

352 cases were identified in total, with 158 matched across all three datasets over an 18 year period (9-19 cases per year). The Welsh rate from the disease registry is 12–14. Since CF is a severely debilitating condition, a greater match was expected. This prompted further investigation of cases which appeared in only one dataset, as these seemed least likely to be true cases. In the 'CARIS only' data, 79% of the admissions, were found coded for respiratory, digestive and health complications not for an E84 CF condition. In the 43 cases in 'GP only' data, and the 19 in 'hospital only' data the events indicated the possible late presentation of CF by older children or children with very mild CF phenotypes.

Conclusion/Implications

Cases of rare diseases like CF can be identified in routine data. Linking across multiple datasets, particularly with specialist datasets like CARIS, help identify potentially misclassified cases. This increases confidence in the data. Future work will include the CF registry, permitting checks against a gold standard data resource.