

International Journal of Population Data Science



Swansea University
Prifysgol Abertawe

Journal Website: www.ijpds.org

Real world performance of privacy preserving record linkage

Irvine, K¹, Smith, M¹, de Vos, R¹, Brown, A², Ferrante, A², Boyd, J², and Thackway, S³

¹Centre for Health Record Linkage, NSW Health

²Curtin University

³NSW Ministry of Health

Introduction

Privacy preserving record linkage (PPRL) using encoded or hashed data has potential to enable large-scale record linkage of previously inaccessible data. With limited real-world evaluation and implementation of PPRL at scale it is challenging for linkage practitioners to judiciously balance data protection with the accuracy and usability of linked datasets.

in Australia but international collaborations are encouraged to more rapidly develop the evidence base and tactical approaches to support real world implementations.

Objectives and Approach

We evaluated the performance of PPRL techniques using Bloom filters for linkage of data across primary and secondary care settings. This technique limits the need to disclose personal information for linkage activities. Primary care data included 272,202 records from 16 general practices in NSW. This was linked to 42.8 million records from a 7 year series of emergency presentations, hospitalisations and death registrations. For the purpose of evaluation, personal information was encoded within the data linkage centre. The quality of PPRL linkage was assessed against the true match status based on a gold standard probabilistic linkage using full personal identifiers.

Results

Compared to the gold standard probabilistic linkage using full personal identifiers, the PPRL techniques produced quality metrics of precision, recall and F measure in excess of 0.90. When configured to leverage pre-existing links between emergency department, hospital and mortality data, quality metrics around 0.98-0.99 were achieved. Lower rates of linkage quality were associated with missing demographic information and some residual variation in linkage quality across practices was observed.

Conclusion/Implications

PPRL using Bloom filters is a promising technique for achieving high quality linkage across primary and secondary care in Australia. Further evaluation will assess scalability and quality

