



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A tool to improve the efficiency and reproducibility of research using electronic health record databases

Citation for published version:

Al Sallakh, M, Rodgers, S, Lyons, R, Sheikh, A & Davies, G 2018, 'A tool to improve the efficiency and reproducibility of research using electronic health record databases', *International Journal of Population Data Science*, vol. 3, no. 2, pp. 71.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

International Journal of Population Data Science

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A tool to improve the efficiency and reproducibility of research using electronic health record databases

Al Sallakh, M^{1*}, Rodgers, S¹, Lyons, R¹, Sheikh, A², and Davies, G¹

¹Swansea University Medical School

²Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh

Background

Interrogation of electronic health record databases often involves time-consuming, manual, repetitive work in developing database queries. We developed a tool to automate this process.

Methods

We identified elementary approaches to query primary care data from the Secure Anonymised Information Linkage databank of Wales. We designed a web-based query builder that allows using combinations of these approaches as 'building blocks' to query complex variables. We created an R programme to automatically generate and execute the corresponding Structured Query Language queries.

Results

The tool allows data extraction using combinations of the following methods: event count (e.g., asthma prescriptions); code/date of earliest/latest event; code/date/value of the event of maximum/minimum value; and frequency of temporally constrained events. Query intervals could be fixed, dynamic, or individualised. The tool integrates with a codeset repository. Data extraction procedures and codesets are saved on a web server as versioned, shareable, and citable objects.

Conclusion

This versatile tool allows rapid and complex data extraction with minimal to no programming skills, reduces human errors, and improves research transparency and reproducibility.

Funding/Support

Health and Care Research Wales, ABMU Health Board, AUK-CAR (AUK-AC-2012-01), Farr Institute of Health Informatics Research (MR/K006525/1-MR/K007017/1).

*Corresponding Author:

Email Address: m.a.alsallakh@swansea.ac.uk (M Al Sallakh)

