

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

Al Sallakh, M<sup>1\*</sup>, Rodgers, S<sup>1</sup>, Lyons, R<sup>1</sup>, Sheikh, A<sup>2</sup>, and Davies, G<sup>1</sup>

<sup>1</sup>Swansea University Medical School

<sup>2</sup>Usher 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](mailto:m.a.alsallakh@swansea.ac.uk) (M Al Sallakh)

