Linked routine data to enhance health-economics analysis

EDITORIAL

Muhammad Jami Husain [1], Sinead Brophy [2], Stefan Siebert [3], Ceri J. Phillips [4]

[1] Keele Management School, Keele University, Staffordshire

[2] College of Medicine, Swansea University, Wales

[3] Institute of Infection Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow

[4] College of Human and Health Sciences, Swansea University, Wales

CONTEXT

The major sources of data in the healthcare research include the data generated from randomised controlled trials (RCTs), observational studies based on patient derived questionnaire data, and routinely assembled data abstracted from the primary and secondary care patient record – popularly known as routine data. Data generated from these routes are usually used as stand-alone datasets, and alongside non-health data such as demographic, socio-economic, and geographical data. The health sector analyses are enriched when the patient-level data generated from these different sources are linked.

In a recent paper, titled "HERALD (Health Economics using Routine Anonymised Linked Data)" [1], we demonstrate the potential methodological advantages in the conduct of health economics analyses using patient-derived questionnaire data linked with routinely collected information and secondary care clinical datasets available in Wales, United Kingdom, with examples from a research cohort.

LINKED ROUTINE DATA IN WALES: THE SAIL DATABANK AND THE PAS COHORT

The Health Information Research Unit (HIRU) at the College of Medicine at Swansea University has set up the Secure Anonymised Information Linkage (SAIL) databank [2,3]. The SAIL databank brings together and links a wide range of person-based data anonymously. The range of complementary sets of data includes clinical data from rheumatologists, existing routinely collected datasets such as the General Practice (GP) records, outpatient (OP) clinical data, inpatient (IP) episodes, accident and emergency (A&E) department, pathology data, NHS administrative register, breast and cervical cancer screening data, all Wales injury surveillance system, all Wales perinatal survey, congenital anomaly register and information service, birth and death data from the Office for National Statistics, and social services databases. The SAIL data systems also allows linkage of data collected through patient questionnaires, which then enables following the patient pathway through the NHS system both retrospectively and prospectively from a reference date (e.g. questionnaire completion date). Separately a cohort of people with ankylosing spondylitis (AS), i.e. the Welsh population-based ankylosing spondylitis (PAS) cohort has been developed using data collected from 500+ patient completed questionnaires, which was then linked within the SAIL data systems [4].

HERALD – POTENTIAL BENEFITS OF USING LINKED DATA

The linkage of routine data, patient completed questionnaires and trial data offers unique opportunities for enhanced health economic analysis, including assessment of the validity, reliability and generalizability of health economic data not possible through the use of traditional isolated datasets. In the HERALD paper [1] we use the PAS cohort by linking patient reported study datasets with the routinely collected general practitioner data; inpatient and outpatient datasets; and Accident and Emergency department data in the United Kingdom, and outline the following potential benefits of using the linked SAIL-PAS cohort data in the conduct of health economic analysis in general, and the cost of ankylosing spondylitis, in particular.

Retrospective and prospective tracking of patient pathways

With data spanning multiple years and the ability to link records across several data-

Corresponding author
Muhammad Jami Husain
m.j.husain@keele.ac.uk

Disclosure
The authors declare they have no fnancial competing interests

 47

sets, it is possible using SAIL to track the healthcare utilisation history of patients in receipt of some form of intervention for a given condition across multiple healthcare sectors both before and after their index/reference healthcare event. Therefore, the SAIL data linkage system allows tracking of the patient pathways, both retrospectively and prospectively. Linkage with GP data system provides information about patients' primary care events going back many years including previous diagnoses, referrals, presenting symptoms, investigation results and previous medications. This dataset can also be used to follow the patient at every visit to the GP and therefore record the development of associated conditions and use of co-medications. Linkage with IP data will record all hospital visits, surgery and hospital treatment. Linkage with the mortality datasets will ensure the dataset remains relevant and can examine survival of included patients. Linkage with A&E datasets will give information on emergency visits.

Validation of patient-reported recall data

The use of linked routine data allows cross-checking of patient-reported recall data with actual healthcare events at the personal level. The inherent limitations (or strengths) of the data quality pertaining to the survey questionnaires under the recall method can be flagged; and an assessment of the generalisability of the patient-reported data can be made. On the other hand, data obtained from routinely collected data systems often require careful interpretation with respect to their quality, validity, timelines, bias, confounding and statistical stability [5]. With the triangulation of datasets in the SAIL system, the validity and reliability of single datasets can be assessed [4]. The triangulation process will at least flag the discrepancies, and we can then have an idea about any quality issue pertaining to both the routine and questionnaire data.

Objective measure of the cost of illness

The use of linked data enhances the precision of the healthcare use information and the timelines within which the costs incurred; and therefore will help provide an objective estimate of the cost and burden of diseases to the funders, health service (NHS in the United Kingdom), society and the individual at each stage of disease over a prolonged period of time. In many conditions there is a delay between the onset of symptoms and establishing a diagnosis, during which period the patients still utilise healthcare resources. The

linked routine data can provide information about the patients' visits to healthcare facilities during this symptomatic pre-diagnosis period, when the requirement for diagnostic investigations is often greatest. For example, within the SAIL data system, we can identify patients from a cohort of any particular disease who were diagnosed during a reference time (as indicated by first appearance of a specific diagnostic read code); and link those with various datasets (e.g. GP data, IP hospital admissions, OP, A&E data etc.) to track their pre-diagnosis visits to healthcare facilities since the date the symptom onset (as reported by the patients or established from the GP or A&E records). This allows one to compare the extent of health service utilisation, and therefore related costs, before and after the symptoms developed. In addition, one can also calculate the costs as a result of delayed diagnosis. This provides a more objective estimate of the actual costs of chronic conditions and any interventions.

Healthcare pathways and referral history

A retrospective analysis of the patient's healthcare history can identify the types of referrals to healthcare services made at different points in time facilitating assessment of health service usage and recommendations for improving patient care pathways, and thus maximise clinical efficiency and efficiency in the use of resources. Such analysis makes it possible to deduce whether the anonymised individuals in question were suffering any common co-morbidities, in receipt of healthcare treatment prior to the occurrence of the reference event, whilst it also allows any frequent complications requiring medical attention in the days, months, years following the event (which could be an intervention or questionnaire) to be identified. The temporal aspects of the linked data sets also help conducting event history analysis, survival analysis and other relevant statistical and econometric models.

Profiling of patients

The linked data includes diverse sets of information, which enables profiling and stratification of patients relating to disease manifestations and severity, lifestyle, co-morbidities, and associated costs. Additionally, given that many chronic conditions have heterogeneous manifestations with a variable course and unpredictable episodes of exacerbation, the analysis can be carried out under several person stratification schemes based on severity of disease, various demographic attributes, and socio-economic conditions. This stratification

will facilitate early targeting of interventions to patients at highest risk, thereby improving the cost-effectiveness of these interventions.

CONCLUSION

From the methodological perspective, any linkage system would add new dimensions and perspective to traditional health related research (e.g. complement and enhance the results of RCTs), as a resource for clinical audits, and in a variety of health impact assessment exercises. Linked routine data provides many opportunities for enhanced healthcare research and allows evaluation of impacts beyond the limited primary outcomes of interventional studies. These linkable databases provide factual and continuous information with rich clinical and non-clinical details, which offers wide ranging opportunities in the realm of conducting evaluative research, clinical epidemiology, trial recruitment, genetic research, basic research of biological markers, stratified medicine, post-trial sur-

veillance, risk assessment, service delivery evaluation, resource use, decision analysis, identification of early disease predictors, and the identification of subjects for prospective studies. This data system also offers the opportunity for post-marketing surveillance and pharmacovigilance of new expensive, and often potentially dangerous, healthcare interventions in real-life settings. Complementing this resource with targeted health economic analysis, as proposed in the HERALD methodology, offers a unique opportunity to deliver the level of health economic data required to evaluate and drive forward cost-effective modern healthcare services.

ACKNOWLEDGEMENTS

REFERENCES

1. Husain MJ, Brophy S, Macey S, et al. HERALD (Health Economics using Routine Anonymised Linked Data). *BMC Med Inform Decis Mak* 2012; 12: 24; http://dx.doi.org/10.1186/1472-6947-12-24

2. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009; 9: 3; http://dx.doi.org/10.1186/1472-6947-9-3

3. Ford DV, Jones KH, Verplancke JP, et al. The SAIL databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009; 9: 157; http://dx.doi.org/10.1186/1472-6963-9-157

4. Atkinson MD, Brophy S, Siebert S, et al. Protocol for a population-based Ankylosing Spondylitis (PAS) cohort in Wales. *BMC Musculoskelet Disord* 2010; 11: 197; http://dx.doi.org/10.1186/1471-2474-11-197

5. McLean G. A review of the use of routine data on CHD in general practice research in Scotland. *Scott Med J* 2005; 50: 109-13