



Llywodraeth Cymru
Welsh Government

www.cymru.gov.uk

Data Linking Demonstration Project - Journey Mapping for Patients with Multiple Chronic Health Conditions

Research Summary

Social research

Number: 76/2013

1. The Multiple Chronic Health Conditions (MCHC) demonstration project was one of three data linking projects delivered under the Welsh Government Programme to Maximise the Use of Existing Data. The projects were delivered by a Knowledge Transfer Research Fellow jointly funded by the Welsh Government (WG) and the Economic and Social Research Council (ESRC) working on a one-year fellowship which ran on a part time basis over the period from October 2011 to March 2013.
2. Data Linking is a technique for creating links between data sources so that anonymised information that is thought to relate to the same person, family, place or event can be connected for research purposes.
3. The suite of three data linking demonstration projects utilised the Secure Anonymised Information Linkage (SAIL) database at Swansea University. This is a

Welsh Government (WG) funded facility which contains a growing number of anonymised health and social policy-relevant datasets that, subject to ethical and data controller approval and under to strict information governance arrangements, can be made available for analysis by approved researchers. The projects examined the anonymisation and linkage process, from acquiring additional data to carrying out analysis on new data sets created by linking existing administrative data.

Aims

4. To demonstrate the unique contribution the linking of routinely collected data can make to the evidence base.
5. To stimulate the engagement of appropriate WG officials with regard to information governance and practical issues around acquiring, processing and analysing new linked data sets.

6. To examine linked electronic health records for insights about patient pathways that will be useful in designing a more efficient and effective health service for the future.
7. To address a number of methodological challenges, as follows:
 - a. Select a set of patients with a single chronic health condition and find a way to align their records by the stage of the condition e.g. at point of first emergency admission, so that their pathways become comparable.
 - b. Determine whether 'health events' routinely recorded in primary and secondary care records (e.g. prescriptions for blood thinners) can be used to split the set of patients into subsets who have similar 'patient pathways'.
 - c. Explore how the different patient pathways vary according to risk factors (e.g. smoking, weight, BP etc) and the socio-demographic characteristics of patients (e.g. gender and age group).
 - d. Find a way to add a single, chronic complicating condition (or co-morbidity) into the analysis as a first step in developing a method for the analysis of multiple chronic health conditions.

Methodology

8. Originally, the project aimed to develop patient pathways for patients with a number of chronic conditions. The initial exploration of the data indicated that the number of variables involved and consequently the vast numbers of potential patient groups that could be examined was far more than could be handled within the limited scope of a demonstration project. It was therefore agreed with WG policy leads to concentrate the analysis on patients who had suffered a stroke and patients with diabetes and to develop a methodology that could potentially be used for other combinations of chronic conditions in the future.
9. A patient pathway is the route a patient takes through different health services, usually from diagnosis to completion of treatment. In this case, we are looking back at the pathways patients have taken in the period leading up to the diagnosis in order to e.g. identify protective interventions or groups of patients GPs might wish to target.
10. The SAIL hospital admissions data was used to select a set of patients who had experienced an emergency admission for stroke. To 'qualify', the records had to indicate an emergency admission with the principle diagnosis 'Stroke' between the dates of 01.01.2007 and 31.12.2009. This provided an anonymous group of 18,744 stroke patients.
11. The hospitalisation history of the stroke group prior to the qualifying event was

examined to determine the first emergency stroke admission date for each individual; this was used as a benchmark date to align the patients' medical histories to the same stage of the condition: their first emergency admission for stroke (FEAS).

12. Using linked datasets, a number of health and demographic indicators were developed for the four years prior to the FEAS and for the period between the FEAS and the date when the most recent records were available in SAIL (31/12/2012).
13. The resulting dataset was compared in a number of ways to try to identify patient groups based in similar demographic characteristics and 'patient pathways'. Where appropriate, we checked whether any differences in the characteristics of stroke patients or any variations in patterns of stroke were big enough for us to be confident that they did not happen purely by chance – such changes are referred to as 'statistically significant' or 'significant'.

Key Findings

14. The project identified a set of 18,744 individuals who were admitted between 01.01.2007 and 31.12.2009 for whom the principal diagnosis was 'Stroke', 8,962 (48%) were men and 9,782 (52%) were women.
15. The incidence of 'Stroke' varied by age and gender, with significantly more men

than women having strokes before the age of 70 years

16. Emergency admissions for 'Stroke' varied by deprivation¹, with significantly more strokes among people from the most deprived fifth than the least deprived fifth of the stroke population. This was true for both men and women in all age groups under 70 years.
17. During the four years prior to the FEAS, and including the stroke admission itself, men spent on average 44 days in hospital, whereas women spent significantly more, on average 61 days. During these admissions, in addition to codes used to describe their stroke - a total of 1,330 further diagnostic codes were used 97,574 times to record additional health conditions for the stroke patients (on average 11 codes per patient). These codes include common conditions such as hypertension, other conditions diagnosed on admission for stroke as well as previously diagnosed conditions reported as part of a patient history.
18. Just under half of the Project Set of stroke patients who died before 31.12.2011 died from the effects of their stroke and a further 18% died of 'other circulatory conditions'. A further 10% of those who died before 31.12.2011 died of cancer and 8% from respiratory causes. Cross checking hospital and

¹ Welsh Index of Multiple Deprivation (WIMD)

ONS mortality data revealed that about 2% of stroke deaths did not appear in the hospital data; these are most likely to be cases where a patient was pronounced dead before reaching hospital.

19. A total of 18% of the stroke patients had the complication of being diabetic. Data from the Welsh Health Survey indicates that the prevalence of diabetes in the general population of Wales is 6%. The greater prevalence of diabetes in stroke patients is in line with what is known about diabetes symptoms being risk factors for stroke².
20. There were gender differences in the diabetic stroke patients. A significantly greater proportion of the diabetic stroke patients were male (53% were male compared with 47% female). Female diabetics who had strokes presented with a stroke when they were about, on average, 5 years older than male diabetic stroke patients. Given that they were on average older, it is not surprising to find that a significantly higher proportion of the diabetic stroke patients who died before 31.12.2011 were female (23% female compared with 16% male). The average length of stay in the four years prior to and including the FEAS was significantly higher in diabetic men than non-diabetic men, especially for

those with Type1 diabetes; no similar variation in average length of stay was seen in women.

21. Cluster analysis was undertaken on the Project Set of Stroke patients for whom primary care event records were available (6,921 or 39%³). A set of primary and secondary care events chosen for their likely relevance to stroke were used to identify groups of patients with similar pathways. The events chosen were, for each patient: the total number of primary care events, the number of prescriptions for statins, anticoagulants and nicotine replacement therapies, the total number of blood test results and blood pressure, body weight and BMI measurements, the number of times smoking status was recorded as well as the 'current smoking status' and BMI measurement closest (but prior) to stroke. Indicators from secondary care records related to their history of hospital admissions prior to and following their first stroke emergency admission (e.g. number of admissions, total length of stay, recorded diagnoses, the elective and non-elective nature of the hospitalisation).
22. A very simple cluster analysis method was used due to the limited time available for the demonstration project. The findings should therefore be

² <http://link.springer.com/article/10.1007/s00125-006-0493-z#page> Study of stroke risk in type 2 diabetes using General Practice Research Database in the UK.

³ NISCHR has recently invested further funding for GP engagement posts with the aim of significantly increasing the number of GP practices contributing to the SAIL databank.

considered experimental and indications are given in the main report for how the methodology could be improved upon. Further discussion with policy colleagues would also help to decide the kinds of events that should be included in future, taking into account the consistency of recording of those events. The findings should therefore be viewed with caution.

23. As we would expect, given the findings reported above, some of the clusters of stroke patients created varied according to age group and gender - the younger patients tended to appear in distinct clusters and some clusters were strongly gender-specific.
24. The 'pathways' identified using the cluster analysis suggest a distinction between those patients who were relatively highly engaged with primary care and those who had little or no engagement with primary care prior to their stroke event.
25. One of the clusters identified contained mainly patients with relatively minor stroke events and no diabetes but high levels of interaction with health services. This suggests that other conditions not included in the analysis were having a significant effect on the health service use in the 'patient pathway' for this cluster. This may be due to the presence within the cluster of patients with just the kinds of multiple chronic conditions the project was originally designed to analyse; however, additional work would

be required to explore this possibility further.

Next Steps

26. The complex nature of the data meant that a significant amount of project time went into developing an understanding of the data itself e.g. how to manipulate and recode it into suitable forms to use as indicators. As a result, the focus of the Project was narrowed to focus on patient interactions with primary and secondary care before the stroke event. Consequently there is scope for significant further work.
27. The cluster analysis has shown that meaningful groups of patients with similar 'pathways' prior to stroke can be identified. Further work could be done to refine the solution for stroke patients. However, the principle having been established that this kind of analysis can identify meaningful groups, it could be used to undertake analysis for further main conditions and to examine the effect of including more than one co-morbidity in the analysis.
28. Significant further development work would be required to get to the point of having a tool to e.g. compare the pathways for clusters of patients with a particular combination of chronic conditions to determine whether key events exist, the presence or absence of which might be protective e.g. reducing length of stay or other patient outcome.

The demonstration project has proceeded far enough to enable the informed writing of a specification for such a tool. In 2012, SAIL was successful in bidding to lead the Centre for the Improvement of Population Health through e-Records Research (CIPHER), one of the four major centres that will make up the Farr Institute, a MRC-funded UK health informatics research institute. The creation of the Farr Institute will present opportunities to identify projects that can build on the methods this Project has developed.

Views expressed in this report are those of the researchers and not necessarily those of the Welsh Government

Authors:

Martin Heaven

Health Information Research Unit, Swansea University / Data Linkage Fellow, Welsh Government

Sarah Lowe

Knowledge and Analytical Services, Welsh Government

Contact:

Sarah Lowe

Phone: 029 2082 6229

Email: sarah.lowe@wales.gsi.gov.uk

ISBN 978-1-4734-0675-9

© Crown Copyright 2013