International Journal of Population Data Science

Journal Website: www.ijpds.org

Lessons learned from linking to 2001 census in Scotland

Povey, Chris

Objectives

SHELS (Scottish Health and Ethnicity Linkage Study) linked Scotland's 2001 census to various hospital and death data sets with national coverage. Census ethnicity data were assigned to the study records to build a cohort of most of the Scottish population; included in the cohort were people with no health records.

Approach

Create a lookup table of a person's census index to the Scottish eHealth index, the CHI, equivalent of English new national health number. A modified version of the eHealth administrative matching system was used to satisfy census confidentiality requirements. There were two linkages performed in 2004 and 2008. 2004 was a feasibility run; the 2008 applied lessons learned from the previous linkage and used much more completely indexed health records.

Results

The first linkage produced match rate of 95% of 4.9 million 2001 census entries; the second 96%.

Conclusion

Lessons learned. Linking datasets using indexes is the most accurate and efficient way to produce study cohorts. Indices change over time; a methodology called 'reconciliation' was devised to retrospectively and continually adjust previously indexed (linked) records.

How to Track members who migrate out of the cohort.

A linkage resource called a residential events dataset (RESEVENT) was built for the 2008 linkage run; it holds merged history of linkage identifier fields by date from January 2000 to the present based on GP registrations.

This introduces a time dimension to indexed linking.

How to build RESEVENT like linkage resources; should they be census based? What should they contain? How to do daily national census and select controls for case/control cohorts from RESEVENT resource.

How postcode changes over time can be handled (reconciled) - same address, different postcode, but no address present.

Proposal for an index of national indices based on national administrative datasets starting with NHS number (new and old NHSCR) and NI number to make linking even more efficient - this is not a RESEVENT resource; this resource would mean data need be matched to index only once, all subsequent linkages would be deterministic links of reconciled indices.

*Corresponding Author:

Email Address: chrispov153x@gmail.com (P. Povey)