

Using Residential Anonymous Linking Fields to Identify Vulnerable Populations in Administrative Data

Hollinghurst, J¹, Fry, R², Akbari, A³, and Rodgers, S¹

¹Swansea University Medical School

²National Centre for Population Health and Wellbeing Research

³Health Data Research UK - Wales and Northern Ireland, Swansea University Medical School

Introduction

Demographic profiling is an important aspect of anonymised healthcare research to identify the population of interest. Typically, administrative data is used in conjunction with patient registers to create cohorts, but it can be a time consuming process. We describe a method using routinely collected health data to identify vulnerable populations.

Objectives and Approach

Using existing longitudinal data and the Residential Anonymised Linking Field (RALF) we aim to identify institutions linked to vulnerable populations. We search for specific characteristics of these institutions including the age of occupants, number of current residents, and rate of change of occupants. We also aim to compare our method to a pseudonymised national registry for care homes to ensure it is accurate. This can effectively reduce the need for repeat pseudonymisation of institutions, which is both expensive and time consuming.

Results

To implement our method we found the most recent address for living individuals aged 65-95. This produced 202,640 residences from 1,330,335. Of the 202,640 residences, 1347 had four or more cohabitants aged 65-95, and 172 had exactly three residents with ten or more distinct individuals registered over a 10-year period. Our final synthetic dataset therefore had 1519 unique potential care homes to compare to the national registry, which contains 1525 registered care homes.

We can now link the synthetic dataset to individuals to flag their residential status, which may be a defining factor in their level of care. Furthermore, we can answer specific research questions relating to their residency, such as the time it takes to move to a care home following a hospital admission.

Conclusion/Implications

By using quantifiable characteristics of care homes we were able to create a synthetic care home register by searching existing data. This is a reproducible process that would be of particular benefit for projects where a registry is not available, or where time or cost would limit the availability.

