International Journal of Population Data Science

Journal Website: www.ijpds.org

A longitudinal, historical model of households using linked Welsh administrative data

Tingay, K¹, Anastasiou, A², and Musselwhite, C³

¹Swansea University

²bibInsight, Swansea University Medical School

³Centre for Innovative Ageing, College of Human and Health Sciences, Swansea University

Introduction

Routine data sources are increasingly being explored as an alternative to surveys for governmental censuses. Censuses are useful, but only capture one snapshot in time. For research, changes in state over time can be important. This is particularly true for households, where changes in both composition and location are routine.

Conclusion/Implications

The ability to have a longitudinal record of households can provide additional detail to research which can be costly to obtain otherwise. This method shows promise as a useful tool for linked population data research. Further validation against surveys, and work to identify and model communal residences is needed.

Objectives and Approach

The Secure Anonymised Information Linkage (SAIL) databank at Swansea University holds anonymised address data tied to Welsh General Practice (GP) registrations. The dataset has the individual's unique identifier (ALF), address identifier (RALF), and moving in and moving out dates. This can be securely linked to other demographic and health data.

Using this data, a finite, deterministic household model was developed. This model implements a set of rules for establishing, advancing, and terminating households. The rules cover different types of households, including transient and multi-generational, but not communal residences such as student halls or residential care homes.

Results

The model was implemented in Python, and operates over finite moving in/out transactions undertaken by actors (ALFs) on specific properties (RALFs). The roles also help to diminish data issues, such as infants moving into an address before their parents.

The model creates and updates histories of households for each RALF. This allows researchers to identify unstable versus stable households, or to look at possible environmental health factors over time.

The script also provides a running count of the total number of residents at each RALF at any given time. This makes it useful for time-specific research, or for linking to censuses.