International Journal of Population Data Science

Journal Website: www.ijpds.org

Validating Synthetic Longitudinal Populations for evaluation of Population Data Linkage

Dalton, T^{1*}, Kirby, G¹, Dearle, A¹, Akgün, Ö¹, and Mackenzie, M¹

¹University of St Andrews

Background

'Gold-standard' data to evaluate linkage algorithms are rare. Synthetic data have the advantage that all the true links are known. In the domain of population reconstruction, the ability to synthesize populations on demand, with varying characteristics, allows a linkage approach to be evaluated across a wide range of data. We have implemented ValiPop, a micro-simulation model, for this purpose.

Approach

ValiPop can create many varied populations based upon sets of desired population statistics, thus allowing linkage algorithms to be evaluated across many populations, rather than across a limited number of real world 'gold-standard' data sets.

Given the potential interactions between different desired population statistics, the creation of a population does not necessarily imply that all desired population statistics have been met. To address this we have developed a statistical approach to validate the adherence of created populations to the desired statistics, using a generalized linear model.

This talk will discuss the benefits of synthetic data for data linkage evaluation, the approach to validating created populations, and present the results of some initial linkage experiments using our synthetic data.

*Corresponding Author:

Email Address: tsd4@st-andrews.ac.uk (T Dalton)