International Journal of Population Data Science

Journal Website: www.ijpds.org

Finding Maternal Siblings in Birth Registration Data to form a Pregnancy Spine – Data Linkage & Graph Based Methods for Unknown Cluster Sizes

Tomlin, C[1]*, Gammon, S[1], Morris, C[1], and O'Brien, C[1]

[!]Office for National Statistics

We have developed an innovative methodology to link maternal siblings within 2000-2005 England and Wales Birth Registration data, to form a Pregnancy Spine, a unification of all births to each unique mother. Key challenges were Blocking & Cluster resolution.

To optimise geographic blocking, Internal Migration data was incorporated to map likely geographic movement of mothers between births.

Following probabilistic linkage, sibling clusters were modelled as a graph and their structure optimised using community detection methods. Childhood statistics data relating to child DOB were incorporated to evaluate accuracy and remove false links.

Our development has resulted in a new blocking and cluster resolution method. We developed new ways to assess sibling group accuracy, beyond traditional classifier metrics, and infer error rates.

We applied our method to Registration Data used in earlier studies for QA of our methods.

Using this, and other maternal sibling composition statistics, we present results showing that a high degree of accuracy was obtained for standard and new evaluation metrics.

These methods will improve other linkage projects linking unknown clusters sizes/multiple datasets, or longer time-period longitudinal linkage. To this Spine, researchers can append and link other data sources to answer questions about maternal and child health outcomes.

*Corresponding Author:
Email Address: charles.tomlin@ons.gov.uk (C Tomlin)