

International Journal of Population Data Science

Journal Website: www.ijpds.org



Identifying and linking individuals participating in multiple longitudinal population studies in the UK Longitudinal Linkage Collaboration

Samuel Neale¹, Richard Thomas¹, Jacqui Oakley¹, Hazel Lockhart-Jones², Chris Orton², Simon Thompson², and Andy Boyd¹

¹University of Bristol, Bristol, United Kingdom

²Swansea University, Swansea, United Kingdom

UK Longitudinal Linkage Collaboration (UKLLC) is the national Trusted Research Environment (TRE) for record linkage in longitudinal research, partnering with SeRP and >20 Longitudinal Population Studies (LPS). Participating in LPS is rewarding, with many participants enrolling into multiple studies (e.g., ~8% of ALSPAC mothers are also enrolled into UK Biobank). It's scientifically important to account for this in pooled and meta LPS analysis as most statistical assessments assume independence of sample membership. Currently, LPS records are treated separately, leading to potential duplication and over-counting of exposures/outcomes in research, where individuals participate in multiple cohorts. We used probabilistic record linkage to identify individuals across multiple cohorts. We engaged study Data Managers through a consensus-building workshop to reconcile governance issues.

Of ~570,000 participants from 22 partner LPS in UKLLC, we've currently identified 4,785 individuals in two cohorts, 155 in three, and <10 in four. These numbers are likely to increase as UKLLC scales to support larger studies (with an anticipated 2m participants hosted by 2027). Mappings of individuals belonging to multiple cohorts were delivered to end-users via our data provisioning pipeline in a manner compatible with the dynamic nature of the pooled UK LLC hosted sample. This linkage will support accurate pooled- and meta-analysis within UKLLC. Ensuring the governance challenges are accounted for is essential for the acceptability of our community LPS governance framework. Extending this linkage of individuals across different LPS and platforms (e.g., to UK Biobank) will be necessary for robust federated analysis across TREs.

