International Journal of Population Data Science

Journal Website: www.ijpds.org

Enhancing Environmental data Resources in Cohort Studies: ALSPAC exemplar (ERICA)

Boyd, A¹

¹ALSPAC, University of Bristol

Introduction

Novel data platforms are needed to expedite the linkage of geospatial and natural environment data with longitudinal population study(ies) (LPS) data. However, the intrinsic relationship between geospatial data and participant identities raises confidentiality concerns amongst participants which need to be accommodated within platform designs to ensure acceptable data use.

Objectives and Approach

We aimed to establish generalisable mechanisms for linking spatial records into a cohort study LPS databank. Using 'Data Safe Haven' (Burton et al. 2015) approaches we developed a pipeline of technical processes (e.g. geocoding participants' residential and school address records, developing an 'engine' for linking exposure data to LPS data), which we tested for participant acceptability (through a focus group testing acceptable use of personal identifiers in this context) and tuned to meet regulatory requirements (e.g. privacy impact assessments, information security accreditation). We demonstrated our approach through an exemplar investigation assessing in utero NO₂ exposure with later childhood respiratory outcomes.

Results

Participants' expressed clear expectations that the research use of location data should be restricted to trusted study staff (as distinct from the wider research community), although this expectation is context specific and does not represent a carte blanche for using granular information (e.g. GPS tracing data). This necessitated a 'split stage' protocol, where personal identifiers are handled separately from data/analysis. Daily NO₂ exposure was modeled using road, other local and regional records; with validation exposure data collected from city-based sensors. Participants' location information was geocoded and linked to NO₂ using the ALGAE (ALgorithms for Generating address histories and Exposures) privacy-preserving geocoding engine. We will summarise participant views, our

exemplar findings, and describe the linkage engine we have developed and its availability via an open-source repository.

Conclusion/Implications

Participants' expect that studies control for confidentiality risks introduced by research using spatial identifiers. It is not realistic to expect that all LPS have the capacity to undertake specialist spatial linkages. Our generalisable approaches and open-source software could provide the basis for geospatial and natural environment epidemiology platforms in LPS.