SaxInstitute | 45 and Up Study

Technical Note: Geocoded 45 and Up Study Data

Version: February 2023

Overview

45 and Up Study participant addresses are geocoded based on residential address information at the time of each questionnaire. Broad level geography based on geocoded data at SA2 level and above can be requested as part of a standard data supply. In addition to broad level geography, researchers with ethics approval for access to lower level geography (small areas) can be provided with SA1, postcode, latitude, longitude and mesh block, along with other low-level geography. This is provided in a separate SURE workspace with different IDs that are linked back to the questionnaire data by the 45 and Up Study Coordinating Centre.

Broad level geography

Broad level geography is geocoded data at SA2 level and above. Researchers can request the Variable Group 'Broad Level Geography' from the 45 and Up Study Dataset and Variable Selection Spreadsheet. Variables in this group include:

- Statistical Area Level 2 (SA2)
- Statistical Area Level 3 (SA3)
- Statistical Area Level 4 (SA4)
- Local Government Area (LGA)
- Local Health District (LHD)
- Primary Health Network (PHN)
- Accessibility/Remoteness Index of Australia (ARIA)
- Greater Capital City Statistical Areas (GCCSA)
- Modified Monash Model (MMM)
- Socio-Economic Indexes for Areas (SEIFA)
  - Decile of Index of Education and Occupation (IEO)
  - Decile of Index of Economic Resources (IER)
  - Decile of Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)
  - Decile of Index of Relative Socio-economic Disadvantage (IRSD)

Low level geography

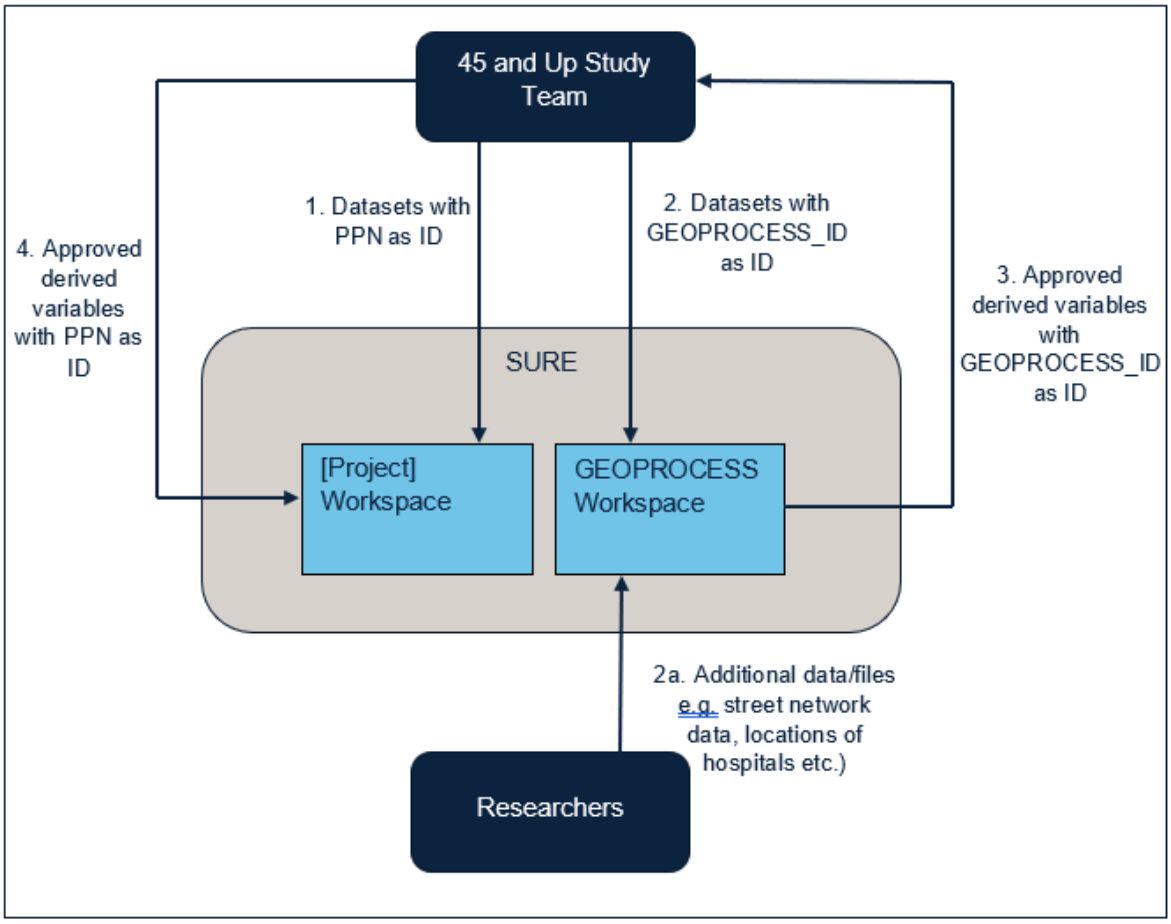
Low level geography is geocoded data at SA1 level and below. Researchers can request this data on the 45 and Up Study Research Project Expression of Interest and Application form by selecting 'Small area geography (residential location) at SA1 or below' in Section 2. Researchers must explain why this level of detail is required. All variables from Broad level geography are included as well as the following:

- Geocodes (latitude/longitude)
- Mesh blocks (MB)
- Geocoder used (Febrl, Callpoint Spatial)
- Match status (exact address/match, locality match, street match, intersection match, average address)
- Statistical Area Level 1 (SA1)
- SA1 Usual Resident Population
- Postcode
- Socio-Economic Indexes for Areas (SEIFA)
  - SA1 Index of Education and Occupation (IEO)
  - SA1 Index of Economic Resources (IER)
  - SA1 Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)
  - SA1 Index of Relative Socio-economic Disadvantage (IRSD)

GEOPROCESS SURE workspace

Low level geography data is kept separate from other 45 and Up Study data to prevent reidentification of participants. New IDs are assigned by the 45 and Up Study team and the data is made available in the GEOPROCESS SURE workspace (see Figure 1). All data entering or exiting the GEOPROCESS workspace is curated by the 45 and Up Study team. This workspace was created to allow researchers the ability to derive "value-added" variables associated with exact geographic location of a participant's address, such as distance to nearest hospital, while meeting the Study's commitments to participants regarding privacy and to minimise the risk of re-identification.

Figure 1: Data flow for researchers accessing the GEOPROCESS SURE workspace



Importing data to GEOPROCESS

Researchers can upload additional data or files (e.g. street network data, locations of hospitals or other health care facilities) to the GEOPROCESS workspace. These files will be reviewed by the 45 and Up Study team. Information on what and how that data will be used must be emailed to 45andup.data@saxinstitute.org.au for a 45 and Up Study data manager to review.

Exporting data from GEOPROCESS

Only derived variables may be transferred from the GEOPROCESS workspace to the project's regular SURE workspace. After uploading derived variables for export from GEOPROCESS, researchers must email 45andup.data@saxinstitute.org.au explaining what the derived variables are and how they were derived. The data will be reviewed by a 45 and Up Study data manager to assess whether it's suitable to be taken out of the GEOPROCESS workspace. Revisions may be recommended if not. No potentially identifiable data can be exported from GEOPROCESS, i.e. derived variables must not be too small a geographic area (SA2 is lowest allowable in a regular SURE workspace).

Geocoding of data

Two sources of geocoding have been used:

1. Freely extensible biomedical record linkage (Febrl)
   In 2009, approximately 240,000 full addresses of 45 and Up Study participants were geocoded by NSW Health using the Febrl program. The technical details relating to this program are in the document: http://pdf.aminer.org/000/295/338/a_probabilistic_model_for_postcode_recognition_a_first_step_towards.pdf. NSW Health produced files containing the latitude, longitude, area health service (AHS), census district (CD), statistical local area (SLA), local government area (LGA), and match status. The match status after each record was matched was given as exact address, exact intersection or average address. The latitude and longitude is in the middle of the locality if not exact.
2. Callpoint Spatial
   The remaining baseline addresses and subsequent follow up data have been geocoded using Callpoint Spatial (http://www.callpointspatial.com.au/). The Callpoint Spatial geocoder provides latitude, longitude, meshblock, and Statistical Area Level 1 (SA1). Two broad levels of mapping were applied (exact and street) which mapped to meshblock and locality (mainly PO boxes) which mapped to SA1. The match status after each record was matched has been given as exact match, intersection match, street match or locality match. The latitude and longitude is in the middle of the intersection, street, locality if not exact.

Notes on geocoded data

- If there are no geoprocessing variables or the geocoder was not able to map to a single meshblock or SA1 then the record has been removed and no higher-level geography or Socio-Economic Indexes for Areas (SEIFA) or postcode are provided
- Where the population of an SA1 is too small, SEIFA has not been assigned as these scores are not released by the Australian Bureau of Statistics (ABS)
- Only NSW addresses have been included
- Higher level geography has been produced using shape files provided by the ABS and the Ministry of Health (Local Health District and Primary Health Network)
- Geocoding of addresses where a property name has been given rather than a street address may not be as robust