Data Access Committee Protocol

EOIs requesting potentially identifying variables

Preamble

There are strict legal and ethical conditions under which ALSWH is permitted to use the Study's own survey data and external linked health record data in analyses. One important condition is that datasets must not include potentially personally identifying. Therefore, geographic variables, including Statistical Local Areas (SLAs), Health Service Districts, Postcodes*, SEIFA index scores*, and latitude/longitude must be coded or categorised, and sufficiently aggregated, to minimise risks to participant privacy.

*_ARIA+, and SEIFA percentiles, are non-sensitive alternatives which are routinely provided in ALSWH datasets._

Procedures

1   Researchers must provide adequate detail in their EoI about how geographic variables will be coded, to enable risk assessment by the Data Access Committee.

   a.   Where existing or published coding schema will be used, the schema and references must be supplied.

   b.   Where ad hoc coding is planned, the proposed procedure should be described as fully as possible. If proximity to a specific exposure will be measured (e.g., kilometres from power station), please explain how the sites of the exposures will be ascertained, and provide any available references to justify the methodology.

   c.   Please note that new variables derived from geographic variables may be made available to other data users as part of the ALSWH survey dataset.

2   Coding of geographic variables will be performed as a separate step, before survey and linked data are supplied.

   a.   The ALSWH Data Manager (UQ) will generate a unique set of temporary IDs (TempID) for the cohorts involved, and send the key between the IDAlias and the TempID to the ALSWH Operations Manager (UN).

   b.   If longitudes and latitudes of addresses are being used then the ALSWH Operations Manager will attach the geographic variables, strip off the IDAlias, and supply the dataset to the designated researcher for coding. The ALSWH Operations Manager will also provide an accuracy code to indicate what address was used to produce the latitude and longitude (for example, whether it was based on database residential address, database postal address, survey residential postcode, survey postal postcode, overseas – not geo-coded, or not geo-coded due to privacy reasons, not geocoded due to other reasons).

   c.   After coding according to the agreed schema, the researcher will return the dataset containing TempID, geographic variable/s and the new coded variable/s, to the ALSWH Operations Manager.

   d.   The ALSWH Operations Manager will re-apply the IDAlias, strip off the TempID and the original geographic variable/s, and send the dataset (containing only IDAlias and the coded variable/s) to the ALSWH Data Manager.

   e.   The ALSWH Data Manager will review the coded data, to ensure that it does not pose a re-identification risk. The ALSWH Data Manager may request that categories be further aggregated or re-named (for example, due to small cell sizes). Once satisfied, the data will be integrated with survey (and, if applicable, linked health record data) and made available to researchers as per the standard ALSWH Data Access Protocol.