

International Journal of Population Data Science

Journal Website: www.ijpds.org



Scotland's Integrated Demographic Dataset and Administrative Data Based Ethnicity Dataset

David Rowley¹, Caroline Ellis¹, and Lisa Whyte²

¹National Records of Scotland, Edinburgh, United Kingdom

²Scottish Government, Edinburgh, United Kingdom

Objectives

The census is among the richest available population-scale, individual-level datasets. Scotland's Integrated Demographic Dataset (SIDD) is a population-scale individual-level dataset using only administrative data. The Administrative Data Based Ethnicity Dataset (ABED) is based on the SIDD and includes ethnicity information.

Conclusion

The SIDD and ABED are large datasets, approximately covering the Scottish population, with the SIDD accurately capturing most individuals. Administrative data can provide substantially more population coverage for ethnicity than could be achieved using surveys, but lower than what could be provided by a census.

Methods

Several administrative datasets were linked together, including those from health, education and the electoral register. The records were resolved into individuals to produce the Administrative Data Record Set. This was trimmed down using activity-based business rules to a set of individuals (the SIDD) believed to be usual residents on the reference date. The SIDD was linked to the census dataset to test its quality. Ethnicity information from the linked administrative datasets was added to the SIDD to produce the ABED. In cases where ethnicity information is not available, this was supplemented with ethnicity information from the 2011 census dataset.

Results

The SIDD is broadly comparable to the census dataset. 95.6% of census records appear on the SIDD, while 84.7% of SIDD records appear on the census. 89.4% of individuals appear at the same address on the census and the SIDD. 98.2% appear in the same local authority, although this is lower for individuals in their 20s. LA differences are not unbiased, e.g. there are more people in their late 20s in a suburban LA on the SIDD and another LA on the census, than vice versa.

When using only the currently available linked administrative data, a stated ethnicity can be found for 58.2% of individuals. When administrative data is combined with the Census 2011 data, a stated ethnicity can be found for 76.6% of individuals.



Temporary page!

$\LaTeX$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because $\LaTeX$ now knows how many pages to expect for this document.