

International Journal of Population Data Science

Journal Website: www.ijpds.org



Linking administrative and Census 2021 data in Wales, UK: A cross-sectional study examining completeness and representativeness for population linkage analytics.

Rhodri Johnson¹, Jane Lyons¹, Michael Edwards¹, Samantha Turner¹, Richard Fry¹, Lucy Griffiths¹, and Ronan Lyons¹

¹Swansea University, Swansea, United Kingdom

Objectives

To explore the completeness and representativeness of Census 2021 data linkage within the Secure Anonymised Information Linkage (SAIL) Databank for research on the population of Wales, UK and, understand which subgroups of the population are disproportionately represented in data linkage population-wide studies.

Methods

An observational, population-wide cross-sectional comparison study, utilising administrative demographic data and decennial survey data held in SAIL. Two linked data sources, the Welsh Demographic Service Dataset (WDSD) and Census 2021, were used to create and compare two cohorts of the resident population of Wales, UK, on 21st March 2021.

The two cohorts were linked together to provide understanding on how many individuals from Census 2021 can be successfully linked within SAIL and found across both sources. We utilised logistic regression models to analyse the variation in the linkability of the survey data within SAIL by various demographic and household characteristics.

Results

In total, 3,090,976 individuals were present in the WDSD population, 2,965,196 individuals in the Census population, 2,440,191 individuals found in both, with 650,785 and 525,005 individuals found only in WDSD and Census respectively. Focussing on the multivariate logistic regression analysis ($n = 2,415,260$, aged 16+ and non-communal establishment resident), being male (OR=1.28 [95%CI 1.28,1.32]), aged 75+ years (OR=1.27 [95%CI 1.25,1.29]), of Asian ethnicity (OR=1.27 [95%CI 1.24,1.30]), a more recent migrant (arriving to UK after 2000) (OR= 1.30 [95%CI 1.28,1.32]), member of the LGBTQ+ community (OR=1.29, [95%CI 1.25,1.29]) or not disclosing LGBTQ+ status (OR=1.41 [95%CI 1.39,1.43]), separated, divorced or widowed (OR=1.28 [95%CI 1.27,1.29]), or living in rental accommodation

(OR=1.47 [95%CI 1.45,1.48]) were the characteristics associated with the highest odds of not having Census linkable data in SAIL.

Conclusion

Results show that certain personal characteristics and subgroups of the population of Wales are disproportionately represented when combining population estimates and utilising Census data in data linkage population-wide studies in SAIL. This is an important finding for researchers to understand when carrying out future linked research on the Welsh population.

