

International Journal of Population Data Science

Journal Website: www.ijpds.org



Linking of the 2021 Census to massive linked administrative data to understand coverage and quality

Josie Plachta¹ and Sarah Collyer¹

¹Office for National Statistics, Newport, United Kingdom

The Office for National Statistics has built a vast, composite dataset for population statistics by linking data from health, education, and employment sources, known as the Demographic Index (DI). It attempts to contain a record 'cluster' for each person in England and Wales. To understand the coverage and quality of the DI, it has been linked to the 2021 Census to a high standard, enabling review of those captured incorrectly: and over- and undercoverage.

Massive data techniques were used to apply deterministic, probabilistic, and associative methods. High quality was achieved by applying clerical matching methods to cases that could not be confirmed by automatic techniques. Due to resource limitations, only a subsample was linked to this high standard. The resulting links were flagged to indicate cases where the DI had correctly captured persons or had made errors. Errors included capturing persons at the wrong address, accidentally splitting a person's records across two clusters, or incorrectly capturing two persons in the same cluster. Unlinked records were flagged as under-coverage (census) or over-coverage (DI).

The 2021 Census was linked to the DI with an estimated precision of 99.4%-99.7% and recall of 99.1%-99.7%. This exceptional quality allows ONS analysts to use this dataset with high confidence in analysing the quality of the DI and its impact on statistics. In general, DI under-coverage was low, with 0.9% of Census records in the subsample not present on the DI. However, DI over-coverage was much higher, with 29.5% of DI records in the subsample not present on the census. 2.3% of census persons in the subsample had been incorrectly split across multiple clusters, and 0.3% had been merged into a cluster with multiple other persons.

The ONS successfully linked the 2021 Census to the DI to a high quality. The linkage suggests that the DI captures most of the current population correctly but captures many persons that are not. These insights must be considered by any users of the data.

