Producing Official Statistics for and about New Zealand's Rainbow Population

Digby Carter, Statistics New Zealand

This paper seeks to provide an overview of major design decisions, challenges, and quality considerations during the design and data evaluation for the inclusion of demographic concept data relevant for rainbow populations into Stats NZ's census of population and dwellings.

Stats NZ has undertaken a major update of the way we collect and produce data across the concepts of gender, sex, variations of sex characteristics and sexual identity. Updating the way we capture these concepts in some of our major collections has occurred alongside and during the redevelopment of the updated statistical standards for gender, sex and variations of sex characteristics published in 2021, and the statistical standard for sexual identity published in 2019.

The recent 2023 Census is the first New Zealand census to reflect the updated statistical standards for gender, sex at birth, variations of sex characteristics and sexual identity. This includes the first inclusion of a question on variations of sex characteristics in Stats NZ collections. Collecting the full suite of concepts allowed us to produce the widest range of outputs on the rainbow population of New Zealand from the census outputs.

The coverage provided by the census means it is the best source of information on smaller population groups across both concepts related to the rainbow community, but also other identity-based concepts such as ethnic and religious affiliation. Data is available at a granularity not possible from other existing surveys or administrative data sources.

The inclusion of these concepts into the census necessitated a number of design decisions including selecting subject populations, whether to require mandatory response, form placement, question design, and methods of alternative data sourcing where possible. Considerations of respondent burden or refusal, along with the need to obtain responses had to be balanced.

Two of the concepts (gender and sexual identity) included write in response options, which required processing written responses for the concepts for a collection at the scale of the census for the first time. Some changes to the coding of synonyms were required post-collection, during the data evaluation phase.

The 2023 Census dataset is produced through a combined model of survey responses and administratively enumerated records. This provided a unique challenge for the new and updated concepts. While a historic sex concept was collected in previous censuses, this, or available administrative data sources with broad coverage did not directly correlate conceptually to the new gender concept, or the updated sex at birth concept. These concepts are currently are a key input into may other outputs including our family and household census outputs, as well as our official population estimates. As we required full coverage for these concepts within the census dataset, a new statistical imputation methodology was produced to balance retaining individual accuracy where possible, representativeness of the output categories at a population level, along with appropriateness of use and ethical considerations. No alternative data sources were suitable or provided relevant coverage for the concepts of sexual identity and variations of sex characteristics.

The census provides quality assessment of all output concepts across three metrics of 1) data sources and coverage, 2) consistency and coherence, and 3) accuracy of responses. Quality assessment of the new and updated concepts had to balance a range of considerations , including: differences in concept, differences in collection method for comparable sources, and any potential bias in the census responding population compared with administratively enumerated records. Specific quality issues were identified with variations of sex characteristics response data which have resulted in some restrictions on the output data for this concept.