

UN FOC Group on social and demographic statistics webinar: “Building Modern and Resilient Population Data Systems to Enhance Data Quality, Improve Cost Efficiency, and Policy Relevance” – 3 October 2024

– Vince Galvin

Challenges and opportunities in New Zealand’s population data system setup – combined population data system

- **Brief History**

Response rate in the 2018 Census was lower than expected. A judgement was made that the response rate (85%) was too low to use traditional imputation.

We used the 10 years of research we had done into measuring the population with admin data to produce a census file by statistically linking the Census file to our admin population files.

With a few exceptions the Census file was used for its main purposes.

In the 2023 Census we designed this solution into our systems. We are just about to do the main data release for this Census.

- **Original Challenges**

The approach to 2018 developed in the context of a Crisis so we had to deal with external scepticism.

Rapid, unannounced transition to using administrative data raised concerns about social license and privacy to national attention.

In some of our user community, there was some distrust of the quality of administrative data.

- **Challenges in designing and implementing the approach**

Importance of keeping the big picture view of how the addition of admin data works as a system - it’s a little easy to get very engaged in looking at the process decision by decision.

The last few percent to be added to the file were for populations that were not so well reflected in the admin data – the quality issues were not clear cut.

The thresholds for inclusion into the file must be determined – how will this be governed to preserve external confidence?

We do not have a unique identifier we can use so we must make some random deletions to account for missing links – this complicated assessing quality.

The file will have types of errors that Census files have never had before (especially with geography of individual records) so users need to be advised that this is going to happen.

The file will be more variable in quality across the file than any previous file so there is a need to find simple ways of explaining this.

- **External Confidence**

Criticality of explaining the approach to as much of the user community as possible.

Transparency of decisions taken and having material available on sensitivity (of the whole process) are important.

We established an External Review Committee and the key figures within the committee had to be available to the media.

We developed and tested a variable quality rating system to understand “order of magnitude” quality issues.

- **Benefits of our Situation**

We have run the 2023 Census using a mixed model by design and we have (by applying our quality rating system) made useful improvements over 2018.

I think our population measurement is more robust – a greater range of data is confronted in building the file.

Administrative data has some unique opportunities, timeliness, longitudinal measures and (ironically) some measurement possibilities.

We are working towards being able to move away from directly collecting Census data.