

Adaptive Design for the 2023 Census

~~Coverage~~ methodology

Nikolai Kondratev, Naeimeh Abi

Population support and Coverage | Kapinga Taupori

Methods & Design | Ngā Ritenga & Hoahoatanga

Stats NZ | Tatauranga Aotearoa | stats.govt.nz

The journey

- What is Census and Census Coverage?
- Post-Enumeration Survey (PES)
- Standard Design of PES
- Adaptive design of PES
 - Non-response patterns
 - Late Census responses
 - Total uncertainty of coverage measurement
 - Census minimum data capture
- Coverage measurement for the admin-based census

What is Census and Census Coverage?

Census and Coverage Survey

- A census is the official count of the people and dwellings
 - Along with other information of people residing in the country
- 2023 Census is a combination of responses and admin data
- All types of the census have some degree of coverage issues, such as:
 - Under-coverage – people missed by the census
 - Over-coverage – people counted more than once, or should not have been included into census at all
- To adjust census counts, Stats NZ conducts Post-Enumeration Survey (PES) as a coverage survey
 - To measure census under-coverage and over-coverage rates
 - To produce an input for estimated resident population

Post-Enumeration Survey

- Selecting sampling areas
- Over-sample the areas with historically low census coverage

- Finding which PES records were in the census file

- Modelling the relationship between coverage indicators and demographic variables
- Adjusting census counts using measured coverage rates

Sample design

Data collection

Linking to census

Derivation of under- or over-coverage indicators

Model-based estimation

- Dwelling enumeration
- Interview of sampled dwellings

- Identifying in-scope records
- Assigning under- or over-coverage indicators

Post-Enumeration Survey and its Standard Design

In order to make this method work, we need to ensure that we meet following **statistical assumptions**:

- **Causal independence** – the likelihood of being in Census, doesn't affect the likelihood of being in PES
- **Homogeneity of capture** – all records have the same likelihood to be captured in a given list (this is achieved sub-setting the dataset into sub-populations)
- **Perfect linking** – no missed links between same individuals in census and PES, and no incorrect links
- **No erroneous inclusions** into the census file and population counts
- **Closed population** – no entries or exits to or from the target population between Census and PES (not fully achievable, but relatively ignorable)

Standard design

The standard design includes all statistical design components involved in coverage estimation, as well as any dependencies between the statistical components and the sample survey operation.

- Producing high-quality estimates for Māori
- Scope and target population definitions
- Sample design
- Questionnaire content
- Dwelling enumeration method
- Linking methodology
- Treatment of missing data
- Coverage estimation methodology

Standard design was published on 22 March 2023

<https://www.stats.govt.nz/methods/2023-post-enumeration-survey-standard-design-for-coverage-estimation/>

Post-Enumeration Survey outputs and KPIs

Post Enumeration Survey produces:

- Under-coverage and over-coverage rates
- Resident population counts on census night
- Produced counts have uncertainty in them
- More uncertainty, means less precise results
- Having a good enough sample, good response rates, and high-quality data ensures that we meet our precision targets for all sub-populations (*e.g.*, Māori-descent, Pacific, young adults)

Adaptive Design Topics

Adaptive design

A lesson from 2018 Census Coverage. It's about being as proactive as possible!

The purpose of the adaptive design is

- To consider the risks associated with the various statistical assumptions made in the standard design
- To detect when a risk becomes an issue
- To have prepared statistical responses

Adaptive design: non-response patterns

- Decrease of PES response rates reduces the precision of predicted population counts
- Response rate decrease has worse effects on smaller sub-populations
- Not only lower PES response rates, but also lower census response rates deteriorate the final precision of PES results
- Dependent non-response in PES and census leads to the bias and under-estimated population counts

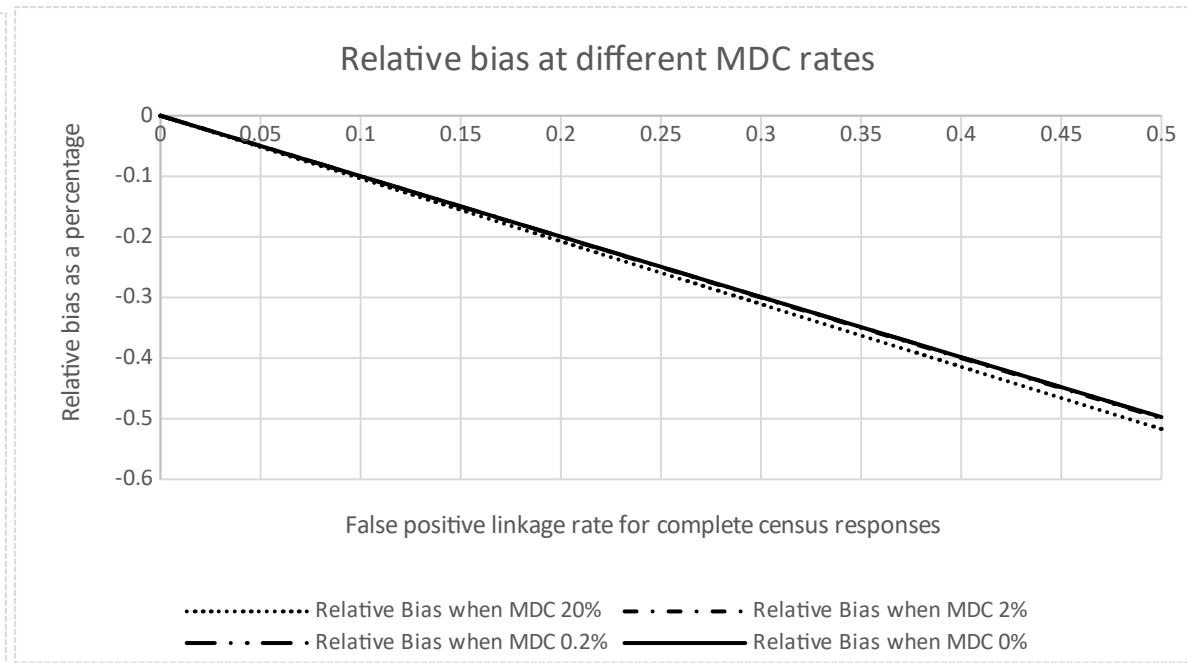
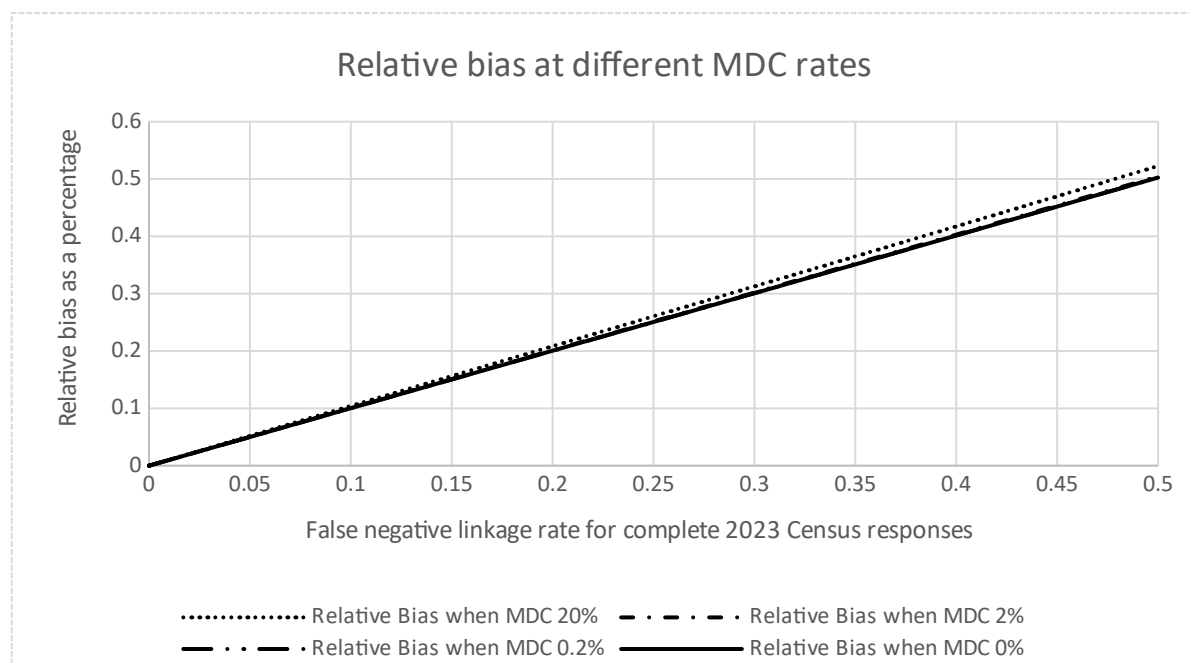
Adaptive design: late census responses

- Any census response submitted after start of the PES is labeled a "late response"
- The main concern is that these responses may violate the assumption of independence between census and PES:
 - *PES stimulates people to fill out the census*
 - *People are less likely to fill out late response because they already responded to PES*
- We've developed a plan for screening late response patterns and detecting possible dependencies in the census and PES data sets
- We also checked if the Tropical Cyclone Gabrielle, which affected Census activity, was likely to cause unexpected late responses and abnormal patterns

Adaptive design: minimum data capture (MDC)

- A strategy of completing census paper forms to facilitate respondents to fill out a reduced number of fields to count as a completed response.
- The concern for PES methodology is that this reduction of data might increase the amount of linkage error.
- We investigated the impact of different level of MDC for different rate of false negative/positive linkage rate on the relative bias of population estimates.

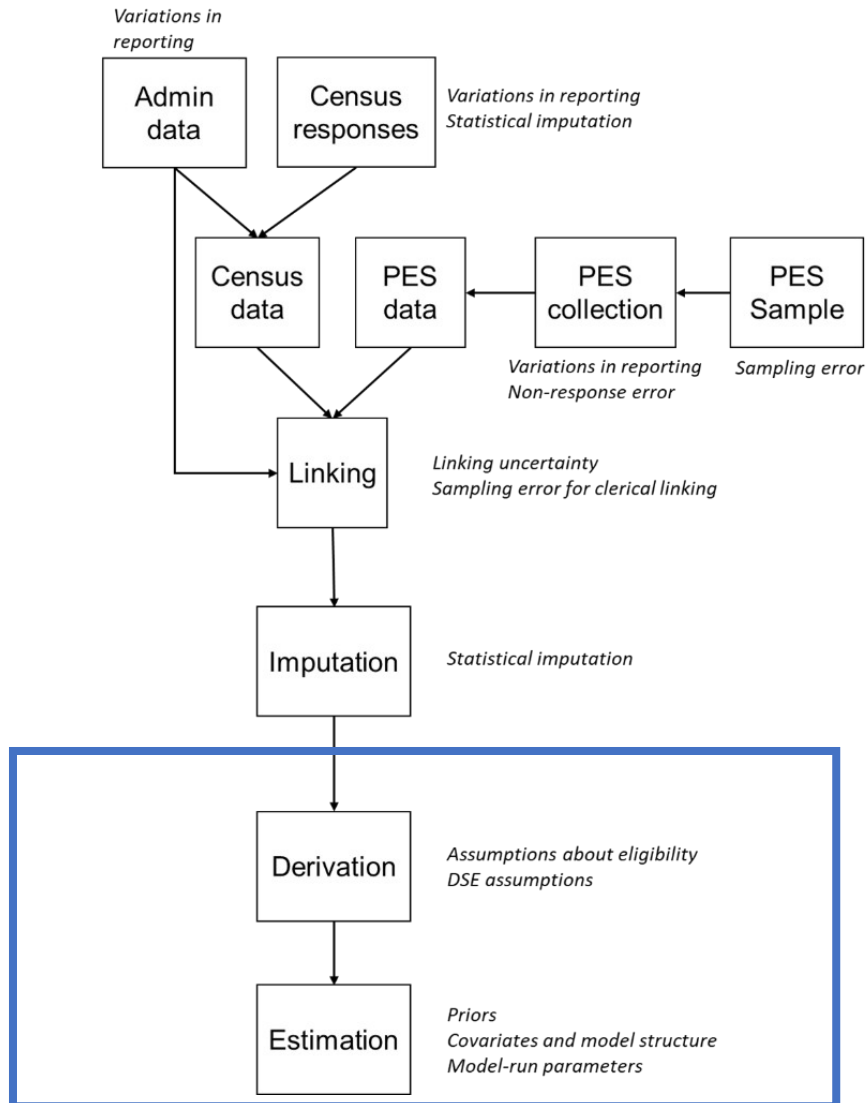
Adaptive design: minimum data capture (MDC)



- Results show that the increase in linkage error caused by MDC is negligible and well within the margin of error for census coverage estimates.
- We recommend that MDC responses are linked in the standard way, taking care to reduce false positive and false negative linkage error as described in the PES 2023 Standard Design

Adaptive design: uncertainty measures and total uncertainty of the process

PES uncertainty sources



- Current methodology does not reflect the uncertainty in data inclusion and estimation model selection
- The literature and statistical methodology of uncertainty in data inclusion is currently limited
- The uncertainty in model selection can be accounted for using Bayesian model averaging or stacking approaches

Adaptive design: toolset to handle the “extreme” coverage patterns

- By extreme coverage patterns we mean the unexpected low census coverage rates in particular sub-groups
- For example, events like the Tropical Cyclone Gabriel may have affected the census respondent behaviour
- These situations can be solved introducing the elements into the model to account for the large differences in census coverage between different sub-populations

Adaptive design: review and publication

- Quality Assurance Panel (QAP) – comprising of internal and external experts to endorse our methodology

Dr James Brown – Professor of Official Statistics, University of Technology Sydney

Andrew Sporle – Director, iNZight Analytics, and Department of Statistics, University of Auckland

Dr Wing Cheuk Chan – Public Health Physician, Population Health – Counties Manukau DHB

Dr Hannes Diener – Senior Design Analyst, Stats NZ Research and Development Hub

Megan Parry – Senior Insights Analyst, Stats NZ Te Tohu Rautaki - Angitu Maori

Christine Bycroft – Principal Statistician, Stats NZ Methods & Design team

Kim Dunstan – Senior Insights Analyst, Stats NZ Population Insights team

- Adaptive design will be published on 6 December 2023

Future of PES in the admin-based population estimation

- PES initially was designed to measure the response-based census coverage
- In 2018 PES measured the coverage of the combined response + admin census; same is planned for 2023
- PES-like surveys can be used to measure the coverage of admin lists, and also to assess and improve the quality of attributes, and the research in this area is currently happening at Stats NZ
- If you're interested in non-coverage-survey-based estimation methodology, you can attend a talk by Lucianne Varn tomorrow