2023 Post-enumeration Survey: Adaptive design for coverage estimation

Citation Stats NZ (2023). *2023 Post-enumeration Survey: Adaptive design for coverage estimation*. Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104963-6

Published in December 2023 by
Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

Contact
Stats NZ Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4600
www.stats.govt.nz

Contents

List of tables and figures

List of tables

List of figures

Purpose and executive summary

Purpose

This paper discusses the adaptive design developed by the 2023 Census Coverage project for the 2023 Post-enumeration Survey. The adaptive design aims to be proactive in assessing the potential risks associated with the statistical assumptions made in the standard design. The focus of this work is to plan to identify when external factors to the project might put key assumptions at risk, and to ensure that these risks can be mitigated.

The Post-enumeration Survey (PES) is a household survey undertaken shortly after the census to evaluate the completeness of census coverage, and to report on census response rates. All the key design components of the 2023 PES (hereafter referred to as 'PES') are published in 2023 Post-enumeration Survey: Standard design for coverage estimation. The 2023 PES standard design (hereafter referred to as the 'standard design') outlines the design decisions made for the PES along with how the data would be processed and used to develop coverage estimates for the 2023 Census.

Nevertheless, it should be acknowledged that each census cycle is unique, and as a result, not all aspects of the standard design may be achievable in the 2023 Census (hereafter we will call it 'census') and in the PES and its resulting coverage estimation. Therefore, the 2023 Census Coverage project has also developed an adaptive design.

The adaptive design was refined based on lessons learned during the standard design process. The most important area of focus is addressing the statistical risks and assumptions present in the standard design. This includes the fundamental assumptions that form the basis of the dual system estimation (DSE[1]) methodology, as well as additional assumptions related to sampling.

The 2023 PES standard design specifies work items to include in the scope of the adaptive design (Stats NZ, 2023 a). Within the adaptive design paper, some of these items have been thoroughly investigated, while others have been deferred for future consideration based on their prioritisation within the given time frame. A summary of work pieces considered for scope of adaptive design is presented in Table 1. The last column of Table 1 shows the status of consideration for each item in the adaptive design.

[1] A summary of the DSE assumptions can be found here: Why do we do the Post-enumeration Survey (PES), and why is it hard? (Stats NZ, 2021)

Table 1 Summary of work considered for scope of adaptive design (copied from PES standard design)

Work piece	Description	Included in scope	Status of consideration in the adaptive design
Risks and assumptions	This work focuses on the statistical risks and assumptions associated with the standard design. Two key components of the work are to develop mechanisms for assessing the risks and assumptions (identifying when it shifts from a risk to an issue) and developing contingency methods for a subset of the more likely scenarios.	Yes – priority 1	Investigated
Treatment of late census returns	This work investigates three different approaches for treating late census returns.	TBC[1]	Investigated
Gender research	This work explores the options available for producing population estimates for all three gender categories should the PES not have sufficient sample of all three categories to produce high quality estimates.	Yes – priority 2	Investigated the impact of PES sample on variance. Did not investigate alternative designs as concluded to accept the variance impact.
Uncertainty measures	This work responds to a recommendation from a 2018 PES external reviewer to explore ways of better communicating the subjective nature of balancing bias in PES estimation	Yes – priority 3	Investigated
Differences in reporting of demographics	This work builds on research completed as part of the standard design. Completion of this research requires collaboration and partnership with external communities.	No	
Investigation into census combined model components	This work involves understanding the different components of the census combined model, in particular admin enumerations into dwellings, and admin enumerations into meshblocks.	No	
Investigation into usual residence concept	This work involves investigating the various conceptual and operational definitions of usual residence used by census, PES, and the ERP. The goal of the research was to understand how the different definitions relate to and impact PES.	No	

[1] At the time of drafting the PES standard design, the treatment of late census returns was deemed stable and not identified as a significant risk, hence its priority was labelled as 'TBC' (to be confirmed). However, this prioritisation shifted when PES was implemented in the field, particularly following the impact of Cyclone Gabrielle, which caused widespread disruptions to parts of the North Island of New Zealand in February 2023.

In addition to the items we investigated in Table 1, we also studied the impact of using 'minimum data capture' in the census on the PES. Minimum data capture is a method to facilitate census respondents completing a reduced number of fields on paper forms to count as a census response. The intention of this is that it gets a response where otherwise one might not be possible. This report summarises the work done to simulate potential risks and evaluate how they impact the PES estimates.

Furthermore, we considered the possibility of using a mixture of distributions for TALB[1]-level random effects in the model. Preliminary investigations showed that that method is difficult to implement, so would offer little value to continue for the 2023 PES.

Summary of contents

The work is separated into four major areas, each comprising one chapter in this report:

1. Impact of PES and census challenges on the estimates
2. Treatment of census late responses
3. Uncertainty measures for estimates
4. Census minimum data capture.

Impact of PES and census challenges on the estimates

Chapter 1 focuses on the impact of unexpected outcomes of the PES and census data on the estimates. This work will be specifically looking at the impact of low response rate for the PES on coverage estimation and estimating small sub-groups such as 'another gender' subgroup. This chapter also investigates unexpected outcomes in the 2023 Census data. For this work we specifically look at the impact of low census coverage on the estimates, the risk of having dependency between the PES and census, and the impact of the existence of heterogeneity (non-homogeneity) of capture in the census on the estimates.

The outcome of work conducted in this chapter has demonstrated that the PES design is robust enough to provide reliable estimates for large subpopulations, even in situations with low PES response or census coverage[2]. However, it is important to note that estimates related to small subgroups, such as 'another gender', have higher uncertainties. Furthermore, the findings suggest that to minimise bias in the final estimates, it is crucial to include all significant covariates, such as age groups, in the model estimations.

[1] Territorial Authority Local Board (TALB). TALBs are geographical units that contain multiple primary sampling units (PSU).

[2] PES 2023 had a successful data collection phase (more than 80 percent national achieved sampling rate).

Treatment of census late responses

Chapter 2 provides details of treatment of census late responses. In addition to investigating the treatment of late census responses as a general issue, this chapter also discusses the impact of the discrepancy in PES field interviewing start dates (one for Auckland and the other for the rest of the country) on the number of late responses and, consequently, our estimates.

The primary outcome of this study recommends retaining the standard design approach of 'remove and add' as the default option. However, a final decision on the treatment of late responses for the 2023 cycle will depend on what we find in the 2023 Census file and 2023 PES responses. Furthermore, our findings indicate that the misclassification of late/on-time responses due to considering two collection start dates is negligible, suggesting that there is no immediate need for alterations to the estimation process.

Uncertainty measures for estimates

Chapter 3 reviews all major uncertainties in PES and provides detailed descriptions of various ways of measuring and presenting uncertainties in the PES estimation process.

Outcomes of this chapter show that for 2023 coverage estimation, it is advisable to employ model averaging when multiple models yield suitable fits to the PES dataset and provide viable outputs. This approach helps account for uncertainty in the model selection process. However, it's important to note that uncertainties related to unmeasurable assumptions, such as the probability of late census responses violating the census-PES independence, cannot be quantified or incorporated into the analysis.

Census minimum data capture

Chapter 4 focuses on the impact of census minimum data capture on the PES. This chapter aims to identify the possible increase in linkage error caused by the use of minimum data capture as a tool in the census and what this increased linkage error will do to the total estimated resident population.

The results obtained from the work conducted in this chapter indicate that the rise in linkage error resulting from minimum data capture is minimal. The primary recommendation is to pay close attention to false positive and false negative linkage errors, as outlined in the standard design (Stats NZ, 2023 a) and the linking design paper (Stats NZ, 2023 b), and to account for this through the planned clerical linking process.

Executive summary

The findings from this study demonstrate that the works in this study align with the key deliverable items stated in the standard design.

Stocktake and risk assessment of assumptions: Adaptive design assesses various assumptions made in the standard design, such as the robustness of the PES design, the impact of low PES response, the impact of low census coverage rate, and the treatment of late responses. These assessments prioritise resource allocation and focus.

Methods and diagnostic tests: Adaptive design does not explicitly mention diagnostic tests; however, we recommend methods such as model averaging to account for uncertainty and address assumptions' validity.

Methodology options for repair: We suggest retaining the standard 'remove and add' approach as the default option, which can be seen as a methodology option for handling late responses.

Framework for extreme assumption violations: In the adaptive design we imply a framework for determining whether to make more significant changes to the estimation framework, such as the treatment of late responses. This involves considering additional information before making a final decision.

Chapter 1: Impact of PES and census challenges on the estimates

1.1 Introduction

In recent years, there has been a steady decline in response rates across almost all household surveys as well as the census (Stats NZ, 2020). The potential effect of these low and unequal response rates in the PES and census on the accuracy of coverage estimates is a source of concern for the PES project, particularly if the achieved PES sample is not representative of the population. In this chapter, we examine how non-responses in both the PES and census may affect the quality of the PES estimates, which is typically measured in terms of precision and bias. Precision reflects how close the multiple measured values are to each other, and bias shows the difference between the true and estimated values. Addressing an uneven sample necessitates the use of specific methods and adjustments in estimation to account for the situations, when the initial design assumptions do not hold anymore. PES estimations are constructed using Bayesian models in conjunction with dual system estimation (DSE), and in this context, the issues of uneven response and violation of representation assumptions are connected to the estimation methods through independence and homogeneity of capture DSE assumptions.

Therefore, in addition to the impact of low PES response rates and low census coverage rates, two additional risks were considered in this paper for PES estimates: the dependency between the PES and the census response patterns, as well as the existence of unexpected variations in census coverage, which are not accounted for in the model.

This chapter focuses on assessing the impact of these challenges on the accuracy of PES estimates. The first part (1.3 PES response rates and small population size) examines the influence of changes in PES dwelling response rates on coverage estimates and the level of uncertainty in the estimates. This section also explores the accuracy of PES estimates for smaller sub-groups, such as individuals with an 'another gender' response.

The second part of the chapter (1.4 Challenges in census) delves into the uncertainties and biases in the estimates if the census encounters the aforementioned challenges.

1.1.1 Summary of findings

There were several key outcomes of this chapter:

- It was observed that even in cases of low PES response, the credible intervals (CIs) of the estimates were within the bounds of the PES Key Performance Indicators (KPI)[1]. Since the CIs

[1] Key performance indicators (KPIs) for the 2023 PES are determined by the 2023 Census Coverage project in consultation with 2023 Census and the team responsible for production of the ERP. The proposed KPIs for 2023 are a development of the targets set in 2018 PES. These KPIs cover survey response, completeness of the sample-

are likely to be within KPI bounds even during low PES response, we recommend no adjustment to the standard design in this scenario.

- It was observed that even in cases of high PES response rate or census coverage, it is expected that there will be wide credible intervals for coverage estimates of the 'another gender' group, therefore greater uncertainty for the estimates of this group. In fact, we will accept the risk of greater uncertainty of this group's estimates and can conclude that the size of a sub-population in relation to the total population has an important impact in the width of credible intervals.
- Findings from the unaccounted heterogeneity simulation showed a bias in the final estimates when a well-understood covariate (for example, age group) has been removed from the model. When we left out a key factor like age group from our model, it didn't noticeably affect the overall national estimates, but it did cause some inaccuracies when we looked at smaller sub-groups. So, general suggestion is to be careful and include all important factors when building the initial estimation model. If we keep observing biased or implausible estimates for some subgroups, our strategy is to improve the model using different attributes and to draw from the experience gained in previous PES cycles.

1.2 Simulating the target population

The investigations in this chapter primarily rely on simulation studies. To conduct these studies, we initially constructed a simulated target population based on a 2018 Census file. A similar semi-synthetic population was previously created to determine an appropriate sampling design for the PES (Stats NZ, 2023 a).

We first describe the construction of the simulated population and its census and then run our simulation study on it.

1.2.1 Simulating a target population

The first step of creating a simulated population was to create a target population which would accurately represent the dwelling and individual population structure of New Zealand. Our priority when doing this was to preserve the complexity of the NZ population's sociodemographic and geographic structure, and this was prioritised over matching the current population counts.

---

frame, and precision of survey outputs, and they will be used to inform the sample design, field operational planning, and modelling of coverage data.

2023 PES KPIs are:

- achieved sample rate at national level ≥ 82 percent
- achieved sample rate at subnational (regional council areas) level ≥ 80 percent
- survey response rate at national level ≥ 90 percent
- dwelling-frame accuracy (enumeration) ≤ 2 percent dwelling under-coverage (Stats NZ, 2023 a).

To create the simulated target population, we select all permanent private dwellings from the 2018 Census dataset using the same criteria used for creating the Household Sampling Frame (HSF)[1]:

- dwelling type code ‘1’ – private dwellings
- dwelling status code ‘11’ or ‘31’ – occupied or under construction dwellings
- dwelling description code not ‘06’, ‘07’, ‘08’, ‘09’ - exclude mobile dwelling not in a motor camp improvised dwelling, or shelter, roofless or rough sleeper, non-private dwelling.

We then populate the dwellings with the usual residents from 2018 Census dataset. The final population of dwellings and individuals had the following limitations and simplifications:

- did not include non-private dwellings, and all included dwellings were in scope for PES, which will not be the case for the real survey
- did not include empty dwellings, since they would not be able to be used for simulated census, PES, and coverage estimation processes
- did not include admin-enumerated records from the 2018 Census dataset that were not associated with a dwelling. Allocating these records into a ‘synthetic dwelling’ may have distorted the real population structure and would come with technical complexity.

1.2.2 Creating PES non-response

In the second step of creating the simulated target population, some dwelling records need to be removed from the file to reflect the PES non-response. The process of removing the records was carried out under the assumption that the patterns of non-response in the PES would be similar to those observed in the 2018 PES.

Using the 2018 PES datasets, a response indicator was assigned to each dwelling record. This binary indicator, represented by 1 for ‘responded’ and 0 for ‘did not respond’ was determined using a Bernoulli random variable. The probability of a dwelling responding was specific to each Territorial Authority/Local Board (TALB) area using the below equation:

$$\left[R_{th} | p_{response_{th}}\right] \overset{indep}{\sim} Bernoulli\left(p_{response_{th}}\right); \qquad h = 1, \dots, N_t^{dwell}, \quad t = 1, \dots, N^{TALB}$$

Where R_{th} is PES response indicator for a dwelling h in a TALB t, $p_{response_{th}}$ is the probability of response of dwelling h in a TALB t, N_t^{dwell} is the number of dwellings in TALB t, and N^{TALB} is number of TALBs in the population.

1.2.3 Simulating PES

Once the PES response indicator was incorporated into the simulated target population, synthetic survey data could be chosen. The survey collection process replicated the actual PES process and followed a multi-level collection procedure. All primary sampling units (PSUs) were allocated across the designated

[1] The HSF is the standard sampling frame Stats NZ uses to select samples and manage overlap control between a variety of household-based surveys which run either with Stats NZ or another government department.

strata. The required number of PSUs was sampled from each stratum to reach the desired number of dwellings. A cluster size of 11 dwellings per PSU was selected, which was identified as an optimal size in the 2018 PES sample design. Subsequently, non-responding houses were excluded from the sample.

It is important to note that this simplified approach did not address all potential issues, including:

- dwelling enumeration errors
- individual level non-response
- data reporting and collection problems
- PES to census linking errors.

1.2.4 Simulating census coverage

To simulate census coverage in the target population, each individual record was assigned binary indicators for under-coverage and over-coverage. First, over-coverage indicators were assigned to the records. Then, under-coverage indicators were assigned to all records that were not flagged as over-coverage. The probabilities of each record experiencing under- or over-coverage were determined using the outputs of the PES 2018 and varied depending on TALB, four binary ethnicity variables (Māori, Pacific, Asian, Other), binary Māori-descent indicator, the three-categories of gender (male, female, another gender), and six variables representing 15-year-age groups. The under-coverage and over-coverage used in the simulation study are formulated as below:

$$\left[O_{ti} | \boldsymbol{X_{ti}}, p_{over_{ti}}\right] \overset{indep}{\sim} Bernoulli\left(p_{over_{ti}}\right); \qquad i = 1, \dots, N_t^{ind}, \quad t = 1, \dots, N^{TALB}$$

$$\left[U_{ti} | \boldsymbol{X_{ti}}, p_{under_{ti}}, O_{ti} \neq 1\right] \overset{indep}{\sim} Bernoulli\left(p_{under_{ti}}\right); \qquad i = 1, \dots, N_t^{ind}, \quad t = 1, \dots, N^{TALB}$$

Where O_{ti} and U_{ti} are over-coverage and under-coverage indicators for an dwelling i in a TALB t respectively, $p_{over_{ti}}$ and $p_{under_{ti}}$ are the probability of an individual i in a TALB t being under-coverage and over-coverage respectively, N_t^{ind} is the number of individuals in TALB t, and N^{TALB} is the number of TALBs in the population.

This simulation tries to achieve the lowest level of granularity in the coverage patterns that can be obtained from the 2018 PES outputs.

In this simulation step the following assumptions were made:

- over-coverage and under-coverage are mutually exclusive and independent as observed in the 2018 estimation process
- the coverage patterns in the simulated target population align with the patterns observed the 2018 Census.

1.3 PES response rates and small population size

To investigate the impact of low PES response rates on the estimates, simulation studies were initially conducted at the national level. This allowed for an understanding of the relationship between changes in PES response rates and their effects on estimate uncertainty at the national level to be explored. Then, simulations were carried out for the following sub-groups:

- Māori ethnicity and Māori descent
- Pacific ethnicity
- age groups: 0–4, 5–14, 15–29, 30–44, 45–64, 65–74, 75+
- three-category gender (that is, *'male', 'female',* and *'another gender'*).

The objective of conducting these simulation studies for sub-groups was to explore how the sizes of sub-group populations and unequal response patterns among them influence the uncertainty of the estimates.

Variables that were used in the simulation studies are listed as follows.

- **Explanatory variable: dwelling response rate**

  In the simulation studies, the dwelling response rate served as the manipulated explanatory variable. The simulations were run by changing the response rate and observing its impact on uncertainty measures. In most cases, a 90 percent response rate was used as the base result, then the response rate was systematically decreased by even intervals to see the effect on the response variable.

  Instead of directly referring to the response rate, in this paper we will use the term 'non-response rate' as it aligns better with the context of the conducted simulation studies.

  It is important to note that the non-response rates were determined at the dwelling level. If a dwelling did not respond, all individuals within that dwelling were excluded from the final PES sample. Additionally, the exact number of individuals residing in non-responding dwellings could not be determined due to the lack of contact. Therefore, the response rates were considered at the dwelling level.

  While this study did not consider dwellings with a mixture of responding and non-responding individuals, the coverage estimation model was run on individual records. Hence, the estimation process required knowledge of the number of people within each sub-group. In the simulations, we have used response rate (except for gender, where we directly set the proportion of each group) as the explanatory variable. Nevertheless, the primary focus was on the number of records available for modelling and how that quantity influenced the estimates and uncertainty.

- **Response variable: credible interval width**

  The main focus of the simulations conducted in this study was to assess the level of uncertainty in the estimates. A commonly used measure of uncertainty in Bayesian inference is the Bayesian credible interval. In this section, the provided plots illustrate the relative width of a 95 percent credible interval, which serves as an indicator of uncertainty.

In this study, we calculate a 95 percent credible interval by determining the range between the 2.5 percent and 97.5 percent percentiles of the posterior distribution. This measure aligns with the uncertainty measure for PES KPIs. The simulations aim to investigate the extent of change in response rates that would still allow us to meet the desired PES KPIs.

1.3.1 Results of the simulation studies

PES response rate

The results of the simulation studies are presented in graphs found in **Appendix 1**. These graphs depict the relationship between the response rate (x-axis) and the width of the credible interval (y-axis).

The findings from the simulation studies demonstrate that as the response rate decreases, the uncertainty in the estimates increases. This holds true for all the considered sub-groups. However, the increase in uncertainty intervals is more noticeable for smaller population groups (for example, 'another gender') compared to larger population groups (for example, 'male' or 'female') as the response rate declines. An increase in non-response will increase the uncertainty according to the power law[1]. For instance, a decrease in non-response from 95 percent to 90 percent has a lesser effect compared to a decrease from 85 percent to 80 percent.

Furthermore, the findings reveal that PES non-responses have a greater impact on estimates for population groups that are not mutually exclusive (for example, ethnicity where an individual can have multiple ethnicities), with smaller population groups (for example, Māori ethnicity and Pacific ethnicity) being more affected than larger population groups (for example, European ethnicity). This is due to factors such as individuals having multiple ethnicities and the methodology used to define ethnicity at the household level. On the other hand, when observing estimates for mutually exclusive groups like age or geographical location, only the uncertainty of the modified group is affected, while the other groups remain unchanged. This is logical since an individual cannot belong to more than one age group, and thus changes in one group's rate should not affect the others.

The results indicate that even in cases of extreme non-response, the relative uncertainty remains stable. This suggests that for most population groups, it is still likely to meet the PES KPI targets. From our experience in household surveys, we know achieving a response rate below 50 percent is unlikely and it is important to note that this is a simplified simulation and depends on various factors, including census performance.

In conclusion, it is improbable that a lower-than-expected non-response rate would necessitate statistical mitigation. The findings emphasise that in the event of inadequate response, it is important to target on responses in areas with groups that constitute a smaller proportion of the population.

[1] In statistics, a power law is a functional relationship between two quantities, where a relative change in one quantity results in a proportional relative change in the other quantity, independent of the initial size of those quantities: one quantity varies as a power of another.

PES non-response as the source of bias

High PES non-response, which is independent of census non-response, will affect our ability to meet precision KPIs, and increase the uncertainty of our predictions, but should not cause bias in the coverage estimates, if the sample is still representative of the true population. The PES sample may have biased capture of census coverage, if PES non-response is correlated with census non-response. We run the simulations where we introduced a dependence component between PES non-response and census coverage (**Appendix 2**). The effect of the dependence component was increased in the simulation runs, thus, making PES sample include fewer under-coverage records in each run. The observations showed that the larger dependence of PES non-response from census non-response leads to the under-estimation of the subgroups with higher non-response.

Small population size

Estimates for a third gender (that is, 'another gender') are inevitably going to have high uncertainty due to the initial small sample size compared to the other two gender groups (that is, 'male' and 'female'). Generally, analyses (including imputation, coding rules, and any differences in reporting between census and PES) of small subgroups can produce spurious results. In this situation with a wide margin of uncertainty it would be difficult to report and give an interpretation of the results.

1.4 Challenges in census

The accuracy of the PES estimates is heavily reliant on the performance of the census. This section of the study examines the influence of census-related challenges on the PES estimates. The investigation is divided into three distinct parts:

- assessing the effects of variations in census coverage rates on the PES estimates
- analysing the impact of interdependencies between two lists, specifically the census dwelling frame and the PES dwelling frame, on the PES estimates
- investigating the consequences of unexpected non-homogeneity of capture on the PES estimates.

Each part will be discussed in detail in the subsequent sections.

1.4.1 Changes in census coverage rates

Investigating the influence of changes in census coverage rates on the PES estimates closely resembles the work conducted to explore changes in PES response rates. So here we use almost similar analysis that was done above. This section of the study primarily focuses on the impact of census coverage rate on the PES estimates for the following sub-groups:

- Māori ethnicity and Māori descent – these sub-populations were selected because they have the narrowest KPI target range (+/- 1.0 percent). Focusing on these groups allows for a closer examination of the impact of census coverage rates on achieving the desired level of accuracy for Māori-related estimates.

- 'Another gender' – this group was chosen based on the findings from previous simulation studies conducted on PES response rates. The results indicated that estimates for this category had higher levels of uncertainty due to its smaller proportion within the population. Therefore, investigating the impact of census coverage rates on this group helps understand the effects on estimates of population size where the population had low representation.

In both cases, the simulation studies aim to compare scenarios when the census coverage rate is high, for example, 90 percent, and then systematically decrease the coverage rate by even intervals to find the effect of census coverage rate on the estimate uncertainty measures.

The explanatory variable and response variables considered in the simulation studies are as below:

- **Explanatory variable: coverage rate**

  In this simulation, the variable that is manipulated is the census coverage rate. The simulations were conducted by altering the coverage rate and examining its impact on uncertainty measures, specifically the width of the credible interval. The simulation study began with a base scenario where the coverage rate was set at 99 percent, indicating that 1 percent of the population was not accounted for in the census. Subsequently, the coverage rate was reduced to 95 percent in the next step of the simulation. Further iterations involved systematically decreasing the coverage rate in 5 percent increments to observe the corresponding effect on the response variable.

- **Response variable: credible interval width**

  Like the previous simulation studies, the credible interval width, as an uncertainty measure, is considered as the response variable.

1.4.1.1 Results of the simulation studies

This section provides an overview of the results obtained from the simulation studies, and the complete set of results can be found in **Appendix 2**.

The findings reveal that estimates derived from a census with higher under-coverage have greater uncertainty in the PES estimates. Consequently, if the census has lower coverage rates, it will have an impact on the ability of the PES to meet its KPIs. The results specifically demonstrate that the 'another gender' group follows the same pattern, whereby decreasing the census coverage leads to an increase in the uncertainty of estimates for this group.

It is worth noting that even with 99 percent census coverage, achieving a high level of precision in the estimates (as set for the 'male' and 'female' genders) is challenging. This finding is crucial as it indicates that even when the census performs better than the census KPI for national coverage, producing estimates with a low level of uncertainty is difficult. It is important to emphasise this outcome when discussing expectations with customers. Additionally, while there is no specific KPI set for the 'another gender' group like there is for the 'male' and 'female' groups, it is crucial to acknowledge this outcome in discussions with relevant stakeholders.

1.4.2 Dependency between two lists

In this section, the potential risk of dependency between PES and census is investigated. A dependency issue can occur when an eligible person who did not respond to the census has a higher probability of being non-response in PES when compared to someone who did respond to the census. This correlation of census and PES non-response may lead to the PES sample not being representative of true census coverage patterns in the target population. This overlapping under-coverage problem would result in a bias in the coverage estimate and an underestimation of the true population[1].

In the simulation study, we studied the impact of this dependency of the uncertainty of the estimates. Descriptions of the considered explanatory variable and response variable are shown below:

- **Explanatory variable: coverage rate**

  This variable is the same as the one used as the explanatory variable in Section 1.4.1. However, the key difference in this set of simulations is the introduction of an additional parameter, γ (gamma), to measure the level of dependence. This parameter will be varied in the different simulations. The focus is on modifying γ to reduce the probability of individuals, who were not captured by census, to appear in the PES sample. This is a direct violation of the independence assumption, when PES should have the same sampling probabilities for census respondents and non-respondents. The simulation study commences with a low level of dependence and subsequently increases the level of dependence in subsequent simulations. The aim of these simulations in these sections is to explore the interaction between a declining coverage rate and a worsening dependence situation.

- **Response variable: relative bias**

  The relative bias measures the difference between the estimate of the population and its true value. This is represented as a percentage difference and can be either a positive or negative value.

The simulation study for this part of the study used the median of the posterior distribution of predicted population as a point estimate of the population estimates. To find the relative bias, this point estimate was subtracted from its true value in the population, as shown in the below equation:

$$Relative\ bias\ \% = \frac{(Q50 - True\ count)}{True\ count} \times 100$$

[1]During the simulations, the true population can be designated as the benchmark against which measurements are compared. The true population represents the accurate count of individuals in the population. The objective of the PES is to estimate the size of the true population by accounting for coverage errors in the census. Hence, if all the assumptions related to Dual System Estimation (DSE) are satisfied, the coverage-adjusted estimate should align with the true population count.

1.4.2.1 Results of the simulation studies

The results of the simulation study are reported in **Appendix 2**. The findings show that dependence between census and PES non-response will cause negative bias, resulting in an underestimation of the final estimates, which is further exacerbated by lower coverage rate. The simulation does highlight the importance of maintaining independence between the census and PES. The simulation could be helpful if we observe an unexplained underestimation of our coverage estimations (either at a national or subnational level) and may provide evidence for further exploration of an independence violation.

1.4.3 Unexpected non-homogeneity

In this section, we explored another potential risk associated with ignoring heterogeneity of capture in the population, which was not accounted for through model covariates. Homogeneity is a key DSE assumption – it does not hold at the population level so we use covariables to create the most homogenous groups we can. In the Bayesian context, this has an added benefit of supporting granular outputs. The Bayesian model uses covariates to separate the population into groups with differences in coverage patterns, and the model will adjust to the heterogeneity of capture. This allows us to provide a coverage estimate to a low level of granularity for everyone in the census eligible list. The challenge of adding covariates to the model is finding all the relevant ones and ensuring the created new groups are not too small to apply Bayesian modelling. In this part of the study, a covariate that has a well-understood effect on the model (for example, age group) has been removed. This means that one of the important covariates is not included in the estimation process. This simulation aims to observe this practice on the estimates. For this, we were increasing the under-coverage of the young adult population (15–29-year-olds). Then we fit two models – one which uses the binary covariate to indicate if the individuals belong to the young adult group (correct model) and the other which does not have a youth-related covariate (incorrect model). We ran all simulations with a base coverage of 95 percent, and youth coverage decreased from that value.

The explanatory variable(s) and response variable(s) required for this simulation study are listed below.

- **Explanatory variable: coverage rate**

  Explanatory variables for these simulation studies are the same as those used in the previous simulations.

- **Response variables: credible interval width and relative bias**

  These simulations use the two response variables mentioned in previous sections. These two response variables are: the uncertainty measure that was introduced in Section 1.4.1, and the relative bias, which was introduced in Section 1.4.2.

1.4.3.1 Results of the simulation studies

The findings from the simulation studies, as detailed in **Appendix 2**, indicate that both the correct and incorrect models exhibited a similar range of relative bias in national-level estimates. While the correct model resulted in slightly higher uncertainty, the difference in uncertainties between the correct and incorrect models was not remarkable. This suggests that ignoring a covariate is unlikely to affect the precision of national-level estimates, and any notable differences would only emerge when examining estimates at a more granular level.

However, there is a notable distinction between the correct and incorrect models when providing estimates for different age groups. The findings demonstrate that the incorrect model introduces bias in estimates across all age groups, underestimating the youth age group while overestimating the others. The bias is particularly pronounced in the youth group, with relative bias worsening as the coverage rate decreases. In contrast, the correct model incorporates a covariate specific to the youth age group, resulting in higher uncertainty in the estimates compared to other age groups. This effect is further exacerbated when the coverage rate is lower for the youth age group. There are two reasons for this:

- We introduced a new covariate to identify young adult members of the population, and the youth subgroup represents a relatively small group.
- The coverage rate for the youth group is lower than that of other groups, leading to higher uncertainty. This observation aligns with the findings from section 1.4.1, where lower census coverage corresponded to increased uncertainty.

The unaccounted coverage heterogeneity contributes to bias in the final estimates. Incorporating a new covariate or variable into the estimation model to represent this heterogeneity helps mitigate the bias but also increases uncertainty for the subgroups defined by that new variable. Despite the increase in uncertainty, the simulated outputs still meet the KPIs when the census coverage is sufficiently high (> 85 percent in these simulations).

1.5 PES sample selection on areas affected by Cyclone Gabrielle

The 2023 PES sample was designed and tested to ensure it contained sufficient representation of different subpopulations in New Zealand. According to the PES sampling design, the first stage of sample selection involves the selection of a sample of Primary Sampling Units (PSUs) from the HSF to enable the approved sample design to be achieved. This process was completed by selecting 1,500 sample PSUs out of 23,174 PSUs available in the country.

The second stage of the sample selection consists of selecting eligible dwellings within the selected PSUs. The required PES sample size is 16,500 dwellings and can be reached by selecting an average of 11 dwellings per PSU.

To select the sampling dwellings, we need to have a reliable and accurate frame of permanent private dwellings – a ‘dwelling frame’. The PES dwelling frame was created by the coverage team in 2022. It was designed in such a way that the chance of an individual or dwelling being found by PES is unrelated to the chance of that same individual or dwelling being found by census – ensuring that we meet our ‘causal independence’ coverage assumption.

Cyclone Gabrielle devastated parts of the North Island of New Zealand in February 2023 and caused significant damage and disruption. Many of these cyclone-affected areas would not be accessible for census teams, so there was a risk associated for the coverage project of losing much of the PES sample (either losing whole PSUs or the majority of sampling dwellings in a PSU) due to the limited ability to collect information on uninhabitable dwellings in the cyclone-affected areas.

To overcome this problem, it was decided to select more sample dwellings from the cyclone-affected areas.

Chapter 2: Treatment of census late responses

2.1 Introduction

2023 Post-enumeration Survey: Standard design for coverage estimation (Stats NZ, 2023 a) stated that PES uses Dual System Estimation (DSE) methodology to estimate census under- and over-coverage. One of the assumptions for applying DSE is causal independence between two lists. It can be defined as the likelihood of being recorded on one list (that is, census) having no relationship with the likelihood of being recorded on the other list (that is, PES). To make this assumption hold, the likelihood of being recorded in one list should not influence the likelihood of being recorded on the other list.

The coverage project, which is responsible for measuring over- and under-coverage in census, minimises the risk of causal dependence between PES and census by ensuring operational independence (separate staff, different sources for address information, no overlap of PES field operations with census field operations, strict embargo of PES areas from census). Despite the steps to ensure operational independence, there is still a chance that a respondent could be influenced by PES to participate in census. This problem may specifically occur among people who are considered as late census respondents. A census late response is a census form that is returned to the census team after the PES has started field interviews in any geographic area, regardless of whether the response was from a PES area.

This risk increased as PES ran two collection start dates, one national start date and another a month later for Auckland. PES postponed the start date in Auckland due to additional activity of the Whānau Ora Commission Agency (formerly Te Pou Matakana)[1] working to increase response rates among Māori and Pacific households in Auckland.

In this chapter, we explore the impact of late census responses on estimates and discuss the most effective approach to handle them. The chapter is divided into two parts: the first part investigates the treatment of late census responses as a general issue, and the second part focuses on the impact of the discrepancy in PES field interviewing start dates (one for Auckland and the other for the rest of the country) on the number of late responses and, consequently, our estimates.

2.1.1 Summary of findings

In this chapter, we examined the following three primary methods for handling census late responses in PES estimation.

- Lates as under-coverage: this approach simply removes late responses from the census list and relies on the PES sample and estimation to recover and reflect them in the estimates.
- Remove and return: this method involves temporarily removing late responses from the estimation model and then adding them back into the population after estimation.

[1] Whānau Ora | Home (whanauora.nz)

- Lates as coverage: here, late responses are retained in both the census and PES datasets. Late responses in this method are treated the same as the on-time census responses.

Our study revealed that when there is no sampling bias in PES, all three methods provide unbiased estimates. However, the performance of the methods varies when patterns among late responses exist.

Historically, the 'remove and return' approach was chosen as the default treatment in the PES standard design (Stats NZ, 2023 a), and it produced more reasonable estimates for the 2018 Census and PES datasets than treating lates as under-coverage. Yet, we cannot be certain if this default method will perform equally well for the 2023 Census coverage. Therefore, we plan to evaluate the effectiveness of different approaches in the 2023 PES to ensure accurate treatment of late responses. We will analyse the PES data to see if there are any potential patterns among on-time/ late census responses.

Additionally, our findings in this chapter indicate the following:

- late responses are unlikely to be influenced by or prevented by PES activity
- late responses exhibit a correlation with under-coverage
- the PES sample contained proportionally fewer late responses than the census for specific groups, such as young males and individuals of Pacific ethnicity.

Furthermore, we discovered that the misclassification of late/on-time responses due to the consideration of two collection start dates is not significant. This indicates that there is likely to be no need to make any adjustments to the estimation process based on this factor.

2.2 Treatment of census late responses as a general issue

When census responses are received late, there is a possibility that the knowledge of PES occurring in the field could influence a PES respondent to complete a census form (when they had not already done so). This means that the PES process may inadvertently impact the census returns. To avoid violating the causal dependence assumption, the PES methodology treats on-time and late census responses separately.

There are three possible methods to handle late census responses in coverage estimation.

1. Lates as under-coverage: Late responses are excluded from the census list. This means that we consider the number of census responses to be the number of census responses minus the number of late census responses, and the under-coverage count to be the under-coverage count plus the number of late census responses.
2. Remove and return: Removing late responses from coverage model estimations and then adding them back to the total population estimates at the end: Late responses are considered out of scope in both the census and PES lists, but they are added back to the population estimates after the coverage adjustments have been applied.
3. Lates as coverage: Treating late responses as coverage: Late responses are kept in both the census and PES lists, assuming they are as valid as on-time responses and follow all the assumptions about PES independence.

In theory, if all the assumptions in DSE and sampling hold, all three methods should lead to unbiased population estimates. However, in practice, this may not always be the case.

In the 2018 PES, the original method was to use the first approach (lates as under-coverage). However, the results were not feasible when they were compared with the other sources, leading to a change in methodology to the second approach (remove and return), which produced more feasible results. As a result, the 'remove and return' approach became the default design method for the 2023 PES.

The 2018 situation demonstrated that although theoretically, all three methods should have produced similar results, there were differences in practice. This led to the following three research questions.

- Research question 1: In a controlled scenario where all variables meet the DSE assumptions, do the three methods yield similar results? How do various biases in PES affect each method?
- Research question 2: Systematically exploring the relationship between census late responses, PES activity, and under-coverage in the 2018 data to determine if there is evidence of dependencies that could cause bias in estimates.
- Research question 3: Relating the results from the previous questions to guide the selection of a suitable method for the 2023 PES. The goal is to establish a framework for deciding when a change from the standard design methodology is necessary and identify the factors to consider in making an informed decision.

In this section we will investigate the above research questions.

2.2.1 Research question 1: estimation methods in absence and presence of PES sampling biases

We created a set of simple simulations to demonstrate how three possible methods introduced in Section 2.2 behave in the different scenarios, where PES has a biased capture of lates. All simulations were run with the following population dynamics:

- total population of 5 million individuals
- 80 percent on-time census respondents
- 10 percent late census respondents
- 10 percent under-coverage

Then we changed each assumption and estimated the population using the classic Lincoln Petersen (Petersen, 1896; Lincoln, 1930) formula:

$$\widehat{N}_{total} = \frac{N_{Census} \times N_{PES}}{N_{Census\ responses\ in\ PES}}$$

Where:

N_{Census} is the number of individuals who responded to the census,

N_{PES} is the number of individuals who replied to PES, and

$N_{Census\ responses\ in\ PES}$ is the number of individuals who responded to both census and PES.

In the simulation studies, we considered these three different types of potential biases.

- Scenario 1: PES sample design under-samples or over-samples late responses. This scenario can happen if the PES responses are driven by census activity, and late responses are more or less likely to respond in PES.
- Scenario 2: PES prevents late responses or converts under-coverage into lates. In this scenario, PES activity affects the census response patterns, and individuals after responding to PES may change their decision about the census late response.
- Scenario 3: PES has biased capture of both lates and under-coverage. This is the scenario when PES is biased towards on-time census responses, and PES non-response has dependence with census non-response.

The simulation study results demonstrate the impact of potential biases on each method and are summarised in Table 2. For a detailed explanation of the simulations, please refer to **Appendix 3**.

Table 2 Impact of potential biases on each estimation method.

Bias		Method		
Type of bias in PES	**Direction of bias in PES**	**Lates treated as under-coverage**	**Remove and return (2018 method)**	**Lates treated as coverage**
No bias	No bias	No bias	No bias	No bias
Scenario 1: PES under- or oversamples lates but samples under-coverage correctly	Fewer lates than in population	Strong underestimation	Small underestimation	No bias
	More lates than in population	Strong overestimation	Small overestimation	No bias
Scenario 2: PES under- or oversamples under-coverage and lates, but captures on-time correctly	PES prevents late responses (Fewer lates in PES)	No bias	Overestimation	Overestimation
	PES stimulates late responses (More lates in PES)	No bias	Underestimation	Underestimation

Scenario 3: PES has biased capture of both lates and under-coverage	Fewer lates and under-coverage than in population	Strong underestimation	Underestimation	Underestimation
	More lates and under-coverage than in population	Strong overestimation	Overestimation	Overestimation

To summarise the results, we found that:

- when there is no bias in PES sampling, all three methods yield unbiased estimates
- when PES misrepresents the proportion of late responses and under-coverage in the population, treating lates as under-coverage is the most susceptible to bias
- remove and return method produced biased estimates but performed better than method 1
- treating lates as coverage (method 3) worked well only when PES accurately captured under-coverage rates, but the proportion of lates in PES was not accurate.

2.2.2 Research question 2: investigation of the relationship between late census responses and PES in 2018 Census data

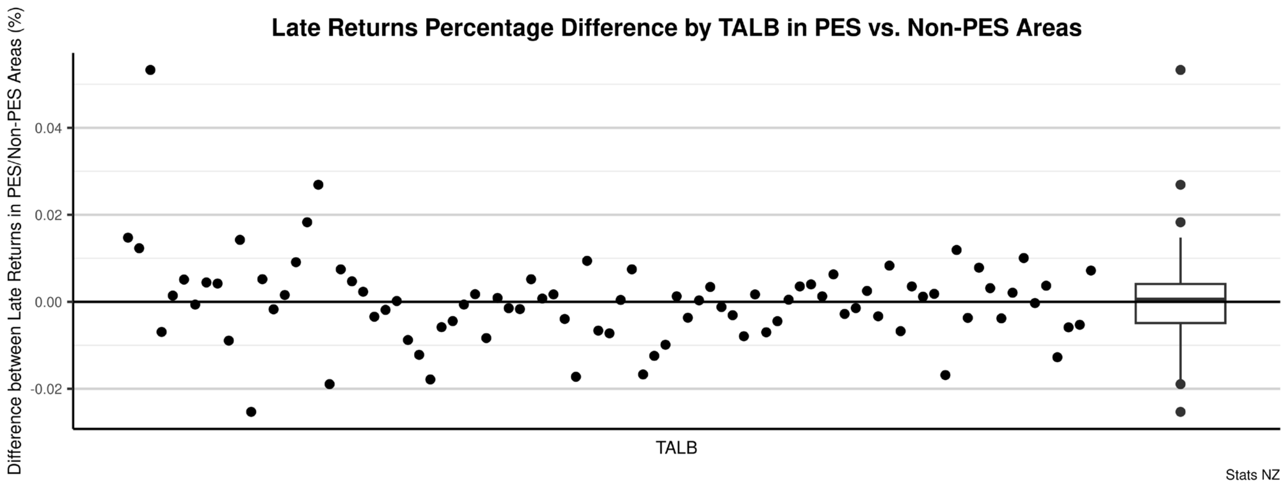
2.2.2.1 Checking the relationship between late census responses and PES activity

In this part of the study, we analysed the data from the 2018 Census and PES to investigate possible bias scenarios that could have affected the population estimation results. We were particularly interested in understanding why the 'remove and return' method produced more accurate population counts compared to treating late responses as under-coverage.

First, we examined whether the 2018 PES activity might have influenced or deterred late census responses. We hypothesised that areas with PES activity would show different proportions of late responses compared to areas without PES activity. To test this, we compared the percentage of late census returns in both PES areas and non-PES areas within the same Territorial Authority Local Board (TALB). The PES interviews a sample of PSUs in each TALB, resulting in PES areas (PSUs visited by PES) and non-PES areas (PSUs not visited by PES). We analysed the data for the entire New Zealand population, as well as individuals of Māori descent and Pacific ethnicity.

Figure 1 displays the difference in the percentage of census late returns between areas visited by PES and areas not visited by PES. The difference is calculated by subtracting the percentage of late returns in non-PES areas from that in PES areas. Each dot in this figure represents a TALB that was included in the 2018 PES sampling. A boxplot on the right summarises the overall distribution of dots. The average percentage of late responses in PES and non-PES areas for New Zealand population are both 1.98 percent.

Figure 1 Difference of census late response percentage between PES and non-PES areas



Similarly, Figure 2 and Figure 3 show the same comparison, but for the Māori descent and Pacific ethnicity subgroups, respectively. In these figures, each dot represents a TALB sampled in the 2018 PES. A boxplot on the right in the figures summarises the overall distribution of dots. In Figure 2, for individuals of Māori descent, the average percentage of late responses in PES areas (2.64 percent) is slightly lower compared to non-PES areas (2.76 percent). In Figure 3, for individual of Pacific ethnicity, the average percentage of late responses in PES areas (3.21 percent) is lower compared to non-PES areas (4.22 percent).

Figure 2 Difference of census late response percentage between PES and non-PES areas for individuals of Māori descent.

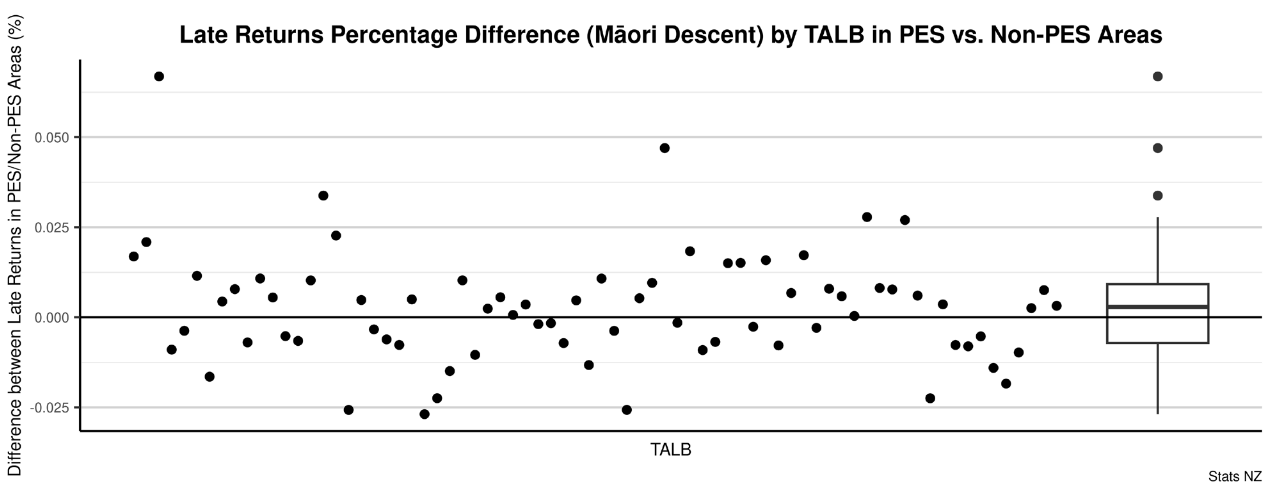
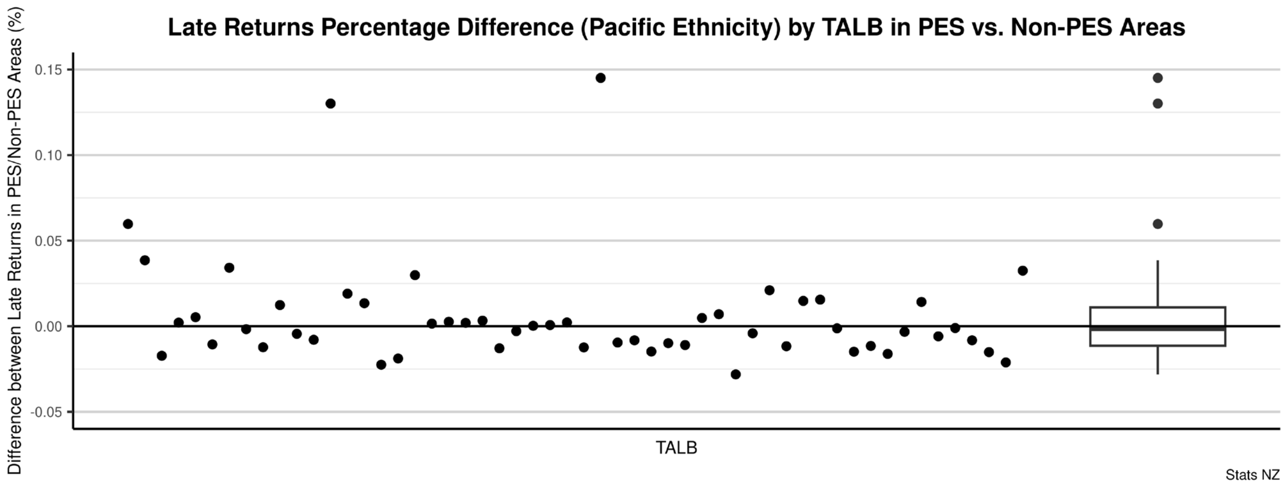


Figure 3 Differences of census late response percentage difference between PES and non-PES areas for individuals of Pacific ethnicity.



Figures 1, 2, and 3 do not show a significant difference in the prevalence of late returns between areas visited by PES and areas not visited by PES for the total New Zealand population, individuals of Māori descent, and individuals of Pacific ethnicity, respectively.

To further investigate the results at a more granular level, we analysed late response patterns at the PSU level using a regression model. The model considered the proportion of late responses in a PSU as the response variable and the proportion of Māori-descent and Pacific ethnicity individuals in a PSU as predictors. Additionally, we included an indicator to check the effect of PES activity on late responses. Details of the regression model can be found in Appendix 4.

The results of the regression study revealed that late census responses in the 2018 data were correlated with the proportions of Māori-descent and Pacific ethnicity individuals in a PSU, but not associated with PES areas. Therefore, there was no evidence of a causal dependence between late responses and PES activity for the 2018 Census and PES.

As a result of this investigation, we did not find evidence that PES activity stimulated or prevented late census responses, which would have been observed in the census file. Combining this result with the findings from the previous section (refer to Table 2) would help us understand why the 'remove and return' method produced more accurate population counts in PES 2018 compared to treating late responses as under-coverage. As shown in Table 2, treating lates as under-coverage leads to unbiased results when PES activity induces late census responses. The lack of any patterns in over/under sampling of lates in the 2018 PES explains why attempting to treat lates as under-coverage during the 2018 PES estimation did not yield plausible results. Therefore, the likely explanation was that late rates were distorted in the PES sample, as we showed in scenario 1 in simulations.

In the following sections, we investigate the sampling rates of lates in PES data and the relationship of lates with under-coverage.

2.2.2.2 Checking the correlation between late census responses and under-coverage

The previous analyses showed that late responses in the 2018 Census were associated with Māori-descent and Pacific populations. These populations are historically associated with lower census coverage rates. We decided to check if we can observe any correlations between late responses and Māori descent/Pacific ethnicity who are historically under-covered in censuses. The source of the under-coverage data is 2018 PES. PES survey data has under-coverage records which were collected; PES outputs contain estimated under-coverage counts which were predicted using the coverage model. Below are the analyses of PES data and PES population output analyses.

Correlation between lates and under-coverage in PES data

To examine the relationship between late response patterns and under-coverage, we conducted a logistic regression analysis using the PES data. The analysis involved fitting a model that used the 'late response' indicator (1 or 0) as the response variable, and it included various individual characteristics such as sex, age, Māori descent, Pacific ethnicity, and the percentage of under-coverage in each PES PSU as covariates. Details of the regression analysis can be found in **Appendix 5**.

Results of the logistic regression model (detailed in **Appendix 5**) align with the findings from our previous analysis of the census data, where we compared PES PSUs with non-PES PSUs (section 2.2.2.1). The regression analysis indicates a strong correlation between late census responses, under-coverage, and individuals of Māori descent in the 2018 PES data. Additionally, there is a moderate correlation between late responses and the Pacific ethnicity population, as well as the youth population (ages 15–29) and young individuals of Māori descent. These results suggest that there is a connection between late responses and under-coverage, particularly among specific demographic groups in the 2018 PES data.

Correlation between lates and under-coverage in PES outputs

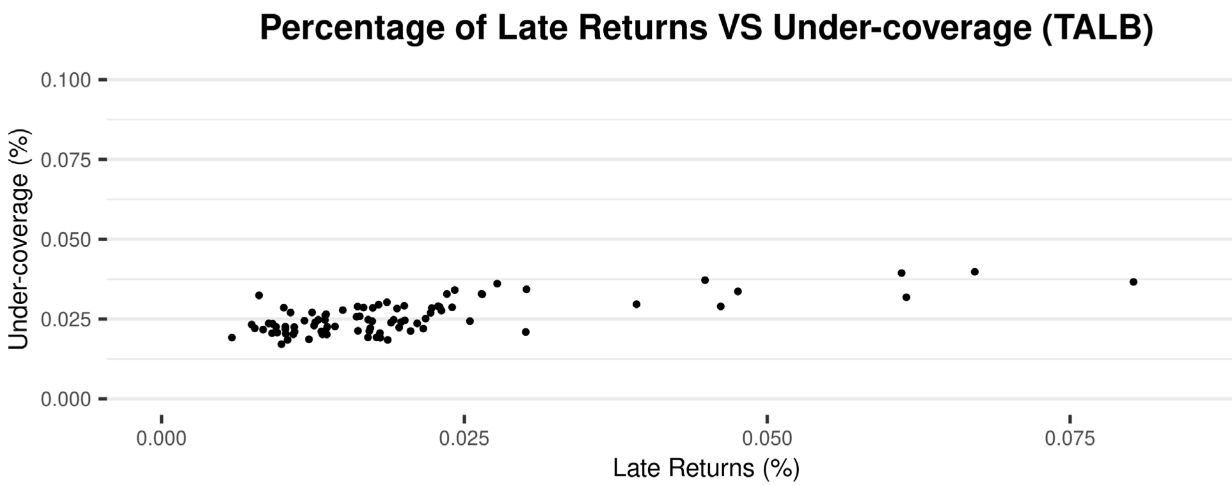
In the second part of our analysis, we focused on the PES outputs and examined the relationship between estimated model-based under-coverage rates and the proportion of late census responses at the TALB level.

Figure 4 illustrates the correlation between estimated under-coverage and the proportion of lates for usual residents at the national level and for each demographic subgroup considered in the study. Each dot on the graphs represents a TALB, with the y-axis representing the proportion of under-coverage estimated by PES, and the x-axis showing the proportion of late census returns in the same population.

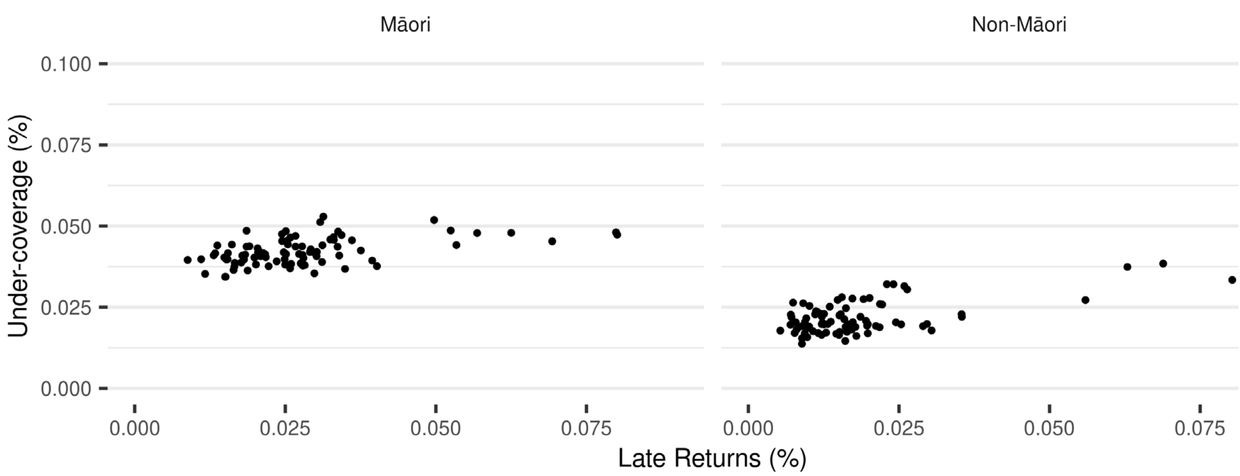
The results depicted in Figure 4 indicate a clear correlation between under-coverage and lates in all cases. These findings support our initial conclusion made during the production of the 2018 PES results, suggesting that census late responses are indeed correlated with under-coverage. This correlation further reinforces the importance of understanding and accounting for the impact of late responses in the estimation process.

Figure 4 Proportion of individuals of late responses and under-coverage in each TALB. (A) total population level results, (B) Māori-descent and non-Māori-descent populations, (C) Pacific ethnicity and non-Pacific ethnicity population, (D) 15-29 population and other ages population

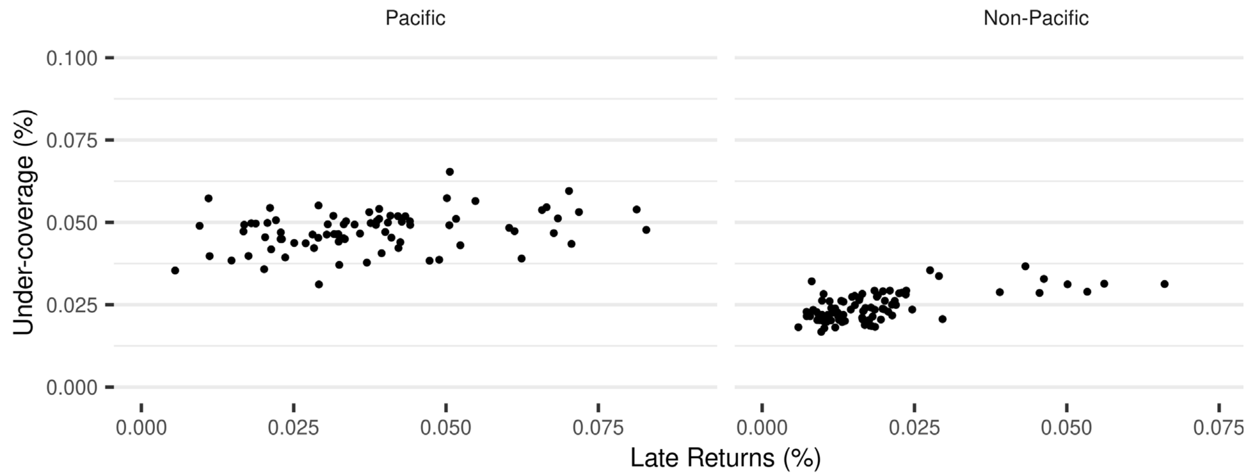
(A)



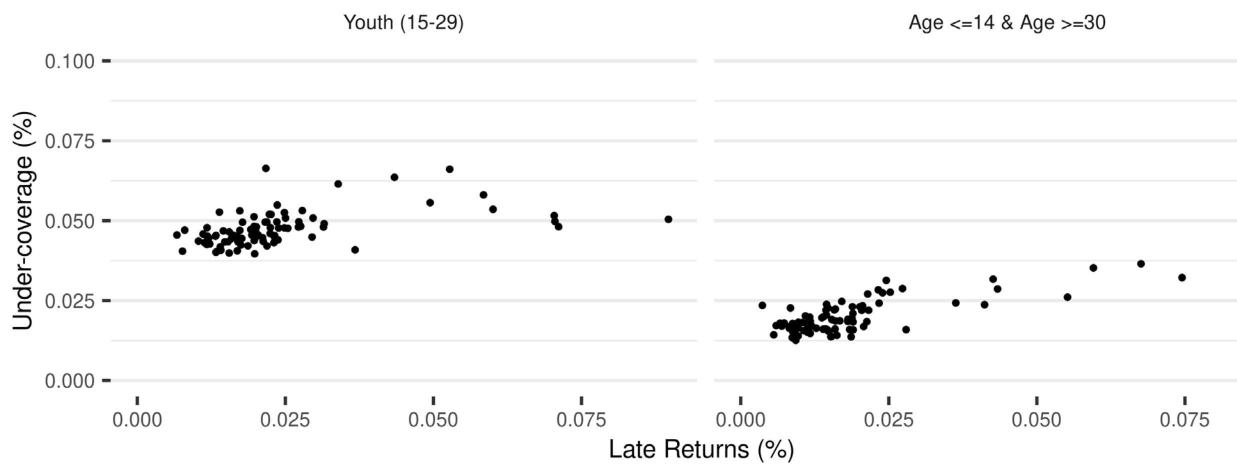
(B)



(C)



(D)



2.2.2.3 Checking the sampling rate of late responses by PES

So far, our analysis revealed that late responses in the 2018 Census data are unlikely to be influenced by PES activity, and they are highly correlated with under-coverage. Based on the simulations results, we found that treating lates as under-coverage can introduce bias when the proportion of lates in PES differs from the true proportion in the population (**Appendix 3**, Scenarios 1 and 3). This led us to investigate how PES captures late responses compared to the census data. To do this, we calculated the ratio of late to on-time responses in both the census and PES areas (Table 3[1]). A ratio of 1 indicates that the percentage of late returns is the same in both datasets. Ratios greater than 1 suggest that the

[1] All counts have had RR3 applied for confidentiality.

census has a higher proportion of late returns, while ratios less than 1 indicate that the census has a lower proportion of late returns compared to PES areas.

Table 3 Late and on-time proportions in 2018 Census and PES datasets.

Group	Late / On-time in census	Late / On-time in PES	Late proportion in census (column 2) / Late proportion in PES (column 3)
National	0.021	0.020	1.05
Male	0.021	0.020	1.05
Female	0.021	0.019	1.11
Young male	0.027	0.025	1.08
Māori-descent	0.033	0.034	0.97
Young Māori descent	0.034	0.031	1.10
Young male Māori descent	0.031	0.036	0.86
Pacific ethnicity	0.062	0.047	1.32
Young Pacific ethnicity	0.064	0.041	1.56
Young male Pacific ethnicity	0.061	0.036	1.69

Our findings (Table 3, column 4) show that for Māori descent subgroup and young male Māori descent subgroup, PES areas have a higher proportion of late census returns compared to the census areas. This is expected since PES is designed to capture a higher number of individuals in the low response rate group, particularly in populations with historically low response rates. However, we found some surprising results for certain subgroups, such as Māori descent subgroup, and Pacific ethnicity subgroup. In these cases, PES sampled a lower percentage of late returns compared to the census.

These findings provide valuable insights into how PES captures late responses and help us better understand the relationship between PES and census data.

These observations are in line with the simulated scenarios where PES was under-sampling lates or both lates and under-coverage. As we saw in both scenarios, PES having a lower rate of lates leads to the method of treating lates as under-coverage approach to under-estimate the population. Therefore, in the 2018 situation, treating lates as under-coverage would have been the most bias-prone approach (Simulation scenarios 1 and 3, **Appendix 3**).

2.2.3 Research question 3: a suitable method for the 2023 PES data

First, we need to state that in the absence of the biases in the PES sample, all three methods lead to the unbiased results.

Treating lates as under-coverage tends to create larger biases in the estimates (higher sensitivity to our assumptions) and works well only when there is a causal dependence between late responses and PES activity. The historical 'remove and return' approach was prone to biases in all scenarios but performed better in the 2018 estimation. These observations show that we will need more information about 2023 to make a final estimation decision. Below we are reiterating the table with estimation approaches and add the 2018 data observations supporting or rejecting the tested scenarios, and if these scenarios can explain observed biases.

Table 4 Summary of performance of tested estimation approaches and observations from 2018 data

Bias		**Method**			**Support from 2018 data**
Type of bias in PES	**Direction of bias in PES**	**Lates treated as under-coverage**	**Remove and return (2018 method)**	**Lates treated as coverage**	
No bias in PES	No bias	No bias	No bias	No bias	Two methods lead to the different results indicating the presence of some degree of bias in 2018
Scenario 1: PES under- or oversamples lates	Fewer lates than in population	Strong underestimation	Small underestimation	No bias	May have happened because PES had fewer lates for young male and Pacific ethnicity individuals
	More lates than in population	Strong overestimation	Small overestimation	No bias	
Scenario 2: PES prevents late responses or converts under-coverage into lates	PES prevents late responses (Fewer lates in PES)	No bias	Overestimation	Overestimation	No strong evidence to show that PES activity causes or prevents lates
	PES stimulates late responses (More lates in PES)	No bias	Underestimation	Underestimation	
Scenario 3: PES has biased capture of both lates and under-coverage	Fewer lates and under-coverage than in population	Strong underestimation	Underestimation	Underestimation	Not tested in this paper due to lack of the information about the true census under-coverage in the population, which is also independent from PES
	More lates and under-coverage than in population	Strong overestimation	Overestimation	Overestimation	

2.3 Treatment of census late responses given different start dates for different regions

The standard design for the 2023 PES assumed collections would commence at the same time for all areas. However, work by the Whānau Ora Commission Agency commissioned by Stats NZ to increase the census response rates among Māori and Pacific households in Auckland meant PES has had to postpone the collection start date in Auckland. The advantage of the work of Whānau Ora in census is an important part of the effort to increase census response rates for Māori and Pacific, and the responses they have obtained will help in providing better estimates for Māori and Pacific populations.

To account for Whānau Ora activity in Auckland, the PES 2023 collection ran from 2 June–31 August for any area outside Auckland and 1 July–30 September in Auckland[1]. This subsection details the best way to account for late responses when there are two collection start dates and to obtain an accurate estimate of a population given this discrepancy in dates.

2.3.1 Classification of late responses in census and PES

As explained earlier, to maintain the independence between census and PES, we need to determine whether a census response is late or on time. Determining the classification of a census response as ’late’ or ’on time’ is straightforward if the usual residence (UR) on census night matches the location from where the response was submitted. A census response is considered ’on time’ if it is submitted before 2 June, or between 2 and 30 June from a person whose UR on census night is in Auckland. However, complications arise when people move between regions during three different periods: census night, the start of PES interviews in zone 1 (covering the whole country except Auckland), and zone 2 (Auckland region).

Considering the time of each person's movement and their usual residence location, we can classify census responses into late and on-time categories. All the potential scenarios for classifying census responses are detailed in Table 5.

Table 5 Classification of census late response.

Date of census submission	Census UR	UR when census submitted	Did they move? When did they move?	Is it late?
Before 2 June	Anywhere	Anywhere	NA	No
After 2 June and before 1 July	Auckland	Auckland	NA	No
After 2 June and before 1 July	Auckland	Not Auckland	Before 2 June	Yes
			After 2 June	Yes
After 2 June and before 1 July	Not Auckland	Auckland	Before 2 June	No
			After 2 June	Yes

[1] The final PES interviews in Auckland were finished on 31 August, and the field interviews in the rest of the country were finished almost a week earlier.

After 2 June and before 1 July	Not Auckland	Not Auckland	NA	Yes
After 1 July	Anywhere	Anywhere	NA	Yes

Unfortunately, we do not have access to information about an individual's movement between regions from census responses for the period between census and PES. Therefore, the only way to classify a response received in June as late is by assuming that the UR at the time of the census is the same as the UR when the census response is submitted.

A person who completes both the census and is interviewed for PES provides more information about their movements, which allows us to improve our classification of census responses for those interviewed for PES. The improved classifications are shown in Table 6.

Table 6 Classification of census late response using PES information.

Date census submitted	UR census	UR PES	PES Address	Is it late?
Before 2 June	Anywhere	Anywhere	Anywhere	No
After 2 June and before1 July	Auckland	Auckland	Auckland	No
After 2 June and before 1 July	Auckland	Auckland	Not Auckland	Sometimes
After 2 June and before 1 July	Auckland	Not Auckland	Auckland	Sometimes
After 2 June and before 1 July	Auckland	Not Auckland	Not Auckland	Sometimes
After 2 June and before1 July	Not Auckland	Auckland	Auckland	Sometimes
After 2 June and before 1 July	Not Auckland	Not Auckland	Auckland	Sometimes
After 2 June and before 1 July	Not Auckland	Auckland	Not Auckland	Sometimes
After 2 June and before 1 July	Not Auckland	Not Auckland	Not Auckland	Yes
After July	Anywhere	Anywhere	Anywhere	Yes

However, even with this additional information, there is still no way of definitively determining when a census response is late. For instance, consider two people with a census usual residence in Auckland, who both submitted a census response in July and are interviewed for PES outside of Auckland in August. If the first person moved at the end of June, then their census response submitted in July should be considered on-time. If the second person moved at the end of April, then their census response in July should be marked as late. However, we will give the same classification (either on time or late) to both as PES and census information have no way of telling the two situations apart. We worked through the different permutations, based on how people could potentially move in and out of census and PES areas during these times. The logic of this is explored in **Appendix 6**.

Calculations in **Appendix 6** show that, if we adjust the census late responses to reflect their address in PES, approximately 20 percent of the time these adjusted classifications will be incorrect. For example, if we reclassify 100 responses as late, around 20 of those responses should have been classified as on time. It was also shown that in the situation that a census respondent who submits their census response in June, and they were outside of Auckland on census night and then moves to Auckland before the end date of PES interviewing period, approximately 33 percent of the time these adjusted classifications are incorrect.

2.3.2 Possible impact of misclassified late/on-time responses on estimates

To assess the impact of misclassification on our model estimates we must first determine the expected number of misclassified late responses or misclassified on-time census responses in the general population and in the target subpopulations. Misclassification may occur if a person relocated to/from Auckland between census night and a pes interview, resulting in the incorrect assignment of their response as late or on-time as discussed in the previous section.

For the general population, we assume:

- 100,000 people submitted a census response in June – less than 2 percent of the population
- fewer than 6,000 people move out of Auckland each month – this is about 0.12 percent of the population, or just under 0.5 percent for four months (March, April, May, June)
- fewer than 6,000 people move into Auckland each month.

Assuming that submitting a response in June is independent of relocating between census and PES, we would expect less than 0.1 percent of people to fall into the class of those who submit a census response that is misclassified as on time, and the same for a census that is misclassified as on time.

We now consider two possible methods for dealing with the misclassified late/on time responses caused by the different start date of PES in regions 1 and 2.

- Method 1 adjusts the late classification of census responses for those respondents selected for PES.
- Method 2 does not adjust the late classification of census responses.

Appendix 7 includes a comparison of relative biases at various response rates and sampling rates for each method. The differences in relative bias between estimates obtained through the different methods are not substantial, whether viewed as a percentage increase or an absolute value. This indicates that there is minimal distinction between the two methods. Since Method 1 does not exhibit significantly less relative bias than Method 2, and considering that Method 2 is simpler, we recommend not making adjustments to census late classifications based on information from PES.

Chapter 3: Uncertainty measures for estimates

3.1 Introduction

Statistical estimates, which rely on samples, inherently have random variation, which introduces uncertainties of the true values. Uncertainties can be measured and presented in various ways, such as confidence intervals or standard error bounds. These measures provide a means to quantify and express the level of uncertainty associated with statistical estimates.

In 2018 PES, census coverage estimates were produced using a Bayesian estimation model. The same methodology will be used for the 2023 PES as described in the 2023 PES Standard Design (Stats NZ, 2023 a). The uncertainty of PES estimates, similar to other Bayesian statistics, is typically reported as credible intervals. A credible interval is a range which the value of a parameter falls within given a particular probability (Ward, Harold, & Leonard, 1963). These intervals provide a range of plausible values for the estimated parameter, conveying the level of uncertainty associated with the estimate. We usually report 95 percent credible intervals, producing the 2.5 and 97.5 percentiles of the posterior distributions of the predicted populations. Narrower credible intervals indicate that the predictions have higher levels of precision.

While credible intervals effectively report the uncertainty tied to sample-based estimates, they do not provide a comprehensive picture of all potential sources of error in PES. To gain a more complete understanding of the overall accuracy and reliability of the estimates, it is essential to consider Total Survey Error (TSE). TSE is a systematic framework designed to holistically evaluate and understand these errors, taking into account various components (Biemer & Lyberg, 2020). In the context of PES, TSE encompasses a wide range of error sources that extend beyond sample error.

The purpose of this chapter is to review all major uncertainties in the PES estimation process. The paper will focus on describing and measuring the uncertainties that arise during data collection, processing, and modelling in the PES-based coverage estimation process. It will outline which uncertainties are already considered and reflected in PES outputs, and which ones still need to be incorporated. Additionally, in this chapter we will review the methodology used to account for the modelling uncertainty and show the possible uses in PES on a small-scale example, to demonstrate the potential applications in PES.

3.1.1 Summary of findings

In this chapter we identify three main approaches to deal with derivation and estimation uncertainty in PES:

1. Generate all possible population count outputs under all possible estimation design options. Then select the most accurate output according to independent population benchmarks. For example, conduct the estimation with and without late census responses and select more plausible output.

2. Embed the potential source of uncertainties as model parameters. Create a single model, which will include components and parameters to represent all sources of uncertainty. For example, the model can take into account linking error uncertainty and missing data imputation uncertainty, and also include a parameter to represent the PES response rate bias for late census responses.
3. Produce several structurally different models and apply model averaging techniques (for example, fitting models with different sets of covariates and averaging the results).

Findings in this chapter show that approach 1 does not reflect the uncertainty of our option selection, while approaches 2 and 3 introduce the data-related and model-selection uncertainties in our predictions by design. We will not be able to reflect the uncertainties around the DSE assumptions, which are not measurable by design. For example, we cannot assign a probability of late census responses violating the census-PES independence.

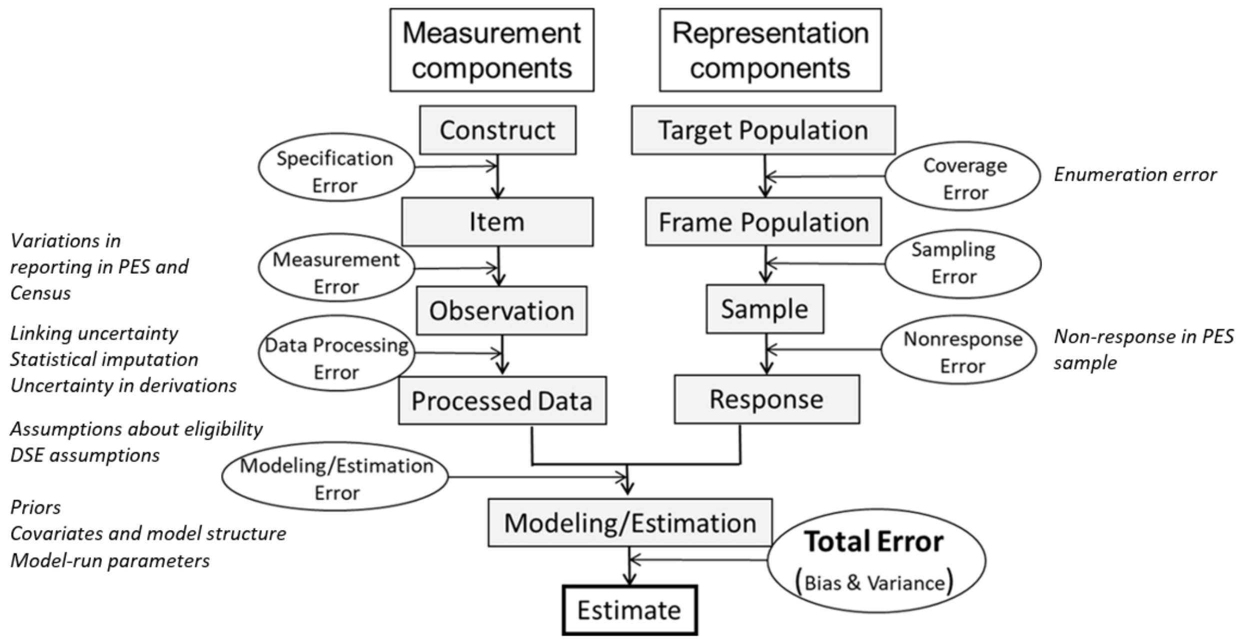
In the 2023 coverage estimation, we recommend the use of some form of model averaging to represent the model selection uncertainty in the scenario. If more than one model shows the good fit to PES dataset and produces feasible outputs, we represent the uncertainty in our model selection process.

3.2 Total survey error and uncertainty in PES

All statistics, which are based on data collected from surveys, are accompanied by measures of uncertainty. These uncertainties aim to represent survey errors, which may occur from sampling, collection, data processing, and estimation processes (Figure 5). The total survey error (TSE) framework is used to understand the possible sources of errors in a survey to better control the quality. The total error has two main components: sampling error and non-sampling error. Non-sampling error includes coverage, non-response, measurement, processing, and estimation errors. In Figure 5, the components of the TSE framework are distributed into the survey measurement of attributes and survey representation error sources.

PES is a statistical system including multiple stages, starting from sample selection to production of coverage estimates. Each of these steps has the potential to introduce uncertainty into the final estimates, as illustrated in Figure 5, using the TSE framework. In this section, we provide a brief overview of each step involved in the PES and discuss the potential sources of uncertainty that may contribute to the coverage estimates. Aside from the sampling error, this includes uncertainty inherited from census, PES uncertainties that might occur in the data collection and linking stages, and uncertainties that are caused by the statistical assumptions and estimation model choice.

Figure 5 Total survey error framework and uncertainty sources in the 2023 Post-enumeration Survey. The figure adopted with modifications from Biemer & Lyberg, 2020



3.2.1 Uncertainties in the data collection and linking stage

Uncertainty at PES sampling stage. As with any sampling survey, PES data collection will have sampling uncertainty. Sampling uncertainty arises due to the fact that we can only collect data from a subset of the population, rather than the entire population, leading to potential uncertainty in the estimates. The sampling design of PES is already imbedded into our estimation model structure; thus, this type of uncertainty is already represented in the uncertainty of our final estimates.

Uncertainty in the data collection stage. Uncertainty in the collection step of the PES arises due to potential measurement errors caused by imperfect data capture and non-response errors, where certain individuals or households may not provide the required information, leading to incomplete data and potential biases in the estimates. The missing attribute values may be mitigated through imputation or by filling in gaps using data from alternative sources, while complete unit non-response leads to an increase in this form of uncertainty or, if correlated with census non-response, to bias in the final estimates.

Variations in reporting in census and PES. The same person in census may give a different response for a given attribute than they did in PES. For example, a person may report different usual residence addresses in census and PES, because they changed their location. A person may also choose to report or not report all the ethnicities they identify with. We looked at the uncertainty caused by these variations in the standard design, and found they were small but important to monitor (Stats NZ, 2023 a).

Uncertainty in linking stage. The final PES estimation dataset consists of multiple linked data sources. The data linking process may miss out true links or create false links.

PES responses are linked to the combined census file which include census responses and admin enumerations for the people that did not have responses. That linking and combining process will have a level of under- and over-coverage associated, which the PES automated and clerical linking processes will attempt to reveal. Stats NZ is currently developing a framework to express uncertainty created by linking. This is developed with admin-data-based population estimates in mind but can be extended to talk about uncertainty in the final PES output caused by linking error.

Uncertainty in statistical imputation in census and PES. Statistical imputation is a method used to replace missing data with estimated values. The uncertainty of this method may be incorporated via the production of the different versions of the missing data in census and PES and checking the variability of the final estimates. However, the census is provided with only a single imputation and historically, item non-response is minimal in the PES. For fitting of the PES coverage models only item-non-response in PES is relevant. Census item-non-response in attributes is, however, relevant to the construction of the census usually resident population: having a correct attribute value for a census record allows the application of the correct predicted adjustment rate. Uncertainty due to attribute imputations in census may be incorporated into estimation at that stage.

3.2.2 Uncertainties in the derivation stage and model selection

Once the PES dataset is linked to the census dataset, we derive the coverage indicators (for example, a record is flagged as an under-coverage or an over-coverage) and fit a model to find the relationship of these indicators with the demographic variables. Both steps (that is, deriving the coverage indicators and fitting the model) are the source of the uncertainty which is not currently represented in PES outputs. In this sub-section we will explain these two sources of uncertainty.

Variations of deriving eligible people for PES. To achieve a reliable population estimate, we need to ensure that we are adjusting the census counts of people who are eligible to be counted at census night. This implies that these populations meet all statistical requirements of all DSE assumptions. However, in practice there are situations that may violate these assumptions. This results in increased uncertainties in the estimates.

During the derivation of coverage indicators for individuals in the 2018 PES process, we reached questions of how to treat the following situations to maintain independence between census and PES.

- How to treat late census responses (people whose census forms are returned after the PES has started field interviews). These can either be treated as under-coverage, or removed from the list of eligible people for estimation because they violate the assumption of the independency between the census and PES
- Whether residents-temporarily-overseas (RTO) records should be treated as over-coverage or removed from the list of eligible population for estimation
- How to adjust the population counts obtained from corrections facilities and defence non-private dwelling administrative records. These records violate the homogeneity assumption

as those in these NPDs are almost certain to have been captured in census and, when appropriate, PES. The question is whether or not to adjust these records

- Setting the linking acceptance criteria – a key step in balancing linking and over-coverage detection.

In the 2018 PES process, we examined all these options, and finally selected the inclusion criteria that led to the most plausible result. These criteria took into account alternative population estimates that were available as well as expert demographic opinion. However, other defendable choices were also explored and led to somewhat different estimates. The statistical literature does not provide much guidance on assessing this type of 'data-inclusion' uncertainty because it is not clear how to correctly average the combination of designs and datasets, and how to assign weights to each design-data combination. In contrast, the methodology for calculation of model weights, given the fixed dataset is developed and will be reviewed further in this paper.

Uncertainty in the model selection. After deriving the coverage indicators, we fit a logistic hierarchical Bayesian regression model with varying intercepts. In 2018, we fit a series of models – tested models – varying the parameters as follows:

- Including or excluding a particular covariate and interactions between covariates in the model
- Adding or removing the levels of the geographical predictor in the hierarchical model (for example, meshblock-level random effects).

Each of the checked model options was chosen if it produced a good fit (using posterior predictive checks for the ability to reproduce the initial PES dataset), and, as with the data variations, we tested to see if the model outputs aligned with some population benchmarks. However, model choice uncertainty was not reflected in the final outputs.

In the following section we will review methods, which can be used in PES to better reflect the uncertainties related to both data design stage and modelling stage.

3.3 Approaches to represent design and modelling uncertainties in PES outputs

As mentioned in the previous section, uncertainty reported by PES does not currently account for the uncertainties associated with defining record eligibility, statistical assumptions, and modelling. In this section, we will discuss various methods which can be used to better reflect these uncertainties in our predictions process and reporting outputs, along with their potential benefits and limitations.

3.3.1 Producing and reporting separate outputs for different assumptions

This approach was already used in 2018 Census and PES, when we were producing a particular dataset for a set of assumptions and analysing the feasibility of the results against our expectations. This analysis helped to understand the performance of tested designs and our confidence in the design assumptions. If the analyses suggest that the assumptions are not holding, it is possible to adapt the design to better support the assumption. Choosing a suitable solution to treat the census late responses and making a decision to not adjust the counts from defence and corrections admin records are two examples of this approach from PES 2018.

In the case where we do not have any benchmarks, in order to check which assumptions lead to plausible results we may need to report several series of output. This approach is often used in demographic projections. For example, the United Nations publishes population projections under different fertility, mortality, and migration scenarios [World Population Prospects - Population Division - United Nations].

Generally, we want to avoid this approach for PES results, because having several output series will make it harder for customers to use the population estimates.

3.3.2 Meta-analysis and hierarchical models

In the previous approach we described a situation when the uncertainty is represented with different outputs. Here, we introduce another approach which is based on combining different outputs into one. This approach, known as meta-analysis, is a method to combine the results of multiple statistical analyses, which came from different scientific studies. The procedure aims to derive a pooled estimate of the common unknown parameter. The approach is widely used when researchers combine the survey results from different countries to make conclusions about worldwide trends (for example, research on the global prevalence of autism spectrum disorder (Jinan Zeidan, 2022). Bayesian hierarchical models are well-suited for meta-analyses because they allow for random effects and report uncertainty of the outputs from separate studies.

In the context of census coverage estimation, it would be difficult to apply standard meta-analysis methodology because meta-analysis usually makes conditional independence assumptions which would be hard to justify for a set of PES estimates obtained by varying data inclusion criteria and model structure. While there is some work published on dependent meta-analysis, it would take further work to understand whether these methods were applicable in a situation like PES.

3.3.3 Model averaging

Another possible solution for reporting uncertainties in PES is the model averaging approach. The model averaging only deals with uncertainty in model structure and not the other components of data structure uncertainty. In this approach we fit a series of models on our data and then combine them to obtain the final estimates and then report the corresponding uncertainties.

In the coverage estimation, after receiving the PES input data, we will need to fit a series of models to infer the relationship of the census coverage with different demographic variables and then select the model with the “best” fit. There are several methods which are applied in the situation when you cannot select a single model. Below we are giving a brief description of these methods.

Model comparison

In this method, we can compare the models, which were fit to PES data, by using a combination of information criteria (such as Watanabe-Akaike or leave-one-out cross validation information criteria) and posterior predictive checks. Information criteria are metrics that help to select the model, which is a good fit and does not have excess covariates. Posterior predictive checks allow us to test if the chosen model produces plausible predictions (plausibility in this case can be assessed against external benchmarks or the observed dataset). This approach was used for 2018 PES estimation and stays our default option of choosing the best estimation model.

Bayesian Model Averaging (BMA) and pseudo-BMA

The idea of model averaging in statistical analysis has been extensively studied. Notable work that has led to BMA includes work by Roberts (1965), who suggested a distribution which combines the opinions of two experts (or models). This is similar to BMA when only two models are considered. Leamer (1978) expanded the work of Roberts and developed the basic framework for BMA, emphasising the fundamental idea that BMA can mitigate the uncertainty involved in selecting the model.

The idea of BMA is to consider every plausible model. The model is treated as a parameter and particular model specifications are values of this parameter. As with any parameter in Bayesian methodology, each model is assigned a prior probability. The data is then used to update our belief about these models, according to Bayes' theorem, and so to obtain posterior probabilities of these models. These posterior probabilities of the models are then used to produce weighted averaged estimates from the plausible models. This approach helps to account for the uncertainty of the model selection process and can often improve the accuracy of the predictions. The limitations of the approach are the computational complexity and the problem of choosing a good prior distribution for each model. A good basic description of BMA is available in (Hinne, Gronau, van den Bergh, & Wagenmakers, 2020).The statistical information criteria in this method are used to calculate model weights, which in turn are used to average the outputs of the several models.

Pseudo-BMA is similar to BMA. While BMA starts with a prior distribution and updates this to get a posterior distribution, pseudo-BMA does not require the specification of prior probabilities for the models. It instead weights predictions by what is known as a pseudo-bayes factor. This weighting is based on the idea that the weight of a model can be estimated from the leave-one-out cross-validation (LOO-CV) scores of the models (Höge, Guthke, & Nowak, 2020). One big advantage of pseudo-BMA is that it has an easy implementation in R and stan and it does not require the specification of prior probabilities of a model. However, pseudo-BMA can be sensitive to the choice of LOO-CV scoring rule.

Model stacking

Model stacking is used in machine learning to combine the predictions from the different models. In the first step, a single dataset is used to train a set of base models. On the second step, the predictions of the base models are used to train a high-level model. The high-level model learns how to combine the predictions from the base level models to produce more accurate final predictions. Model stacking can help to improve accuracy and reduce the variance of predictions but is computationally expensive and sensitive to the choice of the base models. In the case of PES, stacking may be used to average the posterior distributions produced from the different models (Yuling Yao A. V., 2018).

Bayesian hierarchical stacking

Hierarchical stacking allows us to assign different stacking weights to the different parts of the dataset. This way, each subset of data will be estimated with the model which works best for it (Yuling Yao G. P., 2022). This has the potential to be very useful in the 2023 population estimates, as behaviour in cyclone affected areas may differ from that of the rest of the population. However, unlike pseudo-BMA and stacking, there are no ready-to-use packages to implement this method.

Pros and cons of different model combination methods

In all the methods described above the first step is to decide on the collection of models to combine/ average over. This is called the model space. Once we have done this, we have three cases:

- the true model is in the model space.
- the true model exists outside the model space.
- the true model does not exist.

In the case of modelling census non-response, we are likely to be in one of the latter two cases.

BMA performs well in the case where the true model is contained in the model space – and will asymptotically select the true model. In this context, asymptotically means that with the increase number of repeats, the true model will be selected more and more times. However, in the case where the true model is not contained in the model space, BMA still asymptotically chooses a single model. This behaviour can lead to estimates that are not as good as estimates that could be obtained by a combination of models in the model space (Monteith, Carroll, Seppi, & Martinez, 2011). However, we do not always elicit asymptotic behaviour from BMA, and BMA is shown to perform favourably when compared to using a single model (Hinne, Gronau, van den Bergh, & Wagenmakers, 2020).

However, BMA is computationally complex and involves the evaluation of complex integrals that can only ever be approximated.

Pseudo-BMA eliminates the need for a prior probability to be assigned to a model. In the case, where the true model is not contained in the model space, it often does not make sense to assign a model a probability being the 'true' model. A drawback to pseudo-BMA is that it can tend towards more complex models. For modelling census non-response pseudo-BMA reduces model uncertainty and is easy to implement within our existing systems (Höge, Guthke, & Nowak, 2020).

Model stacking is introduced in (Yao, Vehtari, Simpson, & Gelman, 2018). It is shown to perform well in comparison to BMA and pseudo-BMA in many situations where the true model is not contained in the model space. It can be implemented within the existing system using R and Stan.

Bayesian hierarchical stacking is a generalisation of stacking where the model weights vary as a function of the data. Bayesian hierarchical stacking is more computationally complex than many other options and does not currently have an easy implementation in R and Stan. However, if there is significantly different response behaviour in different regions (for example in cyclone affected areas), it may be worth further investigation. Example code is available in (Yao Y. &., 2021) and with dedicated time and resources it may be possible to implement this for coverage estimation of the 2023 Census.

In the next chapter we will show a simulated case, how model averaging can be used in PES estimation.

3.3.4 Simulation of model averaging scenario in PES

In the previous section, we gave an overview of methods, which can be used to reflect the uncertainty in the PES data preparation and estimation decisions. In this section, we provide an example of how a model averaging approach can be implemented using Stan and R environment – tools we use for the PES estimation. In these simulation studies, we focused on implementing Stacking and Pseudo-BMA methods.

3.3.4.1 Scenario 1 – the true model is presented among the tested models

To demonstrate the use of the model averaging, we first created a simulated population using 2018 Census data. For each record i, in a household h, in a TALB t, we simulated the binary census under-coverage indicator using the following model:

$$U_{01_{ih}} \sim Bernoulli\left(p_{ucov_{ih}}\right), i = 1, \dots, N_h^{ind}; h = 1, \dots, N_t^{hhld}$$

$$logit\left(p_{ucov_{ih}}\right) = -4 - I_{island_{ih}} + 0.5\, I_{Maori\ descent_{ih}} - 0.25\, I_{sex01_{ih}} + \alpha_h^{hhld}$$

$$\alpha_h^{hhld} \sim N(\alpha_t^{ta}, 1), h = 1, \dots, N_t^{hhld}$$

$$\alpha_t^{ta} \sim N(0, 0.1), t = 1, \dots, N^{TALB}$$

Where:

$U_{01_{ih}}$: under-coverage indicator for an individual i in a household h,

$p_{ucov_{ih}}$: under-coverage probability for an individual i in a household h,
$I_{island}, I_{Maori\ descent}, I_{sex01}$: binary indicators for the North or South Islands, Māori descent and sex,

α_h^{hhld} : household-level coverage effect,

α_t^{ta} : TALB-level coverage effect.

Note that these covariate effects (slopes of the linear model) and random effect variances are similar to values we observed in 2018 data, and we are using the type of the model used in 2018 PES estimation.

After simulating the under-coverage indicator for each individual, we sampled 180 dwellings using simple random sampling method from each TALB to simulate PES. The simulated PES data then was used to fit the models and predict population estimates.

Through the simulation studies, we tested four models, including the true model, that is the model which was used to generate synthetic coverage data. This true model appears as the first model, model A, in our list. Table 7 shows the possible covariates for the models, which covariates were used in the models, and their predicted model weights.

As mentioned before, model averaging approach combines the considered models by introducing weights to each model. The weights that are assigned to each tested model by each considered model averaging method are also listed in Table 7.

Table 7 Covariates used in the model testing in Scenario 1 and weights of the tested models. "+" covariate was included, "-" covariate was not included.

Model	**Covariates**					**Model weights**	
	Island	**Māori-descent**	**Binary sex**	**Other ethnic group**	**Māori-descent * Sex**	**Stacking weight, %**	**Pseudo-BMA weight, %**
A (true model)	+	+	+	-	-	90.97	99.05
B	+	-	+	+	-	7.39	0.67
C	+	-	-	+	-	0.00	0.27
D	-	+	+	-	+	1.63	0.00

As expected, both stacking and pseudo-BMA approaches assign relatively large weights to the true model – model A.

We also produced averaged predictions using the stacking and pseudo-BMA weights. The predicted number of individuals missed in the simulated census data for the Māori-descent and non-Māori-descent populations are reported in Figure 6.

Figure 6 Predicted counts of individuals missed by the simulated census when the tested models included the true model. X-axis shows the count of individuals, and Y-axis shows the models used for the prediction. Model A is a true model, which was used to generate the simulated data. The vertical line shows the true under-count – number of individuals, who did not respond in the simulated census. Distributions show the density of the posterior distribution of the predictions produced by fitted models, bold and thin lines show 66% and 95% credible intervals, respectively.

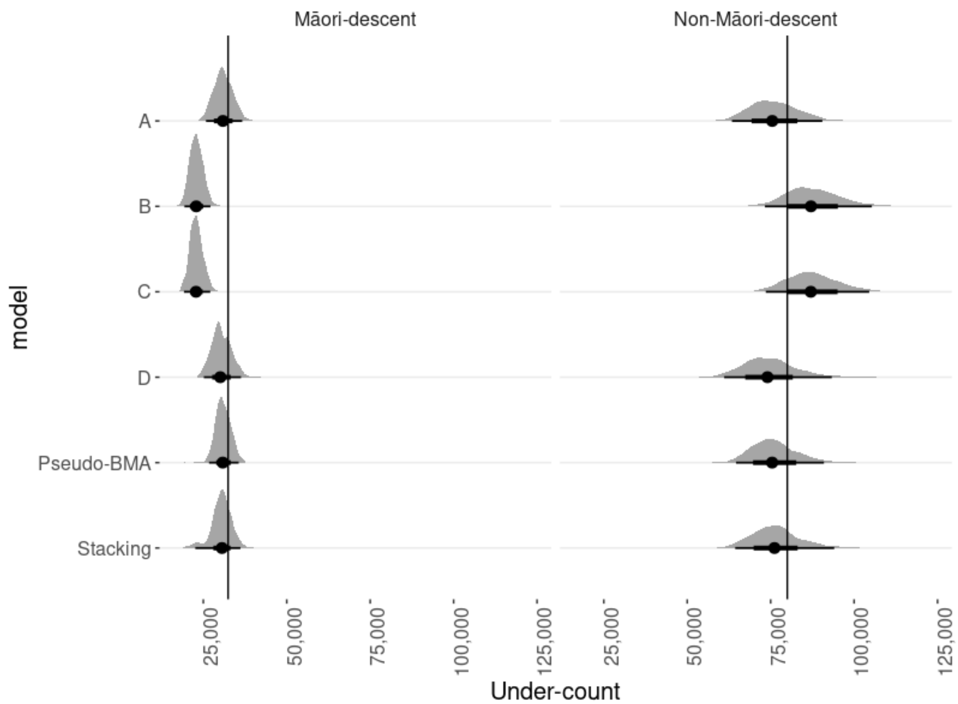


Figure 6 shows that the worst predictions for the Māori-descent variable under-counted population were made by the models B and C – the ones which do not include Māori-descent covariate. Both models tend to underestimate the number of people missed by the census for the Māori-descent population. Models A (true model) and D both had a covariate to account for the under-coverage effect of the Māori-descent variable, thus produced more accurate predictions for the subpopulations. Since the true model, model A, got the highest weights in both stacking and pseudo-BMA approaches, averaged models produced predictions which were very close to those of the true model.

In this case, when the true model was contained in the set of models over which we are to average, the true model received the highest weight. However, it should be noted that, while BMA supports interpreting model weight as the probability of that model being the true model, pseudo-BMA does not support such an interpretation (Höge, Guthke, & Nowak, 2020).

3.3.4.2 Scenario 2 – the true model is not among the tested models

In this scenario, again we first simulated a population using 2018 Census data. This time, we assigned a binary census under-coverage indicator to each individual using the following model:

$$logit\left(p_{ucov_{ih}}\right) = -4 - I_{island_{ih}} + 0.5\, I_{Maori\ descent_{ih}} + 0.1\, I_{male_{ih}} + \ 0.2\, I_{youth_{ih}} \\ + I_{Maori\ descent} \times I_{male_{ih}} \times I_{youth_{ih}} + \alpha_h^{hhld}$$

As shown in the equation, we introduced the effects of the single covariates: island, Māori-descent, male and youth (ages 15–29). We also introduced a three-way interaction between the covariates

Māori-descent, male, and youth which significantly increases the probability of being missed by the census. After assigning the census under-coverage indicator, we run four incorrect models to provide population estimates. Covariates considered in each model and the weights assigned to each method of model averaging approach are shown in Table 8.

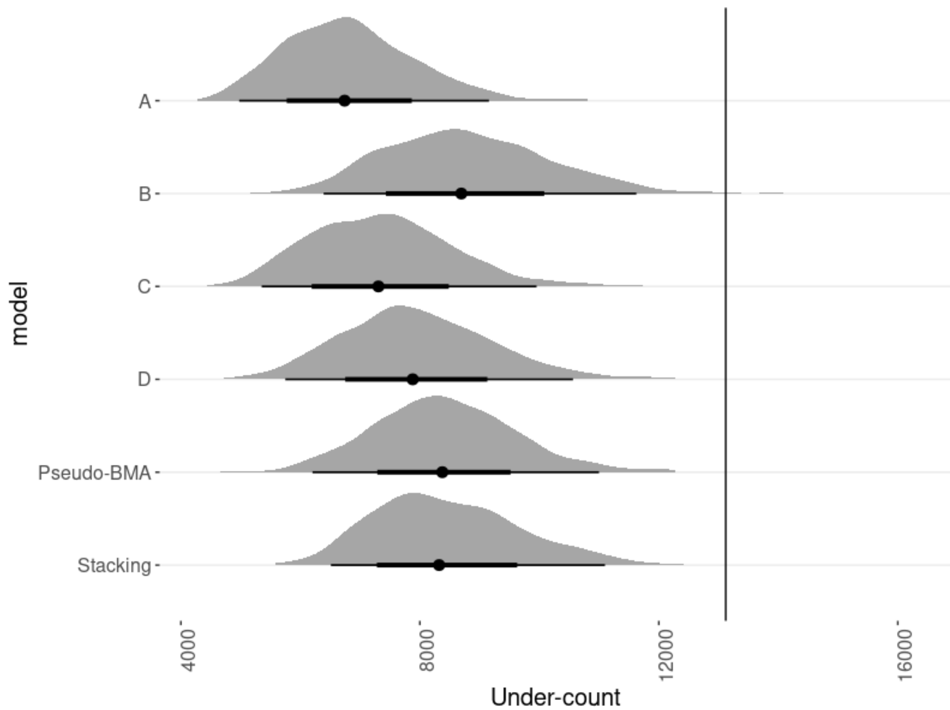
Table 8 Covariates used in the model testing in Scenario 2 and weights of the tested models. "+" - covariate was included, "-" - covariate was not included.

Model	Covariates							Model weights	
	Island	Māori-descent	Male	Youth	Māori-descent * youth	Māori-descent * male	Youth * male	Stacking weight, %	Pseudo-BMA weight, %
A	+	+	+	+	-	-	-	0.00	0.72
B	+	+	+	+	+	-	-	68.81	81.01
C	+	+	+	+	-	+	-	11.42	7.70
D	+	+	+	+	-	-	+	19.78	10.57

As reported in Table 8, none of the models was the model used to generate the initial dataset: none of them contains a three-way covariate interaction, but models B-D include two-way interactions.

In Figure 7 we are showing the under-count predictions from each of the models and averaged predictions. We are showing the results for the young male Māori-descent individuals, because it is the group with the largest bias by design in our simulations. In this figure, X-axis shows the count of individuals, and Y-axis shows the models used for the prediction. The vertical line shows the true under-count – number of individuals, who did not respond in the simulated census. Distributions show the density of the posterior distribution of the predictions produced by fitted models, bold and thin lines show 66 percent and 95 percent credible intervals, respectively.

Figure 7 Predicted counts of individuals missed by the simulated census of the young male Māori-descent population.



As shown in Figure 7, all predictions tend to underestimate the amount of under-coverage in this subpopulation, reflecting the model misspecification introduced by design in this simulation experiment. Model B had the closest predictions (median of the posterior distribution is closest to the vertical line), while model A (with no covariate interactions) showed the largest under-estimation.

Considering Table 8 and Figure 7 together, we can find that both stacking and pseudo-BMA approaches assigned the highest weight to the best-performing model, and the lowest weight was assigned to the worst one. As a result, averaged predictions performed better than the models A, C, and D, but were slightly worse than the best-performing model, model B. This indicates that in the absence of the true model, stacking and pseudo-BMA approach tend to give the highest weight to the model with the best predictions.

Overall, the observations from two simplified simulated scenarios show that model comparison using weights and model averaging can be useful tools for 2023 PES estimation, if we are facing large uncertainty in model selection.

Chapter 4: Census minimum data capture

4.1 Introduction

Minimum data capture (MDC) is a method to facilitate census respondents completing a reduced number of fields on paper forms to count as a census response. The intention of this is that it gets a response where otherwise one might not be possible. The 2023 Census program plan for minimum data capture applied to homeless/rough sleepers, non-private dwelling elevated care, and refusals – cases where a respondent may be willing to answer a limited number of questions. These fields were:

- Q1 – Name
- Q2 - Date of Birth
- Q5 - Usual residence address
- Q12 - Māori descent
- Q13 - Iwi affiliation.

After consultation with the 2023 Census Coverage project team, Q8 – Ethnicity and Q9 - Country of Birth were noted as the next questions to be asked once MDC was completed. As it is, a MDC response should meet the PES linkage acceptance criteria (defined in 2023 Post-enumeration Survey: Linking design (Stats NZ, 2023 b)), if the responses to name, date of birth, and usual residence address are full and complete.

These variables should allow for linking to admin data to enable missing information for questions that were not answered to be filled using alternative data sources (taken from administrative and historical census data). However, linking between the census responses and administrative data spine may be made more challenging due to linking variables such as country of birth and sex at birth not being available for these MDC records. It is also possible that the admin data may not exist for these variables.

The concern for PES methodology is that this change will increase the amount of linkage error within the census dataset, along with the amount of linkage error between PES and the census dataset. The anticipated increased error between PES and the combined census file is due to both country of birth and gender (while census uses sex at birth as a linking variable when linking to the IDI spine, the 2023 PES linking design uses gender) not being asked for these records. This may violate the perfect linking DSE assumption – that we can be confident that every PES record can be linked to every census record that refers to the same person, and none that do not. The thresholds to meet the perfect linking assumption are outlined in the 2023 Post-enumeration Survey: Linking design (Stats NZ, 2023 b)

Given the importance of the perfect linking assumption, understanding the impact that the increased number of MDC records caused by the census design change will have on our coverage estimation is essential. This chapter carries out a sensitivity study into what impact MDC will have on our final population estimates.

4.1.1 Summary of findings

The outcome of the sensitivity study carried out for this chapter shows that the increase in linkage error caused by MDC, when including refusals, 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 (Stats NZ, 2023 a).

4.2 Linkage error and MDC

Linkage error is made up of two broad groups; false positive error when two records are incorrectly linked, and false negative error when two records that refer to the same person are not linked. The use of linkage acceptance criteria should reduce the rate of both kinds of error as it removes all records from PES, census, and the IDI spine that we cannot be confident will link correctly. However, both country of birth and gender are both linking variables that can help us to be surer about a link.

False positive error between PES and the census file will result in an underestimation of the population due to a PES record being incorrectly assumed to have been counted in the census. False negative error between PES and census will lead to an overestimation of the population due to a PES respondent not being recognised as having been counted in the census.

While it is out of scope for this work, false negative error within the census file (between census responses and the IDI spine) will not lead to an overestimation of the population if PES can link to each duplicate record. False positive error within the census file is less problematic for PES as the sample design accounts for most forms of bias in census under-coverage.

As with all responses, there is the chance that MDC responses will have false information included in them. The census processing system will remove records with obviously false names, however, there is the chance that more subtle incorrect information may filter through to the combined census dataset that is used in PES linkage. Additionally, false positive linkage error within the census dataset may mean that incorrect information from alternative administrative data sources may filter through to the census dataset that we receive for linking. We are unable to estimate the impact that this erroneous information will have on PES linking, however we anticipate it is negligible.

We do not have a clear indication of the effect of MDC on linkage error, as MDC has not been used in previous years. However, *census partial responses* (specifically used to refer to census records that only appear on a household set-up form or dwelling form) are expected to have a similar linkage error rate to MDC responses. A *census partial response* gives the following information:

- name
- census night address
- age
- gender (not used in the linking process between census responses and the Integrated Data Infrastructure (IDI) spine (ever-resident population list derived solely from administrative sources), however it is used between PES responses and the combined census file).

We expect that MDC records will link slightly better than census partial responses due to these records containing a full date of birth and having usual residence address available. However, we think that census partial responses provide a good metric for how well we can expect MDC records to link. In 2018 the linkage error rates for *census partial responses* were estimated as follows:

- 20 percent more likely to have false positive error
- 20 percent more likely to have false negative error.

As mentioned above, we expect a similar linkage error rate to apply to MDC responses, and thus use the numbers above when calculating the predicted impacts of MDC.

These rates assume that a clerical review is not carried out by PES to resolve this linkage error caused by the automated linking process (outlined in 2023 Post-enumeration Survey: Linking design). The 2023 PES clerical linkage design is deliberately robust to resolve linkage error and ensure that it is accurately measured to ensure that the perfect linking assumption is achieved. However, it is key that we understand the impact of MDC on the results of our automated linking so that we can resource clerical linking appropriately.

The estimated false negative linkage error rate when including MDC responses is:

$$\gamma = \alpha + \frac{\alpha\left(0.2N_{MDC} + 0.2N_{partial}\right)}{N_{census}}$$

where α represents the rate of false negative linking of complete census responses, N_{MDC} is number of MDC responses, $N_{partial}$ is number of partial responses, and N_{census} is number of census responses, including partial and MDC responses.

We estimated the false positive linkage error rate when including MDC responses is:

$$\acute{\gamma} = \beta + \frac{\beta\left(0.2N_{MDC} + 0.2N_{partial}\right)}{N_{census}}$$

Where β represents the rate of false positive linking of complete census responses, and N_{MDC}, $N_{partial}$, and N_{census} are as above.

Table 9 and Table 10 give the results of a sensitivity study on the impact that different rates of MDC responses will have on our linkage error rates. We anticipate that the number of these records in the 2023 Census dataset will be much lower than 100,000. However, for completeness we have included a range of values to ensure robustness. For completeness, we have also included the impact that an increase in *census partial responses* from 200,000 to 400,000 will have on our linkage error.

An error rate of 1 (shown in Table 9 and Table 10) assumes that the perfect linking assumption would have been met if it were not for the inclusion of census partial responses and MDC records. The next section explores the impact that this increase in linkage error has on our population estimates if it is not resolved.

Table 9. False negative error rates for different (minimum data capture) MDC rates and two partial response rates.

MDC rate	False negative error rate given a *census partial response* rate of 4%	False negative error rate given a *census partial response* rate of 8%
0%	1.008α	1.016α
0.02% (~1,000)	1.00804α	1.01604α
0.2% (~10,000)	1.0084α	1.0164α
2% (~100,000)	1.012α	1.02α
3% (~150,000)	1.014α	1.022α

Table 10. False positive error rates for different (minimum data capture) MDC rates and two partial response rates

MDC rate	False positive error rate given a *census partial response* rate of 4%	False positive error rate given a *census partial response* rate of 8%
0%	1.008α	1.016α
0.02% (~1,000)	1.00804α	1.01604α
0.2% (~10,000)	1.0084α	1.0164α
2% (~100,000)	1.012α	1.02α
3% (~150,000)	1.014α	1.022α

4.3 Impacts of MDC on population estimates

This section outlines the impact that the increase in linkage error resulting from different numbers of MDC would have on our population estimates if it was not resolved. We have split this into the impacts of false negative error and false positive error. In reality, these forms of linkage error do interact to have an impact on population estimates if they are not resolved. However, the work in this section indicates that bias resulting from MDC responses is negligible.

False negative linking

When census and PES are correctly linked, the Lincoln-Petersen estimator (Petersen, 1896; Lincoln, 1930) gives a population estimate, $\widehat{N}_{correctLinking}$, of

$$\widehat{N}_{correctLinking} = N_{census} + \widehat{N}_{ucov}$$

Where N_{census} is the number of people captured in the 2023 Census dataset and $\widehat{N}_{ucov}$ is the estimated number of people in the target population that were not counted in 2023 Census.

Assuming a false negative linkage rate of γ, Lincoln-Petersen estimator gives an estimation, $\widehat{N}_{falseNegative}$, of the population size as

$$\widehat{N}_{falseNegative} = \frac{N_{census} + \widehat{N}_{ucov}}{1 - \gamma}$$

Calculating the relative bias of the population estimate gives

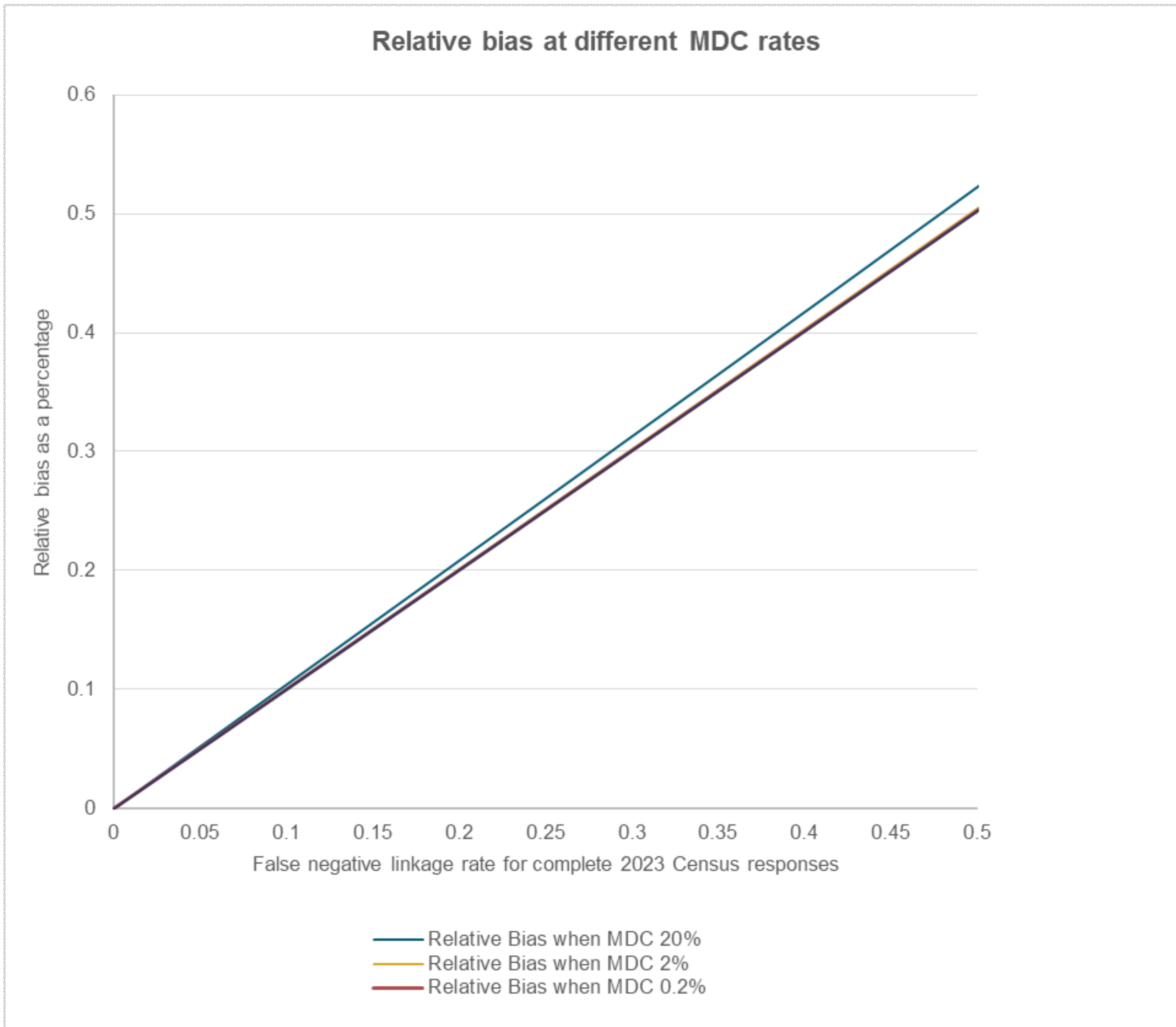
$$Relative\ Bias_{FN} = \frac{\gamma}{1-\gamma}$$

Assuming there are no partial responses in census, estimated false negative linkage error rate caused by MDC can be rewritten as $\gamma = \alpha + N_{MDC} \times 0.2\alpha$. Notice that the standard design for PES requires $\alpha < 0.5\%$ of the population estimate[1] (Stats NZ, 2023 b).

Figure 8 displays the impact of different level of MDC for different rate of false negative linkage rate on the relative bias of population estimates. The lines in Figure 5 are so close together as to be almost indistinguishable. While a large value of α may increase uncertainty to unacceptable levels, the increase caused by MDC is still minimal.

[1] Less than 0.5 percent of PES records with one link are assessed as a false positive link, less than 0.5 percent of PES records with more than one link are assessed as having at least one false positive link, less than 0.5 percent of PES records with no link are assessed as a false negative link, less than 0.5 percent of PES records with one link are assessed as having a false negative link to an additional record.

Figure 8 The impact of different level of MDC and different rate of false negative linkage rate on the relative bias of population estimates



False positive linking

Assuming a false positive linkage rate of $\acute{\gamma}$, Lincoln-Petersen estimator gives an estimation, $\widehat{N}_{falsePsitivee}$, of the population size as

$$\widehat{N}_{falsepositive} = \frac{N_{census} + \widehat{N}_{ucov}}{1 - \acute{\gamma}}$$

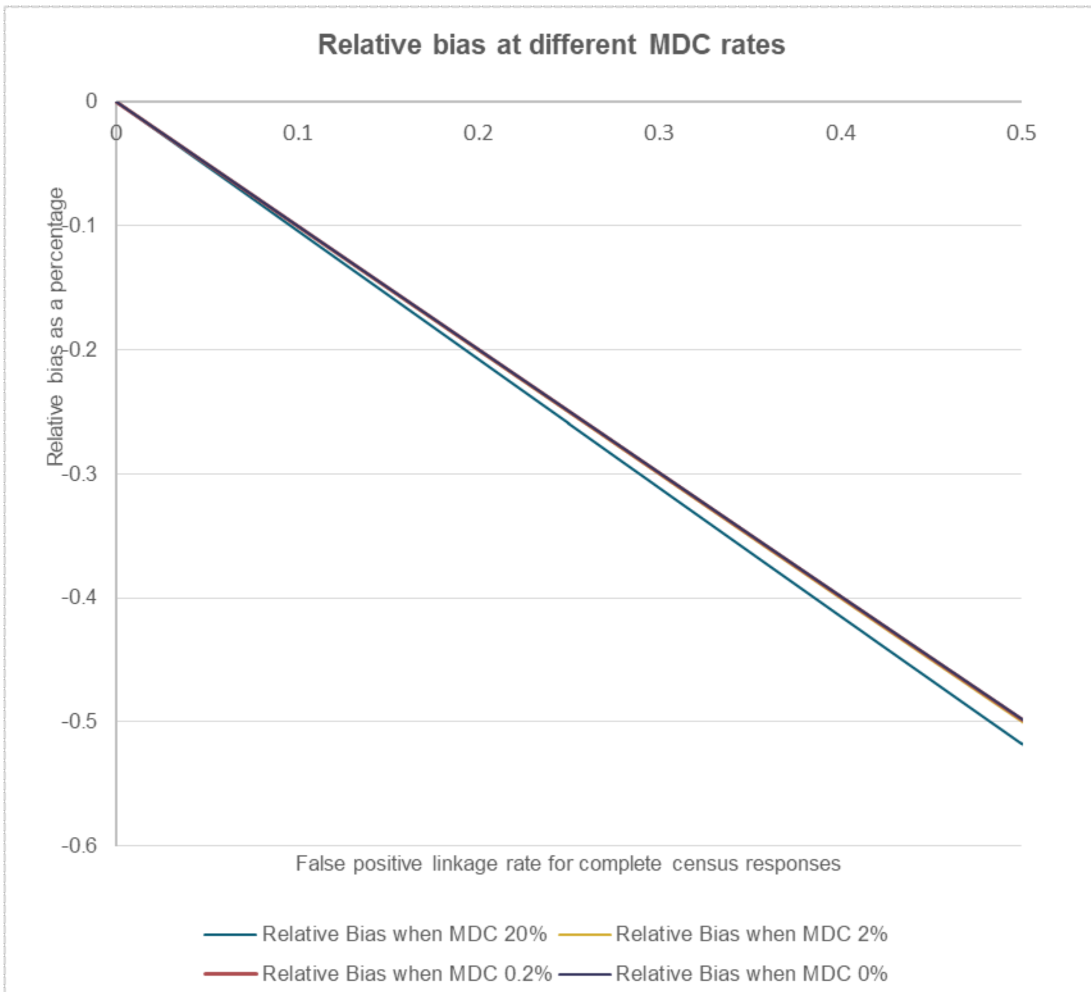
Calculating the relative bias of the population estimate gives

$$Relative\ Bias_{FP} = \frac{-\acute{\gamma}}{1 - \acute{\gamma}}$$

Assuming there are no partial responses in census, estimated false positive linkage error rate caused by MDC can be rewritten as $\acute{\gamma} = \beta + N_{MDC} \times 0.2\beta$. Notice that the standard design for PES requires $\beta < 0.5\%$ of the population estimate.

Figure 9 displays the impact of different level of MDC for different rate of false positive linkage rate on the relative bias of population estimates.

Figure 9 Impact of different levels of minimum data capture (MDC) and different rates of false positive linkage rate on the relative bias of population estimates



Again, we can see that all lines in Figure 9 are so close together as to be almost indistinguishable. While a large value of β may increase uncertainty to unacceptable levels, the increase caused by MDC is still minimal.

It can be seen that MDC is unlikely to have an impact on population estimates at the anticipated rates it will be used. However, it is still recommended that error rates for these records is noted in the final linkage report.

Glossary

admin enumeration – use of administrative data to add people to the usually resident census population when a census response has not been received.

causal independence – the likelihood of being recorded on one list has no relationship with the likelihood of being recorded on the other. For example, there is no structural or operational reason PES would be more or less likely to capture those records already captured by census.

census coverage (rate) – compares the number of people who were counted in the census with the number who should have been counted (as estimated by the PES).

census response (rate) – compares the number of people who responded to the census with the number who should have responded (as estimated by the PES). Historically a 'response' has been defined as the completion and return of an individual form. However, 2018 PES also produced a census response rate calculation using a minimum information definition of response as well.

clerical linking – the process of manually reviewing records not able to be linked by automatic processes and searching for missed links. In clerical linking, an analyst decides which record pairs are links and which are non-links. This process can allow for more nuance to be considered in the linking outcome than can be accounted for in automatic processes.

combined census– a census in which some information on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes, but where other information that is not available from such sources is collected directly from individual persons and households by means of full or partial field enumeration or from other sample surveys.

confidentiality – the protection of data from, and about, individuals and organisations; and how we ensure that data is not made available or disclosed without authorisation.

deterministic imputation – missing response information is obtained directly from linked information relating to the same individual.

false negative link – two records that should have been linked because they correspond to the same unit (that is, they are a true match) but were not linked.

false positive link – two records that were linked in error and do not correspond to the same unit (i.e., they are a not a true match).

homogeneity of capture – all records have the same likelihood of being captured in a given list. This is not achievable at a population level, so we use characteristics (such as age, sex) to group records until they have equal likelihoods of capture.

information criteria – likelihood-based measures of model fit that include a penalty for the number of the parameters in the model.

over-sampling – the goal or action of sampling a particular group of interest at a higher rate than they contribute to the wider population. At Stats NZ, we use techniques such as stratification, allocation, and sampling selection methods to implicitly over-sample groups of interest. Less commonly, we may also use additional techniques such as booster panels to explicitly over-sample groups of interest.

perfect linking – error-free linking between the two lists including no missed links, and no records incorrectly linked.

PES sampling frame – a list of all private dwellings within PES geographic areas that are available for sample selection. This is the key mechanism through which PES accesses people in the PES target population.

posterior probability distribution – the distribution, which describes the quantity or the parameter of interest after observing and analysing data. In our simulations, we looked at the posterior distributions of census coverage rates and population count, after analysing PES dataset. Posterior distributions reflect the uncertainty in our knowledge about the parameters of interest.

prior probability distribution – the distribution, which describes the quantity or the parameter of interest before observing data.

relative bias – The relative bias provides a measure of the magnitude of the bias. Mathematically the gap between observed and predicted data is divided by the observed data to get the relative bias.

statistical imputation – missing response information is filled using response information from another unit, the donor. Typically, the donor is chosen in such a way that it resembles the imputed unit as much as possible on one or more background characteristics.

References

Biemer, P., & Lyberg, L. (2020, January 15). *Total survey error.* doi:https://doi.org/10.4135/9781526421036888444

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science (3(2))*, 200-215.

Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999). Bayseian Model Averaging: A Tutorial. *Statistical Science (14, 4)*, 382-417.

Höge, M., Guthke, A., & Nowak, W. (2020). Bayesian Model Weighting: The Many Faces of Model Averaging. *Water 2020, 12, 309.*, 12(2), 309.

Jinan Zeidan, E. F. (2022, May). Global prevalence of autism: A systematic review update. *Autism research, 15*(5), 778-790. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/aur.2696

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data* (Vol. 53). New York: Wiley.

Lincoln, F. C. (1930). Calculating waterfowl abundance on the basis of banding returns. *US Department of Agriculture (No. 118)*, 1-4.

Monteith, K., Carroll, J. L., Seppi, K., & Martinez, T. (2011). Turning Bayesian Model Averaging Into Bayesian Model. *The 2011 International Joint Conference on Neural Networks*, (pp. 2657-2663). San Jose.

Petersen, C. G. (1896). The yearly immigration of young plaice in the Limfjord from the German sea. . *Report of the Danish Biological Station, 6*, 1-48.

Roberts, H. V. (1965, March). Probabilistic Prediction. *Journal of the American Statistical Association, 60*, 50-62. Retrieved from https://doi.org/10.2307/2283136

Stats NZ. (2020). *Post-enumeration survey 2018: Methods and results.* Retrieved from www.stats.govt.nz: httpshttps://www.stats.govt.nz/methods/post-enumeration-survey-2018-methods-and-results/

Stats NZ. (2021, February 18). *Why do we do the Post-enumeration Survey (PES), and why is it hard?*

Stats NZ. (2023 a, March 22). *2023 Post-enumeration Survey: Standard design for coverage estimation.*

Stats NZ. (2023 b, March 22). *2023 Post-enumeration Survey: Linking design.* Retrieved from www.stats.govt.nz: https://www.stats.govt.nz/methods/2023-post-enumeration-survey-linking-design/

Ward, E., Harold, L., & Leonard, J. S. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70(3)*, 193–242.

Yao, Y. &. (2021). Bayesian Hierarchical Stacking: Some Models Are (Somewhere) Useful. *Bayesian Analysis*.

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using Stacking to Average Bayesian Predictive Distributions (wuth Discussion). *Bayesian Analysis*, 917-1003.

Yuling Yao, A. V. (2018, September). Using Stacking to Average Bayesian Predictive Distributions (with Discussion). *Bayesian Anal., 13*, 917-1007. Retrieved from https://projecteuclid.org/journals/bayesian-analysis/volume-13/issue-3/Using-Stacking-to-Average-Bayesian-Predictive-Distributions-with-Discussion/10.1214/17-BA1091.full

Yuling Yao, G. P. (2022, December). Bayesian Hierarchical Stacking: Some Models Are (Somewhere) Useful. *Bayesian Anal., 17*, 1043-1071. Retrieved from https://projecteuclid.org/journals/bayesian-analysis/volume-17/issue-4/Bayesian-Hierarchical-Stacking-Some-Models-Are-Somewhere-Useful/10.1214/21-BA1287.full

Appendix 1: Results of the simulation studies – PES response rate

We have presented the results of these simulations in the form of a graph, where the response rate is plotted on the x-axis and the width of the credible interval is plotted on the y-axis. The purpose of the graph is to illustrate the changes in credible intervals resulting from variations in PES response rates. Each point on the graph represents a simulation run at a specific response rate.

For each response rate interval, we conducted the simulation ten times[1]. Each run at the same response rate yielded a different estimate, leading to variation in the plotted dots. To visualize the relationship between decreasing response rates and their impact on the credible intervals, we added a trendline. The grey area surrounding the trendline indicates the variation in our estimates across the model runs for each response rate.

The y-axis is particularly important for us as it helps us assess whether we meet our target. For instance, if our KPI target is +/- 0.5, a width of 1.0 in the credible interval would meet the KPI requirement. Therefore, if our simulation results yield a maximum width of 0.8, we would have successfully met our KPI target.

Changing response rate at national level

Figure 10 focuses on the population at national level and examines the effect of reducing response rates uniformly across all groups. We conducted an analysis to explore the relationship between non-response rates and uncertainty, represented by the width of credible intervals. It is important to note that we examined unrealistic scenarios with extremely low response rates, specifically 10 percent and 30 percent.

In these scenarios, non-response was introduced at the national level for all households. The precision target for the national KPI is +/- 0.4 percent (equivalent to a credible interval width of 0.8 percent), which is expressed as the width of the credible interval.

[1] Due to computation requirements of running the model.

Figure 10 Relation between PES response rate and uncertainty of estimates at the national level (note that the axes are in logarithmic scale)

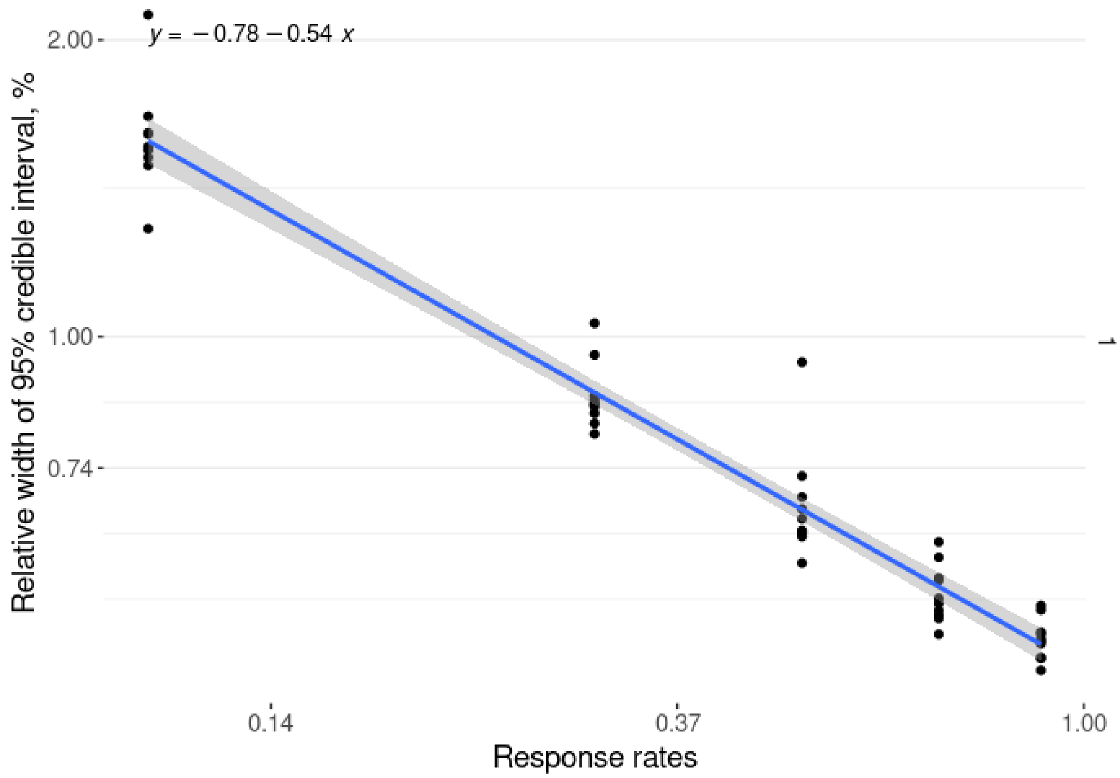
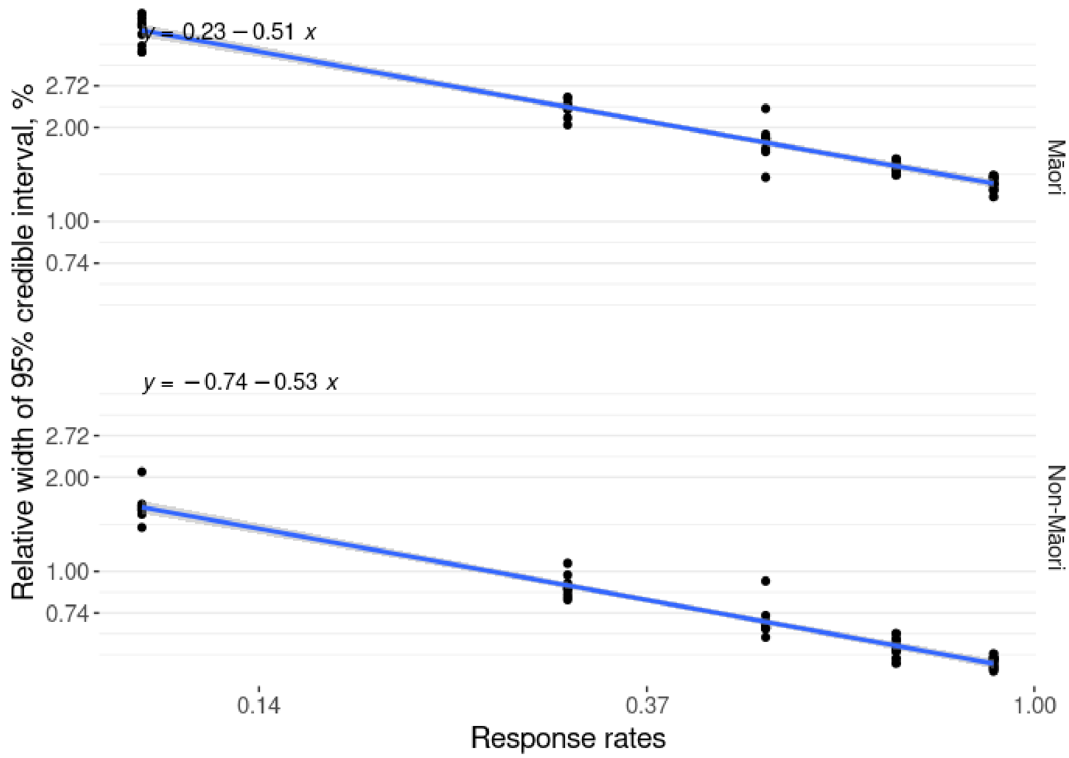


Figure 10 depicts that the uncertainties in the estimates do exceed the target KPI of +/- 0.4 percent (equivalent to a credible interval width of 0.8 percent) but only in the extreme cases of non-response, specifically when the response rate falls below approximately 35 percent. The results of the simulation study indicate that as long as there is a response rate of at least 35 percent at the national level, we would still be able to meet the KPI requirements.

Figure 11 shows the relationship between the PES response rate and the uncertainty of estimates for two distinct subpopulations: Māori and non-Māori. As observed in Figure 11, the coefficients in the trendlines for both groups are similar, with only a slight variation. However, there is a difference in the intercepts of the trendlines. This indicates that both subpopulations were affected by non-response at a similar rate, but they experienced different levels of uncertainty due to their differing initial population sizes.

Figure 11 Relation between PES response rate and uncertainty of estimates for Māori and non-Māori ethnicity subpopulations



Note: the x-scale ticks on Figure 10 and Figure 11 are different to the other plots because of the double log transformation in ggplot2. We are still able to meet KPI for Māori if the response rate at a national level is higher than 50 percent.

Changing response rate for ethnicity subpopulations

In this section, we examined the effect of PES non-response on the uncertainties of specific ethnic groups within the PES. It is important to mention that we only changed the response rate of one group at a time while keeping the response rate of the other group constant at 90 percent. The graphs in this section specifically focus on response rates ranging from 50 percent to 90 percent, rather than the broader range of 10 percent to 90 percent.

Figure 12 illustrates the relationship between changes in the response rate for the Māori ethnicity group and the resulting uncertainties in estimates for both the Māori and non-Māori groups.

Figure 12 Relation between Māori response rate and uncertainty of estimates for Māori and non-Māori ethnicity subpopulations

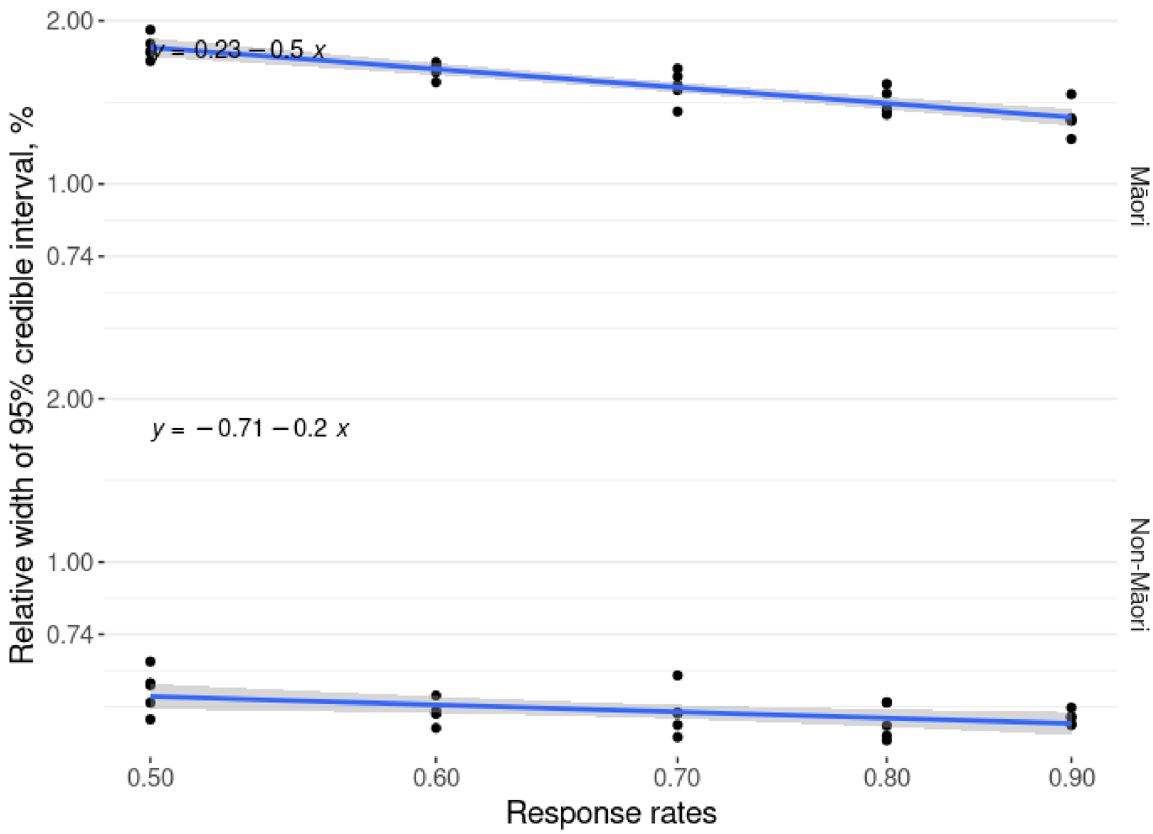
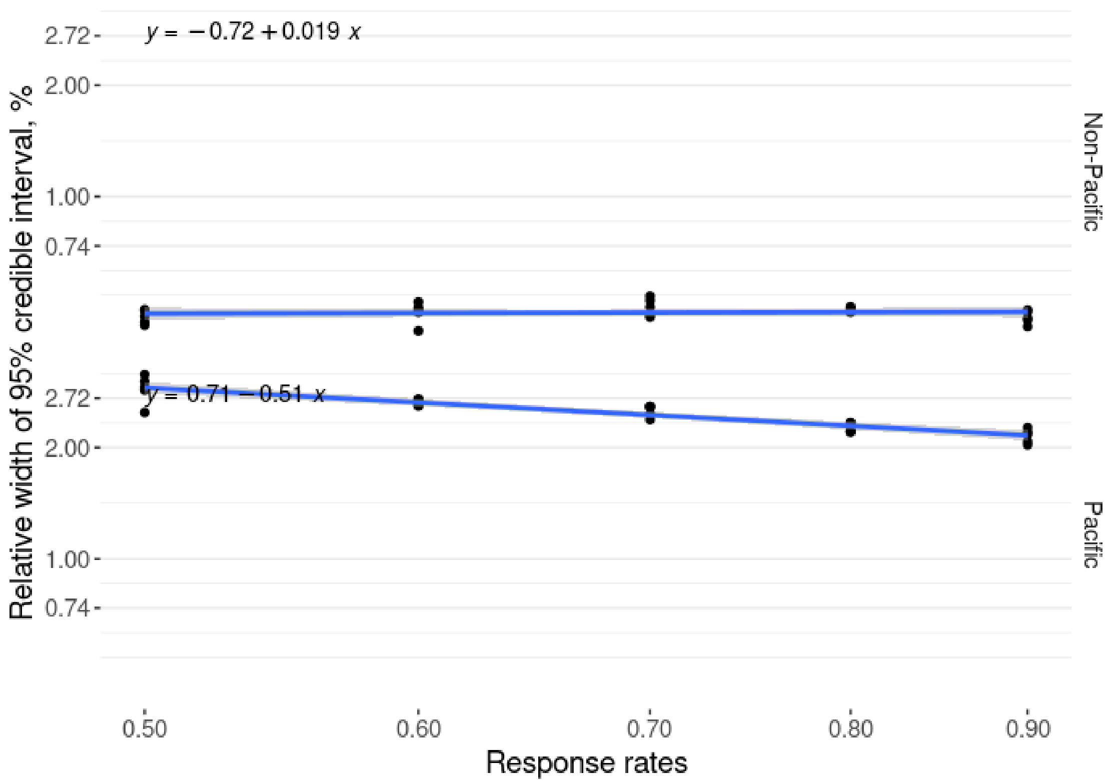


Figure 12 illustrates that as the non-response rate increases for the Māori ethnicity group, the width of uncertainty in estimates increases more for the Māori group compared to the non-Māori group. This outcome is not surprising since when non-response is higher in one group than the other, it is expected that the uncertainty will have a greater impact on that particular group with the higher non-response. This finding aligns with our observations from the national response rate analysis. Furthermore, even at the lowest response rate, the estimates remained within the target range of uncertainty (+/- 1 percent or 2.0 percent of the credible interval width).

Moving on to Figure 13, it examines the relationship between changes in response rates for the Pacific ethnicity group and the resulting uncertainties in estimates for both the Pacific and non-Pacific groups.

Figure 13 Relation between Pacific response rate and uncertainty of estimates for Pacific and non-Pacific ethnicity subpopulations



The results for the Pacific ethnicity simulations exhibit similarities to the Māori ethnicity results. The coefficient values for both groups are close, with 0.46 for Pacific ethnicity and 0.5 for Māori ethnicity. However, the intercept for the Pacific ethnicity group is greater than the intercept for the Māori ethnicity group. This finding is consistent with our previous simulations, where groups with smaller population counts tend to have wider uncertainty intervals. Importantly, the simulations demonstrate that even at low response rates, the uncertainty intervals remain within the predetermined targets (+/- 1.5 percent or 3 percent width of the credible interval).

Regarding the effect on the non-Pacific population, a similar trend was observed in the Māori ethnicity simulations when we modified the response rate of one group. In this case, by modifying the response rate for the Pacific ethnicity group, there is a small effect on the uncertainty of the non-Pacific group. This effect may be attributed to individuals who identify with multiple ethnic groups, as decreasing the response rate in one group can inadvertently affect another.

Furthermore, the effect on the non-Māori group is greater than the effect on the non-Pacific group. One possible explanation is that individuals who identify as Māori are more likely to identify with another ethnic group compared to those who identify as Pacific. This observation is supported by the simulation data, where 65 percent of the Pacific population exclusively identified as Pacific, while only 45 percent of the Māori ethnicity population solely identified as Māori ethnicity. This difference in identification patterns likely accounts for the effect observed in the larger non-focus group.

Changing response rate for different age groups

In this section, we examined the impact of changes in response rates for selected broad age groups. Specifically, we focused on two groups for comparison: the older age group (65+) with typically higher response rates, and the younger age group (15–29) with generally lower response rates. In the first simulation (Figure 14), we only manipulated the response rate of the older age group (65+), while keeping the response rate of the younger age group constant. For Figure 15, we solely modified the response rate of the younger age group (15–29). These simulations allowed us to observe the effects of changing response rates within each age group and understand the corresponding uncertainties in the estimates.

Figure 14 Impact of changing response rates of older age group (65+) on the width of credible intervals.

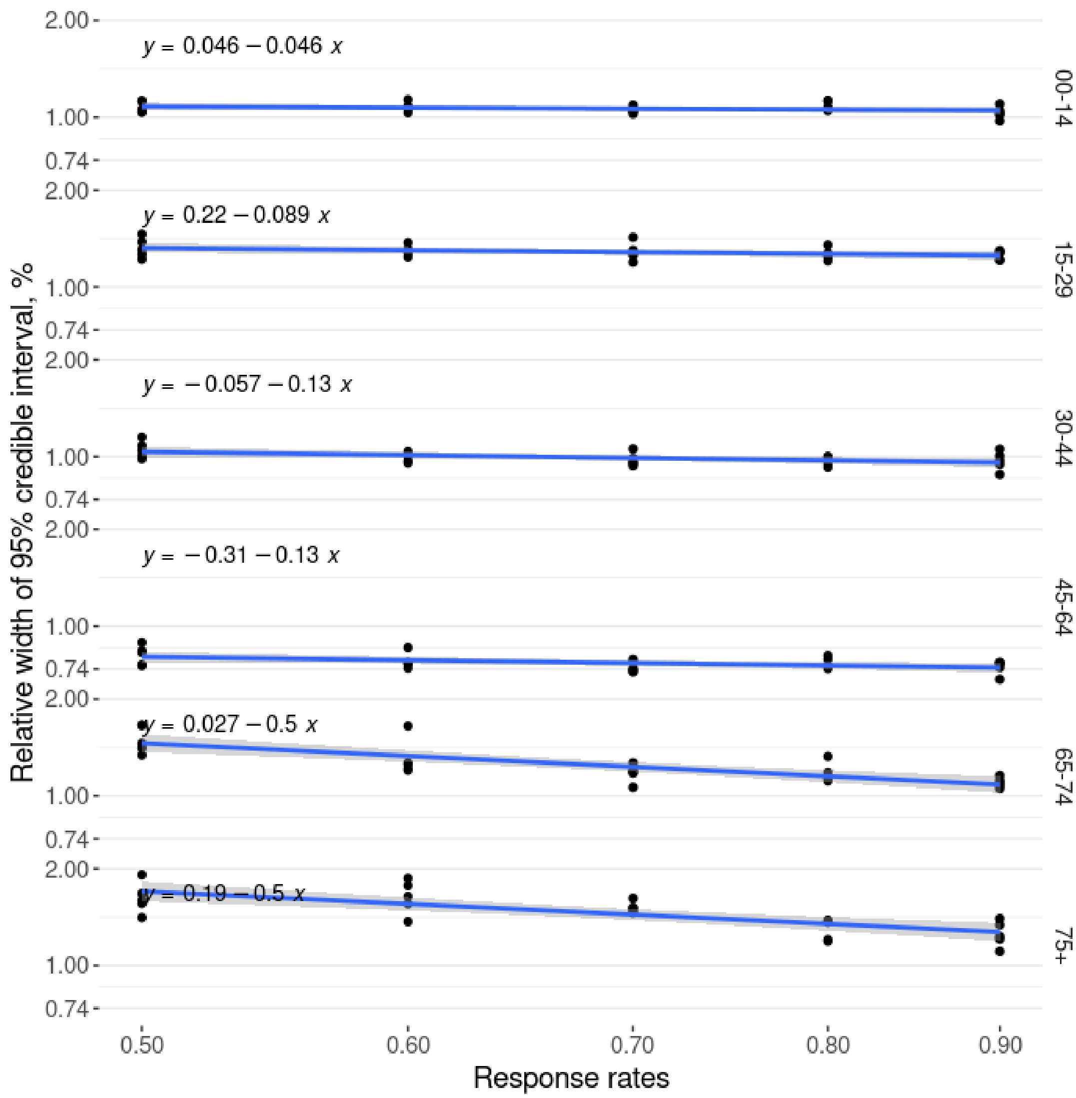
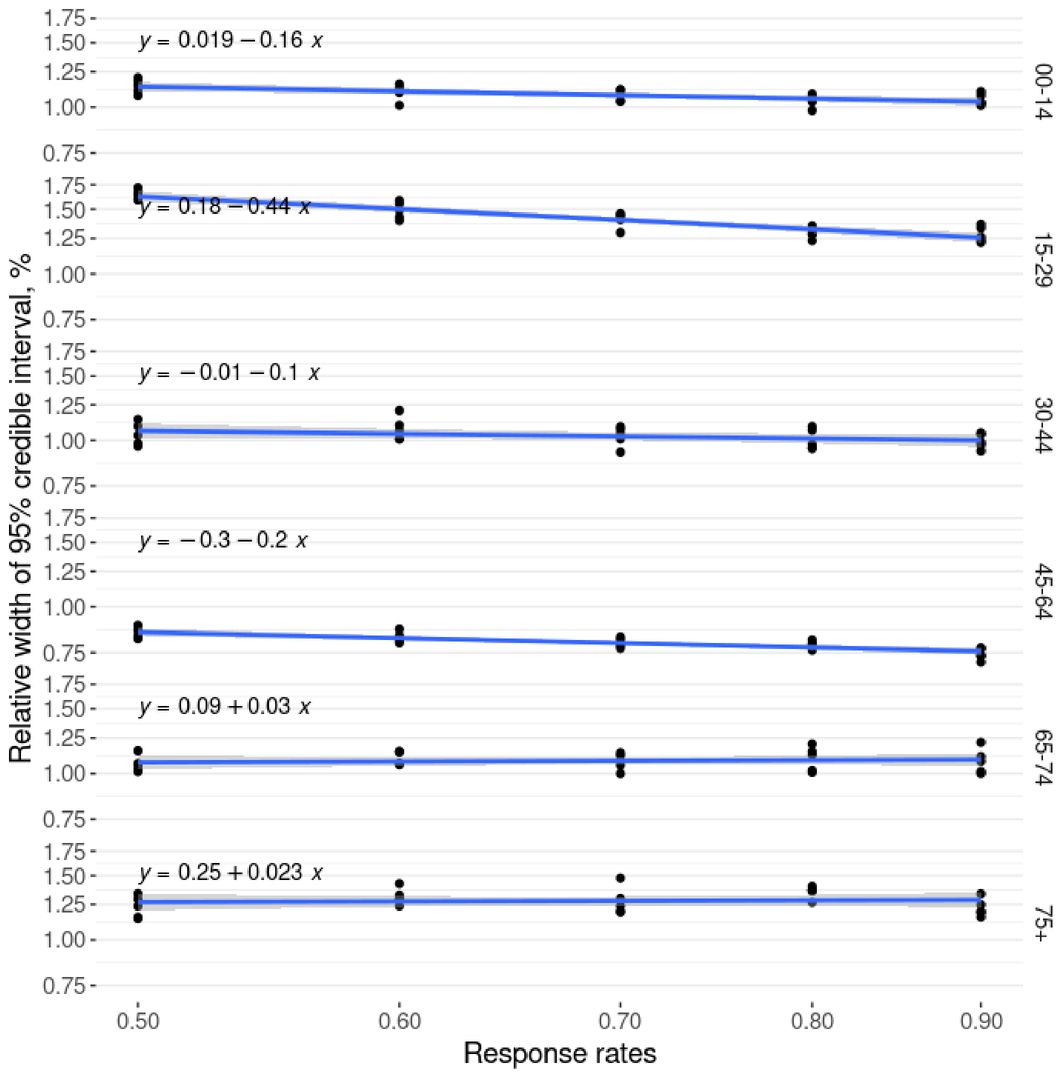


Figure 15 Impact of changing response rates of young age group (15-29) on the width of credible intervals



One notable difference between the age group simulation and the previous ethnicity simulation is that changes to one age group do not have an influence on other age groups. In contrast, in the ethnicity simulation, modifying one group had a small effect on others due to the possibility of individuals reporting multiple ethnicities, making ethnicities not mutually exclusive. However, this is not the case for age groups since an individual cannot belong to multiple age groups.

In the age group simulation, we observe that only the changed age group experiences a change in response rate. As with the other scenarios, a decrease in response rate leads to an increase in the width of relative uncertainty. However, there is one exception in Figure 14, where the younger age groups are modified and the older age groups exhibit an increasing trend. Nevertheless, the positive coefficient is low, resulting in an almost flat trend line. This trend is likely a characteristic of the older population groups and should not be attributed to the effect on the younger age groups, as age groups are mutually exclusive and should not influence each other.

The most important finding from this simulation is that even when the response rate is low, such as 50 percent, the relative width of uncertainty remains within the bounds of our established KPIs. This observation is consistent with the findings from the other simulations.

Effect of PES response rate on small sub-populations

Figure 16 shows the results of the simulation study of exploring the effects of non-responses of PES on a small sub-group. The red horizontal line in Figure 16 denotes the KPI for gender of 0.8 percent (+/-0.4 percent). The KPI is only for the male and female categories, there is no KPI set for the 'another gender' category.

Figure 16 Uncertainty of gender estimation

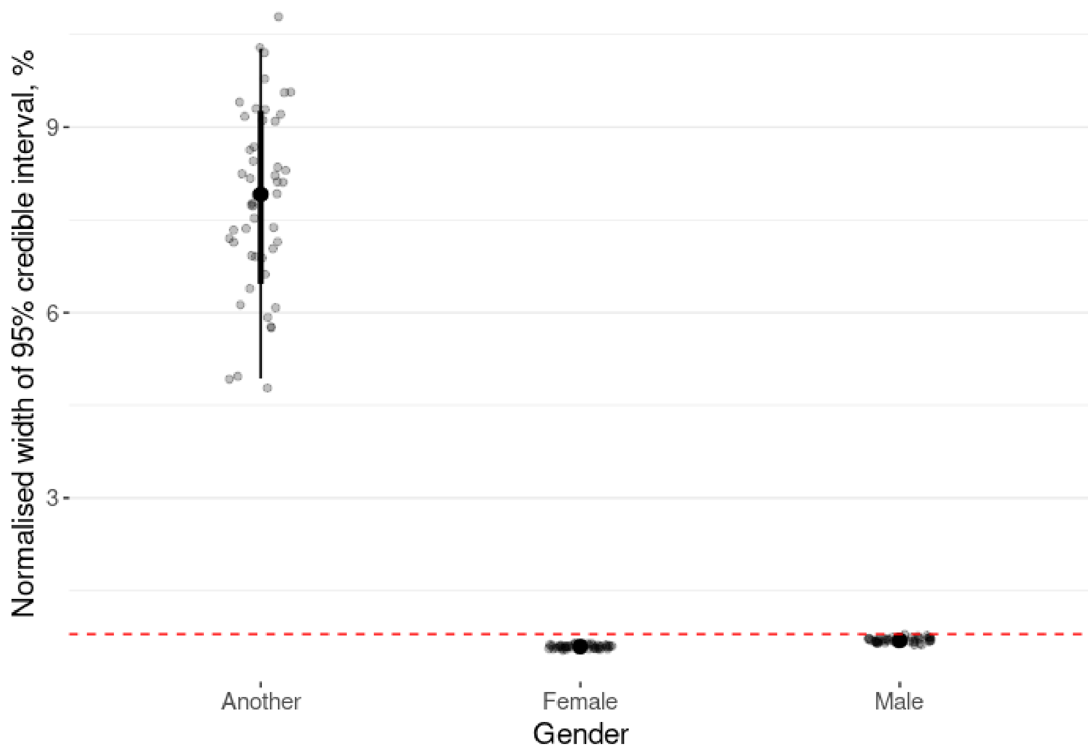


Figure 16 shows that the findings are consistent with what we found when we simulated a decrease in response in bigger groups: the smaller the sample size, the more significant the increase in uncertainty. However, in the case of another gender, the uncertainty is extreme.

Appendix 2: Results of simulating census challenges

Census coverage rate

The effect of measured census under-coverage on the uncertainty of final population estimates of the Māori-ethnicity and Māori descent population are shown in Figure 17 and Figure 18, respectively. Each point shows the result of one simulation run. The horizontal red dashed line marks the PES KPI for Māori ethnicity and Māori-descent population.

Figure 17 Uncertainty of Māori ethnicity population estimates under different levels of census coverage rates

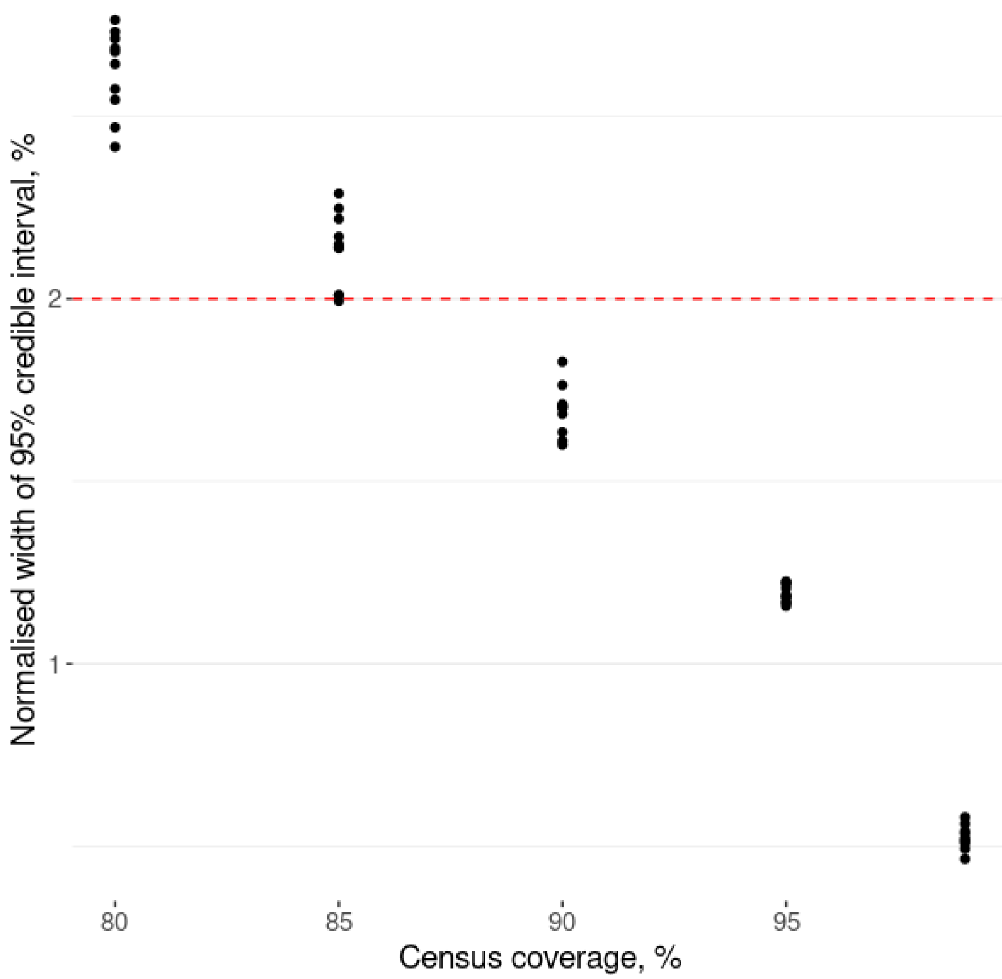
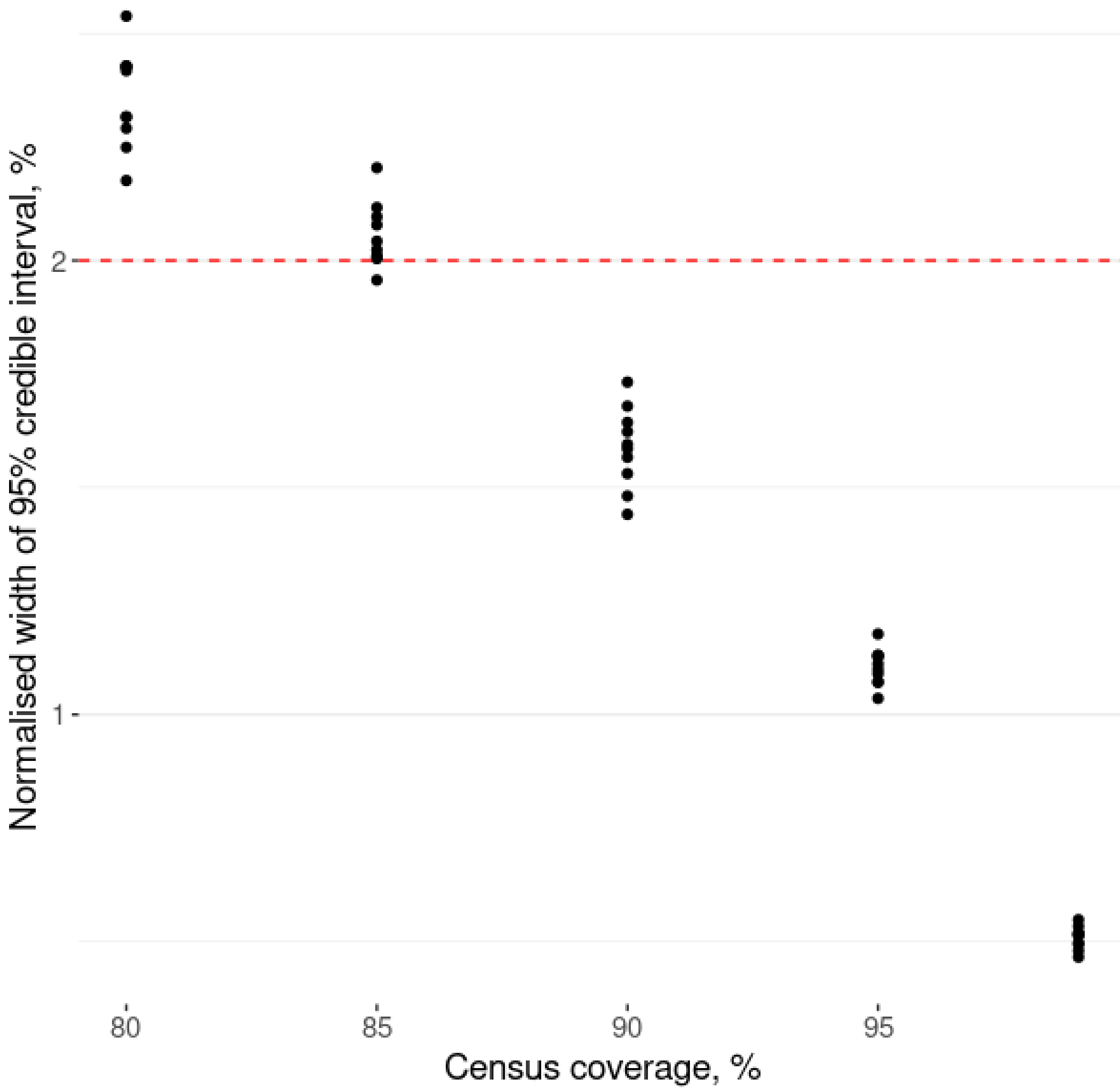
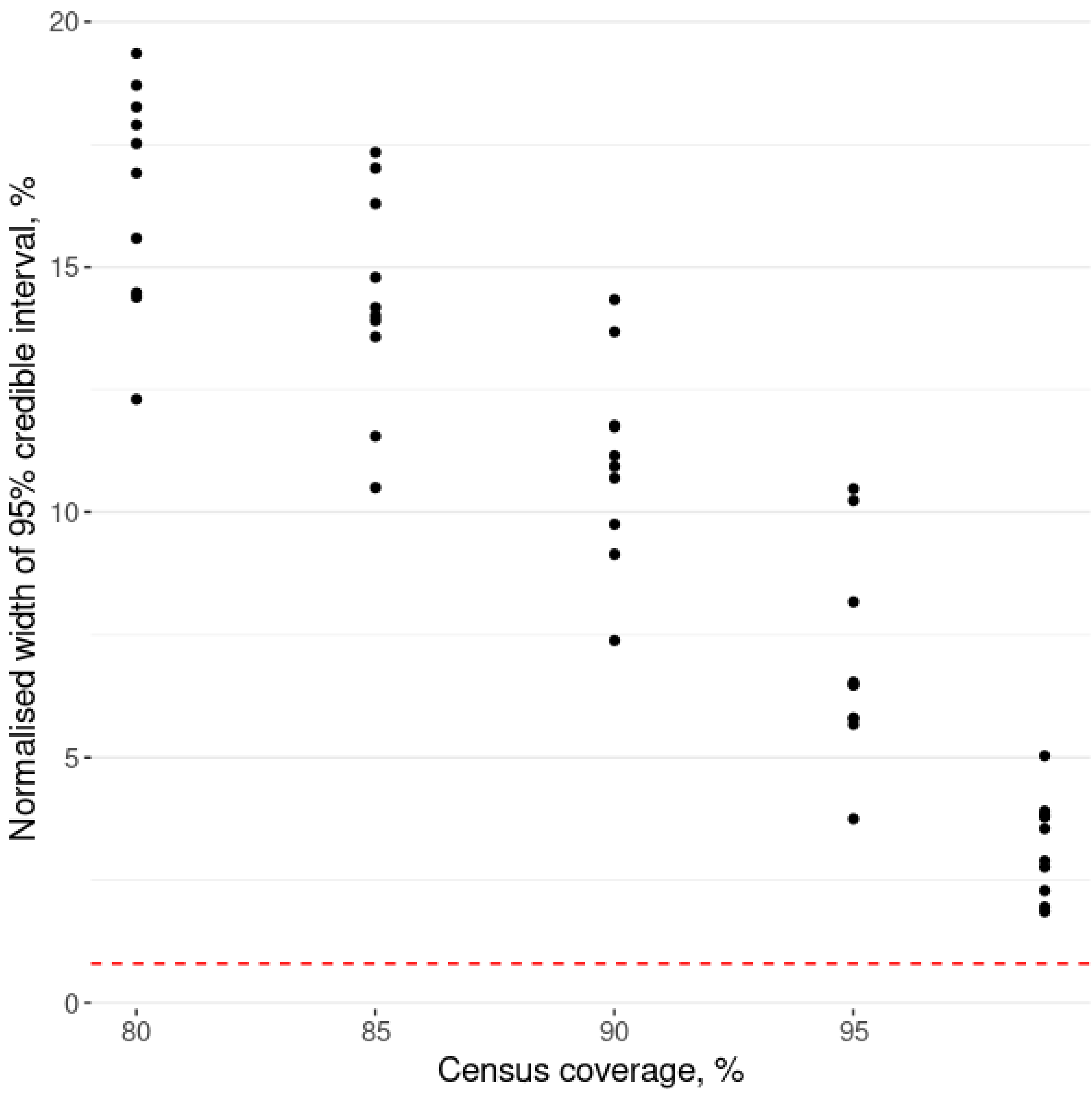


Figure 18 Uncertainty of Māori descent population estimates under different levels of census coverage rates



The effect of measured under-coverage on the uncertainty of final population estimates of the 'another gender' population is shown in Figure 19. The x-axis shows the tested census coverage rate, and y-axis shows the width of the credible intervals. Each point shows the result of one simulation run. The horizontal red dashed line marks the PES KPI for the male and female genders.

Figure 19 The effect of measured under-coverage on the uncertainty of final population estimates of the 'another gender' population.



Dependency between PES and census non-response

For the correct estimation of census coverage rates, we need to ensure that the PES sample is representative of the true population. In the situation where PES non-response is associated with census non-response, we expect bias in the final estimates. To model this situation, we simulated a situation where PES non-response was correlated with census non-response. This scenario represents the violation of DSE assumption about the independence of two lists: probability of being recorded in PES is dependent on the probability of being in census. Figure 20 illustrates the bias resulting from census under-coverage and PES non-response, specifically examining the impact of different levels of dependency (represented by γ) between the two lists at the individual level. The focus is on the Māori and non-Māori ethnicity sub-populations. The γ values indicate the strength of the dependence between individuals who are under-coverage in the census and non-response in the PES at the individual level. On the X-axis, we have the Māori ethnicity census coverage, while the coverage of the non-Māori ethnicity group remains constant at 99 percent. Each data point on the graph represents the result of one simulation iteration, providing insights into the bias caused by the interplay of under-coverage and non-response.

Figure 20 Bias resulting from census under-coverage and PES non-response at the individual level.

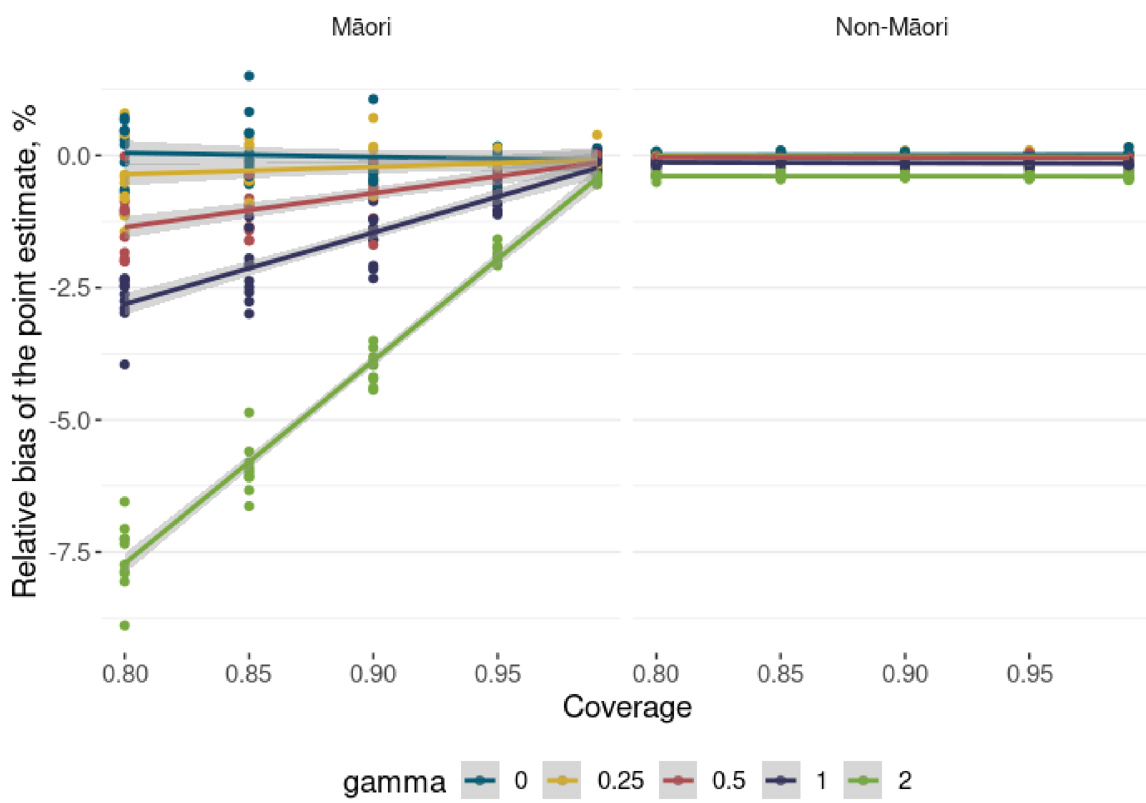
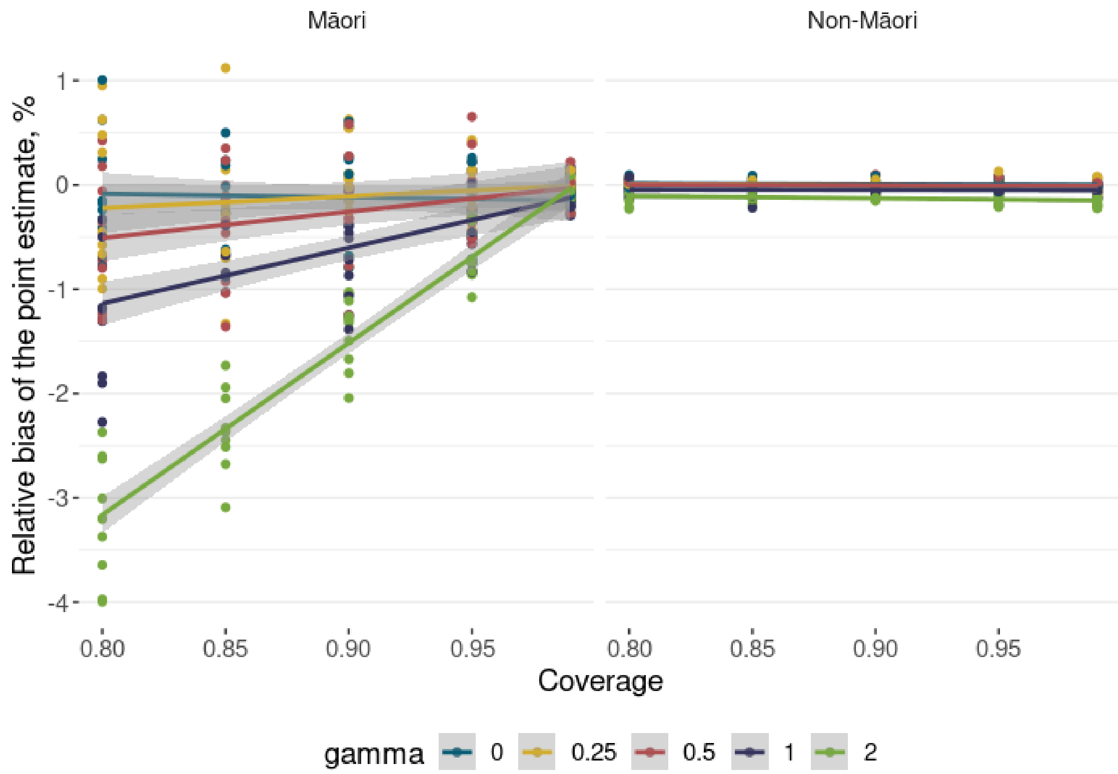


Figure 21 is similar to Figure 20, but in this case, the γ values represent the strength of dependency between the two lists at the dwelling level. The focus is still on the Māori and non-Māori ethnicity sub-populations. Each data point on the graph represents the result of one simulation iteration. The purpose of Figure 21 is to examine how the strength of dependency at the dwelling level impacts the bias caused by census under-coverage and PES non-response. By analysing the relationship between the γ values and the resulting bias, we gain insights into the interplay between these factors and their effects on the estimates for the different ethnicity sub-populations.

Figure 21 Bias resulting from census under-coverage and PES non-response at the dwelling level.



Both Figure 20 and Figure 21 demonstrate an interaction between decreasing coverage rates and the dependency between the two population lists of PES and the census. In the right subplot (non-Māori), where the coverage rate remains constant at 99 percent, changes in the dependency parameter (gamma) do not have an impact on the relative bias of our estimates. However, in the left subplot (Māori), we observe an increasing amount of negative bias in our estimates, indicating an underestimation of the true value, as the coverage rate decreases. This bias is further intensified by higher gamma values, which indicate a higher likelihood of individuals (or dwellings) missed by the census also not responding in the PES.

Unexpected non-homogeneity

Figure 22 displays the precision of national population count estimates specifically for the young adult population (aged 15–29) at different levels of census coverage. Each data point represents the result of one iteration, and the colour of the dots indicates whether the estimation was conducted using a model that accounted for the under-coverage of youth (red, correct) or a model without the inclusion of the 'youth' covariate (blue, incorrect).

Figure 22 Precision of the national population count estimates.

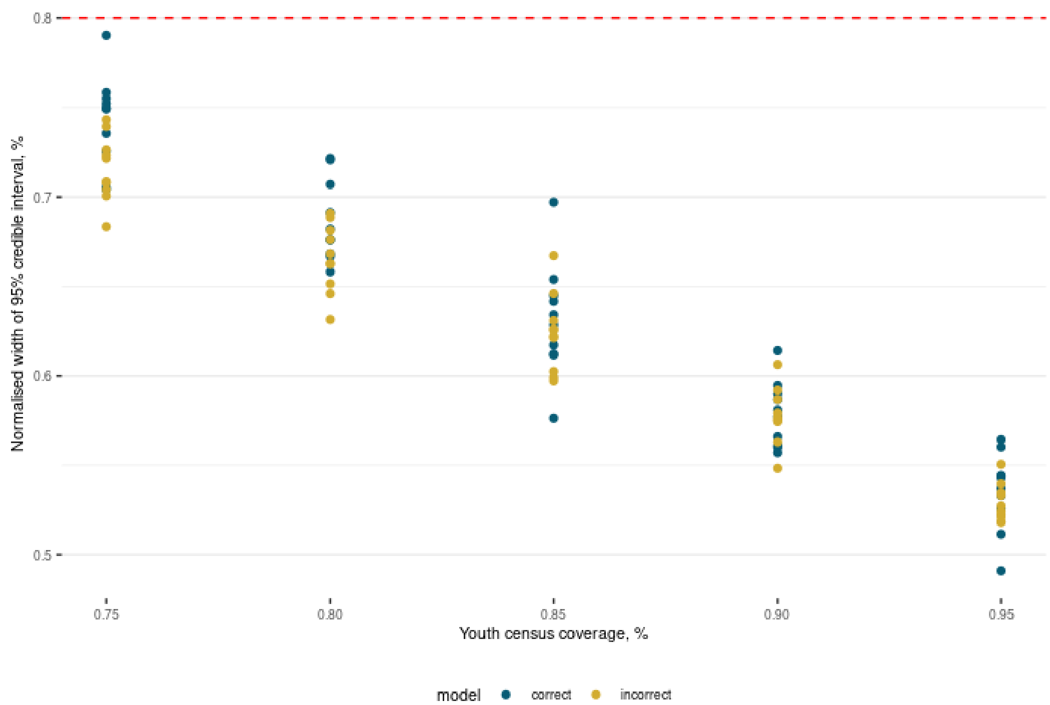


Figure 23 depicts a similar analysis as Figure 22, but focuses on different age subgroups. It presents the precision of population count estimates for each age subgroup at varying levels of census coverage. In Figure 23, each dot represents one iteration result, the colour of the dots shows if we performed the estimation with the model accounting for the youth under-coverage (red, correct) or the model without 'youth' covariate (blue, incorrect).

Figure 23 Precision of the age subgroup count estimates.

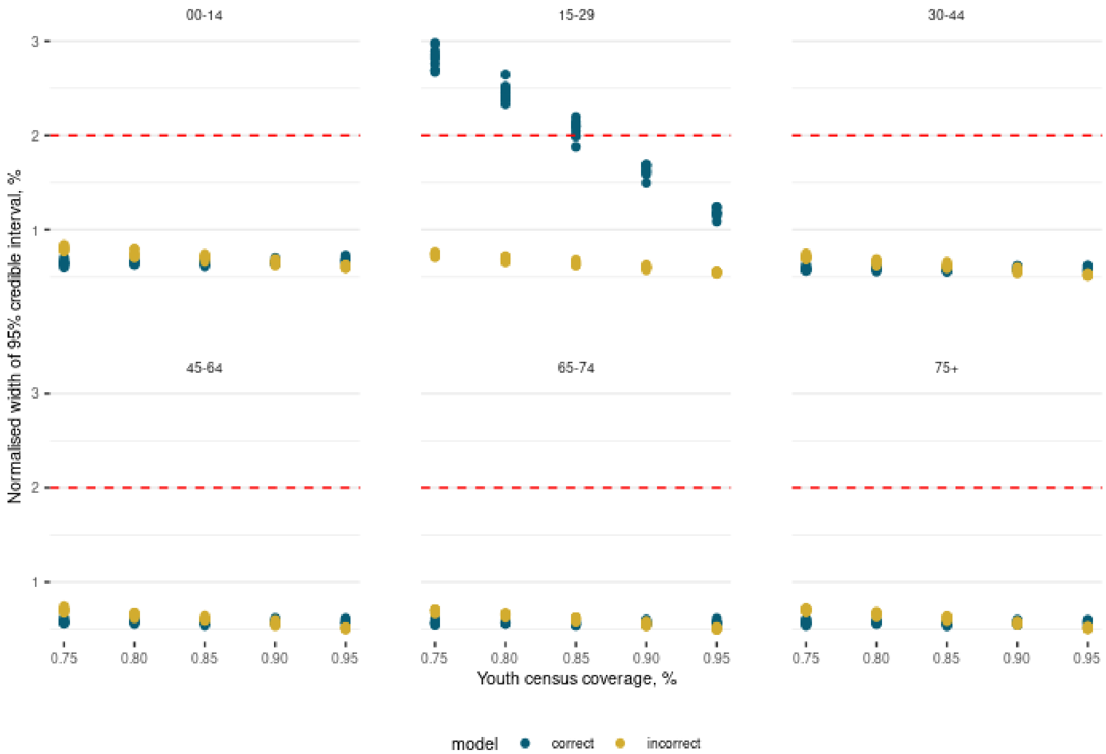


Figure 24 and Figure 25 show the bias of population counts at the national level and for each age subgroup, respectively. Again, each dot represents one iteration result, the colour of the dots shows if we performed the estimation with the model accounting for the youth under-coverage (red, correct) or the model without 'youth' covariate (blue, incorrect).

Figure 24 Bias of the national population count estimates.

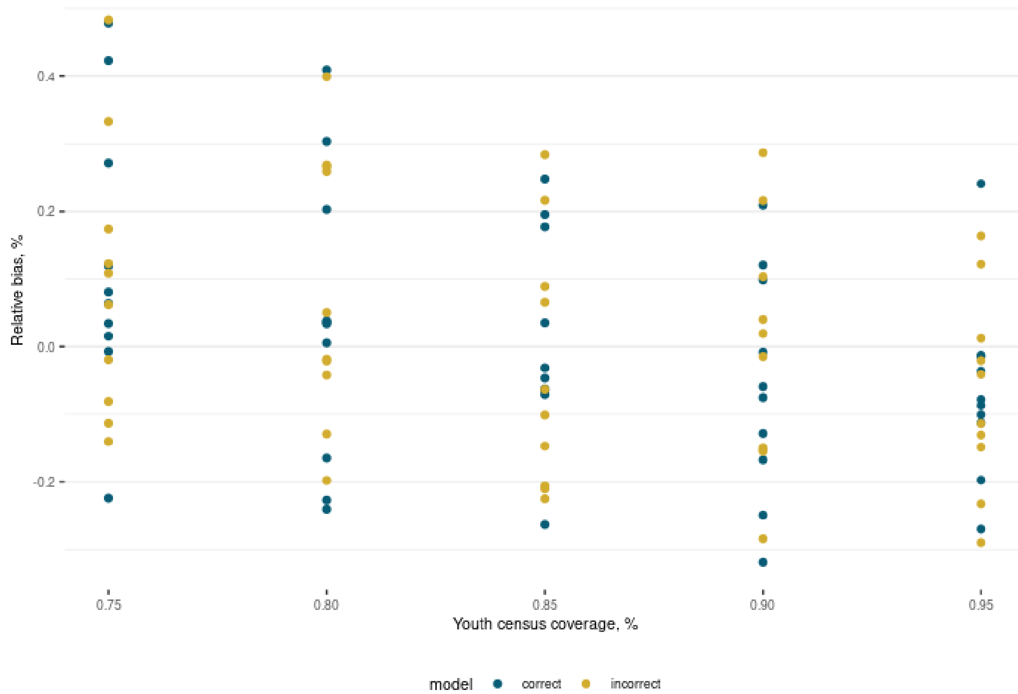
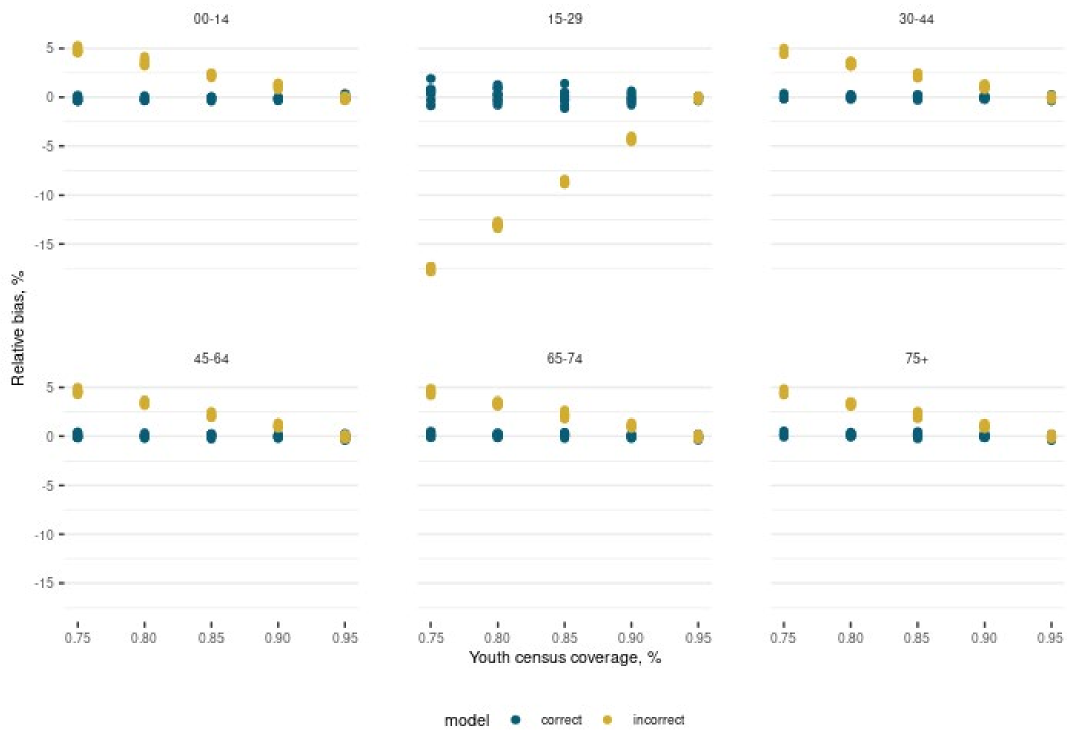


Figure 25 Bias of the age subgroup count estimates.



Appendix 3: Simulation of late census responses

Scenario 1: PES captures biased proportion of late census responses, but captures under-coverage correctly

In this scenario, we simulated the situation when PES correctly captures the proportion of under-coverage. However, the bias is caused by incorrect proportions of on-time and late responses. Parameters used in the simulation studies are shown in Table 11.

Table 11 Parameters used in the simulation studies in Scenario 1

	Total count	5,000,000										
True population	On-time, %	80										
	Lates, %	10										
	Under-coverage, %	10										
PES Sample	On-time, %	90	88	86	84	82	80	78	76	74	72	70
	Lates, %	0	2	4	6	8	10	12	14	16	18	20
	Under-coverage, %	10	10	10	10	10	10	10	10	10	10	10
Bias type		PES under-samples lates					No Bias	PES over-samples lates				

For each simulation study, we estimated the population. Results of the simulation study is shown in Figure 26.

Figure 26 Simulated population estimation results in scenario 1

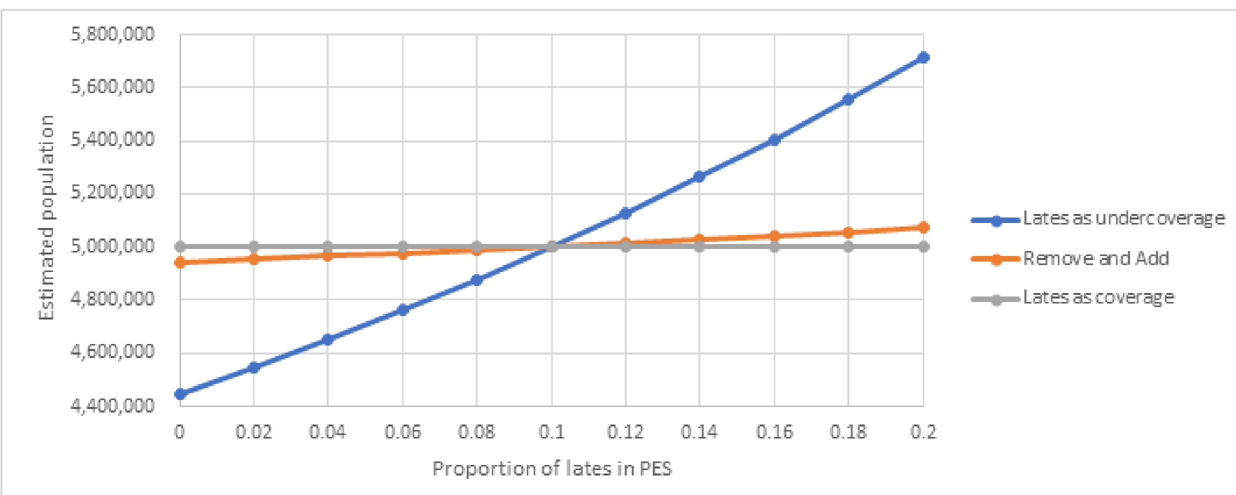


Figure 26 shows that the method where lates are treated as coverage (‘lates as coverage’ method) performed the best in this scenario, while treating lates as under-coverage showed the highest susceptibility to bias. The 2018 method, remove and add, exhibited significantly lower bias compared to treating lates as under-coverage.

Scenario 2: PES captures biased proportion of under-coverage and lates, but captures on-time correctly

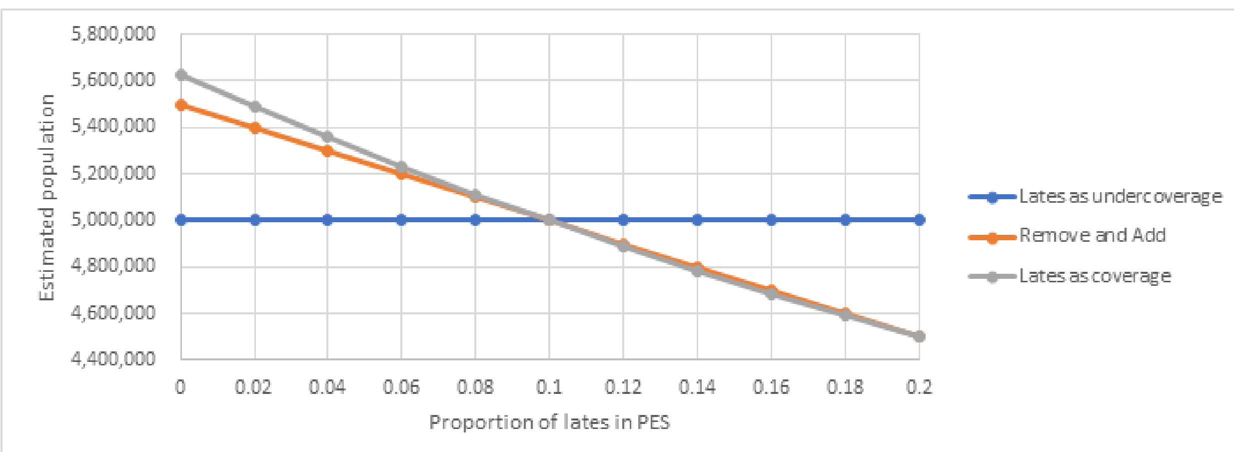
In this scenario, we examined a bias scenario where PES activity can either discourage or encourage late census responses. Parameters used in the simulation studies are shown in Table 12.

Table 12 Parameters used in the simulation studies in Scenario 2

	Total count	5,000,000										
True population	On-time, %	80										
	Lates, %	10										
	Under-coverage, %	10										
PES Sample	On-time, %	80	80	80	80	80	80	80	80	80	80	80
	Lates, %	0	2	4	6	8	10	12	14	16	18	20
	Under-coverage, %	20	18	16	14	12	10	8	6	4	2	0
Bias type		PES activity prevents late submissions					No Bias	PES activity converts under-coverage to late responses				

Results of the simulation study is shown in Figure 27, Simulated population estimation results in Scenario 2.

Figure 27 Simulated population estimation results in Scenario 2



As Figure 27 *shows that* treating lates as under-coverage gives unbiased results in Scenario 2. The method used in 2018 performs slightly better than the method where we treat lates as coverage.

Scenario 3: PES captures biased proportion of on-time responses, but captures under-coverage and lates correctly

In this scenario we simulated the situation when PES under- or over-samples lates and under-coverage people with the same rate. Parameters used in the simulation studies are shown in Table 13.

Table 13 Parameters used in the simulation studies in Scenario 3

	Total count	5,000,000										
True population	On-time, %	80										
	Lates, %	10										
	Under-coverage, %	10										
PES Sample	On-time, %	100	96	92	88	84	80	76	72	68	64	60
	Lates, %	0	2	4	6	8	10	12	14	16	18	20
	Under-coverage, %	0	2	4	6	8	10	12	14	16	18	20
Bias type		PES oversamples on-time responses, under-samples lates and under-coverage on the same rate					No Bias	PES under-samples on-time responses, over-samples-samples lates and under-coverage on the same rate				

Results of the simulation study is shown in Figure 28.

Figure 28 Simulated population estimation results in Scenario 3

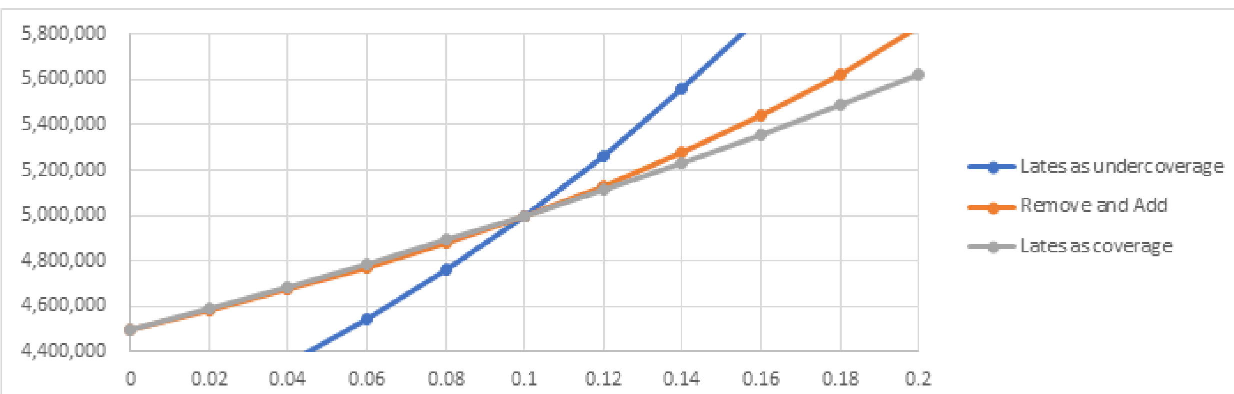


Figure 28 shows that in Scenario 3, none of the methods could recover true population in the presence of the PES sampling bias.

Appendix 4: Regression analysis of the relationship between 2018 late census responses and PES activity

Here, we denote the proportion of lates in a PSU i as y_i. We then fit a zero-inflated beta regression model, where proportion of lates was a response variable and PSU characteristics were predictors. The choice of the model was made because of the large numbers of PSUs with 0 late responses, and because the beta-distribution is a common approach to model proportions. Logistic regression with the individual probabilities of late responses was not practically possible due to the large size of the census dataset.

$$f_Y(y_i|\mu,\phi,\pi) = \begin{cases} \pi_i & if\ y_i = 0 \\ (1-\pi_i)\dfrac{y_i^{\mu_i\phi_i-1}(1-y_i)^{(1-\mu_i)\phi_i-1}}{B(\mu_i\phi_i,(1-\mu_i)\phi_i)} & if\ y_i > 0 \end{cases}$$

$$logit(\pi_i) = \boldsymbol{X_i}\beta_\pi$$

$$logit(\mu_i) = \boldsymbol{X_i}\beta_\mu$$

$$log(\phi_i) = \boldsymbol{X_i}\beta_\phi$$

The model was fitted with the R package *brms* with default set of priors. List of PSU covariates $\boldsymbol{X}$ used in the model are listed in the Table 14.

Table 14 Variables used in the analysis.

Covariate	Values
Intercept	Constant 1 for each census PSU
Proportion Māori-descent individuals in a PSU	Continuous: 0-1
Proportion Pacific ethnicity in a PSU	Continuous: 0-1
PES area indicator	Binary: – selected for PES, 1 – not selected for PES
Proportion of Māori-descent individuals and PES area indicator interaction	Interaction variable

Proportion of Pacific individuals and PES area indicator interaction	Interaction variable

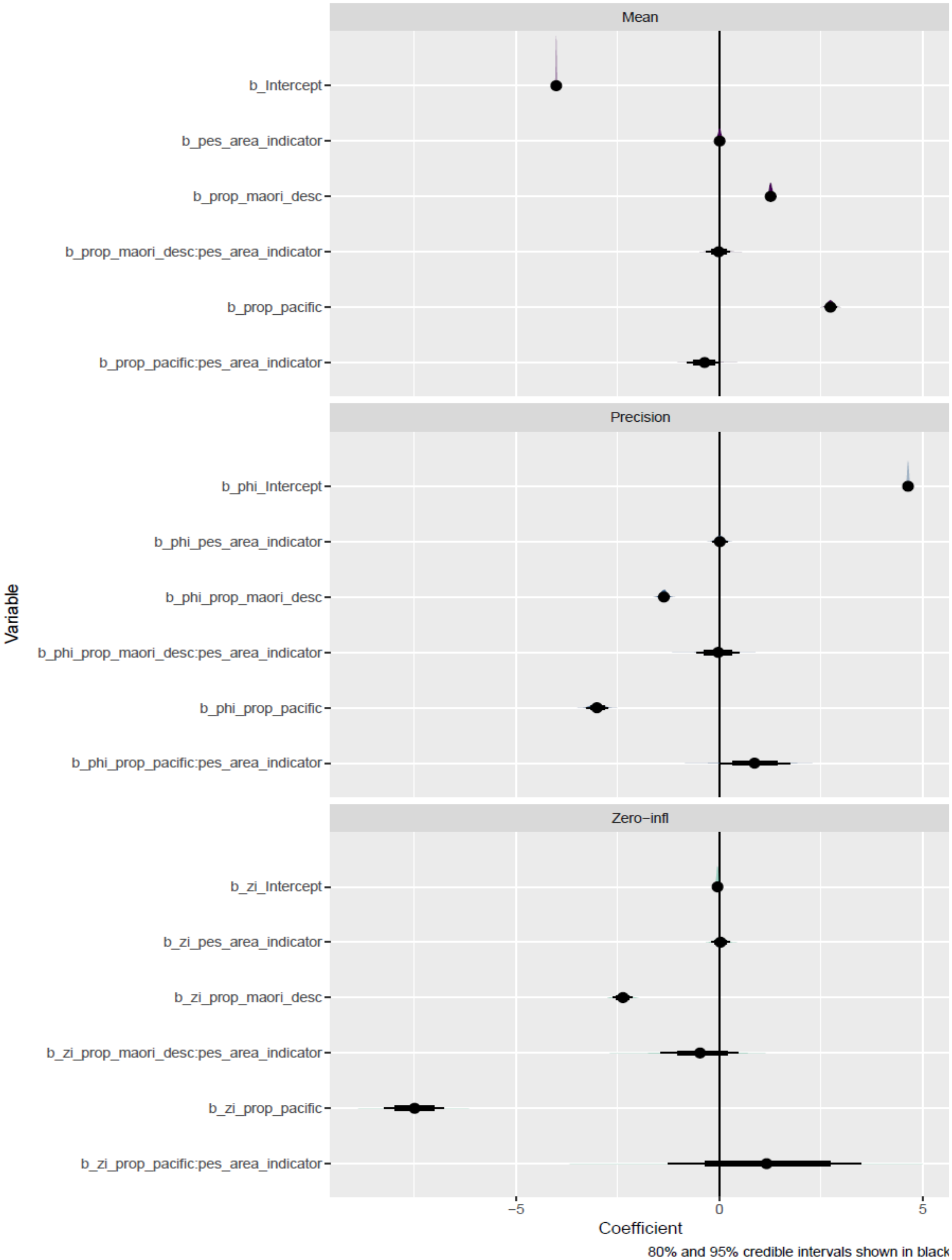
Table 15 and Figure 29 below shows the statistics of the posterior distributions of model parameters and their interpretation. In Figure 29, Y-axis shows the variable, x-axis shows the values, the points show the median of the posterior distribution, intervals show 80 percent (thicker line) and 95 percent (thinner line) credible intervals.

Table 15 Estimated covariate effects on late response proportions in PSUs.

Parameter	Estimate	Lower bound, 95% CI	Upper bound, 95% CI	Interpretations of the effects of the significantly contributing variables
Mean (how high is the proportion of lates in average)				
Intercept	-4.01	-4.03	-3.99	
Proportion Māori-descent	1.26	1.19	1.34	On average, Māori-descent population has higher proportion of lates
Proportion Pacific	2.73	2.59	2.89	On average, Pacific ethnicity population has higher proportion of lates
PES area indicator	0.01	-0.07	0.10	
Māori-descent and PES area interaction	-0.01	-0.29	0.27	
Pacific and PES area interaction	-0.36	-0.76	0.06	
Size (how wide is the distribution of the number of lates)				
Intercept	4.64	4.60	4.68	
Proportion Māori-descent	-1.36	-1.50	-1.22	Proportion of Māori-descent individuals affects the shape of lates distribution

Proportion Pacific ethnicity	-3.01	-3.27	-2.74	Proportion of Pacific ethnicity individuals affects the shape of lates distribution
PES area indicator	0.01	-0.15	0.19	
Māori-descent and PES area interaction	-0.03	-0.57	0.47	
Pacific and PES area interaction	0.86	-0.01	1.71	
Zero-inflated model (probability to have no lates in a PSU)				
Intercept	-0.04	-0.09	0.01	
Proportion Māori-descent	-2.37	-2.61	-2.13	High proportion of Māori descent individuals in a PSU reduces the probability to have no lates
Proportion Pacific	-7.49	-8.24	-6.77	High proportion of Pacific ethnicity individuals in a PSU reduces the probability to have no lates
PES area indicator	0.03	-0.19	0.25	
Māori-descent and PES area interaction	-0.48	-1.47	0.45	
Pacific and PES area interaction	1.13	-1.37	3.43	

Figure 29 Posterior distributions of covariate effects.



Appendix 5: Regression analysis of the relationship between late census responses and under-coverage in 2018 PES data

The model was fitted using R package *brms*.

First, a binary indicator of late response, $Y_{late_{it}}$ (1 – late, 0 – on-time) was assigned to each record i in a TALB t. Then, the following regression model was fit.

$$Y_{late_{it}} \sim Bernoulli(p_{late_{it}})$$

$$logit(p_{it}) = \mu + \boldsymbol{X_{it}}\beta + \alpha_t$$

Where, $p_{late_{it}}$ is a probability of the late response, μ – is a global mean, $\boldsymbol{X_{it}}$ – a covariate matrix, β- covariate effects, α_t- random effect of TALB, σ_t^2- variance of TALB effects.

Table 16 shows a list of covariates used in the model.

Table 16 Variables used in the regression analysis of late and under-coverage relationship.

Covariate	Values
Intercept(TALB)	Each TALB has a different intercept
Sex	A binary variable where 1 represents 'male' and 0 represents 'female'
Māori Descent	A binary variable where 1 represents 'Māori ' and 0 represents 'Non-Māori'
Pacific Ethnicity	A binary variable where 1 represents 'Pacific' and 0 represents 'Non-Pacific'
Youth	A binary variable where 1 represents 'people with age between 15-29' and 0 represents 'others'
prop_ucov	A continuous variable between 0-1 that shows the percentage of under-coverage per PSU

The statistics of posterior distributions of covariate effects are in Table 17.

Table 17 Summary table of fixed effects.

	Estimate	Lower Bound 95% CI	Upper Bound 95% CI

Intercept	-4.5	-4.79	-4.24
Maori_desc (= 1)	0.68	0.42	0.94
Youth (= 1)	0.66	0.34	0.97
Sex (= 1)	0.19	-0.04	0.41
Pacific (= 1)	0.62	0.3	0.91
Maori_desc:youth =(1:1)	-0.77	-1.31	-0.27
Maori_desc:sex =(1:1)	-0.27	-0.64	0.1
Youth:Sex =(1:1)	-0.42	-0.86	0.03
Sex:Pacific =(1:1)	-0.09	-0.5	0.32
Youth:Pacific =(1:1)	-0.59	-1.18	-0.04
Maori_desc:Youth:Sex =(1:1:1)	0.38	-0.41	1.17
Youth1:Sex:Pacific =(1:1)	-0.15	-1.03	0.75

Appendix 6: Mathematical calculations of classification of late responses in PES

Suppose the total number of people who fall into one of the cases described in Table 7 is k. Suppose equal numbers move in each month - that is $\frac{k}{5}$ people move out of Auckland each month. The cases that should not have been classified as late are AAAAAB and AAAABB. Suppose PES interviews a third of those selected for PES each month. Thus, AAAAAB has a $\frac{1}{3}$ chance of being classified as late, and AAAABB has a $\frac{2}{3}$ chance of being classified as late. Thus, we would expect a misclassification of $\frac{k}{5} \times \frac{1}{3} + \frac{k}{5} \times \frac{2}{3} = \frac{k}{5}$ – so 20 percent of the time we are misclassifying on-time responses as late.

Suppose the total number of people who fall into one of the cases described in Table 8 is k. Suppose equal numbers move in each month - that is $\frac{k\prime}{6}$ people move out of Auckland each month. Suppose PES interviews a third of those selected for PES each month. Thus, BBBBBBA has a $\frac{1}{3}$ chance of being classified as on time, and BBBBBAA has a $\frac{2}{3}$ chance of being classified as on time and BBBBAAA will always be classified as on time. Thus, we would expect a misclassification of $\frac{k}{6} \times \frac{1}{3} + \frac{k}{6} \times \frac{2}{3} + \frac{k}{6} \ = \frac{k}{6}$ – so 33 percent of the time we are misclassifying late responses as on time.

Appendix 7: Change in relative bias for different treatment of misclassified late responses.

The simulations are based on using Linkoln-Petersen estimator, which includes the correction for lates/on-time misclassification (Method 1), and without the correction (Method 2). The relative bias was calculated as:

$$\frac{(N_{estimate} - N_{true})}{N_{true}} \times 100\%$$

- Method 1 adjusts the late classification of census responses for those respondents selected for PES. The relative bias for this method is $\frac{\kappa N_{ucov}}{1-\lambda\mu} - N_{ucov}$ where κ is the selection rate for PES for under-coverage, normalised by the selection rate for PES for on-time census responses, λ is the selection rate for PES for late responses, normalised by the selection rate of PES for on-time census responses, and μ is the number of responses that are misclassified as on-time and changed to late by Method 1, divided by the number of on-time census responses. This is a little indigestible but is written in this way to minimise the number of variables.
- Method 2 does not adjust the late classification of census responses. The relative bias calculation in this case is $(\kappa - 1)N_{ucov}$.

To give some context, if we have a non-response rate of 50 percent, $\lambda \times \mu = 0.1/50$, where 0.1% is an upper bound for the general population for the percentage of people to fall into the class of those who submit a census response that is misclassified as on time. Thus, for the general population, we would, at a very conservative estimate, expect $\lambda \times \mu < 0.2\%$.

When $\kappa < 1$, the relative bias is negative, meaning the estimates are underestimates. In this case, for the values we are interested in, Method 1 provides a relative bias that is closer to zero than method 2. The size of this difference is shown in Figure 30 and in Figure 31.

When $\kappa > 1$, the relative bias is positive, and Method 2 provides better results. These graphs are omitted as they are similar to those above and it is unlikely that $\kappa > 1$.

To allow for subpopulations where relocation rates into/out of Auckland are high, we have allowed $\kappa \times \mu$ to be as large as 1 percent. This is very high compared to the expected value for the general population where we would expect $\lambda \times \mu < 0.2\%$.

Figure 30 Relative bias for different under-coverage rates when the selection rate for PES for under-coverage is 75%

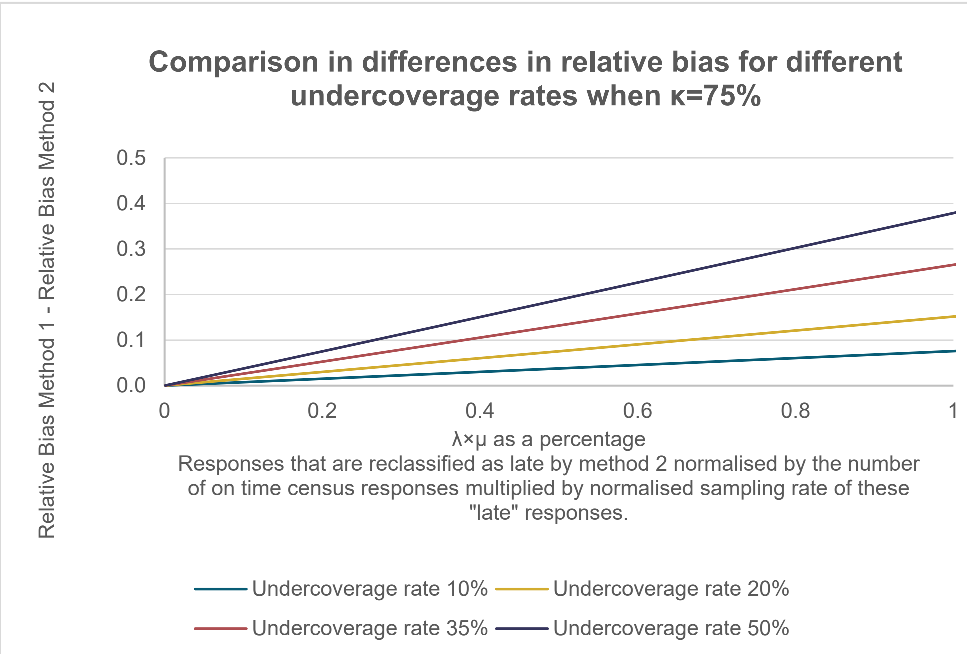


Figure 31 Relative bias for different values of under-coverage