Robust Statistical Estimation for Capture-Recapture using Administrative Data[1]

Abstract

The first key use of a nation's Census is to count its resident population. A Census will have counting errors, often referred to as over-coverage and under-coverage. So it is common practice in many countries to conduct an independent count of its residents, a so-called *coverage survey*, and estimate or adjust for these counting errors within the capture-recapture framework. In recent times, many censuses and coverage surveys have faced challenges in counting the population efficiently and effectively due to rising costs, declining response rates and respondent burden. This has led to a shift towards exploring the role that administrative registers could play in counting the population within the capture-recapture framework. Administrative registers are relatively inexpensive and can have high coverage of a nation's population. This paper explores methods to overcome common problems with the use of administrative registers within this framework, including linking errors and scoping the register to only capture residents. These methods are empirically assessed in the context of the Australian population.

Keywords: Capture-recapture, administrative registers, multi-system estimation, resident population counts

1   Introduction

1.1   Background

The first key use of a nation's population Census, and one that we focus on here, is to count its resident population. However, the count of Census responses is not equal to the count of residents (or *in-scope individuals*), according to the definition adopted by the Census. For instance, someone who is not present in the country on census night might be counted or a resident present in the country might not be counted. Over-coverage error in the count occurs if an out-of-scope individual is counted, i.e., *erroneous enumeration*, or if an in-scope individual is counted more than once (e.g., two Census forms may be submitted for the same person). Under-coverage occurs if an in-scope individual is not counted (e.g., non-response).

It is common practice in many countries to conduct an independent count of residents, a so-called *coverage survey*, and estimate or adjust the census under-count within the capture-recapture framework. (A separate sample survey of the census returns may be deployed to deal with the erroneous enumeration error.) In the original development of capture-recapture methods in application to wildlife population measurement (see Seber 1986), animals were captured, marked and recaptured resulting in two incomplete lists of the population: one list is used to estimate the capture rate of

---

the other list and the estimate of the population count is equal to the count of a list divided by its estimated capture rate.

Estimation of the unknown population size relies on a set of intertwined assumptions so that a failure of any one can invalidate the others leading to biased estimates (International Working Group for Disease Monitoring and Forecasting 1995; Zhang 2019a). These assumptions are:

- no change in the population between captures (i.e., the population is closed, or there is no in- or out-migration)

- individuals can be matched from capture to recapture (without error)

- homogeneity of capture or recapture (i.e., on each sampling occasion all individuals have the same capture probability)

- independence between the capture and recapture processes.

While it was recognised from the earliest papers (e.g., Sekar and Deming, 1949) that, in applications to human populations, the failure of the distributional assumptions about the capture-recapture events leads to incorrect population estimates, it was not until later that it was termed 'correlation bias' (Alho et al. 1999; Brown et al. 2006). They also noted that this bias can be as a result of two types of dependence

- List dependence: the act of being included in the first list makes an individual more or less likely to be included in the second list, i.e. inclusion in the first sample has a causal effect on inclusion in the second sample. This is sometimes referred to as *causal dependence* (Wolter, 1986; Brown et al. 2006).

- Heterogeneity: even if the two lists are independent, the lists may become dependent if the capture probabilities are not the same (i.e. not homogenous, or are heterogenous) amongst individuals. This is sometimes referred to as *apparent (or autonomous) dependence* (e.g., Wolter, 1986; Coull and Agresti 1999).

In official statistics, capture-recapture methods are commonly referred to as Multi-System Estimation (MuSE). In census applications there are often only two captures or lists (Dual System Estimation, DSE), namely the Census and coverage survey. For country specific studies, see e.g. Thomas (2008) for the USA, Brown et al. (2011) for the UK, Chipperfield et al. (2017) for Australia, Statistics Canada (2019) and Statistics New Zealand (2019, 2020). Use of three (Triple System Estimation, or TSE) or more lists to some degree allows relaxation of the list dependence assumption. See Nirel and Glickman (2009) for an overview and see Zaslavsky and Wolfgang (1993), Darroch et al. (1993), Baffour (2009), Baffour et al (2013) and Griffin (2014) and Fienberg (1972), for more specific details about the population estimation.

In recent times, censuses have faced challenges in counting the population efficiently and effectively, due to its cost, declining response rates, respondent burden, and its ability to correctly measure complex living arrangements. This has led to a shift towards reliance on administrative registers: by 2010 there was a clear shift towards register-based censuses in countries such as Switzerland,

Netherlands, Belgium and Slovenia (Skinner 2018). For other countries, engaged in their respective census transformation programmes, this led to considering the possibility of replacing the traditional coverage survey or Census with administrative registers. For example, the Central Bureau of Statistics of Israel conducted a Register Survey for its current round of population census (see also Bernardini, et al, 2022; Zhang, 2022), Office for National Statistics (2013, 2017) and ISTAT (Bernardini et al 2021) discuss replacing a Census with administrative registers. Natural disasters (e.g. Covid-2019) have led to the delay of censuses in many nations. In contrast, other countries (e.g. Scandanavian) have long-standing practices of register-based censuses and so are not under the same contemporary pressures.

In the context of official statistical agencies using registers (Zhang, 2019a, 2022) for MuSE, two preconditions are critical:

- lists contain no erroneous enumeration or duplicates,

- no errors occur in the *linkage* of the lists

These errors are far from trivial in official statistics if one lacks a centralised population register and its unique person identification number (such as the case in the Nordic countries) which can be used to link all the lists. These are discussed in more detail in the context of this present study below.

Though not our perspective here, the on-going development of completely register-based population size estimation methods should also be mentioned. Statistics New Zealand (2021) counts the number of administrative records that have a level of activity that is consistent with a resident, though it is experimental at this stage. Statistics Estonia (Tiit and Maasing, 2016) and Central Bureau of Statistics of Latvia (2019) derive residency scores for an Extended Population Register. Van der Heijden et al (2021) use 5 registers to estimate the Māori population in New Zealand. Central Statistical Office of Ireland (Dunne and Zhang, 2023) develop a system of population estimates compiled on administrative data only.

1.2   Outline of Study

In the context of Australia, since 1971, the resident population has been estimated by a DSE, using the Post-Enumeration Survey (PES) and the Australian Census of Population and Housing. Much has changed in the data landscape in recent years, most notably the creation of Australia's "Person Linkage Spine", or "Spine", that is the union of three administrative registers from 2006 to present: Medicare (Services Australia), Centrelink (Department of Social Services) and Personal Income Tax (Australian Taxation Office).

This study assesses the role that the Spine could play, within the MuSE framework, to estimate the resident population over the next 5 to 10 years. This assessment is complicated by the fact that the MuSE preconditions mentioned above are not easily satisfied. Erroneous enumeration arises since there is no error-free indicator on the Spine that a record belongs to a resident, however defined.

There are also two problems with linkage at the time of writing this paper:

i. lack of Spine-PES linkage,

ii. false negative of Census-Spine linkage (i.e., missing Census-Spine matches).

Another important issue in our study is when a subpopulation classification variable for Indigenous status is not observed for all records on the Spine. Instead a predicted status is used in MuSEs.

We develop alternative DSE and TSE, where Spine erroneous enumeration is dealt with by trimming (Zhang and Dunne, 2017; Dunne and Zhang, 2023), and we handle the linkage problems by False-Negative adjustments in estimation. Notice that, while the details differ, Statistics New Zealand (2019) applied a DSE in the presence of similar problems of Spine over-coverage and false-negative Census-Spine linkages, which was an inspiration to our work. We refer to the methods developed and implemented in this paper as a practical approach to MuSE, in the presence of erroneous enumeration *and* linkage error, noting that there have been some recent developments in the modelling approach to deal with erroneous enumeration and undercounting jointly; see e.g. Di Cecco et al (2018), Zhang (2015, 2019b) and Ballerini (2021). The practical approach to MuSE developed in this paper may be relevant, wherever the investigation is complicated by the failure of preconditions.

The rest of the paper is organised as follows. In Section 2 we set out the basics of MuSE and, in particular, the notations for DSE and TSE, where the preconditions are assumed to be satisfied. In Section 3, we detail the methods for dealing with the Spine erroneous enumeration and linkage problems, as well as the estimation of subpopulation. A case study applying these methods in combination with the Australian Census and PES in 2016 is presented in Section 4. Some final remarks and future research topics are given in Section 5.

2   Basics of MuSE, DSE and TSE

We consider a closed population with (unknown) N individuals. In full generality, let us define $K \geq 2$ as the number of incomplete lists taken from the population. Define $\boldsymbol{\delta}$ for an individual to be a K-vector where the kth element is equal to 1 if the individual is captured by the kth list and is equal to zero otherwise. For example, in the two-list case, $\boldsymbol{\delta} = (0, 1)$ indicates that the individual is not captured (missed) in the first list, but is captured in the second list. Similarly, for the three-capture case, $\boldsymbol{\delta} = (1, 0, 1)$ indicates that the individual is captured in the first and third list, but missed in the second. Individuals with capture patterns $\boldsymbol{\delta} = (0, 0)$ and $\boldsymbol{\delta} = (0, 0, 0)$ are by definition unobserved, for the two-list and the three-list cases, respectively.

The target parameter of MuSEs is N. We can estimate this by modelling the probability distribution of the observed capture patterns, $f(\boldsymbol{\delta})$. The probability distribution of the unobserved pattern $f(\mathbf{0})$, where $\mathbf{0} = (0, 0, ..., 0, ..., 0)$, is assumed to be a function of the observed capture pattern. Noting that the observed counts of $\boldsymbol{\delta}$ can be written in the form of an incomplete contingency table, Fienberg (1972) outlines log-linear modelling for the estimation of the unknown population size, where the counts $n_{\boldsymbol{\delta}}$ follow the multinomial distribution. The log-linear modelling framework for capture-recapature data is intuitively appealing since it allows for explicit considerations of de-

pendence among lists and heterogeneity of capture. Moreover, given any model, either a closed form estimator exists for N or it can be obtained through iterative techniques; see Chapter 6 of Bishop et al. (1975).

Under this framework, the goodness of fit of alternative models can be formally tested, although a good fit does not guarantee the validity of a given model. Typically in MuSE, the best model is taken to be the one with the fewest possible parameters that allows for some dependency amongst the lists. This model is then used to predict the missing cell count, and subsequently estimate the population size. Note that the motivation for selecting the most parsimonious model is to reduce the variance of the estimator of N, since usually the simpler the model, the smaller the variance.

Below we describe some necessary details of the DSE and TSE for this study.

2.1 DSE

Here there are two incomplete lists of units that cover to the target population. For argument's sake call one the S-list and the other the C-list. Let $s = 1$ if a population unit is on the S-list and $s = 0$ otherwise. Similarly, define c for the C-list. All units in the population belong to one of four cells (c, s) for $c, s = 0, 1$. Let the number of units in the (c, s)-cell be n_{cs}. The target population size can be expressed as $N = n_{11} + n_{10} + n_{01} + n_{00}$. The count n_{00} is not observable from the two lists and the aim is to calculate its estimate, $\hat{n}_{00}$, using the three observed cells and to estimate N by $n_{11} + n_{10} + n_{01} + \hat{n}_{00}$. The notation of cell counts and other totals are summarised in Table 1.

Table 1: Two-list counts, n_{00} unobservable, $N = n_{++}$

	S-List Captures		
C-List Captures	$s = 0$	$s = 1$	$s = +$
$c = 0$	n_{00}	n_{01}	n_{0+}
$c = 1$	n_{10}	n_{11}	n_{1+}
$c = +$	n_{+0}	n_{+1}	n_{++}

Now, provided the first precondition holds, the population can be regarded as *closed* since each record on a list belongs to one and only one unit in the target population. That is, there are no units on a list that are outside the target population. Provided the second precondition holds, the count n_{11} can be observed without errors, whatever the marginal list-counts n_{1+} and n_{+1}, so that n_{01} and n_{10} are also observed without errors. Then, the two distributional assumptions of causal and autonomous independence together imply that, for any $i \in U$, where U is the population we wish to count,

$$\Pr(\delta_{i1}\delta_{i2} = 1 \mid i \in U) = \Pr(\delta_{i1} = 1 \mid i \in U) \Pr(\delta_{i2} = 1 \mid i \in U) := \pi_1 \pi_2.$$

If we let $\mu_{cs} = E(n_{cs})$, it follows that

$$\frac{\mu_{00}\mu_{11}}{\mu_{01}\mu_{10}} := \frac{E(n_{00})E(n_{11})}{E(n_{01})E(n_{10})} = 1$$

and an estimator of μ_{00} is $\hat{\mu}_{00} = n_{01}n_{10}/n_{11}$. An estimator of N follows as $\hat{N} = n_{1+}n_{+1}/n_{11}$ with

the variance estimator $\hat{V}(\hat{N}) = n_{+1}n_{1+}n_{01}n_{10}/n_{11}^3$ which can be derived using the *Delta* method (see e.g. p.162, Baffour, 2009).

As noted by Zhang (2019a), the same $\hat{N}$ and $\hat{V}(\hat{N})$ follow by the method of moment, conditioning on one of the lists, say, $\{\delta_{i2} : i \in U\}$ and assuming that the other list has a constant capture probability (regardless of δ_{i2}), i.e.

$$\pi := \Pr(\delta_{i1} = 1 \mid i \in U) = \Pr(\delta_{i1} = 1 \mid i \in U, \delta_{i2})$$

such that

$$E(n_{1+})/N = \pi = E(n_{11} \mid n_{+1})/n_{+1}.$$

2.2   TSE

Now we bring in a third list, called the P-list. Let $p = 1$ if a population unit is on the P-list and $p = 0$ otherwise. All the population units belong to one of the 8 cells (c, p, s) for $c, p, s = 0, 1$. Let the number of units in the (c, p, s)-cell by n_{cps} (as in Table 2). The population size can be expressed as $N = \sum_{c,p,s} n_{cps}$, where n_{000} is unobservable and so the TSE aims to estimate it using the seven observed cell counts.

Table 2: Three-list counts, n_{000} unobservable

		$c = 0$				$c = 1$		
		$s = 0$	$s = 1$	$s = +$		$s = 0$	$s = 1$	$s = +$
$p = 0$		n_{000}	n_{001}	n_{00+}	$p = 0$	n_{100}	n_{101}	n_{10+}
$p = 1$		n_{010}	n_{011}	n_{01+}	$p = 1$	n_{110}	n_{111}	n_{11+}
$p = +$		n_{0+0}	n_{0+1}	n_{0++}	$p = +$	n_{1+0}	n_{1+1}	n_{1++}

Let $\mu_{cps} = E(n_{cps})$ be the expected number of individuals in the (c, p, s) cell of the $2 \times 2 \times 2$ contingency table, then the ('saturated') log-linear model can be specified as

$$\log \mu_{cps} = \lambda + \lambda_c^{(1)} + \lambda_p^{(2)} + \lambda_s^{(3)} + \lambda_{cp}^{(12)} + \lambda_{cs}^{(13)} + \lambda_{ps}^{(23)} + \lambda_{cps}^{(123)}, \tag{1}$$

where $\lambda_c^{(1)}$, $\lambda_p^{(2)}$, $\lambda_s^{(3)}$ are the main effect terms, $\lambda_{cp}^{(12)}$, $\lambda_{cs}^{(13)}$, $\lambda_{ps}^{(23)}$ are the two-way interaction terms, and $\lambda_{cps}^{(123)}$ is the three-way interaction term.

When we have an incomplete $2 \times 2 \times 2$ contingency table, with μ_{000} for the unobserved ('missing') cell, the saturated model is not identifiable in that we have eight parameters but seven observable cell counts. The implication of considering only hierarchical models, is that the highest order interaction, the three-way term $\lambda_{cps}^{(123)}$, is set to zero, and in practice the 'saturated model' becomes

$$\log \mu_{cps} = \lambda + \lambda_c^{(1)} + \lambda_p^{(2)} + \lambda_s^{(3)} + \lambda_{cp}^{(12)} + \lambda_{cs}^{(13)} + \lambda_{ps}^{(23)}. \tag{2}$$

Given $\lambda_{cps}^{(123)} = 0$, each of the two-factor effects (i.e. $\lambda_{cp}^{(12)}$, $\lambda_{cs}^{(13)}$, and $\lambda_{ps}^{(23)}$) is unaffected by the level of the third variable, such that

$$\frac{\hat{\mu}_{001}\hat{\mu}_{010}\hat{\mu}_{100}\hat{\mu}_{111}}{\hat{\mu}_{000}\hat{\mu}_{011}\hat{\mu}_{101}\hat{\mu}_{110}} = 1 \tag{3}$$

6

where $\hat{\mu}_{cps} = n_{cps}$ (e.g., Bartlett 1935; Fienberg 1972; Baffour et al. 2021), from which $\hat{\mu}_{000}$ follows. It can be seen that the effect of the no-three-way interaction assumption for the TSE is analogous to the independence assumption for the DSE.

Moreover, it now becomes possible to define various unsaturated hierarchical models by setting some of the rest λ-terms to be equal to zero. The restriction for all models under consideration to be hierarchical implies that when a particular λ-term is set to zero then all of the higher-order relatives are also zero. Closed form solutions, and their variances, exist for all models, apart from when all three lists are independent for which case the iterative proportional fitting algorithm can be used. We refer to Baffour (2009) for the details including the variance estimators.

3 Methods

For this study MuSE preconditions involving the Spine are initially unsatisfied. Below we describe how to deal with this, as well as the models for the TSE without the Spine-PES linkage.

3.1 Trimming the Spine

The PES and Census collect information on whether a person is a resident, so as to exclude non-residents from official resident counts. However, the Spine does not have residency status for all the records and so will include nonresidents. For example, a person may have emigrated from Australia a year before Census but, for various administrative reasons, their migration record was not linked to the Spine to reflect this. An easy way to trim the erroneous records on the Spine is to apply common sense rules (e.g., must have had at least 2 signs of administrative activity in the last 6 months). We apply a more refined scoring approach described below.

Let $r = 1$ if a Spine record belongs to a resident and $r = 0$ otherwise. Ideally, we would trim the records if $r = 0$. But r is unknown. Define $u = E(r|A)$ to be the probability that a record on the Spine belongs to a resident conditional on A, which is a vector of covariates that are related to resident status. Here A includes covariates for the recency of interactions with the Social Security, Medicare and Personal Income Tax systems (e.g., was there an interaction within 1, 6, or 12 months on the Medicare system), death status, migration status, age, sex and state.

As r is not observed we cannot fit a model to r to get $\hat{u}$, which is the estimate of u for each Spine record. However, we can approximate it by r^*, where $r^* = 1$ if a Spine record is linked to a Census record and $r^* = 0$ otherwise, and fit the model to r^* conditional on A to get $\hat{u}^*$ as an estimate of $u^* = E(r^*|A)$. Because of linkage error and because Census is not a complete list, $u^* \neq u$ generally. Nevertheless, we do not need unbiased predictions in order to trim the records with the largest u. Trimming is useful in reducing bias of MuSEs as long as $\hat{u}^*$ is reasonably correlated with $\hat{u}$.

We scale $\hat{u}^*$, where the scaling factor varies by age, sex and state, such that the scaled $\hat{u}^*$s sum to the projected resident population estimates from the previous Census. The scaling is simply for convenience so that we may use a single threshold, ϵ, for trimming across all age, sex and states.

Having prepared the scores $\hat{u}^*$, we then trim records from the Spine if

$$\hat{u}^* < 1 - \epsilon$$

where ϵ is a chosen positive number close to 0. For the applications in Section 4, trimming was applied with $\epsilon = 0.001$. This strict cutoff (i.e., $1 - \epsilon = 99.9\%$) means that only a negligible number of nonresidents would remain after trimming.

Notice that by using a model estimated from the Spine-Census linkage, as described above, one is introducing a dependence between the trimmed Spine and the Census via the model parameters. That is, for any target population unit, its inclusion indicators in the two lists will no longer be exactly independent, assuming they were independent without trimming. However, by and large this dependence can be ignored since the trimming induced covariance of the two indicators is of the order $O(p/m)$, asymptotically as m tends to infinity while p is held fixed, where p is the number of model parameters and m is the number of linked Spine-Census records.

We shall refer to the trimmed version of the Spine as *SpineSure* when describing the estimates to be evaluated in Section 4. Whereas, the term Spine is used in the rest of this Section for simplicity, where by stipulation all the lists are free of erroneous enumeration, such as when next describing the models for 3-list capture-recapture data.

3.2   TSE models without Spine-PES linkage

Let $c = 1, 0$ denote whether a resident is captured by the Census or not; similarly, let $p = 1, 0$ for the PES and $s = 1, 0$ for the Spine. Consider modelling (c, p, s) in the presence of linkage problem (i) – Section 1.2, assuming the *other* preconditions are satisfied. Adjustment for the linkage problem (ii) is given in Section 3.3 next.

Given the Census-Spine linkage and the Census-PES linkage, but not the linkage between Spine and PES, the counts n_{cps} of the cross-classified list domains and the relationships between them are given in Table 3, similarly to Table 2, where the counts in parentheses are unobserved in addition to n_{000} that is missing in any case.

Table 3: Cross-classified counts without Spine-PES linkage, all counts in parentheses are unobserved

	$c = 0$				$c = 1$		
	$s = 0$	$s = 1$	$s = +$		$s = 0$	$s = 1$	$s = +$
$p = 0$	(n_{000})	(n_{001})	(n_{00+})	$p = 0$	n_{100}	n_{101}	n_{10+}
$p = 1$	(n_{010})	(n_{011})	n_{01+}	$p = 1$	n_{110}	n_{111}	n_{11+}
$p = +$	(n_{0+0})	n_{0+1}	(n_{0++})	$p = +$	n_{1+0}	n_{1+1}	n_{1++}

Let $q_{cps} = E(n_{cps})$ to emphasise the distinction to μ_{cps} in the usual setting. Denote by $[CP][CS][PS]$ the basic model (without 3-way interaction), under which we have

$$q_{000} = \left(\frac{q_{111}q_{100}}{q_{110}q_{101}}\right)\frac{q_{010}q_{001}}{q_{011}} = \left(\frac{q_{111}q_{010}}{q_{110}q_{011}}\right)\frac{q_{100}q_{001}}{q_{101}} = \left(\frac{q_{111}q_{001}}{q_{101}q_{011}}\right)\frac{q_{100}q_{010}}{q_{110}} \tag{4}$$

Note that each term in the parentheses above corresponds to an odds ratio conditional on $c = 1$, $p = 1$ or $s = 1$, respectively, denoted by ρ_1^C, ρ_1^P or ρ_1^S. In other words, the equation (4) can equally be obtained from one of the following assumptions:

$$\rho_1^C = \rho_0^C \qquad \text{or} \qquad \rho_1^P = \rho_0^P \qquad \text{or} \qquad \rho_1^S = \rho_0^S$$

Now that this basic model cannot be identified given the linkage problem (i), we shall consider the models given by any of the independence assumptions below:

$$[C][PS] \qquad \text{or} \qquad [P][CS] \qquad \text{or} \qquad [S][CP]$$

Notice that any of these is a stronger assumption than the corresponding constant-odds-ratio assumption, such that the equation (4) still holds under any of them. Once the missing cell counts $(n_{001}, n_{010}, n_{011})$ due to the lack of complete three-way linkage have been estimated, one can plug them into (4) to estimate n_{000} and N.

Model $[C][PS]$ Under this model, we have

$$\Pr(s = 1 | p = 1, c = 0) = \frac{q_{011}}{q_{01+}} = \frac{q_{111}}{q_{11+}} = \Pr(s = 1 | p = 1, c = 1)$$

as well as

$$\Pr(p = 1 | s = 1, c = 0) = \frac{q_{011}}{q_{0+1}} = \frac{q_{111}}{q_{1+1}} = \Pr(p = 1 | s = 1, c = 1).$$

Either of them yields an estimator of n_{011} directly. One can also consider a combination of the two estimators. However, since the Census is much larger than the PES in size, we would simply use the estimator

$$\hat{n}_{011} = n_{111} n_{01+} / n_{11+}$$

given which we obtain $\hat{n}_{010} = n_{01+} - \hat{n}_{011}$ and $\hat{n}_{001} = n_{0+1} - \hat{n}_{011}$.

Model $[P][CS]$ This model allows the Census and Spine captures to be correlated, which is an added value compared to the Census-PES DSE otherwise. We have

$$\frac{q_{c0s}}{q_{+0+}} = \frac{q_{c1s}}{q_{+1+}}$$

such that

$$\frac{q_{011}}{q_{+1+}} = \frac{q_{001}}{q_{+0+}} = \frac{q_{0+1} - q_{011}}{q_{+0+}} \qquad \Leftrightarrow \qquad \frac{q_{+0+}}{q_{+1+}} = \frac{q_{0+1} - q_{011}}{q_{011}}$$

as well as

$$q_{+0+} = q_{10+} + q_{00+} = q_{10+} + q_{001} + q_{000} = q_{10+} + (q_{0+1} - q_{011}) + \frac{q_{010}}{q_{+1+}}q_{+0+}$$

$$= (q_{10+} + q_{0+1} - q_{011}) + \frac{q_{+0+}}{q_{+1+}}(q_{01+} - q_{011})$$

$$\Leftrightarrow \quad \frac{q_{+0+}}{q_{+1+}} = \frac{q_{10+} + q_{0+1} - q_{011}}{(q_{+1+} - q_{01+}) + q_{011}} = \frac{q_{10+} + q_{0+1} - q_{011}}{q_{11+} + q_{011}}.$$

Thus, we obtain

$$\frac{q_{10+} + q_{0+1} - q_{011}}{q_{11+} + q_{011}} = \frac{q_{0+1} - q_{011}}{q_{011}}$$

$$\Leftrightarrow \quad (q_{10+} + q_{0+1})q_{011} - q_{011}^2 = q_{0+1}q_{11+} + (q_{0+1} - q_{11+})q_{011} - q_{011}^2$$

$$\Leftrightarrow \quad (q_{10+} + q_{11+})q_{011} = q_{0+1}q_{11+} \quad \Leftrightarrow \quad q_{1++}q_{011} = q_{0+1}q_{11+}.$$

An estimator of n_{011} follows as

$$\hat{n}_{011} = n_{0+1}n_{11+}/n_{1++}$$

given which we obtain $\hat{n}_{010} = n_{01+} - \hat{n}_{011}$ and $\hat{n}_{001} = n_{0+1} - \hat{n}_{011}$.

Model $[S][CP]$ Under this model, we have

$$\frac{q_{cp1}}{q_{++1}} = \frac{q_{cp0}}{q_{++0}}$$

such that

$$\frac{q_{011}}{q_{++1}} = \frac{q_{010}}{q_{++0}} = \frac{q_{01+} - q_{011}}{q_{++0}} \quad \Leftrightarrow \quad q_{++0}q_{011} = q_{++1}q_{01+} - q_{++1}q_{011}.$$

Moreover, we have (4) since $\rho_1^S = \rho_0^S$. Rewriting (4) as $q_{000} = \rho_1^C q_{010}q_{001}/q_{011}$, where $\hat{\rho}_1^C = (n_{111}n_{100})/(n_{110}n_{101})$ is available, we have

$$q_{++0} = q_{1+0} + q_{0+0} = q_{1+0} + q_{010} + q_{000}$$

$$= q_{1+0} + (q_{01+} - q_{011}) + \rho_1^C(q_{01+} - q_{011})(q_{0+1} - q_{011})/q_{011}$$

$$\Leftrightarrow \quad q_{++0}q_{011} = (q_{1+0} + q_{01+})q_{011} - q_{011}^2 + \rho_1^C\big(q_{01+}q_{0+1} - (q_{01+} + q_{0+1})q_{011} + q_{011}^2\big).$$

Thus, we obtain

$$a_2 q_{011}^2 + a_1 q_{011} + a_0 = 0$$

where

$$\begin{cases} a_2 = 1 - \rho_1^C \\[2mm] a_1 = -(q_{++1} + q_{1+0} + q_{01+}) + \rho_1^C(q_{01+} + q_{0+1}) \\[2mm] a_0 = q_{++1}q_{01+} - \rho_1^C q_{01+}q_{0+1}. \end{cases}$$

An estimator of n_{011} follows, provided the equation of q_{011} has at least one positive root, on replacing the q-totals by their observed values and ρ_1^C by $\hat{\rho}_1^C$.

Goodness-of-fit Given each possible model above, one can check the goodness-of-fit against the corresponding assumption. For instance, given the model $[C][PS]$, we have

$$\hat{n}_{+ps} = n_{1ps} + \hat{n}_{0ps}$$

for $p, s = 0, 1$. This yields 8 cell-specific discrepancies depending on $c = 0$ or 1, i.e.

$$\hat{n}_{cps}/\hat{n}_{c++} - \hat{n}_{+ps}/\hat{n}_{+++}$$

where some of the estimates are observed directly. One can obtain the Pearson or Kullback-Leibler divergence measure to choose between model specifications. Similarly for the other two models. However, model selection cannot be entirely based on the principle of parsimony, although it is justified to disregard those models that fit badly to the observed data.

Finally, given the missing cell counts caused by the lack of complete three-way linkage, it is unclear to us whether any standard software packages for fitting log-linear models can be used to calculate the non-standard estimates discussed above or if the associated variance estimation procedure can account for the missing information (a referee mentioned the R-package CAT for the 'Analysis and Imputation of Categorical-Variable Datasets with Missing Values'). In any case the estimators can easily be computed by writing one's own code. For an estimator that has a closed form, the Delta method for variance estimation is quite straightforward. Otherwise, or generally, bootstrap under the estimated multinomial distribution would seem most convenient in practice.

3.3 Adjusting for false negative of Census-Spine linkage

The linkage problem (ii) is such that about 5% of Census records were False Negatives (FN) in its linkage to the Spine. Here we discuss how this was estimates and handled in MuSE.

Let $l = 1, 0$ indicate whether a Census record is linked to its matching Spine record. Let z be a vector of indicator variables for whether or not the linking variables (name, age, date-of-birth, address) are missing on the Census file. Let y be age and sex. Let $a = 1, 0$ indicate whether a person is born in Australia. For any resident i in the Census list, we assume

$$w_i^{-1} = E_c(l_i \mid z_i, y_i) = E_c(l_i \mid z_i, y_i, a_i = 1) \tag{5}$$

with respect to the linkage error distribution. That is, the false-negative probability, conditional on (z, y), does not depend upon whether a resident is Australian born. (By definition w_i is the weight that is used below to correct for the effects of false negatives.) If we also assume that all Australian born people, denoted by $a = 1$, have a Spine record (assuming that a birth would, with certainty, generate a Medicare record that will appear on the Spine), then (5) is identifiable. We found that the false-negative rates are higher for Aboriginal and Torres Strait Islander Peoples and people living in remote or rural areas, which is plausible.

Regarding the model (5), we notice that it is reasonable to assume all the native born persons $(a = 1)$ have a Spine record provided a country (like Australia) has universal birth registration, school and health services. We also notice that although there may be other factors for false negative linkage beyond (z, y), there is a reason why these are not actually used as the key variables for record linkage, such as non-negligible missing-ness or measurement errors, which would need to be handled if such variables were included as additional covariates for l.

Let n_{11s} be the number of links between Census and Spine, where $s = 1, 0$. We estimate the corresponding number of Census-Spine matches, denoted by m_{11s}, based on

$$E_c(n_{11s}) = m_{11s}\bar{w}_{1+s}$$

where $\bar{w}_{1+s}$ is the mean of w_i among the Census records that are linked to the PES if $s = 1$, or those unlinked if $s = 0$. This yields the estimated match counts, denoted by $\tilde{n}_{111}$, $\tilde{n}_{110}$ and $\tilde{n}_{11+}$, and then $\tilde{n}_{101} = n_{1+1} - \tilde{n}_{111}$ and $\tilde{n}_{100} = n_{1+0} - \tilde{n}_{110}$. Together with the other observed counts, these estimated counts can be plugged into the TSE models (Section 3.2) and a relevant DSE involving the Spine.

3.4   Subpopulation size estimation

Let $X = 1, 0$ be a subpopulation classifier. Denote the subpopulation size of $X = x$ by $N_x = n_{11x} + n_{10x} + n_{01x} + n_{00x}$ in the two-list case, with the corresponding subpopulation cell counts. The DSE of n_{00x} is $\hat{n}_{00x} = n_{10x}n_{01x}/n_{11x}$ and that of N_x is $\hat{N}_x = n_{1+x}n_{+1x}/n_{11x}$, provided the DSE assumptions hold conditional on $X = 1, 0$.

Now, suppose X is unavailable in the second list, so that n_{1+x} is directly observed but not n_{+1x}. Instead, let X^* be a predicted value of X which is available. Let n^*_{+1x} be predicted subpopulation count on the second list according to X^*, where $n^*_{+1x} = n^*_{01x} + n^*_{11x}$. It is easy to show that, provided $E(X^* - X) = 0$ in all the 4 cells, then we can obtain a face-value DSE of N_x simply by treating $X^* = X$. That is, substituting n^*_{01x} for n_{01x} yields the estimator

$$\hat{N}^*_x = n_{11x} + n_{10x} + n^*_{01x} + \hat{n}^*_{00x} = n_{11x} + n_{10x} + n^*_{01x} + n_{10x}n^*_{01x}/n^*_{11x}.$$

Moreover, the idea extends to the TSE, where X is not available on one of the lists. Provided the relevant assumption $E(X^* - X|c, p, s) = 0$, we can replace the unobserved n_{cpsx} by n^*_{cpsx} in the TSE.

In our study, X is Indigenous status and X^* is the predicted Indigenous status on the Spine. We derive X^* by a model, where Census Indigenous status is the dependent variable and the independent variables are taken from the Spine. Here we have ignored any uncertainty in MuSEs due to the prediction of Indigenous status. Notice that Zwane and Van der Heijden (2006) and Van der Heijden et al (2021) use other methods for dealing with such missing classification variables in the context of MuSE, albeit without the other problems also present here such as false negative linkage and erroneous Spine records.

4 Case study: Australian population 2016

4.1 Set-up

The Australian Census counts people who are present on Census night, are referred to as the *Present Population*. For our purpose the 2016 Census counted 21.2 million residents, defined by Census as a person who lives or intends to live in Australia for at least 6 months. Shortly after the 2016 Census, the PES counted 110,000 people who were present. The PES was linked to the Census in a high-quality clerical and automatic process which is assumed to be without error. ABS strives to maintain independence of Census and PES counting processes so as to justify the independence assumption. The official estimate of the Present Population size is 23.6 million using an Instrumental Variable Regression estimator (Chipperfield et al, 2017). To this is added an administrative count of 0.6 million Residents Temporarily Overseas (RTO) to get the Final Census Night Resident Population (24.2 million). RTOs are defined by a '12/16 month rule'. For a person to have immigrated or emigrated, they must have stayed in, or were absent from, Australia for a period of 12 out of 16 months. After further minor demographic adjustments and smoothing to reduce sample errors at fine levels we arrive at the Official Estimated Resident Population (ERP).

For the purpose of this study, the Spine includes about 33 million administrative records for people who were 'ever resident' in Australia from 2006 to 2020, including people who have died and who have temporarily or permanently left Australia. After trimming the Spine (Section 3.1), there are 14.5m records on SpineSure.

Now that conceptually the Census and PES do not enumerate the RTOs, we did not remove the RTOs from SpineSure, in order to see whether MuSE involving SpineSure can potentially target the ERP directly. For example, holding the Census as fixed despite its non-coverage of the RTOs, the Census-SpineSure DSE is unbiased for the Resident Population provided SpineSure has a constant capture probability of the Resident Population, across RTOs and non-RTOs. Although the RTOs can appear on SpineSure, the probability is likely to be lower than that for the non-RTOs, given that trimming (Section 3.1) is based on an estimated probability of being counted in the Census. Also, on average the more recent immigrants are likely to have a somewhat lower probability to appear on SpineSure due to having less time to interact with government services.

Regarding the FN adjustment using the method of Section 3.3, the Census-SpineSure linkage is estimated to have about 5% FN errors overall, which vary significantly across the different states (e.g., up to 30% in Northern Territory) or other breakdowns of the population.

The results to be presented include:

- Official estimate, or simply ERP in what follows;

- DSE based on Census and PES, or simply Census-PES DSE;

- Census-SpineSure DSE similarly, with FN adjustment as in Section 3.3;

- PES-SpineLink DSE, where SpineLink contains the SpineSure records that are linked to the

Census, and the links between PES and SpineLink are identified given the links between PES and Census. No adjustment for FN is required;

- TSE under the model $[C][PS]$ (Section 3.2) with FN adjustment (Section 3.3). We do not present results for TSE under the other models in Section 3.2 because, as will become clear below, they make little differences in this study.

The following details are worth noting. First, the RTOs will be added to the Census-PES and PES-SpineLink DSE, in order to make the presented estimates comparable (i.e. include RTOs). Second, all MuSEs stratify by state, age and Indigenous status, where we use predicted Indigenous status (Section 3.4) for SpineSure. Predicted status is used for illustrative purposes here- all official estimates based on administrative status would require extensive consultation with Aboriginal and Torres Strait Islander peoples as the ABS places high importance on reporting statistics by Indigenous status in a culturally sensitive manner.

Third, if a person is counted by multiple lists and there is inconsistency in lists' post-strata, then post-strata is assigned according to the PES, Census and Spine (in that order of preference). For example, if a person's PES state and Census state are different, the person would be assigned to a strata based on the PES state. Finally, although the PES sampling design weights are not accessible for the purpose of this project, we know they depend on state and whether an area is expected to contain a high proportion of Aboriginal and Torres Strait Islander Peoples. Hence, we proceed with the PES source under the working assumption that its sampling design is ignorable when state and Indigenous Status are used in post-stratification.

Fourthly, we do not explore here in detail the stability of MuSEs with respect to the cut-off applied when trimming the Spine, although this has been exercised for the equivalent 2021 versions of the Census, Spine and PES. For example, in 2021 the trimmed Spine had 17.3 million Spine records (instead of the 14.5 million in 2016), and changing the cut-off to allow 8% more Spine records (17.3 to 18.7 million) only resulted in an increase of 0.4% in the Census-Spine DSE. This suggests that the MUSEs are stable around the chosen cut-off threshold for trimming.

4.2   National and State Estimates

Table 4 gives the different estimates at the National and State levels, in comparison with the Official ERP and its associated standard error (SE). Now that the ERP and the Census-PES DSE are based on the same sources, the difference between them quantifies the combined impact of methods, including demographic adjustments, smoothing and the tailored estimator (Chipperfield et al, 2017). Whereas the differences among the various DSEs and the TSE reflect chiefly the effect of different sources.

First, it can be seen that the TSE and the Census-SpineSure DSE are very close to each other. The main reason is that the PES is a much smaller list and contains little extra information in this setting. To illustrate, Table 5 gives the national cross-classified counts in the TSE setting, where about 0.869 million people are missed from all three lists and only 4 thousand people were captured

Table 4: National and State ERP and MuSE ($\times 10^3$) given post-stratification by 5-year age, state and Indigenous status. TSE using Census, PES and SpineSure; DSE otherwise.

Level	ERP $\pm 2\,$SE	Census-PES	Census-SpineSure	TSE	PES-SpineLink
National	24 180 $\pm$ 96	24 082	23 959	23 956	23 420
NSW	7 730 $\pm$ 62	7 566	7 679	7 676	7 387
Victoria	6 170 $\pm$ 50	6 011	6 083	6 082	5 950
Queensland	4 840 $\pm$ 45	4 802	4 799	4 804	4 690
South Australia	1 670 $\pm$ 17	1 678	1 710	1 710	1 688
Western Australia	2 560 $\pm$ 28	2 475	2 526	2 524	2 551
Tasmania	520 $\pm$ 7	520	515	515	504
Northern Territory	250 $\pm$ 7	238	237	235	244
Australian Capital Territory	400 $\pm$ 10	397	407	407	400

by the PES but missed by both the Census and SpineSure. Notice that post-stratification is only by state and single-year age for Table 5, and the estimated total population size (2.36×10^6) differs to the TSE in Table 4 by about 300 thousand. In other words, the choice of post-stratification has a much greater effect than dropping the PES and only using Census and SpineSure.

Table 5: Cross-classified counts ($\times 10^3$) of Census (c), SpineSure (p) and PES (s), either known (in italics) or estimated (given post-stratification by state and single-year age)

	$c = 0$				$c = 1$		
	$p = 0$	$p = 1$	$p = +$		$p = 0$	$p = 1$	$p = +$
$s = 0$	869	1 539	2 408	$s = 0$	8 173	12 945	*21 119*
$s = 1$	4	7	*11*	$s = 1$	36	59	*97*
$s = +$	873	1 546	2 419	$s = +$	8 210	13 005	*21 215*

Next, from Table 4, the PES-SpineLink DSE is noticeably lower (and significantly different) than the other two DSEs nationally as well as in the three largest states. We shall return to the main reason for this later, but concentrate on the other two DSEs here, which are reasonably aligned with each other. Comparing the differences between the DSEs by Census-SpineSure and Census-PES to the Standard Error (SE) of latter, we note that the national estimate and 4 of the 8 state estimates are marginally outside the 95% confidence interval. Of those 4 outside the confidence interval, 3 of the Census-SpineSure DSE estimates are closer to the ERP than Census-PES DSE.

Notice that we do not present any variance estimates for the MuSEs above, due to the complexity caused by false negative linkages, inconsistent stratification variables in different lists and the complex design of the PES. An appropriate treatment of all these effects for variance estimation is beyond this paper. However, while we do not present them here, the standard errors of the Census-SpineSure DSE or TSE when ignoring the above mentioned complexities are over 10 times smaller than those of the ERP, because the Spine is much larger than the PES sample (which largely determines the variance of ERP). Thus, more accurate variance estimation is unlikely to change the main conclusions we draw from scoping the alternative MuSEs.

The fact that nationally the Census-SpineSure DSE is lower than the Census-PES DSE (and

the ERP) suggests that the erroneous enumeration problem of the Spine can be largely removed by trimming. Although the FN adjustment we applied might not fully account for the heterogenous FN probability, we can be quite confident about the total of FN links, such that the main cause for the difference in these two national estimates is likely due to the treatment of RTOs. Whereas the RTOs are added to the Census-PES DSEs to yield the estimates presented here, no such direct addition of RTOs is performed for the Census-SpineSure DSE. However, as mentioned before, SpineSure is likely to have a lower capture probability among certain groups of individuals, such as the RTOs and the recent immigrants. It is possible SpineSure nearly fails to cover certain Residential Population groups below the post-stratum level, for which adding the RTOs outside SpineSure would be the only viable adjustment.

Table 6: National and State ERP and MuSE ($\times 10^3$) of Aboriginal and Torres Strait Islander Peoples (same setting as Table 4)

Level	ERP $\pm 2\,\mathrm{SE}$	Census-PES	Census-SpineSure	TSE
National	798 ± 40	757	751	749
NSW	266 ± 22	252	246	246
Victoria	58 ± 13	55	56	56
Queensland	221 ± 22	207	215	215
South Australia	42 ± 8	42	40	40
Western Australia	101 ± 18	96	92	91
Tasmania	28 ± 4	28	26	26
Northern Territory	75 ± 4	69	68	67
Australian Capital Territory	7.5 ± 1.5	6.5	7.4	7.3

Finally, Table 6 shows the estimates for the Aboriginal and Torres Strait Islanders population by state and at the national level, except for the PES-SpineLink DSE. The differences between the Census-SpineSure and Census-PES DSEs are well within the 95% confidence intervals derived from the ERP. In terms of absolute values, the two DSEs agree better with each other than in Table 4 although, relatively speaking, the DSEs differ more to the ERP here than in Table 4, which is not unexpected given the observed level of inconsistency between a person's PES and Census' Indigenous status, which the ERP accounts for but the DSEs here do not. Nevertheless, it is seen that using the predicted Indigenous status on SpineSure in cases where it is missing from the PES and Census does not create major problems for estimates at these levels.

4.3   State and Age Distributions

We now consider some more disaggregated results. First, by age group, Figure 1 shows MuSEs against the 95% confidence interval derived from the ERP. We note that

- The Census-PES DSE tracks the ERP equally well in all the age groups. The Census-SpineSure DSE also performs about equally well in all the age groups, perhaps with the exception of 20–35-year-olds, possibly because SpineSure 'over-trims' 20–35-year-old RTOs, as mentioned before. There is little difference between the Census-SpineSure DSE and the TSE, except in the two

oldest age groups, where the PES enumeration is relatively too small to support the three-way cross-classification.

- The PES-SpineLink DSE is biased downward as noted before. Though not presented here this downward bias disappears at the national level if state is removed from post-stratification. This suggests that disagreement between state on Spine and that observed in PES may be the most important cause for the bias.

Figure 1: By age group, MuSE against 95% confidence interval derived from ERP
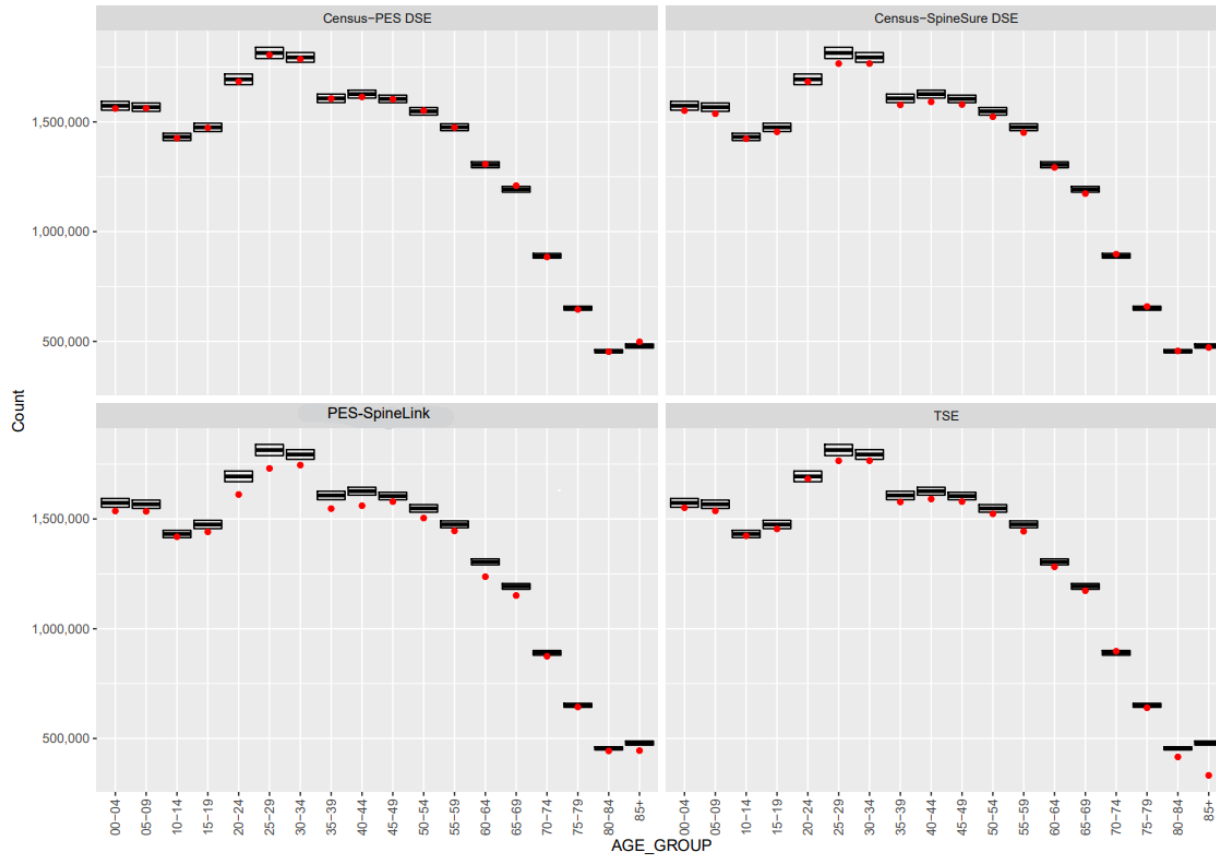
Legend:
- Multi-system Estimator
□ 95% Confidence about Official Estimated Resident Population

Next, Figure 2 gives the state by age-group breakdowns of the Census-SpineSure DSE. All the DSEs are within (or just outside) the 95% confidence interval derived from the ERP, except in NT where many fall outside the confidence intervals. The main reason is likely to be inadequate adjustment of heterogenous false negatives, the probability of which in NT is estimated to be among the highest across the country.

Finally, Figure 3 shows that the Census-SpineSure DSE of the Aboriginal and Torres Strait Islanders population is smooth and tracks the ERP well. The official ERP for the Aboriginal and Torres Strait Islanders Population is smoothed to reduce volatility due to the small PES sample size. In any case, using the predicted Indigenous status on SpineSure in cases where it is missing otherwise

does not seem to be a major statistical cause of concern at such disaggregated levels either.

5   Final remarks

Historically, many nations have used a population census as a basis to count their resident population and used a coverage survey to correct for census counting errors. Currently, many of them are planning to reduce reliance or phase out census collections given pressures to reduce costs and response burden. This paper presents findings of a study conducted by the Australian Bureau of Statistics into the role that administrative data, here the so-called *Spine*, could play to count residents in the presence or absence of a Census or PES. This study was conducted within the Multi-System Estimation framework.

Despite the lack of preconditions, our approach enables us to study the Census-Spine DSE that simulates population estimates without a coverage survey. The Census-Spine DSE achieves comparable estimates to the Census-PES DSE. Even though further research is required, this evidence may be useful in considering future directions for ABS Census transformation.

References

[1] Ballerini, V. (2021). *The Fisher's Noncentral Hypergeometric distribution and population size estimation problems.* DOI: https://arxiv.org/pdf/2210.08346.pdf (accessed March 2024).

[2] Baffour, B, Brown, J and Smith, P. 2013. "An investigation of triple system estimators in censuses", *Statistical Journal of the IAOS* 29 (1): 53-68. DOI: 10.3233/SJI-130760

[3] Bishop, Y., S. Fienberg, P. Holland, J. Richard, and F. Mosteller. 1975. *Discrete multivariate analysis: theory and practice.* MIT Press, Cambridge, Massachusetts.

[4] Baffour, B. 2009. Estimation of Population Totals from Imperfect Census, Survey and Administrative Records. *PhD Thesis*, University of Southampton. Available at: https://eprints.soton.ac.uk/72367/ (accessed March 2024)

[5] Bernard B., Brown, J.J. and Smith, P. W.F. 2021. "Latent Class Analysis for Estimating an Unknown Population Size – with Application to Censuses". *Journal of Official Statistics* 37 (3) :673-697, DOI: http://dx.doi.org/10.2478/JOS-2021-0030.

[6] Bernardini, A., Cibella, N., Gallo, G., and Al. 2021. "Empirical evidence for population counting: the combined use of administrative sources and survey data". *ESS Workshop*. Available at: https://ec.europa.eu/eurostat/cros/system/files/gerardo-gallo empirical-evidence-population-counting istat.pdf (accessed March 2024)

[7] Brown, J., Abbott, O., and Smith, P. A. 2011. "Design of the 2001 and 2011 census coverage surveys for england and wales." *Journal of the Royal Statistical Society Series A* 174 (4): 881-906. DOI: https://doi.org/10.1111/j.1467-985X.2011.00697.x

[8] Bernardini, A., Brown, J., Chipperfield, J., Bycroft, C., Chieppa, A., Cibell, N., Dunnet, G., Hawkes, M.F., Hleihel, A., Law, E.C., Ward, D., and Zhang, L.C. 2022. "Evolution of the Person Census and the Estimation of Population Counts in New Zealand, United Kingdom, Italy and Israel."*Statistical Journal of the IAOS*, 38(4):1221-1237. DOI: 10.3233/SJI-220018

[9] Chipperfield, J., Brown, J., and Bell, P. 2017. "Estimating the count error in the Australian census." *Journal of Official Statistics* 33 (1): 45-59. DOI: https://doi.org/10.1515/jos-2017-0003

[10] Darroch, J., Fienberg, S., Glonek, G., and Junker, B. 1993. "A Three-Sample Multiple- Recapture Approach to Census Population Estimation with Heterogeneous Catchability." *Journal of the American Statistical Association*, 88: 1137-1148. DOI: https://doi.org/10.2307/2290811

[11] Di Cecco, D. , Di Zio, M. , Filipponi, D. and Rocchetti, I. 2018. "Population Size Estimation Using Multiple Incomplete Lists with Over-coverage." *Journal of Official Statistics*, 34 (2): 557-572 DOI: http://dx.doi.org/10.2478/JOS-2018-0026.

[12] Dunne, J. and Zhang, L.-C. 2023. "A system of population estimates compiled from administrative data only (with discussions)." *Journal of the Royal Statistical Society, Series A* 187 (1): 3-21. DOI: https://doi.org/10.1093/jrsssa/qnad065

[13] Fienberg, S.E. (1972). "The multiple recapture census for closed populations and incomplete 2k contingency tables." *Biometrika*, 59 (3): 409-439. DOI: https://doi.org/10.2307/2334810

[14] Griffin, R.A. 2014. "Potential Uses of Administrative Records for Triple System Modeling for Estimation of Census Coverage Error in 2020". *Journal of Official Statistics* 30 (2) :177-189. DOI: https://doi.org/10.2478/jos-2014-0012

[15] Nirel, R. and Glickman, H. 2009. "Sample surveys and censuses", in *Sample Surveys: Design, Methods and Applications*, Vol 29A (eds D. Pfeffermann and C.R. Rao), Chapter 21, North Holland, pp. 539-565.

[16] Office of National Statistics. 2013. *Beyond 2011: Producing Population estimates Using Administrative Data: In Theory.* Available at: http://webarchive.nationalarchives.gov.uk/20150109083021/http:/www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011–producing-population-estimates-using-administrative-data–in-theory–m8-.pdf . (accessed March 2024)

[17] Office of National Statistics (2017). *Research Outputs: Coverage-adjusted administrative data population estimates for England and Wales, 2011.* Available at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/methodology/researchoutputscoverageadjustedadministrativedatapopulationestimatesforenglandandwales2011 (accessed March 2024)

[18] Sekar, C. C., and Deming, W. E. 1949. "On a Method of Estimating Birth and Death Rates and the Extent of Registration" *Journal of the American Statistical Association* 44: 101-115. DOI: https://doi.org/10.2307/2280353

[19] Skinner, C.J 2018. "Issues and Challenges in Census Taking". *Annual Review of Statistics and Its Application* 5(1): 49-63. DOI: https://doi.org/10.1146/annurev-statistics-041715-033713

[20] Statistics Canada. 2019. *Coverage Technical Report.* Available at: https://www12.statcan.gc.ca/census-recensement/2016/ref/98-303/98-303-x2016001-eng.pdf (accessed March 2024).

[21] Statistics New Zealand 2019. *Dual system estimation combining census responses and an admin population.* Available at: https://www.stats.govt.nz/assets/Uploads/Methods/Dual-system-estimation-combining-census-responses-and-an-admin-population/dual-system-estimation-combining-census-responses-and-an-admin-population.pdf (accessed March 2024)

[22] Statistics New Zealand 2020. *Post-enumeration Survey 2018: Methods and Results.* Available at: https://www.stats.govt.nz/assets/Uploads/Reports/Coverage-in-the-2018-Census-based-on-the-New-Zealand-2018-Post-enumeration-Survey/Downloads/Post-enumeration-survey-2018-Methods-and-results-Stats-NZ.pdf (accessed March 2024)

[23] Statistics New Zealand 2021. *Experimental administrative population census.* Available at: https://www.stats.govt.nz/experimental/experimental-administrative-population-census (accessed March 2024)

[24] Thomas, M. 2008. Census Coverage Measurement, *Memorandum Series 2010-E-18, US Census Bureau.* Available at: https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/ccm-workshop/2010-e-18.pdf (accessed March 2024)

[25] Van der Heijden, P.G.M., Cruyff M., Smith P. A., Bycroft C., Graham P., Matheson-Dunning, N. 2021. "Multiple system estimation using covariates having missing values and measurement error: estimating the size of the Māori population in New Zealand". *Journal of the Royal Statistics Association Series A* 185 (1): 156-177. DOI: https://doi.org/10.1111/rssa.12731

[26] Wolter, K. 1986. "Some Coverage Error Models for Census Data". *Journal of the American Statistical Association* 81(394): 338-346. DOI: https://doi.org/10.2307/2289222

[27] Zaslavsky, A., and Wolfgang, G. 1993. "Triple-System Modelling of Census, Post-Enumeration Survey, and Administrative-List Data". *Journal of Business and Economic Statistics* 11(3): 279-288. DOI:https://doi.org/10.2307/1391952

[28] Zhang, L.C. (2022). "Complementarities of Survey and Population Registers". *Statistics Reference Online*, Wiley. https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat08352

[29] Zhang L. and Dunne J. 2018. "Trimmed dual system estimation" in Bohning, D. (Ed.), van der Heijden, P. (Ed.), Bunge, J. (Ed.). (2018) *Capture-Recapture Methods for the Social and Medical Sciences.* New York: Chapman and Hall/CRC

[30] Zhang, L.-C. 2015. "On modelling register coverage errors". *Journal of Official Statistics* 31 (3) 381-396. DOI: https://doi.org/10.1515/jos-2015-0023

[31] Zhang, L.-C. and Dunne, J. 2017. "Trimmed Dual System Estimation". In *Capture-Recapture Methods for the Social and Medical Sciences*, eds. D. Böhning, J. Bunge and P. v. d. Heijden, Chapter 17, pp. 239-259. Chapman & Hall/CRC.

[32] Zwane, E. N. and van der Heijden, P. G. M. 2015. "Analysing capture–recapture data when some variables of heterogeneous catchability are not collected or asked in all registrations" . *Statistics in Medicine* 26(5): 1069-1089. DOI: https://doi.org/10.1002/sim.2577

[33] Zhang, L.-C. 2019a. "A note on dual system population size estimator". *Journal of Official Statistics*, 35 (1): 279-283.

[34] Zhang, L.-C. 2019b. "Log-linear models of erroneous list data". In *Analysis of Integrated Data*, eds. L.-C. Zhang and R.L. Chambers. Chapter 9, pp. 197-218. Chapman & Hall/CRC.

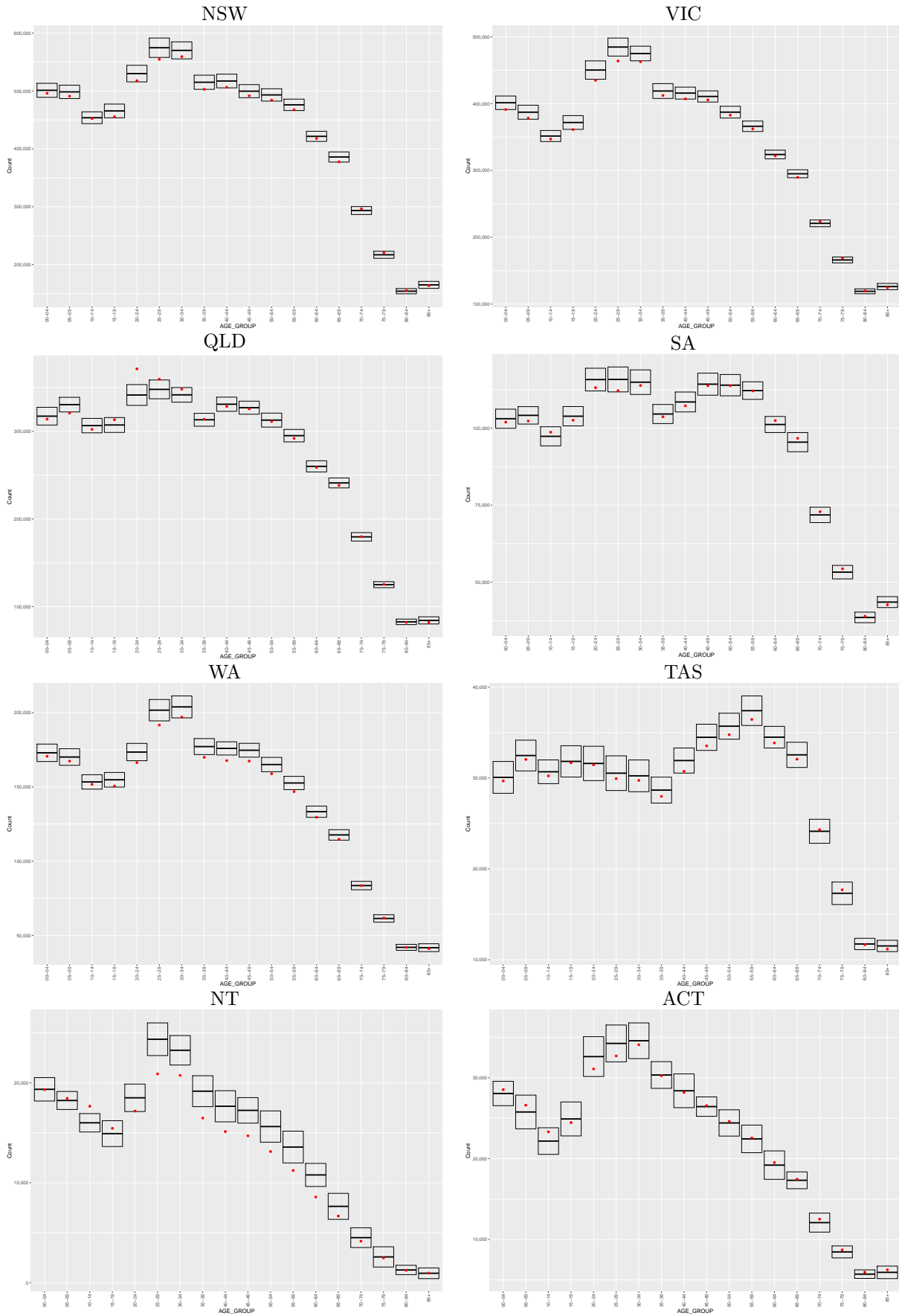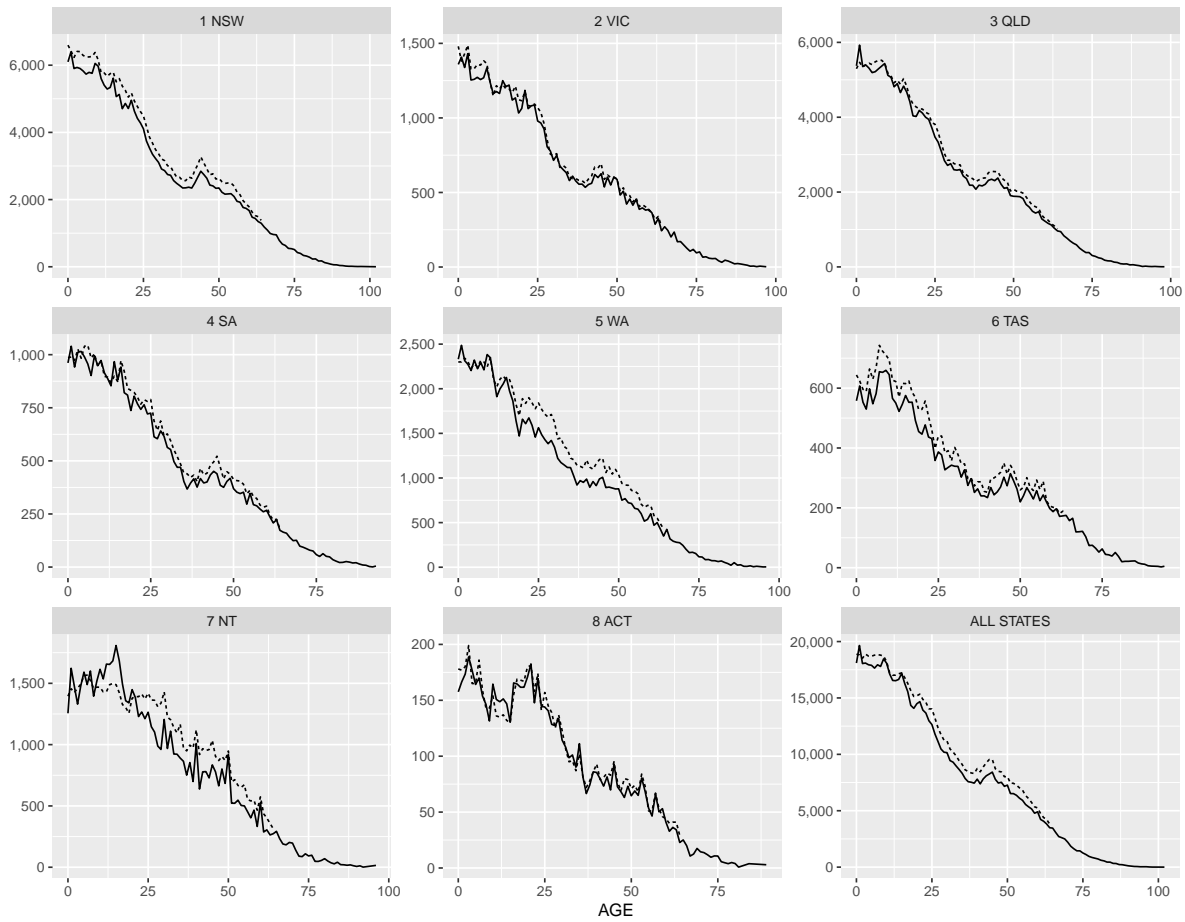Figure 2: Census-SpineSure DSE against 95% confidence interval derived from ERP

Figure 3: Census-SpineSure DSE and ERP of Aboriginal and Torres Strait Islanders Population

Legend:
- - - Census-SpineSure Dual System Estimator
— Official Estimated Resident Population (ERP)