A horizontal teal bar with a white circular icon on the left side.

Coverage assessment in an administrative census: A progress report on issues and methods

Census Transformation programme

Tracey Savage
Christine Bycroft



Crown copyright ©

This work is licensed under the [Creative Commons Attribution 3.0 New Zealand](https://creativecommons.org/licenses/by/3.0/nz/) licence. You are free to copy, distribute, and adapt the work, as long as you attribute the work to Statistics NZ and abide by the other licence terms. Please note you may not use any departmental or governmental emblem, logo, or coat of arms in any way that infringes any provision of the [Flags, Emblems, and Names Protection Act 1981](https://www.legislation.govt.nz/act/public/1981/0049/01.01/01.01/). Use the wording 'Statistics New Zealand' in your attribution, not the Statistics NZ logo.

Disclaimer

This report represents the views of the author. It does not necessarily represent the views of Statistics NZ and does not imply commitment by Statistics NZ to adopt any findings, methodologies, or recommendations. Any data analysis was carried out under the security and confidentiality provisions of the Statistics Act 1975. Unless otherwise stated, results presented are the result of data analysis undertaken by the author.

Liability

While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation

Savage, T, & Bycroft, C (2014). *Coverage assessment in an administrative census: A progress report on issues and methods*. Available from www.stats.govt.nz.

ISBN 978-0-478-42916-9 (online)

Published in September 2014 by

Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact

Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4610
www.stats.govt.nz



Contents

List of tables and figures	4
1. Background	5
Census Transformation	5
About this paper.....	5
2. Introduction	7
What do we mean by coverage assessment?	7
Population estimation and the current New Zealand census.....	7
Scope and aims	8
3. Methods	9
Coverage issues in New Zealand administrative data sources	9
Coverage issues and assessment methods at other NSOs	10
4. Key findings	11
Coverage issues in New Zealand administrative data sources	11
Coverage issues and assessment methods at the Office for National Statistics.....	13
Coverage issues and assessment methods at other NSOs	15
Coverage issues in an administrative census	17
Methods for dealing with coverage issues in an administrative census	20
Overarching implications for an administrative census in New Zealand.....	27
5. Conclusion	29
References	31
Appendix: Coverage issues and assessment methods at NSOs with an administrative census.....	33
Austria.....	33
Germany	34
Israel	36
Singapore.....	37
Switzerland	38



List of tables and figures

Tables

1. Census models in selected countries	15
2. Approaches for dealing with coverage issues when constructing the administrative population	20
3. Approaches for estimating and adjusting for remaining coverage issues	23
4. Approaches used for coverage estimation	25



1. Background

Census Transformation

In March 2012 the New Zealand Government agreed to a Census Transformation strategy. This strategy has two strands:

- a focus in the short-to-medium term on modernising the current census model and creating efficiencies.
- a longer-term focus on investigating alternative ways of producing small area population and socio-demographic statistics, including the possibility of changing the census frequency to every 10 years, and exploring the feasibility of a census based on administrative data (Statistics NZ, 2012a).

The main emphasis of the longer-term strand of the Census Transformation programme is on the feasibility of producing census information from existing administrative data sources, as this aspect is the least understood. In addition, exploring the suitability of existing data sources to produce official statistics is part of our drive to be “an administrative data-first statistics office” (Statistics NZ, 2012b). The investigation takes a phased and iterative approach.

The first phase of the investigation (which this paper is part of) is designed to provide evidence that will inform decisions on the preferred direction for future development of the New Zealand census (see Statistics NZ, 2014 for an overview). The early focus is on developing an understanding of future census information requirements, and whether existing administrative sources can meet those requirements.

Our current understanding indicates that two approaches appear to offer feasible solutions for an administrative census if barriers can be overcome (Statistics NZ, 2014a).

The first approach is to create a national population register. This approach is the basis for successful international examples. However, New Zealand does not currently have a population register, and privacy, legal, and cost concerns mean that the creation of a register in the near future is unlikely. Therefore we assume there will be no New Zealand national population register in the next 10 to 20 years.

Another approach is to link multiple existing data sources. This shows promise as a statistical solution to produce population counts that would avoid the need to survey everyone. This approach is the current focus of investigations, including this paper.

An assessment of coverage issues is integral to the development of a census based on administrative data sources, and methods to measure and adjust for coverage errors need to be developed.

About this paper

Coverage assessment in an administrative census: A progress report on issues and methods forms part of the first phase of the Census Transformation programme. The paper reports on preliminary investigation into issues and methods for coverage assessment, if Statistics New Zealand moves to a Census of Population and Dwellings based primarily on administrative data.

The paper describes coverage issues already identified in New Zealand administrative sources. It also summarises the experience of countries that have recently moved to, or are considering a move to, a census that relies on population registers or other administrative sources to count their population.

The purpose of this paper is not to make final decisions about coverage assessment for a New Zealand administrative census, but rather to help guide decisions about where to direct more in-depth investigations in the second phase of work.



2. Introduction

What do we mean by coverage assessment?

Coverage assessment aims to measure errors relating to **who** is counted in the census population.

No approach to producing population counts is fully accurate. This is true whether Statistics NZ produces population counts via a full-enumeration census, as we do currently, or via an administrative based census, as we are considering in the Census Transformation programme.

In any census model there are two broad coverage issues that lead to inaccuracies in census population counts:

- People who should be counted in the census, but we do not count for some reason – for example, new born babies who are missed by the census. This results in census undercount.
- People who should not be counted in the census, but are counted for some reason – for example, people who are counted more than once, or people who have left New Zealand to live overseas. This results in census overcount.

At the national level, perfect coverage (ie no undercount and no overcount) means that we count every person who should be counted, and no one that should not – we count all the right people.

At the subnational level, perfect coverage means that we count every person within each geographic area, and no one from outside that geographic area – we count all the right people, in the right place.

Population estimation and the current New Zealand census

New Zealand's current census addresses coverage errors through a coverage survey called the Post-enumeration Survey (PES). The PES is an independent survey of a small sample of the population, which measures the accuracy of the census population count. The PES measures people missed by the census, people counted more than once in the census, and the resulting net census undercount.

Statistics NZ produces official national population estimates and subnational population estimates to territorial authority (TA) and area unit (AU) level.¹ These official estimated resident population (ERP) figures are based on the New Zealand resident population concept, defined as all individuals who usually live in a given area at a given time.²

¹ For definitions and explanations of Statistics NZ geographies, including territorial authorities and area units, see www.stats.govt.nz/Census/about-2006-census/2006-census-definitions-questionnaires/definitions/geographic.aspx

² For a definition of the Estimated Resident Population, see <http://datainfolus.stats.govt.nz/Item/nz.govt.stats/7751f101-7b2d-4e97-a487-3ac4126d22d4>

The five-yearly New Zealand Census of Population and Dwellings forms the basis for the ERP. The base ERP estimate that is calculated each census year includes:

- all residents present in New Zealand, and counted by the census – the census usually resident population count
- residents who are temporarily overseas, who are not included in the census
- an adjustment for residents missed or counted more than once by the census – the net census undercount, as measured by the PES.

Visitors from overseas are excluded from the base ERP calculation.

In between censuses, the estimated resident population for a given date is calculated by updating the base ERP for births, deaths, and net external migration (arrivals less departures) of residents during the period between census night and the given date.

Just as the PES is used to measure and adjust for inaccuracies in the current census, we assume that any census based on administrative data will require an independent check of the accuracy of the population count constructed from the administrative data sources. We assume this independent check will be some form of coverage survey.

A coverage survey will be required to measure the levels and patterns of coverage errors in the population constructed from administrative sources. For example, overcount, undercount, and resulting net undercount by age, sex, ethnicity, and geographic location. These measures will feed into an estimation model that provides the ERP for a given reference date.

Scope and aims

The overarching questions for the Census Transformation programme looking at the potential of an administrative census, in relation to coverage, are:

- What kind of coverage survey should be used with the administrative census models being considered?
- How should coverage survey results be used in the estimation process to adjust the administrative data counts?

The design of coverage surveys and population estimation methods is a complex area and raises difficult technical issues.

This paper aims to:

- provide an overview of the coverage issues likely to be encountered in a New Zealand census based on administrative data
- present methods used, or being investigated, overseas in similar situations.

The specific research questions addressed in this paper are:

1. What coverage issues might we encounter in moving to an administrative census, and how do these differ from the coverage issues we currently face in our census?
2. What coverage adjustment and assessment methods might we need to develop and carry out for an administrative census, in order to deal with these coverage issues?
3. What are the implications for a New Zealand administrative census, and for a coverage survey? In particular, will our PES suffice or do we need to change our approach to coverage estimation?



3. Methods

This chapter outlines the two strategies we used to identify potential coverage issues in an administrative census. We looked at:

- coverage issues in New Zealand administrative data sources
- coverage issues and assessment methods at other national statistics offices.

Coverage issues in New Zealand administrative data sources

Statistics NZ has evaluated the potential of administrative data sources currently available in New Zealand for producing population estimates, for both **single administrative data sources** (Statistics NZ, 2013) and a **linked administrative data source** (Gibb, & Shrosbree, 2014).

The single administrative data sources were evaluated to assess their potential for producing subnational population estimates in between censuses. Four data sources - health, tax, education, and electoral - were investigated.

The linked administrative data source that we evaluated was Statistics NZ's Integrated Data Infrastructure (IDI), where rules for combining linked tax, education, and migration data were developed to produce estimates of the usual resident population.

Several quality measures were used to assess these administrative data sources, including some related to coverage. The administrative population concepts were compared to the statistical usual resident population concept used for the estimated resident population (ERP), and aggregate population counts were compared between the administrative data sources and the ERP. Comparisons were made at both the national level and for subnational geographic areas, as well as key breakdowns by age and sex.

For the linked administrative data in the IDI, linking quality was used as a quality measure as this also relates to coverage. As there is no common unique identifier used across government agencies in New Zealand, many of the IDI linkages use probabilistic techniques, using variables such as name, date of birth, and sex (Statistics NZ, 2014b).

Errors in the linkage can result in both undercoverage and overcoverage for population counts derived from the IDI. For example, if records for the same individual that should be linked together are not linked (false negative links), that individual is counted twice.

Comparing aggregate population counts produced from administrative data sources with ERP counts has provided an initial, high-level indication of the coverage issues likely to be encountered with most of the main administrative sources available for an administrative census in New Zealand.

However, even if the population counts from an administrative source are close to the ERP, there is no guarantee that the administrative data includes all the right people (ie only those in the New Zealand resident population). To accurately work out how well administrative data sources cover the New Zealand resident population, we need to check coverage at the individual level. We plan to investigate this in the next phase of the Census Transformation programme.

Coverage issues and assessment methods at other NSOs

In this paper we have drawn on work undertaken by the United Kingdom's Office for National Statistics (ONS) in its Beyond 2011 programme.

The ONS has been "taking a fresh look at options for the production of population and small area socio-demographic statistics" (Ralphs & Staples, 2012). This includes examining the statistical challenges of using administrative data-based models to produce population and small area socio-demographic statistics. Their research identified the main issues for coverage assessment, and approaches to overcoverage adjustment and coverage survey design.

We also identified other National Statistics Offices (NSOs) that have recently made the move from a full-coverage census to an administrative census.

Those NSOs that seemed most relevant to New Zealand were selected for further follow-up work to understand the coverage issues they encountered, and the coverage assessment methods they used.

The selected NSOs were:

- Austria
- Germany
- Israel
- Singapore
- Switzerland



4. Key findings

Coverage issues in New Zealand administrative data sources

Single administrative data sources

Statistics NZ (2013) evaluated four administrative data sources: primary health organisation (PHO) enrolments, tax data in the linked employer-employee data (LEED), school roll returns data, and electoral enrolment data.

A major conclusion was that none of these administrative data sources can produce sufficiently accurate population estimates on their own. In terms of coverage compared with the estimated resident population (ERP), good results were found for some groups in some sources. For example, the national coverage rate in school roll returns data is close to 100 percent between the compulsory school ages of 6 and 16 years, and both electoral enrolment and PHO data show national coverage rates close to 100 percent above age 35.

Overall, there is a tendency for undercoverage in these administrative sources compared with the ERP, and results differ between males and females, particularly in the PHO data and LEED tax data.

In summary, each of the data sources evaluated has limitations, with the key coverage issues being:

- poor coverage of young adults. None of the data sets evaluated has sufficient coverage of the 17–30-year-age-group.
- poor coverage for subnational territorial authority areas and area units. High rates of undercoverage are found in many areas, as well as some overcoverage in other areas. Low quality address data in the administrative sources directly affects the quality of subnational estimates.

Linked administrative data source

Gibb and Shrosbree (2014) examined the potential for linked administrative data sources to provide accurate population estimates by age, sex, and geographic area using Statistics NZ's Integrated Data Infrastructure (IDI)³ as a test environment.

The IDI continues to evolve as new datasets are added, and significant expansion is planned from the latter half of 2014. Only one of the administrative sources described above (the LEED tax data) was available in the IDI during the Gibb and Shrosbree study.

The IDI is not designed to provide population counts. The IDI contains records for individuals who have **ever appeared** in the contributing data sources, and therefore the IDI includes people who have died (and the death may not be recorded), and people who have, for example, worked and paid tax in New Zealand and have since left the country. Simply using the IDI, as it stands, to count the resident population at a given point in time would lead to gross overcoverage errors. To illustrate, more than 5.6 million individuals had activity recorded in the tax or education datasets in the IDI as at 30 April 2013, which is far greater than New Zealand's ERP of approximately 4.5 million as at 31 March 2013 (Gibb & Shrosbree).

³ For more information about the Integrated Data Infrastructure, see www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx

To reduce the level of overcoverage, Gibb and Shrosbree took a 'signs of life' approach using tax data, tertiary education enrolment data, and secondary school achievement data. Individuals were included only if they were active in either the tax or education data in the five years before the reference date. External migration data, plus date of death where available, was used to remove individuals that had moved overseas or died. The resulting dataset was termed the 'IDI-ERP'.

Aggregate population counts produced from this IDI-ERP were then compared with ERP counts, using the same broad 'coverage rate' measure used to evaluate the single administrative data sources. The IDI-ERP total population for ages 15 years and over was 98 percent of the ERP, and coverage rates were consistently good for all ages between 15 and 75 years at the national level. However, the rules used in the 'signs of life' approach resulted in very poor coverage for individuals aged under 15 years.

Other findings related to coverage were:

- considerable overcoverage for individuals aged 75 years and over (likely due to deaths within the last five years not recorded by Inland Revenue)
- some undercoverage among women aged 25–65 years, and among men approaching 65 years (superannuation age)
- subnational coverage rates were variable, with substantial undercoverage in some territorial authorities – likely due to the low quality of address data in the IDI.

The results obtained from 2013 IDI data could be improved by adding datasets such as birth registrations, school rolls, health data (to increase the coverage of children), and death registrations (to remove deceased individuals).

It seems likely that the major gaps in younger age groups and overcoverage of older ages can be greatly improved as the IDI expands. There is also the possibility of adjusting the 'signs of life' rules that have been applied to minimise overcoverage, or alternatively to minimise undercoverage.

Implications for a coverage survey

The findings described above have several implications for a coverage survey of an administrative census.

It seems likely that population counts derived from a linked data source such as the IDI would still have inaccuracies, due to both undercoverage errors and overcoverage errors for different age and sex groups.

Linkage errors may also contribute to overcoverage. No estimate of false-negative linking errors is available. However considering the education-tax linkages used in the IDI-ERP, even a moderate false-negative rate of 5 percent would have a substantial impact on the total national population, increasing it by 2.1 percent. And for the 15–29 year age group in which this data is concentrated, it would lead to 3.9 percent overcoverage.

Perhaps the most problematic coverage issue in the IDI-ERP is the substantial undercoverage in some territorial authorities, which becomes even more pronounced for the smaller area unit geographies.

Coverage issues and assessment methods at the Office for National Statistics

Beyond 2011 programme

Statistical challenges of using administrative sources

In its early stages, the Beyond 2011 programme at the UK's Office for National Statistics (ONS) identified two key challenges in using administrative data to estimate populations (Ralphs & Staples, 2012).

Both challenges related to the impact of time lags on administrative data:

- Under-reporting or delays in reporting emigration and deaths (unrecorded emigration, unrecorded deaths). This can lead to 'list inflation' and national overcoverage of the population, as individuals who left the country or died are not removed from the administrative data source in a timely manner, if at all.
- Address information not being kept up-to-date when individuals move home (unrecorded internal migration). This can lead to people being counted 'in the wrong place', which the ONS refers to as local (rather than national) over and undercoverage,⁴ or being recorded in more than one place – resulting in overcoverage.

Ralphs and Staples also identified other potential coverage issues, including:

- undercoverage of population groups that have minimal contact with administrative systems, or are missed altogether (the 'hard-to-count'). For example, young healthy males and people not eligible for NHS services
- undercoverage due to time lags in reporting of births (unrecorded births)
- list inflation, or local over and undercoverage, due to the handling and coverage of communal establishments differing across administrative sources. For example, people in student halls may have multiple locations where they could be resident, such as a term-time address and a home address
- coverage issues due to linking errors that result when two or more administrative data sources are matched together.

In summary, Ralphs and Staples noted that in the types of broad coverage administrative sources most likely to be used in an administrative census, overcoverage (eg due to list inflation) is likely to be a much more significant issue than in a full-enumeration census. In fact, so much so that overcoverage has the potential to obscure or out-weigh any undercoverage issues.

Under the Beyond 2011 programme, the ONS has developed, and is using, an administrative data quality framework to rigorously assess and better understand the scope, coverage, and statistical data quality of administrative data sources on a case-by-case basis.

Addressing the statistical challenges of using administrative sources

The Beyond 2011 programme identified that a coverage survey will almost certainly be needed to robustly adjust administrative data sources, and improve the quality of population estimates (Ralphs & Staples).

⁴ The Israeli Central Bureau of Statistics also refers to this as local over and undercoverage, and this is the term used throughout the rest of this paper. Note that local coverage issues can be thought of as a measurement error or misclassification with respect to location, and so ideas and methods from this field of statistics might help us address these issues.

In addition to a coverage survey, Ralphs and Staples identified a range of potential approaches and methods to address coverage issues.

These approaches are summarised below:

- Robust and reliable data linking methods to manage the impact of linking errors on coverage.
- Overcoverage adjustment processes to correct administrative sources for list inflation before any coverage survey.
- Coverage measurement methods to estimate and correct for overcoverage, in addition to undercoverage, through a coverage survey.
- Modelling and weighting methods to extrapolate the results of coverage estimation to geographic areas not included in the coverage survey, and produce population estimates.
- Evaluation methods to assess the quality of and validate the population estimates produced.

From their review of international approaches to estimation and adjustment for census coverage errors (Elkin, Dent & Rahman, 2012), the Beyond 2011 programme identified some potentially useful approaches to consider.

In relation to the use of dual system estimation (DSE), Elkin et al noted that the UK, Israel, US, and Australia all adjust census data for national overcoverage, before using an independent coverage survey and DSE to estimate undercoverage.⁵ This separate adjustment for overcoverage – either through use of additional data sources and/or further enumeration or surveys – ensures that the negligible overcoverage assumption underlying the DSE method is met.

In particular, Elkin et al noted the US and Israel's use of a second sample survey, involving direct sampling from census and administrative records respectively, to separately estimate and adjust for overcoverage.

⁵ Though there are variations in how overcoverage is included and how the DSE is estimated.

Coverage issues and assessment methods at other NSOs

We now consider the experience of five other NSOs that have recently moved to an administrative census model: Austria, Germany, Israel, Singapore & Switzerland.

Census models in selected countries

Table 1 shows how the NSOs of five selected countries have shifted from a full-enumeration census to a census based on administrative data.

Among these NSOs the availability and quality of population registers varies. There are also variations in the presence and type of sample surveys used to supplement information available from administrative sources. Address registers and building registers are also used in their census models.

Table 1
Census models in selected countries

	Previous census model	Year of last traditional census	Year of first administrative census	Administrative census model
Austria	Full enumeration census (10-yearly)	2001	2011	Register-based census <ul style="list-style-type: none"> • Central Population Register • other linked administrative sources
Germany	Full enumeration census	1987	2011	Combined census <ul style="list-style-type: none"> • a centralised population register • supplemented by surveys
Israel	Full enumeration census <ul style="list-style-type: none"> • short form to 100% • long form to 20% 	1995	2008	Combined census <ul style="list-style-type: none"> • the Population Registry • supplemented by a 17% survey of households
Singapore	Full enumeration census (10-yearly)	1990	2000 (repeated in 2010)	Combined census <ul style="list-style-type: none"> • Household Registration Database • supplemented by a 20% survey of households
Switzerland	Full enumeration census	2001	2011	Combined census – a centralised population register <ul style="list-style-type: none"> • supplemented by annual household surveys (approx 5%)

Coverage issues in selected countries

All five countries that we investigated experience issues of overcoverage, undercoverage, and poor address quality in their administrative census. In addition, Germany and Israel experienced issues with counting institutional populations, such as people in prison.

Switzerland reports that undercoverage and overcoverage remain in the registers, even after adjustment procedures are applied (Anne Massiani, personal communication, 28 March 2013). Germany has relatively large variance in overcount rates between municipalities, even after adjustment procedures are applied (Federal Statistical Office of Germany, 2006).

[See Appendix: Coverage issues and assessment methods at NSOs with an administrative census](#) for detailed descriptions of the administrative census in each of the five selected NSOs, with an emphasis on coverage.

The following sections summarise the coverage issues and the methods used to deal with these issues across the five selected countries, the ONS Beyond 2011 programme, and New Zealand's experience. We address each of the following questions:

1. What coverage issues might we encounter in moving to an administrative census, and how do these differ from the coverage issues we currently face in our census?
2. What coverage adjustment and assessment methods might we need to develop and carry out for an administrative census, in order to deal with these coverage issues?
3. What are the implications for a New Zealand administrative census design, and for a coverage survey? In particular, will our current PES suffice or do we need to change our approach to coverage assessment?

Coverage issues in an administrative census

Taken together, the findings from Statistics NZ's evaluation of administrative data sources, the ONS Beyond 2011 programme, and the administrative census experiences of the five selected countries suggest answers to the question:

1. What key coverage issues might we encounter in moving to an administrative census, and how do these differ from the coverage issues we currently face in our census?

Overcoverage is potentially a more significant issue

In a full-coverage census, coverage adjustment is primarily about controlling for undercoverage, while overcoverage is considered or assumed to be a relatively minor issue (Ralphs & Staples, 2012).

This is true of New Zealand's current census, where the main coverage issue is undercoverage – people who should be included and counted in the census, but that we miss for some reason – and overcoverage is a relatively minor problem (though the PES does measure people who are counted more than once in the census, and adjusts for both types of error).

In broad coverage administrative sources, which are likely to be used in an administrative census, the reverse situation is often true – overcoverage is likely to be a much more significant issue (Ralphs & Staples).

The ONS Beyond 2011 programme identified time lags in data reporting in administrative data sources to be one of the key coverage challenges (Ralphs & Staples). This can lead to 'list inflation' and overcoverage of the population because of underreporting of emigration and deaths (ie individuals who have left the country or died are not removed from the administrative data source in a timely manner, if at all).

This scenario was a common coverage issue identified in all of the other NSOs investigated. For example, Statistics Singapore found unreported emigration and deaths abroad to be a source of list inflation for their elderly population (aged 75+) (Singapore Department of Statistics, 2002).

In the New Zealand context, overcoverage is a known and significant issue in the integrated data infrastructure (IDI), with administrative sources containing unrecorded deaths, unrecorded emigration, and inactive individuals. Reducing this overcoverage is the reason why the 'IDI-ERP' was generated from the IDI to approximate the New Zealand resident population. Such an approach to adjust for overcoverage is likely to be needed regardless of the New Zealand administrative data sources used in an alternative census model.

Undercoverage is still an issue

Despite overcoverage being potentially a more significant issue in an administrative census, there will still be some undercoverage of people who are not recorded on any of the available administrative sources, or are not represented on the population list generated from administrative sources.

This is certainly true of New Zealand administrative data sources. Our evaluation of single administrative data sources found that none of the data sets evaluated had high coverage of the young adult age group (17–25 years) (Statistics NZ, 2013). The linked IDI data showed use of high school and tertiary enrolments significantly improved results for this age group. The administrative data available in the IDI at the time of the evaluation also showed some undercoverage among women aged 25–65 years, and for men approaching 65 years.

Undercoverage is also a common theme across the other NSOs investigated. The ONS noted that young, healthy males are likely to be missing from their administrative data sources (Ralphs & Staples). Both the Austrian and Israeli NSOs recognised that those in the country but not on their population registers (eg foreigners without a residence permit, certain groups of homeless people) would not be counted (Statistics Austria, 2012a and Elkin, Dent & Rahman, 2012). Austria and Switzerland also noted that time lags in the reporting of births resulted in undercoverage in their population registers (Statistics Austria, 2012a and Swiss Federal Statistics Office, 2012).

One concern identified by the ONS is the danger that over-coverage due to list inflation may potentially obscure or out-weigh problems of undercoverage (Ralphs & Staples). Therefore methods for dealing with coverage issues in a New Zealand administrative census need to be able to account for both over- and undercoverage.

Poor quality address data is a key coverage issue

Even if coverage is good at the national level, poor quality address data can result in both over- and undercoverage issues at the subnational level.

The ONS identified address information not being kept up-to-date when individuals move within the country – or unrecorded internal migration – as a key coverage challenge in using administrative data to estimate populations (Ralphs & Staples, 2012). This can lead to people being counted ‘in the wrong place’, which the ONS refers to as local (rather than national) over- and undercoverage.

Unrecorded internal migration also posed a major challenge for Israel’s Central Bureau of Statistics in moving to their Integrated Census in 2008. Approximately 25 percent of people are recorded in the population register at an address other than at which they reside.

Statistics Austria also found that time lags in reporting change of address resulted in undercoverage in their population register (Statistics Austria, 2012a).

Germany and Switzerland maintain population registers at a regional level (eg canton, municipality), and individuals need to de-register and re-register when they move region. In this scenario, time lags in updating address information can affect one or both regional registers, potentially resulting in three different types of coverage issue.

These three coverage issues are:

- local over- and undercoverage that balances out at the national level – when a person has moved, but has not de-registered or re-registered yet (recorded in the wrong place)
- overcoverage (national and local) due to duplicates – when a person who has moved registers in their new region, but has yet to de-register in their old region
- undercoverage due to missing records – when a person who has moved de-registers in their old region, but has yet to register in their new region.

Our evaluation of administrative data sources has identified poor quality address data as a key issue for New Zealand. The issue was identified in many of the single administrative data sources evaluated (Gibb, 2013), and also in the IDI-ERP linked administrative data, where the quality of address data was too low to produce sufficiently accurate estimates for subnational areas (Gibb, 2014).

Poor quality address data is likely to result in a mix of local over- and undercoverage that balances out (due to time lags in updating address when a person moves), and local undercoverage (due to missing address data for a person, or address data that is not detailed enough to geo-code to the desired level). Therefore, methods for dealing with coverage issues in a New Zealand administrative census need to be able to account for poor quality address data.

Non-private dwellings create unique problems

Israel, Germany, and the ONS identified issues around coverage of special facility addresses, institutions, or communal establishments. These types of dwellings are equivalent at some level to non-private dwellings (NPDs) in the New Zealand census. We have yet to investigate coverage of NPDs in New Zealand administrative data sources, but plan to in the next phase of the Census Transformation programme.

Linkage errors need to be managed

While linking multiple administrative data sources can potentially help improve coverage (eg by removing list inflation, filling gaps in data), it also adds another element of complexity to the coverage problem.

The impact of linking errors on coverage – both undercoverage and overcoverage – and on population estimates has been identified as an issue by the ONS and Statistics NZ.

Both agencies are investigating the use of probabilistic linking techniques to combine multiple administrative data sources, create an administrative census population list, and produce population estimates.

Linking errors will also be an issue for any agency using probabilistic matching techniques without a data source acting as a fixed population spine (ie using the union of multiple datasets), for example Germany.

Therefore measuring linking errors, and assessing their impact on coverage and population estimates, needs to be part of our investigations in the next phase of the Census Transformation programme.

Methods for dealing with coverage issues in an administrative census

Some form of quality assurance or coverage checks and adjustments will be required to address or manage coverage issues in the administrative data sources used to produce population counts, and to improve the quality of the population counts.

We now address the second question:

2. What coverage adjustment and assessment methods might we need to develop and carry out for an administrative census, in order to deal with these coverage issues?

Processes used to manage quality

The NSOs we investigated use or considered using, a range of approaches and methods to deal with data quality and coverage issues in the data sources underlying an administrative census.

Table 2 summarises the approaches and methods that countries have in common, or are most relevant to New Zealand, given the coverage issues identified in our existing administrative data sources.

Table 2
Approaches for dealing with coverage issues when constructing the administrative population

	Adjust for overcoverage	Improve the quality of address data	Carry out additional field work to confirm or update administrative data	Acknowledge but do not estimate or adjust for undercoverage
Austria	Yes	...	Yes	Yes
Germany	Yes	...	Yes	...
Israel	No	Yes	Yes	Yes
Singapore	Yes	Yes	Yes	...
Switzerland	Yes	...	Yes	...
UK	Yes
Symbol: ...not applicable				

Adjust for overcoverage

Overcoverage is a common issue across all the NSOs we investigated, mainly due to time lags when updating registers and administrative data sources (eg unrecorded deaths and unrecorded migration).

Useful approaches to adjusting for overcoverage include:

- match with or adjust for death data, to remove unrecorded deaths
- match with or adjust for external migration data, to remove unrecorded migration
- apply filtering rules to include or exclude individuals from the target population (eg a signs of life approach)
- check for and remove duplicate records
- compare data at two different time points to correct for temporary overcounts.

If such adjustments are applied aggressively, the problem of overcoverage can potentially be converted to one of undercoverage. If the rules are applied more leniently, however, overcoverage is likely to remain in the administrative list. Some of these adjustment approaches were used in our evaluation of the IDI to generate the IDI-ERP population, with the aim of striking a balance between overcoverage and undercoverage.

Improve the quality of address data

Poor quality address data is another common problem across the NSOs we investigated. To address this, both Israel and Singapore link data sources with higher quality address information in to their population registers.

Carry out additional field work to confirm or update administrative data

Additional surveying – either one-off or ongoing – to address identified coverage issues is another common theme across NSOs.

Examples of this include:

- a focused census of those aged 75 years and over (Singapore – one-off)
- full-enumeration of special facility addresses or institutions (Germany and Israel – ongoing)
- contacting individuals, where needed, to confirm their residence as at the census reference date (Austria – ongoing)
- contacting owners of houses and dwellings in small municipalities to confirm the number of inhabitants / occupants, where a discrepancy between data sources has been identified (Germany – ongoing).

Minimise the impact of linking errors on coverage

As noted above, linking errors can have an impact on coverage, an issue identified by both the ONS Beyond 2011 programme, and the Census Transformation programme. To minimise these impacts, we need to achieve reliable, accurate matches, with low levels of linking error.

We therefore need to manage various factors that contribute to the level of linking error, for example, the quality of the variables to be used as matching keys,⁶ the probabilistic linking technique, and the setting of the linking parameters.

⁶ If addresses are going to be used as a matching key, this provides another compelling reason to improve the quality of address data in administrative sources.

However it will not be possible to eliminate linking errors altogether, so we also need to find ways to measure or estimate the level of linking errors, to work out their impact on population estimates and to adjust for this.

Other approaches

- Work with agencies to address or resolve quality and coverage issues, and improve the source administrative data
- Include additional administrative datasets to address known gaps or undercoverage (eg linking in datasets that have high coverage of undercovered age groups)
- Delay the extraction of administrative data for a period following the census reference date (eg six months) to avoid coverage issues due to time lags.

Acknowledge known coverage issues, but do not adjust or estimate for them

Finally, it is possible to acknowledge that some coverage errors exist, but that no action will be taken to remove them. For example, the Israeli Central Bureau of Statistics acknowledges undercoverage in its population register (people without an ID number but who are resident in Israel for 12 months or more), but makes the explicit assumption that anyone not on the population register should not be counted in the census.

Approaches to a coverage survey and coverage estimation

Even after checks and adjustments to the underlying administrative sources have been completed, as outlined in the previous section (Processes used to manage quality in administrative data sources), ideally an independent check on the accuracy of the resulting population list will still be done to assess any remaining over- and / or undercoverage.

Of the NSOs investigated, most do (or plan to do) this independent check by running a coverage survey. Only two – Austria and Singapore – do not, as they believe that the population counts produced using their population registers are sufficiently accurate and reliable for their purposes.

However, Austria and Singapore still use comparison registers and administrative data to independently validate their population counts. Other NSOs that have a census based on population registers they believe to be reliable and of high quality (eg Sweden, Finland, and the Netherlands) take the same approach in not running a coverage survey.

Table 3
Approaches for estimating and adjusting for remaining coverage issues

Table 3 summarises the approaches taken to estimate and adjust for remaining coverage issues once the best administrative population list has been achieved.

	Independent validation checks using other admin sources	Estimate level of undercoverage	Estimate level of overcoverage	Adjust for coverage errors
Austria	Yes	No	No	...
Germany	No	Yes – by an independent coverage survey	Yes – by an independent coverage survey	Yes – by imputing and deleting individual records in the population register
Israel	No	Yes – by an independent coverage survey	Yes – by a dependent coverage survey	Yes – by generating and attaching a weight to each record in the population register
Singapore	Yes	No	No	...
Switzerland	No	Yes – by an independent coverage survey	Yes – by matching a sample from the population list back to the full population list	No adjustment ¹
UK	Yes	Yes – by a coverage survey	Yes – by a coverage survey	Yes – by use of modelling and weighting methods
Symbol: ...not applicable				
¹ Rather, the Swiss Federal Statistical Office plans to publish a separate report on the quality and comprehensiveness of their register-based census, including under- and overcoverage estimates.				

In New Zealand, significant undercoverage and overcoverage, both national and local, is expected to remain for any population constructed from administrative sources. Just as we do with our current census, an independent check on the accuracy of the population list constructed from the administrative data source(s) – in the form of a coverage survey – will be required for any New Zealand administrative census model.

However, given that the coverage issues in an administrative census differ from those in a full-enumeration census, new approaches to a coverage survey and coverage estimation will need to be developed.

We need to consider approaches that deal with both undercoverage and overcoverage, and with poor address data quality. Our current census coverage survey (PES) is unlikely to be sufficient – it is highly likely we will need to adapt or significantly change it, or adopt a different coverage survey approach and design altogether.

Design of a coverage survey

The approach to, and design of, a coverage survey depends on what coverage issues exist in the administrative census population list. Just as the PES is tailored to our current

census model and its coverage issues, a coverage survey for an administrative census also needs to be tailored to the model used and the coverage issues in the population list constructed from administrative data sources.

While the most obvious approach at this stage is a survey to assess coverage of the resulting population list, we also need to consider other options (eg assessing coverage of each of the administrative data sources that contribute to the list).

Understanding the level and pattern (eg by age, sex, ethnicity, geographic location) of coverage errors in the population list can also help us target the sample toward processes or population groups that we know to be hard to count, and maximise efficiency of the coverage survey.

The approach and design also needs to take into account the estimation process or model that will be used to adjust the census population counts, and provide the base estimated resident population for census year, given coverage estimates feed into this.

Therefore the approach to, and design of, a coverage survey depends on:

- the administrative census model used
- how many and which administrative data sources contribute to the model
- how the administrative data sources are combined to create the population list – eg linking methodology, parameters and errors, processes used to remove overcoverage from the population list
- the specific under- and overcoverage issues that remain in the resulting population list
- whether and how quickly the coverage issues change over time (which may impact on how often a coverage survey needs to be run)
- the level of detail at which coverage estimates and population estimates are required – eg sex, age, ethnicity, geographic area
- the accuracy required of the coverage estimates and population estimates.

Many of the factors listed above are unknown at this point, but will be further refined in the next phase of the Census Transformation programme.

Estimating coverage components separately

A common approach among the NSOs that we investigated is to separately estimate for the two components of coverage – overcoverage and undercoverage. Estimates may be produced from a coverage survey, or by some other means.

Of the NSOs that carry out coverage estimation for their administrative census, three different approaches are used, as summarised in Table 5.

Table 4
Approaches used for coverage estimation

	Approaches to estimating undercoverage	Approaches to estimating overcoverage
Germany	Independent coverage survey linked back to individuals in the administrative population list at the sampled address ¹	
Israel	Independent coverage survey linked back to the administrative population list at the individual level ²	Dependent coverage survey, starting from a sample of the administrative population list, to track and trace selected individuals ³
Switzerland	Independent coverage survey linked back to the administrative population list at the individual level ⁴	By matching a sample drawn from the administrative population list back to the rest of the list, to identify duplicates and other erroneous records

¹ The survey is carried out in large municipalities only. Linking back to individuals at the sampled address allows them to identify both people who were undercounted and overcounted at that address.

² This is similar to our current PES approach.

³ This involves sampling in the opposite direction to our PES.

⁴ This is similar to the approach used by the ONS and Canada to estimate overcoverage in their full-enumeration census models.

In the Israeli approach, the two surveys are carried out in the same geographic areas (ie inter-leaved), leading to efficiencies in the field work required. This is similar to the E-sample and P-sample approach used by the US Bureau of the Census. Both Israel and Germany also collect additional attribute (long-form) information via their coverage survey.

Other approaches are also possible – for example, estimating for undercoverage due to unrecorded births using birth records (an option noted by the ONS) – as are different combinations of these. In the next phase of the Census Transformation programme, we will determine which, if any, of these approaches is best suited to a New Zealand administrative census.

If Statistics NZ uses dual system estimation (DSE) methods to estimate undercoverage, the two-component approach outlined above allows adjustment for national overcoverage first. This is a necessary requirement in order to meet the underlying DSE assumption of negligible overcoverage. Israel also extended the DSE method for its Integrated Census in 2008 to adjust for both local under- and overcoverage due to poor address data quality.

Other potential differences compared with the PES

The information collected in a coverage survey for an administrative census may need to be different from that collected in our current PES, in order to measure coverage. For example, address history information may need to be collected for individuals (as trialled by the ONS Beyond 2011 programme in their Population Coverage Survey Test) to correct for poor quality address data.

In addition, the types of dwelling included in a coverage survey may also need to be different from those in New Zealand's current PES. For example, the ONS noted that communal establishments may need to be included given the potential coverage issues in these.

Approaches for population estimation

Given that coverage issues and coverage estimation approaches in an administrative census differ from those in a full-enumeration census, new approaches for population

estimation would need to be developed if New Zealand moved to an administrative census model.

Among the investigated NSOs that carry out coverage estimation for their administrative census, three different approaches to population estimation are taken:

- no population estimation – the administrative census population counts are not intended to be adjusted using the coverage estimates (Switzerland)
- imputation and deletion of individual records in the population register to adjust for estimates of under- and overcoverage (Germany)
- generation of a ‘census weighting’ – based on the coverage estimates – for each record in the population register (Israel).

All these approaches result in a person-level census dataset being available for analysis and production of population statistics, rather than producing only aggregate population estimates.

The ONS Beyond 2011 programme is also considering various weighting and modelling methods to produce population estimates.

If New Zealand moved to an administrative census, the population estimation approach would need to adjust for linking errors – those that arise in constructing the population list, and in linking the coverage survey to it – in addition to under- and overcoverage.

The population estimation approach would also need to be able to provide measures of uncertainty (eg confidence intervals) for the resulting population estimates. The Census Transformation programme is researching Bayesian methods and models for population estimation, which have the ability to meet both of these needs.

Overarching implications for an administrative census in New Zealand

Implications for an administrative census in New Zealand have been noted in the previous sections where they relate to a key finding about issues and methods for coverage assessment.

We now address the third question:

3. What are the overarching implications for a New Zealand administrative census, and for a coverage survey? In particular, will our current PES suffice or do we need to change our approach to coverage assessment?

Administrative census models are dependent on the wider context in which an NSO operates

Administrative census models, and their coverage assessment issues and methods, vary between NSOs. Each NSO operates in a unique context that depends on a range of factors, many of which are non-statistical – including the political environment, legislative issues, privacy concerns, historical issues, and cultural norms.

These factors can affect the administrative census model used by an NSO in a variety of ways. For example the acceptance of, or appetite for, population registers and compulsory registration; availability of unique identifiers; the level of influence the NSO has over agencies supplying administrative data; and the NSO's ability to accurately measure external migration.

It is not always easy to identify the underlying factors or rationale behind each NSO's administrative census model, and the related coverage assessment issues and methods. NSOs are not always transparent about these factors, and methodological documentation tends to focus more on the statistical rationale, rather than the wider NSO context.

A New Zealand administrative census would need a tailored coverage adjustment and assessment solution

While there will be many similarities between a New Zealand administrative census, and administrative censuses at other NSOs, aspects of the context in which we operate are still unique.

For example:

- Some coverage issues faced by other NSOs are much less of an issue for New Zealand. For example, we are in the relatively unique situation of having high quality external migration data available, which we can readily use to eliminate overcoverage due to emigration.
- Some coverage adjustment methods proposed by other NSOs may not be very useful in a New Zealand administrative census. For example, if we were to take a stringent rules-based approach to reducing overcoverage, we are still likely to need an external check for overcoverage to confirm that it has been eliminated. Also, taken to the extreme, this approach may make the undercoverage problem even worse (ie eliminate valid records from the population).

New Zealand's uniqueness means that there is no 'one-size-fits-all' solution to coverage adjustment and assessment for a New Zealand administrative census.

Therefore, it is unlikely Statistics NZ would be able to adopt the approach of any single NSO in its entirety and apply it wholesale to an administrative census in New Zealand. Instead, we would need to take a mixed approach, selecting various methods from across a range of NSOs. This approach would allow us to tailor an appropriate solution for New Zealand, while drawing on the knowledge, experience, and methods of NSOs facing similar coverage issues.

Our research so far has identified a suite of potential methods for us to consider and select from, or adapt, in the next phase of our investigations. In addition, innovative new methods will also need to be developed to address emerging coverage issues (eg population estimation in the presence of linking errors).

The most relevant NSOs for us to focus on in the next phase of research are the ONS, Israel, and Germany, given the following similarities to New Zealand's situation:

- All three NSOs have investigated or put in place methods to deal with overcoverage in addition to undercoverage.
- In the absence of a unique identifier, Germany has used probabilistic linking to combine multiple administrative data sources to produce population counts, and the ONS is considering this approach as an option.
- Israel has developed methods to deal with poor quality address data in its population register, and the ONS is also researching this area.



5. Conclusion

Key coverage issues

Coverage assessment in an administrative census: A progress report on issues and methods identifies the following key coverage issues, in the event that Statistics NZ moves to a Census of Population and Dwellings based primarily on administrative data:

- Significant overcoverage in administrative data sources, in addition to undercoverage.
- Poor quality address data resulting in coverage errors at the subnational level.
- Linking errors that arise when joining multiple administrative data sources, which can affect both coverage and population estimates.

These coverage issues are similar to those faced by other national statistical offices (NSOs) that have recently moved to an administrative census, or are currently considering doing this.

However, these coverage issues differ from the issues faced in New Zealand's current full-enumeration census.

Managing and measuring coverage issues

Our research of other NSOs also identified a range of potential methods for managing and measuring coverage errors in an administrative census. We have already applied some of these methods (eg 'signs of life' rules) during our initial evaluation of the potential of the Integrated Data Infrastructure (IDI) to produce population estimates.

However, given that each NSO operates in a unique context, there is no single NSO model we can follow exactly. If New Zealand moves to a census based primarily on administrative data, we need to develop a tailored solution for New Zealand, drawing on the knowledge, experience, and methods of other NSOs where relevant.

In conclusion, the evidence gathered in the first phase of our coverage assessment research clearly signals the areas we need to investigate further in the second phase of the Census Transformation programme.

Specifically, to address the type and extent of coverage issues identified in existing New Zealand administrative data sources, we need to focus on developing:

- processes to address or manage quality and coverage issues in the data sources underlying an administrative census
- an independent check, in the form of a coverage survey, as part of any administrative census model
- new, and perhaps significantly different, approaches to a coverage survey and methods for coverage estimation, given our current coverage survey (Post-enumeration Survey) is unlikely to be sufficient
- new methods for population estimation that account for both coverage and linking errors, and provide measures of uncertainty for the estimates produced.

Next steps

To expand on the preliminary investigations in this paper, we recommend the following next steps:

- Investigate the quality and coverage issues in linked administrative data at a more detailed (unit record) level, by linking 2013 Census data to the IDI

- Undertake more detailed research into the coverage estimation and adjustment methods being used or investigated at other NSOs
- Determine the best approaches and methods for dealing with quality and coverage issues in a New Zealand administrative census
- Research and broadly develop coverage estimation and population estimation methods for a New Zealand administrative census
- Produce a high level coverage survey design for a New Zealand administrative census.

However measurement of, and adjustment for, coverage issues cannot be considered in isolation. They need to be factored into the end-to-end design of the administrative census model.

Three inter-dependent streams of work to be considered are:

- constructing the population list to manage known coverage issues, for example unrecorded deaths and false-negative matches
- designing a coverage survey and estimation methods to measure coverage errors in the population list
- developing population estimation methods that include measures of coverage error and linking error.

An iterative approach will be required as these inter-connected work streams progress, and are refined.

Research questions to consider

As we carry out the next steps recommended above, we also need to consider several open research questions:

- What is the best way to deal with the anticipated problems of high overcoverage in addition to undercoverage, and poor address data quality?
- What is the best approach for designing a coverage survey, and for coverage estimation?
- How should coverage survey results be used to adjust counts from the administrative population list?
- What is the impact of linking errors, and are there methodological approaches for making inference about population estimates from data linked with errors?
- What other population estimation methods are needed to meet requirements for population statistics?



References

- Bechtold, S (2013). [The new register-based Census of Germany - a multiple source mixed mode approach](#). Paper from Federal Statistical Office of Germany for ISI 2013. Available from www.statistics.gov.hk.
- Bycroft, C (2014). [Statistics New Zealand: An overview of progress on the potential use of administrative data for census information in New Zealand](#). Available from www.stats.govt.nz.
- Central Bureau of Statistics Israel (2012). [The integrated rolling census in Israel](#). Note for Conference of European Statisticians, 60th Plenary Session – Paris, 2012. Available from www.unece.org.
- Central Bureau of Statistics Israel (2013). [Population census](#). Retrieved 26 June 2013. Available from www.cbs.gov.il.
- Crothers, C, Woodley, A, & Davies, T (2007). [Utilisation of official statistics in the Auckland region](#). Official Statistics Research Series. Available from www.statisphere.govt.nz.
- Elkin, M, Dent, P, & Rahman, N (2012). [Beyond 2011: A review of international approaches to estimating and adjusting for under- and over-coverage](#). Available from www.ons.gov.uk.
- Federal Statistical Office of Germany (2006). [Planning the register-based Census 2011 in Germany](#). Note for Conference of European Statisticians, 54th Plenary Session – Paris, 2006. Available from www.unece.org.
- Federal Statistical Office of Germany (2012). [Lessons learnt from a mixed-mode census for the future of social statistics](#). Note for Conference of European Statisticians, 60th Plenary Session – Paris, 2012. Available from www.unece.org.
- Gibb, S, & Shrosbree, E (2014). [Evaluating the potential of linked data sources for population estimates: IDI as an example](#). Available from www.stats.govt.nz.
- Ralphs, M, & Staples, V (2012). [Beyond 2011: Exploring the challenges of using administrative data \(M2\)](#). Available from www.ons.gov.uk.
- Renaud, A (2007). [Estimation of the coverage of the 2000 Census of Population in Switzerland: Methods and results](#). *Survey Methodology*, vol 33, no. 2, 199–210. Available from www5.statcan.gc.ca.
- Singapore Department of Statistics (2002). [Singapore Census of Population 2000: administrative report](#). Available from www.singstat.gov.sg.
- Singapore Department of Statistics (2003). [Singapore register-based census – lessons learnt and challenges ahead](#). Paper for 21st Population Census Conference – Kyoto, 2003. Available from www.ancsdaap.org.
- Singapore Department of Statistics (2009). [Singapore's Census of Population 2010](#). Available from <http://unstats.un.org>.
- Statistics Austria (2012a). The register-based census in Austria. PowerPoint presentation for the Beyond 2011 International Review in June 2012.
- Statistics Austria (2012b). [Challenges in the transition from a traditional to register-based census in Austria](#). Note for Conference of European Statisticians, 60th Plenary Session – Paris, 2012. Available from www.unece.org.

Statistics New Zealand (2012a). [Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy](#). Available from www.stats.govt.nz.

Statistics New Zealand (2012b). [Statement of Intent 2012–17](#). Available from www.stats.govt.nz.

Statistics New Zealand (2013). [Evaluation of administrative data sources for subnational population estimates](#). Available from www.stats.govt.nz.

Statistics New Zealand (2014a). [An overview of progress on the potential use of administrative data for census information in New Zealand](#). Available from www.stats.govt.nz.


Statistics New Zealand (2014b). [Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project](#). Available from www.stats.govt.nz.

Swiss Federal Statistics Office (2011). [The new census](#). Available from www.bfs.admin.ch.

Swiss Federal Statistics Office (2012). [The Swiss census system: a comprehensive system of household and person statistics](#). Note for Conference of European Statisticians, 60th Plenary Session – Paris, 2012. Available from www.unece.org.

United Nations Statistics Division (2010). [Post enumeration surveys: Operational guidelines – technical report](#). Available from <http://unstats.un.org>.

Zadka, P (2012). 2008 Integrated Census – Israel. PowerPoint presentation for the Beyond 2011 International Review in June 2012.



Appendix: Coverage issues and assessment methods at NSOs with an administrative census

This appendix provides more information about the administrative census carried out by each of the five NSOs covered in this paper. The summary for each NSO includes:

- a short history of its administrative census
- the type of administrative census model that it uses
- the coverage issues it faced when developing and implementing the administrative census model
- the approaches it used to deal with coverage issues.

Austria

Statistics Austria moved to a register-based census in 2011, having carried out its last 10-yearly full-enumeration population census in 2001 (along with a building and housing census and a census of local units of employment).

Before 2001, Austria did not have an interconnected network of population records. Instead, each municipality administered its own records. However, several data sources that would help enable the move to a register-based census, including a population register and a housing register, were created or evolved using data from the 2001 Census.

The spine for Statistics Austria's register-based census is their Central Population Register (CPR). This is linked to six other base registers – the Housing Register of Buildings and Dwellings, the Business Register of Enterprises and their Local Units, the Register of Educational Attainment, the Central Social Security, the Tax Register, and the Unemployment Register. These registers are backed-up by seven comparison registers that are used to assure the quality of the census results (eg via cross-checks), and to supply additional attribute information not available or only partly available in the base registers.

Key coverage issues

Key coverage issues that Statistics Austria faced in setting-up its register-based census, and the approaches it developed to deal with these issues, included:

- overcoverage in the CPR due to duplicates, 'list inflation' etc. This is minimised via a complex 'analysis of residence process' that includes checking the CPR for multiple records, deaths, and signs of life, and in some cases sending out a letter to confirm whether the person was a usual resident as at the reference date. Erroneous inclusions identified by this process are excluded from the census counts
- undercoverage in the CPR due to time lags in records being updated, for example births and changes of address. This is avoided by using CPR data extracted six months after the census reference date, which allows additional time for records to be updated
- undercoverage of foreigners without a residence permit and certain groups of homeless people. This is noted as an issue with the register-based census, but there are no specific processes or coverage assessment measures in place to address the issue.

Austria's register-based 2011 Census did not include independent assessment or measurement of coverage errors (eg via a post-enumeration survey), though neither was this a feature of their full-enumeration census before 2011 (UNSD, 2010). Presumably

they are confident that the base CPR, coupled with the approaches they have put in place to address the known coverage issues with this, results in population counts that are sufficiently accurate for their purposes.

Sources: Statistics Austria (2012a & 2012b).

Germany

The Federal Statistical Office of Germany based its 2011 Census on a new model. This was Germany's first census in nearly 25 years following the last traditional census in 1987.

The 2011 Census was a combined census model, consisting of linked administrative registers – including population registers (decentralised), the employment register, and a register of addresses and buildings (set up in 2008 in preparation for the 2011 Census) – supplemented by surveys (both full-coverage and sample surveys), including a census of buildings and housing (postal survey to all owners).

The centralised population register serves as the spine and primary base for the population counts, with data from the other registers and supplementary surveys linked to it. One of the biggest challenges for the Federal Statistical Office of Germany was that this linking has to be done without the use of unique person or building identifiers. Instead, exact and probabilistic linking is done using auxiliary variables such as address, name, sex, and date of birth.

Key coverage issues

Key coverage issues that the Federal Statistical Office of Germany identified in developing its new census model included:

- poor data quality and a high level of coverage error in the population register for special facility addresses – for example, prisons, psychiatric hospitals, care homes, and student residences
- strong variation in coverage error rates for municipalities of differing population sizes, with smaller municipalities having higher quality registers and lower rates of over- and undercoverage than larger municipalities
- coverage errors due to combining decentralised population registers – each subject to time lags in being updated – into one central register. This means that a person may be counted several times (overcounting) or not at all (undercounting), in the wrong place, or with a wrong residence status (eg emigrants)
- discrepancies identified at the address level between the number of inhabitants in the population register (following a household generation process), and the number of occupants indicated by the property owner in the census of buildings and housing
- remaining coverage errors in the population register even after adjustment procedures are applied. In particular, a relatively large variance in overcount rates between municipalities.

The Federal Statistical Office of Germany carried out a range of approaches to manage and measure these coverage issues, including:

- full-enumeration of special facility addresses as part of the new census model
- matching the combined population register to itself, to identify and correct for people with duplicate / multiple entries
- comparing population register data for two different reference dates to identify and correct for 'temporary' overcount
- a supplementary survey in small municipalities (less than 10,000 inhabitants), limited to addresses with only one occupied living quarter, in which a discrepancy was identified in the number of inhabitants / occupants. Based on the responses, corrections are made to the register for surveyed addresses where needed
- a household sample survey in large municipalities (more than 10,000 inhabitants) to both assess coverage of the population register, and cover variables and population groups not contained in the registers.

One of the aims of the household sample survey was to measure the extent of overcoverage and undercoverage in the population register. The sample frame – restricted to municipalities with more than 10,000 inhabitants – consisted of addresses with housing space and occupied living quarters on the central register of addresses and buildings.

A random sample – stratified by geographic region and size of household – was taken, covering nearly 10 percent of the population. Selected addresses were surveyed to identify existing occupants who should be counted in the census. The individual level survey data was then linked to records on the population register at the same address to identify matching persons (up-to-date entries), missing persons (undercount) and non-active persons (overcount).

Based on the results of the household survey, the population register was adjusted in two ways:

- Population register data was corrected for the sampled addresses at which people were identified as being over- and/or undercounted.
- Coverage error estimates from the sample survey⁷ were also used to correct population register data across all large municipalities (ie extrapolating coverage results from the sample to the whole survey population).

In both cases, overcoverage was corrected by deleting people from the population register, and undercoverage was corrected by imputing people into the population register. The driver behind this approach was the need to output a person-level census dataset, in order to carry out planned analysis and also to meet user demands.

German inter-censal population estimates are currently based on the latest census, updated for births and deaths, and migration (arrivals and departures). In the medium-term, the Federal Statistical Office will attempt to replace this with regular evaluations (at least once a year) of the population registers, although this is dependent on certain pre-conditions being met (eg register coverage errors being within tolerable limits).

Sources: *Bechtold, S (2013); Federal Statistical Office of Germany (2012 & 2013).*

⁷ Corrections were made based on the marginal distributions of estimated register coverage errors for demographic core variables, and variables of employment statistics. Adjustments to the distribution of variable combinations (eg age group by sex) were also done where register errors at that level could be estimated with sufficient accuracy.

Israel

The Israeli Central Bureau of Statistics conducted its first integrated census based on administrative data in 2008. Before then, Israel had conducted five population censuses – the last in 1995 – using traditional census methods, with full enumeration of dwellings, a short-form completed for all residents, and a long-form completed by a 20 percent sample of households.

The 2008 Integrated Census used a combined census model. This comprised a register-based census using Israel's Population Registry, plus full-enumeration of the institutional population, supplemented with two interlaced but independent sample surveys.

The census was designed to:

- collect detailed demographic, social, and economic information (ie long form information)
- evaluate the quality of the population registry, plus calculate measures of under- and overcoverage that can be used to correct population estimates based on the registry.

Key coverage issues

Key coverage issues that the Israeli Central Bureau of Statistics identified in its population registry included:

- national undercoverage: it does not include people who do not have ID numbers who are residents in Israel continuously for a year or longer, legally or illegally – for example, foreign workers
- national overcoverage: it contains emigrants who no longer live in Israel, and some unrecorded deaths, due to time lags in updating the registry
- local under- and overcoverage: due to the poor quality of the address information, most coverage errors occur at the local level, with approximately 25 percent of people recorded in the population registry at an address different from the one where they actually reside (ie people counted in the wrong place).

The Israeli Central Bureau of Statistics did not estimate or adjust for national undercoverage – it was assumed that anyone not on the population registry (ie people without an ID number) should not be counted in the census.

To deal with the remaining coverage issues (national overcoverage and local under- and overcoverage), the ICBS first linked in address information from other administrative data sources to update and improve the quality of the population registry address data. This resulted in what is called the Improved Administrative File (IAF).

They then carried out two interlaced but independent sample surveys to estimate national overcoverage and local under- and overcoverage in the IAF, and also to collect long-form information.

To estimate coverage, the Israeli Central Bureau of Statistics used a modified capture-recapture (dual system estimation or DSE) method, which was extended to include an evaluation of overcount (the traditional DSE method is designed to evaluate undercoverage, and assumes that overcoverage is minimal).

The sample design involved two independent samples focused on the same enumeration areas:

1. a sample of enumeration areas to estimate undercoverage (U sample). Approximately 17 percent of the population were enumerated, and surveyed to collect long-form information.
2. a sample of individuals from the Improved Administrative File (IAF) to estimate overcoverage (O sample). Approximately 20 percent of the population were selected from the IAF – people with addresses registered in the same enumeration areas as sampled in the U sample, plus a sample of individuals with address not geo-coded.

The result was two lists of people for each sampled enumeration area, generated independently – people listed in the IAF at the addresses in the enumeration area (O sample), and those actually found in the enumeration area during the U sample field enumeration. These lists were linked after the U survey was conducted, via exact and probabilistic linking methods especially developed for the integrated census.

For people on the IAF list, but not enumerated in the field (ie not matched to a person in the U sample list), additional follow-up was then done in the field or via phone. These people were traced, located, and surveyed to confirm their presence (or not) and location in Israel.

Once national overcoverage was estimated and accounted for, a generic DSE model was used to estimate local undercoverage / overcoverage (which balance across the country) for each statistical area. Based on these coverage estimates, a census weight was then generated for each record (person) in the IAF, and these weights could then be used to produce estimates for any population group.

During 2012, the ICBS began an integrated rolling census for the first time. The integrated rolling census approach is based on the 2008 Integrated Census, but with census operations carried out over a 10-year period (similar to France) and several modifications – including a reduced sample size for the two coverage surveys, and use of an enhanced labour force survey to help in the estimation of undercount.

However, due to methodological problems identified upon examination of the estimates, data collection for the integrated rolling census was suspended in 2013 and the census method is being re-evaluated. In the meantime, inter-censal population estimates are being updated the same way as previously, using 2008 integrated census data as a base.

Sources: Central Bureau of Statistics (2012 & 2013); Elkin et al (2012); Zadka, P (2012).

Singapore

In 2000, the Singapore Department of Statistics adopted a combined census model for the first time, moving from a de facto to a de jure population concept, and continued with this approach in 2010. Before this, the department ran a traditional full-enumeration census every 10 years, with developments since the 1980s allowing them to increase the use of administrative data in the census over time.

In Singapore's combined census model, basic population counts and characteristics are compiled from administrative registers (in census and inter-censal years), and supplemented with a sample survey (approximately 20 percent of households) that collects additional information not available from administrative sources (in census years only).

The basic population data is extracted from two registers:

- Household Registration Database (HRD) – a population register containing basic personal information on Singapore citizens and permanent residents (set-up using information sourced from administrative registers and the 1990 Census)
- National Database on Dwellings (NDD) – a register of all residential addresses.

Together, the HRD and NDD give a physical location for every person and household in Singapore.

Key coverage issues

During development of the register-based census for 2000, various data coverage issues were identified. These were primarily due to time-lags in households and individuals notifying the relevant administrative agency of changes – for example, change of address, demolition of existing dwelling units, and construction of new dwelling units. Some instances of 'list inflation' were also identified – for example, the department found that the elderly in the population were overcovered in administrative data sources compared with their official population estimates.

These coverage issues were mostly resolved during development of the 2000 Census by:

- comparing population counts compiled from the administrative registers with the official population estimates produced from the full-enumeration census
- matching and verifying census data with other administrative registers – for example, to update addresses in the HRD and NDD
- conducting additional fieldwork or small-scale surveys before the census year to confirm the presence of people and dwelling units, and update the registers (HRD and NDD)

This additional fieldwork carried out before the census year included a 'focused' census for people aged 75 years and over, which was done via mail questionnaires in 1998/99 to verify the addresses and locations of the elderly. The exercise confirmed that some of the elderly had either migrated, passed away, changed address, or were residing in facilities for seniors. The returns were used to update the HRD, and therefore produce more accurate counts of the elderly population before the Census 2000 launch.

Additional fieldwork before the census also included surveys to update the physical status of dwelling units in the NDD, including fieldwork to check demolitions, and contact with major property developers.

However, some of the processes described above have been continued – either on a regular or ad-hoc basis – to preserve the integrity of the registers (HRD and NDD), and ensure good quality data. The Singapore Department of Statistics also continues to work with various agencies to resolve data issues, including data coverage, for both existing and new data items.

Singapore's combined census does not include independent assessment or measurement of any remaining coverage errors (eg via a post-enumeration survey). The Singapore Department of Statistics is confident the quality of the resulting administrative data is sufficiently high to produce an accurate count of the population and its basic characteristics.

Sources: Singapore Department of Statistics (2002, 2003, & 2009).

Switzerland

The Swiss Federal Statistics Office conducted its last 10-yearly full-enumeration census in 2001, moving to the new Swiss Census System in 2011. This new system uses a combined census model, in which existing registers are supplemented with annual sample surveys. The Swiss census is conducted and evaluated on an annual basis.

The existing administrative registers used are the population registers of the various Swiss communes and cantons, and key federal person data registers, and the residential buildings and dwellings register (the Federal Register of Buildings and Dwellings, built with data from the Buildings and Dwellings Survey of the 2000 Census).

Three annual sample surveys collect necessary information not contained in the registers, each with a specific focus, and designed to provide results to different levels of granularity. In total, about 5 percent of the population is surveyed in writing or by telephone each year. Person-level data from the registers and sample surveys is linked using a unique social security number.

Key coverage issues

Key coverage issues that the Swiss Federal Statistics Office identified during development of its new census, and the approaches it used to deal with these issues, include:

- Overcoverage due to people, particularly foreign nationals, appearing in several registers. This is addressed in two ways:
 - By prioritising the source registers (based on their timeliness, completeness, reliability, and other factors) and then, for each category of persons, selecting the source on which the number of people to be counted is to be based.
 - By using the social security number to link the records of one individual in different sources, in order to define the target population for the census, and to avoid duplicates.⁸
- Under- and overcoverage due to time lags between the point in time when an event (birth, death, migration) occurs, when it is reported to the authorities, and when it is updated in the relevant register(s). To the extent possible, the numbers are adjusted based on information received afterwards (ie after the data has been sent to the SFSO for the census reference day).
- Even after taking the above steps, some under- or overcoverage is still expected to remain in the register(s) for some sub-populations, though these may differ from the sub-populations undercovered / overcovered in Switzerland's traditional census. In order to assess this, the SFSO is conducting the first independent coverage survey of its new census in 2013, with results available in 2014.

The coverage survey will measure the quality and comprehensiveness of the register-based component of the new census. As in their 2000 census, which was the first census for which the SFSO had run a post-enumeration survey, the survey procedure will be based on a capture-recapture / dual system estimation method.

For the 2000 census, this involved two separate components – one to estimate undercoverage, using an independent sample survey matched back to the census (as in New Zealand's current PES), and one to estimate overcoverage, matching a sample drawn from the census list to the rest of the census list to identify duplicates and other erroneous records. The coverage analysis was supplemented by a study of matches from the undercoverage component to work out potential measurement, classification, and location errors in the census data.

For the 2013 coverage survey, the SFSO will be adapting this procedure to account for the experience gained from the 2000 Census, and the new Swiss census system introduced in 2011. However, at the time of our last correspondence with the SFSO, the specifics of the 2013 coverage survey estimation approach were still being finalised.

⁸ If unable to match on social security number (eg due to a number not yet being assigned to a person, or needing to be corrected), the SFSO attempt to match on other identifiers – name, date of birth, address. The processing of remaining unmatched cases is very time-consuming.

From the coverage survey, the SFSO plans to produce a report on the quality and comprehensiveness of the register-based component of their combined census, including under- and overcoverage estimates. However, they do not intend to use these estimates to correct or adjust the population counts – the published register-based counts will remain the official population counts.

Sources: Renaud, A (2011); Swiss Federal Statistics Office (2011 & 2012).