



# Comparing income information from census and administrative sources

Census Transformation

Sandy Swei



**Crown copyright ©**

This work is licensed under the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) 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](#). Use the wording 'Statistics New Zealand' in your attribution, not the Statistics NZ logo.

### **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.

### **Disclaimer**

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics NZ. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the author(s), not Statistics NZ.

Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation. The results in this paper have been confidentialised to protect these groups from identification.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the [Integrated Data Infrastructure extension: Privacy impact assessment \(6th ed\)](#) available from [www.stats.govt.nz](http://www.stats.govt.nz).

Note: All figures presented in this paper have been rounded to the nearest hundred to protect confidentiality.

### **Citation**

Suei, S (2016). *Comparing income information from census and administrative sources*. Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).

ISBN 978-0-908350-76-6 (online)

### **Published in October 2016 by**

Statistics New Zealand  
Tatauranga Aotearoa  
Wellington, New Zealand

### **Contact**

Statistics New Zealand Information Centre: [info@stats.govt.nz](mailto:info@stats.govt.nz)  
Phone toll-free 0508 525 525  
Phone international +64 4 931 4600  
[www.stats.govt.nz](http://www.stats.govt.nz)

# Contents

<b>1 Background</b>	<b>5</b>
Census Transformation in New Zealand	5
About this paper	5
<b>2 Introduction</b>	<b>6</b>
Census income information	6
Other sources of income information	6
Aims and scope	7
<b>3 Method</b>	<b>8</b>
<b>4 Statistical standard and classifications for income</b>	<b>10</b>
Sources of personal income	10
Total personal income	10
<b>5 Data sources</b>	<b>12</b>
New Zealand Census of Population and Dwellings	12
The Integrated Data Infrastructure (IDI)	13
Populations used in analysis	15
Income information in the IDI – Inland Revenue data	16
Derivation of income information in the IDI	17
<b>6 Results</b>	<b>20</b>
Comparison of concepts and definitions	20
Coverage differences	21
Results for total personal income	21
Results for sources of personal income	24
Cross comparisons	26
<b>7 Conclusion</b>	<b>30</b>
Summary	30
Limitations	31
Discussion	31
<b>8 References</b>	<b>32</b>
<b>9 Appendixes</b>	<b>33</b>
Appendix 1: Income source classification – definitions	33
Appendix 2: Income bands classification	33
Appendix 3: Census 2013 income questions	34
Appendix 4: Counts and percentages by income band for census and tax data	35

# List of tables and figures

## List of tables

1	Quality information for the income variables, 2013 Census .....	13
2	Variables relating to deriving income information available in the IDI as at May 2015 compared with that collected by 2013 Census .....	18
3	Count and percentage of each individual source of income recorded in 2013 Census and IDI-URP .....	24
Appendix table 1	Income bands classification.....	33
Appendix table 2	Counts by income band for census and tax data.....	35

## List of figures

1	Structure of the Integrated Data Infrastructure in May 2015 .....	15
2	Percentage with non-missing total personal income, by age, 2013 Census and IDI-URP .....	22
3	Distribution of income bands, 2013 Census and IDI-URP.....	22
4	Difference in income bands, 2013 Census and IDI .....	23
5	Distribution of income source, 2013 Census and IDI-URP.....	25
6	Percentage agreement with census income source, IDI tax-derived income source categories .....	26
7	Distribution of census income bands, for those in IDI income band \$15,000–20,000 receiving NZ Superannuation in IDI and investment income in census .....	27
8	Distribution of income bands for wage and salary earners, 2013 Census and IDI .....	28
9	Distribution of income bands, for census zero income and IDI-URP missing income .....	29
Appendix figure 1	Census 2013 income questions .....	34

# 1 Background

## Census Transformation in New Zealand

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 making it more efficient.
- A longer-term focus on investigating alternative ways of producing small-area population and social and economic statistics. This includes the possibility of changing the census frequency to every 10 years, and exploring the feasibility of a census based on administrative data (Statistics New Zealand, 2014a).

The next census in 2018 will be significantly modernised, including an online completion target of 70 percent and using administrative data to support collection and processing.

Continuing to meet critical information needs must underpin decisions on the future of census. Investigations into the long-term direction for census are focused on developing an understanding of future census information requirements, and the ability of administrative data sources to meet those requirements.

[Read more about Census Transformation in New Zealand.](#)

## About this paper

This paper is one of a series of investigations by the Census Transformation programme to identify and explore the potential for administrative data sources to provide census-type information.

One important reason for having a census is to provide information about the characteristics of the population, including information about total income and source of income. Previous Census Transformation work has identified the potential for Inland Revenue tax data to provide census-type income information.

This paper compares information about 'total personal income' and 'sources of personal income' from the 2013 Census with experimental estimates produced from the tax sources available in Statistics NZ's Integrated Data Infrastructure (IDI).

Overall, the administrative sources show good potential for providing income information for those who have interacted with the New Zealand tax system. Income information derived from the tax data is more precise than that obtained through the census questionnaire. However, most investment income and non-taxable income information has not been available in the IDI for this investigation, and would need to be included to provide a more complete picture of total personal income. With the administrative data, it is not possible to positively identify those earning zero income.

The comparisons in this paper provide a better understanding of the quality of the administrative data available about income, and contribute to Statistics NZ's work to transform the census model.

## 2 Introduction

This paper describes our preliminary analysis of total personal income and source of income information available from administrative sources. It builds on findings from O’Byrne et al (2014) in the context of future censuses.

### Census income information

Income information currently collected by the census has a range of uses important to social and economic policy. For example, income is used as an indicator for poverty statistics and as a component of the New Zealand Index of Deprivation, and the Ministry of Social Development uses income as an indicator for social assistance policy settings. Household income is an important measure derived from personal income, and is used for income inequality studies.

Census income information consists of:

- sources of personal income
- total personal income
- income sources and total income for households and families.

Census income variables are available for small geographic areas and can be cross-classified by information about the individuals in the household, and about the dwelling. The census is unique in its ability to provide these kinds of detailed breakdowns and information for small geographic areas across the whole country.

### Other sources of income information

Although Household Economic Survey (HES) and New Zealand Income Survey (NZIS) produce more detailed information about income, they are designed primarily to produce national-level data and are limited in the amount of regional detail they can provide.

The concept of using administrative data to produce income statistics is not new in New Zealand. Statistics NZ has published official income statistics using the linked employer-employee data (LEED) since 2004. The LEED dataset is created by linking a longitudinal employer series from the Statistics NZ Business Frame to a longitudinal series of Employer Monthly Schedule (EMS) payroll data from Inland Revenue ([see Linked Employer-Employee Data](#)).

However, there are several differences between the LEED data and census income information. LEED income sources do not include investment income or non-taxable income. Aggregated earnings are published only for those in paid employment, and outputs are produced at the territorial authority level, but not for smaller geographic areas. The LEED population includes all those with taxable income over the period of a year, while the census includes only those residents in New Zealand at a given date (census day). Finally, LEED does not include standard demographic variables such as ethnicity or qualifications, or income for households and families, which are provided by the census.

[User guide for wage and income measures](#) has more information about the design and purpose of the several income and wage measures produced by Statistics NZ.

In a first broad look at the potential for administrative data to produce the social and economic information currently provided by the census, O’Byrne et al (2014) assessed ‘personal income’ and ‘income source’ as likely to be satisfied by administrative data. The main source identified was the tax data from Inland Revenue (IR). This first assessment

using scores against five quality measures was based on metadata and intended to be indicative only.

## Aims and scope

Our overall aim for this investigation was to analyse to what extent census income information can be derived from existing administrative data. Three main research questions guided this work:

- What is the quality of income information in administrative data?
- To what extent can the data requirements for income attributes currently obtained from census be satisfied using available administrative data?
- What would be required to improve the potential for administrative sources to estimate census income variables?

This report provides reference information about the statistical concepts and administrative data sources that are relevant to personal income and source of income. It presents findings from analysis comparing income information in the census with income information derived from linked administrative data sources.

Our investigation focused on the potential for administrative data to provide the same income information as produced by the census, with the 2013 Census as the reference. We restricted the categories of income sources to those in line with the statistical standard. We recognise that income information needs may change over time, and that the census may not fully meet existing needs.

The administrative sources we considered were limited to those available in Statistics NZ's Integrated Data Infrastructure (IDI) as at May 2015. These sources include all the main government sources related to income, and allow estimates of income variables to be derived. We have limited this study to taxable income obtained from Inland Revenue data. We are currently developing methodologies to include non-taxable social transfers into the estimates, but do not report on this here.

We limited our analysis to the national level by age and sex, with no breakdown by ethnicity or geography. To be useful in the context of census, income information derived from administrative data needs to be linked with other data on demographic and other census characteristics. The potential for administrative data sources to produce other census variables is discussed in other work (O'Byrne et al 2014, Shrosbree 2015, Bycroft et al 2016). In particular, the quality of geographic information in the IDI is examined in Gibb and Das (2015), and the quality of ethnicity information in the IDI is examined in Reid et al (2016).

We have also excluded household income from this analysis, as the available administrative data does not currently allow us to confidently determine the construction of households (Gibb and Das, 2015).

The estimates of total personal income and income sources we derived using administrative sources are experimental and are for the purpose of comparison with census information only.

## 3 Method

The method used to evaluate the potential to produce income information from administrative data sources, specifically the IDI, involves:

1. Describing the formal statistical concepts (section 4) relevant to income used in official statistics. Statistical standards and classifications provide the concepts and definitions against which both census and administrative sources are compared.
2. Describing the data sources (section 5) used in this investigation – 2013 Census and the IDI – and developing a method for deriving estimates of ‘personal income’ and ‘income source’ from the data available in the IDI.
3. Comparing the two data sources at three levels (section 6).
  - i. **Comparison of concepts and definitions.** We compare the concepts and definitions used in the census and the IDI, and identify differences. These comparisons are guided by the descriptions of each data source.
  - ii. **Comparison of aggregate counts and estimates.** We compare census results for income information with their IDI equivalents.
  - iii. **Comparison of individual-level information.** We compare census responses to the income questions with the information available in the IDI.

The aggregate-level comparison is based on the census usual resident population (Census-URP) at 5 March 2013 and an equivalent population constructed from the IDI (the IDI-URP).

The individual-level comparison is restricted to anonymised individuals in a dataset constructed from linking the 2013 Census to the IDI.

The concepts of coverage error and measurement error provide a framework for assessing the accuracy of data sources (Statistics NZ, 2016; Zhang, 2012).

Coverage relates to the target population from which observations for a particular topic can be drawn. For census income variables, the population of interest is people aged 15 years and older (15+) in the New Zealand resident population. Understanding administrative sources and the aggregate-level comparisons are most useful in providing insight into differences in coverage.

Measurement errors occur when the response provided differs from the targeted measure. Such errors may be random, or they may result in a systematic bias if they are not random. Validity errors may occur when administrative definitions and concepts do not align well with the statistical concept being measured. Measurement errors in the census and administrative data may also be due to errors in collection and processing systems, and may result in missing or incorrect information. The individual-level comparisons can inform our understanding of measurement error.

Close agreement of responses in administrative data and the census provides strong support for good measurement in both sources. However, when we get different responses, it is harder to determine which is more likely to be correct. This will depend on a range of factors and requires a deep understanding of the mechanisms underlying the particular administrative data collection, and of how people respond to survey questions.

Being able to integrate information with other sources by linking the same units also affects accuracy. It can result in linkage errors, which are of two types:

- links may be missed, for example if the name of a person is recorded differently on different files
- two different people may be wrongly linked, for example if their names and dates of birth are very similar.

Linkage errors may reduce the effective coverage of an administrative source, as no income information is available if links are not made when they should be. If the wrong people are linked together, this may inflate measurement errors if each person has a different response.

## 4 Statistical standard and classifications for income

Statistical standards and classifications (Statistics NZ, nd) provide definitions for the key concepts in this investigation. These statistical standards and classifications are designed for use in official statistics collections and are those used in the 2013 Census.

[Statistical standard for income bands](#) describes the key concepts, definitions, and classifications for 'sources of personal income' and 'total personal income' as measured by the census. The concept currently used to collect income band information is the gross annual income.

### Sources of personal income

The census variable 'sources of personal income' identifies the various sources from which individuals aged 15 years and over (15+) received their total personal income in the 12 months preceding census day.

In the census, it is only realistic to collect information on money income. This is income that the individual respondent can recall or retrieve from their financial records. Money income is money flow from the deployment of one's labour or entrepreneurial skills, assets, and transfers received. The concept of money income therefore relies on identifying the sources from which money income is derived.

Excluded are income in kind, imputed income, unrealised income, contingent income, tax credits, and reimbursements of expenses.

Seven categories for income sources are stated in the glossary for the statistical standard for income bands:

- Wages and salaries
- Self-employment income
- Investment income
- Private superannuation income
- New Zealand Superannuation and war pensions
- Other government benefits
- Other sources of regular and recurring income.

The definitions of these categories are given in Appendix 1.

### Total personal income

Total personal income in the census represents the before-tax income for the respondents in the 12 months ending 31 March.

Income bands are a convenient and alternative way of collecting income information when income is not the prime focus of a survey, as is the case for the census. Income collected as an income range rather than an actual dollar income is designed to overcome collection difficulties for what has historically been a sensitive topic for respondents. Bands reduce respondent burden by providing a simple way to obtain contextual information about income, while providing statistics on the distribution of income across the population.

Tax data collects income as actual dollar values, but these may be output as income bands.

The income bands within the classification are determined by the analysis of income data collected by Statistics NZ's detailed income collections. This analysis identifies emerging

income trends in areas such as middle and upper income earners, and benefit levels. These bands are reviewed periodically to remain relevant to societal trends. The first two categories reflect income loss and zero income; there are 14 income bands starting from \$1 to \$5,000, and moving up to \$150,001 or more; and five residual categories, eg 'not stated'. The income bands classifications used by the 2013 Census are given in appendix 2.

## 5 Data sources

This section describes the two data sources compared in this investigation: the New Zealand Census of Population and Dwellings, and Statistics NZ's Integrated Data Infrastructure (IDI). We describe the construction of the Census-IDI linked dataset via linking the 2013 Census to the IDI and the populations used in this analysis. Specifically, we describe the information about income in each source.

### New Zealand Census of Population and Dwellings

The Census of Population and Dwellings is the official count of people and dwellings in New Zealand. It provides a snapshot of New Zealand at a point in time, and measures social and economic change. Census information has a wide range of uses within and outside government. Historically, the census has been held every five years, with some exceptions. The latest census was held in March 2013.

[Find more information about the census.](#)

While the census aims to count everyone who is in New Zealand on census night, for this investigation, we are only interested in New Zealand residents aged 15+, as income information is only collected from this population.

#### Income information in the census

The statistical standard for income information currently produced by the census is as described in Section 4: Statistical standard and classifications for income. The income questions asked in the 2013 Census are shown in Appendix 3: Census 2013 income questions. Information on income sources was first collected in 1981 to focus respondents on providing accurate total personal income, but has since become useful in its own right.

The key limitations of census income information are the inability to determine the amount received from each source, at which period, or for how long. The broad income bands are also not ideal for complex analysis, although this was not the intended purpose.

Item non-response and substitute records (records representing individuals for whom no census forms were received) collectively make up total 'non-response' to the income questions.

Table 1 presents the quality information for the income variables in the 2013 Census (Statistics NZ, 2013). A non-response rate over 10 percent is considered relatively high. Substitute records contribute to 4.9 percent of the non-response rate.

**Table 1**  
**Quality information for the income variables, 2013 Census**

Variable	Quality rating	Item non-response rate (%)	Substitutes (%)	Non-response rate (%)
Sources of personal income	High	2.3	4.9	7.2
Total personal income	Moderate	4.8	4.9	9.7
Household income	Poor	-	-	15
Family income	Moderate	-	-	11.2
Source: Statistics New Zealand				

The highest non-response rates were for:

- the Pacific ethnic group
- those not in the labour force
- people aged 24 years and under
- people aged 70 years and above.

This non-response by individuals flows through to the ability to derive household income, and results in a 15 percent non-response rate for household income.

Item non-response is possibly due to a combination of factors including not understanding the question, ability to recall, and respondent burden. Those not working may consider this question irrelevant. For some respondents, there may be an unwillingness to answer what may be considered intrusive questions.

However, there is a marked difference in non-response between modes, with 6.9 percent item non-response for paper forms and a much smaller 1.4 percent for online forms. While this difference may be partly due to the characteristics of people who responded online, it may also suggest that difficulties with the paper questionnaire may be more important factors than unwillingness to respond.

[2013 Census QuickStats about income](#) provides more detail on 2013 Census income results and data quality.

## The Integrated Data Infrastructure (IDI)

Statistics NZ developed the IDI as an environment in which to link multiple data sources in a systematic and secure way. It was developed to produce official statistics and to allow Statistics NZ staff and external researchers to conduct policy evaluation and research on people's transitions and outcomes. The IDI contains de-identified administrative and survey datasets, linked at the individual level. This research is used to better inform decision-makers to improve outcomes for all New Zealanders. The IDI continues to evolve as new datasets are added.

[Read the latest information about the Integrated Data Infrastructure.](#)

Here we have used the IDI as a test environment for examining the potential of linked administrative data sources for producing income information.

This section describes the structure and content of the IDI as at May 2015.

The structure of the IDI (figure 1) can be described as a central 'spine' to which a series of data collections are linked. The spine forms the conceptual centre of the IDI and all other datasets are linked to it. Broadly, the target population for the spine is all individuals who have ever been residents of New Zealand.

Three data sources are linked together probabilistically to create the spine:

- a list of all IRD numbers that have been issued by Inland Revenue (IR)
- a list of all births registered in New Zealand since 1920
- a list of all visas granted to migrants from 1997 (excluding visitor and transit visas).

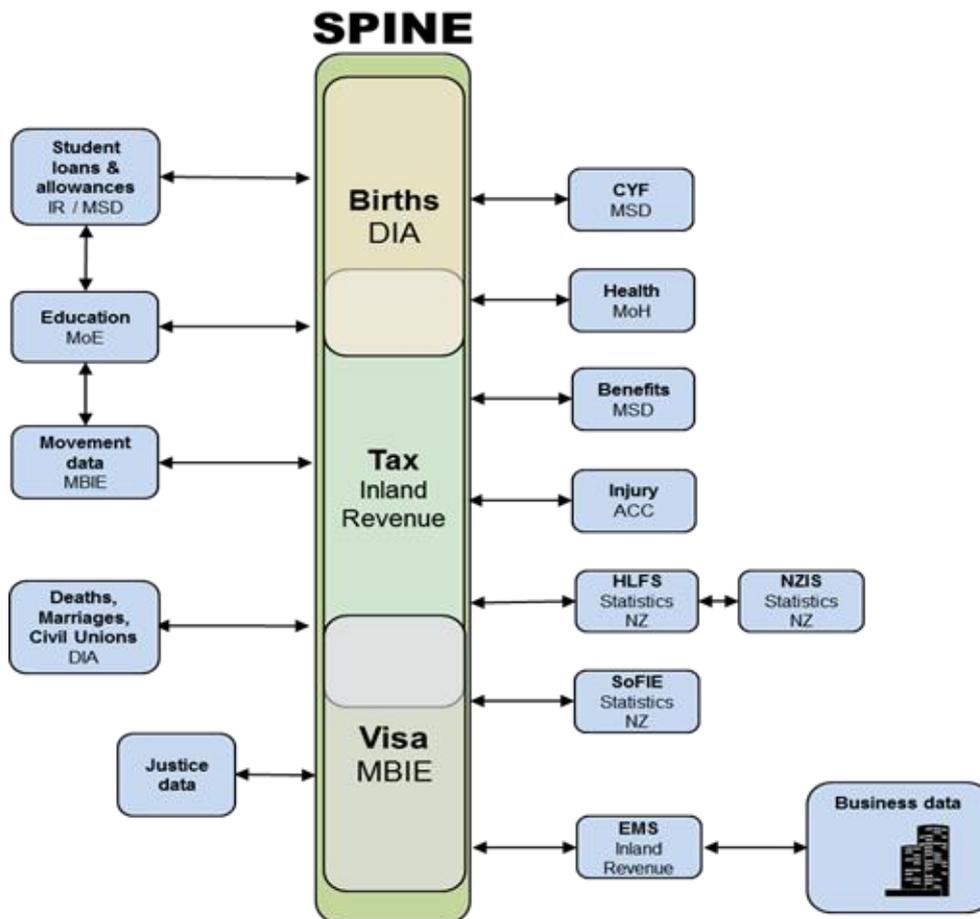
Other datasets are linked to the IDI spine; the linking methodology is described in Statistics NZ, 2014b. Priority is placed on obtaining a high precision rate, ie minimising creating erroneous links, with the trade-off that more correct links may be missed. In practice, linkages are designed so that under 2 percent of links made are erroneous.

The linked datasets cover a wide range of subject areas and include:

- employer and employee job and earnings information based on IR tax data
- health information including general practitioner enrolment and hospital visits from the Ministry of Health
- education data from the Ministry of Education
- benefit dynamics data from the Ministry of Social Development
- student loans and allowances data from several sources
- migration movements data from the Ministry of Business, Innovation and Employment
- the Household Labour Force Survey and New Zealand Income Survey data from Statistics NZ.

The IDI also contains several summary tables that provide core information about individuals (age, sex, ethnicity, and geographic information) summarised from available data sources.

**Figure 1**  
**Structure of the Integrated Data Infrastructure in May 2015**



Source: Statistics New Zealand

Values for the income variables can be derived from relevant information within the data sources linked in the IDI. In this investigation, we used the tax year summary table made from only IR data to derive these values. We note that because tax registrations form part of the spine, no linkage error should be introduced by linking tax records directly to the spine through the tax number. In 2013, there was around 0.1 percent missing information of the wages and salaries earners for whom no tax number is available (unpublished results from LEED).

The following sub-sections describe the relevant information used in this derivation process:

1. the populations selected for comparison
2. the data sources and variables we used to derive values for total personal income and source of income.

## Populations used in analysis

To obtain the most relevant comparison between the census and IDI information, we used different populations of IDI and census records for different parts of the analysis in this investigation.

### The IDI usually resident population (IDI-URP)

The aggregate-level analyses in this investigation used only individuals in the IDI aged 15+ who are identified as New Zealand resident, and present in New Zealand on 5 March

2013 (census day). As well as selecting only New Zealand residents, we also used migration data to remove New Zealand residents who were temporarily overseas on the census night. That is, the definition we used for the administrative sources is the same definition as the census usually resident population.

These individuals form an experimental population, referred to here as IDI-URP, which aims to reflect as closely as possible those who would also be included in the definition of the census usually resident population count for the purposes of valid comparisons. The IDI-URP is derived from the IDI spine using information on recent activity from tax, health, education, and ACC data, and removing those who have left the population due to migration or death. More detail about the selection of a resident population from the IDI is available in Gibb et al (2016).

Given that receipt of income is one of the criteria for inclusion in the IDI-URP, the coverage of this group should align well with income variables measured in this investigation. Comparisons between the census and the IDI-URP will however be affected by census undercount, and by people who are incorrectly included or excluded when deriving the IDI-URP. Net census undercount is estimated at 2 percent, and there is an estimated net overcoverage of 2 percent in the IDI-URP.

## **The Census-IDI linked population**

The 2013 Census was linked to the IDI spine as at May 2015 for the purposes of understanding possible future census models (Statistics NZ, 2014c). This linked census data was used by our Census Transformation team only and not available to other IDI researchers at this time. Records were linked probabilistically using name, date of birth, sex, usual residence, and country of birth. The individual-level analyses in this investigation use this linked Census-IDI dataset.

This linked population includes all census records for usual residents aged 15+ for which a suitable link in the IDI has been found. An IDI link was found for 93 percent of the 3,376,400 census usual residents aged 15+. Some 13,500 individuals who responded to the census income question (0.4 percent of usual residents aged 15+) could not be found in the IDI.

In some cases, a link may be made between two different individuals. This is estimated to have occurred for less than 1 percent of the links made between the census and IDI. This means that linkage error could explain a small proportion of cases where income information is found to be different between the census and the IDI.

## **Income information in the IDI – Inland Revenue data**

Inland Revenue (IR) is the most extensive source of income information in the IDI. This source is described here, and used to derive income variables from the IDI, which are later compared with the census.

Inland Revenue is the New Zealand government's revenue collection agency. The raw IR datasets in the IDI contain information about income from four main IR income tax return forms:

- Employer Monthly Schedule (EMS) provides gross earnings where PAYE is deducted at source. The EMS consists of all wage and salary earners, withholding payments, government transfer payments, and payments from ACC. It includes categories for government benefits, student allowances, paid parental leave, and New Zealand Superannuation payments. The EMS is filed monthly by the employers and provides pay details of employees who work for them.
- IR3 for self-employment (filed annually by sole traders) which includes non-zero partnership, self-employment, or shareholder salary income, as well as rental income.

- IR4S filed by companies includes remuneration income paid to shareholders, directors, and relatives of shareholders (filed annually)
- IR20 (formerly IR7) for partnership and look-through companies (filed annually).

Income information can be estimated using the IR-derived tables recently developed for the IDI. There is one IR-derived table for each of these four sources of income information.

[See IDI Data Dictionary: Tax derived](#) tables for details about data coverage, quality, and business rules.

Inland Revenue holds information on income from taxable investments such as interest and dividends. Although this was not available in the IDI at the time of this study, activities are underway to include this data in the IDI, which will substantially address the issue of lack of investment income.

Some Inland Revenue datasets include information for some of those reporting zero income; however, this information was not available at the time of this study.

### **Tax year summary table**

Individuals are identified by the IRD number used across the tax system. This means that information about the same individual can be accurately combined from different sources of taxable income.

The tax year summary table used in this investigation is an IR-derived table as at May 2015. This table comprises all records combined from all the four aforementioned datasets that contain income information.

The records have been arranged into the granularity of one record per individual per year. This table orders the monthly data into tax years, so that April is effectively month 1 in this table. Total annual income is derived for the year 1 April 2012 to 31 March 2013, consistent with the census question.

Variables used in this analysis include the tax year, all income sources and total income (before tax) from each individual source. Sex and age information was sourced from the IDI central table.

### **Income information – other sources**

Some of the information collected by IR is also received from other government agencies such as claims from ACC; and benefits, student allowances, and paid parental leave from the Ministry of Social Development (MSD). Gross taxable income is available in the IDI from all of these sources. However, we have treated the IR tax data as the authoritative source of non-zero taxable income, and other agency sources have not been used to derive values for income variables. Details for these sources are included in table 2.

MSD also holds information about non-taxable government transfers such as the accommodation supplement. However, there was insufficient data available in the IDI to readily estimate values for any non-taxable income.

## **Derivation of income information in the IDI**

This section describes how we derive income information using the tax year summary table in the IDI.

## Income source

Information about income sources for individuals who are usual residents aged 15+ are directly extracted from the tax year summary table.

Some census income source categories are available directly from the tax data. New Zealand Superannuation payments, student allowance payments, and ACC payments are each obtained from the EMS, and are compared with the same census categories.

We note that veterans pensions are reported in the same tax data as the New Zealand Superannuation, but are not able to be differentiated. Paid parental leave (from the EMS) is the only 'other government benefit' available in the tax data in addition to the main working-age benefits listed below.

In other cases we have combined categories to create consistent groupings for comparison purposes. Income sources in the census and IDI were combined as follows:

- Wages and salaries, and withholding payments from the EMS are combined into 'wages and salaries'.
- 'Self-employment' income is a combination of income from annual tax returns: all income from the IR3 tax form for sole traders (with the exception of rental income), company director/shareholder income from the IR4S, and partnership income from the IR20. Income from the EMS is also included here where this has been identified as income paid to a sole trader or company.
- Rental income from IR3 is the only source coded as 'investment' income in the IDI.
- Unemployment, sickness, domestic purposes, and invalid's benefit in the 2013 Census are combined into one 'benefit' category. The same government transfers are identified from the EMS tax data under the category 'benefits'.
- Other categories in the census not included above are combined into one 'other' category. These are private superannuation, government benefits other than working-age benefits and student allowances, and other sources of income such as support payments from outside the household. These are not available in the IDI.

Table 2 shows the relationship between income sources as collected by the census and corresponding sources in the IDI.

**Table 2**  
**Variables relating to deriving income information available in the IDI as at May 2015 compared with that collected by 2013 Census**

Sources of income	IDI at May 2015	2013 Census	Other sources
<b>Wages and salaries (WAS)</b>	Wages and salaries, withholding payments, commissions, bonuses	Wages, salary, commissions, bonuses, etc, paid by my employer	Employer contribution to KiwiSaver from IR
<b>Self-employment income (SEI)</b>	Sole trader, company director/shareholder, partnership, commissions, bonuses	Self-employment, or business I own and work in	-
<b>Investment income (INV)</b>	Rent (IR3)	Interest, dividends, rent, other investments	Interest, overseas and other

			investments from IR
<b>New Zealand Superannuation (NZS) and war pensions</b>	New Zealand Superannuation and veterans pension	New Zealand Superannuation or veterans pension	-
<b>Private superannuation income</b>	-	Other superannuation, pensions, or annuities (other than New Zealand Superannuation, veterans pension, or war pensions)	-
<b>Other government benefits</b>	Benefit (BEN)	Unemployment benefit, sickness benefit, domestic purposes benefit, invalid's benefit	-
	Student allowance (STU)	Student allowance	-
	Paid parental leave (PPL)	Other government benefits, government income support payments, or paid parental leave	Non-taxable government benefits from MSD
<b>Other sources of regular and recurring income</b>	Accident Compensation Corporation (ACC)	Regular payments from ACC or a private work accident insurer	-
	-	Other sources of income, counting support payments from people who do not live in my household	-
<b>No source of income during that time</b>	-	No source of income during that time	Inactive records from IR

## Total personal income

Total personal income is derived from the tax year summary table by summing all income from each source. This gives an annual total income before tax for each person in the tax data. Actual dollar amounts are then converted into income bands using the statistical classification so they can be compared directly with the 2013 Census income bands.

The distance between income bands, rather than a dollar amount, is used for comparison. We calculate the distance between income bands by taking the difference between the numerical codes assigned to the income band for a respondent in the census and their tax-derived income band. A negative value indicates the personal income in the IDI is higher than that in the census for an individual.

## 6 Results

Results are presented in four subsections in line with the method:

1. comparisons of concepts and definitions between income information from the census and the IDI
2. coverage differences
3. comparisons of aggregate counts and estimates between the census and the IDI
4. comparisons of individual-level records.

We discuss first the concepts, definitions, and coverage of income sources and total income together. We then present results from analysis on ‘total personal income’, followed by those for ‘income source’. Finally we show some examples of analysis of income by income source that the detail provided in the tax data allows.

### Comparison of concepts and definitions

The concept of income source that both the tax-derived data in the IDI and the census attempt to capture is similar to the statistical standard – that is, various sources from which individuals received their personal income in a tax year. There are differences in the range of income sources being captured, as listed in table 2.

The definition of total personal income in both the tax-derived data in the IDI and what the census attempts to capture are similar to the statistical standard for income bands – that is, the before-tax income over a 12-month period. The most obvious difference between the two sources is that census captures personal income in bands, while tax data contains the actual dollar amount. The actual dollar amount is easily aggregated to the income bands used by the census.

The census is a self-completed questionnaire, while the IR tax data records formal interactions with the tax system. These different collection methodologies may lead to differences in the estimates of income sources and total personal income.

- In the census, we rely on the respondent’s correct interpretation of the income source question, their correct recall, and correct identification of the source(s) of income, whereas the IDI provides the official records of these sources of income through the tax system.
- In the census, we rely on the respondent’s interpretation and calculation of their total gross income, ability to recall all of their income over the previous year, rounding, and choosing the correct income band. In contrast, in most cases, the IDI provides the official records of income from the tax system. A small amount of income information in the IDI is from detail provided by the individuals about their self-employment. Although this information relies on respondent’s interpretation, it still has more legal constraints than the census.
- Incentives to avoid paying tax mean that some people may minimise the income they report in their tax returns. These incentives are not present in the census.
- The concepts that shape people’s view of their income source may in some cases not align well with tax definitions and the tax forms used for filing. Self-employment may be particularly affected by these differences in interpretation.
- Coding decisions affect comparability. While this investigation attempts to re-code all tax income sources to the standard classifications, this has not been possible in all cases.

These factors contribute to measurement error. Comparison of individual responses will shed light on whether these differences result in random variation or whether there are systematic differences between the two sources.

## Coverage differences

Sources of coverage error in the administrative sources available data in the IDI include:

- The administrative sources investigated only provide positive information about the presence of taxable income. They do not provide information for individuals with no income at all. In contrast, the census provides a respondent declaration of zero income, and no source of income.
- In the IDI, IR tax data provides the main stream of income information, and hence all information about income and income sources reported here is from taxable income. Non-taxable income source categories are not available in the tax data: for example, non-taxable benefits (including the accommodation supplement) and 'support payments from people who do not live in my household'.
- Income information for people who participated in the labour market, but do not participate in the tax system, is not available in the IR data.
- Investment income from interest and dividends is not available in the IDI (although it is collected by IR). Individuals for whom this is their only source of income will not be included in tax-derived estimates. Counts of those with investment income will be biased downward, and total income for those with income from interest and dividends will also be too low.

Other sources of income not available in the IDI are:

- income from private superannuation
- overseas income that is not taxed in New Zealand

We can expect to see some systematic differences between the census and tax-derived income information due to these differences in coverage.

## Results for total personal income

### Aggregate comparisons for total personal income

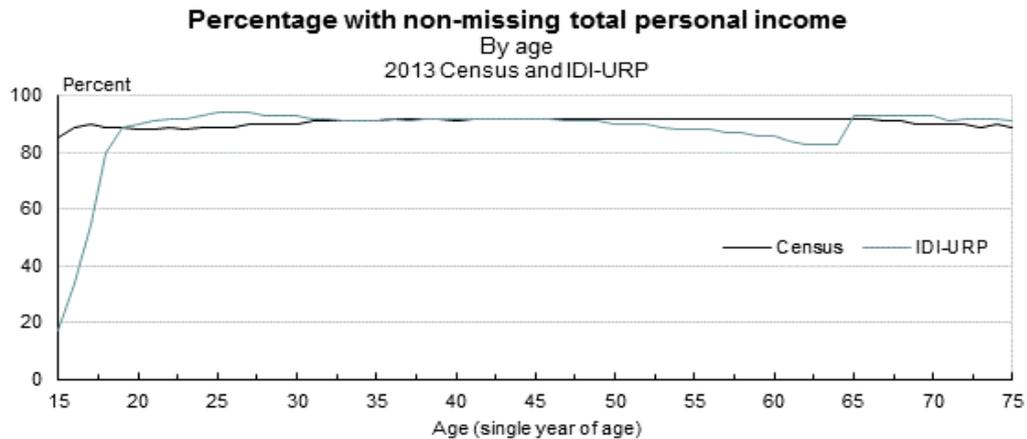
This section presents the results of comparisons between the IDI-URP estimates and the census results for total personal income. Deriving income information from administrative sources for the IDI-URP population gives us the distribution of total personal income that would be obtained if a census were based solely on administrative sources (and targeted at the same usual resident population as our current census).

Personal income was derived for 88 percent of the IDI-URP aged 15+ (of 3,519,100 in the IDI-URP). This compares with 90 percent of the census usual resident population aged 15+ (of 3,376,400 in the 2013 Census), including those who reported zero income.

Figure 2 shows the availability of income information by age. The most striking feature is a drop in coverage in the tax income data for the youngest ages. This is likely due to most of the younger age population being still in school and not working or receiving taxable income. Another drop in coverage in the tax income data is observed for the ages approaching 65 years. These are likely to be people who have retired from paid work, but are not yet receiving superannuation. In the IDI, the coverage for those aged 65 and over (65+) is slightly higher than census. High coverage is expected for ages 65+ in the tax data, since nearly all New Zealand residents aged 65+ are eligible for superannuation.

Census response rates are very consistent at around 90 percent over all ages, though with a slight drop at ages 15 and 16 years.

**Figure 2**



Source: Statistics New Zealand

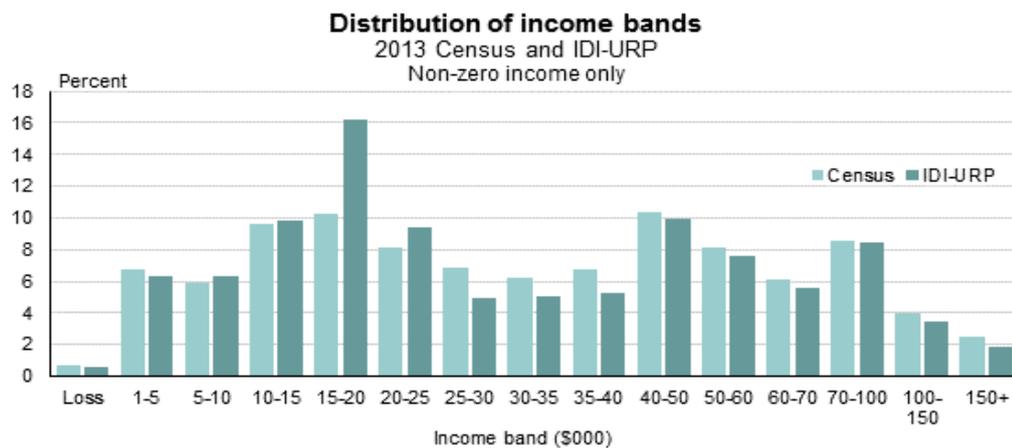
In the 2013 Census, 242,500 individuals (7.2 percent) have zero income, with 49 percent of 15- to 18-year-olds reporting zero income. This compares with no one in the IDI-URP recorded as receiving zero income from the tax data.

Overall, the tax data includes more responses than the census for those with non-zero income, while the census includes more people with valid responses, since the census includes those with zero personal income.

The following analyses exclude the zero income category.

Figure 3 shows the percentage of people reported in each income band for the census and the IDI-URP (excluding zero income and missing responses). The distribution across income bands is largely comparable between the IDI and the census, except for income between \$15,001 and \$20,000, where the proportion in the IDI-URP (16 percent) is significantly higher than for the census (10 percent).

**Figure 3**



Source: Statistics New Zealand

The \$15,001–\$20,000 income band largely coincides with New Zealand Superannuation rates in 2012 and 2013. Some working-age benefits also fall into this band. Although the marked difference is not limited to age, it is least observed in the younger age groups and most prominent in the older age groups.

[See Benefit rates](#) for more information about working-age benefits.

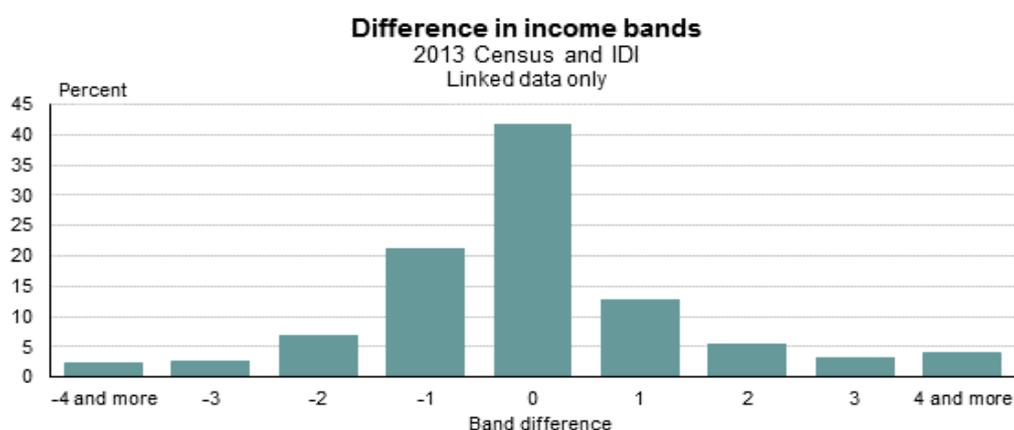
## Comparison of individual-level records for personal income

For the following analysis, we have used the linked Census-IDI dataset, as described above. We compare the total personal income an individual reported in the census with the income derived for the same individual from the tax data in the IDI.

Of the 3,125,100 usually resident individuals aged 15+ in the linked Census-IDI dataset, income information is available for 2,484,300 (80 percent) in both census and IDI. Figure 4 shows the distribution of the discrepancies between income bands derived from census and from tax data. No difference means an individual has the same band in both sources, with 1 or -1 representing one band difference (for any income band category), and so on.

For individuals where a valid income band is available in both sources, 41 percent (1,020,200) have the same income band in both census and IDI tax data, and 75 percent are within one band.

**Figure 4**



Source: Statistics New Zealand

The discrepancies are weighted towards negative differences, where tax income is higher than census income: 33 percent are in higher income bands in the tax data than in the census, while 25 percent have lower income in the tax system than is self-identified in the census. This is despite investment income being unavailable from tax sources for this study.

The underreporting of income is not unique to the New Zealand census, and has been reported by other international agencies. In the USA, researchers have reported the long-term differences between census and Bureau of Economic Analysis (BEA) measures of income, where BEA relies on administrative records and the Census Bureau relies on sample surveys (Katz, 2012). This report suggests that the substantial conceptual differences between the two sources may be a key contributing factor.

A full matrix of the counts by income band reported by individuals in the census and their income band as derived from the tax data is available in appendix 4. This shows a tendency to higher rates of agreement with higher income bands. However, it is difficult to eliminate the effect of more consistent reporting from the increasing band width for higher incomes.

The pattern of more census responses reporting a lower income band than the tax data (rather than differences in the opposite direction) is seen across all income bands. This may be partly due to census responses incorrectly reporting net instead of gross income.

# Results for sources of personal income

## Aggregated comparisons for income source

This section presents the results of comparisons between the IDI-URP estimates of the numbers of those with income from each source and the 2013 Census results.

A positive indication of income source was derived for 88 percent of the IDI-URP aged 15+ (of 3,519,100 aged 15+ in the IDI-URP). This compares with 93 percent of the census usual resident population aged 15+ (of 3,376,400 in the census-URP) who had a positive identification of income source.

Table 3 shows the number and percent of people in each of the main income source categories for the 2013 Census and for the IDI-URP population as measured by the IDI tax data. Individuals can have more than one income source. In table 3, people are counted once in each income source category, so that total responses are greater than the number of people. Proportions in each category are calculated based on the total number of people with stated responses.

There are 233,600 individuals (7 percent) reporting no source of income in the 2013 Census, very similar to the number reporting zero income. This compares with no one in the IDI-URP, since many of those with no income have no requirement to be included in tax returns.

By far the largest group in both sources are people earning income from wages and salaries. In the 2013 Census, the second most common source of personal income was investment income ('interest, dividends, rent, and other investments'), followed by New Zealand Superannuation and self-employed income. The most significant differences between census and the tax data are the proportions in investment income and the combined 'other' income sources, which are both very small in the tax data.

**Table 3**  
**Count and percentage of each individual source of income recorded in 2013 Census and IDI-URP**

Source of income	IDI-URP		Census		Ratio
	Count	Percent	Count	Percent	
Wages and salaries (W&S)	2,040,500	66	1,809,500	58	1.1
Self-employment income (SEI)	528,500	17	483,500	15	1.1
Investment income (INV)	36,400	1	655,100	21	0.1
Regular payments from ACC (ACC)	70,500	2	36,300	1	1.9
New Zealand Superannuation (NZS)	584,900	19	526,400	17	1.1
Main working-age benefits (BEN)	497,200	16	314,500	10	1.6
Student allowances (STU)	104,000	3	89,400	3	1.2
Other government benefits, superannuation, or sources of income (Other)	31,900	1	263,600	8	0.1
No source of income	0	0	233,600	7	0.0
Total people stated	3,080,700	126	3,133,700	141	1.0

The final column in table 3 shows the ratio of IDI-URP total responses to the census total responses for each income source category. A ratio close to one shows that census and

tax-based estimates for a category are highly consistent, independent of what other income sources may be reported. Wages and salaries, self-employment, New Zealand Superannuation, and student allowances are the most consistent of all income sources, with ratios of 1.1 and 1.2.

The low ratio for investment income (0.1) is as expected, due to the lack of information on interest and dividends available in the IDI. The low ratio of 0.1 in the IDI-URP 'other benefits' is also expected, since only paid parental leave is available in the tax data.

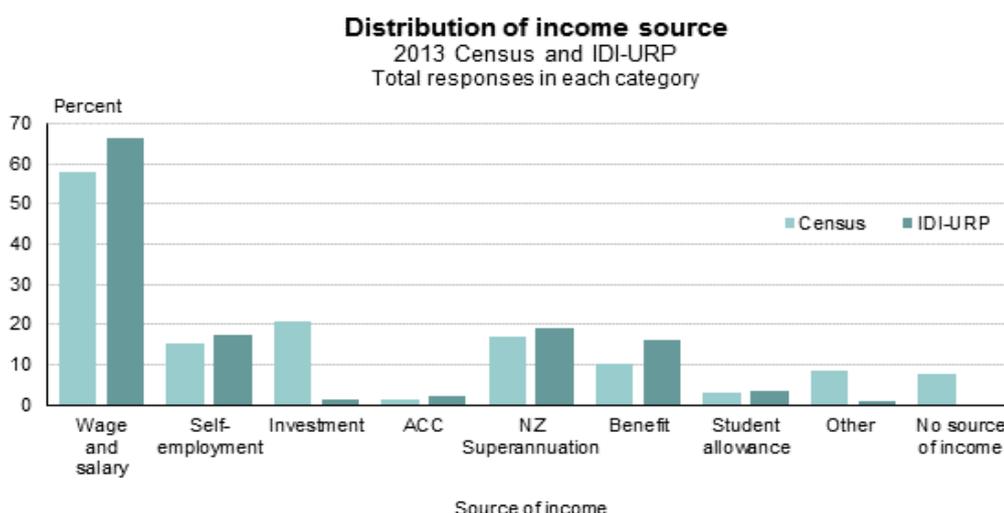
Some other categories also show large discrepancies. For example, there are many more people with ACC and the main working-age benefits income sources in the IDI-URP than there are in the census.

The tax data is a formal record of ACC payments. The low census responses to ACC as an income source may be affected by respondent recall or by how the census question is presented or interpreted. Taxable ACC payments are a replacement of work income, and are typically small amounts – half (55 percent) are less than \$5,000, and almost all (82 percent) are less than \$20,000. This would suggest that any impact on reporting of total income in the census is likely to be small.

The high ratio of 1.6 for the main working-age benefits is more surprising, and appears to reflect under-reporting of benefits as a source of income in census.

Figure 5 shows the percentages of each individual source of income recorded in the 2013 Census, and for the IDI-URP population as measured by the IDI tax data.

**Figure 5**



Source: Statistics New Zealand

### Comparison of individual-level records for income source

For the following analysis we have used the linked Census-IDI dataset to compare the income source information provided by an individual in the census with the income sources derived for the same person in the IDI.

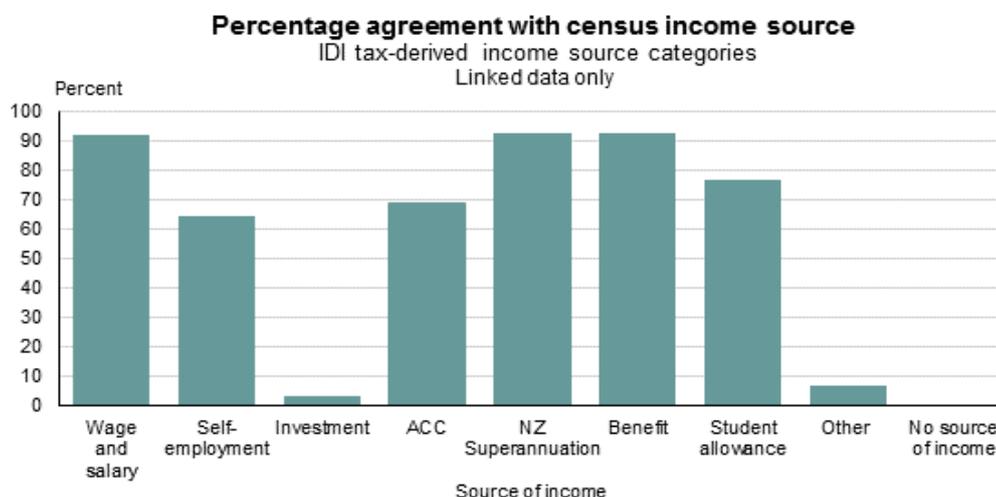
Of the 3,125,100 records for usual residents aged 15+ with a link between IDI and census, 83 percent have information about income source in both the census and the IDI. Overall, 56 percent have the same combination of source of income in both census and IDI, but consistency varies by income source category.

Figure 6 shows the proportion of individuals who reported a given income source in the census who also reported the same income source in the IDI. When someone reports in the census that they have income from wages and salaries, New Zealand Superannuation, and the main working-age benefits, more than 90 percent of the time the tax data agrees. Agreement is somewhat lower for student allowances (77 percent) and ACC (69 percent). In each of these categories, we expect the tax data to be a reliable record of receipt of this income source.

While there is good agreement at the aggregate level for self-employment, there is more discrepancy evident at the individual level, where only 64 percent of those reporting self-employment in the census have been coded as self-employed in the tax data. This may reflect differences between people’s interpretation of being self-employed as expressed in the census, and our coding of self-employment through tax filing.

As expected, only very few of those reporting investment income and ‘other benefits’ in the census have these income sources in the IDI.

**Figure 6**



## Cross comparisons

The differences in reported income sources may have varying impacts on the total personal income recorded in the census compared with total personal income estimates derived from administrative tax data. The tax data, unlike the census data, has income recorded separately for each income source, and the linked Census-IDI data allows us to compare at the individual level.

In this section, we show three examples of cross-comparison between total personal income and sources of personal income to elaborate on some of the main findings from this report.

These examples illustrate the kinds of analysis that can be undertaken.

### The \$15,001 to \$20,000 income band

Figure 3 showed that the distribution across income bands is largely comparable between the IDI and the census, except for the \$15,001 to \$20,000 income band. A total of 310,000 individuals in this income band in the tax data are in a different income band in the census.

This income band also coincides with working-age benefits and New Zealand Superannuation rates. Analysis on people in the IDI income band \$15,001 to \$20,000 confirms that 80 percent of this clustering is made up of individuals receiving either working-age benefits (22 percent) or New Zealand Superannuation (58 percent).

The census shows the proportion of people reporting income from interest, dividends, rent, and other investments steadily increased with age group. It was most common for people aged 65+ (40 percent) (Statistics NZ, 2013b). Since investment income is known to be missing in the tax data, we would generally expect total income for older people to be lower in tax data than the census.

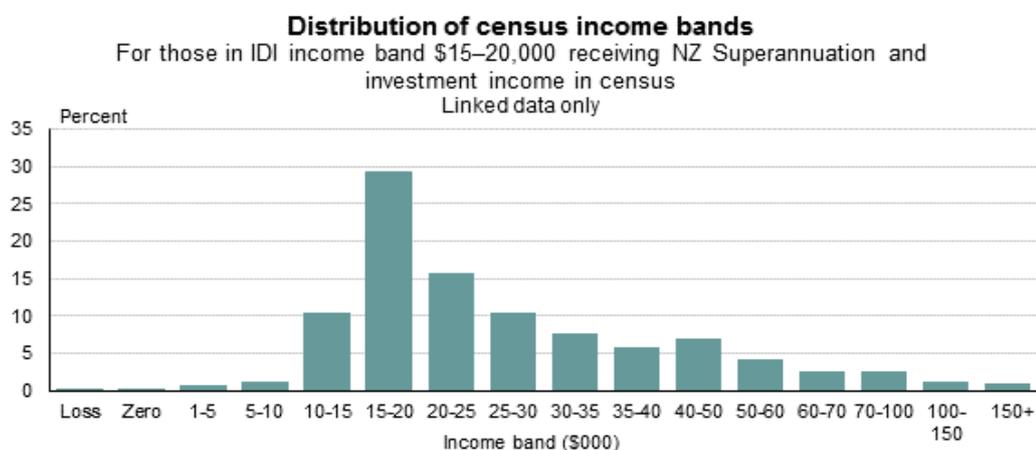
We now look to see whether this is one of the reasons for the markedly different distributions seen for the \$15,001 to \$20,000 income band.

One-third of people in the IDI income band \$15,001 to \$20,000 report receiving investment income in the census that is not available in the IDI. Of those, 87 percent are New Zealand Superannuation recipients.

Figure 7 shows the distribution of census income bands for these New Zealand Superannuation recipients who report investment income in the census and are in the IDI income band \$15,001 to \$20,000. This figure reveals that most of the individuals in the IDI income band \$15,001 to \$20,000 report the same or a higher income band in the census.

While the reporting of receipt of New Zealand Superannuation is very consistent between the census and the IDI, the total personal income derived from the tax data is consistently lower than reported in the census. As the census includes investment income, while the tax data largely does not, it seems likely that the census total personal income is the more accurate source for this group.

**Figure 7**

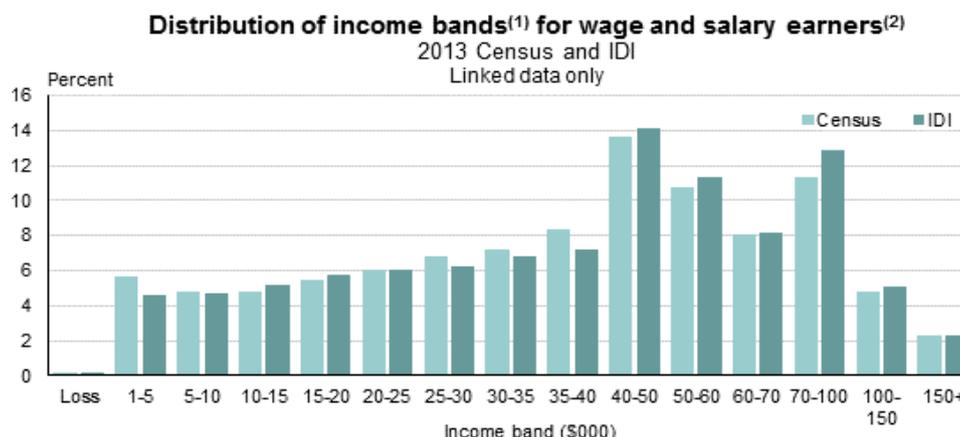


Source: Statistics New Zealand

## Wages and salaries

Wages and salaries is the most consistently reported source of income between the census and the IDI. Over 90 percent of respondents who reported wages and salaries as the only source of income in the census also earn wages and salaries in the IDI. Here, using the Census-IDI linked dataset, we look at the income distribution for those who reported wages and salaries as their only income source in both the 2013 Census and tax data. Figure 8 shows strong consistency in the distribution of wage and salary income between census and the IDI across all income bands.

**Figure 8**



1. Non-zero income only.

2. Those with wages and salaries as sole income source in both census and IDI.

Source: Statistics New Zealand

When we compare the difference in income bands for wage and salary earners, we find 49 percent in the same band in both the census and the IDI, and 82 percent within one band. This example demonstrates that earners of wages and salaries not only have good consistency in identifying this income source, but also report total personal income more consistently between census and IDI than they do other sources of income.

### Zero-income earners in census and IDI

In the Census-IDI linked population, there are around 232,700 zero-income earners in the census, but almost none in the IDI. As well, 434,300 individuals have no income information in the tax dataset.

Figure 9 combines two distributions that compare those with no income reported in the census, and missing income data in the IDI. Figure 9 shows:

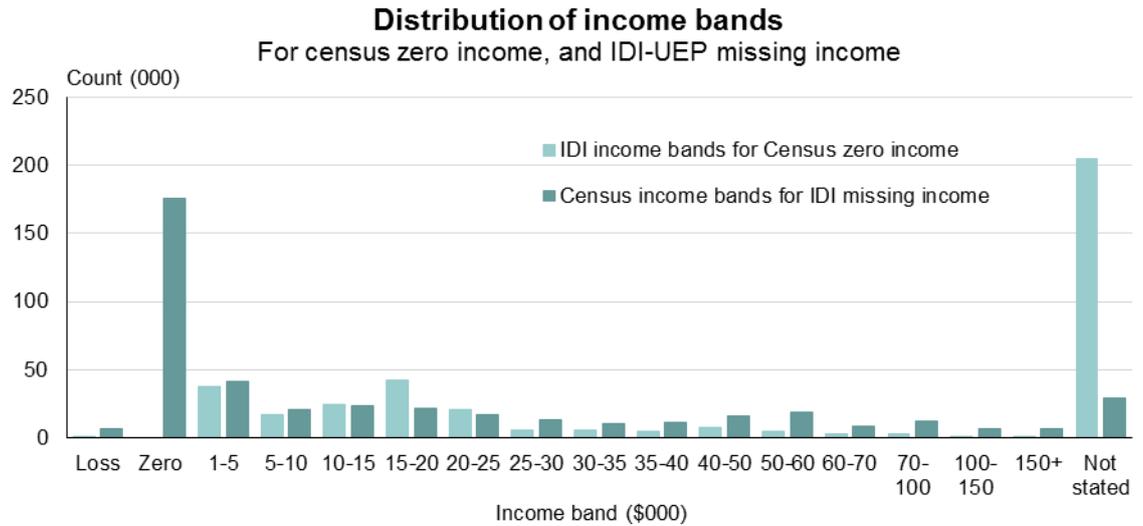
- the distribution of census zero-income earners across IDI income bands
- the distribution of people who have no income information in the tax system across census income bands.

Around 75 percent (175,600) of zero-income earners in the census also have no income information in the tax system. The remaining zero-income earners in the census are mainly in the lower IDI income bands. A census response of 'no income' thus largely corresponds to missing income in the tax data.

Conversely, just 40 percent of those who have no income information in the tax data also reported zero income in the census. The remaining 60 percent are in an income band other than zero in the census, and spread across all census income bands. Thus, missing income in the tax data cannot be assumed to be zero total personal income.

While 'no income' reported in census is largely supported by the administrative data, the reverse is not true.

**Figure 9**



Source: Statistics New Zealand

Median income is a common measure of the income distribution. It is difficult to calculate a comparable median income for census data and tax data. This is because people with zero income are not explicitly identified by IR in the tax data, and because the census median income uses an approximation.

The 2013 Census median income was estimated at \$28,500, which includes those who reported zero income. The median income for the 2013 IDI-URP is estimated to have a lower bound of \$20,800, assuming those with missing income have zero income, and an upper bound of \$28,500, excluding all those with missing income.

# 7 Conclusion

## Summary

This paper presents our investigation into the potential for deriving census income information from administrative sources. We compared income variables between 2013 Census with similar information derived from Inland Revenue data available in the IDI in May 2015. Findings are summarised by the relevance of the concepts used in administrative sources, coverage compared with the census, and sources of measurement error.

The concepts of money income calculated before tax over an annual period, as used by the census, are readily derived from Inland Revenue tax data.

Census total personal income includes income from all sources. While the administrative sources available in the IDI for this study include taxable income from most income source categories, the data available does not include investment income from interest and dividends, and also does not include non-taxable government transfers, other non-taxable income sources, or income earned and taxed overseas. It is not possible to positively identify those with no income from the administrative data.

Census population coverage is reduced due to non-response (2 percent net census undercount and 5 percent substitutes), and by design census does not include residents temporarily overseas. The target population for the tax data does include all New Zealand residents, but coverage is reduced because those with no income are not included, and the tax data will be missing those for whom investment, non-taxable, or overseas income are their only sources of income.

Overall, 75 percent of individuals fall in the same or an adjacent income band in both sources. Measurement error from the census is most likely due to respondent error, and missing data from the 5 percent of those who respond to the census, but do not answer the total personal income question. This individual non-response leads to 15 percent non-response for household income, and is known to disproportionately affect certain groups.

Income recorded through the taxation system can be taken as a formal record with minimal measurement error. Measurement error in the administrative sources is expected to be mainly due to the lack of information for some income sources. Investment income is the second most common source of income reported in the census, and administrative estimates of total income will be biased downwards for these people. The lack of non-taxable government transfers is particularly significant when considering the income of working-age beneficiaries, since supplementary benefits are often paid in conjunction with core benefits and can form a significant proportion of social assistance income.

Comparisons of income bands between the census and an equivalent administrative population show a reasonably consistent distribution, except for a large discrepancy for the \$15,001 to \$20,000 income band. Most tax-derived income in this band is from government benefits and New Zealand Superannuation, and the discrepancy appears to reflect lack of coverage in the tax data of additional income sources for these groups. A comparison of income distribution restricted to only wages and salaries shows very close agreement between the census and the tax data.

The results, for the most part, are encouraging for the quality of income data provided by census respondents.

## Limitations

The IDI-URP population shows the income distributions that would be achieved if only administrative data were available. The IDI-URP is constructed to be as close as possible to the census population. However, the aggregate comparisons between the census and the IDI-URP are affected by errors in both sources – undercount and missing responses in the census, and incorrect inclusions and exclusions in the IDI-URP. These population differences may be causing some of the observed inconsistencies, rather than real differences in income.

In contrast, the linked Census-IDI dataset removes any population differences, but slightly reduces the census population available, and results may be affected by linkage errors between the census and the IDI.

## Discussion

Our broad aim has been to understand the potential for deriving census income information from administrative sources. We have based our investigations on the concepts and classifications of income information that have been used by the census for almost a century. We have not considered here whether the same approach would continue to best meet requirements for income information in future censuses. Further work is needed to clarify future requirements for census income information in the light of the strengths and limitations of administrative data.

Administrative data from the taxation system on income and income sources has several advantages over survey collection. The tax data is clearly more *precise* than census data. Tax data has income in dollar values rather than income bands, it distinguishes income by each source, and does not suffer from respondent recall or misinterpretation. Because Inland Revenue is the official tax collection agency, we can expect taxable income to be measured reliably on the whole. Tax data is available annually, more frequently than is possible in a periodic survey-based census.

However, a lack of key income sources – most investment income, and non-taxable income – means that the available administrative data overall is less *accurate* than the census. While not available in the IDI at the time of this study, much of the information needed to improve the measurement of total income is in fact held within government by Inland Revenue and the Ministry of Social Development. Work is underway to include this data in the IDI, which will substantially address the issue.

Some data gaps remain in the administrative data: income earned overseas, non-taxable income from sources outside of government, and taxable income not reported to government would not be included. Partly because of these gaps, it is not possible for government sources to positively identify all those with no income.

Overall, the administrative sources investigated show good potential for providing census information about total personal income and income source, provided income information in the IDI can be extended to include IR tax data on interest and dividends, and MSD data on non-taxable social transfers. However, extending the available data for personal income will not provide robust family or household income. Improvements in the ability to construct households and families in the IDI are needed to open up some of the main uses of income data.

## 8 References

- Bycroft, C, Reid, G, McNally, J, & Gleisner, F (2016). [Identifying Māori populations using administrative data: A comparison with the census](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Gibb, S, Bycroft, C, & Matheson-Dunning, N (2016). [Identifying the New Zealand resident population in the Integrated Data Infrastructure \(IDI\)](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Gibb, SJ, & Das, S (2015). [Quality of geographic information in the Integrated Data Infrastructure](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Katz, AJ (2012). [Explaining long-term differences between census and BEA measures of household income](#). Retrieved from [www.bea.gov](http://www.bea.gov).
- Ministry of Social Development (nd). [Benefit rates](#). Retrieved from [www.workandincome.govt.nz](http://www.workandincome.govt.nz).
- O'Byrne, E, Bycroft, C, & Gibb, S (2014). [An initial investigation into the potential for administrative data to provide census long-form information: Census Transformation programme](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Reid, G, Bycroft, C, Gleisner, F (2016). [Comparison of ethnicity information in administrative data and the census](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Shrosbree, E (2015). [Comparing education and training information in administrative data sources and census](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics New Zealand (nd). [Classifications and related statistical standards](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics New Zealand (2016). [Guide to reporting on administrative data quality](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics NZ (2014a). [An overview of progress on the potential use of administrative data for census information in New Zealand](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics New Zealand (2014b). [Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics New Zealand (2014c). [Privacy impact assessment for linking census and Post-enumeration Survey data to the Integrated Data Infrastructure](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics New Zealand (2013a). [2013 Census information by variable](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).
- Statistics NZ (2013b). [2013 Census QuickStats about income](#). Retrieved from [www.stats.govt.nz](http://www.stats.govt.nz).

## 9 Appendixes

### Appendix 1: Income source classification – definitions

- **Wages and salaries:** income received from all current and previous salary and wages held over the reference period, and any job-related bonuses, commissions, redundancies, or other taxable income such as honoraria or directors fees.
- **Self-employment income:** net profit or loss received from all current and previous self-employment over the reference period, including drawings (cash or goods the respondent takes out of the business instead of receiving a ‘wage’).
- **Investment income:** net profit or loss received from investments such as rent, Māori land or other leased land, dividends from New Zealand companies, royalties, and interest from the following: banks, other financial institutions, bonds, stocks, money market funds, debentures or securities.
- **Private superannuation income:** includes income received from both job-related superannuation schemes and other private schemes.
- **New Zealand Superannuation and war pensions:** in addition to New Zealand Superannuation, this category also includes veterans, war, disablement, and surviving spouse pensions.
- **Other government benefits:** all family assistance payments, such as those receiving: the ‘working for families’ package; main benefits (eg unemployment); sickness, domestic purposes, and invalids benefits; student allowances; emergency benefits; and supplements.
- **Other sources of regular and recurring income:** includes income received from Accident Compensation Corporation (ACC) and private compensation providers, trusts, annuities, alimony, educational scholarships, and income protection insurance.

### Appendix 2: Income bands classification

The statistical standard uses the Census 2013 classification, with 14 bands.

Appendix table 1 Income bands classification

Classification	Income bands – standard classification 2009
Abbreviation	INCOMEBANDS
Version	V1.0
Effective date	April 2009

#### INCOMEBANDS V1.0

11 Loss	22 \$50,001–\$60,000
12 Zero income	23 \$60,001–\$70,000
13 \$1–\$5,000	24 \$70,001–\$100,000
14 \$5,001–\$10,000	25 \$100,001–\$150,000
15 \$10,001–\$15,000	26 \$150,001 or more
16 \$15,001–\$20,000	44 Don't know
17 \$20,001–\$25,000	55 Refused to answer
18 \$25,001–\$30,000	77 Response unidentifiable
19 \$30,001–\$35,000	88 Response outside of scope
20 \$35,001–\$40,000	99 Not stated
21 \$40,001–\$50,000	

Source: Statistics New Zealand

## Appendix 3: Census 2013 income questions

**30** Mark as many spaces as you need to show all the ways you yourself got income in the 12 months ending today.

DON'T count loans because they are not income.

wages, salary, commissions, bonuses, etc, paid by my employer  
 self-employment, or business I own and work in  
 interest, dividends, rent, other investments  
 regular payments from ACC or a private work accident insurer  
 New Zealand Superannuation or Veteran's Pension  
 other superannuation, pensions or annuities (other than NZ Superannuation, Veteran's Pension or war pensions)  
 Unemployment Benefit  
 Sickness Benefit  
 Domestic Purposes Benefit  
 Invalid's Benefit  
 Student Allowance  
 other government benefits, government income support payments, war pensions, or paid parental leave  
 other sources of income, counting support payments from people who do not live in my household

or  no source of income during that time

Source: Statistics New Zealand

**31** From all the sources of income you marked in question **30**, what will the total income be:

- that you yourself got
- before tax or anything was taken out of it
- in the 12 months that will end on 31 March 2013

loss  
 zero income  
 \$1–\$5,000  
 \$5,001–\$10,000  
 \$10,001–\$15,000  
 \$15,001–\$20,000  
 \$20,001–\$25,000  
 \$25,001–\$30,000  
 \$30,001–\$35,000  
 \$35,001–\$40,000  
 \$40,001–\$50,000  
 \$50,001–\$60,000  
 \$60,001–\$70,000  
 \$70,001–\$100,000  
 \$100,001–\$150,000  
 \$150,001 or more

See the Guide Notes to help work out your income

[Go to the next page in English](#)

# Appendix 4: Counts and percentages by income band for census and tax data

The figure below shows a full matrix of the counts and percentages by income band reported by individuals in the census, and their income band as derived from the tax data.

**Appendix table 2**  
**Counts by income band for census and tax data**

		IDI income band (\$000)																Not stated	Total
		Loss	0	1-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-50	50-60	60-70	70-100	100-150	150+		
Census income band (\$000)	Count																		
	Loss	1,500	0	2,000	1,300	1,400	1,200	500	400	200	200	300	200	100	100	0	0	6,500	15,900
	0	1,000	30,200	8,300	6,100	5,600	2,200	900	600	500	800	400	200	200	100	0	0	175,600	232,700
	1-5	1,600	67,000	33,300	16,700	11,000	4,100	1,700	1,300	900	1,100	400	200	200	0	0	0	41,600	181,100
	5-10	1,400	19,700	44,500	38,800	20,400	7,400	2,700	1,400	900	900	400	200	200	100	0	0	20,800	159,800
	10-15	1,400	10,400	24,600	78,200	86,900	21,100	6,100	2,900	1,600	1,600	600	300	300	100	0	0	23,800	259,900
	15-20	1,300	6,600	12,100	39,100	117,300	52,700	13,500	6,500	3,300	2,700	900	400	300	100	0	0	22,200	279,000
	20-25	1,100	4,700	7,200	16,400	51,500	67,800	26,700	14,700	6,900	5,100	1,500	600	500	200	0	0	17,000	221,900
	25-30	1,000	3,600	5,000	9,300	31,700	28,800	31,700	30,100	15,500	11,000	2,800	1,100	700	200	100	0	13,200	185,800
	30-35	800	2,500	3,400	5,300	20,100	15,100	16,100	34,800	31,400	21,700	5,000	1,700	1,000	200	100	0	11,000	170,200
	35-40	700	2,200	2,600	3,800	13,400	10,400	9,400	17,700	41,800	51,900	11,900	3,700	1,900	300	100	0	11,100	182,900
	40-50	900	2,300	2,800	3,700	13,000	9,200	7,200	10,600	20,200	122,300	55,200	13,000	6,000	800	200	0	16,100	283,500
	50-60	700	1,400	1,600	1,900	6,800	4,400	3,100	3,900	5,500	26,800	95,700	41,300	14,400	1,400	300	0	11,900	221,100
	60-70	500	900	1,000	1,100	4,200	2,600	1,600	1,900	2,500	7,800	20,400	66,100	43,400	2,400	500	0	9,000	165,900
	70-100	600	1,100	1,100	1,300	4,200	2,500	1,700	1,700	2,000	6,000	8,200	20,400	148,800	20,200	1,700	0	12,400	233,900
	100-150	400	500	400	500	1,800	1,000	600	600	700	1,800	2,000	2,700	16,700	63,700	9,000	0	6,500	108,900
	150+	600	500	400	500	1,400	700	500	400	500	1,100	1,100	1,500	4,000	9,400	39,200	0	6,400	68,200
Not stated	500	7,700	8,900	18,900	36,700	19,000	5,200	4,900	4,400	6,900	4,400	2,700	3,300	1,100	500	0	29,200	154,300	
Total	16,000	163,300	158,500	243,000	427,200	249,500	129,100	134,200	138,800	269,800	211,100	156,200	242,000	100,300	51,700	0	434,300	3,125,000	

Source: Statistics New Zealand

**Appendix table 3**  
**Percentages by income band for census and tax data**

		IDI income band (\$000)																Not stated	Total	
		Loss	0	1-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-50	50-60	60-70	70-100	100-150	150+			
Census income band (\$000)		Percent																		
	Loss	9	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
	0	6	0	18	5	3	1	1	1	0	0	0	0	0	0	0	0	0	40	7
	1-5	10	0	41	21	7	3	2	1	1	1	0	0	0	0	0	0	0	10	6
	5-10	9	0	12	28	16	5	3	2	1	1	0	0	0	0	0	0	0	5	5
	10-15	9	0	6	16	32	20	8	5	2	1	1	0	0	0	0	0	0	5	8
	15-20	8	0	4	8	16	27	21	10	5	2	1	0	0	0	0	0	0	5	9
	20-25	7	0	3	5	7	12	27	21	11	5	2	1	0	0	0	0	0	4	7
	25-30	6	0	2	3	4	7	12	25	22	11	4	1	1	0	0	0	0	3	6
	30-35	5	0	2	2	2	5	6	12	26	23	8	2	1	0	0	0	0	3	5
	35-40	4	0	1	2	2	3	4	7	13	30	19	6	2	1	0	0	0	3	6
	40-50	6	0	1	2	2	3	4	6	8	15	45	26	8	2	1	0	0	4	9
	50-60	4	0	1	1	1	2	2	2	3	4	10	45	26	6	1	1	1	3	7
	60-70	3	0	1	1	0	1	1	1	1	2	3	10	42	18	2	1	1	2	5
	70-100	4	0	1	1	1	1	1	1	1	1	2	4	13	61	20	3	3	7	
	100-150	3	0	0	0	0	0	0	0	0	1	1	1	2	7	64	17	1	3	
	150+	4	0	0	0	0	0	0	0	0	0	0	1	1	2	9	76	1	2	
Not stated	3	0	5	6	8	9	8	4	4	3	3	2	2	1	1	1	1	7	5	
<b>Total</b>		100	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	

Source: Statistics New Zealand