Stats NZ
Tatauranga Aotearoa

Estimating income from linked admin data: Impact of new sources

Census transformation

Te Kāwanatanga o Aotearoa
New Zealand Government

Disclaimer
Access to the data used in this study was provided by Stats NZ under conditions designed to give effect to the security and confidentiality provisions of the Data and Statistics Act 2022. These are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ. For more information about the IDI please visit https://www.stats.govt.nz/integrated-data/.
The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994 for statistical purposes. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Citation
Welsh, I (2023). *Estimating income from linked admin data: Impact of new sources.* Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104947-6 (online)

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

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

www.stats.govt.nz

Contents

Tables and figures

List of tables

List of figures

Purpose and summary

Purpose

Estimating income from linked admin data: Impact of new sources provides an evaluation of the quality of admin-sourced income data for two main personal income measures in the census: total personal income, and sources of personal income. It updates earlier work using additional tax and benefit information that has recently become available.

Summary

This paper is one of a series of investigations by Stats NZ's census transformation programme aimed at identifying and exploring the potential for admin data sources to provide census-type information. This work informs the direction of future censuses, but also underpins the use of administrative data in the 'combined' census model of the 2018 and 2023 Censuses and is the basis for inclusion of variables in the experimental administrative population census (APC).

Income is an important topic in the census. Previous Stats NZ work (Suei, 2016; Zabala, 2016) found there was good potential for admin data to provide personal income information, using data from those who had interacted with the New Zealand tax system. However, there were significant gaps for some income sources, as most investment income and non-taxable income was not available in Stats NZ's Integrated Data Infrastructure (IDI) at the time. There are now more types of income data available in the IDI from both the taxation and benefit system.

We compared results from the 2013 Census and 2018 Census with estimates produced from linked administrative sources in the IDI. We used a combination of taxable income data and benefit receipt data to determine the sources of income that an individual received. The sum of the total payments received then gives the individual's total personal income.

The availability of additional data sources and the derivation for zero income have resulted in substantial improvement in the areas that were of concern in the previous 2016 investigation. Coverage is now high with around 97 percent of the admin population assigned income information.

The distribution of income bands and the prevalence of most income sources is broadly similar between the census and the admin-derived population. However detailed comparisons reveal issues with census results for some categories. Income recorded through the taxation and benefit systems 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, which is now a small component of income received. In contrast, census responses are constrained by the limitations of a self-complete questionnaire and rely on a respondent's willingness and ability to provide accurate information.

Administrative data offers several advantages. Income is recorded in dollar values rather than income bands and distributions can include higher income categories. Income distributions can also be provided by income source. There is potential to extend to other concepts such as net income after tax. The data can also be produced annually, as already done in the experimental administrative population census (APC).

The administrative sources now available through government tax and benefit systems provide high quality, more detailed, and more frequent information about total personal income and income sources that goes beyond what can currently be achieved through a census questionnaire.

Introduction

The paper provides an evaluation of the quality of admin-sourced income data for the two main personal income measures in the census: **total personal income**, and **sources of personal income**. It builds on earlier work by Suei (2016) in the context of future censuses.

Census transformation in New Zealand

Stats NZ's census transformation programme (CT) is looking at the future direction of the New Zealand census, and working towards a census based on administrative data and supported by sample surveys. The work not only informs future censuses, but also underpins the use of administrative data in the 'combined' census model of the 2018 and 2023 Censuses, where the traditional full field enumeration approach is supported by the use of administrative data when responses are missing. Census transformation research has also been applied in the experimental administrative population census (APC), an annual time series that demonstrates the census-type information than can currently be produced using administrative sources.

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 admin data sources to meet those requirements.

The census provides a wide range of information about the characteristics of New Zealand's population, households and families, and dwellings. This paper is one of a series of investigations published by the census transformation programme exploring the potential for admin sources to provide census-type information.

Census income information

Income information currently collected by the census has a range of uses important to social and economic policy. Census data on total personal income and sources of personal income is collected for the usually resident population aged 15 years and over, enabling detailed geographic and demographic breakdowns (for example, by sex, age, and ethnic group). This data shows the distribution of total personal income and types of income received across the population. It is frequently combined with other census data on work to understand how income varies by work and labour force status, occupation, and industry. It is also used to derive household income and family income. Key users of census income data include central government agencies, local authorities, private organisations, and researchers. Information on household income and benefits (from sources of income) are inputs into the NZDep Index. This index is widely used for a range of research and policy work, and for targeting services and spending to help New Zealand's most vulnerable people.

Census income data is produced only every five years, and the level of detail available is constrained by the information people can reasonably be expected to provide in the context of a self-complete questionnaire. For example, total income is reported in income bands, and it is not possible to determine the income for each income source. The key value of census income data is its ability to provide information for small groups. Other surveys provide much more detailed income information.

Other sources of income information

The Household Labour Force Survey (HLFS) (Income), which is collected once a year, tracks the income and demographics of 15,000 households. It was formerly the stand-alone New Zealand Income Survey but is now an HLFS supplement. The survey's large sample population, and the inclusion of information about income from paid employment, self-employment, and government transfers, means income by income type can be compared across a range of demographics, such as sex, age, ethnicity, disability status, and highest qualification.

The Household Economic Survey (HES) collects detailed information on all sources of income and is used to report on child poverty. Every three years, HES produces a comprehensive report on household expenditure, making it the best measure for comparing household income and spending. From the 2018/2019 HES onwards, administrative data has been used to replace the following sources of income for all eligible individuals: income from wages and salaries; benefits; and other payments received from the New Zealand Government.

Although HLFS and HES produce more detailed information about income than census, HLFS through personal income outputs and HES through various types of household 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. Stats 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 Stats NZ Business Frame to a longitudinal series of Employer Monthly Schedule (EMS) payroll data from Inland Revenue.

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 Stats NZ.

Previous work

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 was based on metadata and intended to be indicative only.

Suei (2016) examined the use of administrative data for deriving personal income compared with the 2013 Census. Overall, the administrative sources showed good potential for providing income information for those who have interacted with the New Zealand tax system. Income information derived from the tax data was found to be more precise than that obtained through the census questionnaire. However, some gaps in the data available in the IDI at the time were evident, such as a lack of investment income from interest and dividends, non-taxable government transfers, other

non-taxable income sources, and income earned and taxed overseas. Suei (2016) also noted that while the administrative sources provide positive evidence of income received, the available data did not identify those with no income.

Aim and scope

Our overall aim for this investigation was to analyse to what extent census income information can be derived from existing administrative data. In particular, we focused on answering the questions:

- How do recent changes to the income data available in the IDI affect our ability to measure personal income?
- What are the benefits of using admin data for income beyond what census data can currently provide?

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 compares like-for-like replacement of data collected by the 2013 Census and 2018 Census with administrative data linked in the IDI, as well as identifying opportunities provided by admin-derived income data. Categories of income sources are those used in the 2018 Census questionnaire.

The administrative sources we considered were those available in Stats NZ's Integrated Data Infrastructure (IDI) as of June 2022. We derived income variables based on taxable income and working for families (WFF) data supplied by Inland Revenue (IR) and data about benefit payments supplied by the Ministry of Social Development (MSD).

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. We limited our analysis to the national level by age and sex, and for ethnicity and Māori descent, using data obtained from the administrative population census (APC). Comparisons are made with 2013 Census and 2018 Census data.

In an extension of previous work, we investigate a method for determining those with zero income. Additionally, the administrative data has some reporting time lag. We investigate what impact this time lag could have on the use of admin income sources in the census context.

Analysis of household income is out of scope for this paper and will be investigated in future work.

Results provided in this paper are not official statistics. They are published as an example of the type and quality of information about income that can currently be obtained from admin data sources.

Method

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

1. describing the formal statistical concepts 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 used in this investigation and developing a method for deriving estimates of **personal income** and **income source** from the admin data available in the IDI.

3. comparing the census data and the admin sources at three levels:

    a. **concepts and definitions**. We compare the concepts and definitions used in the census and the admin sources, with what is ideally being measured as described in statistical standards.

    b. **aggregate counts and estimates**. We compare census distributions for income information with distributions derived from the admin sources for a similar population.

    c. **individual-level information**. We compare census responses to the income questions with the equivalent information for the same individual derived from the admin sources.

The concepts of representation and errors of measurement provide a framework for assessing the accuracy of data sources (Stats NZ, 2016; Zhang, 2012).

Coverage is our main measure of representation. Coverage is the proportion of individuals from the relevant target population that we can derive the admin attributes for. For census income variables, the population of interest is people aged 15 years and over in the New Zealand resident population. Understanding administrative sources and the aggregate-level comparisons are most useful in providing insight into differences in coverage.

Where individuals could be linked between the census and admin data, we compared their information in both sources to evaluate consistency between the administrative values and census responses. These individual-level comparisons provide insight into potential errors of measurement that may result from differing statistical concepts, errors in collection and processing systems, or from linkage errors. Errors of measurement can occur in both census responses and administrative values.

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.

Linkage error

Agreement between sources can also be affected by the methodology used to link individuals across data sources. Specifically, two types of linkage error will affect comparisons using linked data:

- 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 coverage of an admin source (no information is available if links are not made when they should be), or they may introduce measurement errors if the wrong people are linked together. The June 2022 IDI refresh has reported successful linkage rates of 93.6 percent for MSD benefit dynamics data, and 94.6 percent and 94.7 percent for 2013 Census and 2018 Census respectively. Because IR data is used to compile the IDI spine, all tax records can be linked directly to the spine through the IR number.

The false positive rate is an estimate of links made where two records are linked together, but the records do not belong to the same person. This is estimated to have occurred at a rate of 0.9 percent for the census data, and 1.3 percent for MSD data. 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.

Statistical standards and classifications

Statistical standards and classifications 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 and 2018 Censuses.

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 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 received their total personal income in the 12 months preceding census day.

In the census it is generally only realistic to collect information on money income. This is income that a person can normally recall or can readily retrieve from their financial records. Money income is money flow from the deployment of one's labour, entrepreneurial skills, and assets; and from transfers received. So, the concept of money income relies on identifying its sources.

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

- **Investment income.** Net profit or loss received from investments such as rent, Māori land or other leased land, dividends from New Zealand companies, royalties, interest from the following: banks, other financial institutions, bonds, stocks, money market funds, debentures or securities.
- **New Zealand superannuation and war pensions.** In addition to New Zealand superannuation, this category also includes the veteran's, war disablement, and surviving spouse pensions.
- **Other government benefits.** All family assistance payments such as those under the working for families package are included in this source category, as well as main benefits (for example, unemployment benefit), student allowances, emergency benefits and supplements.
- **Other sources of regular and recurring income.** Includes income received from trusts, annuities, alimony, educational scholarships, and income protection insurance.
- **Private superannuation income.** Includes income received from both job-related superannuation schemes and other private schemes.
- **Self-employment income.** Net profit or loss received from all current and previous self-employment jobs held over the reference period, including drawings (cash or goods the respondent takes out of the business instead of receiving a 'wage').
- **Wages and salaries.** Income received from all current and previous wage and salary jobs held over the reference period, and any job-related bonuses, commissions, redundancies or other taxable income such as honoraria or directors fees.

Census Sources of Personal Income Classification is a flat classification with 15 categories in 2013 and 14 categories in 2018.

There are no conceptual differences between the 2013 Census and 2018 Census but minor changes to the classification were made for the 2018 Census to reflect government changes to benefits, effective from 15 July 2013:

- Unemployment benefit and sickness benefit are reclassified in 2018 Census as jobseeker support.
- Domestic purposes benefit is renamed in 2018 Census to sole parent support.
- Invalids benefit is renamed in 2018 Census to supported living payment.

The categories for the classification as used in the 2018 Census are:

- **00** No source of income during that time
- **01** Wages, salary, commissions, bonuses etc paid by my employer
- **02** Self-employment or business I own and work in
- **03** Interest, dividends, rent, other investments
- **04** Regular payments from ACC or a private work accident insurer
- **05** New Zealand superannuation or veteran's pension
- **06** Other superannuation, pensions, or annuities (other than NZ superannuation, veteran's pension or war pensions)
- **07** Jobseeker support
- **08** Sole parent support
- **09** Supported living payment
- **10** Student allowance
- **11** Other government benefits, government income support payments, war pensions or paid parental leave
- **12** Other sources of income, including support payments from people who do not live in my household
- **99** Not stated

Multiple responses could be provided for the sources of income question, which means that an individual could be counted two or more times and percentages calculated across income sources on the total population will add up to more than 100 percent.

Total personal income

The census variable **total personal income** identifies the before-tax income for respondents in the 12 months ending 31 March of the census year.

The concept currently used to collect income information is gross annual income. This is defined as income received by the individual, family, or household before the deduction of income tax, levies or withholding payments, and includes such items as income sourced from wages and salaries, self-employed income, property and rental income, dividends and investments, social insurance, superannuation, government assistance schemes and private transfers such as child support. It does not include social transfers in kind such as public education or government-subsidised health care services. Also excluded are reimbursement of expenses, money received from borrowing, contingent

income, and unrealised income. Irregular payments such as lump sum inheritance payments are excluded.

To overcome collection difficulties, in the census, total personal income is collected as an income range rather than an actual dollar income. Total personal income is also aggregated to form the following income outputs: total household income; total family income; combined parental income for couples with child(ren); total extended family income.

The income bands within the classification are determined by the analysis of income data collected by Stats NZ's detailed income collections. This analysis identifies emerging income trends in areas such as benefit levels, and middle and upper income earners. These bands are reviewed periodically to remain relevant to societal trends.

The income band data is collected for gross income where detailed information is unable to be collected or where it is not required. This classification has been developed to primarily cater for self-administered collections and collections requiring less detailed income band data. Collections that obtain detailed income data may output income within bands appropriate to the collection.

Total personal income in the census is a flat classification with income band categories. The first two categories reflect income loss and zero income. The income band categories used in the 2013 and 2018 Censuses (Census Income bands V1.0.0) are:

- **11** Loss
- **12** Zero income
- **13** $1–$5,000
- **14** $5,001–$10,000
- **15** $10,001–$15,000
- **16** $15,001–$20,000
- **17** $20,001–$25,000
- **18** $25,001–$30,000
- **19** $30,001–$35,000
- **20** $35,001–$40,000
- **21** $40,001–$50,000
- **22** $50,001–$60,000
- **23** $60,001–$70,000
- **24** $70,001–$100,000
- **25** $100,001–$150,000
- **26** $150,001 or more
- **99** Not stated

These bands have been updated for the 2023 Census, which has raised the upper income band to $200,001 or more, and combined income bands 13 and 14 into one band (Census Income Bands V2.0.0).

Data sources

This section describes the data sources used in this investigation: the New Zealand Census of Population and Dwellings, IR tax data, working for families (WFF), and MSD benefits data, with a focus on the income data in each source. We also describe the method used to derive total personal income and income sources from the admin sources, and the populations used in this analysis.

New Zealand Census of Population and Dwellings

The census is an official count of how many people and dwellings there are in New Zealand and captures a snapshot of who is living in New Zealand. The data helps the government plan services. These include hospitals, kōhanga reo, schools, roads, and public transport. Councils, Māori and iwi, businesses, and other organisations also use the data to work out the needs in their area.

The census aims to count everyone in New Zealand on census night. Overseas visitors are included while New Zealand residents who are not in New Zealand on census night are excluded. The target population for income measures is the 'census usually resident population' aged 15 years and over.

Historically, the census has been held every five years, with some exceptions. This investigation uses data from censuses held in 2013 and 2018 to compare against the results of the admin derivation. Results from the 2023 Census were not available for this analysis.

The 2013 Census was a traditional full field enumeration census. The 2013 Census includes a unit imputation process in the form of 'substitute' records, which represent individuals counted in the census but for whom no census forms were received.

The 2018 Census marks a significant step forward in the use of admin data (Stats NZ, 2019a, 2019b). The use of admin data for census attributes was partly as planned, but the role of administrative data was significantly expanded as a result of a lower-than-expected response rate. Admin enumerations were added to the census file when there was evidence the individual had not responded, and we had confidence in the quality of their admin record. Administrative data and the previous 2013 Census, as well as statistical imputation, were used where possible for social and economic characteristics where these were missing. The 2023 Census uses a combined model by design.

Income information in the census

The statistical standard for income information currently produced by the census is as described in Statistical standards and classifications. The income questions asked in the 2013 and 2018 Census are shown in Appendix A.

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.

Sources of personal income and **total personal income** were each collected from a single question in both 2013 Census and 2018 Census. It is not possible to determine the amount received from each source.

The responses obtained for **sources of personal income** were used to derive the sources of family income, sources of extended family income, and sources of household income variables. It is also output numerically as the number of different sources of income and number of income support sources (excluding ACC payments and NZ superannuation). Responses obtained for **total personal**

income are used to derive the following outputs: combined parental income for couples with child(ren), total extended family income, total family income, and total household income.

Both variables **sources of personal income** and **total personal income** were priority level 2 variables in 2013 Census and 2018 Census, and are again in the 2023 Census.

The mode of collection influences how the question can be answered. The online census form applies some constraints to prevent invalid responses. For the **sources of personal income** question, the online form allowed multiple responses to be selected as on the paper form – though 'No source of income' could not be selected with any income source and vice versa. The paper form allowed selection of 'No source of income' as well as other sources. The **total personal income** question only allowed a single response on the online form whereas the paper form did not prevent respondents from ticking multiple income bands. Additionally, both questions only allowed respondents aged 15 and over usually resident in New Zealand to respond on the online form. Edits are used to resolve inconsistent responses on the paper form and may lead to a code of 'response unidentifiable'.

In the 2013 Census, missing income responses were assigned to residual codes and grouped as 'not stated' in outputs. For the 2018 Census, Inland Revenue data available at the time, and statistical imputation were used where there was no valid questionnaire response for income. Table 1 shows the percentage of responses obtained from the various data sources for 2018 Census.

Table 1. Percentage of total personal income data and source of income, by data source

Percentage of total personal income data and source of income, by data source		
Source	**Total personal income**	**Sources of income**
Response from 2018 Census	81.2 percent	83.6 percent
2013 Census data	0.0 percent	0.0 percent
Administrative data	16.5 percent	14.1 percent
Statistical imputation	2.3 percent	2.1 percent
No information	0.0 percent	0.2 percent
Total	100 percent	100 percent
Source: Stats NZ DataInfo+		

Quality rating for the income variables

For the 2013 Census, the Stats NZ overall quality rating for **sources of personal income** was 'high', meaning it was fit for use with minor data quality issues only. The non-response rate was 7.2 percent, of which 4.9 percent were substitute records.

Total personal income in 2013 was given an overall quality rating of 'moderate', meaning it was fit for use with some data quality issues. The non-response rate was a 9.7 percent, of which 4.9 percent were substitute records. Because of the high non-response rates to total personal income, total household income and total extended family income are considered 'poor' quality.

For 2018, information on quality ratings was provided for three metrics: data sources and coverage, consistency and coherence, and data quality. The lowest metric determined the overall quality rating. Both personal income variables were given 'high' quality ratings for each metric, meaning the

overall quality rating was 'high'. Data sources and coverage was assigned a 'high' quality rating due to the use of administrative data and imputation. Without this data use, the quality rating would have been 'poor'.

The 2018 Census saw an improvement in derived variables which were mainly graded moderate, although income for extended families was still poor. However, the quality ratings in 2018 were affected mainly by missing household and family information, not because of missing personal income information.

Further information about the quality of census income information can be found in DataInfo+ 2013 Census: Total Income and DataInfo+ 2018 Census: Total personal income, Sources of personal income, Families and households income.

Integrated Data Infrastructure

Stats NZ's Integrated Data Infrastructure (IDI) was used to access the admin data sources. The IDI is a large research database. It holds de-identified microdata about people and households. The data is about life events, like education, income, benefits, migration, justice, and health. It comes from government agencies, Stats NZ surveys, and non-government organisations (NGOs). The data is linked together, or integrated, to form the IDI. Researchers use the IDI to conduct cross-sector research that provides insight into our society and economy.

Source agencies provide data periodically, and the IDI is updated three times a year. This update is known as the IDI refresh. This paper primarily used the June 2022 refresh in the IDI, with additional refreshes (all refreshes available from January 2020 to October 2022) used to investigate the effect of reporting lag.

Admin populations used in analysis

In this investigation, an admin population is derived that uses the individuals aged 15 years and over who were usually resident in New Zealand on 31 March of each year from 2006 to 2021. This population is obtained using the APC methodology (Stats NZ, 2022).

The reference date of 31 March was chosen as it aligns with the reference date for the **total personal income** question, and the tax year, so allows for a simpler derivation of income data from administrative records.

The admin population used in this analysis consisted of 3,527,646 individuals in 2013, 4.5 percent higher than in the 3,376,419 people in the 2013 Census usually resident population aged 15 years and over, and 3,868,881 individuals in 2018, 2.5 percent higher than the 3,776,355 people in the 2018 Census usually resident population aged 15 years and over. These differences in population are due to a combination of factors: slightly different reference dates (6 March for census, and 31 March for the admin population), census net undercount, people incorrectly included or excluded when deriving the admin population, and residents temporarily overseas who are included in the admin population but are not in the census usual resident population.

The individual analyses in this investigation use the linked census-admin population. This linked population includes all census records for usual residents aged 15 years and over for which a suitable link in the IDI has been found, and valid income information is available in both sources. Overall 86.2 percent of the census subject population were linked to the admin population in 2013 and 91.7 percent in 2018.

In 2018, the linked population used for the individual analysis is limited to those with a valid response provided on the census form and excludes records with income derived from admin data or statistical imputation.

Inland Revenue

Inland Revenue (IR) is the New Zealand Government's revenue collection agency. All sources of taxable income are required to be reported to IR, for example income earned through work, investments, rental property, or government benefits. IR data differentiates between wage and salary earners and self-employed persons because each group is treated differently for taxation purposes. Most self-employed income tax returns are filed annually.

The IR datasets in the IDI contain information about income from seven main IR sources:

- **Employer monthly schedule (EMS)** was used to report income taxed at source until April 2019. It was primarily designed for employers to deduct 'pay as you earn' (PAYE) tax from wages or salaries of their employees. All employers are responsible for filing the EMS, deducting PAYE tax, and paying it to IR. The month associated with an EMS is not always the same as the month in which a person was employed because it records the month in which they were paid. Each record in the EMS corresponds to a job (an employer-employee relationship) and includes the employee's tax code and employment start or end dates if they are in the month in which they were paid.

  Regular earnings of some self-employed persons can be reported in the EMS as wages or salaries. Most of these cases are identified in the IDI tax tables with the earnings correctly classified as self-employment earnings.

  Independent contractors are classified by IR as self employed. Those who perform a duty listed in the IR340 tax form are reported in the EMS where withholding tax from their earnings is deducted by their employer. The earnings of independent contractors in the EMS are known as withholding or schedular payments.

  The EMS form is also used by government agencies to report government transfer payments to individuals that are taxed at source. They include income-tested benefits, New Zealand superannuation, student allowances, paid parental leave, and accident compensation payments.

  The monthly EMS reporting was replaced in April 2019 by a system of filing by pay period.

- **Pay period tax filing.** From April 2019, all employers had to file PAYE tax returns by pay period rather than monthly. As part of the refresh cycle, the pay period records are rolled up to monthly returns for inclusion in the IDI EMS table.

- **Personal tax summary (PTS)** is a tax return for individual taxpayers to show their individual income and tax deduction details for a given tax year. Details are based on information about their taxable income provided to IR each month by their employers or payer and a range of tax rebates and tax credits entitlement. The PTS has been superseded by the Automatic Assessment compiled by IR.

- **Automatic assessment (AA).** Since 2019, IR has issued individuals with an automatic assessment of annual income from employment, investment, and benefits. This includes income from salary or wages, portfolio investment entities (PIE) including KiwiSaver, New

Zealand superannuation, schedular payments, income-tested benefits, interest or dividends, taxable Māori authority distributions, and benefits under an employee share scheme.

- **IR3** is the income tax return for individuals used to confirm the amount of personal income tax to be paid at the end of each tax year. The main purpose of IR3 is to include any monetary payments that have not been taxed at source. IR3 is used for self-employment (filed annually by sole traders) and includes non-zero partnership, or shareholder salary income, as well as rental income.

  Starting from July 2017, IR has extended the supply of IR3 information to the IDI to include further information about investment income, including earnings from interest, dividend, estate trust, overseas, and other sources. Also included are non-taxable income from a range of tax credit entitlements, tax rebate, and student loan-liable income. IR3 is filed annually. There may be a substantial lag in reporting IR3 data as individual businesses and those using tax agents have up to a year after the end of the financial year to file an annual tax return with IR.

- **IR4S** is filed by companies and includes remuneration income paid to shareholders, directors, and relatives of shareholders. IR4S is filed annually.

- **IR7** (named **IR20** in IDI tables) is for partnership and look-through companies. IR7 is filed annually.

Timeliness

Data from processed tax forms is supplied by IR to Stats NZ each month and can be accessed from the IDI in the following refreshes.

EMS data is timely because large employers and other employers must file the tax form by the 5th and 20th of the following month respectively. The pay period tax filing is also timely as it coincides with the pay period.

Data from annual tax returns is not timely because an extension to their filing deadline is given by IR if the business uses a tax agent to complete the forms. It can take more than 12 months from the end of the tax year for data from a processed annual tax return to be supplied to Stats NZ.

Working for families (WFF)

Working for families (WFF) is assistance for families that is delivered jointly by MSD and Inland Revenue. The datasets Inland Revenue provides to the IDI contain records of all recipients of any WFF main components.

- WFF tax credits (family tax credit, in-work tax credit, minimum family tax credit, parental tax credit)
- Accommodation supplement
- Childcare assistance (childcare subsidy, OSCAR subsidy).

These data are available from 1999.

The family details table provides entitlements and amounts paid to the primary caregiver. This includes tax credit entitlements from the family support tax credit (FSTC), parental tax credit (PTC), child tax credit (CTC), family tax credit (FTC), in-work tax credit (IWTC), and best start tax credit

(BSTC), as well as child support payments received. Though WFF entitlements are dependent on family income, it is paid to the primary caregiver so for income derivation purposes it is regarded as income only for the primary caregiver.

WFF data is included with monthly supplies of tax data, though is not finalised until annual returns have been filed after the tax year, and income and entitlement can be finally assessed.

Ministry of Social Development (MSD)

Ministry of Social Development (MSD) provides social support to New Zealanders. This includes income support through working age benefits and superannuation services, student allowances and loans, and social housing assistance.

The benefit dynamics data (BDD) supplied by MSD contains information on all people who have been entitled to a working-age social welfare benefit since 1 January 1993. Not everyone is entitled to a main benefit. Common reasons for a person not being eligible for a benefit is if their income or the income of their partner is above the cut-off, or a person is also only eligible for one main benefit. For persons entitled to a benefit, there is usually an initial stand-down period of up to two weeks from the date of entitlement before the first payment is received. In some circumstances, the stand-down period can be 13 weeks.

Each record in the benefit dynamics data corresponds to the period of benefit entitlement. The records of interest are those where the entitlement period fell within the tax year. Income from income-tested benefits is also reported in the EMS at an individual level but does not provide a breakdown by benefit type, and the totals correspond to payments in the month covered by the EMS, rather than the period of entitlement. The BDD is useful because it distinguishes benefit types, and each record has the start and end date of the entitlement period.

The MSD welfare reform policy in July 2013 consolidated benefits into three main categories with differing work obligations: job seeker support, sole parent support, and supported living payment. In the BDD table in the IDI, main benefit records corresponding to entitlement periods that started before July 2013 were allocated to one of three new categories. The MSD datasets within the IDI contain information about government transfer benefits paid to individuals since 1 January 1993 at three tiers.

- First-tier benefits are taxable regular payments made by MSD to an individual. As they are taxable, they are included in the employer monthly schedule (EMS) from Inland Revenue. However, MSD provides more detail on the benefit type than the EMS, more accurately reflects the period when the payment was made, and retrospectively accounts for any payment adjustments. Therefore, MSD data is prioritised over EMS tax data.

  The main benefits or payments included in the first-tier table are:

  - **Jobseeker support** is designed to provide short-term financial assistance to people aged 18 years or older and looking for work, training for work, or temporarily unable to work due to a health condition or disability. It comprises two subgroups: ‘job seeker support – work ready’ and ‘job seeker support – health condition or disability’. A person is not entitled to job seeker support if they are receiving another benefit or are eligible for another benefit.

  - **Sole parent support** payment is available to single parents aged 19 years and over, who are caring for children aged under 14 years. The sole parent support benefit

replaced the domestic purposes benefit (DPB) for sole parents and the widow's benefit for women who had been widowed and cared for dependent children.

  - **Supported living payment** is for people aged 16 years or older who are permanently and severely restricted in their ability to work because of a health condition, injury or disability, or total blindness. It is also for people providing full-time care for someone who would otherwise need to be in a hospital or other care facility. Some of the people receiving the disability benefit, sickness and invalid's benefit, and the DPB benefit who were caring for sick or infirm people were transferred to this benefit under the welfare reform.

  - **NZ superannuation** is a fortnightly payment for individuals aged 65 years and over. To be eligible, an individual must be ordinarily resident in New Zealand or the Realm of New Zealand, have lived in New Zealand for at least 10 years since turning 20, and have lived in New Zealand or the Realm of New Zealand for at least 5 years since turning 50. Entitlements vary depending on the relationship status, living situation, and any overseas benefit or pension entitlement for the individual. Though not a benefit, NZ superannuation is included in the benefit table within the IDI, so is described in this section.

- Second-tier benefits are non-taxable government transfers. Certain payments can be made to service providers. As these are non-taxable transfers, the data is not available in IR data. Often, these are supplementary payments to a main benefit. For example, an individual receiving jobseeker support could also be receiving the accommodation supplement and winter energy payment.

- Third-tier benefits are mainly one-off lump sum payments paid to an individual. Only non-recoverable payments are included in this derivation.

StudyLink is a service of the Ministry of Social Development that helps students pay for their post-school studies through provision of loans or allowances. Part of this includes day-to-day living cost payments. StudyLink is available in the IDI and used here solely as part of the derivation of zero income. Loans are not part of the definition of income, so are not included within this income derivation, and any income received from student allowances is available from the IR EMS table.

The MSD data is timely as data is supplied to Stats NZ on a quarterly basis.

Ministry of Education (MOE)

The Ministry of Education (MOE) is the government's lead advisor on New Zealand's education system. The early childhood and compulsory schooling enrolment data provided by MOE contains data on the enrolment status of individuals at secondary schools, allowing for determination of an individual's schooling status. School enrolment is used here solely as part of the derivation of zero income.

Summary of admin by income sources

Table 2 shows a summary comparison of admin data available in the IDI by the income source categories collected by the census. All sources of income except 'private superannuation income' have admin sources that report income related to the census categories. We also identify additional sources, some of which are not currently available in the IDI.

Table 2. Admin data in the IDI, by income-source classification, June 2022

Admin data in the IDI, by income-source classification, June 2022			
Sources of income (statistical standard)	**Census classification**	**IDI at June 2022**	**Other potential sources**
Wages and salaries	Wages, salary, commissions, bonuses etc, paid by my employer	Wages and salaries, commissions, bonuses	Employer contribution to KiwiSaver from IR
Self-employment income	Self-employment, or business I own and work in	Sole trader, company director/shareholder, partnership, commissions, bonuses, withholding payments	-
Investment income	Interest, dividends, rent, other investments	Rent (IR3), PIE returns (AA), interest, dividends (IR3, PTS, AA), overseas and other investments (IR3) from IR	Foreign investment fund income
New Zealand superannuation and war pensions	New Zealand superannuation or veteran's pension	New Zealand superannuation and veteran's pension	-
Private superannuation income	Other superannuation, pensions, or annuities (other than New Zealand superannuation, veteran's pension, or war pensions)	-	MSD records of overseas income impacting NZ superannuation payments
Other government benefits	Unemployed benefit, sickness benefit, domestic purposes benefit, invalid's benefit (2013) or jobseeker support, sole parent support, supported living payment (2018)	Main taxable benefits (MSD/IR)	-
	Student allowance	Student allowance	-
	Other government benefits, government income support payments, or paid parental leave	Paid parental leave (IR), non-taxable government benefits from MSD, WFF tax credit entitlements	-
Other sources of regular and recurring income	Regular payments from ACC or a private work accident insurer	ACC payments	-
	Other sources of income, counting support payments from people who do not live in my household	Other income (IR3), Māori distribution authority payments (AA), child support payments (WFF)	IR6B estate or trust beneficiary income
No source of income during that time	No source of income during that time	IR data with no income, receiving student loan living costs payment, school enrolment, young age, historical income data available	Inactive records from IR

Derivation of personal income variables in the IDI

Here we outline the methodology developed to calculate the total personal income and derive the sources of personal income variables using admin data available in the IDI. A detailed methodology is given in Appendix B.

Total personal income and income sources derivation

Both variables are derived from a combination of IR, MSD and WFF data.

For every unique individual (snz_uid) in the IDI who is aged 15 or older, for each tax year from 2000:

- Derive income and income sources from IR data.
  - Combine and deduplicate IR3 table from refresh database with those available in the ad hoc database.
  - Use the IDI methodology to combine data from EMS, combined IR3, IR4S, and IR7 (coded as IR20) data tables, with additional IR3 data included. This gives the following sources of income:
    - Wages and salaries; self-employment; ACC payments; NZ superannuation; paid parental leave; student allowance; rental income; interest and dividends from IR3 data; estate trust income; self-declared overseas income; and self-declared other income.
  - Use the PTS and AA tables to obtain income data not already included within the IR tables above. This gives the following sources of income:
    - Interest and dividends; Māori authority distribution income; and PIE income.
  - Combine all IR sources income tables, reconciling any double counting from IR3, PTS and AA tables with priority ordering: IR3 first, then AA, then PTS. This reconciliation applies when the same source of income is present in multiple tables, namely interest and dividends income.
- Derive income and income sources from MSD and WFF data.
  - Sum 'first-tier' taxable benefits from daily gross amounts within the tax year, and code to the appropriate benefit category of the classification. Where an individual has benefit data in both the IR and MSD first-tier tables, the MSD first-tier table takes priority and the IR benefit data is discarded.
  - Sum 'second-tier' non-taxable supplementary benefit amounts from MSD paid to individuals within the tax year.
  - Sum 'third-tier' non-recoverable ad hoc payments from MSD.
  - Collect total tax credit income and child support payments received from WFF data.
- Combine derived incomes to obtain total personal income and sources of income.
  - Total personal income for the tax year is the sum of the income from all income sources.
  - An individual is regarded as having income from a classification source if there is a non-zero amount of income calculated from the data.

As an additional step, a derivation of individuals having zero income is undertaken. This is to fill a major gap in the data available, as having zero income is an important part of income data. An individual is regarded as having zero income – and no source of income – if they do not have any identified income and at least one of the following conditions is met.

- They appear in any of the IR tables.
- They are receiving a student loan payment.
- They are 17 years or under at the end of the tax year, or 18 years and under and enrolled at a secondary school.
- They have income data available for previous tax years.

Results

Results are presented in four subsections. First, we compare concepts and definitions. Second, we compare aggregate counts and estimates. Third, we compare individual-level records. The concepts and definition comparisons are done jointly, and the later comparisons are done separately for total personal income and then for sources of income. Lastly, we present the results of the small investigations undertaken.

Comparing concepts and definitions

The concept of **income source** that both the admin-derived income 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. The only major difference between IDI and census data is that there is no private or overseas superannuation data available within the IDI. Not all admin sources are available for all years, with the most complete data available from 2019.

Additional income data that may be collected by census but not admin data includes non-taxable or cash jobs, transfers between households, and earnings from 'underground' marketplaces, although it is uncertain how much would be reported by census respondents.

The definition of **total personal income** in both the admin-derived income 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 admin data contains the actual dollar amount and the income for each source. The total dollar amount is easily aggregated to income bands such as those used by the census.

Additionally, there is no reconciliation between the two census questions. An individual could report no source of income in census, while also reporting non-zero total income amount. This could be a valid response due to the slightly differing reference dates between the two questions or could be an error. This is not an issue with the admin data.

Measurement errors

The census is a self-completed questionnaire, while the IR tax data and MSD benefits data record formal interactions with the tax and benefit systems. These different collection methodologies may lead to differences in the observed values and 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 admin data provides the official records of these sources of income through the tax and benefit systems.
- 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 admin data provides the official records of income from the tax and benefit systems. A small amount of income information in the admin data 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, although it is possible that some people unwilling to pay tax may also be less willing to correctly answer the census questions.

- 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.

- Recent migrants to New Zealand may have lower income in the admin data than is true for the year. This is because they could have been earning income overseas before moving to New Zealand and this overseas income is unknown to the admin data.

These factors contribute to measurement error in the census responses and the admin sources.

Measurement errors within census for certain income source categories have been revealed by the administrative data. For example, Suei (2016) showed that ACC payments were under-reported in the 2013 Census responses relative to the payments actually made, as evidenced in the IR EMS table.

Coverage differences

Limitations of coverage in admin data sources in the IDI

- The administrative sources investigated only provide positive information about the presence of taxable income or benefit recipients. They do not provide information for individuals with no income at all. The individuals identified as having zero income in admin data are based on an exclusion rule system, rather than an indication of such. In contrast, the census provides a respondent declaration of zero income, and no source of income.

- Income information for people who participate in the labour market, but do not participate in the tax system, is not available in the IR data. It is debatable how much this income information would be available in survey data.

- Investment income from interest and dividends, and PIE services (such as KiwiSaver) are only available in the IDI since 2019 (or 2021 for the PIE data), so have no census data to compare against, except for income declared as part of an IR3 filing.

Sources of income not available in the IDI

- Income from private superannuation or overseas superannuation.

- Overseas income that is not taxed in New Zealand.

Impacts of limited coverage

If these are the only income sources for an individual, they will have missing income. If they have other sources of income, the admin estimate will be too low. We can expect to see some systematic differences between the census and tax-derived income information due to these differences in coverage.

The census has historically had relatively high non-response to total personal income with around 10 percent 'not stated' (9.7 percent in the 2013 Census, 10.2 percent in 2006 and 11.1 percent in 2001) (Data Info+ 2013 Census).

Results for total personal income

Aggregate comparisons for total personal income

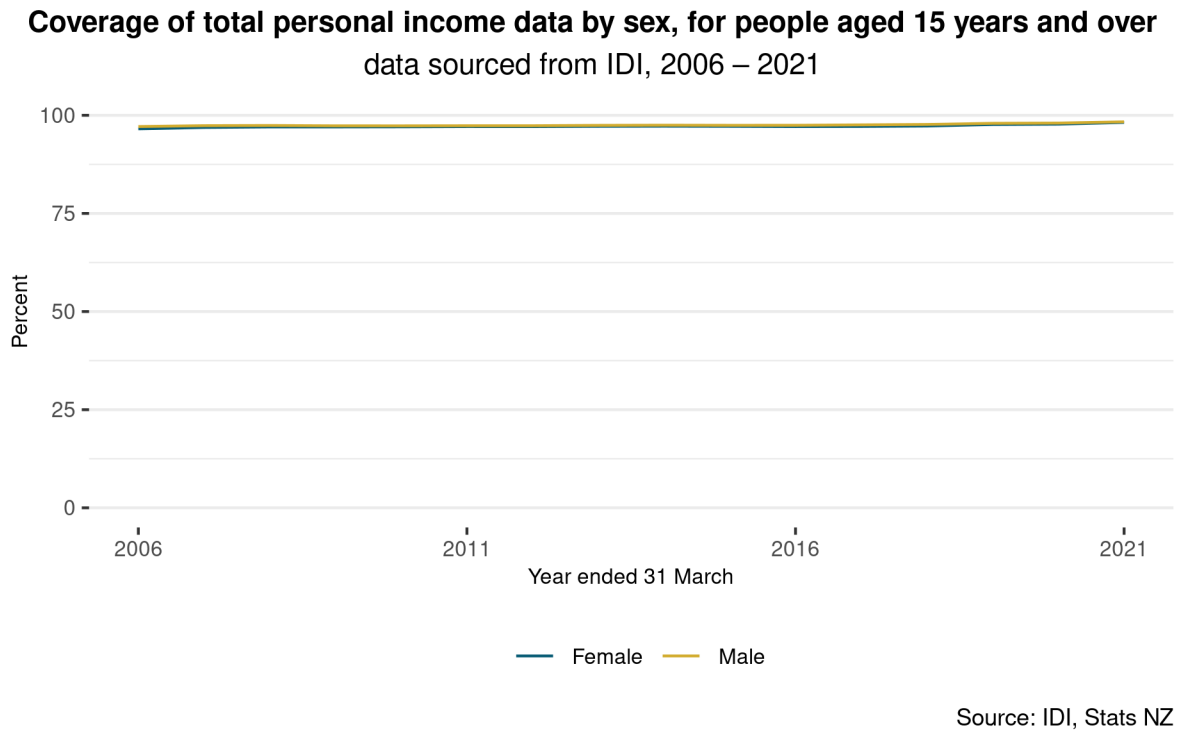
This section presents the results of comparisons between the admin population estimates and the census results for total personal income. Deriving income information from administrative sources for the admin usually resident population gives us the distribution of total personal income that would be obtained if a census were based solely on administrative sources.

Coverage

Since admin-derived total personal income is calculated from the sum of income by each income source, coverage is the same for both income variables. Coverage over time for the income variables by sex is shown in figure 1. It is consistently very high, at around 97.3 percent across all years, with minimal difference between males and females. This is an increase compared with the 88 percent coverage reported by Suei (2016), partly due to improved data availability, but mainly due to the use of a method to determine individuals with zero income. This zero-income derivation method could be improved through access to IR records of inactive IRD numbers.

Coverage of the admin-derived income is higher than historical censuses where it was around 90 percent for 2001, 2006, and 2013 Censuses, for which no statistical imputation was applied. In 2018 Census, responses were provided by 81 percent of respondents but coverage results are 100 percent due to the use of administrative data and statistical imputation to fill gaps in responses.

Figure 1. Coverage of total personal income data by sex, for people aged 15 years and over, data sourced from IDI, 2006–2021

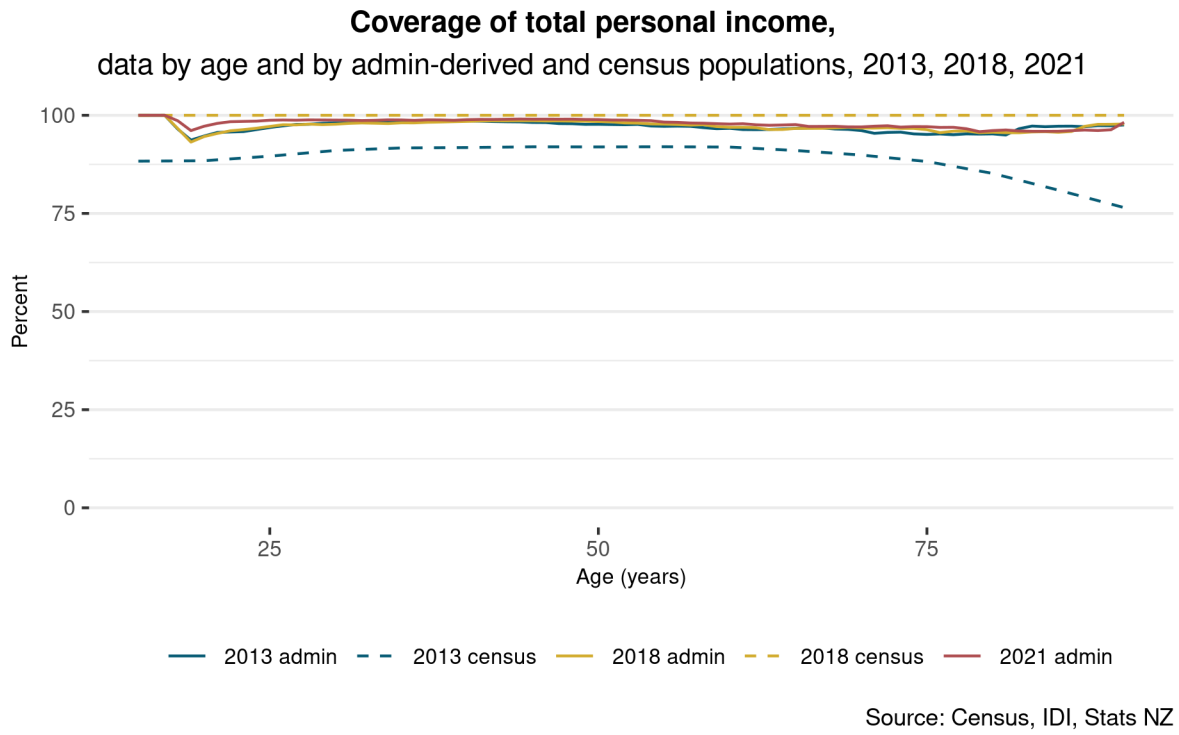


Source: IDI, Stats NZ

Comparing coverage for other demographic variables, we find very little difference in coverage by Māori descent categories, but some small differences by level 1 ethnic group. European, Māori, Pacific, and Other ethnic groups have higher coverage values (consistently around 98 percent over time) than Asian and Middle Eastern/Latin American/African (MELAA) ethnic groups, which generally maintain coverage at around 93.5 percent. However, with the implementation of the IR automated assessment in recent years, this coverage has increased to around 97 percent. See Appendix D.

Figure 2 shows the availability of income information by age, comparing census total personal income with the income for the admin population. Coverage starts at 100 percent for individuals under the age of 18, due to the use of age to help determine zero income, before a sharp dip to around 93 percent for the 19-years age group. The dip is fuelled primarily by recently arrived overseas migrants who are engaged in study. Also of note is the lower dip at 18 for the 2021 curve. This is due to the additional IR automated assessment data source being available from 2019.

Figure 2. Coverage of total personal income, data by age and by admin-derived and census populations, 2013, 2018, 2021



Source: Census, IDI, Stats NZ

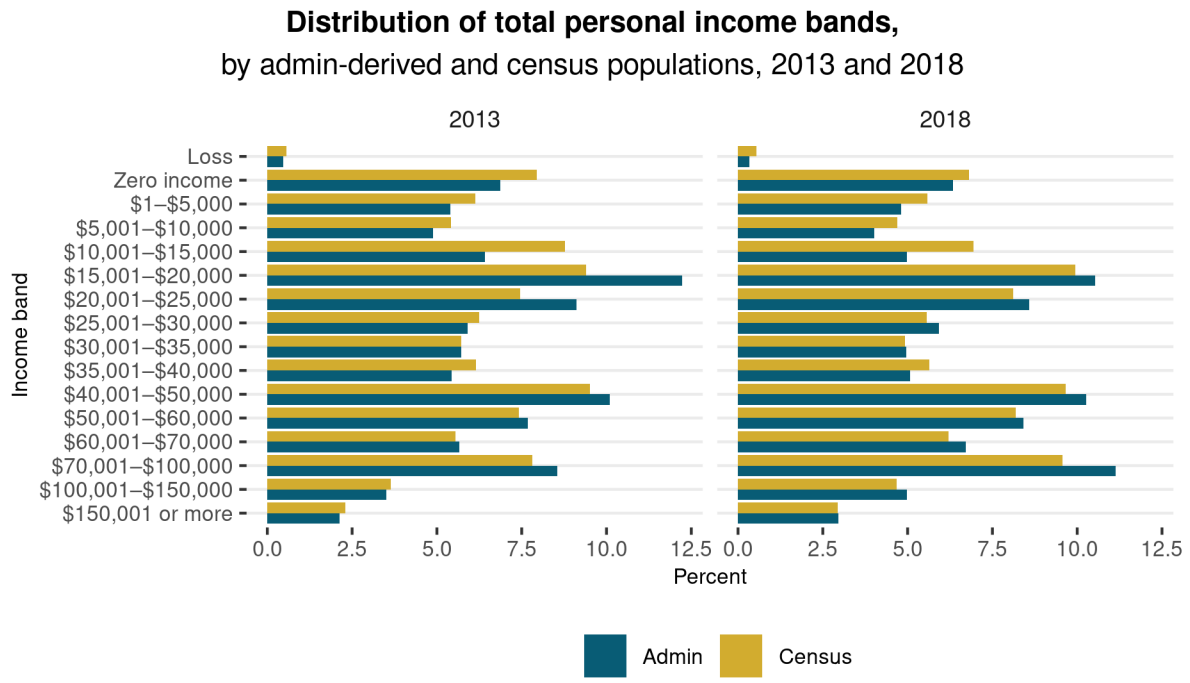
Census coverage of income in 2013 was consistently high at around 90 percent across all ages, though with a slight drop at ages 15 and 16 years. Coverage in 2018 was consistent at 100 percent over all ages, due to the use of administrative data and statistical imputation.

Distributions

The distribution of total personal income by income band is shown in figure 3 for the census and the admin populations, in 2013 and 2018. Distributions are calculated as the percentage of people in each income band out of 'stated' responses, that is, excluding missing data. We note that the 2018 Census dataset overlaps somewhat with the admin-derived population since 16 percent of total personal income is derived from the same admin sources.

The distribution across income bands is largely comparable (less than 1 percent difference between the band occupancies) between the admin population and the census for both 2013 and 2018, with overall agreement being slightly better for 2018 than 2013. Agreement has improved with 2013 Census relative to the paper of Suei (2016), indicating that the improved data sources available within the IDI since that report do improve the ability of administrative data to derive income information, even accounting for the use of the zero-income band in this analysis that was excluded from analysis in the previous work.

Figure 3. Distribution of total personal income bands, by admin-derived and census populations, 2013 and 2018



Source: Census, IDI, Stats NZ

Noticeably, the bands between $10,001 and $25,000 have the largest discrepancies between the admin and 2013 Census populations. The $15,001–$20,000 income band largely coincides with New Zealand superannuation rates pre-tax in 2012 and 2013. After tax, New Zealand superannuation largely falls into the $10,001–$15,000 income band. Some working-age benefits also fall into these bands. 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. This indicates that there was a potential source of error in the census data whereby some individuals receiving New Zealand superannuation as their primary income source reported their after-tax income amount, rather than the pre-tax income amount that census asks for. This is also true to a lesser extent for 2018 data, as New Zealand superannuation and other benefits shifted further away from the income band boundaries meaning less potential for pre-tax and after-tax incomes to fall into different income bands.

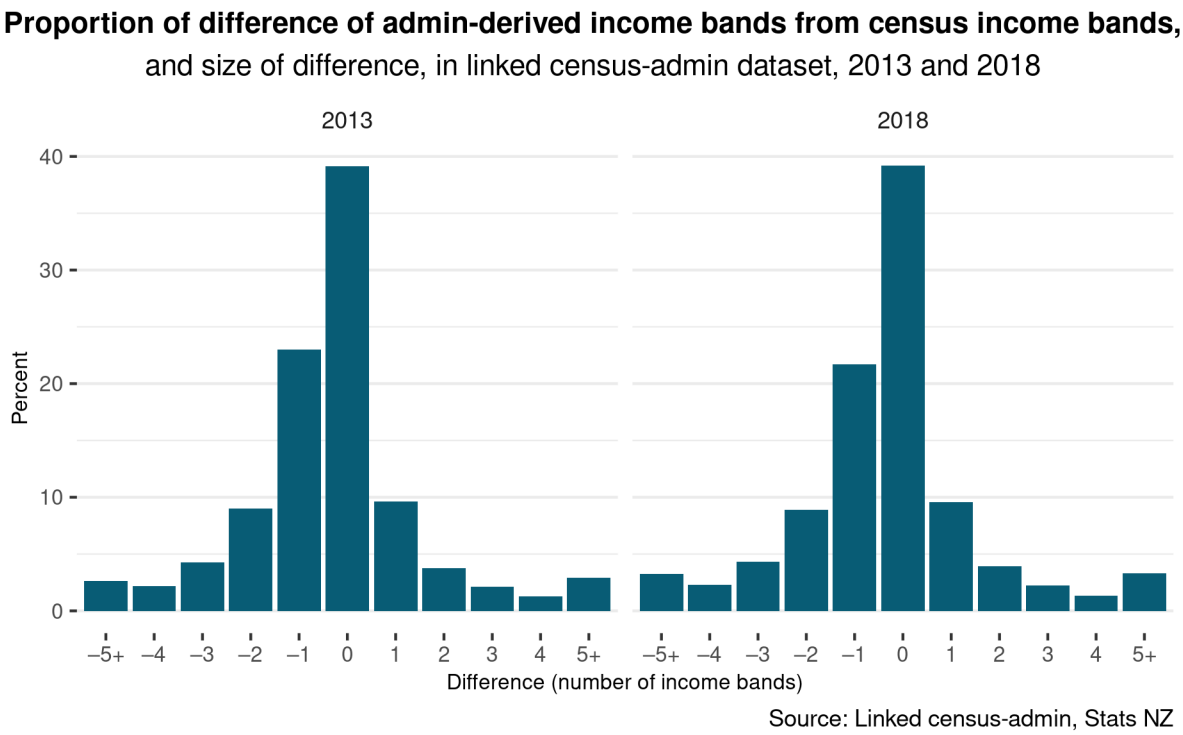
An additional consideration for income band discrepancies between census data and administrative data is the income tables included in the guide notes of the census. As shown in Appendix B, these guide notes provide a correspondence between weekly or fortnightly after-tax income and total pre-tax annual income. However, these tables do not account for any student loan repayments, other debt repayments, child support payments or employee contributions to KiwiSaver, which are automatically deducted from wages and reduce the amount of income an individual will receive in their bank account, and so potentially their perception of their after-tax income. Student loan repayments in particular could cause reporting of lower income than reality as it is repaid at 12 cents of every dollar earnt over a repayment threshold.

Comparison of individual-level records for personal income

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

Of the 3,042,513 in 2013 and 3,547,983 in 2018 usually resident individuals aged 15+ in the linked census-admin dataset, total income information is available for 2,980,584 (98.0 percent) in 2013 and 3,482,409 (98.2 percent in 2018). Figure 4 shows the distribution of the discrepancies between income bands derived from census and administrative data. No difference means an individual has the same band in both data sources, with 1 or -1 representing one band difference (for any income band category), and so on. A negative difference indicates that the admin data income band is higher than the census income band.

Figure 4. Proportion of difference of admin-derived income bands from census income bands, and size of difference, in linked census-admin dataset, 2013 and 2018



The discrepancies are skewed towards negative differences, where admin derived income is higher than census income. Both census years have similar percentage discrepancies. In 2013, 41 percent are in higher income bands in the admin data than in the census (40 percent in 2018), while in both years 20 percent have lower income in the admin data than is self identified in the census. This is a larger weighting towards higher income than previously reported by Suei (2016), showing improved investment income availability having an effect on the derived income.

Full matrices 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 C. 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 admin 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.

The under-reporting of income is not unique to the New Zealand census and has been found 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).

The **quality rating score** for total personal income is derived by comparing the consistency of income bands between individuals with valid values in both census responses and admin data. We allow for misreporting in census around the boundaries of income bands and assume that higher admin-derived incomes are more accurate than census responses. Consistency is defined as an individual's admin income band being within one band of that reported in the census, or two or more bands higher. The quality rating score for total personal income is 0.90 in 2013 and 0.89 in 2018.

The quality rating scores across the three Māori descent categories are similar to the overall quality rating score, being between 0.90 and 0.92 in 2013 and 0.89 and 0.91 in 2018. However, the skewedness of the three categories is different. The Māori descent and 'don't know' categories are more skewed towards negative values than the no Māori descent category (45/46 percent versus 40 percent). A comparable situation is present when these scores are broken down by level 1 ethnic group. All six ethnic groups have quality scores close to the overall score (between 0.88 and 0.92), but the skew towards negative values is more pronounced in the Māori (47 percent) and Pacific (51 percent) ethnic groups.

Results for sources of personal income

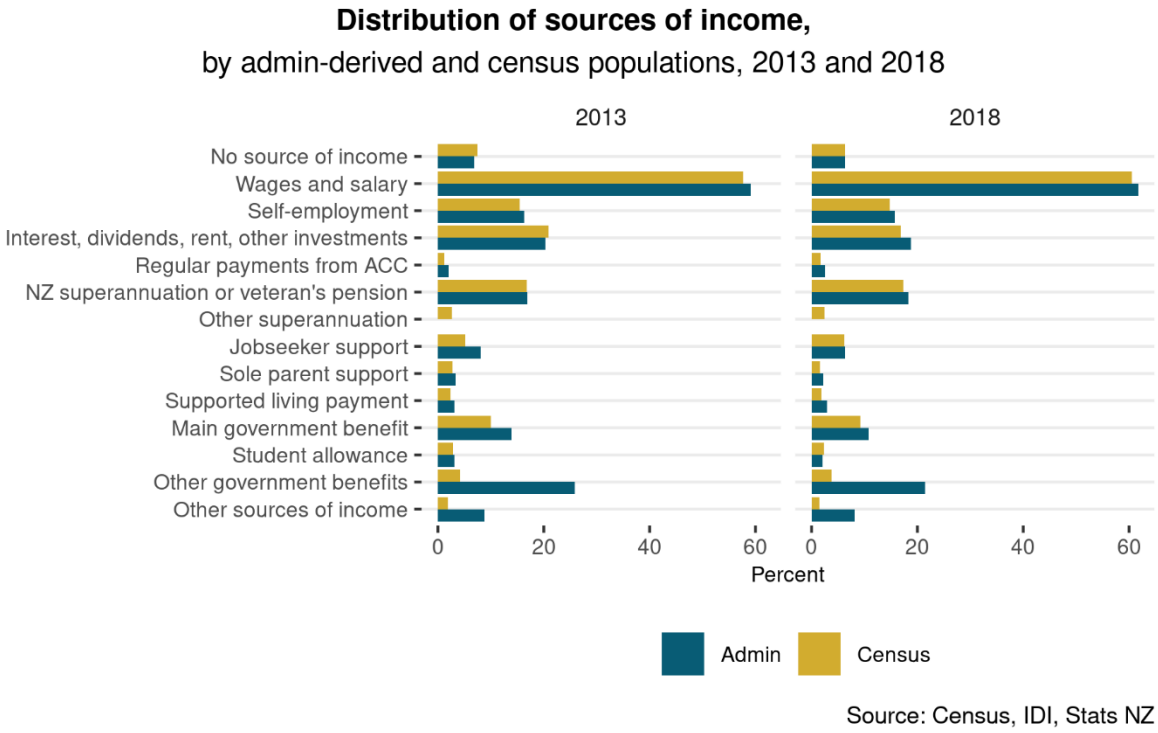
Aggregated comparisons for income source

This section presents distributional comparisons between the admin population estimates and the census results for sources of personal income.

As noted earlier, admin coverage for sources of personal income is the same as total personal income and is consistently high at around 97 percent.

The distribution of sources of personal income is shown in figure 5 for the census and the admin populations, in 2013 and 2018. The benefit categories in 2013 Census have been recoded to match with the categories used in the 2018 Census. Individuals can have more than one income source. 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. Full results with are provided in Appendix C.

Figure 5. Distribution of sources of income by admin-derived and census populations, 2013 and 2018



Around 7 percent of individuals reported no source of income in the two censuses, similar to the number reporting zero income. This is similar to the 6.9 percent in 2013 and 6.3 percent in 2018 derived with no source of income in the admin population, which is much improved compared with previous work that had no individuals identified with zero income.

By far the largest group in both sources are people earning income from wages and salaries, with investment income (interest, dividends, rent, and other investments), New Zealand superannuation, and self-employed income the next most common in both censuses. Similar patterns are observed in the admin distributions. The most significant differences between census and administrative data are the proportions in the ‘other benefits’ and ‘other income’ categories.

Table 3 shows the ratio of the admin population total response percentages to the census total response percentages for each income source category. A ratio close to one shows that census and admin-data-based estimates for a category are highly consistent, independent of what other income sources may be reported.

Table 3. Ratio of admin-derived income to census data, by income source, 2013 and 2018

Ratio of admin-derived income to census data, by income source, 2013 and 2018		
Income source	**Ratio admin-derived/census**	
	2013	**2018**
Wages and salary	1.02	1.02
Self-employment	1.06	1.07
Investments	0.97	1.12
ACC payments	1.78	1.54
NZ superannuation	1.01	1.06
Other superannuation	0.00	0.00
Jobseeker	1.56	1.03
Sole parent support	1.23	1.35
Supported living	1.33	1.62
Student allowance	1.07	0.89
Other benefits	6.19	5.62
Other income	4.56	5.45
No source of income	0.92	1.00
Source: Stats NZ		

Wages and salaries, self-employment, investment, New Zealand superannuation, and no source of income are the most consistent across all income sources and both censuses, with ratios of between 0.92 and 1.12.

It is of note that previous work by Suei (2016) found at the time there was very poor proportional agreement between 2013 Census data and the tax data available for the investment income source. Census results had a much larger proportion of individuals reporting investment income than was detected in administrative data. Due to improvements in the administrative data available, this is no longer the case, with the proportions being much closer to unity.

The very high ratios for ‘other benefits’ and ‘other income’ categories is likely due to misunderstanding or misassigning income sources when answering the census questions.

For example, a major cause of the proportion of the ‘other benefits’ category being higher in administrative data relative to census data, is due to misassignment of supplementary benefits when responding to the census and, to a lesser extent, misassignment of WFF tax credits. If the supplementary benefits an individual receives are recoded the same as the main benefit they are linked to, the proportion of ‘other benefits’ drops to 17.9 percent in 2013 and 14.8 percent in 2018. There is no mention of how to handle these supplementary benefits within the census guide notes, and an individual will generally receive only a single combined payment from MSD. The guide notes do indicate that receiving WFF tax credits should be marked as ‘other benefits’.

With the ‘other income’ category, a potential source of the higher proportion in administrative data relative to census data could be due to not reporting some income sources. For example, there is no mention of overseas income in either the guide notes or the income sources question. In administrative data, self-reported overseas income available from IR3 data is coded as ‘other

income'. In both 2013 and 2018 admin populations, there are more than twice as many individuals with self-reported overseas income found in the IR3 data than there are individuals who report 'other income' as a source in the census data.

Some other categories also show large discrepancies. For example, there are many more people with ACC and the main working-age benefit income sources (except for jobseeker support in 2018) in the admin population than there are in the censuses.

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. Additionally, the employer pays the first week of ACC, so if a person was only off for a few days, they would be paid solely by their employer but they may have thought they were receiving ACC payments. Taxable ACC payments are a replacement of work income, and are typically small amounts – in 2018, 51 percent are less than $5,000, and almost all (80 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 ratios for the main working-age benefits appears to reflect under-reporting of benefits as a source of income in census. The much lower ratio for the jobseeker category in 2018 reflects the use of administrative data sources to help alleviate low response rate, as the majority of responses in the jobseeker category are derived from administrative data. The other main working-age benefit categories did not use administrative data to fill in gaps.

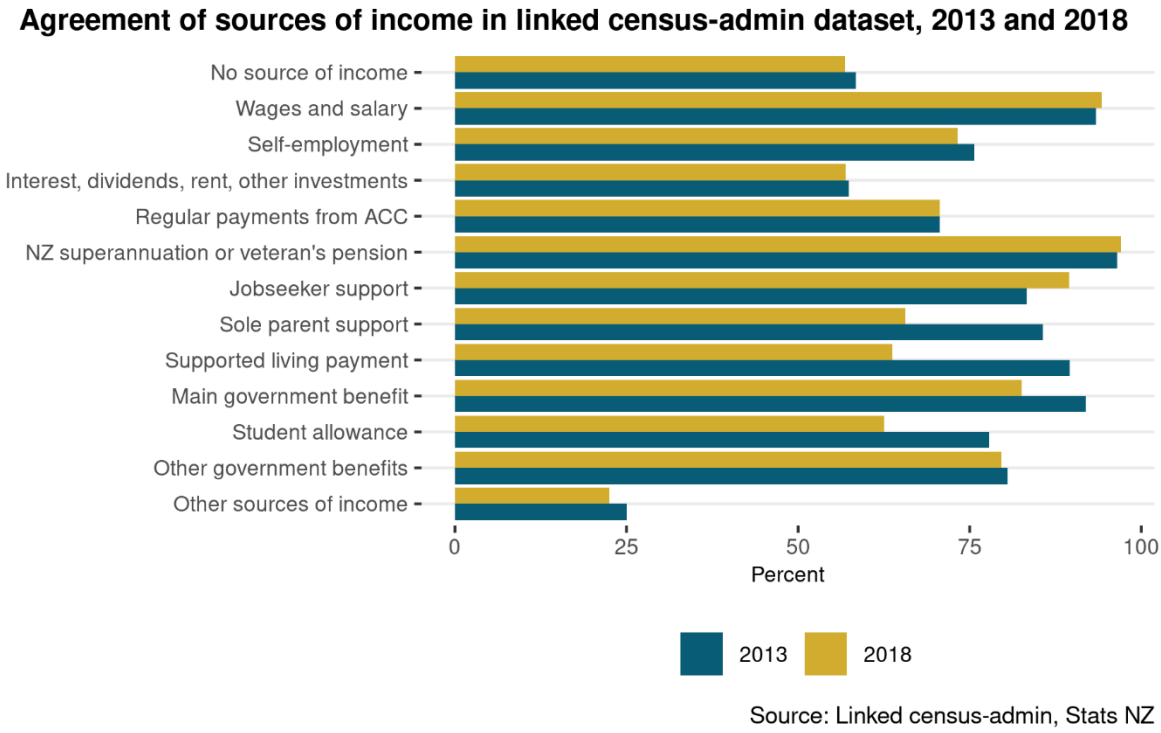
Comparison of individual-level records for income source

For the following analysis we have used the linked census-admin 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 3,042,513 in 2013 and 3,547,983 in 2018 usually resident individuals aged 15 years and over in the linked census-admin dataset, total income information is available for 2,980,584 (98.0 percent) in 2013 and 3,482,409 in 2018 (98.2 percent). Overall, 44.6 percent in 2013 and 41.4 percent in 2018 have the same combination of source of income in both census and IDI, with 89.1 percent in 2013 and 92.1 percent in 2018 having at least one source of income consistent between census and IDI data. However, 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 or New Zealand superannuation, more than 90 percent of the time the administrative data agrees in both 2013 and 2018. Agreement for the main working-age benefits is lower on an individual benefit level in 2013, but when all three (Jobseeker Support, Sole Parent Support, and Supported Living Payment) are combined into a working-age benefit category ('main benefit') the agreement also exceeds 90 percent. In 2018 this is not the case due to much lower agreement with the sole parent support and supported living payment sources.

Figure 6. Agreement of sources of income in linked census-admin dataset, 2013 and 2018



Agreement is somewhat lower for self-employment, student allowances, and ACC. Administrative data is expected to be a reliable record of receipt of student allowance and ACC as income sources. At the individual level of self-employment there is more discrepancy between the income data sources than the good agreement at the aggregate level. 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.

Nearly 60 percent of the population reporting investment income in the census were identified as having investment income from the administrative data. This is a major improvement from previous work where less than 5 percent of the population were identified with investment income in both census and administrative data sources.

The worst performing income source is the 'other income' category, with around 25 percent agreement between census and administrative data sources. This is expected as the overlap between the category in the census question and administrative data is not well defined.

Additional data sources

Since the 2018 Census, newer data sources have become available in the IDI for determining income. As such these data sources have not been compared with census data. The new data is the IR Automated Assessment (AA) table introduced by Inland Revenue from the 2019 tax year, and the best start tax credit (BSTC), introduced in July 2018.

In this derivation, only a few fields from the AA table are utilised, namely those related to income from interest and dividends, Māori authority distributions, and (for records from 2021 onwards), portfolio investment entities (PIE), such as KiwiSaver funds. The BSTC is an additional tax credit controlled jointly by MSD and IR, and is available within the tier-two MSD benefits table, and the IR-supplied WFF table. Counts of each of these income sources for the 2021 year are given in Table 4.

Table 4. Count of new income sources, by data supply and category, 2021 tax year

Count of new income sources, by data supply and category, 2021 tax year			
Data supply	**Source of income**	**Category**	**Count**
IR, Automated Assessment (AA) table	Interest	Investment income	1,643,637
	Dividends	Investment income	137,595
	Portfolio Investment Entities (PIE)	Investment income	1,799,346
	Māori authority distributions	Other income	10,596
IR, Working for families (WFF)	Best start tax credit (BSTC)	Other benefit	92,685
MSD	Best start tax credit (BSTC)	Other benefit	18,789
Source: Stats NZ			

In the AA table, the Māori authority distributions income source is coded as 'other income', while all other sources are coded as 'Investment income'. Like the other tax credits, the BSTC is coded as 'other benefit'.

The inclusion of the Māori authority distributions income source is not expected to have a major influence on the proportion of individuals in the 'other income' category. The number of individuals receiving Māori authority distributions income is low, and the 'other income' source is already a much larger proportion of the population in admin data than in census data. The same is true for the BSTC income source, though in regards to the 'other benefit' census source.

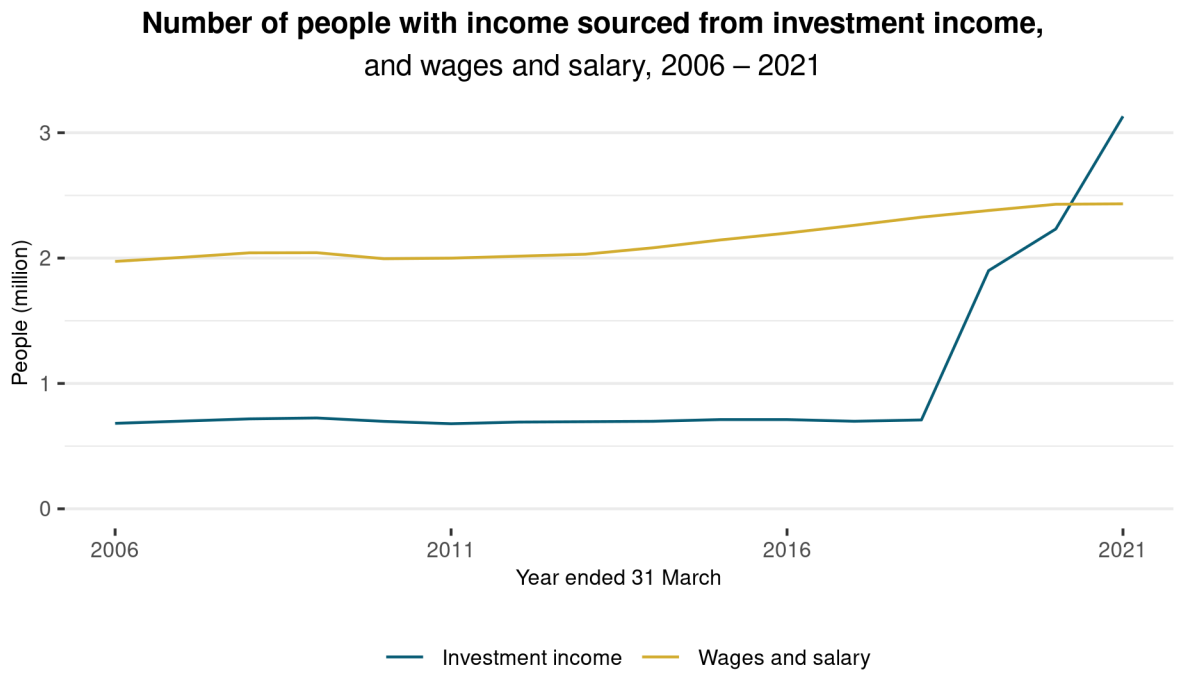
'Investment income' sources in the AA table will result in major differences relative to census data as there are a large number of individuals receiving investment income from interest and PIE.

Interest is income received from various sources over the year, notably including any interest on bank deposits that an individual may have. In 2021, 32.4 percent of individuals with this income source in the admin data earn less than $1 from it, with 78.4 percent earning less than $100 and 90.5 percent earning less than $1,000.

PIE income includes KiwiSaver providers. KiwiSaver was introduced in 2007, and by the end of June 2013, over 2 million individuals were members, reaching over 3 million by the end of June 2022 (Inland Revenue, 2021). The majority of these members would have non-zero balances earning returns.

When combined with the other 'investment income' sources, the 'investment income' category becomes the major populated category from 2019, overtaking the 'wages and salary' category. This is shown in figure 7.

Figure 7. Number of people with income sourced from investment income, and wages and salary, 2006–2021



Source: IDI, Stats NZ

This large increase in the 'investment income' category is a marked difference from what is seen in census data. It suggests that the majority of individuals earning relatively small amounts of income from interest have not reported this as 'investment income' in the census. Similarly, the numbers reporting 'investment income' in both 2013 Census and 2018 Census is far lower than KiwiSaver membership numbers, again indicating that PIE income is not regularly reported as an income source in census. This may be due to the rules on withdrawing funds from a KiwiSaver account meaning that most people do not see this as income because it is not readily available to them.

Zero-income earners in census and IDI

In this section we evaluate the admin derivation of zero-income earners by comparing admin-derived zero income with census responses, using the linked census-admin dataset. We show results only for 2018 Census. Similar patterns were found for comparisons with the 2013 Census.

Since there are few direct indications of zero income in the admin sources, we derive zero income using a set of rules targeting situations which are mostly likely to indicate that people have no source of income. Table 5 shows the cross-tabulation of census responses with admin derivations by zero and non-zero income, for people with zero income in one of the linked datasets. Census responses report more people with zero income than we derive in the admin data. However, almost 40 percent of those reporting zero income in the census do have some admin income. The majority of the admin income amounts are less than $10,000, with 44 percent being less than $5,000 and 58 percent less than $10,000. The main sources of admin income for people who report zero income in the census are wages and salaries and 'other benefits', providing further evidence of under-reporting of income in the census.

Conversely, 63 percent of individuals who identified as having zero income from the admin sources also report zero income in the census. Of those who have admin-derived zero income but non-zero income in the census, 34 percent report less than $10,000 in the census, which includes about 10 percent in the loss category. There is an even spread across the remaining income bands in census. This could indicate that some income sources are not currently being captured in the admin derivation or could be due to census respondent error. Of those where we are still missing admin-derived income information, close to 80 percent do report some income in the census. These results suggest the methods used to derive zero income are well-targeted to those without any income.

Table 5. Comparison of zero income earners between census and admin data, linked data only, 2018

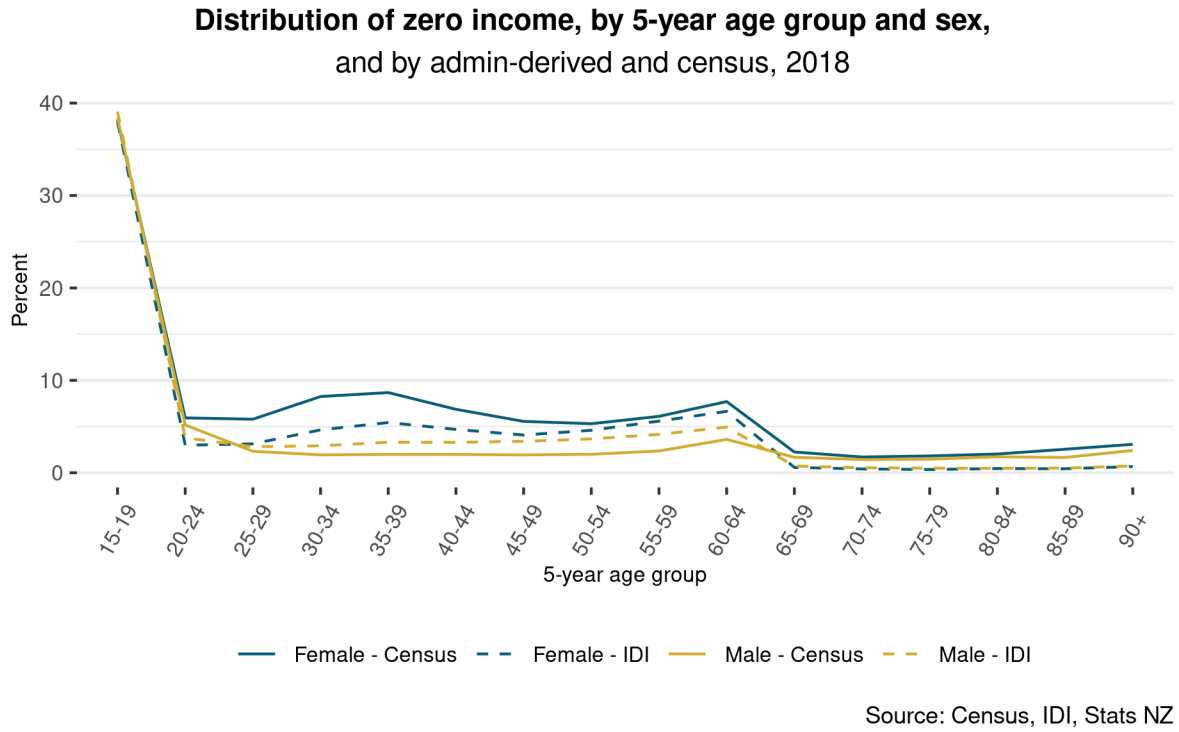
Comparison of zero income earners between census and admin data, linked data only, 2018

Census/admin-derived	Zero income	Non-zero income	Missing	Total
		Count (percent)		
Zero income	126,717 (55.4%)	82,374 (36.0%)	19,785 (8.6%)	228,876 (100%)
Non-zero income	75,771 (62.0%)	…	46,431 (38.0%)	122,202 (100%)
Total	202,488 (57.7%)	82,374 (23.5%)	66,216 (18.8%)	351,078 (100%)

Symbol: … data excluded
Source: Stats NZ

Figure 8 shows the percentage of individuals with zero income, for the census and admin data by 5-year age group and sex. The distributions are largely consistent between census data and admin data. However, there are some differences by gender. The admin data has fewer females with zero income relative to census from 20–24 to around 45–49 age groups. The opposite occurs for males where admin data tends to show a slightly higher proportion of working-age males with zero income than the census.

Figure 8. Distribution of zero income, by 5-year age group and sex, and by admin-derived and census, 2018



Zero-income derivation methods

Of interest is what applied methods resulted in individuals being assigned zero income. Four main methods are used to determine zero income. Individuals may fall under multiple categories, but the methods are applied sequentially, and only those with no currently derived income at each step. Numbers derived by each method are provided in Table 6.

Table 6. Number of individuals assigned zero income by derivation method, 2018

Number of individuals assigned zero income by derivation method, 2018	
Derivation method	Individuals determined
Found in any IR table	29,601
Student loan living costs	10,431
Previous tax system interaction	85,905
Enrolled at secondary school	107,589
Aged 17 years and under	5,307
Source: Stats NZ	

The first method is assigning zero income to those with an entry in any of the IR tables used for the main income derivation, but not having any income assigned. (An entry with no income assigned indicates zero income from that table.) If these entries had non-zero income and were the only source of income for an individual, that individual would be treated as only having income from that source. Thus, applying the same standards with zero income maintains consistency.

An individual receiving living costs payments as part of a student loan is the next method used to assign zero income. Living costs payments are effectively a source of income, but since they are part of a loan, are not classified as income.

The third method assigns an individual zero income if they have had income derived in previous tax years, but it is missing in the current year of interest. This method is the largest group for post-school age individuals. There is a fairly even spread across the age bands from 20–24 years until a marked decrease at 65 years. This drop coincides with the age of qualification for NZ superannuation, which covers most individuals aged 65 and over in New Zealand.

The final method is primarily based on age. A large proportion of individuals with zero income are the younger section of the population. To account for this, two steps are used to decide if a young person should be regarded as having zero income. Firstly, if there is evidence of current enrolment in a secondary school through the Ministry of Education enrolments table, and the individual is 18 years or younger and does not have any derived income for the year, they are given zero income. This accounts for most people not being employed while at school. The last step is a straight age cut-off. If an individual is 17 years or younger, they are given zero income. This last step adds only a few thousand individuals with zero income over and above those added by the school enrolment filter.

Admin-derived zero income counts decrease markedly from 2019 with the availability of the AA data source. Many of those whose only income source is small amounts of interest payments received or PIE income would report zero income in the census. If we wish to maintain consistency with time series, and with what people generally regard as having ‘no income’, some adjustment to how output categories are derived is needed. For example, a lower limit to the amount of certain income sources could be imposed before it is defined as non-zero income.

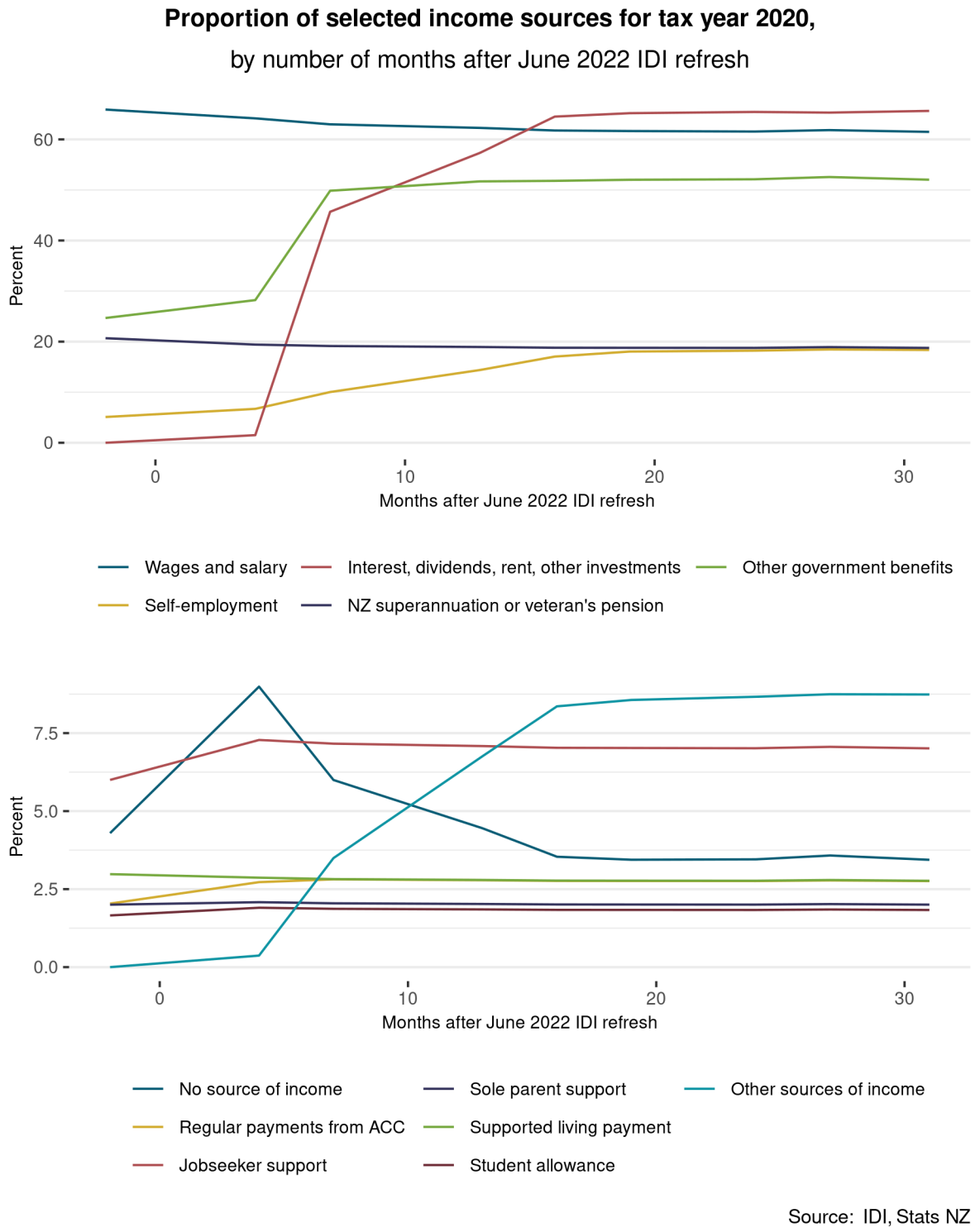
Timeliness of data availability

The IDI is not a real-time data repository. Data providers provide raw data to Stats NZ on a regular basis, each with their own schedule. These are then collated and organised into an IDI refresh, which is released to researchers every three to four months. Due to this update schedule, there is lag inherent in the availability of the data and the time period it is available for. Complicating this is that some sources, such as IR3 tax returns, have a lag in their reporting periods, meaning that the data does not become available to the supplier until some time after the reference period. This data lag will have an impact on the reporting of timely information.

To investigate the impact of time lags, the income derivation code was run for the 2020 tax year and 2021 tax year (when possible) on each refresh from January 2020 to October 2022, covering nine refreshes. We compare the results from each refresh before June 2022 with the June 2022 refresh used for the results in this paper. The October 2022 refresh provides some extension of data, particularly for the 2021 tax year.

Figure 9 shows the proportion of the June 2022 refresh population with income from each income source for the tax year ended 31 March 2020, as calculated over the series of refreshes. Equivalent tables are provided in Appendix C for sources of income and total personal income for the tax years 2020 and 2021.

Figure 9. Proportion of selected income sources for tax year 2020, by number of months after June 2022 IDI refresh



Source: IDI, Stats NZ

Timeliness of data supply falls into two main groups of income sources. The most timely information is for wages and salary, ACC payments, New Zealand superannuation, student allowance, and the

four benefit sources. All data up to the end of the tax year data is available in the June/July refreshes released around four months after the conclusion of the tax year. This time lag is dependent on the supply schedule of the data suppliers, as well as the IDI refresh cycle.

The second group includes self-employment, investment, and other income sources. This group has a time lag of up to 24 months after the end of a tax year, although most data is available within 16 months. This lag is driven by rules around the filing of tax returns, in particular the IR3 return, meaning that they can be filed up to 12 months after the end of a tax year. The other source of delay is due to the IR automated assessment process being undertaken between the end of May and end of July each year, meaning that this data is only available in the IDI at least five months after the end of the tax year.

The distributions of total income depend on the availability of income sources, and so are also affected by time lags. The zero-income category is over-represented in refreshes released closer to the end of the tax year and improves in later refreshes as more data sources become available. This is due to zero income being incorrectly assigned to individuals whose only income source is from annual returns, as those annual returns become available later.

The proportions of those in lower income bands tends to be too high in refreshes released soon after the end of the tax year, while higher income bands are under-represented. This is due to the lower income bands being dominated by individuals earning wages and salary income, or benefits, which are more timely than the self-employment and investment income that dominates in the higher income bands.

Conclusion

This paper presents our investigation into the potential for deriving census income information from administrative sources, updating previous work (Suei, 2016). We focus on understanding how recent changes to the tax and benefit income data available in the IDI affect our ability to measure personal income, and the benefits of using admin data for income beyond what census data can currently provide.

We compared income variables between 2013 Census and 2018 Census with similar information derived from administrative data available in the IDI in June 2022.

Summary of overall results

Concepts

The concepts of money income calculated before tax over an annual period as defined by the statistical standard are readily derived from administrative data. All types of income sources found in the admin data can be mapped to an appropriate category in the statistical standard.

Coverage

Coverage for admin-derived income is around 97 percent, an improvement from the 88 percent found in the Suei (2016) study. This increased coverage is largely due to introducing a derivation for zero income, which is not directly available from admin sources. The admin coverage is higher than the approximately 90 percent item response rate for income achieved by censuses before 2018.

Newly available tax and benefit data has improved measurement of certain income sources, in particular some classes of government benefits and different types of investment income, substantially addressing a major shortcoming identified previously.

The administrative sources available in the IDI now include income from all income source categories, with the exception of private or overseas superannuation data. The most complete data is available from 2019 when Inland Revenue introduced automatic assessment (AA) for individuals that includes all annual income from employment, benefits, and a range of investment types.

Additional income data that may be collected by census but not admin data includes non-taxable or cash jobs, transfers between households, and earnings from ‘underground’ marketplaces, although it is uncertain how much would be reported by census respondents.

Timeliness

Timeliness of income information derived from administrative sources is affected by the time taken for several steps in the process: the time needed for the agency to collect the data, and then to provide the data to Stats NZ; and, in the current system, for the data to be processed in the IDI and made available in regular IDI refreshes. With the current IDI refresh cycle, income data is available in the IDI around four months after the end of the tax year for data that is collected by agencies on a monthly or more frequent basis. This includes income from wages and salary, New Zealand superannuation, and working-age benefits. There is a longer delay of around 16 and up to 24 months before all income derived from annual tax returns is available. This includes self-employment and investment income.

Comparison with the census

Income recorded through the taxation and benefit systems 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, which is now a small component of income received. In contrast, the census is constrained by the limitations of a self-complete questionnaire and relies on a respondent's correct interpretation of the questions on income source and total gross income, their correct recall, correct identification of their source(s) of income, and ability to calculate total gross income.

The distribution of income bands and the prevalence of most income sources is broadly similar between the census and the admin-derived population. However, detailed comparisons reveal issues with census results for some categories. For example, census respondents tend to report lower income than the formal tax system, possibly because some are providing net rather than gross income. The discrepancy is most noticeable for income bands that reflect annual income from government benefits and New Zealand superannuation. Some working-age benefit categories, 'other income', and ACC are under-reported as income sources by census respondents.

The additional information available from the tax system since 2019 for interest income and PIE income such as KiwiSaver is likely to impact comparisons with the 2023 Census. Investment income as measured by the admin sources is likely to be much higher than reported by census respondents, and many of those whose only income is from small amounts of interest payments are likely to report zero income in the census.

Benefits and limitations of administrative-derived income

Getting income data from administrative data sources is both more accurate and more detailed than is possible with a census questionnaire.

Administrative data from the taxation and benefit systems on income and income sources has several advantages over survey collection. Admin data has income in dollar values rather than income bands and distinguishes income by each source. This increases the range of data available, for example, income distributions can include higher income categories, and can be provided by income source. This work has focused on the two current census income variables. However, admin-based income measures have potential to be extended to other concepts of income, such as net income after tax or total disposable income.

Admin data is available in time periods ranging from annually for some tax data, to monthly for wage and salary information, and daily for some benefits data. Outputs can be produced more frequently than is possible in a periodic survey-based census.

Admin-based personal income can also be combined with household and family data to provide household and family income derivations.

The main limitation of the administrative sources is the lack of information about private or overseas superannuation. There are also some admin sources not yet being used in the income derivation, for example, data from IR6B returns relating to income from estate or trusts. These and other data sources may improve accuracy at the margins for certain income categories. There is always likely to be a small amount of missing data, and statistical imputation will be needed to fill remaining gaps.

The other main limitation is the timeliness of income sources that use annual income tax filing.

Admin data for income now better than census

The availability of additional data sources and the derivation for zero income have resulted in substantial improvement in the areas that were of concern in the previous 2016 investigation. The administrative sources now available through government tax and benefit systems provide high quality, more detailed, and more frequent information about total personal income and income sources that goes beyond what can currently be achieved through a census questionnaire.

References

Katz, A. J. (2012). *Explaining long-term differences between census and BEA measures of household income*. Retrieved from www.bea.gov.

O'Byrne, E., Bycroft, C., & Gibb, S. (2014). *An initial investigation into the potential for administrative data to provide census long-form information*. Retrieved from www.stats.govt.nz.

Inland Revenue. (2021). *KiwiSaver member demographics.* Retrieved from www.ird.govt.nz.

Stats NZ. (2016). *Guide to reporting on administrative data quality.* Retrieved from www.stats.govt.nz.

Stats NZ. (2019a). *Data sources, editing, and imputation in the 2018 census*. Retrieved from www.stats.govt.nz.

Stats NZ. (2019b). *Overview of statistical methods for adding admin records to the 2018 census dataset.* Retrieved from www.stats.govt.nz.

Stats NZ. (2022). *Experimental administrative population census: Data sources, methods, and quality (second iteration).* Retrieved from www.stats.govt.nz.

Stats NZ. (nd). *Classifications and related statistical standards*. Retrieved from www.stats.govt.nz.

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

Zabala, F. (2016, June). *Using administrative data to validate income in Statistics New Zealand household surveys*. Retrieved from statswiki.unece.org.

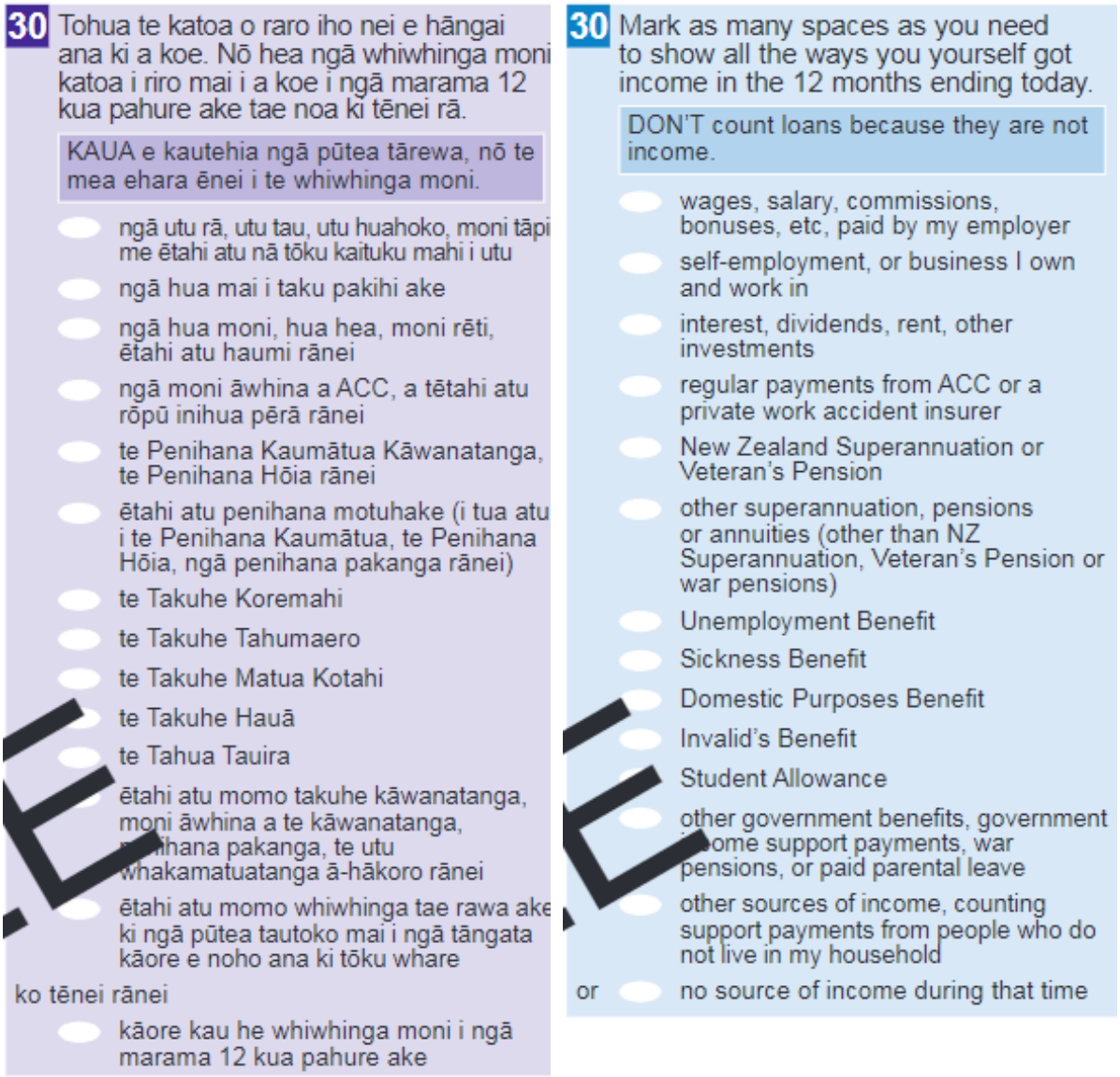
Zhang, L.-C. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*, *66*(1), 41–63. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9574.2011.00508.x

Appendix A: Census income questions

Here we provide the census questions from 2013 and 2018 that relate to the personal income attributes. We also provide the guide notes. These questions are from the paper form.

2013 Census questions

The questions in the 2013 Census were questions 30, about sources of income, and question 31, about total income.



30 Tohua te katoa o raro iho nei e hāngai ana ki a koe. Nō hea ngā whiwhinga moni katoa i riro mai i a koe i ngā marama 12 kua pahure ake tae noa ki tēnei rā.

KAUA e kautehia ngā pūtea tārewa, nō te mea ehara ēnei i te whiwhinga moni.

- ngā utu rā, utu tau, utu huahoko, moni tāpi me ētahi atu nā tōku kaituku mahi i utu
- ngā hua mai i taku pakihi ake
- ngā hua moni, hua hea, moni rēti, ētahi atu haumi rānei
- ngā moni āwhina a ACC, a tētahi atu rōpū inihua pērā rānei
- te Penihana Kaumātua Kāwanatanga, te Penihana Hōia rānei
- ētahi atu penihana motuhake (i tua atu i te Penihana Kaumātua, te Penihana Hōia, ngā penihana pakanga rānei)
- te Takuhe Koremahi
- te Takuhe Tahumaero
- te Takuhe Matua Kotahi
- te Takuhe Hauā
- te Tahua Tauira
- ētahi atu momo takuhe kāwanatanga, moni āwhina a te kāwanatanga, penihana pakanga, te utu whakamatuatanga ā-hākoro rānei
- ētahi atu momo whiwhinga tae rawa ake ki ngā pūtea tautoko mai i ngā tāngata kāore e noho ana ki tōku whare

ko tēnei rānei

- kāore kau he whiwhinga moni i ngā marama 12 kua pahure ake

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

31 Mai i ngā momo whiwhinga moni katoa i tohua e koe i te pātai **30**, e hia te nui o te katoa o ēnei moni:

- i riro mai i a koe
- i mua i te tango mai o ngā tāke me ētahi atu āhuatanga
- i ngā marama 12 tae atu ki te 31 o Poutū-te-rangi 2013

- nui ake te tango moni i te whiwhi moni
- kāore he whiwhinga moni
- $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 nui atu rānei

Tirohia ngā Kupu Whakamārama hei āwhina i a koe ki te whiri i tō whiwhinga moni

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

30 31 Why do you want to know my income?

Income statistics are used for developing social and economic policy, research and monitoring programmes. All of the answers you give are kept confidential.

Remember

- If you and your spouse / partner earn income jointly, only include your part of that income.
- If you received Working for Families payments (including Family tax credit, In-work tax credit, Minimum family tax credit and Parental tax credit), mark 'other government benefits ... '.
- If you received homestay or child support payments, mark 'other sources of income ... '.
- If you did piecework, mark 'wages, salary, commissions, bonuses, etc ... '.

Count any payments that are taken out of your income **before** you get it, such as repayments of student loans, union fees, fines or child support.

DON'T count loans (including student loans), inheritances, sale of household or business assets, lottery wins, matrimonial / civil union / de facto property settlements or one-off lump sum payments.

DON'T count money given by members of the same household to each other. For example, pocket money given to children, or money given for housekeeping expenses by a flatmate.

If you know your weekly or fortnightly income **after tax**, use this table to work out your annual income **before tax**.

Annual income (before tax)

After tax weekly income $	After tax fortnightly income $	Before tax annual income $
up to 86	up to 172	1 – 5,000
87 – 172	173 – 343	5,001 – 10,000
173 – 256	344 – 512	10,001 – 15,000
257 – 335	513 – 671	15,001 – 20,000
336 – 414	672 – 829	20,001 – 25,000
415 – 493	830 – 987	25,001 – 30,000
494 – 573	988 – 1,145	30,001 – 35,000
574 – 652	1,146 – 1,303	35,001 – 40,000
653 – 805	1,304 – 1,610	40,001 – 50,000
806 – 939	1,611 – 1,879	50,001 – 60,000
940 – 1,074	1,880 – 2,147	60,001 – 70,000
1,075 – 1,459	2,148 – 2,918	70,001 – 100,000
1,460 – 2,102	2,919 – 4,203	100,001 – 150,000
2,103+	4,204+	150,001+

2018 Census questions

The questions in the 2018 Census were question 34, about sources of income, and question 35, about total income.

34 Tohua te katoa o raro iho nei e hāngai ana ki a koe. Nō hea ngā whiwhinga moni katoa i riro mai i a koe i ngā marama 12 kua pahure ake tae noa ki tēnei rā.

Kaua e kautehia ngā pūtea taurewa, nō te mea ehara ēnei i te whiwhinga moni.

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.

ngā utu ā-haora, utu ā-tau, utu huahoko, moni tāpiri me ētahi atu nā tōku kaituku mahi i utu	wages, salary, commissions, bonuses, etc, paid by my employer
nā aku mahi tākuhu, nā taku pakihi ake	self-employment, or business I own and work in
ngā hua moni, hua hea, moni rēti, ētahi atu haumi rānei	interest, dividends, rent, other investments
ngā utu auau nā te ACC, nā tētahi atu kamupene inihua hauata mahi rānei	regular payments from ACC or a private work accident insurer
te Penihana Kaumātua Kāwanatanga, te Penihana Hōia rānei	New Zealand Superannuation or Veteran's Pension
ētahi atu penihana motuhake (i tua atu i te Penihana Kaumātua, te Penihana Hōia, ngā penihana pakanga rānei)	other superannuation, pensions or annuities (other than NZ Superannuation, Veteran's Pension or war pensions)
te pūtea tautoko a Jobseeker	Jobseeker Support
te pūtea tautoko mō te Matua Kotahi	Sole Parent Support
he moni āwhina hei tautoko i ngā tāngata hauā, ngā kaiāwhina rānei	Supported Living Payment
te Tahua Ākonga	Student Allowance
ētahi atu momo takuhe kāwanatanga, moni āwhina a te kāwanatanga, penihana pakanga, te utu tiaki pēpi rānei	other government benefits, government income support payments, war pensions, or paid parental leave
ētahi atu momo whiwhinga tae rawa ake ki ngā pūtea tautoko mai i ngā tāngata kāore e noho ana ki tōku whare	other sources of income, counting support payments from people who do not live in my household
ko tēnei rānei	*or*
kāore kau he whiwhinga moni i taua wā	no source of income in that time

35 Mai i ngā momo whiwhinga moni katoa i tohua e koe i te pātai 34, e hia te nui o te katoa o ēnei moni:

- i riro mai i a koe
- i mua i te tango mai o ngā tāke me ētahi atu āhuatanga
- i ngā marama 12 tae atu ki te 31 o Poutūterangi 2018

From all the sources of income you marked in 34 what will the total income be:

- that you yourself got
- before tax or anything was taken out
- in the 12 months that will end on 31 March 2018

moni ngaro	loss
kāore he whiwhinga moni	zero income
$1 – $5,000	$1 – $5,000
$5,001 – $10,000	$5,001 – $10,000
$10,001 – $15,000	$10,001 – $15,000
$15,001 – $20,000	$15,001 – $20,000
$20,001 – $25,000	$20,001 – $25,000
$25,001 – $30,000	$25,001 – $30,000
$30,001 – $35,000	$30,001 – $35,000
$35,001 – $40,000	$35,001 – $40,000
$40,001 – $50,000	$40,001 – $50,000
$50,001 – $60,000	$50,001 – $60,000
$60,001 – $70,000	$60,001 – $70,000
$70,001 – $100,000	$70,001 – $100,000
$100,001 – $150,000	$100,001 – $150,000
$150,001 nui atu rānei	$150,001 or more

34 **Why do you want to know my income?**

35 Income statistics are used for developing social and economic policy, research, and monitoring programmes. The information you provide will be kept confidential.

- If you and your spouse/partner earn income jointly, only include your part of that income.
- If you received Working for Families payments (including Family tax credit, In-work tax credit, Minimum family tax credit and Parental tax credit), mark 'other government benefits…'.
- If you received homestay or child support payments, mark 'other sources of income…'.
- If you did piecework, mark 'wages, salary, commissions, bonuses, etc…'.

Count any payments that are taken out of your income before you get it, such as student loan repayments, union fees, fines, or child support payments.

DON'T count loans (including student loans), inheritances, sale of household or business assets, lottery wins, matrimonial/civil union/de facto property settlements, or one-off lump sum payments.

DON'T count money given by members of the same household to each other. For example, pocket money given to children, or money given for housekeeping expenses by a flatmate.

If you know your weekly or fortnightly income after tax, use the table provided to work out your annual income before tax.

35 Income table

Use this table as a guide to giving your before tax annual income in question 35.

After tax weekly income $	After tax fortnightly income $	Before tax annual income $
up to 86	up to 172	1 – 5,000
87 – 172	173 – 344	5,001 – 10,000
173 – 257	345 – 514	10,001 – 15,000
258 – 336	515 – 672	15,001 – 20,000
337 – 415	673 – 831	20,001 – 25,000
416 – 495	832 – 990	25,001 – 30,000
496 – 574	991 – 1,148	30,001 – 35,000
575 – 653	1,149 – 1,307	35,001 – 40,000
654 – 807	1,308 – 1,615	40,001 – 50,000
808 – 942	1,616 – 1,884	50,001 – 60,000
943 – 1,077	1,885 – 2,153	60,001 – 70,000
1,078 – 1,463	2,154 – 2,926	70,001 – 100,000
1,464 – 2,107	2,927 – 4,215	100,001 – 150,000
2,108+	4,216+	150,001+

Appendix B: Derivation details

Combining IR3 tables

The IR3 table within the IDI refresh, ir_clean.ird_rtns_keypoints_ir3, contains some historical gaps within the data. An issue with the data transference from Inland Revenue meant a large number of IR3 records were not included within the data sent to Stats NZ. This issue was rectified in 2019 for future data transfers. To deal with the existing gaps in the data, missing records were ad hoc loaded into the IDI, but no merging with the refresh data is undertaken. As such, in order for a full accounting of available IR3 data, these ad hoc load tables must be merged with the IR3 table from a refresh. The tables merged are taken from the IDI_Adhoc database and are the clean_read_IR.ir_ir3_2000_to_2014 and clean_read_IR.ir_ir3_2013_to_2020 tables.

Additional IR3 data was ad hoc loaded during the COVID-19 pandemic for supporting the COVID-19 analysis. These tables introduce a few additional records that are not present in either the refresh IR3 table or the historical IR3 tables, and so are also included in the combined IR3 table generated for this derivation.

With the joining of multiple sources of IR3 information (and even within the individual tables), there are individuals with multiple returns for a given year. For a valid derivation, each individual must have only a single IR3 return. As such, the joint tables must be de-duplicated. This proceeds in a number of steps. For individuals with more than one return:

1. The row with the maximum ir_ir3_return_version_nbr is selected.
2. If there are still duplicates, the row with the maximum ir_ir3_ird_timestamp_date is selected.
3. If there are still duplicates, the source table is prioritised. The priority order is: the refresh IR3 table, then the historical ad hoc load table with data from 2000 to 2014, then the historical ad hoc load table with data from 2013 to 2020 and finally the COVID-19 loaded tables, with the most recent loads being prioritised.
4. If there are still duplicates, the row with the maximum ir_ir3_snz_unique_nbr is selected.
5. If there are still duplicates, the final selection criterion is the maximum ir_ir3_location_nbr.

Aggregating IR tables

As part of the IDI refresh process, a derived income table for calendar and tax years is generated. This takes income data from the EMS, IR3, IR4S and IR7 (called IR20 in the IDI) refresh tables, and aggregates most of the income information available within these tables. As the refresh IR3 table is missing some data, it is not possible to use these tables directly. Rather, we perform the same process using the combined IR3 table. In addition, income from the following IR3 columns is added to this aggregated IR table: ir_ir3_gross_interest_amt, ir_ir3_gross_dividend_amt, ir_ir3_estate_trust_income_amt, ir_ir3_overseas_income_amt, and ir_ir3_other_income_amt. The following coded income information is collected from these aggregate IR tables:

- **W&S-EMS** Wages and salary income from the EMS table
- **WHP-EMS** Withholding payments from the EMS table

- **CLM-EMS** ACC payments from the EMS table
- **PEN-EMS** NZ superannuation payments from the EMS table
- **STU-EMS** Student allowance payments from the EMS table
- **BEN-EMS** Main social benefits from the EMS table
- **PPL-EMS** Paid parental leave from the EMS table
- **S01** Sole trader receiving PAYE-deducted income, from the EMS and IR3 tables
- **C01** Company director/shareholder receiving PAYE-deduced income, from the EMS and IR4 tables
- **P01** Partner receiving PAYE-deducted income, from the EMS and IR20 tables
- **C02** Company director/shareholder receiving WHT-deducted income, from the EMS and IR4 tables
- **P02** Partner receiving WHT-deducted income, from the EMS and IR20 tables
- **S02** Sole trader receiving WHT-deducted income, from the EMS and lack of record in the IR20 or IR4 tables
- **S00-IR3** Sole trader income from IR3 table
- **S03-IR3** Rental income from IR3 table
- **C00-IR4** Company director/shareholder income from IR4 table
- **P00-IR7** Partnership income from IR20 table
- **INT-IR3** Interest income from the IR3 table
- **DIV-IR3** Dividend income from the IR3 table
- **EST-IR3** Estate trust income from the IR3 table
- **SEA-IR3** Overseas income from the IR3 table
- **OTH-IR3** Other income from the IR3 table.

Combining IR tax tables

Additional income information is available from the ird_pts (PTS) and ird_autocalc_information (AA) tables as part of the refresh. The AA table is only available from 2019 onwards as it was only implemented by IR in 2019.

Cleaning PTS table

From the PTS table, two income columns are of interest: ir_pts_tot_interest_amt and ir_pts_tot_dividend_amt, coded as **PTS-INT** and **PTS-DIV** respectively. Like the IR3 table, the PTS

table contains duplicate records that must be de-duplicated. The process is similar to that undertaken for the IR3 table, but requiring only two steps.

1. The row with the maximum ir_pts_ird_timestamp_date is selected.
2. If there are still duplicates, the row with the maximum ir_pts_snz_unique_nbr is selected.

Cleaning AA table

The AA table is only available from 2019, but contains some useful income sources, namely income from interest and dividends, Māori authority distributions, and (from 2021) portfolio investment entity (PIE) income. These sources are coded as **INT-AA**, **DIV-AA**, **MAD-AA**, and **PIE-AA** respectively. Again, the AA table contains duplicate records for some individuals that must be removed. The process for this is:

1. The row with the maximum ir_ac_processing_date is selected.
2. If there are still duplicates, the row with the maximum ir_ac_return_version_nbr is selected.
3. If there are still duplicates, the row with the maximum ir_ac_snz_unique_nbr is selected.

Joining the tables

The individual IR tables are then combined into a single table. This combination needs to account for potential double counting of some income sources due to duplication between the income tables. Overlapping income indications are present for the interest and dividend income types, available from the IR3, PTS and AA tables. When joining the three tables (PTS, AA and the aggregated EMS/IR3/IR4/IR20 tables), all interest and dividend income in the aggregated table (code INT-IR3 and DIV-IR3) is taken. Interest and dividend income from the AA table is taken if there is no corresponding income from the aggregated table. Interest and dividend income from the PTS table is taken only if there is no corresponding income from either the aggregated table or the AA table. All other income codes are taken from their corresponding tables. Each income source code is aggregated so that for a given individual, year, and source code, there is only a single record.

Calculating benefit income from MSD

Calculating main benefits from MSD

The MSD first tier table (msd_clean.msd_first_tier_expenditure) contains MSD's record of taxable benefits paid out, which will also be covered by IR data. This is stored as (amongst other data) a start and end date, as well as a daily gross amount. The first step to calculate this benefit income is to filter for ranges that fall within the financial years of interest as well as for active benefits (msd_fte_srvst_code = '3'). For each spell that falls within the financial year of interest (spells which start or end outside of the financial year period are truncated to the financial year), the number of days it is active is multiplied by the daily gross amount (msd_fte_daily_gross_amt) to give the income for that spell. The type of benefit is stored in the msd_fte_serv_code column. These are recoded to either **JOB-MS1**, **SPS-MS1**, **SLP-MS1**, **NZS-MS1**, or **OTH-MS1** based on the metadata concordance information available within the IDI. Although individuals are only entitled to one main benefit at a time, they can have multiple main benefits over the course of a year, or multiple spells receiving a certain benefit. As such, each income source code is aggregated so that for a given individual, year, and source code, there is only a single record.

Calculating supplementary benefits from MSD

The MSD second tier table (msd_clean.msd_second_tier_expenditure) contains information on non-taxable supplementary benefits. Like the first tier data, it is stored as a start and end date as well as a daily gross amount. As well as selecting only those spells that intersect with the financial years of interest, we also filter for active benefits only (msd_ste_srvst_code = '3') as well as for payments to individuals (msd_ste_supp_source_text = 'NON OB/UCB SUP'). This excludes a few benefits, like orphan benefit, unsupported child benefit and child disability allowance, as they are usually given to service providers, not individuals. There is probably nothing smarter that can be done, but it is a significant potential source of bias. The annual amount for each individual is calculated as per the first tier data, with the income source being coded according to the main benefit it is associated with (msd_ste_parent_serv_code), unless the benefit is either family tax credit or best start tax credit, in which case it is coded as **FTC-MS2** or **BST-MS2** respectively. All other codes used are the same as tier one codes, with the change from **-MS1** suffix to the **-MS2** suffix. Again, each income source code is aggregated so that for a given individual, year, and source code, there is only a single record.

Calculating ad hoc benefits from MSD

The MSD third tier table (msd_clean.msd_third_tier_expenditure) contains information on non-taxable one-off and ad hoc payments from MSD. As they are not ongoing benefits, there is no need to worry about spells. The msd_tte_decision_date is used as a proxy for the date of payment and assigned to a financial year accordingly. Additionally, only non-recoverable benefits are selected (msd_tte_recoverable_ind = 'N'). All such ad hoc payments to an individual are aggregated and coded as **OTH-MS3**.

Combining MSD data with IR data

Both MSD and IR records contain benefits data. There are records in both that are not in the other. As it is unknown which of the two sources is more reliable, the MSD data will be overlapped with IR, selecting MSD data when there is overlap. This only applies to MSD first tier data due to the taxable nature of the benefits it relates to. Second and third tiers can be combined in without issue.

Calculating working for families (WFF) income

From the WFF table (wff_clean.fam_return_dtls) supplied by IR, we can calculate income from child support and tax credits. Child support income is coded as **CSP-WFF** and is given in the wff_frd_child_support_rec_amt column. Tax credits are stored as negative values, except for the wff_frd_winz_paid_amt column, and income is generally treated as a positive value. Additionally, there are multiple tax credits available. For the purposes of this derivation, each tax credit is coded individually, giving:

- **WFF-PTC** as the amount of parental tax credit (PTC) the individual is entitled to, taken as the negative of the wff_frd_ptc_entitlement_amt column
- **WFF-CTC** as the amount of child tax credit (CTC) the individual is entitled to, taken as the negative of the wff_frd_ctc_entitlement_amt column
- **WFF-FTC** as the amount of family tax credit (FTC) the individual is entitled to, taken as the negative of the wff_frd_ftc_entitlement_amt column
- **WFF-IWP** as the amount of in-work tax credit (IWTC) the individual is entitled to, taken as the negative of the wff_frd_iwp_entitlement_amt column

- **WFF-BST** as the amount of best start tax credit (BSTC) the individual is entitled to, taken as the negative of the wff_frd_fam_bstc_entitlement_amt column

- **WFF-FST** as the amount of family support tax credit (FSTC) the individual is entitled to less any amount paid by WINZ, taken as the negative of the wff_frd_fstc_entitlement_amt column, less wff_frd_winz_paid_amt. This is only included if the difference is positive.

There are some oddities with the WFF data. To account for these oddities, the following filtering steps are taken:

- When the day and month of wff_frd_return_period_date do not match 31 March, those entries are discarded.

- When snz_uid is the same as partner_snz_uid, those entries are discarded.

- When there are multiple entries for a person/partner/year combination, the most recent wff_frd_updated_date is selected, followed by the maximum wff_snz_unique_nbr.

This data is easily joined with the rest of the income data. However, if an individual has income from both FTC-MS2 and WFF-FTC sources, or BST-MS2 and WFF-BST, the MS2 data is removed from the combination. This is because WFF is jointly administered by both IRD and MSD and aims to avoid some double counting that could occur.

Adding zero-income/no-income source data

An important component of this derivation is the assignment of zero income/no source of income to some individuals where it is warranted. This assignment of zero income is broken into two parts. The first looks for individuals who have some interaction with IR or MSD over the course of the year, but no derived income. The second part requires a target population, utilising age and history to determine if the individual should be regarded as having zero income for a given year.

The initial step for part one of the process is to determine the list of snz_uids present in the tax data each year. If an individual is located in the tax data but does not have any income identified (for example they could have filed an IR3 but had no income declared on it), they are assigned to having zero income for that year.

Students receiving living costs as part of their student loan are another group where a similar process can be undertaken. As the student loan living costs are a loan, they are not regarded as income, but they do indicate interaction with MSD through StudyLink. As such, anyone with a drawn living cost (or allowance paid) amount in the period covering a given tax year, and no other income source derived for that period, is assigned to zero income.

Age-based considerations are the next step. Firstly, the majority of individuals under 18 years of age are undertaking full-time schooling and not working. Thus, if an individual 18 or younger has a current enrolment at a secondary school, as determined by the Ministry of Education student enrolment table at some point during the tax year, and does not have any derived income for that year, they are assigned zero income. Also, if an individual is aged 17 or younger on the last day of a tax year and does not have any derived income for that year, they are assigned zero income for that tax year.

As a final step, to account for individuals moving in and out of the labour force, if an individual has income derived for any year preceding a tax year where they have no income derived, and are part of the population of interest, they are also assigned zero income for that tax year.

Final aggregation

Finally, all detailed admin sources of an individual's income are coded to match the census concept of source of income. The correspondence used is given in table 7 The total personal income for an individual is calculated by summing all source totals.

Table 7. Detailed admin income sources and census income sources

Detailed admin income sources and census income sources					
Detailed admin source	Census source	Detailed admin source	Census source	Detailed admin source	Census source
W&S-EMS	01	NZS-MS1	05	WFF-FTC	11
C00-IR4	02	PEN-EMS	05	WFF-IWP	11
C01	02	JOB-MS1	07	WFF-PTC	11
C02	02	SPS-MS1	08	CSP-WFF	12
P00-IR7	02	SLP-MS1	09	EST-IR3	12
P01	02	STU-EMS	10	MAS-AA	12
P02	02	BEN-EMS	11	OTH-IR3	12
S00-IR3	02	BST-MS2	11	SEA-IR3	12
S01	02	FTC-MS2	11		
S02	02	JOB-MS2	11		
WHP-EMS	02	NZS-MS2	11		
DIV-AA	03	OTH-MS1	11		
DIV-IR3	03	OTH-MS2	11		
DIV-PTS	03	OTH-MS3	11		
INT-AA	03	PPL-EMS	11		
INT-IR3	03	SLP-MS2	11		
INT-PTS	03	SPS-MS2	11		
PIE-AA	03	WFF-BST	11		
SO3-IR3	03	WFF-CTC	11		
CLM-EMS	04	WFF-FST	11		

Appendix C: Data tables

The following tables are available as CSV files in the Download section of the web page.

Table 8. Counts of income band for census and admin data, 2013

Counts of income band for census and admin data, 2013																	
Census/Admin	**Loss**	**Zero**	**$1-$5k**	**$5k-$10k**	**$10k-$15k**	**$15k-$20k**	**$20k-$25k**	**$25k-$30k**	**$30k-$35k**	**$35k-$40k**	**$40k-$50k**	**$50k-$60k**	**$60k-$70k**	**$70k-$100k**	**$100k-$150k**	**$150k+**	**Total**
Loss	1,425	3,153	2,415	1,431	1,368	1,230	813	633	459	327	435	246	165	216	72	60	15,123
Zero	1,512	123,312	37,194	12,360	8,484	7,470	3,879	2,163	1,380	960	1,431	639	675	588	171	144	219,717
$1k-$5k	1,470	20,403	64,014	35,640	18,390	13,392	6,582	3,843	2,622	1,614	2,268	810	645	504	102	60	176,166
$5k-$10k	1,164	4,884	15,336	40,356	38,622	26,199	11,871	5,715	3,306	1,881	2,058	732	483	492	120	72	155,883
$10k-$15k	1,107	3,351	7,449	16,623	54,477	100,836	32,418	13,779	7,473	3,858	3,648	1,257	768	777	219	123	253,443
$15k-$20k	1,092	2,541	4,677	8,190	20,157	106,443	66,714	24,504	13,674	7,641	6,606	2,022	1,104	963	240	165	272,301
$20k-$25k	900	2,172	3,498	5,046	9,519	32,220	70,371	37,473	22,653	12,477	10,680	3,006	1,344	1,275	318	195	216,999
$25k-$30k	813	1,860	2,778	3,573	6,084	17,274	24,579	35,736	36,753	21,603	18,708	5,253	2,136	1,791	441	228	182,190
$30k-$35k	579	1,563	1,959	2,463	3,813	10,146	11,556	16,176	36,426	35,247	30,744	8,346	3,003	2,256	501	255	167,199
$35k-$40k	528	1,617	1,761	1,896	2,724	6,522	7,635	8,859	17,454	40,710	59,931	17,823	5,787	3,810	741	360	180,243
$40k-$50k	726	2,571	1,944	2,067	2,643	6,012	6,225	6,669	10,056	18,924	120,222	67,062	18,732	10,050	1,812	801	279,588
$50k-$60k	597	1,809	1,296	1,248	1,476	2,943	2,841	2,892	3,789	5,211	25,605	93,981	48,576	19,959	2,652	1,113	218,232
$60k-$70k	399	1,230	801	729	798	1,503	1,386	1,482	1,746	2,301	7,686	19,653	65,793	51,069	3,933	1,452	163,662
$70k-$100k	573	1,800	1,035	807	933	1,488	1,365	1,473	1,632	1,848	5,820	8,286	20,220	152,241	24,831	4,014	230,709
$100k-$150k	324	933	570	345	399	636	528	468	549	603	1,812	1,887	2,787	16,896	64,782	12,528	107,202
$150k+	405	735	636	318	339	528	480	351	375	438	1,143	1,011	1,611	4,500	9,147	43,560	66,702
Total	13,998	185,361	153,657	139,227	180,879	365,937	271,260	172,071	167,952	161,559	307,068	236,835	176,859	270,906	111,234	65,820	

Table 9. Counts of income band for census and admin data, 2018

Counts of income band for census and admin data, for 2018																	
Census/Admin	**Loss**	**Zero**	**$1-$5k**	**$5k-$10k**	**$10k-$15k**	**$15k-$20k**	**$20k-$25k**	**$25k-$30k**	**$30k-$35k**	**$35k-$40k**	**$40k-$50k**	**$50k-$60k**	**$60k-$70k**	**$70k-$100k**	**$100k-$150k**	**$150k+**	**Total**
Loss	1,383	4,275	2,442	1,428	1,364	1,611	1,128	861	591	513	705	399	294	381	144	108	18,459
Zero	1,338	126,717	35,043	11,721	8,703	8,748	5,823	2,793	1,635	1,206	1,944	912	975	951	318	264	228,879
$1k-$5k	1,203	24,552	78,072	31,470	15,786	12,261	7,677	4,233	3,024	2,433	3,786	1,869	1,248	1,173	261	153	193,731
$5k-$10k	930	7,914	13,689	51,864	34,476	22,170	11,220	5,718	3,312	2,277	2,901	1,254	966	927	252	174	163,110
$10k-$15k	792	5,340	6,819	13,077	65,979	82,716	31,050	13,977	6,831	4,074	4,656	1,635	1,020	1,143	360	186	244,101
$15k-$20k	876	4,209	4,707	7,278	15,558	153,036	71,436	34,791	19,143	12,240	11,394	3,648	2,112	2,526	786	525	350,658
$20k-$25k	681	3,408	3,381	4,230	8,001	32,739	106,128	55,518	24,924	15,639	14,931	4,722	2,346	2,373	753	429	285,351
$25k-$30k	558	2,892	2,886	3,147	5,076	18,555	24,978	47,748	30,975	22,230	22,377	6,996	2,832	2,445	543	327	197,568
$30k-$35k	426	2,385	1,878	2,268	3,330	11,016	12,579	13,665	44,655	29,865	33,060	10,449	3,807	2,877	669	345	175,575
$35k-$40k	432	2,622	1,860	1,929	2,736	7,413	8,826	9,429	15,462	53,802	58,707	21,705	7,872	5,157	933	411	201,573
$40k-$50k	651	4,422	2,427	2,430	3,081	6,918	7,800	8,175	11,262	19,461	157,077	75,855	25,746	14,691	2,292	957	346,653
$50k-$60k	552	3,651	1,770	1,491	1,818	3,657	3,822	3,720	4,722	6,342	29,142	133,638	62,244	30,618	3,855	1,353	295,212
$60k-$70k	360	2,466	1,176	924	1,038	1,962	1,836	1,839	2,145	2,622	8,997	21,822	98,178	69,438	6,054	1,698	224,679
$70k-$100k	600	3,789	1,593	1,017	1,161	2,088	1,827	1,839	1,983	2,373	7,140	10,149	25,047	238,098	39,918	5,667	347,340
$100k-$150k	408	2,151	957	444	507	864	756	630	636	735	2,220	2,442	3,681	22,761	109,017	19,935	169,701
$150k+	474	1,695	1,062	456	426	747	603	462	426	453	1,413	1,248	2,070	5,907	13,833	73,284	106,038
Total	11,664	202,488	159,762	135,174	169,041	366,501	297,489	205,398	171,726	176,265	360,450	298,743	240,438	401,466	179,988	105,816	

Table 10. Individual sources of income from census and admin data, by count, percent, and ratio, 2013 and 2018

Individual sources of income from census and admin data, by count, percent, and ratio, 2013 and 2018										
Income source	Admin				Census				Ratio	
	Count		Percent		Count		Percent			
	2013	2018	2013	2018	2013	2018	2013	2018	2013	2018
Wages and salary	2,030,847	2,326,389	59.2%	61.7%	1,809,531	2,283,054	57.7%	60.6%	1.02	1.02
Self-employment	560,520	594,516	16.3%	15.8%	483,486	557,667	15.4%	14.8%	1.06	1.07
Investments	694,878	708,876	20.2%	18.8%	655,062	633,951	20.1%	16.8%	0.97	1.12
ACC Payments	70,803	95,952	2.1%	2.5%	36,270	62,157	1.2%	1.6%	1.78	1.54
NZ superannuation	580,404	689,472	16.9%	18.3%	526,437	652,659	16.8%	17.3%	1.01	1.06
Other superannuation	0	0	0.0%	0.0%	83,904	90,756	2.7%	2.4%	0.00	0.00
Jobseeker	278,121	239,847	8.1%	6.4%	162,576	232,956	5.2%	6.2%	1.56	1.03
Sole Parent Support	116,484	81,108	3.4%	2.2%	86,136	60,102	2.7%	1.6%	1.23	1.35
Supported Living	108,666	108,411	3.2%	2.9%	74,499	66,795	2.4%	1.8%	1.33	1.62
Student Allowance	104,874	77,475	3.1%	2.1%	89,361	86,655	2.9%	2.3%	1.07	0.89
Other benefits	888,462	806,160	25.9%	21.4%	131,121	143,394	4.2%	3.8%	6.19	5.62
Other income	300,417	308,364	8.8%	8.2%	60,165	56,562	1.9%	1.5%	4.56	5.45
No source of income	235,554	238,839	6.9%	6.3%	233,628	238,551	7.5%	6.3%	0.92	1.00
Total people stated	3,432,933	3,770,886	100%	100%	3,133,722	3,769,398	100%	100%	1.00	1.00
Missing	93,891	97,239	—	—	242,697	6,954	—	—	—	—
Total people	3,526,824	3,863,625	—	—	3,376,419	3,776,355	—	—	—	—

Table 11. Proportion of income sources, by time since refresh, 2020 and 2021 tax years

Proportion of income sources, by time since refresh, 2020 and 2021 tax years[1]																
Reference date tax year	31 March, 2020									31 March, 2021						
Months after reference	-2 mths	4 mths	7 mths	13 mths	16 mths	19 mths	24 mths	27 mths	31 mths	-5 mths	1 mth	4 mths	7 mths	12 mths	15 mths	19 mths
Income source																
Wages and salaries	65.9%	64.2%	63.0%	62.3%	61.8%	61.7%	61.6%	61.8%	61.5%	57.3%	61.6%	62.0%	61.7%	61.1%	60.7%	60.5%
Self-employment	5.1%	6.5%	9.3%	12.8%	15.0%	15.8%	15.9%	16.1%	16.0%	4.1%	5.4%	6.2%	9.4%	12.6%	15.4%	15.7%
Investment	—	1.1%	42.2%	51.1%	56.8%	57.3%	57.5%	57.3%	57.7%	0.0%	0.0%	1.7%	62.9%	72.1%	79.3%	79.9%
ACC payments	2.0%	2.7%	2.8%	2.8%	2.8%	2.8%	2.8%	2.8%	2.8%	1.5%	2.5%	2.7%	2.8%	2.8%	2.8%	2.8%
NZ superannuation	20.7%	19.4%	19.1%	18.9%	18.8%	18.8%	18.8%	18.9%	18.8%	19.7%	20.2%	20.4%	20.1%	19.9%	19.4%	19.8%
Jobseeker Support	6.0%	7.3%	7.2%	7.1%	7.0%	7.0%	7.0%	7.1%	7.0%	6.4%	8.3%	8.9%	8.9%	8.8%	8.7%	8.7%
Sole Parent Support	2.0%	2.1%	2.0%	2.0%	2.0%	2.0%	2.0%	2.0%	2.0%	1.8%	2.0%	2.1%	2.1%	2.1%	2.1%	2.0%
Supported Living Payment	3.0%	2.9%	2.8%	2.8%	2.8%	2.8%	2.8%	2.8%	2.8%	2.7%	2.8%	2.8%	2.8%	2.8%	2.7%	2.8%
Student allowance	1.7%	1.9%	1.9%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.4%	1.5%	1.8%	1.9%	1.9%	1.9%	1.8%
Other benefits	15.3%	16.3%	31.4%	31.8%	31.8%	31.9%	31.9%	32.2%	31.9%	29.4%	31.9%	32.4%	34.5%	34.8%	34.5%	34.7%
Other income	—	0.4%	3.3%	6.2%	7.6%	7.7%	7.8%	7.9%	7.9%	0.0%	0.0%	0.2%	3.3%	5.6%	7.2%	7.4%
No income	4.3%	9.0%	6.0%	4.4%	3.5%	3.4%	3.5%	3.6%	3.4%	13.0%	10.0%	9.3%	4.8%	3.7%	2.9%	2.4%

Note: 1. Time lags relate to refreshes in January 2020, July 2020, October 2020, April 2021, July 2021, October 2021, March 2022, June 2022, and October 2022

Table 12. Proportion of income bands, by time since refresh, 2020 and 2021 tax years

Proportion of income bands, by time since refresh, 2020 and 2021 tax years[1]																
Reference date tax year	31 March, 2020									31 March, 2021						
Months after reference	-2 mths	4 mths	7 mths	13 mths	16 mths	19 mths	24 mths	27 mths	31 mths	-5 mths	1 mth	4 mths	7 mths	12 mths	15 mths	19 mths
Income band																
Loss	0.2%	0.2%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%	—	—	0.0%	0.3%	0.4%	0.4%	0.5%
Zero income	4.3%	9.0%	6.0%	4.5%	3.5%	3.4%	3.5%	3.6%	3.4%	13.0%	10.0%	9.3%	4.8%	3.7%	2.9%	2.4%
$1 – $5,000	26.7%	8.4%	10.5%	10.1%	10.1%	10.0%	10.0%	9.2%	10.0%	34.5%	8.3%	7.5%	9.9%	9.7%	9.3%	9.6%
$5,001 – $10,000	14.9%	4.4%	4.3%	4.2%	4.2%	4.2%	4.2%	4.2%	4.2%	12.3%	9.5%	3.8%	3.6%	3.6%	3.4%	3.5%
$10,001 – $15,000	8.4%	13.7%	10.5%	9.2%	8.5%	8.4%	8.4%	8.4%	8.4%	12.7%	12.3%	9.0%	7.6%	7.0%	6.5%	6.5%
$15,001 – $20,000	9.8%	10.3%	7.1%	7.2%	7.3%	7.3%	7.2%	7.3%	7.2%	9.4%	8.8%	11.1%	10.3%	9.8%	9.3%	9.3%
$20,001 – $25,000	8.9%	4.0%	7.8%	7.7%	7.7%	7.7%	7.7%	7.7%	7.7%	6.3%	4.5%	8.6%	8.0%	7.9%	7.9%	7.8%
$25,001 – $30,000	6.9%	3.8%	4.6%	4.9%	5.0%	5.0%	5.0%	5.0%	5.0%	4.0%	4.6%	4.5%	4.8%	5.0%	5.2%	5.2%
$30,001 – $35,000	5.2%	4.1%	4.4%	4.6%	4.7%	4.7%	4.7%	4.8%	4.7%	2.4%	5.1%	4.7%	4.2%	4.4%	4.5%	4.5%
$35,001 – $40,000	6.7%	8.9%	9.3%	9.9%	10.1%	10.1%	10.1%	10.2%	10.1%	2.5%	10.0%	9.9%	9.2%	9.5%	9.8%	9.8%
$40,001 – $50,000	3.1%	8.1%	8.5%	8.9%	9.0%	9.0%	9.0%	9.1%	9.0%	1.1%	7.8%	8.4%	8.8%	9.0%	9.2%	9.1%
$50,001 – $60,000	1.7%	6.4%	6.8%	7.3%	7.5%	7.5%	7.5%	7.6%	7.5%	0.6%	5.5%	6.3%	7.1%	7.4%	7.6%	7.6%
$60,001 – $70,000	2.0%	11.0%	11.7%	12.4%	12.8%	12.8%	12.8%	13.0%	12.8%	0.7%	8.5%	10.2%	12.3%	12.8%	13.4%	13.3%
$70,001 – $100,000	0.8%	5.1%	5.4%	5.8%	6.0%	6.1%	6.1%	6.1%	6.1%	0.3%	3.5%	4.5%	6.0%	6.3%	6.7%	6.6%
$100,001 – $150,000	0.2%	1.4%	1.5%	1.7%	1.8%	1.8%	1.8%	1.8%	1.8%	0.1%	0.9%	1.2%	1.7%	1.8%	2.0%	2.0%
$150,001+	0.2%	1.1%	1.2%	1.5%	1.7%	1.7%	1.7%	1.7%	1.7%	0.0%	0.7%	0.9%	1.4%	1.7%	2.0%	2.1%

Note: 1. Time lags relate to refreshes in January 2020, July 2020, October 2020, April 2021, July 2021, October 2021, March 2022, June 2022, and October 2022

Appendix D: Sub-population plots

Figures 10 and 11 show income coverage from the linked census-admin dataset.

Figure 10. Coverage of total personal income data by Māori descent status, for people aged 15 years and over, data sourced from IDI, 2006–2021

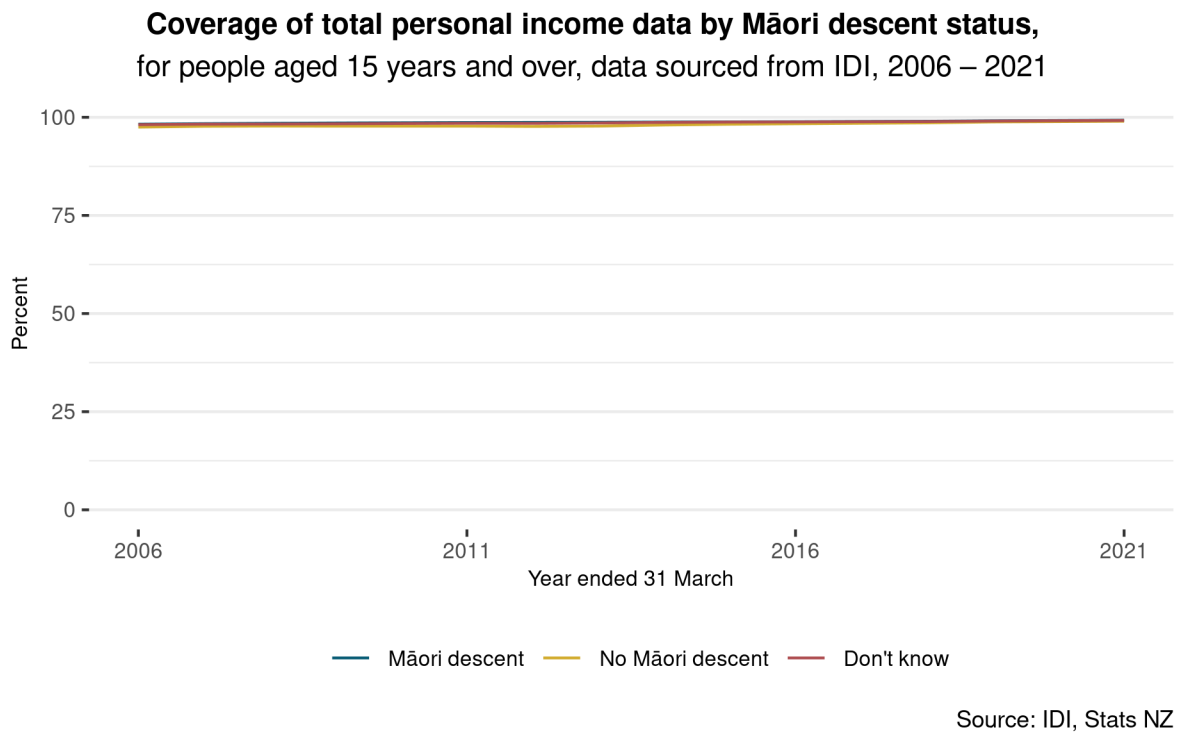
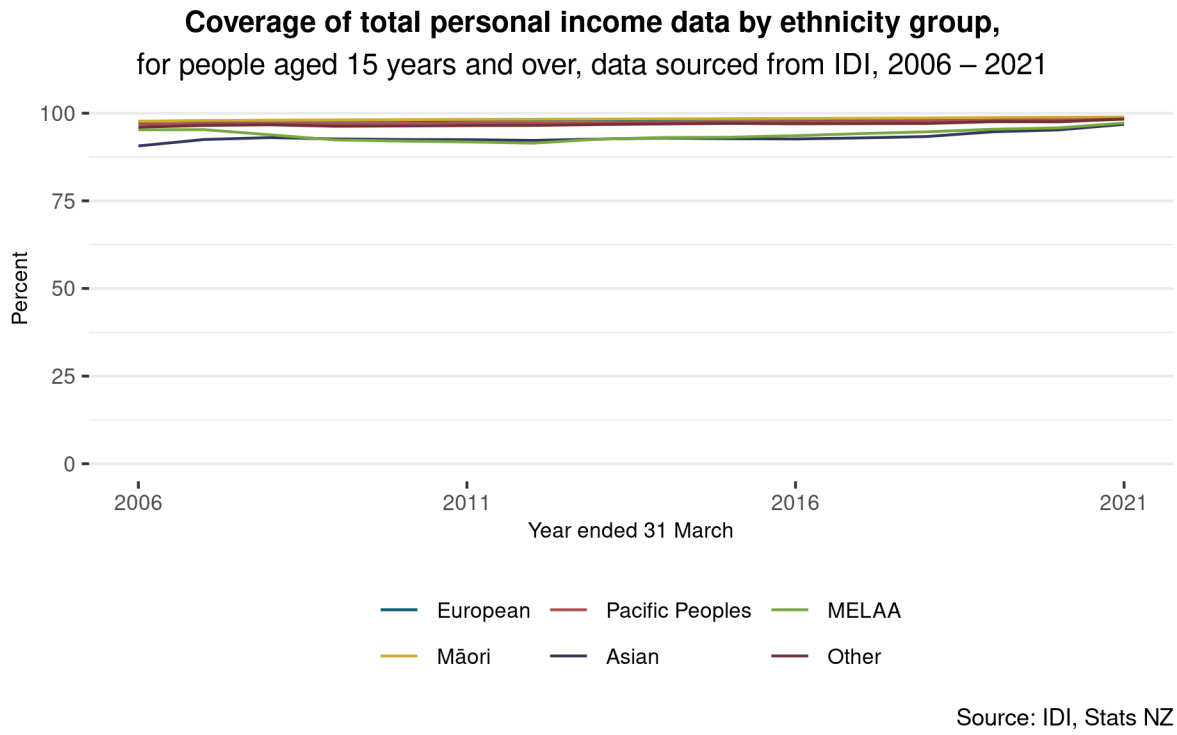


Figure 11. Coverage of total personal income data by ethnicity group, for people aged 15 years and over, data sourced from IDI, 2006–2021



Source: IDI, Stats NZ