Stats NZ
Tatauranga Aotearoa

The potential for linked administrative data to provide household and family information

Megan Gath and Christine Bycroft

New Zealand Government

Disclaimer

The results in this paper are not official statistics. They were created for research purposes from the Integrated Data Infrastructure (IDI) managed by Stats NZ.

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

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI.

Privacy impact assessment for the Integrated Data Infrastructure (available from www.stats.govt.nz) has more information.

Citation
Gath, M and Bycroft, C (2018). *The potential for linked administrative data to provide household and family information.* Retrieved from www.stats.govt.nz.

ISBN 978-1-98-852852-6 (online)

Published in **April** 2018 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

Megan.gath@stats.govt.nz or christine.bycroft@stats.govt.nz

www.stats.govt.nz

Contents

List of tables and figures

List of tables

List of figures

1 Background

Census transformation in New Zealand

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

- a short-to-medium-term focus on modernising the current census model and making it more efficient
- a longer-term focus on investigating alternative ways of producing small-area population and social and economic statistics. This includes the possibility of changing the census frequency to every 10 years, and exploring the feasibility of a census based on administrative data (Stats NZ, 2014a).

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 sources to meet those requirements.

Census transformation – a promising future (a 2015 Stats NZ Cabinet paper) recommended that Stats NZ work actively towards a future census based primarily on Government's administrative data, supported by redevelopment of its household surveys.

Census Transformation in New Zealand has more information.

About this paper

This paper is one of a series of investigations aimed at identifying and exploring the potential for admin sources to provide census-type information.

One important reason for having a census is to provide information about the characteristics of a population, including information about households and families.

This paper's investigation explores the potential for admin data to provide census-type information on households and families by comparing results from the 2013 Census with estimates produced from admin sources. These comparisons give us better understanding of the quality of the admin data and contribute to Stats NZ's work to transform the census model.

Key findings

These are the main findings of this investigation.

- Although there is a small undercount of households based on admin data, the potential exists to remedy this undercount through improved address coding.
- Despite good aggregate statistics for admin households, just under half of households had perfect agreement of household members when compared with census, suggesting the need for improved address information – particularly for younger adults.
- There is not currently sufficient admin data to provide high-quality information on families. Although we combined information from a variety of admin sources to create family nuclei, this methodology resulted in only 60 percent of the census family count.

- When family information was available from admin sources, we found it matched quite well to census family information.
- Future work in this area will require effort to improve the available admin data sources and improvements to some aspects of data coding.

2 Introduction

This paper describes our analysis of households and families information that is available from admin sources. It builds on findings from previous census transformation work in the context of future censuses.

The ability to produce household and family information from admin sources is a key consideration when determining if a census based on admin data is feasible. Census household and family information includes counts of households, household size distribution, and information about the relationships between individuals within households (used to determine family type and household composition). This information is available for small geographic areas and can be cross-classified by information about the individuals in the household and by information about the dwelling.

Family and household information is used to develop and evaluate government policies, such as income support, social housing, housing affordability, family violence, child poverty, household crowding, household income, household expenditure, and household net worth. It is critical that high-quality family and household-level statistics can be produced in future alternative census models.

O'Byrne and colleagues (2014) reported the family type, extended family, and household composition census variables were all unlikely to be satisfied by admin data sources. Previous census transformation work undertook initial explorations of household and family information in admin data. Gibb and Das (2015) constructed households using address information and reported on household size and household membership compared with the 2013 Census. Suei (2016) investigated admin data for legally registered relationships.

We extend this work using the most up-to-date census transformation methodology for address identification, and use a range of sources for family relationships now available in the Integrated Data Infrastructure (IDI).

Aims and scope

Our overall aim was to investigate the extent to which census household and family information can be obtained from existing admin data, building on previous census transformation work in this area.

Four main research questions guided this work.

- Can linked admin data sources be used to construct households? If so:
  - how closely do these households match census households?
  - what types of households can we capture well with admin data?
- Can linked admin data sources be used to construct family nuclei? If so:
  - how closely do these families match census families?
  - what types of families can we capture well with admin data?
- Does the information on households and families currently available in admin data meet census information requirements? If not:
  - what is needed to improve admin data on households and families?

- What are the implications of these findings for using admin data for households and families in future censuses?

This paper provides reference information about the relevant admin data sources, summarises the main concepts from the statistical standards for households and families, and explains how these are implemented in the census. We present findings from analyses comparing information in the 2013 Census with that constructed from the admin sources.

The admin sources we investigated were limited to those available in Stats NZ's IDI at June 2017. These sources include New Zealand registrations of births, marriages, and civil unions from the Department of Internal Affairs, benefits information from the Ministry of Social Development, tax credit information from Working for Families, visa information from the Ministry of Business, Innovation and Employment, and summary tables compiled from various admin sources. We constructed households and families from this data.

Note: estimates of households and families obtained from admin sources and provided in this paper are for comparing with census information only. They do not represent official statistics. For our investigation, 2013 Census data is used as the benchmark against which admin data is compared.

3 Method

This section describes the method we used to evaluate the potential for producing household and family information from admin data. We made comparisons between information from the census and admin sources at three levels:

- between concepts and definitions
- between aggregate counts and estimates
- between individual level records.

In the next section, we summarise the formal statistical concepts and classifications used in official statistics. These provide the concepts and definitions against which both the census and the admin sources are compared. Differing definitions can help explain some of the error when comparing data between sources; we outline these areas of difference between the census and admin data.

The data sources section outlines the relevant information collected by the census and how families and households are constructed in the census. The IDI is described, with a focus on information relating to households and families. We also outline the methodology developed to construct households and families using admin data available in the IDI.

In the results section, we compare census results for household and family information with their admin equivalents, where possible. This includes household and family counts and household size distribution. These aggregate-level comparisons provide an insight into admin coverage, or how well admin data is able to capture the entire census population.

We also provide a comparison at the individual level of census and admin information. Where individuals could be linked between the census and admin data, we compared their information in both sources. This includes an analysis of household membership and family relationships. Individual-level comparisons provide insight for potential measurement error in admin data, which may result from: differing statistical concepts, errors in collection and processing systems, or from linkage errors.

Agreement between sources can be affected by the methodology used to link individuals across data sources and integrate information. 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 an admin source provides (no information is available if links are not made when they should be), or they may introduce measurement error if the wrong people are linked.

4 Statistical concepts for households and families

The statistical concepts, definitions, and classifications for households and families are summarised below.

Appendix 1 has more information.

Definitions for family and household

Stats NZ's statistical standard defines family and household as follows.

- A family (or family nucleus) is defined as a couple, with or without child(ren), or one parent and their child(ren), all of whom usually reside together in the same household. The children do not have partners or children of their own living in the same household.
- A household is one or more people usually resident in the same (private) dwelling, who share living facilities. A household can contain one or more families, or no families at all. A household that does not contain a family nucleus could contain unrelated people, related people, or could simply be a person living alone.

The above definitions, used in census, require information about relationships among people who usually live together. There are currently no admin data sources that collect information according to these definitions (ie on families or on households). However, a number of sources provide relationship information independent of where people live, and information on where people live independent of their relationships. These two types of information can be combined to approximate census definitions.

Classifications

The relevant family and household variables produced by census are:

- census household count
- number of usual residents in household, including further breakdown by age
- family type
- household composition.

These variables are produced based on the formal classifications outlined below.

Census household count

The census household count is a count of all households in New Zealand where at least one person is present at the usual address of the household on a given census night. Households whose members are all away temporarily elsewhere in New Zealand and/or temporarily overseas on census night are excluded, unless there is someone at their usual residence (eg a visitor) to identify them.

Number of usual residents in household

The size of a household is measured by the number of usual residents in that household. The census produces a further breakdown to 'Number of usual residents aged 15 and over in household' and 'Number of usual residents age under 15 in household'.

Family type

The family type classification specifies family membership based on the type of relationship between individuals and shared residence. It is a hierarchical classification with three levels. Level 1 classifies families according to the presence or absence of couples, parents, and children.

A family nucleus can be one of three types, and requires all family members to usually reside together in the same household:

- a couple without children
- a couple with a child (or children)
- one parent and their child (or children).

More detailed classification of families is provided at levels 2 and 3 of the classification, and only level 1 is considered for this investigation.

Appendix 2 has the full classification for family type.

Household composition

Household composition classifies households according to the relationships between usually resident people. Households are classified according to the presence, number, and type of family nuclei, and the presence of related and unrelated people.

Household composition is classified into three levels. Level 1 specifies the number of family nuclei present or, if no family nuclei are present, the number of people present (either a one-person household or a multi-person household). More detailed classification of households is provided at levels 2 and 3 of the classification, and only level 1 is considered for this investigation.

Appendix 3 has the full classification for household composition.

5 Data sources

This section describes the data sources used in this investigation: the New Zealand Census of Population and Dwellings, and Stats NZ's Integrated Data Infrastructure (IDI).

New Zealand Census of Population and Dwellings

The census produces the official count of people and dwellings in New Zealand. It provides a snapshot of our society at a point in time and tells the story of social and economic change in New Zealand. The most-recent censuses were held in March 2013 and March 2018.

New Zealand's census is a de facto census; that is, the census counts people where they are on census night. People who are away from home on census night are counted at their census night address (not their usual residence). The census aims to count everyone who is in the country on census night. Overseas visitors are included in the census, while New Zealand residents who are not in the country on census night are not included.

For this investigation, we are only interested in New Zealand residents ['a person who considers themselves to usually reside in New Zealand'], not those visiting New Zealand temporarily on census night.

Households and families are constructed in the census using information about: the address of the dwelling, the usual residence of individuals, any reported absentees from a dwelling, the relationships of individuals at a dwelling to the reference person (the person who completed the dwelling form), and the living arrangements of individuals (their relationships to the people they usually live with). Unlike most other census variables, absentees are included when constructing households and families – to ensure that households include all members, not just those who were at home on census night.

Appendix 1 has a detailed description of the construction of households and families in the census.

Quality of census data on households and families

In the 2013 Census, the overall quality assessment of household and family variables was high, indicating only minor data quality issues. Substitute households are created when no members of the private dwelling complete a census form. In the 2013 Census, 3.3 percent of households were full-substitute households.

Additionally, when some individuals within a household did not complete a form, substitute records were created for them. In the 2013 Census, 2.1 percent of households contained at least one substitute member (but were not full-substitute households).

The family type variable, as a derived variable, does not have a non-response/not classifiable category because the process of determining whether a group of people constitute a family also involves determining what type of family they form. In 2013, 2.6 percent of households were classified as 'Household composition unidentifiable'.

Integrated Data Infrastructure

The IDI is a linked database that allows policy evaluation and research as well as the production of statistical outputs on the transitions and outcomes of people. The IDI contains many admin and survey datasets, linked at the individual level, and can therefore act as a test environment for examining the potential of linked admin data sources to produce census families and household information.

This investigation is limited to data sources in the IDI at June 2017.

The IDI consists of a central 'spine' to which a series of data collections are linked. The spine is intended to include all individuals who have ever been residents of New Zealand, and aims to include each individual only once.

Three datasets are linked together probabilistically to create the spine:

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

Datasets from different source agencies are linked to the spine using deterministic and probabilistic linking.

Read more about the linking methodology used in the IDI here.

The IDI also contains several summary tables prepared by Stats NZ that provide core information about individuals (age, sex, ethnicity, and geographic information) from across the available data sources.

Following a description of the data sources used to construct households and families, we outline the admin populations used in these analyses.

Admin household information– Address summary table

The IDI includes address summary tables that store address information for individuals from various admin sources. Raw address strings provided by each agency are linked to NZ Post's National Postal Address Database. Where a successful link is made, a unique encrypted address identifier is made available to researchers within the IDI.

Currently, seven address sources are combined in the summary tables.

- 2013 Census – address of usual residence reported on individual census forms
- Inland Revenue – address supplied to Inland Revenue
- National Health Index – address of residence recorded when visiting a hospital or outpatient clinic
- Primary Health Organisation – address of residence recorded when visiting a general practitioner
- Ministry of Social Development – address of residence reported when applying for a benefit, as well as postal address

- Ministry of Education – address of residence reported when enrolling at primary or secondary school (but not tertiary education)
- Accident Compensation Corporation (ACC) – address of residence reported when filing an ACC claim.

All addresses recorded for individuals from these sources are included in the IDI tables.

Admin family information– Department of Internal Affairs

The Department of Internal Affairs (DIA) provides life-event data for integration to the IDI – this includes all births, marriages, and civil unions registered in New Zealand.

Births

Birth registration data is available from 1848 and is provided to Stats NZ each quarter. This dataset has a target population of everybody born in New Zealand and children who were born overseas but adopted in New Zealand. Birth registration data gives access to the biological or adoptive relationships between mothers, fathers, and children.

Information contained in birth registrations is provided by parents following a child's birth. By law, parents of a child born in New Zealand must notify DIA as soon as is reasonably practicable after the birth (deemed by the Registrar-General as generally being within two months of the birth).

In 1998 DIA moved to digital storage of its paper records. Birth registrations occurring after this digital capture include all information from the registration in the digital record, but records before 1998 may be lacking in completeness.

Note: The September 2017 refresh of the IDI included an additional 10 years of back digitisation, resulting in digital capture of birth and marriage registrations from 1988 onwards.

Date of birth for parents on birth certificates has only been captured from 1972. Between 1972 and 1998 there is partial digital capture of parents' birth dates. When DIA have a specific reason to look up a birth certificate (eg someone requests a name change) then the complete data record is digitised at that time. This makes it harder for parents on earlier records to be linked to the IDI spine, as birth date is a linking variable. As a result, some children cannot be linked to their parents because there is not enough information to accurately identify the parents.

Marriages and civil unions

DIA data on registered marriages is available from 1854, but as records have only been digitised since 1998, most data able to be linked is from 1998 onwards. Suei (2016) found some marriages between 1970 and 1997 (approximately 25 percent) were linked to the spine, but there was a large increase in linkages from 1998 (approximately 85 percent of marriages linked to spine).

Civil union data is available from April 2005, the same month civil unions became legal in New Zealand. In August 2013, same-sex marriage was legalised in New Zealand, which means the marriages data from then onwards includes both same-sex and opposite-sex couples (the previous requirement was 'husband and wife'). From 2013, couples in a civil union are able to transfer their civil union to a marriage without first dissolving the civil union, and vice versa.

Information on marriage and civil union registrations includes details about the two people entering into a legally recognised relationship and details about each of their parent(s). Details about the dissolution of marriages and civil unions are also available if dissolved in New Zealand. Thus these datasets provide information on legally recognised relationships (with the exception of de facto relationships). Suei's investigation of legal relationship information in admin data found high agreement with census legal relationship status for people who were married, but less consistency for people in other relationship categories, particularly those separated or widowed (Suei, 2016).

Marriages and civil unions of New Zealand residents occurring overseas are not captured, apart from a small number of marriages in other countries of New Zealand citizens where the ceremony was witnessed by an authorised official from a New Zealand diplomatic or consular post that have been registered with DIA.

Admin family information – Ministry of Social Development

The benefits dynamics data provided by the Ministry of Social Development (MSD) includes information on all people who have received a working-age social welfare benefit since 1 January 1993. Demographic information and changes in benefit status and other circumstances are in this data. Information on the benefit histories of partners and dependent children is also included where applicable.

Time spell information is provided for the relationships between adult partners, and between parents and dependent children (that MSD knows about). Note that dependent children are not necessarily biological children. Thus the MSD dataset has information on the living arrangements of people at a given point in time and indicates the social, but not necessarily legal, relationships among these people (eg between partners, and between children and their parental figures).

As information is only provided for working-age benefits, no partnership information is available from MSD for recipients of NZ Superannuation.

Admin family information – Working for Families tables

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).

This data is available from 1999 and is provided to Stats NZ ongoingly.

The datasets include details about the partners and children of recipients. Time spell information about the dates of these relationships, as known to WFF, are provided. Inland Revenue notes that information on the periods children are with their parents/carers is not always 100 percent correct as it relies on people informing both agencies of when their circumstances change, which doesn't always happen.

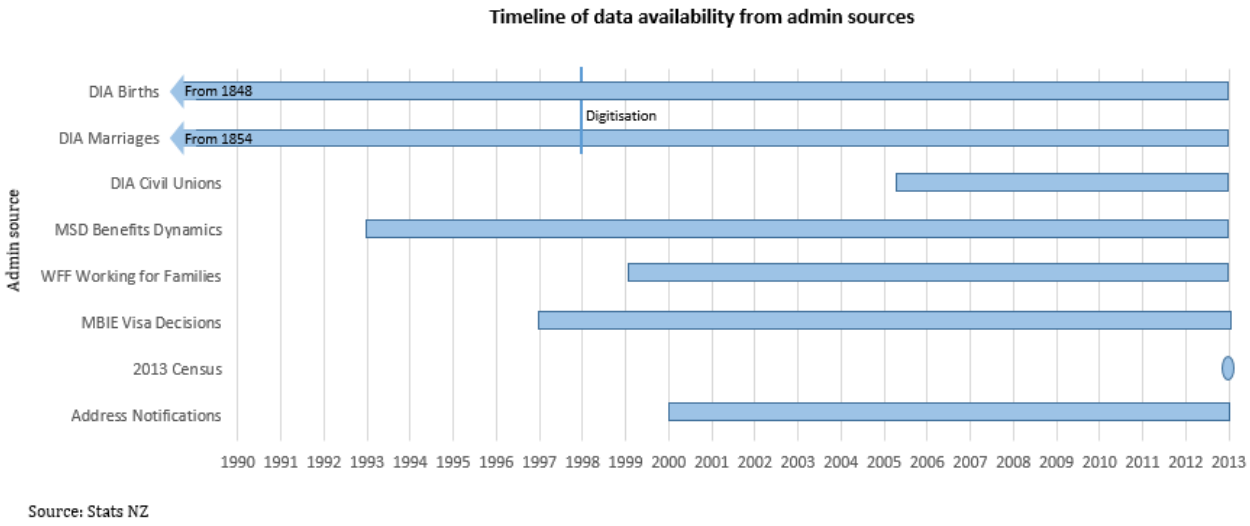
Admin family information – Ministry of Business, Innovation & Employment

The Ministry of Business, Innovation & Employment (MBIE) provides immigration admin data on migrants and international visitors who apply for a visa to enter New Zealand. This includes all resident visa applications, and applications for temporary stay (work, study, or visitor). Data is from 1997 through quarterly updates.

Information from MBIE on visa applications includes the policy under which the application was made and whether an application was approved or declined. Records are provided for all individuals listed on an application, which can be linked through the visa application number.

Figure 1 displays the timeline of admin data availability, by source.

Figure 1



Source: Stats NZ

Additional sources of family information in the IDI

As well as the sources of family information used in this investigation, the IDI has other sources that include survey data from Stats NZ (Household Economic Survey, Household Labour Force Survey, and the Survey of Families, Income, and Employment). These datasets cover small samples of the New Zealand population. We did not include them in the present investigation as they are not strictly admin data sources. However, future work should explore the potential of survey data to supplement admin data.

Housing New Zealand also includes information about family members in data that details applications for social housing. However, because there are known quality issues with this data (eg problems linking legacy identifiers to current identifiers) it was not included in the present investigation.

Finally, a summary relationship table that compiles all parent-child links in the IDI (ie from admin sources, from census, and from survey data) is available in the IDI Sandpit. Unlike the restrictions placed on parent-child links in the present investigation (details in the next section) the relationship table allows all links, which will include duplicated relationships where parents cannot be linked to their spine record (eg parents on birth registrations). As this table does not currently include a time indicator for relationships, we could not identify known relationships (at

census night); for this reason the information in this table was not used in the present investigation.

Admin populations used in analysis

The aggregate-level analyses in our investigation use only individuals who were identified through admin sources as usually resident in New Zealand at 5 March 2013 (census day). These individuals form an experimental population – the IDI-ERP – which aims to reflect as closely as possible those who would also be included in defining the official estimated resident population (ERP). The IDI-ERP is derived from the IDI spine and uses information from tax, health, ACC, education, migration, and deaths data to determine the individuals in New Zealand at a particular reference date (Gibb, Bycroft, & Matheson-Dunning, 2016).

Using the methodology for version 2 of the IDI-ERP (Gibb et al, 2016), the IDI-ERP at 5 March 2013 has 4,479,876 people. The 2013 ERP for the same date is 4,427,948 (based on the 4,242,048 census usual resident count, an estimated net census undercount of 104,200, and an estimated 81,700 residents temporarily overseas on census night; Stats NZ, 2014b).

Thus, the IDI-ERP contains 1.2 percent more people than the ERP. See Gibb et al (2016) for a discussion of the sources of over- and under-coverage in the IDI-ERP.

The IDI-Census linked population

The individual-level analyses in this investigation use data from the 2013 Census that is linked probabilistically to the IDI spine, using name, date of birth, sex, usual residence, and country of birth. The linked Census–IDI-ERP population includes all census records for usual residents for whom a suitable link in the IDI-ERP was found (this was 3,804,234 people, or 90 percent of the 4,242,048 census usual resident count).

It is possible to make an erroneous link between two different individuals (ie a census record and an IDI record are linked via the linking methodology but actually represent two different people rather than the same person). We estimate this occurred for less than 1 percent of the links made between census and IDI. However, it means that linkage error could explain a small proportion of cases where households and family information differs between the census and the IDI-ERP.

In addition to the linked individual population, the set of all households that could be linked between the census and the IDI was used to compare household membership. Households were linked using address information in the IDI. Although addresses in the census and the IDI are both linked to NZ Post's National Postal Address Database, different linking mechanisms are used in the census and the IDI, which could result in some discrepancies.

To create a linked set of households, we excluded substitute census households and households with absentees. This was to ensure a subset of census households in which all members had returned a census form. This resulted in 1,339,653 census households that could be linked to an address register identifier (86.4 percent of all census households). Of these households, 1,242,264 could then be linked to an IDI household (80.2 percent of all census households).

Constructing admin households and families

As households and families are not directly measured in admin sources, we used data from the admin sources detailed above to construct households and families.

Households

Following Gibb and Das (2015), we defined a household as a group of individuals who share the same unique address at a particular point in time. For this investigation, we used the address for census night (5 March 2013), as determined by admin sources, to group people into households.

As it was possible for people to have multiple addresses across admin sources, we implemented a prioritisation methodology to arrive at a single address for each person in the IDI-ERP (Stats NZ, 2017). This method takes the most-recently updated address in any of five sources determined to be of high quality: Inland Revenue, National Health Index, Primary Health Organisations, Ministry of Education, and MSD residential. Two remaining sources, Accident Compensation Corporation and MSD postal, are used only if no address exists in the other five. See Stats NZ (2017) for further information and a detailed analysis of address information from admin sources.

We needed to remove non-private dwellings from the list of addresses used as proxies for households. As there is currently no known information source on types of dwellings in the IDI, we used data from the 2013 Census to identify the addresses associated with non-private dwellings in the census. These were excluded as households.

Families

Families were constructed in two stages. In the first stage, we collected and combined all relevant family information contained in admin data for each person in the IDI-ERP. Known partnerships and parent-child relationships were identified, regardless of where people were living. In the second stage, we used family information in combination with address information to identify family nuclei within households.

Stage 1

We gathered relationship information linking parents and children, and linking partners, from the DIA, MSD, WFF, and MBIE sources described above. Only information known at or before census night (5 March 2013) was used in the investigation. Further, we only included relationship links where both members (parent and child, or partner and partner) were found in the IDI-ERP.

From DIA birth registrations, we obtained data linking children and up to two parents. A partnership between the two parents (if two were listed) was inferred from birth registrations.

DIA marriage and civil union registrations provided data linking the partners entering a marriage or civil union. Although the parents of each partner are included on these registrations, no parents could be linked to the IDI spine (and therefore the IDI-ERP), so no parent-child information could be obtained from legal partnership registrations. All marriages and civil unions dissolved before 5 March 2013 (as indicated in the DIA dataset) were excluded, and if multiple marriages or civil unions were known for a person, we retained only the most-recent partnership.

From MSD and WFF data, links were obtained between parents and children, and between partners. Approximately 1 percent and 3 percent of children had more than two parents recorded in the WFF and MSD data, respectively. As it is possible for someone to have more than two parents (eg, step-parents or adoptive parents and biological parents) we retained all known parents. For people with multiple partners known to WFF or MSD, we selected only the most-recently indicated partnership.

We used MBIE visa application data to infer a relationship between adults and children who were jointly listed on an application. Parent-child relationships were inferred when there was at least a 14-year age gap between the oldest child and the adult on the application, and no more than two adults were included in the application. We used publically available code provided by NZ Treasury to implement these criteria.

Treasury analytics and insights has the code.

For the present investigation, we also inferred a partnership between parents on these applications (ie where a parent-child link was inferred).

Information from all four sources was merged using unique person identifiers provided in the IDI. Where partnership information was available for a person in at least one source, we selected a partner by retaining only the most-recently indicated partner from any source (at 5 March 2013). We retained all unique parents and children known for a person across all sources.

This information was used to form a relationship table that contained, where known, information on partners, parents, and children for each member of the IDI-ERP at 5 March 2013. Table 1 displays the format of this relationship table (with up to two parents and four children displayed).

Table 1

Structure of the relationship table

Person ID	Address	Partner	Parent 1	Parent 2	Child 1	Child 2	Child 3	Child 4
1	1	2	74	81	-	-	-	-
2	1	1	-	-	39	64	-	-
3	2	4	-	-	5	-	-	
4	2	3	88	-	5	-	-	-
5	2	-	3	4	-	-	-	-
6	3	-	101	-	10	-	-	-
7	3	8	-	-	90	-	-	-
8	3	7	44	47	90	-	-	-
9	3	-	-	-	-	-	-	-
10	3	-	6	-	-	-	-	-

Symbol: - none identified; numbers indicate hypothetical person identifiers

Source: Stats NZ

Stage 2

In the second stage, we attached address information to each person in the relationship table (see table 1) to construct families in households. It was first necessary to identify all unique family nuclei. To do this, we created a list of all couples living together (whether or not they were living with children), and all single parents living with at least one child.

For example, in table 1 there are six Person ID-Partner links that represent couples living together – deduplicating these couples results in three family nuclei. Person ID 6 is included in the list of family nuclei as a single parent based on the matching address ID of his/her child (Person ID 10).

Next, we attached children to parents, and to the parents' family nucleus, when their addresses matched. Children were only available to be added to a parent's nucleus if they were not already assigned to a nucleus (ie they were not themselves a parent or living with a partner).

We created variables to indicate, within each family nucleus, the presence of: a couple, a single parent, and children. These variables were then used to recreate level 1 of the family type classification. Looking at family nuclei in addresses (our proxy for households) also let us create a high-level household composition classification.

Tables 2 and 3 show the structure of the family table and person-family linkage table created in stage 2.

Table 2
Family table

Family ID	Address ID	Family nucleus (AT ADDRESS)	Family type
1	1	1	Couple without children
2	2	1	Couple with children
3	3	1	One parent with children
4	3	2	Couple without children

Source: Stats NZ

Table 3
Person-family linkage table

Person ID	Family ID
1	1
2	1
3	2
4	2
5	2
6	3
7	4
8	4
9	-
10	3

Symbol: - none identified

Source: Stats NZ

6 Results

Results are presented in three subsections.

- comparing concepts and definitions
- comparing aggregate counts and estimates
- comparing individual level records.

Comparing concepts and definitions for households and families

While the census definitions of families and households are consistent with the statistical standards described in that section, there is currently no definition for either households or families within admin data.

Our investigation constructed proxy IDI-ERP households using addresses, and used family information collected across sources to group people into families in these households.

Households

As we constructed households and families using information from a variety of admin sources, there are conceptual differences from census that help explain discrepancies found between the two sources. We used address as a proxy for households in admin data, which differs from the census definition that allows multiple households at a single address. This means that some households were wrongly combined in the admin data.

A further problem is the limitations of geocoded address information in the admin data. Even after removing non-private dwellings, we found some addresses had hundreds of 'household' members – likely indicating a building with multiple units (eg apartment building) that did not have unique address identifiers for each dwelling. These people are grouped into a single household in admin data. This results from the current method of geocoding address information in the IDI, which can fail to give unique address identifiers to multiple units at the same street number address.

Households are more than just a group of people who share an address. Census households are made up of individuals who live in the same dwelling and who share living facilities. This type of information is not currently available from admin sources. Census households are also based on usual residence, whereas admin address information does not necessarily indicate residence. For example, some people prefer to receive communication from government agencies at an address other than where they usually reside. These differences help explain why membership in census households and admin households may differ.

Driven by census operational constraints, census households can only be counted for dwellings occupied on census night. There is no similar constraint on admin data. Thus households outside the scope of census could be covered using admin data – being unable to count households temporarily away from home on census night is not a limitation.

Families

In this investigation, families were defined in a similar way to census – as a couple, with or without child(ren), or one parent and their child(ren), all of whom usually reside together in the same household.

The problem of address not necessarily representing usual residence that we noted above also applies here. For example, children who continue to use their parents' address as a contact address after moving out will be incorrectly placed into a family with their parents.

However, the method we used to identify family nuclei within households means that issues noted above with multiple households being rolled into one do not apply to families. For example, families living in an apartment building identified by a single unique address identifier in admin data are coded as one household, but each family living in that building is identified as a distinct family unit (provided the family members have family information available in admin data).

While we derive census family relationships from information reported by household members (see appendix 1), admin family relationships are compiled from relationships known to government agencies.

While the census aims to collect family information for all usually resident individuals living in a family nucleus, family information is only collected in admin sources when the source agency needs it. For example, MSD only collects information on a partnership relationship when a person's benefit entitlement is affected by their partnership status. In contrast, DIA birth registrations place a legal requirement to enter information correctly, which forms strong evidence of a biological relationship between the child and parents. Further, some admin relationships were inferred (such as between parents on birth registrations, and between parents and children on visa applications). These are situations where it is very likely that people are related, but there will be some cases where this inference is incorrect.

Admin families are only created when individuals have family information recorded in the relevant sources. Unlike address information, which is available for the vast majority of the IDI-ERP, a large number of individuals have no family information in admin sources. Some of these people may legitimately not be living in a family, but others will be people in families who simply have not interacted with the source government agencies (or where family information was not collected by the source agencies) or where information is unable to be linked to the IDI spine.

It is also possible that census family information could be inaccurate if people do not accurately report on all individuals in their family, or if missing information means we cannot derive information for the family. Inaccuracies in either the census or admin data will affect comparisons between families.

Comparing aggregate counts and estimates for households and families

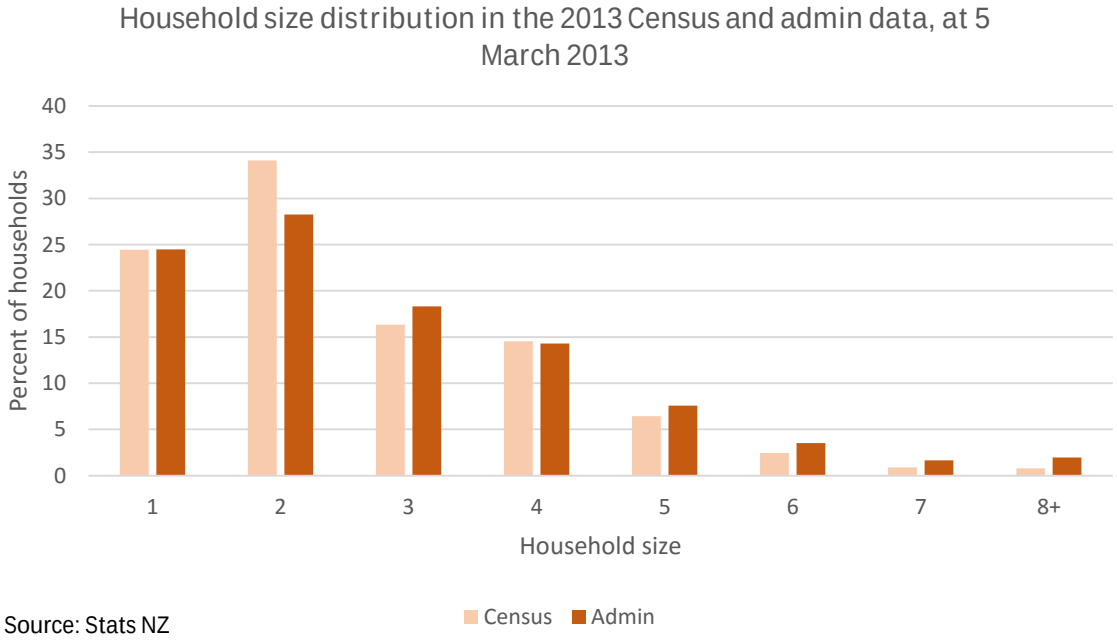
This section presents the results of aggregate comparisons between the admin estimates and the census counts for households and families, and compares coverage of the target population in each source.

Households

In admin data, there were 1,519,233 households, or 98.0 percent of the 1,549,890 households in the 2013 Census. This number is higher than previous investigations into admin households (Gibb & Das, 2015), indicating some improvement in address information in the IDI.

Figure 2 compares household size distribution in census and admin data. Overall the distributions are quite similar, although two-person households are underestimated by admin data and larger households (eg 5-or-more-person households) are overestimated. This finding is consistent with previous census transformation work on the household size distribution of admin households (Gibb & Das, 2015).

Figure 2



Source: Stats NZ

Table 4 presents counts for the number of households and the number of individuals in each household size category.

Table 4
Counts of households and individuals by household size, at 5 March 2013

Household size	Census household count *(Column %)*	Admin household count *(Column %)*	Census person count *(Column %)*	Admin person count *(Column %)*
1	355,284 *(22.9%)*	371,784 *(24.5%)*	355,284 *(8.5%)*	371,784 *(8.6%)*
2	527,673 *(34.0%)*	429,204 *(28.3%)*	1,055,346 *(25.4%)*	858,408 *(19.7%)*
3	254,862 *(16.4%)*	278,058 *(18.3%)*	764,586 *(18.4%)*	834,171 *(19.2%)*
4	235,332 *(15.2%)*	217,065 *(14.3%)*	941,328 *(22.6%)*	868,260 *(20.0%)*
5	106,329 *(6.9%)*	114,924 *(7.6%)*	531,645 *(12.8%)*	574,620 *(13.2%)*
6	41,184 *(2.7%)*	53,442 *(3.5%)*	247,104 *(5.9%)*	320,640 *(7.4%)*
7	15,402 *(1.0%)*	25,077 *(1.7%)*	107,814 *(2.6%)*	175,545 *(4.0%)*
8+	13,824 *(0.9%)*	29,679 *(2.0%)*	157,863 *(3.8%)*	344,325 *(7.9%)*
Total	1,549,890	1,519,233	4,160,970[(1)]	4,347,756[(2)]

1. Note this does not include the 81,078 people living in non-private dwellings.

2. This is the total number of people in the IDI-ERP living in a private dwelling with available address information.

Source: Stats NZ

Using the linked household population (all households we could link between census and admin data) we computed a cross-tabulation of household sizes in the census and admin data (table 5). This provides an idea of the magnitude of difference in household size – when the size determined by admin data differed from the census household size.

Table 5 shows that smaller households were more likely than larger households to have the same size in the admin data as in the census. Across all categories, when there was disagreement in household size the difference tended to be one person in either direction.

Table 5
Cross-tabulation of household size in the 2013 Census and admin data, at 5 March 2013

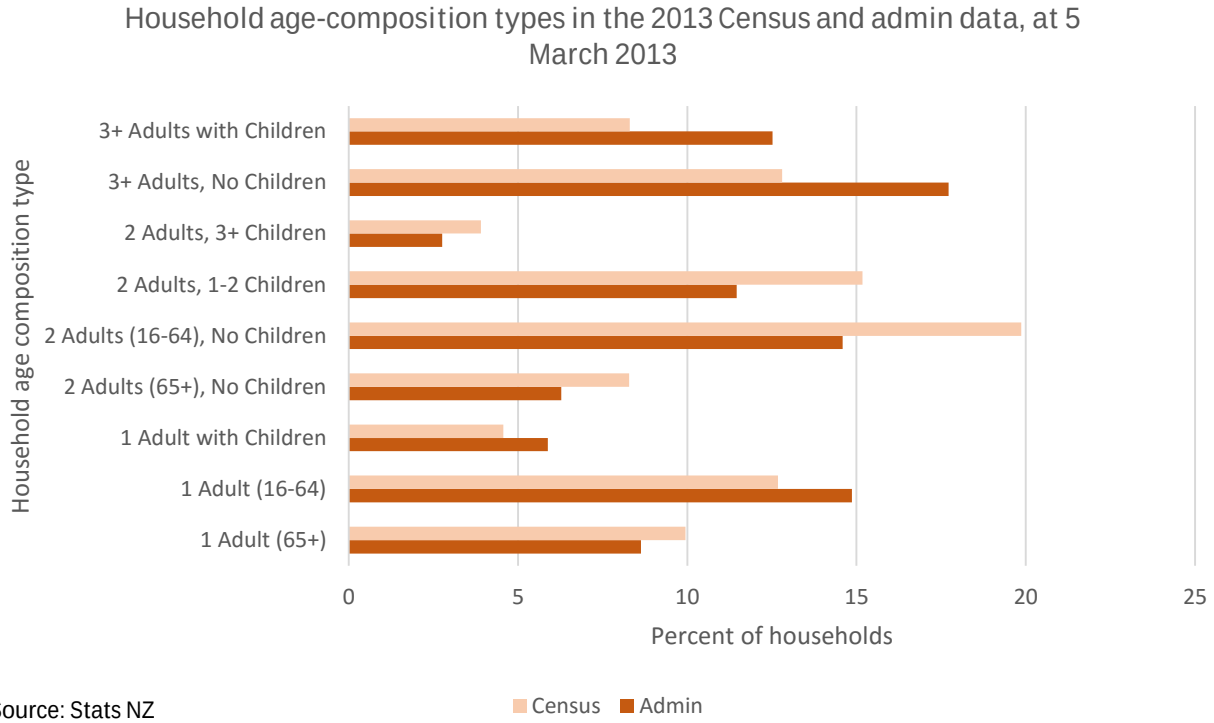
Census household size	Admin household size Count (Percent of census household size category)							
	1	2	3	4	5	6	7	8+
1	**187,521 (72.2%)**	47,892 (18.4%)	14,367 (5.5%)	5,394 (2.1%)	2,235 (0.9%)	987 (0.4%)	498 (0.2%)	717 (0.3%)
2	57,873 (13.4%)	**256,023 (59.1%)**	80,517 (18.6%)	24,930 (5.8%)	8,130 (1.9%)	2,937 (0.7%)	1,185 (0.3%)	1,287 (0.3%)
3	14,334 (6.8%)	36,981 (17.6%)	**97,332 (46.3%)**	39,972 (19.0%)	13,470 (6.4%)	4,713 (2.2%)	1,890 (0.9%)	1,488 (0.7%)
4	7,131 (3.7%)	14,685 (7.5%)	33,213 (17.0%)	**96,438 (49.5%)**	28,452 (14.6%)	9,267 (4.8%)	3,273 (1.7%)	2,526 (1.3%)
5	2,289 (2.6%)	4,209 (4.9%)	7,755 (9.0%)	15,795 (18.2%)	**37,785 (43.6%)**	11,664 (13.5%)	4,134 (4.8%)	2,976 (3.4%)
6	633 (1.9%)	1,257 (3.8%)	1,992 (6.0%)	3,585 (10.8%)	6,348 (19.2%)	**11,640 (35.2%)**	4,386 (13.3%)	3,219 (9.7%)
7	195 (1.5%)	420 (3.2%)	594 (4.6%)	960 (7.4%)	1,476 (11.4%)	2,454 (18.9%)	**3,744 (28.9%)**	3,132 (24.1%)
8+	129 (1.1%)	195 (1.6%)	312 (2.6%)	537 (4.5%)	753 (6.3%)	1,134 (9.5%)	1,599 (13.4%)	**7,302 (61.0%)**

Source: Stats NZ

To compare types of households between census and admin data, we used an approach from the Office of National Statistics (ONS, 2014) to classify households in both datasets, based on the age of household members. Household members were identified as: children (0–15), adults 16–64, and adults 65+. Households could then be categorised into nine household types based on the combination of child, adult 16–64, and adult 65+ members.

Figure 3 compares census and admin data for the distribution of household age-composition types.

Figure 3



Source: Stats NZ

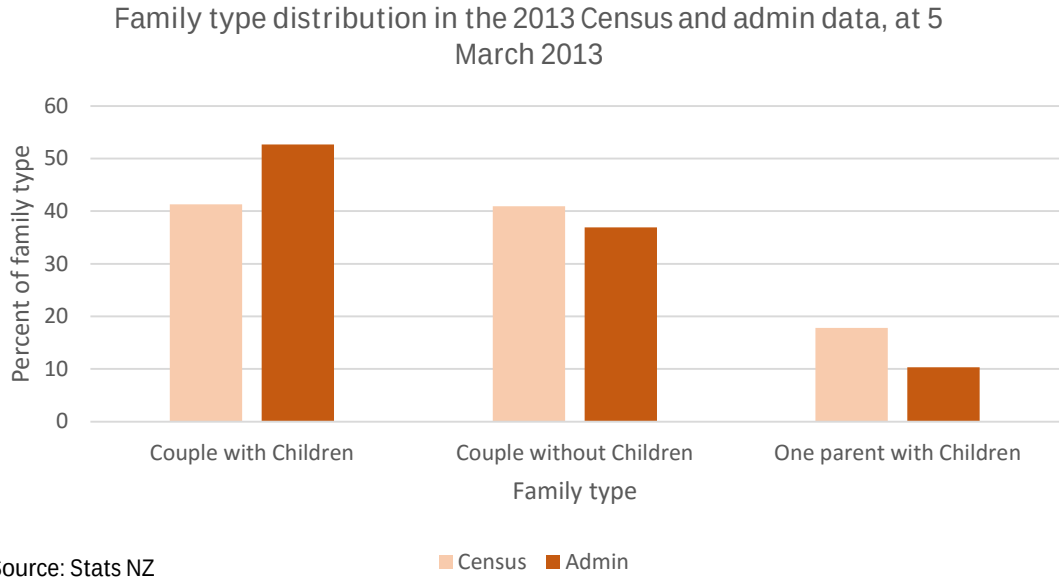
Figure 3 shows that, compared with the census, admin data overestimates the proportion of 3+ adult households (both with and without children), and underestimates the proportion of 2-adult households (both with and without children). The largest gap from census data was for '2-adult (16–64), no children', which indicates the underestimation of these younger adult households is largely to blame for the underestimation of 2-person households seen in figure 2.

Families

While previous census transformation work provided aggregate and individual-level comparisons for households, this investigation is the first to do this for families. Using admin data, we identified 2,063,007 people as living in 672,909 families. This count is 59.2 percent of the 1,136,397 families in the 2013 Census. Clearly there is a substantial amount of family information that is missing from admin data sources.

We compared the distribution of family types at level 1 of the family type classification for census and admin families. Figure 4 shows the results.

Figure 4



Source: Stats NZ

Figure 4 shows that within admin families, the proportion of couple-with-children families is overestimated, corresponding to a slight underestimate of couple-without-children families, and a larger underestimate of one-parent-with-children families.

Note: Figure 4 presents percentages, not counts, so while the proportion of couple-with-children families is overestimated, all three family types are undercounted in admin data. Table 6 presents the counts of families and individuals in each family type.

Table 6

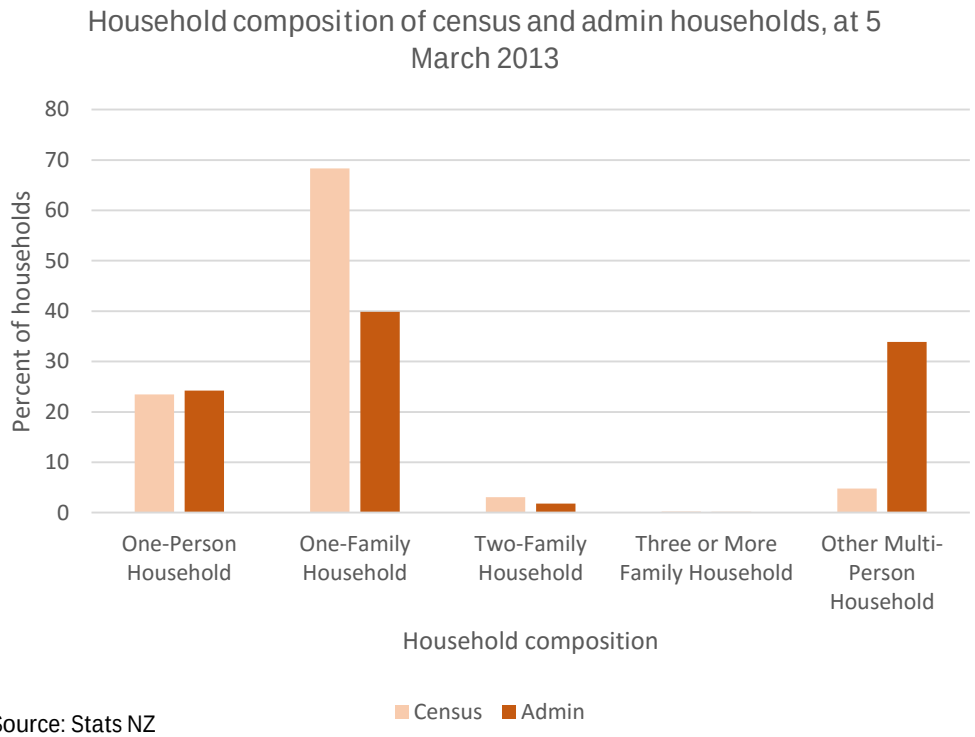
Counts of families and individuals by family type

Family type	Census family count *(Column %)*	Admin family count *(Column %)*	Census person count *(Column %)*	Admin person count *(Column %)*
Couple with children	469,287 *(41.3%)*	354,924 *(52.7%)*	1,846,902 *(55.7%)*	1,394,619 *(67.6%)*
Couple without children	465,303 *(40.9%)*	248,397 *(36.9%)*	930,609 *(28.1%)*	496,794 *(24.1%)*
One parent with children	201,804 *(17.8%)*	69,588 *(10.3%)*	538,608 *(16.2%)*	171,594 *(8.3%)*
Total	1,136,397	672,909	3,316,119	2,063,007

Source: Stats NZ

By grouping admin families into proxy households based on address, it was possible to create a high-level household composition classification. We compared the distribution of household types against census households. Figure 5 shows the results.

Figure 5



Source: Stats NZ

Figure 5 demonstrates the effects of undercoverage of family information in admin data on the distribution of household composition. When family information is missing (for people who really are living in a family, not those who are missing family information because they are not part of a family nucleus), people are instead categorised as a group of people not living in a family nucleus (ie 'other multi-person household'). This leads to a large overestimate of the proportion of 'other' multi-person households and an underestimate of one-family households.

Table 7 shows the counts of households and individuals in each household composition category. Counts are comparable only for one-person households, and the largest disparity is the admin overcount of people living in multi-person households.

Table 7
Counts of households and individuals by household composition type, at 5 March 2013

Household composition	Census household count *(Column %)*	Admin household count *(Column %)*	Census person count *(Column %)*	Admin person count *(Column %)*
One-person household	355,242 *(23.5%)*	368,589 *(24.3%)*	355,242 *(8.9%)*	368,589 *(8.5%)*
One-family household	1,030,497 *(68.3%)*	605,559 *(39.9%)*	3,154,494 *(78.8%)*	2,199,498 *(50.6%)*
Two-family household	47,436 *(3.1%)*	27,798 *(1.8%)*	269,211 *(6.7%)*	184,680 *(4.2%)*
Three or more family household	3,585 *(0.2%)*	2,352 *(0.2%)*	31,899 *(0.8%)*	44,958 *(1.0%)*
Other multi-person household	72,384 *(4.8%)*	514,938 *(33.9%)*	190,905 *(4.8%)*	1,550,034 *(35.7%)*
Total	1,509,144[1]	1,519,233	4,001,751[2]	4,347,756

1. There were 40,746 census households with an unidentifiable household composition.

2. There were 125,724 people with an unidentifiable household composition in the census.

Source: Stats NZ

Comparing individual-level records for households and families

The following analysis used a linked dataset (see Admin household information above) in which census records have, where possible, been linked to admin records. We compare members of households and families in the census and admin data at the individual level for linked records only.

Household membership

Individuals identified as living in households that could be linked between census and admin data (see Admin populations used in analysis above) were the basis of our household membership comparison analysis.

To compare household membership, we used household address information to link census households to the admin proxy households. Next, we could compare the individual members of each household for the two sources by using the IDI's unique person identifier. If household membership was exactly the same in the census and admin data, this was considered 'perfect' agreement. If half or more, but not all, census household members were the same in the admin household, we considered this 'partial' agreement. Census households with fewer than 50 percent of their members contained in the admin household were deemed to have 'poor' agreement.

Membership comparison also accounted for over-assignment in admin data, meaning that if all census household members were present in the admin household but additional members were also assigned to the admin household, the agreement was not perfect.

Table 8 compares household membership for census and admin households.

Table 8
Household membership comparison between 2013 Census households and admin households, at 5 March 2013

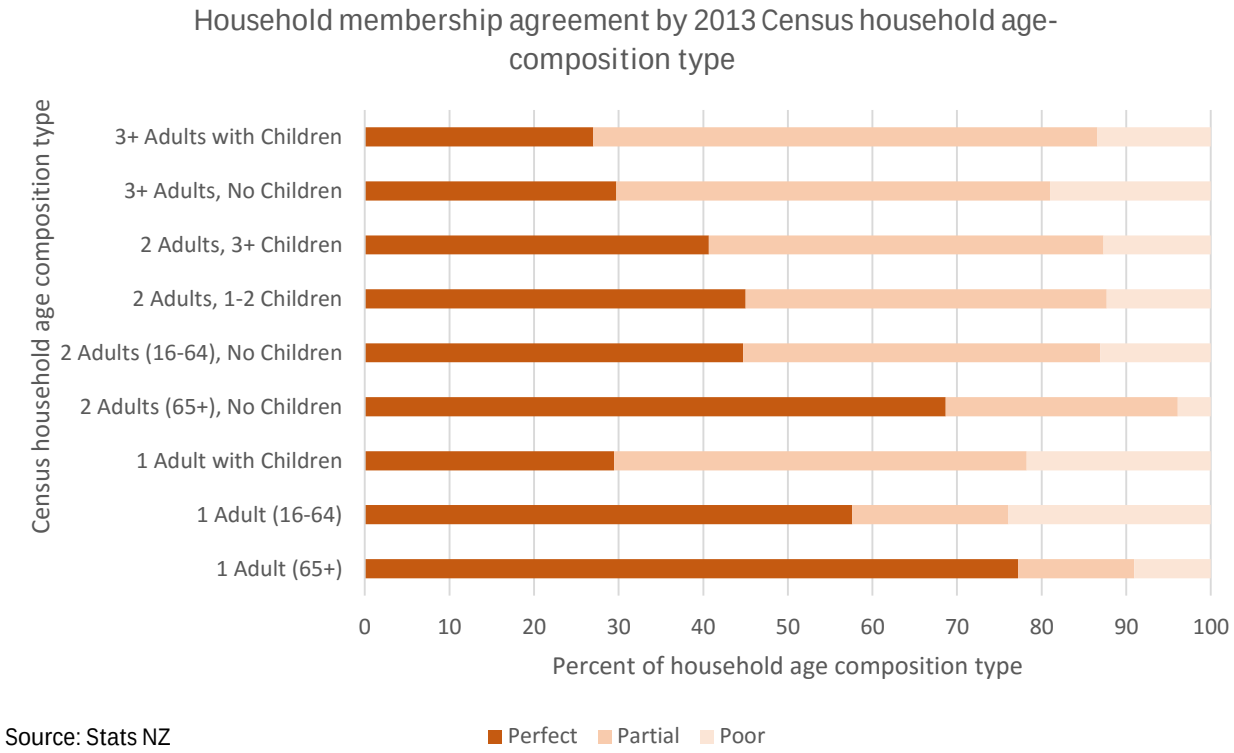
Membership agreement	Households count	Percent of linked census households[1]
Perfect	591,300	47.6
Partial	477,822	38.5
Poor	173,142	13.9
Not linked[2]	97,389	…

1. Refers to the percent of census households that could be linked to an admin household.
2. Refers to the count of census households that could not be linked to an admin household.
Symbol: … not applicable
Source: Stats NZ

Table 8 indicates that almost half of all households that could be linked between the census and admin data contained exactly the same members in both sources.

To determine the types of households with accurate admin household membership, we compared membership agreement across census household age-composition types. Figure 6 shows the results.

Figure 6



Source: Stats NZ

Figure 6 indicates that admin household information is most accurate for households containing 1 adult of any age with no children, and households with 2 adults over 65 with no children. Information is least accurate for households with 3 or more adults (both with and without children). These results can be partly explained by smaller households having fewer members to match, while larger households have more room for error (or missing data) in admin data.

Figure 6 also shows that information on households with older adults (65+) is more accurate than households with younger adults (16–64). This could be due to having better admin address information for older adults, who are less mobile (fewer address changes) and are more likely to interact with certain sources of admin address information (eg health care systems).

Family relationship information

To compare family relationship information for census and admin data, we compared it on an individual basis rather than at the grouped family level (as was done for households). We could not directly link census family IDs to unique families in admin data, hence the use of an individual-level approach.

For all individuals in the linked census-IDI-ERP dataset who we found to be living in a family in both sources (1,733,425 people, or 53.6 percent of all people in census families), we compared all identified family nucleus members for each individual. In other words, for each individual we could link between the census and admin data, and who was placed in a family nucleus in both sources, we compared the other members of that family nucleus.

We calculated agreement levels of family relationship information in an analogous way to household membership: ‘perfect’ agreement for a person indicates that all members of that person’s family nucleus are the same in the census and admin data; ‘partial’ agreement for a

person indicates that half or more, but not all, of that person's family nucleus members are the same in census and admin data; and 'poor' agreement indicates fewer than half of a person's family nucleus is the same across sources.

Table 9 presents family member agreement for individuals linked between census and admin data with family information in both sources.

Table 9
Family nucleus member comparison between 2013 Census individuals and admin individuals, at 5 March 2013

Member agreement	Count	Percent of census individuals[1]
Perfect	1,182,483	68.2%
Partial	426,915	24.6%
Poor	124,026	7.2%
Admin missing[2]	1,474,842	…

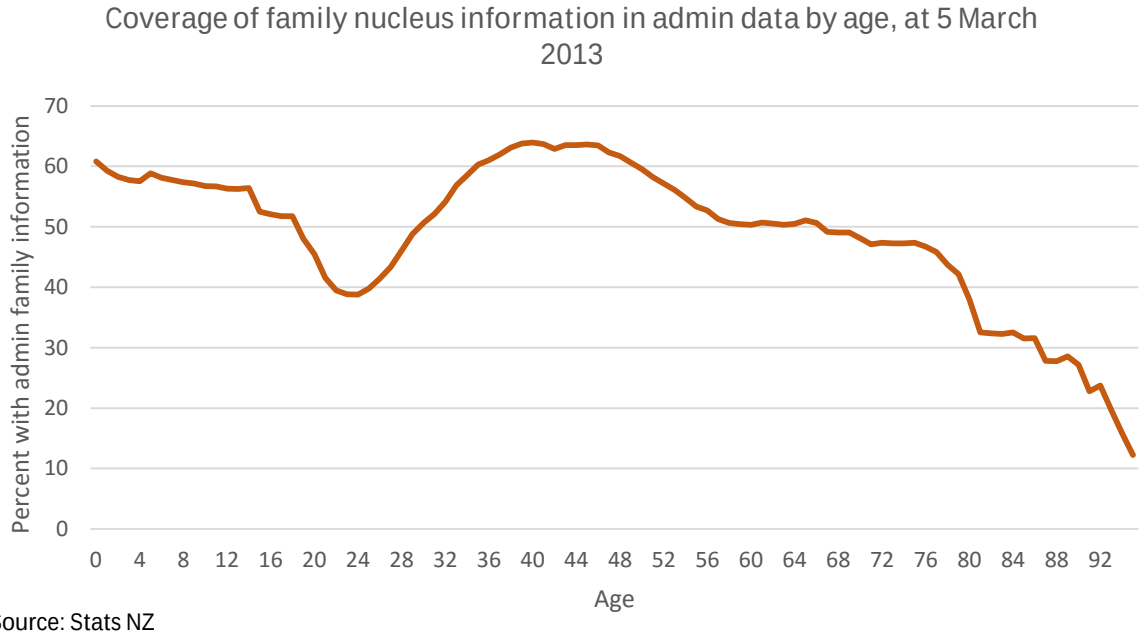
1. Percent of linked census individuals with family information in both census and in admin sources.
2. Count of linked people in census families who are not in an admin family.
Symbol: … not applicable
Source: Stats NZ

Table 9 shows that when we can place individuals into an admin family, more than two-thirds of the time all admin family members perfectly match census family members. A relatively small number of people had a 'poor' rating for admin family member information. This suggests that when family information is available for an individual in admin sources, information is likely to also be available for the other members of their family and this information tends to match that from census.

The primary issue with family nucleus information in admin data may be that it is missing for close to half of all people identified by the census as living in a family. When comparing people in census families who are and are not placed into a family nucleus in admin data, we see that certain types of people are more likely to have admin family information than others.

The presence of family nucleus information in admin data differs by age. Figure 7 shows the percentage of individuals living in a census family who are also placed into an admin family nucleus (by year of age). Family nucleus information is most likely to be available for adults between 32 and 55 years, and children under 15. There is particularly poor coverage of people in their mid-20s and people aged over 80 years.

Figure 7



Source: Stats NZ

People in couple-with-children families were most likely to have admin family nucleus information, followed by couples without children. People in single-parent families were least likely to have admin family nucleus information (see table 10).

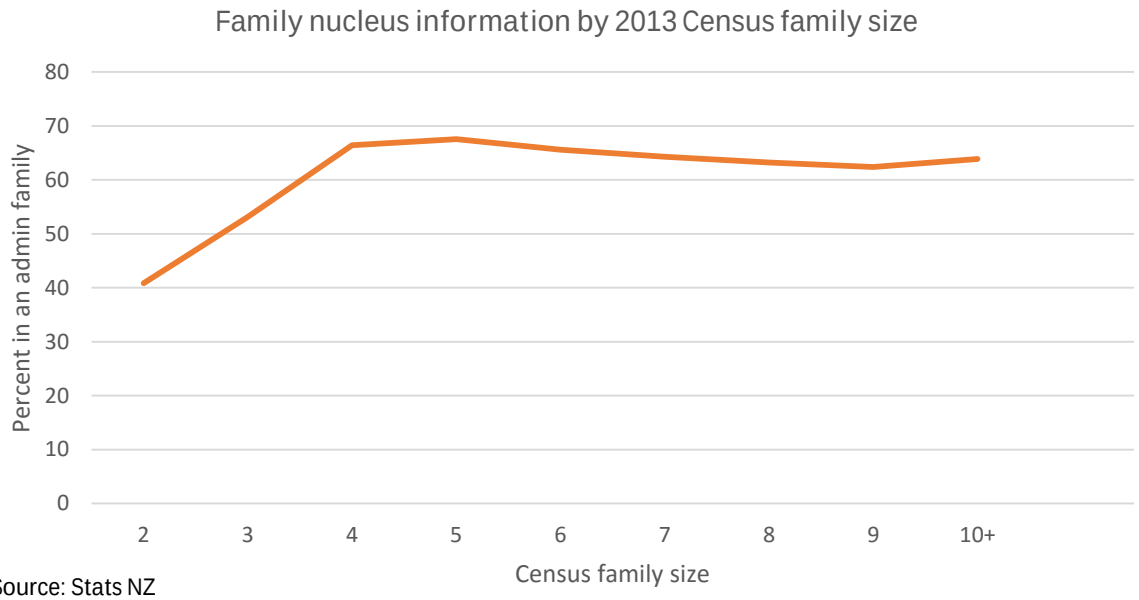
Table 10
Family nucleus information by 2013 Census family type

Census family type	In admin family nucleus n (%)		Not in admin family nucleus n (%)	
Couple without children	398,352	(44.2%)	502,218	(55.8%)
Couple with children	1,214,979	(67.6%)	582,555	(32.4%)
One parent with children	132,837	(26.0%)	377,331	(74.0%)

Source: Stats NZ

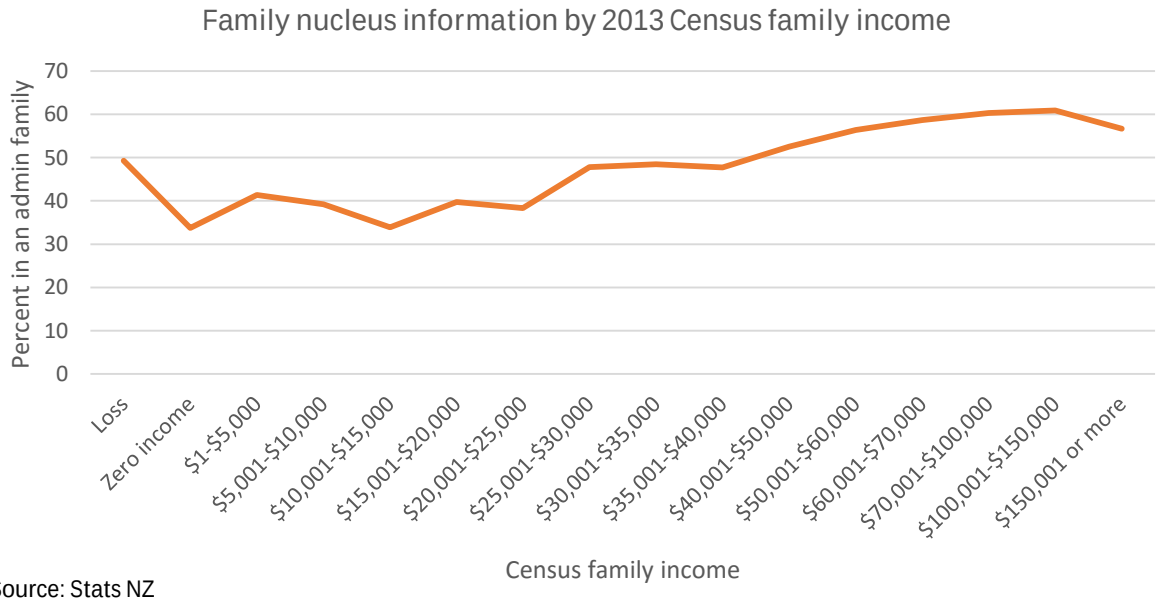
People living in larger census families were more likely to have admin family nucleus information than people living in smaller census families (see figure 8). The likelihood of admin family nucleus information increases with increasing census family income (see figure 9).

Figure 8



Source: Stats NZ

Figure 9



Source: Stats NZ

Poor coverage of family nucleus information for people with low family incomes may seem surprising given the use of two types of financial benefits data (MSD and WFF) as source information. However, further investigation indicated this effect was largely attributable to the 60+ population. Individuals in this age group tended to have low family incomes and were the group of people missing the most admin family information.

Summary of results

Taken as a whole, the results indicate good coverage of address information in admin data, which allows us to calculate aggregate-level household statistics. However, quality issues with address information diminish our ability to correctly place individuals into households, which affects the calculation of any household-level statistics. Households containing only older adults (65+ years) were the ones most likely to have correct membership.

Creating admin families requires both address and relationship information, and the results indicate coverage issues with the latter. Adults aged over 60 years and those in their mid-20s were least likely to have admin family information available. Single-parent-with-children families were the least likely to be identified. However, when family information was available for individuals it tended to match the family information reported in census.

7 Discussion and conclusions

This paper presents an investigation into the potential for obtaining household and family information from admin data currently available in the IDI. We compared 2013 Census data on households and families with similar information obtained from multiple admin sources available in the IDI in June 2017.

Overall, the admin sources investigated show potential for providing household information on an aggregate level, despite some limitations. However, the lack of coverage of families in admin data means the potential for producing census-type information on families is currently minimal. Missing family information also affects our ability to replicate the census household composition variable.

Households

This investigation shows that address information has improved since previous census transformation work on households – the total number of admin households now approximates the census household count (98 percent of census households).

The admin household size distribution is quite similar to the census household size distribution, barring an undercount of 2-person households and an overcount of larger households, which tend to be 2-person young-adult households mistakenly identified as larger households. One explanation for this error could be the greater mobility of younger adults and the lack of need for admin sources to keep up with address changes. This results in additional people being placed at addresses they have actually moved from.

Admin address information is worst for young adults aged about 16–30 years (Stats NZ, 2017), which affects our household information for this group. Address information from NZTA (covering drivers' licences and motor-vehicle registrations) is available in the IDI and had high accuracy for young adults (Stats NZ, 2017). Unfortunately, NZTA data is only available from 2015 and could not be investigated in the present analyses. Future work should explore whether NZTA address data improves our ability to accurately create households, perhaps improving our count of 2-person households.

When we explored the composition of households using categories based solely on the age of household members, the distribution of admin households proved more accurate than when it was created using the traditional census household composition variable that incorporates family information. This was largely due to the extent of missing family information. Creating household age categories presents another way to view household membership when there is insufficient information to recreate the standard household composition variable – an extension of current practice that classifies household members as over and under 15 years.

Not quite half of all admin households had exactly the same household membership as census households, which indicates we still have work to do to improve our ability to place people into households. While our aggregate information on households (using admin data) may suffice for some purposes, incorrect household membership will create problems with variables calculated at the household level, such as household income.

The ability to accurately place people into households depends on correct address information being available from admin sources. Stats NZ (2017) reported the methodology for selecting a

person's address that we used in this analysis will be correct for 81 percent of people. The correct address is available to be selected from any admin source for 86 percent of people. In other words, there are only a very small number of people for whom the wrong address is being selected when the correct one is available – usually their correct address is simply not available from admin sources.

Stats NZ recently developed a Statistical Location Register (SLR) that provides more information on dwellings (not just addresses), including whether those dwellings are private or non-private. Currently this dwelling information is only available through census data. In the SLR, the private and non-private dwelling information is based on 2013 Census data and building consents, and will be updated with 2018 Census dwellings. In future we anticipate IDI addresses to be matched to the SLR reference address ID. This reference ID will enable us to link through to the SLR dwelling indicators.

If the SLR is regularly updated and maintained following the 2018 Census, this would improve our future construction of households and lead to more-accurate counts and more-accurate household statistics. It would also allow private dwellings to be identified through means other than census data.

Families

This investigation has shown that when family information is available in admin data, it tends to be quite accurate. Family nucleus membership was correct for nearly 70 percent of people in admin families (another 25 percent had half or more of their family members correctly identified). However, there are clearly severe coverage issues with admin families. In particular, information is lacking for people in single-parent families, and for people aged 60+ and those in their mid-20s. At present there is not sufficient information from admin sources to provide census-type family information.

However, this investigation shows it is possible to create families using admin data and presents a methodology for doing so. If the underlying data can be improved, this methodology will be useful for constructing admin families for family statistics and for use by IDI researchers.

One explanation for poor family coverage is the lack of historical data from source agencies. DIA data has only been digitised since 1998, restricting the number of familial links that can be made before this time. Further, births, marriages, and civil unions occurring overseas are mostly not captured by admin data. MBIE visa applications data is only available from 1997, so families that arrived in New Zealand before then are not included in the dataset (but might be identified through another source used in this investigation).

The September 2017 refresh of the IDI included an additional 10 years of back digitisation of DIA births and marriages data, which means better identification of relationships (ie more records will be linked to the spine/IDI-ERP and the relationship information on records will become useful). Preliminary analysis indicates only a small improvement to the count of admin families we provided in this paper, as many relationships were identified through other sources. However, this additional DIA information will allow relationships that can be identified across multiple sources to be validated.

Although information on admin families does not currently meet census needs, admin sources could provide an expanded view of families and family statistics beyond that currently collected by census. In particular, it would be possible to explore families within and across households – an

emerging topic of interest given the changing nature of families today. For example, by combining biological relationship information from DIA (birth registrations) with the family nuclei constructed, based on address, we can identify situations where children are living with a birth parent and step-parent (and also identify their biological relationship to other children in the family nucleus). The census does not currently report any statistics on step- or blended families.

Directions for future work in this area include investigating imputation of unknown family relationships, based on known links in the household/family. Modelling and weighting techniques that allow estimation of population-level statistics, using the subset of families that can be identified from admin sources, can also be investigated, as well as making use of family information available from household surveys.

It will also be important to explore ways to enhance and improve coding of the information currently available in the IDI and to investigate potential new sources of admin data that could improve the construction of households and families.

References

Gibb, S, Bycroft, C, & Matheson-Dunning, N (2016). Identifying the New Zealand resident population in the Integrated Data Infrastructure (IDI). Retrieved from www.stats.govt.nz.

Gibb, S, & Das, S (2015). Quality of geographic information in the Integrated Data Infrastructure. Retrieved from www.stats.govt.nz.

Families Commission Act (2003). Families Commission Act. Public Act: 2003 No 128. Retrieved from www.legislation.govt.nz.

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

Office for National Statistics (2014). Beyond 2011: Statistical Research Update (M13) [PDF, 49p]. Retrieved from www.ons.gov.uk.

Stats NZ (2014a). An overview of progress on the potential use of admin data for census information in New Zealand. Retrieved from www.stats.govt.nz.

Stats NZ (2014b). Estimated resident population 2013: Data sources and methods. Retrieved from www.stats.govt.nz.

Stats NZ (2017). Experimental population estimates from linked admin data: 2017 release. Retrieved from www.stats.govt.nz.

Suei, S (2016). *Comparing legally registered relationship status information from census and admin data*. Unpublished document; available on request.

Appendix 1 – Household and family census information

Statistical concepts for families and households

Stats NZ's Statistical Standards and Classifications provide definitions for key household and family concepts. These standards and classifications are designed for use in official statistics collections and are those used in the 2013 Census.

Definitions for family and household

Our statistical standard defines family and household as follows.

- A family (or family nucleus) is defined as a couple, with or without child(ren), or one parent and their child(ren), all of whom usually reside together in the same household. The children do not have partners or children of their own living in the same household.
- A household is one or more people usually resident in the same (private) dwelling, who share living facilities. A household can contain one or more families, or can contain no families at all. A household that does not contain a family nucleus could contain unrelated people, related people, or could simply be a person living alone.

The primary relationship concepts needed to form a family nucleus are partner, parent, and child. These concepts are defined in the Statistical Standard for Relationship.

- A partner is a person with whom another person is: married to or in a civil union with, in a de facto relationship, or in another partnership. These relationships include both same-sex couples and opposite-sex couples.
- A parent is the mother or father (birth/biological, adopted, or step-) of a 'child in a family nucleus'. In the absence of either a mother or father, a 'person in a parent role' may be designated. A person in a parent role is a person who usually resides with the child and who can be considered a parent according to current social norms regarding parenting.
- To be a 'child in a family nucleus', a person must usually reside with at least one parent, and have no partner or child(ren) of their own living in the same household. Note that a 'child in a family nucleus' can be a person of any age.

Definitions of both family and household involve 'usual residence', and the definition of household involves the 'dwelling' and 'living facilities'. These are defined as follows:

- '**Usual residence'** is the address of the dwelling where a person considers themself to usually reside (rather than an address they may be visiting temporarily). Guidelines are provided for a number of specific situations such as students at boarding school and children in joint custody.
- A dwelling is any building or structure, or part thereof, that is used (or intended to be used) for human habitation. It can be of a permanent or temporary nature and includes structures such as motels, hotels, hospitals, prisons, motor homes, huts, and tents. Dwellings are classified as private or non-private.
    - A private dwelling accommodates a person or a group of people, but is not available to the public. Permanent private dwellings include houses and flats, residences attached to a business or institution, and baches, cribs, and huts. Caravans, cabins, tents, and other

makeshift dwellings that are the principal or usual residence of households are classified as temporary private dwellings.

  - All other dwellings used for human habitation (or intended to be used), are non-private and are available to the public. They may be available for use generally, or by virtue of occupation or study, special needs, or legal requirements (ie prisons). Such dwellings may have facilities (such as a dining room) that are for shared use.

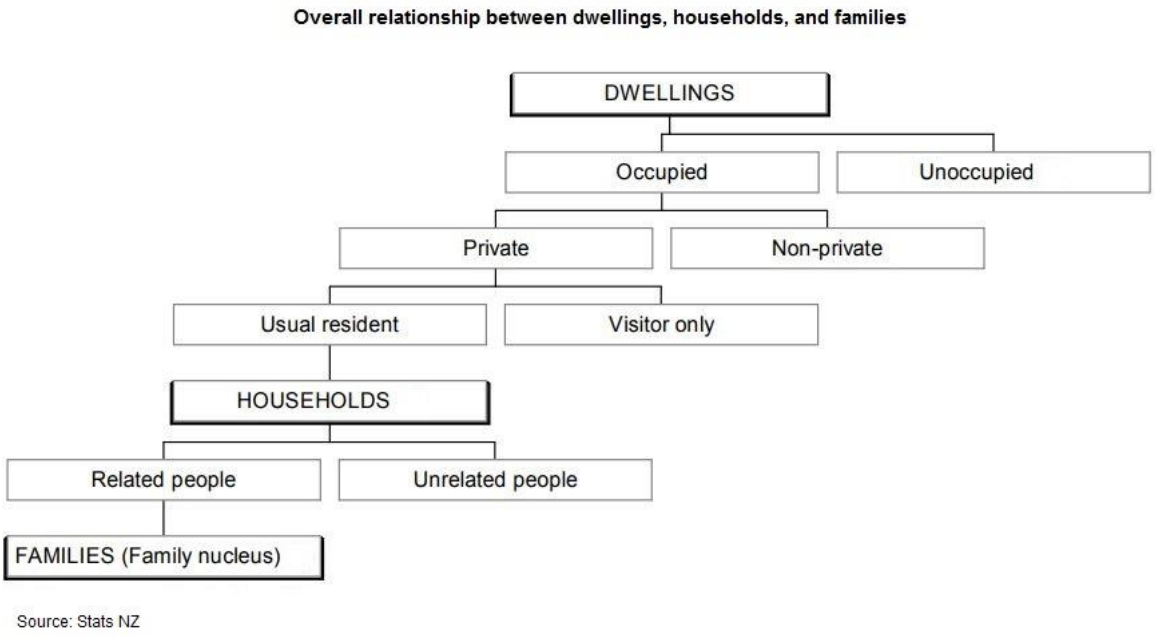
- Sharing living facilities means sharing eating facilities, cooking facilities, bathroom and toilet facilities, and a living area.

A key concept in these statistical standards and in the census is that families and households are formed from people who live together in the same dwelling, and that the dwelling they share must be a private dwelling.

One household occupies one private dwelling. However, not all dwellings have an associated household, because some dwellings are non-private, some are unoccupied, and some are occupied by visitors rather than usual residents.

The overall relationship between dwellings, households and families is shown in Figure 10.

Figure 10



Dwellings, households, and families (standard terms) has more information.

Classifications

Relevant census family and household variables include:

- census household count
- number of usual residents in household, including further breakdown by age
- family type
- household composition.

Each of these variables produced by the census is based on a formal statistical classification. The following sections outline the key features of these classifications.

Census household count

The census household count is a count of all households in New Zealand where at least one person is present at the usual address of the household on a given census night. Households whose members are all away temporarily elsewhere in New Zealand and/or temporarily overseas on census night are excluded, unless there is someone at their usual residence (eg a visitor) to identify them.

Number of usual residents in household

The size of a household is measured by the number of usual residents in that household. The census produces a further breakdown to 'Number of usual residents aged 15 and over in household' and 'Number of usual residents aged under 15 in household'.

Usual residents are people who usually live in the surveyed dwelling.

Family type

The family type classification specifies family membership based on the type of relationship between individuals and on shared residence. Family type is a hierarchical classification with three levels. Level 1 follows the definition of a family above, classifying a family or 'family nucleus' according to the presence or absence of couples, parents, and children. A family nucleus can be a couple without a child, couples with a child (or children), or one parent and their child (or children), all of whom usually reside together in the same household. The children do not have partners or children of their own living in the same household.

Level 2 of family type examines the nature of the biological relationships (eg birth children, step-children, and/or grandchildren), while the third level allows a further classification of step-families. The family type classification can be used along with other classifications (eg household classification or child dependency status) to create new classifications. 'Extended family type' indicates the number of generations in an extended family, which exists when there are family relationships in a household beyond one family nucleus.

Appendix 2 has the full family type classification.

This statistical classification of family type is a relatively narrow definition, and does not fully align with other broader definitions of family. For example the Families Commission Act (2003) defines a family as "a group of people related by marriage, civil union, blood, or adoption, an extended family, two or more persons living together as a family, and a whānau or other culturally recognised family group".

Household composition

Household composition classifies households according to the relationships between usually resident people. Households are classified according to the presence, number, and types of family nuclei, and the presence of related and unrelated people.

Household composition is classified into three levels. Level 1 is the number of family nuclei present, or if none are present, then the number of people present (either a one-person household or other multi-person household).

Appendix 3 has the full household composition classification.

Other derived information

Many other family and household-related variables are derived from census data. These include: type of couple; number of adult children in family; number of children in family; number of dependent children in family; number of people in family; age of family members; sex of sole parent; individual's role in family nucleus; identification of individual's family nucleus; number of dependent children in household; age of youngest child in household; age of youngest dependent child in household.

New Zealand Census of Population and Dwellings

The New Zealand census is a de facto census; that is, the census counts people where they are on census night. People who are away from home on census night are counted at their census night address (not their usual residence). The census aims to count everyone who is in New Zealand on census night. Overseas visitors are included in the census, while New Zealand residents who are not in the country on census night are not included.

Construction of families and households in the census

Following the statistical standard definitions, only people who usually live at the dwelling are included when constructing household and families in the census. Due to the de facto nature of census, usual residents fall into two groups: people who usually live in the dwelling and were present on census night, and those who usually live there but were absent on census night (absentees). Unlike most census variables, absentees are included in constructing the households and families.

The census uses the following information to construct households and families:

- Address. The address of the dwelling (from the dwelling form) establishes the location of the household. The usual residence address of the individual (from the individual form) helps establish which individuals are eligible to be part of a household and any family within a particular dwelling.
- Usual residence indicator. The usual residence indicator describes the relationship between a person's usual residence and their census night address. The categories for usual residence indicator are: same as census night address, elsewhere in New Zealand, overseas, or no fixed abode.
- Absentees. These are people who usually live in a particular dwelling, but who were absent on census night. Absentees are included as members of the household, as long as they were reported as being absent by the reference person on the dwelling form. There is no information about absentees where the whole household was absent on census night. Absentees are recorded as being either elsewhere in, or away from, New Zealand on census night.
- Reference person. The reference person is the individual who completed the dwelling form on census night.

- Relationship to reference person. Relationship to reference person indicates the kind of relationship each person in a defined group of people, family or not, has to the reference person (eg father or boarder). This information helps establish which individuals belong to the same family within a household.
- Living arrangements. Living arrangements describe a person's relationships to all other people (family or not) with whom they usually share a residence.

Address, usual residence indicator, and absentees determine who is eligible to be included as members of the household. Relationships between members of the household are established largely through the information from the relationship to reference person. However, this cannot always identify all the relationships in a given household, especially where a household has multiple families. The living arrangements information is used to further identify relationship information for more complex family situations.

Some surveys gather more comprehensive information through a relationship matrix that allows collection of all relationships between all household members. The New Zealand census has so far not adopted this approach.

Table 11 summarises the census information used to derive census household counts, household size, family type, and household composition. As well, some consistency checks are carried out as part of census processing to ensure families are correctly derived. For example, a check ensures age of parent is greater than age of child and that there is one male and one female if living arrangements indicate an opposite-sex couple. For large households, the name of the respondents and their relationship to the reference person and their living arrangements are used to determine families beyond that of the reference person's.

Table 11
Information used to derive census household and family variables

	Census household count	Number of usual residents in household	Family type	Household composition
Address	x	x	x	x
Usual residence indicator	x	x	x	x
Absentees	x	x	x	x
Reference person			x	x
Relationship to reference person			x	x
Living arrangements			x	x

Source: Stats NZ

Families and households information for longitudinal analysis in the census

The census itself is not longitudinal. However, Stats NZ has separately linked censuses between 1981 and 2006, which allows longitudinal investigation of individuals for whom a link can be made between one census and another. These linked census datasets can provide longitudinal information about families and households, for the census years where links were made.

Appendix 2 – Family type classification

Level 1	Level 2	Level 3	
1			Couple without children
	11		Couple without children
		111	Couple without children
2			Couple with child(ren)
	20		Couple with child(ren) nfd
		200	Couple with child(ren) nfd
	21		Couple with birth/biological, adopted children
		211	Couple with birth/biological, adopted children
	22		Step family
		220	Step-family nfd
		221	Non-blended step-family
		222	Blended step-family
	23		Couple with grandchild(ren)
		231	Couple with grandchild(ren)
	24		Couple with other child(ren) only
		241	Couple with other child(ren) only
3			One parent with child(ren)
	30		One parent with child(ren) nfd
		300	One parent with child(ren) nfd
	31		One parent with birth/biological, adopted child(ren)
		311	One parent with birth/biological, adopted child(ren)
	32		One parent with grandchild(ren)
		321	One parent with grandchild(ren)
	33		One parent with other children only
		331	One parent with other children only

Note: nfd not further defined

Source: Stats NZ

Family type has more information.

Appendix 3 – Household composition classification

Level 1
Level 2
Level 3

1 One-family household (with or without other people)
- 10 One-family household nfd
  - 100 One-family household nfd
- 11 Couple only
  - 111 Couple only
- 12 Couple only and other person(s)
  - 120 Couple only and other person(s) nfd
  - 121 Couple only and other person(s), some or all related
  - 122 Couple only and other person(s), all unrelated
- 13 Couple with child(ren)
  - 131 Couple with child(ren)
- 14 Couple with child(ren) and other person(s)
  - 140 Couple with child(ren) and other person(s) nfd
  - 141 Couple with child(ren) and other person(s), some or all related
  - 142 Couple with child(ren) and other person(s), all unrelated
- 15 One parent with child(ren)
  - 151 One parent with child(ren)
- 16 One parent with child(ren) and other person(s)
  - 160 One parent with child(ren) and other person(s) nfd
  - 161 One parent with child(ren) and other person(s), some or all related
  - 162 One parent with child(ren) and other person(s), all unrelated

2 Two-family household (with or without other people)
- 20 Two-family household nfd
  - 200 Two-family household nfd
- 21 Two 2-parent families
  - 210 Two 2-parent families nfd
  - 211 Two related 2-parent families
  - 212 Two unrelated 2-parent families
- 22 One 2-parent family and a 1-parent family
  - 220 One 2-parent family and a 1-parent family nfd
  - 221 One 2-parent family related to a 1-parent family
  - 222 One 2-parent family unrelated to a 1-parent family
- 23 Two 1-parent families
  - 230 Two 1-parent families nfd
  - 231 Two related 1-parent families
  - 232 Two unrelated 1-parent families
- 24 Other 2-family household
  - 241 Other 2-family household

3 Three or more family household (with or without other people)
- 31 Three or more family household (with or without other people)
  - 311 Three or more family household (with or without other people)

4 Other multi-person household
- 40 Other multi-person household nfd
  - 400 Other multi-person household nfd
- 41 Household of related people

Level 1

- 411 Household of related people
- 42 Household of related and unrelated people
  - 421 Household of related and unrelated people
- 43 Household of unrelated people
  - 431 Household of unrelated people

5 One-person household

- 51 One-person household
  - 511 One-person household

6 Household composition unidentifiable

- 61 Household composition unidentifiable
  - 611 Household composition unidentifiable

Note: nfd not further defined

Source: Stats NZ

Household composition has more information.