

# The Experimental Administrative Population Census (APC)

NZPopCon, Auckland - 29 August 2023

Nadia Andrews, Christine Bycroft, Hannes Diener, Ruby Dixon, Ivan Welsh

# IDI Disclaimer

These results 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/>.

These 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.

# History of Administrative Data and the New Zealand census

Traditional census  
-2013

Combined model  
2018, 2023

APC  
(2021 - 2023)



## Waikato Region

Find 2018 Census data about **Waikato Region** ([find map boundary](#)). Topics include population, ethnicity, religion, health, employment, income, education, and housing.

Search by small area, district or region

On this page

- [Population and dwellings](#)
- [Ethnicity, culture, and identity](#)
- [Health](#)
- [Work, income, and unpaid activities](#)
- [Education and training](#)
- [Transport](#)
- [Housing](#)
- [More data and information](#)

### Population and dwellings

Find our population count by age and sex, and the number of dwellings. Also find where we lived a year before the census, the status of our relationships, and number of children born.

### Population counts

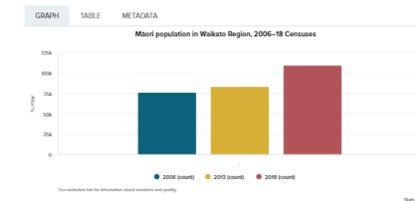
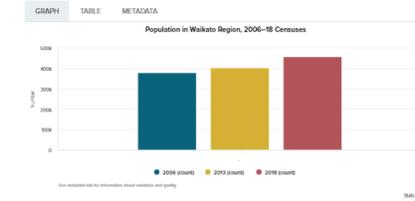
Find information about the number of people in **Waikato Region** for the census usually resident population, how old they are, and how many are male or female.

The census usually resident population count variable is rated as very high quality. [Census usually resident population count](#) has more information about this variable, for example, definitions and data quality.

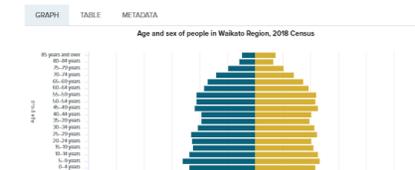
### Quick stats about population counts for Waikato Region (2018 Census)

Number of people	Median age	Males
458,202	37.4 years	226,386
Females	Number of Māori	Mean median age
231,816	109,488	25.0 years

### Population counts, over time



### Population counts, by age and sex



# What is the experimental Administrative Population Census?

*“A **practical** demonstration of what can be achieved **now** with admin data **only**\* to produce census-type information.”*

APC’s purpose:

- Gauge technical feasibility of admin first census to enable decision making
- Evaluate the data and methods as used in the APC
- Enable engagement on whether information needs are met
- Use feedback to direct further development

=> The APC is a conversation **catalyst**

# The APC

- isn't Census 2023 (...)
- isn't Census 2028 (...)
- isn't official statistics (...)
  - ..., but there is knowledge sharing, common methodology, common data sources, and mutual benchmarking.
- does not represent all population measurement research.



# What does the APC include?

Annual time series: 2006 – 2021 (reference date 30<sup>th</sup> June)

Admin New Zealand resident population ("IDI-ERP")

## **19** census variables:

- Population structure: Age, sex, geography (address), usual residence n years ago, number of children born
- Identity and culture: ethnicity (Level 1-4), Māori descent, birthplace, and years since arrival in NZ
- Income: total personal income by income source
- Education: highest qualification, post-school field of study, study participation
- Work: employment indicator, employment status, industry, sector of ownership

# APC Products

- Available to all IDI researchers
  - Unit record data
  - Source code & Metadata
  - ARP spells
  - Unit record quality scores
- Publicly available on [stats.govt.nz](https://stats.govt.nz)
  - Aggregate tables
  - Methodology report
  - Information for Māori
  - Interactive App

# Interactive App

## App Allows

- Limited crosstabs
- Maps to small geography level
- Timeseries
- Customised table download

Select an attribute:  
Select an attribute:  
Income sources

Select geographic level:  
 SA2  TALB  National

Select a year range:  
2006 2021

Select sex:  
 Total  Male  Female

Select age grouping:  
 Single year  Five years  
 Life cycle  No age

Select ages:  
15-19 Years 20-24 Years 25-29 Years

Select source of income:  
Self-employed income

Display as:  
 Counts  Percentage population

Show:  
 Codes  Names  Both

Apply filters

About Map Time Series Tables

About Map Time Series Tables

Search:

year	age_type	age_group	area_name	data_name	Total	total_coverage	TP_total
2006	Five years	15-19 Years	New Zealand	Self-employed income	8730	2.8	311802
2006	Five years	20-24 Years	New Zealand	Self-employed income	18504	6.36	290814
2006	Five years	25-29 Years	New Zealand	Self-employed income	27807	10.51	264693
2007	Five years	15-19 Years	New Zealand	Self-employed income	8988	2.82	318636
2007	Five years	20-24 Years	New Zealand	Self-employed income	19017	6.53	291030
2007	Five years	25-29 Years	New Zealand	Self-employed income	28029	10.44	268542
2008	Five years	15-19 Years	New Zealand	Self-employed income	8700	2.7	322170
2008	Five years	20-24 Years	New Zealand	Self-employed income	18561	6.31	294024
2008	Five years	25-29 Years	New Zealand	Self-employed income	28011	10.24	273570
2009	Five years	15-19 Years	New Zealand	Self-employed income	6780	2.09	324042

Show 10 entries

Showing 1 to 10 of 48 entries

Previous 1 2 3 4 5 Next

Export

filename.csv

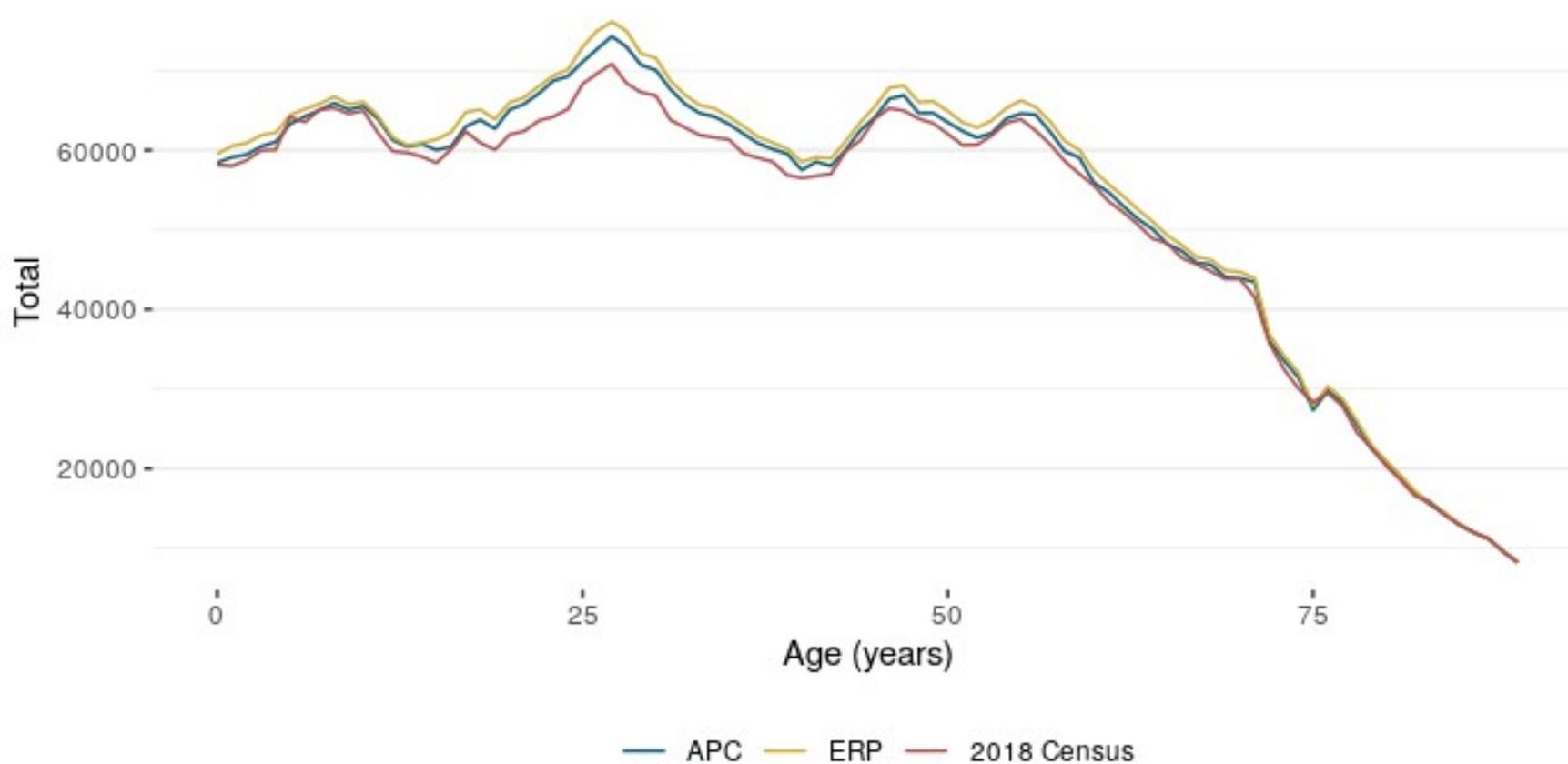
7.9 - 14

## Web App Link

<https://www.stats.govt.nz/experimental/experimental-administrative-population-census>

# What does the admin population look like?

NZ usual resident population,  
for APC and ERP at 30 June 2018 and 2018 Census



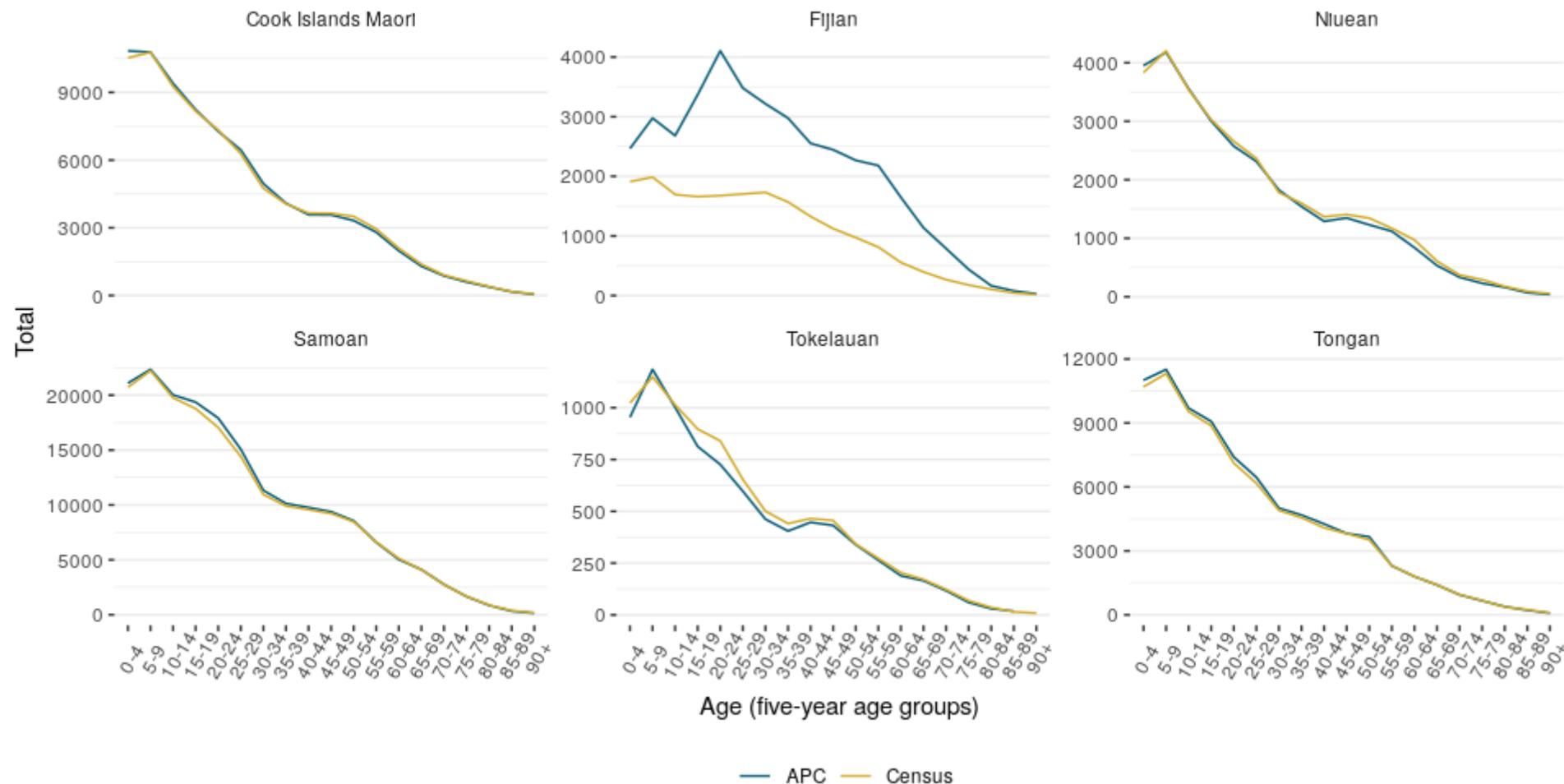
**APC is not quite ready to provide the official estimated resident population**

It needs a population estimation adjustment:

- for under- and over-coverage
- for misclassification and linkage errors
- for error bounds

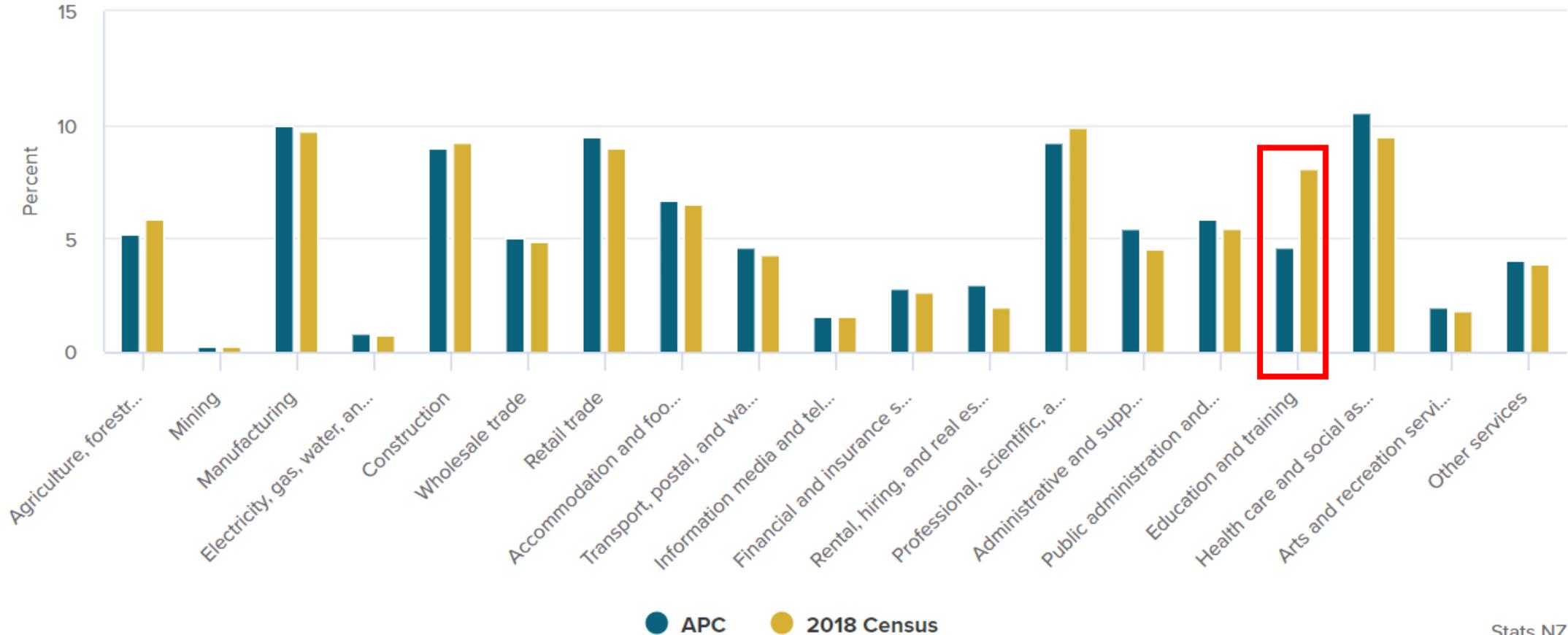
# Age distributions for some Pacific ethnicities, APC and 2018 Census

Counts of Pacific level 2/3 ethnicities, by five-year age groups  
comparison between APC in 2018 and 2018 Census



# What does the data look like?

Distribution of broad industry, for the APC in 2018 and 2018 Census



# Opportunities, that we can see

- Frequency + Detail: Outputs every year (or more often), with all variables available for all reference dates.  
Already a 16 year time series 2006 -2021
- Quality: More precise, more detailed information, or more accurate – not constrained by respondent’s ability to answer questions.
- Resilience: Inherently less exposed to extreme risks or system failure.  
For example, easier to recover from earthquakes, pandemics.
- Longitudinal: The unit record data is longitudinal – this is immensely powerful and largely untapped.  
For example, detailed analysis of flows, cohorts, generational effects.
- Synergies with other research: Potential to be a multiplier for other research.
  - Providing common baselines for researchers in the IDI.
  - **IF** iwi affiliation can be included, one could provide iwi-specific admin censuses.



# He pātai?



# Ngā mihi nui

# Individual form census questions in APC

**14 Where did you usually live 1 year ago, on 7 March 2022?**

not born 1 year ago

at my usual address, given in 5

at another New Zealand address:

street number    flat number

street name

suburb or rural locality

city, town or district

or

I lived overseas 1 year ago. Print the name of the country:

Very **rough** approximation, which does not take into account:

- importance of questions,
- questions that need to be asked on any survey for linking and operational purposes,
- range of the orange category, and
- some details within questions.

concept part of APC

concept potentially available from admin data

concept unlikely to be available from admin data

# How the ARP is created

For any reference date:

1. Start with the spine
2. Remove non-births, + deaths
3. Remove non-residents 12/16 and 6/12
4. Remove non-active units

“Revisions” between releases due to

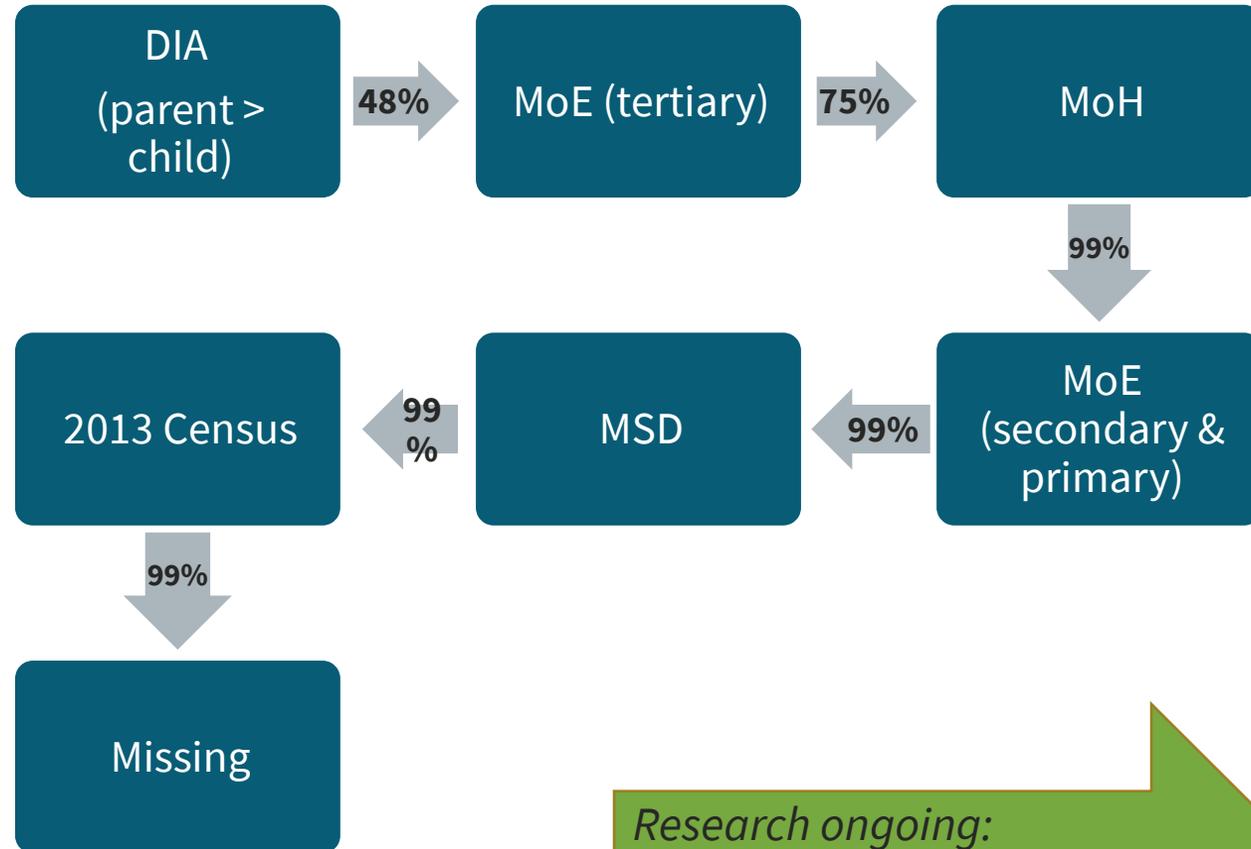
- Changes in the spine, linking, and historical data
- 12/16 availability for recent years
- Changes in methodology (experimental product)

# Quality

Variable	Missing (percent)	Non-missing quality rating	Output quality
Ethnic group level 1: European	< 1	0.96	0.96
Ethnic group level 1: Māori	< 1	0.94	0.94
Ethnic group level 1: Pacific Peoples	< 1	0.92	0.92
Ethnic group level 1: Asian	< 1	0.96	0.96
Ethnic group level 1: MELAA	< 1	0.79	0.79
Ethnic group level 1: Other	< 1	0.26	0.26
Māori descent	14	0.95	0.82
Birthplace	5	0.96	0.91
Years since arrival in NZ	4	0.93	0.89

# How ethnicity is derived

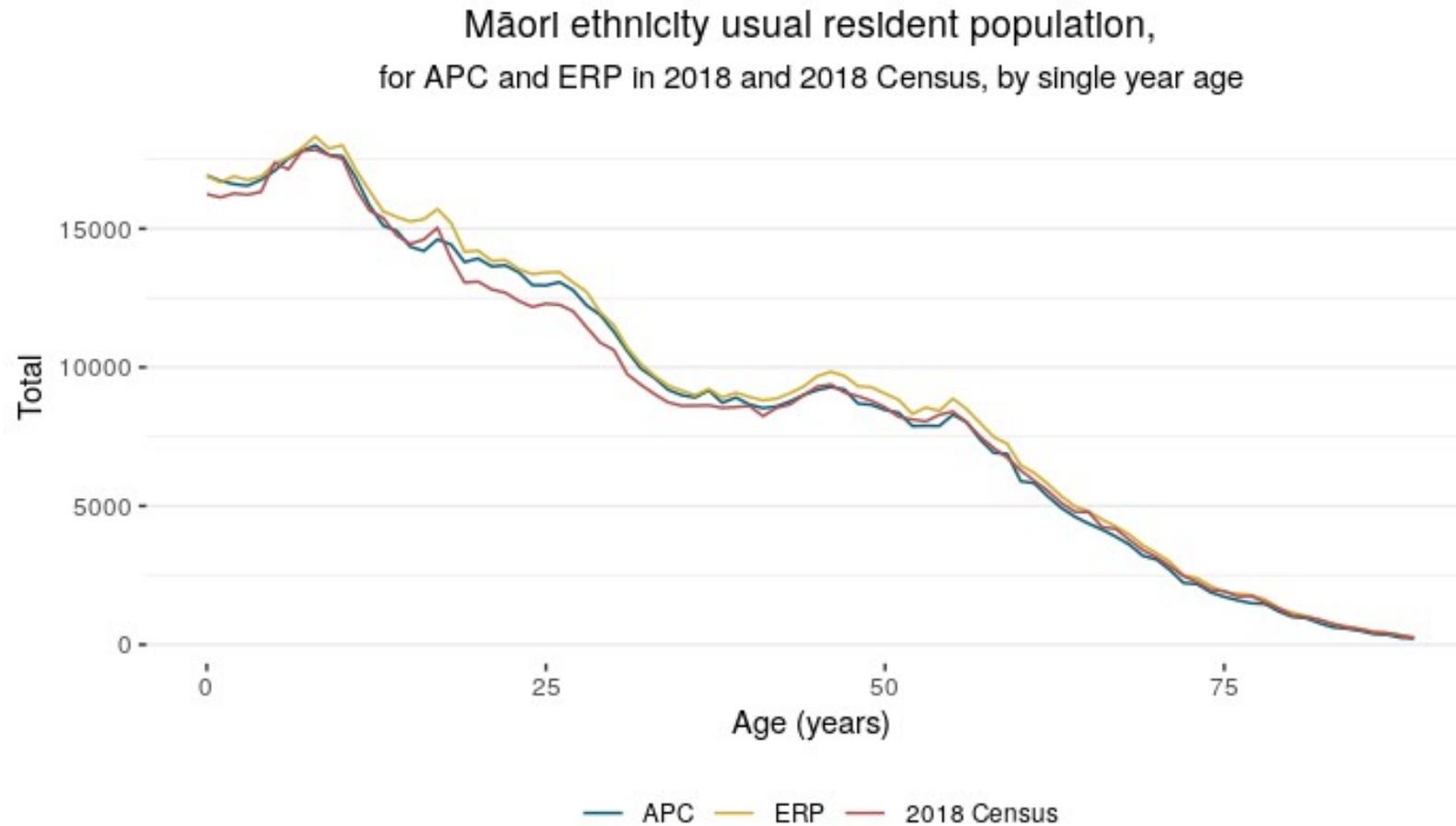
APC uses the source that is most similar to the census first (this is birth registrations).  
If there is no ethnicity from birth registrations, we use the next best source, if it is available.  
We get an (L2) ethnicity for nearly everyone (99 percent) from the top four sources.



Research ongoing:

- Refining method
- Latent class modelling

# What does the data look like?



Source: Stats NZ

**Sector of ownership for APC in 2018 and 2018 Census**

Sector of ownership	Count		Percent of total stated	
	2018 APC (30 June)	2018 Census (6 March)	2018 APC (30 June)	2018 Census (6 March)
Central Government	258,000	334,017	11.9	13.7
Local Government	49,296	37,230	2.3	1.5
Private	1,886,349	2,073,897	87.0	84.8
Total people stated	2,168,322	2,445,141	100.0	100.0
Missing	543,819	0		
Total people	2,712,141	2,445,141	...	...

**Note:** This data has been randomly rounded to protect confidentiality. Individual figures may not add up to totals, and values for the same data may vary in different tables.

**Symbol:** ... not applicable

**Source:** Stats NZ

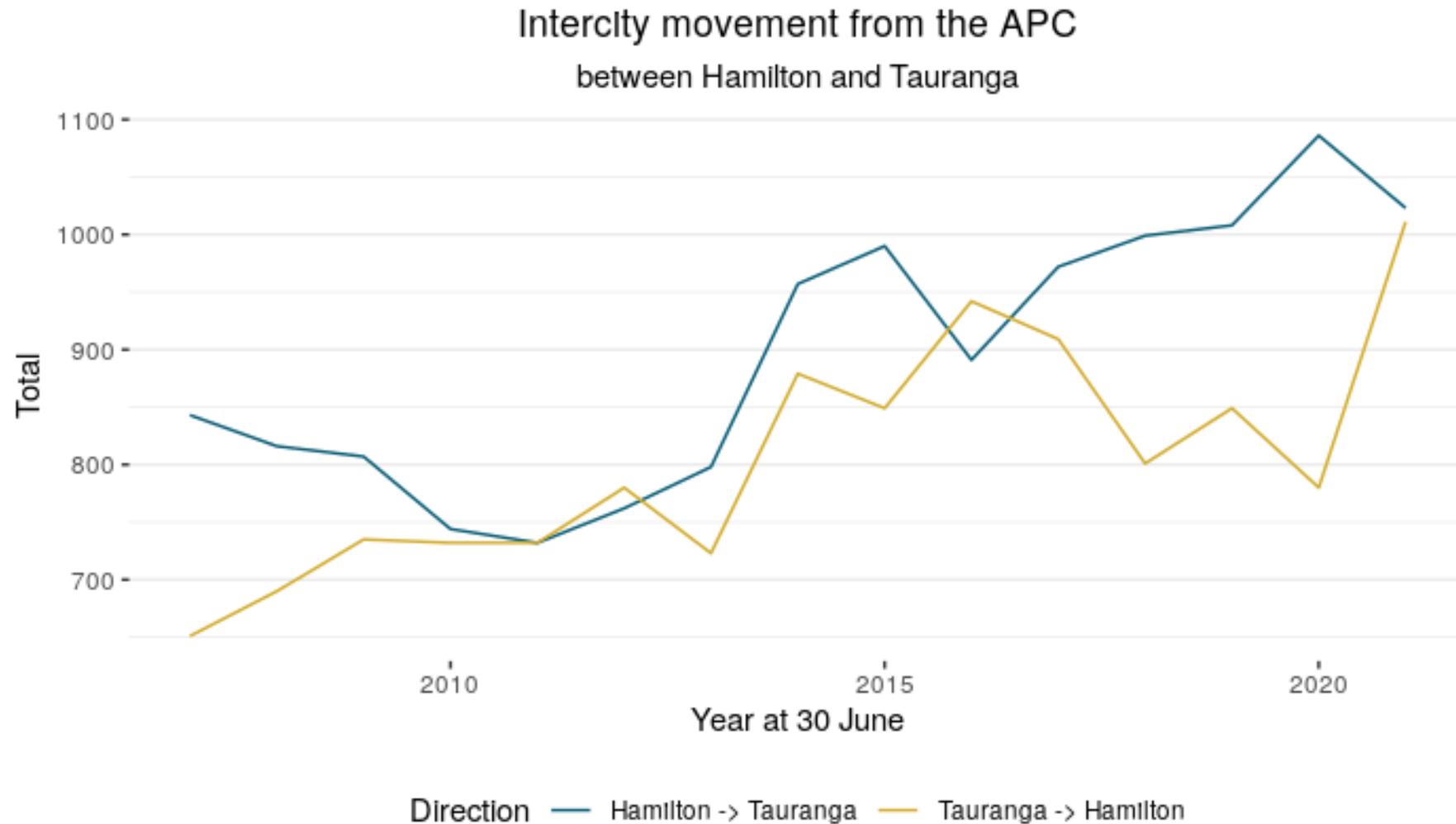
# Census concepts not well covered in admin data

- Activity limitations (disability) information that is not captured by the health system
- Iwi affiliation
- Gender and sexual orientation
- Languages spoken
- Religious affiliation
- Some work and labour force information ( ‘unemployed/nilf’, occupation, unpaid work)

**Some** strategies to mitigate these limitations:

- Improved agency collections, and new data sources
- Improved methods and statistical mitigations
- Combining APC data with hhld surveys or a census attribute survey

# Longitudinal data





# Scoping ideas for the third iteration

## Other R&D projects

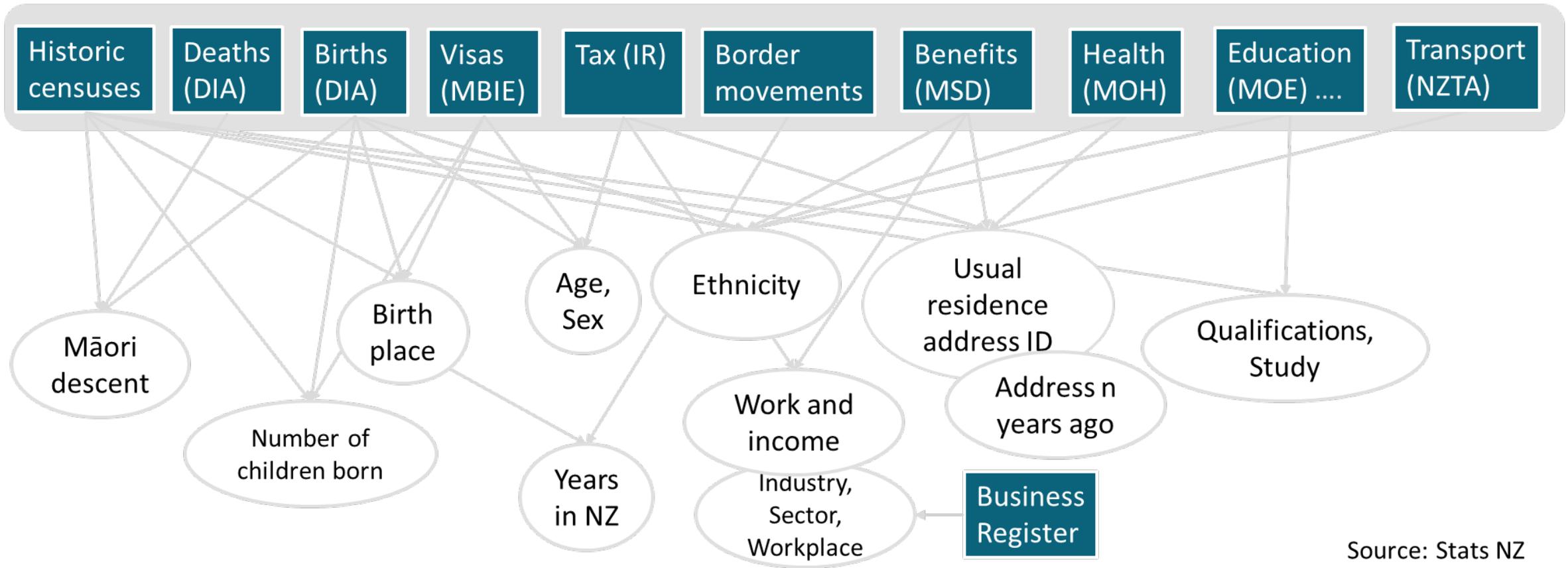
- HHLDs and families from admin data
- Dwellings from admin data
- Iwi affiliation from admin data
- Undercoverage adjustments (DSE)
- Design of a census attribute survey

## Planned scope

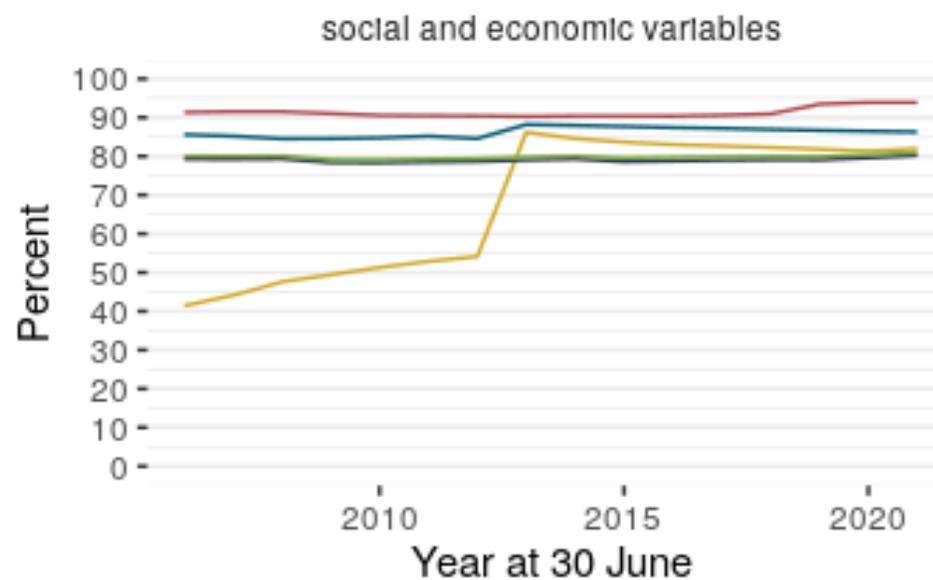
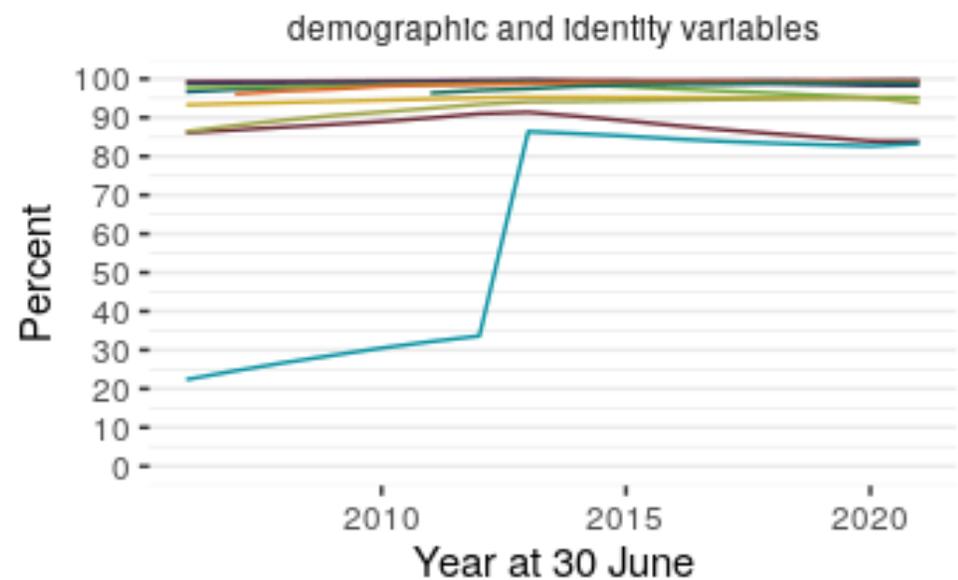
- More variables and refinements
  - Current research: relationship status, family relationships, years at usual residence
  - Non-census variables: Mother's age at first birth, and years since highest degree attained, Income by income source
  - Activity, ethnicity, and address review
- Statistical Imputation
- More timely data, extended time period
- More detailed public releases
  - Flows data
  - Outputs by L2-ethnicity
  - Crosstabs, non-standard geographies
  - (Even) more information on quality
- Broader engagement with data users/treaty partners/stakeholders

# Where does the data come from?

## All the data sources are linked in the IDI



### Coverage for each APC variable, 2006 to 2021



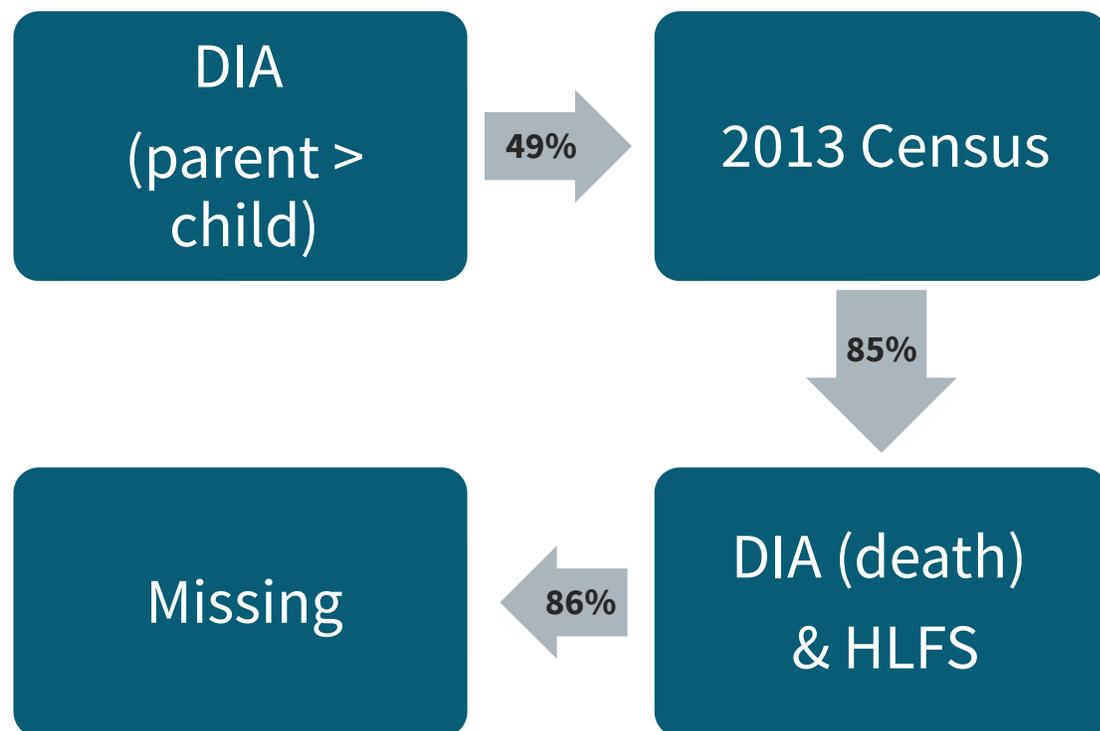
- Address (mb)
- Birthplace
- Ethnicity (L1/2)
- Ethnicity (L3)
- Ethnicity (L4)
- Maori descent
- Number of children born
- UR 1 years ago
- UR 5 years ago
- YsAN
- Field of study
- Highest qualification
- Income
- Industry
- Sector of ownership

# How Māori descent is derived

Department of Internal Affairs birth registrations is used first. (The question in its current form has been asked since 1995.)

The 2013 Census: is used for people who are not in the birth registrations data.

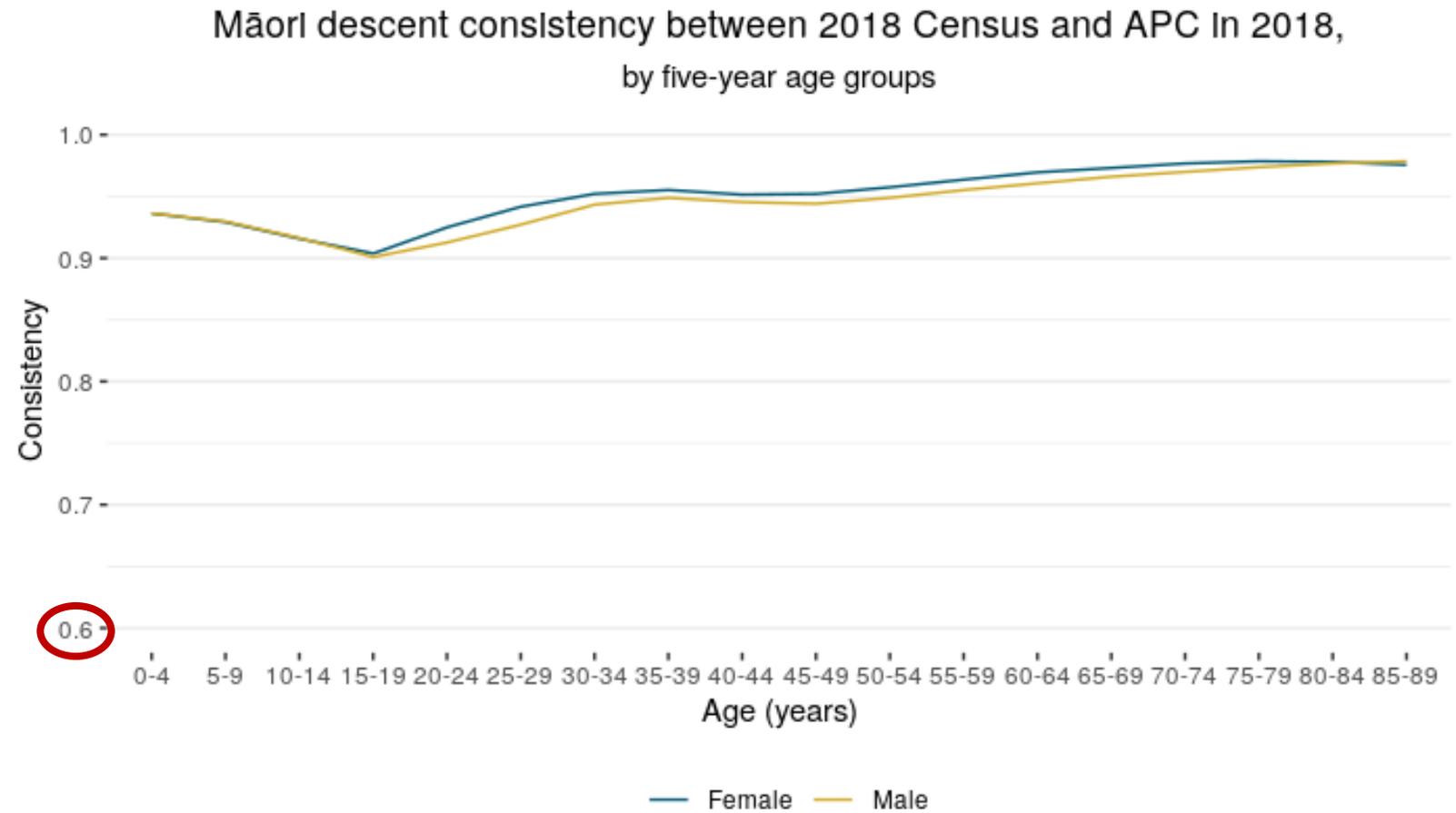
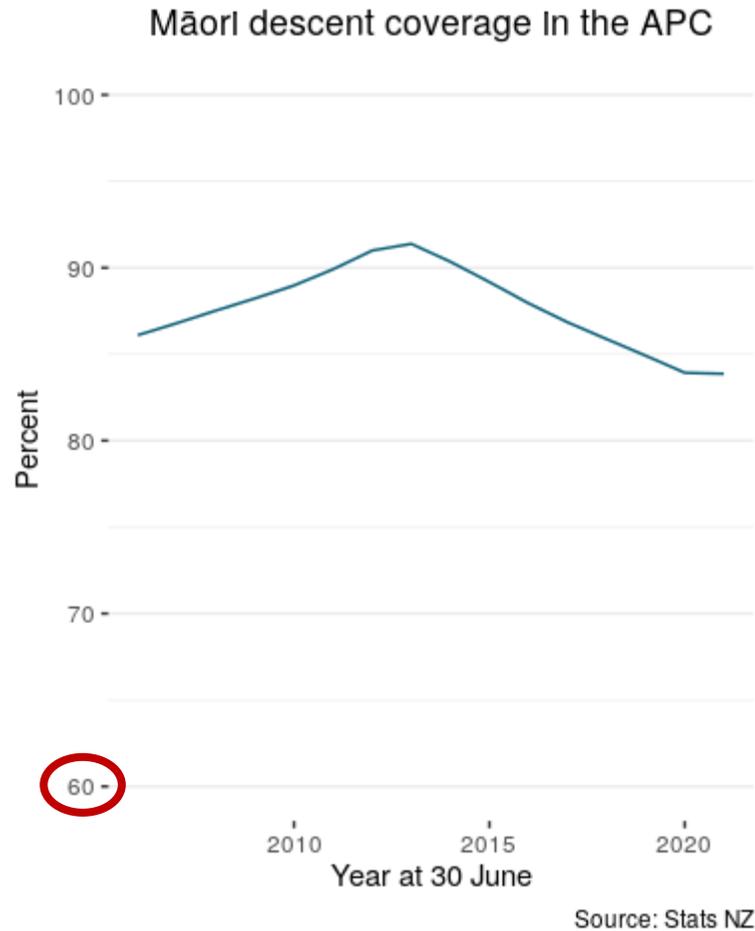
Māori descent from DIA death registrations and the Household Labour Force Survey fills in information for a very small number of people who did not have Māori descent in the other sources.



*Research ongoing:*

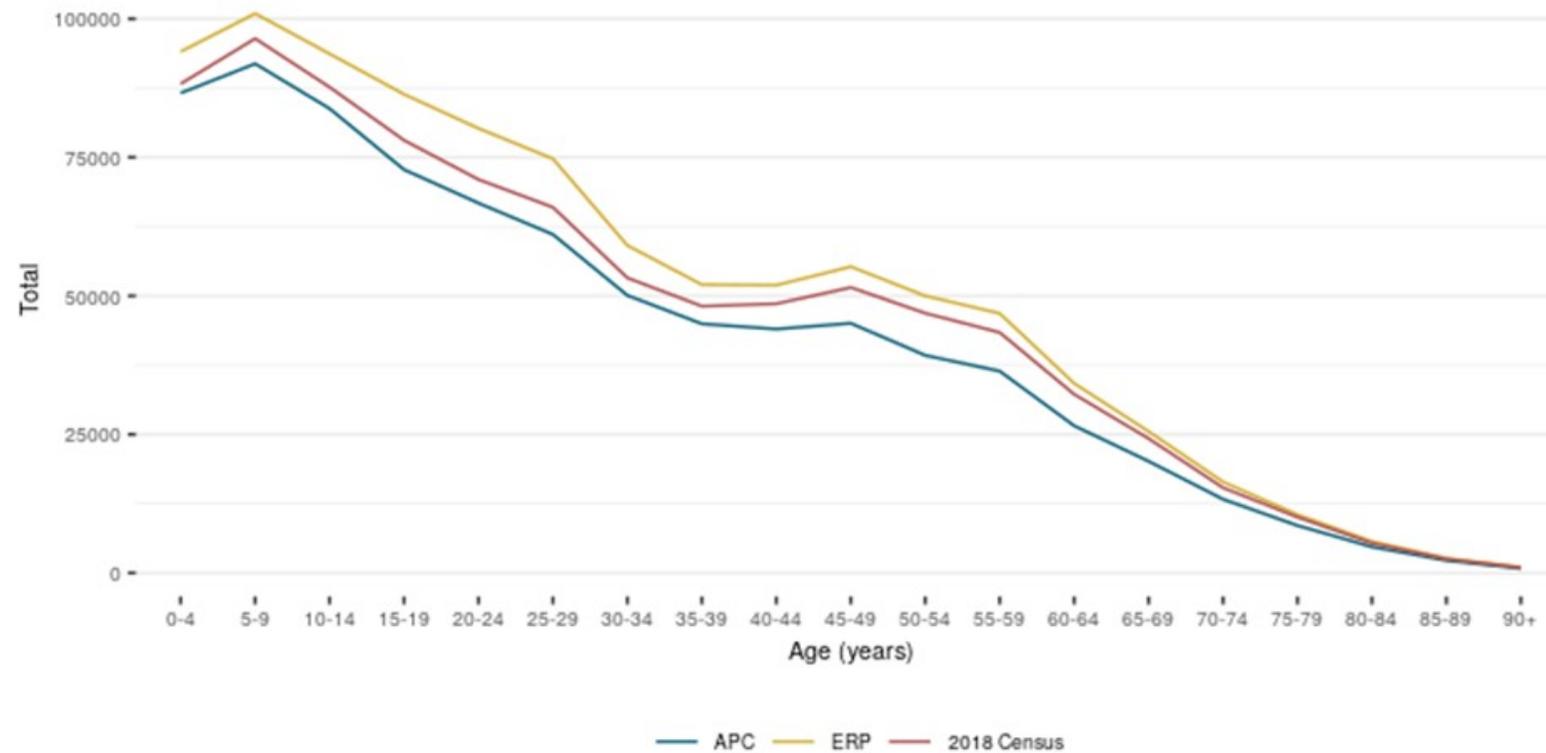
- Potential of including electoral roll data

# Quality of Māori descent data



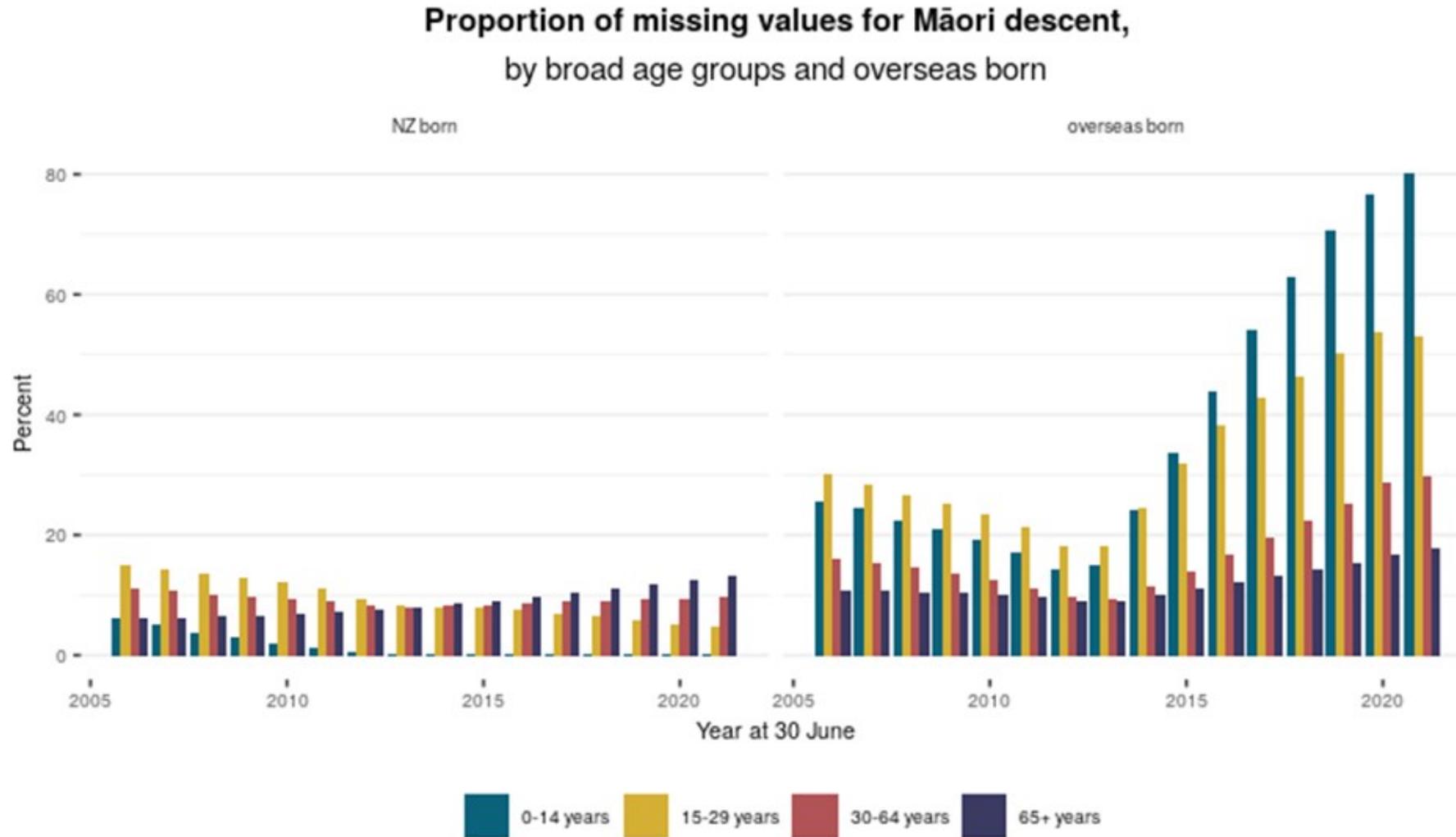
# Quality: Māori descent

**Māori descent usual resident population,**  
for APC and ERP in 2018 and 2018 Census, by five-year age groups



Source: Stats NZ

# Quality: Māori descent



# More frequent measurements

Monthly admin usual resident population,  
for the APC Jan 2006-Dec 2020



# Dempster Shafer theory



Compare selected value, with available values in sources (not necessarily the same sources as used.)

- The more sources, the higher the score
- The more agreement, the higher the score
- The more sources disagreeing, the lower the score

These values are added in the unit record data for:

- Māori descent
- Birthplace
- Years since arrival in NZ
- Level 1 ethnicities (all 6)

Scores for any breakdown can be produced simply by averaging the individual scores!

# Dempster Shafer theory

Snz_uid	APC	Source 1	Source 2	Source 3	APC QM↓
12345	A	A	A	A	.95
23456	B	-	B	-	.9
34567	C	C	C	A	.7
45678	A	B	-	A	.4
56789	B	A	B	C	.3
...					