

Stats NZ – 2023 Census

Combined Census ‘by design’

Rochelle Morgan, Senior Manager – Census Methodology

Methods & Design

Station 10



The Stats NZ Combined Census 'by design'

- Introduction to the combined census model
- 2023 Combined Census 'by design'
- Why we use administrative data
- New Zealand's source of administration data
- Combined Census approach and decision framework
- Use of admin data for units (admin enumeration)
- Alternative sources for census attributes
- Coverage and quality implications for response rates

The Combined Model

‘Information on individuals and households is collected by combining data collected from one or more surveys or full field enumeration with administrative or statistical registers’ [\(UN, 2017\)](#)

In our context, information comes from:

- Full enumeration census
- Administrative data – primarily via the Integrated Data Infrastructure (IDI)

Key changes from 2018 Census

- 2018 Census was a combined census implemented ‘on the fly’ (in recovery) to mitigate low response
- 2023 Census is combined model and use of alternative data ‘by design’
- More research, testing and analysis to improve:
 - quality of admin variables
 - quality of admin enumerations
 - detailed methods and modelling
- Framework for admin enumerations
- Sensitivity testing for different response rates

The 2023 Census Model

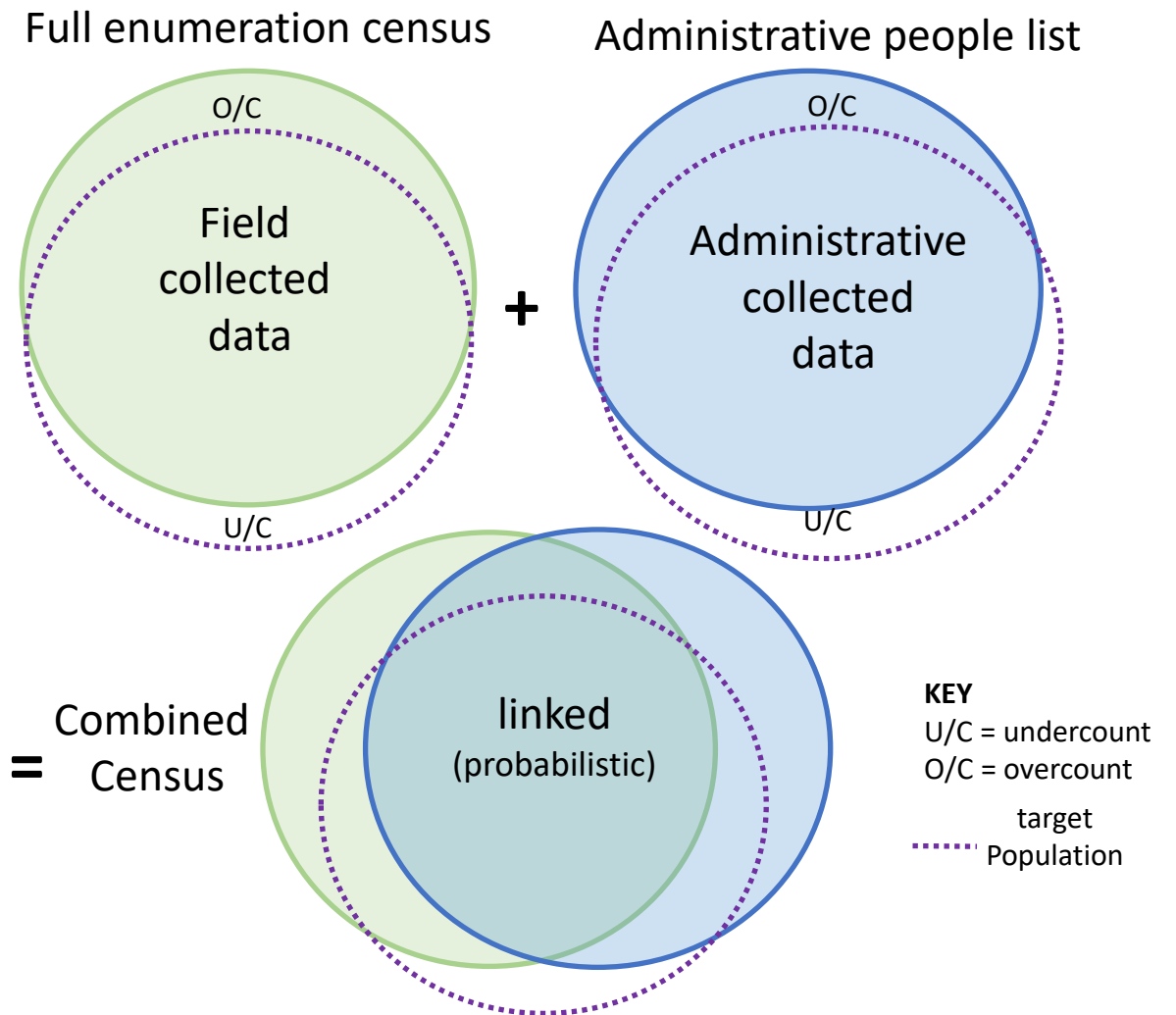
Combined Census ‘by design’

High level design – scope statement:

“Administrative and alternative data sources will be used to make improvements to existing data processes and output quality, to reduce or resolve known issues in processing, and to support the processing of data under a combined model approach.

The specific uses of administrative and alternative data sources for units and characteristics will be determined by the ‘combined model’ design.

Administrative and alternative data sources will also be used in the development of interim population estimates to support analysis and quality assurance.”



Why use admin data?

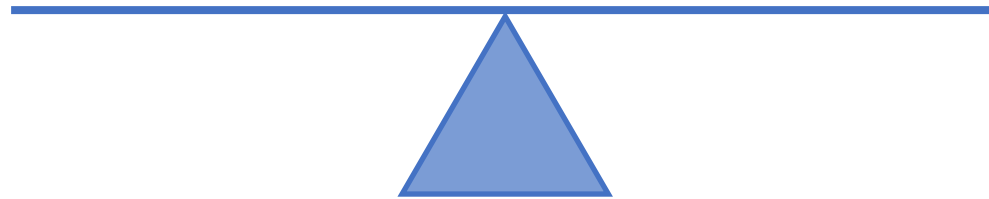
Leverage the strengths of admin data to get the best count of people and dwellings

- overcome or 'design around' known limitations: coverage issues, differences in population structure, location accuracy for small geographies, forming households
- Add missed people (non-respondents) to the census file and fill in missing information to improve overall value of census outputs.

Two main uses of admin to fill data gaps:

- Missing people – adding administrative records to the Census dataset (unit non-response)
- Missing characteristics – using information from multiple sources (item non-response)

Right to
privacy

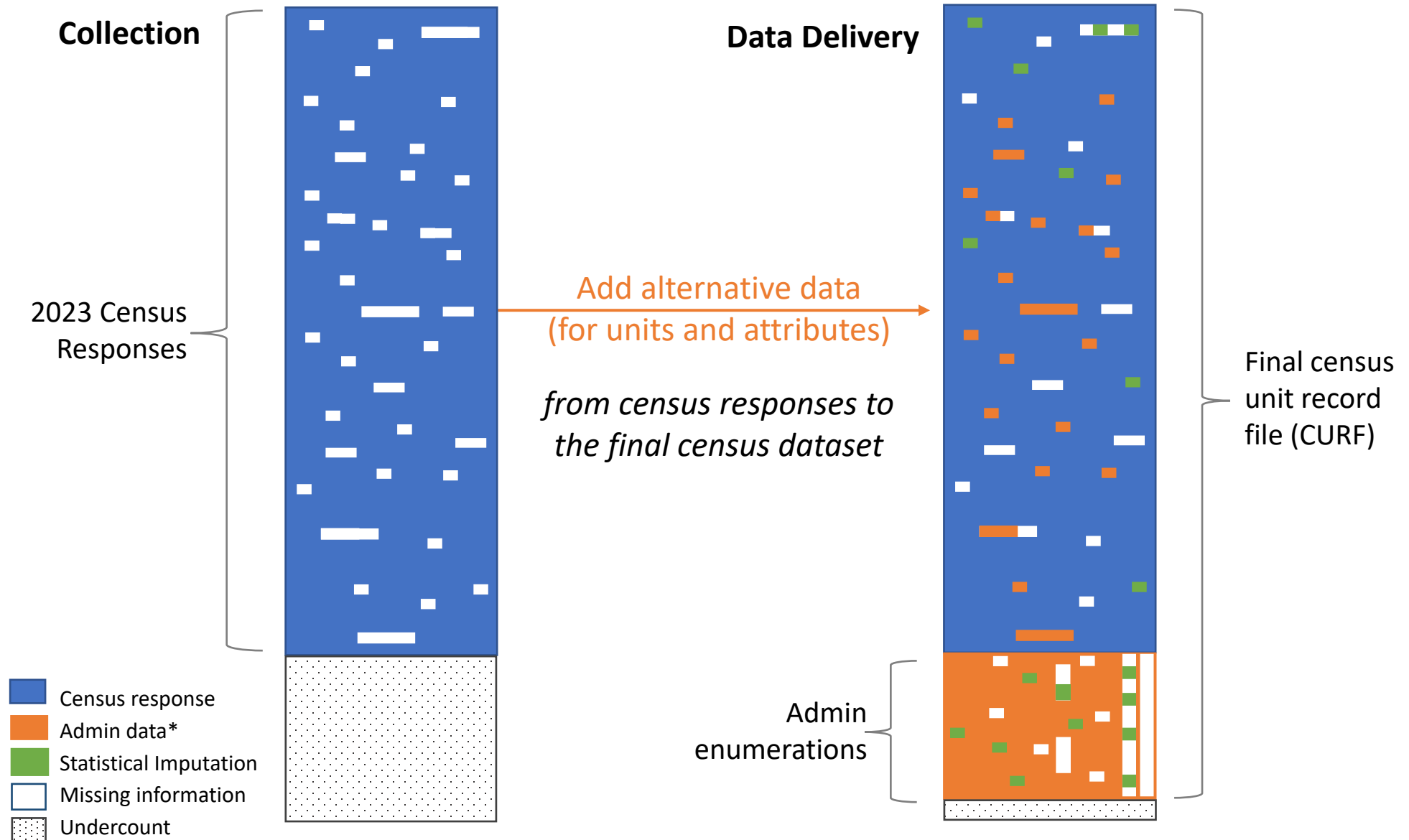


Right to live in an
informed society

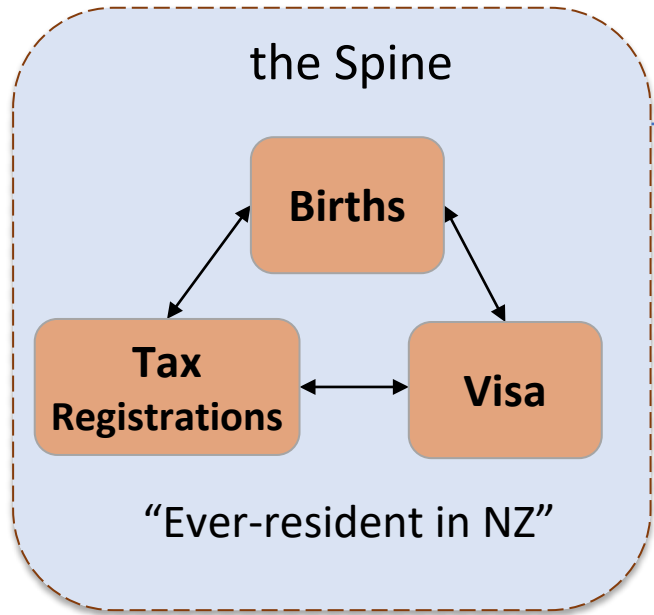
Balancing privacy and benefits

From census responses to the final census dataset

The combined census model 'by design' combines an admin population list, equivalent to the [Admin Population Census](#) (APC), with census collected response data to achieve high coverage and quality, including rich attributes information.

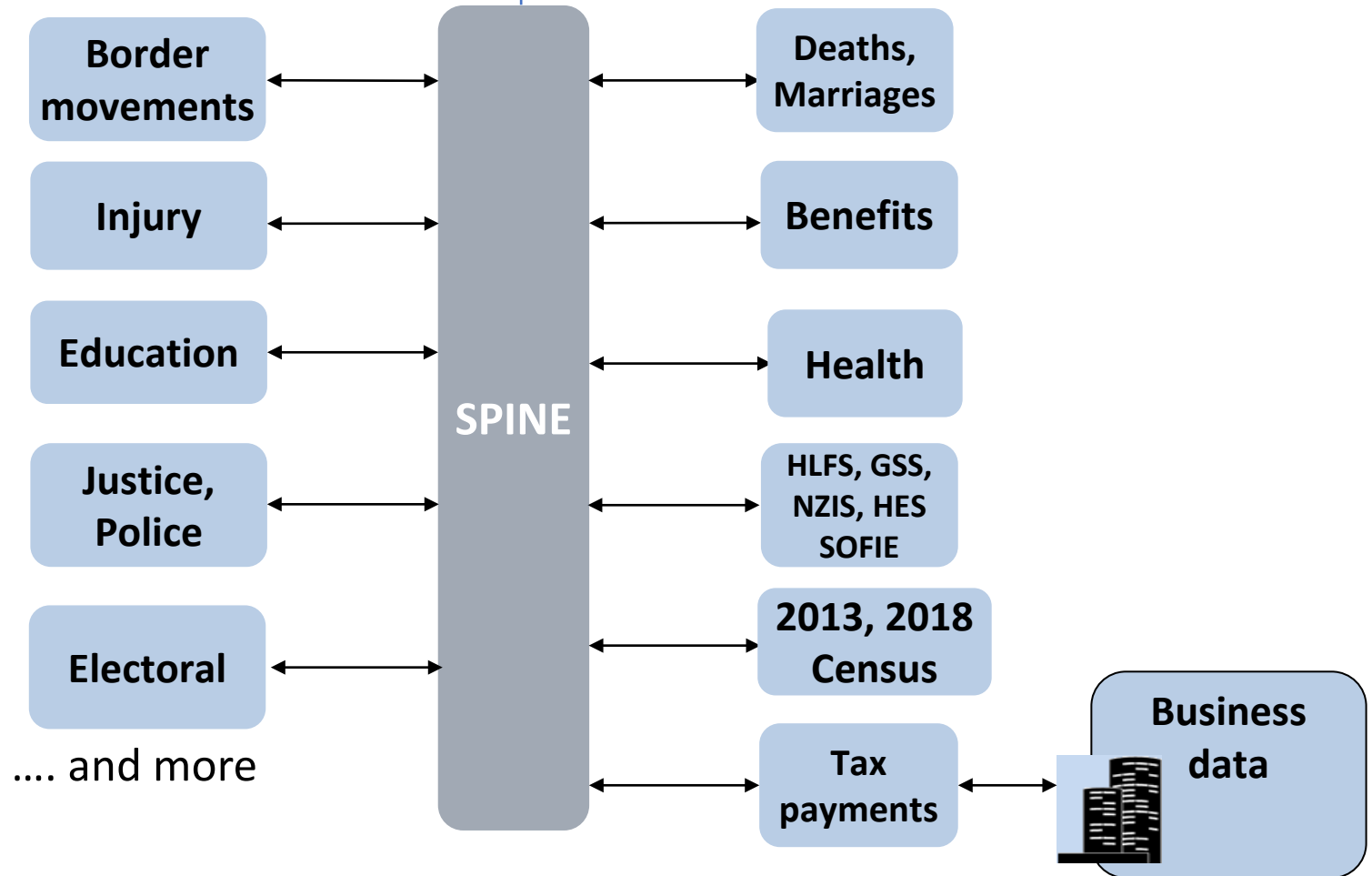


The Integrated Data Infrastructure (IDI)



Administrative data is sourced from Stats NZ IDI

Around 10m individuals 'ever resident in NZ'



Combined Census Approach & Decision Framework

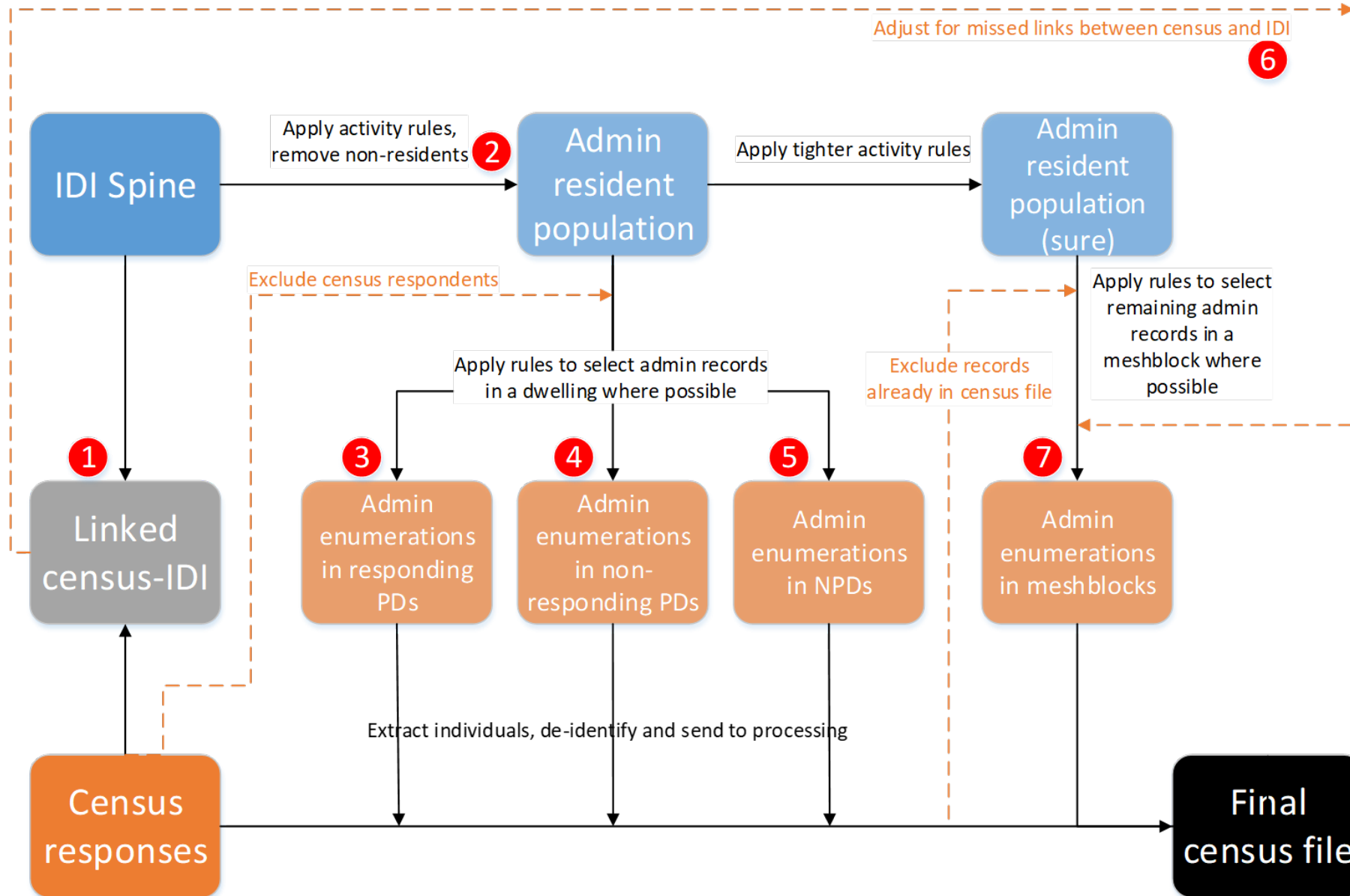
Approach:

- Prioritise including admin enumerations in dwellings where possible.
- Add records where we are confident in their quality and that they will improve census outputs

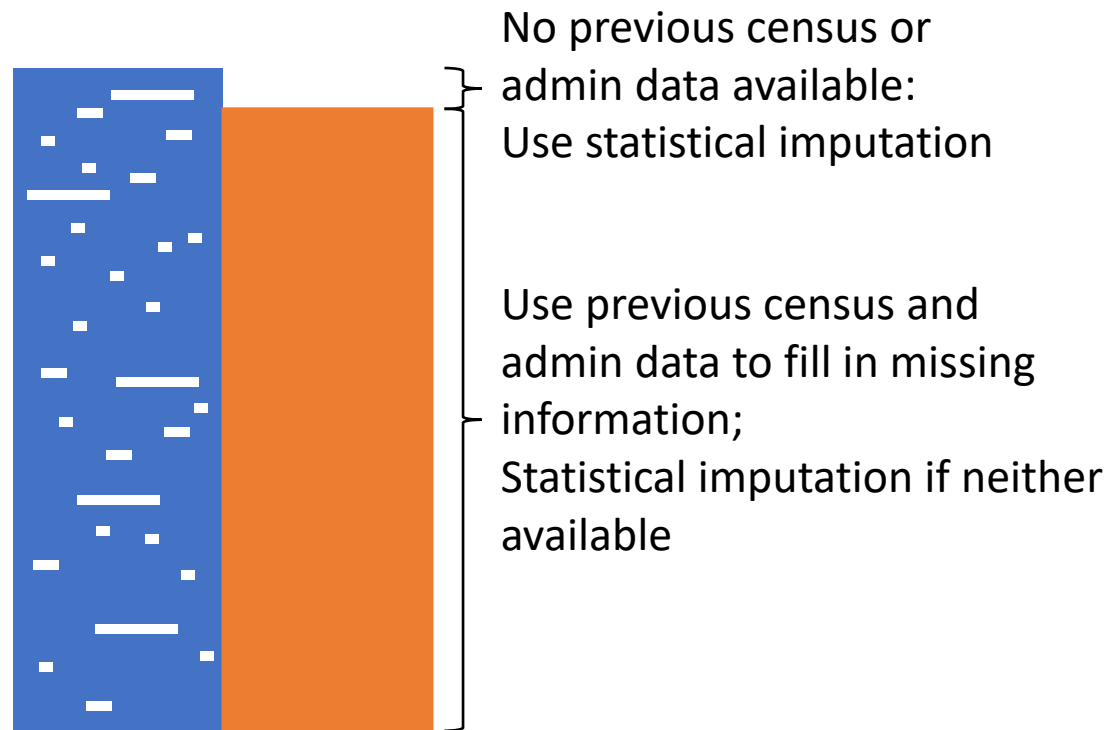
Decision Framework:

- Thresholds are set to determine which admin records are 'high-quality' and therefore able to be included in the census file
- These thresholds are flexible to different response scenarios. Specific criteria could be loosened if the response rate is lower than desired:
 - This would improve coverage and representativeness for affected areas
 - But may lead to some lower accuracy in individual records at a detailed level
- Guiding principals:
 - 2023 Census responses take priority over admin records
 - including admin enumerations in a dwelling is more valuable than including admin enumerations in a meshblock
 - including admin enumerations in a meshblock is more valuable than not including admin enumerations at all
 - including admin enumerations should reduce, rather than increase, biases in the census dataset
 - some error at the individual level is acceptable if overall distributions are improved

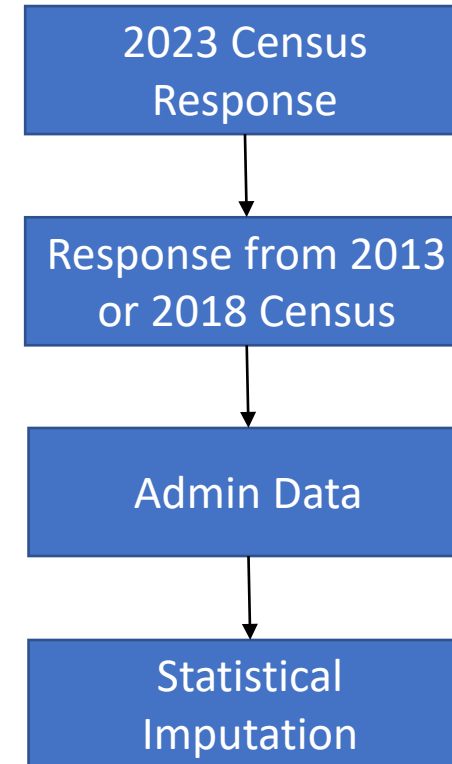
Admin enumerations process



Use alternative data to fill in missing information



Hierarchy of sources for filling in a variable:



NOTE: Not all data sources will be used for all variables. Details will vary based on availability and appropriateness.

Use alternative data to fill in missing information

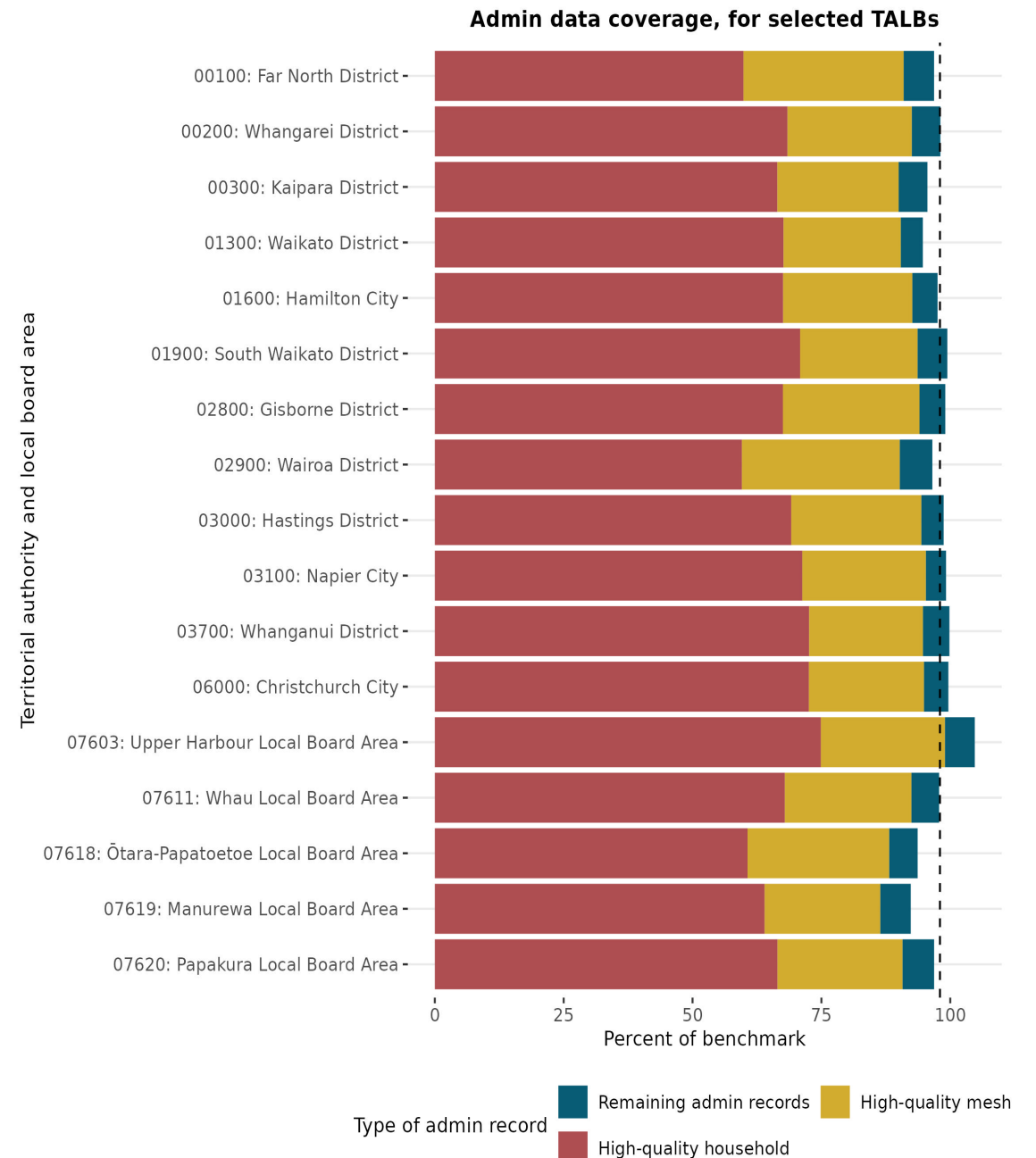
Topic	Census attribute	Use of data source			
		2023 Census response	Historical census response	Admin data	Statistical Imputation
Population structure	Date of birth/Age	Y	Y	Y	Y
Population structure	Sex at birth	Y	?	?	Y
Population structure	Gender	Y	?	?	Y
Location	Usual residence address	Y	N	Y	Y
Location	Usual residence one year ago	Y	N	Y	N
Location	Usual residence five years ago	N	Y	Y	N
Location	Years at usual residence	Y	Y	Y	N
Population structure	Number of children born	Y	Y	Y	N
Ethnicity and culture	Ethnicity	Y	Y	Y	Y
Ethnicity and culture	Sexual identity	Y	N	N	N
Ethnicity and culture	Māori descent	Y	Y	Y	?
Ethnicity and culture	Iwi affiliation	Y	Y	?	?
Ethnicity and culture	Birthplace	Y	Y	Y	N
Ethnicity and culture	Years since arrival in New Zealand	Y	Y	Y	N
Ethnicity and culture	Languages spoken	Y	Y	N	Y
Ethnicity and culture	Religious affiliation	Y	Y	N	Y
Education and training	Highest secondary school qualification	Y	Y	Y	N
Education and training	Level of post-school qualification	Y	Y	Y	N
Education and training	Post-school field of study	Y	Y	Y	N
Education and training	Study participation	Y	N	Y	Y
Transport	Main means of travel to education	Y	N	N	Y
Income	Total personal income	Y	N	Y	Y
Income	Sources of personal income	Y	N	Y	Y
Work	Industry	Y	N	Y	Y
Work	Occupation	Y	N	N	Y
Work	Sector of ownership	Y	N	Y	Y
Work	Status in employment	Y	N	Y	Y
Work	Unpaid activities	Y	N	N	N
Transport	Main means of travel to work	Y	N	N	Y
Health and disability	Smoking status	Y	Y	N	Y
Health and disability	Disability / activity limitations	Y	N	N	N
Population structure	Relationship status - legally registered	Y	Y	Y	N
Population structure	Relationship status - partnership status in current relationship	Y	Y	Y	N

Implications of response on Coverage – example only

Scenario based on no census response for an area

- Most areas have high coverage in admin data already, although there can be some variation across areas, and across different sub-populations (see the graph on the right for a selection of areas).
- However, not all admin records would meet our criteria of “high quality”. Therefore, not all records will be added as admin enumerations under existing methods
- Without any adjustments, this could lead to some areas or populations having lower coverage than others. But this can be mitigated through the use of additional data sources and/or different quality thresholds

**NOTE: these coverage estimates intended to be indicative only
– data is made up**



Implication of response rate on data quality – example only

Data quality for a variable depends on both the response rate and the quality of alternative data for the variable

For example:

Age has very reliable alternative data

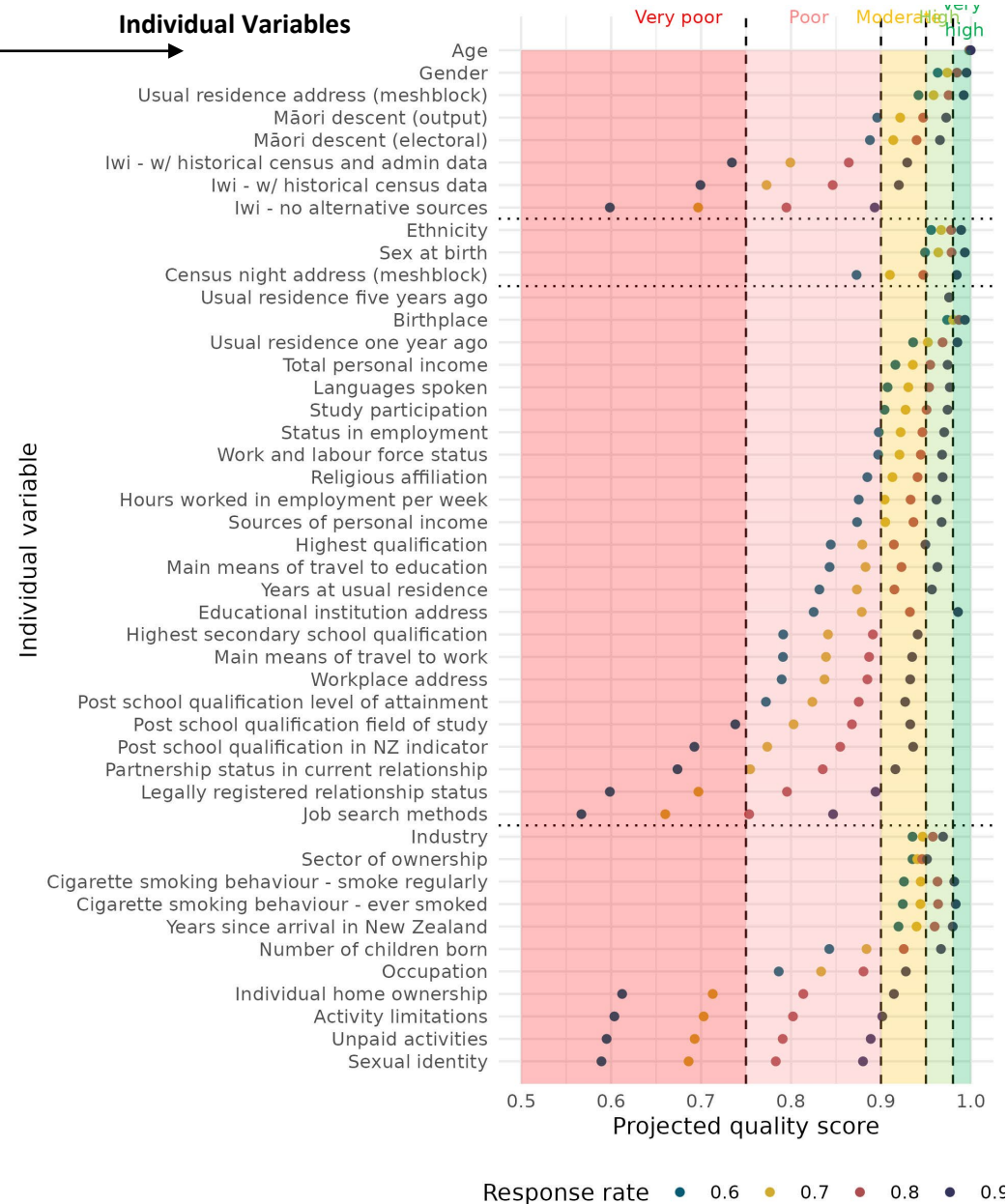
- Quality likely to be **Very High** regardless of response rate

Iwi variables have less reliable alternative data

- Quality may be **Poor** to **Moderate** at a 90 percent response rate

Note: data quality projections are indicative only – data is made up

EXAMPLE ONLY Data Quality (Metric 1) Projections



Tatau tātou - All of us count

Ngā mihi
Thank you

