

## Three Population Milestones: Some Comments and Cautions

RICHARD BEDFORD\*

### Abstract

This research note reviews the work of an independent panel of experts in the production and use of population statistics that was set up by Stats NZ in August 2018 to provide the Government Statistician with advice about methodological and substantive issues relating to the quality of the 2018 Census data. This advice was sought because of a much lower than anticipated participation by the public in the census enumeration in March 2018. The release of a considerable amount of 2018 Census data between April and July 2020 provides a useful opportunity to reflect on aspects of the panel's findings with regard to methodology and data quality.

Three critical issues relating to 2018 Census data surfaced quickly in the panel's work with statisticians from Stats NZ: 1) variability amongst subgroups in the population and across areas in participation in the enumeration, 2) issues related to the social licence for use of a wide range of administrative data to fill gaps in census responses, and 3) the challenge of comparing 2018 Census data with data from the 2013 and 2006 censuses. These issues are discussed at length in the External Data Quality Panel's (EDQP) reports but some comments on and cautions relating to each one are provided in this research note. The primary objective is to indicate where important information for understanding strengths and limitations of 2018 Census data can be found in reports produced by Stats NZ and the EDQP.

**Keywords:** 2018 Census; data quality; census enumeration; social licence; ethnicity; Treaty obligations

Three important developments relating to New Zealand's population have occurred during May 2020 as the country transitions out of one of the world's most restrictive lockdown responses to the coronavirus (COVID-19) pandemic. The first is the passing of a population threshold by the end of March 2020 that has featured in a range of ways in projections

---

\* Richard Bedford is a research associate at the National Institute of Demographic and Economic Analysis. He is also an emeritus professor at the University of Waikato and Auckland University of Technology.  
Email: [richard.bedford@waikato.ac.nz](mailto:richard.bedford@waikato.ac.nz)

for New Zealand's population for at least 50 years – five million usual residents. The second is the release, or impending release, of a great deal of the data relating to individuals, families, households and dwellings from the latest Census of Population and Dwellings just over two years after the population was enumerated in March 2018. The third is Stats NZ's announcement that, through the Government's May 2020 budget, it currently has NZ\$210 million (exclusive of GST) available for the preparation and conduct of the 2023 Census of Population and Dwellings.

This research note comments briefly on these three milestones with particular reference to the work of the 2018 Census External Data Quality Panel (EDQP) that Stats NZ established in August 2018. This panel was set up to provide the Government Statistician with advice about methodological and substantive issues relating to the quality of the 2018 Census data in the light of a much lower than anticipated participation by the public in the census enumeration in March 2018. The release of a considerable amount of 2018 Census data between April and July 2020 provides a useful opportunity to reflect on aspects of the panel's findings with regard to methodology and data quality. The Government's decision to fund a slightly enhanced Stats NZ 'minimum cost' option for the budget for the 2023 Census (not Stats NZ's preferred option) has major implications for the extent to which the Department can deliver on its Treaty obligations to Māori – an area where it failed badly in the 2018 Census.

## **A demographic milestone, March 2020**

At 31 March 2020, the resident population of Aotearoa New Zealand is estimated to have reached 5,002,100 (Stats NZ, 2020a). The country joins some of its frequent comparators in the 5+ million club – Norway (5.4 million), Finland (5.5 million), Singapore and Denmark (5.8 million each). The very high net migration gains during the six years ending March 31 between 2015 and 2020 (all in the 49,000–71,000 range) have been exceptional in numerical terms in Aotearoa New Zealand's history. They have accelerated the achievement of the 5 million milestone and contributed to the long-standing decline in fertility rates in the total population.

Stats NZ (2020a, p. 2) cautions that the net migration estimates for the year ended March 2020 (71,500) have been impacted in several ways by the international border closures linked with the COVID-19 pandemic. The

provisional estimate for the resident population at 31 March is likely to be adjusted once visitors, international students, temporary workers and New Zealand citizens, who have had their arrivals and departures disrupted by border closures, are able to make their international moves again. The 71,500 net migration gain for the year ending March 2020 is likely to be adjusted later in the year, but it is not expected that this adjustment will mean the 5 million milestone is not achieved around March 2020.

The population estimate of 5,002,100 is based on an estimated resident population (ERP) that was derived from the population enumerated in the Census of Population and Dwellings in March 2013. The ERP based on the resident population that has been determined for March 2018, the month when the last census enumeration was held, has yet to be derived. The results of the Post-Enumeration Survey (PES) following the 2018 Census have not been finalised and, until they are, an ERP based on the results of this census, cannot be established. The 2018 Census resident population has been set at 4,699,755 on the basis of a complex set of calculations and adjustments following a much less complete enumeration of the total population than in previous censuses (Stats NZ, 2019a).

## **A 2018 Census data milestone, March–June 2020**

The 5 million milestone followed news on 1 May about the impending release of a significant amount of detailed data from the 2018 Census (Stats NZ, 2020b). *Customer update on 2018 Census data and activities – May 2020* is an important document for those who have been using, or have plans to use, 2018 Census data. It contains:

- 1) an update on the 2018 Census variables, including revisions to earlier assessments of the quality of the data for a number of these variables
- 2) an update on the 2018 Census data releases and products
- 3) a summary of upcoming data and product releases, and
- 4) a summary of recent and forthcoming activities linked with the release of 2018 Census data.

Including in the latter summary is information on data quality training events (disrupted by COVID-19 restrictions) and details of several publications in Stats NZ's suite of technical methods papers that will be published between the end of April and the end of June 2020. Included in

the technical papers is an important one addressing revisions to Stats NZ's initial quality assessment of the variables relating to families and households in the 2018 Census (Stats NZ, 2020c). This was a particularly challenging suite of variables to adjust for missing data given the fact that alternative sources of comparable data on family and household structures are not readily available to fill gaps in the census enumeration. These data are revisited briefly later in this research note.

With the resident population now around five million, and a significant amount of detailed 2018 Census data about to be released, there is likely to be considerable interest amongst researchers and policy analysts in using census data to examine characteristics of the residents in different parts of the country and changes in these populations between 2013 and 2018. In this context it is timely to draw attention again to the extensive work that a mix of statisticians, academics and local government users of census data undertook between August 2018 and January 2020, at the request of Stats NZ, on the quality of the 2018 Census data.

This research note provides a selective outline of some of the work done by the 2018 Census External Data Quality Panel (EDQP) that Stats NZ established to provide advice to the Government Statistician on a range of issues linked with the production of the 2018 Census data file.<sup>1</sup> The operational problems that resulted in a much lower than expected participation by sections of the country's population in March 2018 have been dealt with elsewhere (Jack & Grazidei, 2019). In the sections that follow, the focus is on the quality of the 2018 Census data and a range of considerations that emerged from 18 months of engagement by the EDQP with a team of senior statisticians and data specialists in Stats NZ.

## **A very different census data file**

All users of 2018 Census data will be aware by now that the information they are working with comes from a range of sources, not just the census enumeration in March 2018. The data file containing the 2018 Census variables comprises a mix of information from four main sources: the 2018 Census, the 2013 Census, a range of administrative data sources many of which can be accessed via Stats NZ's integrated data infrastructure (the IDI), and from statistical imputation (Stats NZ, 2019a, 2019b; EDQP,

2019a). The relative importance of these sources varies considerably by variable as well as by area.

This is a very different data file from those that contain the results of earlier censuses. While there has been some limited imputation of missing information in earlier censuses, as well as the use of empty dummy variables, it has never been Stats NZ's practice to fill gaps using data from the previous censuses or from administrative sources. The 2018 Census data files are therefore not directly comparable in terms of their sources, or their coverage of groups within the population, with the numbers from previous census data files.

The first major task that confronted the EDQP was getting a good understanding of what was involved in constructing a data file for the 2018 Census. Three critical issues surfaced quickly in this context: 1) variability amongst subgroups in the population and across areas in participation in the enumeration, 2) issues related to the social licence for use of a wide range of administrative data to fill gaps in census responses, and 3) the challenge of comparing 2018 Census data with data from the 2013 and 2006 censuses. These issues are discussed at length in the EDQP's reports but some comments on and cautions relating to each one are provided here with the primary objective of indicating where important information for understanding strengths and limitations of 2018 Census data can be found in reports produced by Stats NZ and the EDQP.

## **Variable participation in the census – quality considerations**

There was considerable variation in participation in the 2018 Census enumeration by members of different ethnic groups, amongst younger adults and across different regions. There is always some variability in participation in the census but the nature and extent of variations in the 2018 Census are of a different order of magnitude to those found in earlier censuses. There are several reasons for this and a major one was operational failures in the follow-up with people who had not completed the census online. This matter is discussed at length in an independent assessment of the organisation and delivery of the 2018 Census by Jack and Grazidei (2019). Len Cook (2020), in his article on lessons for the 2023 Census in this issue of the journal, also touches briefly on operational issues with the 2018 Census enumeration.

The variability in participation in the 2018 Census enumeration is something that users must keep in mind at all times when using the published census data. It affects every variable in different ways, and when combinations of variables are used simultaneously (for example, age, sex, ethnicity, occupation, income), some understanding of the mix of sources of information that have been used to generate the data for each variable is required.

### *Assessing data quality for variables with information from several sources*

A valuable interactive web-based facility has been developed by the EDQP, with the support of Stats NZ, which allows users to see in graphs or tables the percentages of responses for a particular variable in each region and for each major ethnic group (level one of Stats NZ's ethnic classification). These percentages have been obtained from the four primary sources used to compile the census dataset: the 2018 Census, the 2013 Census, administrative data sources, and by imputation using statistical procedures. This facility (EDQP, 2020a), along with notes on how to use it, can be accessed at <https://www.stats.govt.nz/reports/2018-census-external-data-quality-panel-data-sources-for-key-2018-census-individual-variables>.

The data sources facility complements and extends the main Stats NZ sources of information on the census variables which are contained within their metadata, DataInfo+ and Stats NZ (2019a & 2019b). Stats NZ has derived a quality assurance framework and a quality rating scale for assessing the quality of data generated for each of these variables (Stats NZ, 2019c). The quality ratings assigned to data for specific variables range from “very high” to “very poor”, and the shares of data for each variable that have come from the four different sources are one of several factors that play an important role in determining the overall quality rating.

The EDQP has produced an independent assessment of the quality of data relating to 31 of the variables containing information on individuals and dwellings. These independent assessments, along with links to the appropriate DataInfo+ pages and reference to the Stats NZ quality rating for these variables, are contained in the EDQP's (2019b) *Assessment of Variables* report. The panel used the same basic approach as Stats NZ to quantifying the quality of data that underpins their quality rating scale, but placed greater emphasis on variations in quality for data by ethnic group, at

lower levels of aggregation in a variable’s coding classification and at lower levels of spatial aggregation for the data. This resulted in some differences between the panel’s assessments of quality on a number of variables, particularly due to the differential response by ethnicity, a variable that is used frequently in analyses of census data in Aotearoa New Zealand. These differences are identified and briefly explained in the *Assessment of Variables* report.

### *The example of ethnicity*

Where the panel differs from Stats NZ in rating key variables, like ethnicity, there are specific sections addressing these in the EDQP’s initial (September 2019 (EDQP, 2019b)) and final (February 2020 (EDQP, 2020b)) reports. Two of the sections in the *Initial Report* contain detailed assessments of the data relating to Māori descent, and ethnicity (sections 4 and 5, EDQP, 2019a, pp. 34–61). The Stats NZ quality rating of “high” for the ethnicity variable is based primarily on the scores that are achieved at level two of the ethnicity classification (the 14 major specified ethnicities in the population) at the national level and reflects the ability of the data sources used to identify an individual’s ethnicity.

The panel has taken a broader view of the needs of users of ethnicity data than simply the ethnicity variable itself. The EDQP’s quality rating of “moderate” for this variable also takes into consideration much more variability in participation in the census, especially by members of the Māori and the different Pacific populations, at subnational levels. In the panel’s view, the “high” rating disguises the very significant under-enumeration of some ethnic groups in different parts of the country – something that users need to be very aware of when using these data. These variations are clearly apparent when reference is made to the mix of sources of ethnicity data that are used to derive the ethnic populations in different regions, local authority boards (Auckland), and statistical areas (SA2). Some of these variations are illustrated in the graphs and tables that can be obtained from the interactive web-based facility referred to earlier (EDQP, 2020a).

In the panel’s *Final Report*, census data on ethnicity and birthplace are examined further with particular reference to level four (L4) in the ethnicity classification (EDQP 2020b, pp. 45–59). Here a more refined set of quality assessments by Stats NZ for specific major ethnic groups is considered. Again, some differences between the panel’s quality ratings and

Stats NZ's ratings are discussed with particular reference to information relating to specific L4 ethnic categories. In concluding its review of data on ethnicity, the panel observed that:

... the full range of quality categories from very high to very poor apply when assessing ethnicity data at the L4 level of the classification. This variability in data quality for particular ethnic groups or clusters of specific ethnicities, in turn, has an important equity dimension. The lower quality data are generally for the non-European ethnic groups, many of whom are not very visible in many of New Zealand's statistical databases. Poor quality census data for these groups means that they are further disadvantaged and marginalised in the one source that aims to produce high-quality data on all ethnic groups. (EDQP, 2020b, p. 58)

The panel's focus on ethnicity has relevance for users of 2018 Census data because Stats NZ has recently released a major data package which they say "allows users to dive deep into New Zealand's diversity" (Stats NZ, 2020d, 2020e). The package of tables covering ethnicity, birthplace, Māori descent and a host of other variables relating to identity and culture is a significant addition to published 2018 Census data (Stats NZ, 2020e). Stats NZ is aware that users may want to check on some of the quality-related issues surrounding variables like ethnicity, and at the end of their "deep dive" release (Stats NZ 2020d), they include a reasonably comprehensive comment about, and links to, the reports produced by the EDQP on data quality.

### *Data on small areas, families and households*

The panel's *Final Report* also contains discussions of issues relating to the quality of small-area data, especially data at the SA2 level (EDQP 2020b, pp. 75–81), and the families and households data that, at the time the report was released, were being re-assessed by Stats NZ in terms of their quality (EDQP 2020b, pp. 65–73). The small area data are essential to the work of local authorities and community organisations. These users need to be aware that the uneven participation of the population in the 2018 Census enumeration has a strong spatial dimension. Many of the small group of SA2s that were worst affected by low census participation are in South Auckland. At this level of spatial aggregation, some variables, like occupation, had under 50 per cent of the data coming from responses to the census forms – very large shares of the data for occupation were imputed (EDQP 2020b, p. 79).

When the panel's *Final Report* was published at the beginning of February 2020, Stats NZ was in the process of undertaking a detailed assessment of the quality of the data on families and households in order to understand sources of error and to identify particular groups that were more likely to have poor quality data. This assessment has now been completed and instead of an overall quality rating of “very poor” for the 46 variables in the families and household suite, Stats NZ (2020c, p. 29) have determined that quality ranges from “high” (one variable) to “very poor” (four variables). The great majority (32 of the 46 variables) have been re-classified as being of “moderate” quality using the three metric assessment system in the Stats NZ quality rating scale (Stats NZ, 2019c).

The panel has not assessed these new ratings for variables relating to families and households, but it welcomes the additional work that has been done by Stats NZ to address issues relating to these very important data. As the panel notes in its *Final Report*: “Information about families and households that can be obtained from the Census of Population and Dwellings is vital for public policy, for meeting Treaty obligations to Māori, for population projections and for the derivation of analytical measures such as household crowding and social deprivation” (EDQP, 2020b, p. 65).

### **Accessing data to fill gaps in the 2018 Census data file**

By the time Stats NZ established the 2018 Census External Data Quality Panel, it had become very clear to the Department's statisticians that data from the 2018 Census enumeration was going to have to be augmented from other sources before it could be released to the public. Although the magnitude of under-enumeration could not be specified very clearly at the time the panel was formed, it was obvious from provisional figures for the total and Māori populations that there were major gaps in the data. It was later established that some information on around 83 per cent of the total population could be obtained from the census individual forms. For the Māori population the proportion was much lower – around 68 per cent. For Pacific populations it was only 65 per cent.

#### *Ethical and privacy issues surrounding use of administrative data*

During the meetings of the EDQP and senior statisticians and census managers between late August 2018 and early April 2019, a major topic of

discussion was ethical and privacy issues linked with accessing data from a range of administrative sources to fill gaps in the census data file. Stats NZ had access to a great deal of data on the people of Aotearoa New Zealand, and a considerable amount of this information was already within the integrated data infrastructure (IDI). What the panel wanted reassurance about was the extent to which the individuals providing the information that had been obtained from different government agencies were aware that this information might be used to fill gaps in the 2018 Census. This was of particular importance to many Māori given their concerns about data sovereignty and the right to decide how information on them will be used (Kukutai & Taylor, 2016; Te Mana Raraunga, 2018).

The previous census was an obvious source of information to fill gaps in some variables relating to people aged 5 years and over who were resident in New Zealand in March 2013. 2013 Census data, along with birth and death records and information from arrival and departure records between March 2013 and March 2018, enabled the coverage for several variables to be improved significantly. However, the range of information available on individuals from birth, death and international migration records that was directly comparable with the 2018 Census data was quite limited.

The panel had few issues with drawing on these sources for suitably comparable and timely data when filling gaps in certain 2018 Census variables. An obvious example is country of birth (a census question), which is, or should be, consistent for a person through their lives. Another is age which, while changing through time, can be updated once age at a specified time or date of birth can be obtained. The 2013 Census and the birth, death and migration records all contain information on age and birthplace.

Other variables that are subject to change through time, such as relationship status, occupation, ethnicity and iwi affiliation, cannot be assumed to be stable or to change in a consistent way over time. The timeliness of data obtained from alternative sources is very important if it is to be used as a substitute for responses to census questions in 2018. Timeliness aside, it quickly became apparent that there were a range of privacy and ethical issues that would have to be addressed when drawing data from alternative administrative sources such as health, education, welfare and IRD records, as well as records maintained on prisoners and members of the defence forces who had not completed census forms.

The panel had some robust discussions with Stats NZ about the extent to which they could assume they had social licence to use data from a wide range of administrative sources to fill gaps in the census. The Government Statistician assured us that their ways of accessing these data, and the use to which they would put the information when compiling the census data file, was legal and had been cleared by the Privacy Commissioner. Their initial document addressing these issues was released in November 2018, revised and updated in April 2019, and updated and released as a second edition in November 2019 (Stats NZ, 2018, 2019d, 2019e).

*Social licence, data sovereignty and the need for consultation before the 2023 Census*

Concerns about social licence and data sovereignty persisted throughout the panel's considerations of methodologies Stats NZ was employing to ensure the census data file had much greater and more comprehensive coverage of the national population than was available from the enumeration in March 2018. In late March and early April 2019, the panel prepared a series of internal discussion papers for Stats NZ on four topics, documenting many of its concerns. These addressed:

- 1) fitness for purpose of iwi data from Census 2018 (20 March)
- 2) use of administrative data to fill gaps in responses to the 2018 Census (4 April)
- 3) issues to consider when assessing data quality and inter-censal comparability (5 April)
- 4) clarification of multiple meanings of 'IDI' in current usage by Stats NZ (9 April).

Stats NZ appreciated the efforts of the panel in providing this advice which assisted in their consideration of these complex issues.

With just under a third of Māori not captured in responses to the census, there were clearly major challenges facing Stats NZ in meeting its 2018 Census Strategy (Stats NZ, 2016) with regard to "improving Māori national and subnational response rates" by "focus[ing] on Māori when making modernization changes" (p. 8). The multiple failures of the 2018 Census in delivering reliable data in a timely fashion to Māori as part of the

Government's Tiriti o Waitangi partnership with Māori is well documented by Kukutai and Cormack (2018). Cook (2020) also reviews the several obligations Stats NZ has under the Statistics Act (1975) and the Electoral Act (1993), amongst other legislation, to deliver data that only the census currently can deliver on Māori descent, te reo Māori and iwi affiliation.

Not surprisingly, sourcing data on Māori from administrative databases has raised many questions for Māori, amongst others, on the right that government agencies have to share information about clients that was given for a specific purpose. This is a standard requirement of informed consent that ethics committees look for when approving research projects involving collection of data from human subjects. In its recently released business case supporting a budget bid for funding for delivery of the 2023 Census,<sup>2</sup> Stats NZ (2019f) acknowledged that: "One of the enduring concerns about administrative data is the extent to which New Zealanders and the organisations that collect the data for operational reasons feel comfortable about how Stats NZ re-purposes it to produce official statistics. Stats NZ will continue to engage with stakeholders to promote the understanding of, and secure support for, the increased use of administrative data, and enable input from technical specialists and other stakeholders" (p. 28).

Stats NZ has made delivery on the needs of Māori and iwi one of its six key objectives for the 2023 Census. In this regard, they make the following commitment in the business case:

The next census must support the commitment under Te Tiriti o Waitangi that establishes Māori as partners with the Crown. The importance of this is reflected in the Data Strategy, which recognizes that having the right data available is essential to support Crown–Māori relations.

The low response rate for Māori in the previous census must be lifted – otherwise critical data such as iwi counts will not be available, and the benefits of rich, localized data about Māori will be lost. (Stats NZ, 2019f, p. 36)

At the same time, Stats NZ will continue to work with Māori and iwi and across government to improve the collection of Māori and iwi data in administrative sources.

These are big commitments for a census which has been funded at slightly above the lowest-cost alternative that Stats NZ offered in the business case. The \$210 million that is currently available for the 2023 Census is very close to what Stats NZ (2020f) regards as their bottom line:

“Stats NZ will not conduct a census with a budget less than the ‘Do the minimum’ approach, as this would set the country up for another failure. There would be a high likelihood that the unacceptable result of the 2018 Census is repeated” (p. 7). This is not the budget alternative Stats NZ was hoping for, especially given the need to invest significantly in building partnerships with Māori as Treaty partners and with other hard-to-reach populations that were poorly enumerated in 2013. It will be interesting to see how much investment can now be put into meeting the commitments to Māori made in the business case.

### **Comparing 2018 Census data with data from earlier censuses**

The 2018 Census breaks a long tradition of reporting results based almost entirely on census forms collected at the time with little statistical adjustment. The 2018 Census data are not directly comparable with data from earlier censuses for two main reasons:

- 1) The 2018 Census data come from a range of sources, not just from the forms used for the census enumeration, and this has resulted in some differences in the coverage of the population, the type of information that is reported for some variables, and the dates the data were collected.
- 2) The greater coverage allowed by use of alternative data sources, as well as extensive use of imputation, has resulted in Stats NZ removing the “not stated” or “not elsewhere included” categories from many variables in the 2018 Census that, in recent censuses, have often been categories with quite sizeable numbers.

These issues are addressed at the level of the variable in the panel’s *Assessment of Variables* report (EDQP, 2020b). The panel’s overall views on the statistical limitations of data added to mitigate non-response are presented in the *Initial Report* (EDQP, 2019a, pp. 28–29). The lack of direct comparability of 2018 Census data with data from the 2013 and 2006 censuses is something users interested in examining trends in particular characteristics of the population will need to keep in mind at all times. There is a real risk that false conclusions about patterns of change could be drawn from direct comparisons between data collected on certain variables in the 2018 Census and data from earlier censuses unless users remain alert to the differences in the way the 2018 Census data file has been created.

In this context, it is appropriate to conclude this research note by repeating the “Guidelines for use of 2023 Census data” that are contained in the panel’s *Final Report*:

When using the 2018 Census data the panel suggest:

- Read the individual assessments of variables written by the panel in either the panel’s *Initial Report*, in the separate *Assessment of Variables* report, or in the *Final Report*. These provide the panel’s overall assessments of the variables, provide key background information (including changes to the question or coding) and often contain caveats about the level at which the data can reliably be used.
- Read the relevant Stats NZ DataInfo+ page. There are links to these in the *Assessment of Variables* report. Also read Stats NZ’s (2019a, 2019b) reports on data sources, editing and imputation in the 2018 Census.
- Consider whether the use relies on 2018 Census data alone (i.e. cross-sectional analyses), or uses changes in one or more variables between censuses. In the latter case, be aware of the changes in methods in the 2018 Census and their impact on variables.
- Users wishing to understand specific communities and small areas should consult the data sources facility (EDQP, 2020a) to get an indication of data quality for the variables in which they are interested.
- Some variables contain high levels of imputed data (e.g. occupation 20.3 per cent; main means of travel to work, 19.0 per cent; hours worked in employment, 18.7 per cent). CANCEIS imputations are designed to be unbiased but they do increase uncertainty.<sup>3</sup> Data users who are using individual level micro-data (e.g. in the IDI) will face different problems and options compared to users who are using published tabulations.
- CANCEIS imputations will get many individual responses wrong, but these errors will be more-or-less offsetting so that the overall results are acceptable at a level of aggregation that is high enough for the errors to cancel out. Tabulations using variables with high levels of imputation will therefore be reasonably accurate for tabulations with large counts, but not for small cell counts (e.g. in small areas).
- Analyses at the individual level, such as statistical models fitted to individual-level data, may be very badly affected if these involve variables with high levels of CANCEIS

imputation. Anyone planning to do individual-level analyses on data with substantial non-response needs to address these issues directly – e.g. by restricting the analyses to subpopulations with low levels of imputation, or by not using the CANCEIS values and using alternative missing-data techniques such as multiple imputation.

- If cross-tabulating variables, or carrying out complex analyses, consider the quality of each 2018 Census variable in isolation and then when taken together. For example, while data on age is of very high quality, variables analysed by age cross-tabulated by ethnicity may not be. The quality of such analyses will critically depend on which level of the ethnicity classification is used, and which specific ethnicities are being analysed.

(EDQP, 2020b, pp. 43–44)

## Acknowledgements

This research note contains some personal reflections on aspects of the work of a panel that was initially appointed for a period of around 8 months (late August 2018 to late March/early April 2019) but which ended up working with Stats NZ on 2018 Census data quality issues until early February 2020. It was a privilege working with such a talented group of specialists in and users of census data. As co-chair of the panel, I especially want to acknowledge the tireless work Adele Quinn (Manager, Census Analytics, Stats NZ) and Ian Cope (Census Consultant, United Kingdom – EDQP panel member) did for the panel throughout the 18 months of its operation. The knowledge we all gained from sustained engagement with a host of challenging issues relating to census data quality was enormous. Useful feedback on an earlier version of this paper was received from Christine Bycroft, Len Cook, Ian Cope, Robert Didham, Adele Quinn and Carol Slappendel. I remain responsible for any errors or omissions.

## Notes

- 1 The members of the 2018 Census External Data Quality Panel and the panel's main objectives are summarized in the appendix. A link to the Terms of Reference for the Panel, which can be accessed on the Stats NZ website, is provided in this appendix.
- 2 Access to the text of the business case was proactively released on 21 May in a media release announcing a new governance advisory board and the Budget 2020 decision regarding funding for the 2023 Census (Stats NZ, 2020f).

- 3 CANCIES (CANadian Census Edit and Imputation System) is the name of the imputation system used by Stats NZ to fill gaps in variables that could not be filled from alternative sources (such as 2013 Census or administrative records). The imputation of missing values is discussed in Section 3 Statistical Methods in the panel's *Initial Report* where the methodology employed by Stats NZ to produce the 2018 Census datafile is endorsed (EDQP, 2019a, pp. 21–22).

## References

- Cook, L. (2020). The Census of Population and Dwellings – where to in 2023? *New Zealand Population Review*, 46, [this issue].
- External Data Quality Panel (EDQP) (2019a). *Initial report of the 2018 Census External Data Quality Panel*. <https://www.stats.govt.nz/reports/initial-report-of-the-2018-census-external-data-quality-panel>
- (2019b). *Assessment of variables*. <https://www.stats.govt.nz/reports/2018-census-external-data-quality-panel-assessment-of-variables>
- (2020a). *Data sources for key 2018 Census individual variables*. <https://www.stats.govt.nz/reports/2018-census-external-data-quality-panel-data-sources-for-key-2018-census-individual-variables>
- (2020b). *Final report of the 2018 Census External Data Quality Panel*. <https://www.stats.govt.nz/news/final-report-from-2018-census-external-data-quality-panel-now-available>
- Jack, M., & Grazidei, C. (2019). *Report of the independent review of New Zealand's 2018 Census*. <https://www.stats.govt.nz/reports/report-of-the-independent-review-of-new-zealands-2018-census>
- Kukutai, T., & Cormack, D. (2018). Census 2018 and implications for Māori. *New Zealand Population Review*, 44, 13–151.
- Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous data sovereignty: Toward an agenda*. ANU Press.
- Stats NZ. (2016). *2018 Census strategy*. <https://www.stats.govt.nz/methods/2018-census-strategy>
- (2018). *Privacy impact assessment for using admin data in 2018 Census data*. <https://cdm20045.contentdm.oclc.org/digital/collection/p20045coll1/id/2693/rec/2>
- (2019a). *Overview of statistical methods for adding admin records to the 2018 Census dataset*. <https://www.stats.govt.nz/methods/overview-of-statistical-methods-for-adding-admin-records-to-the-2018-census-dataset>
- (2019b). *Data sources, editing, and imputation in the 2018 Census*. <https://www.stats.govt.nz/methods/data-sources-editing-and-imputation-in-the-2018-census>

- (2019c). *Data quality assurance for the 2018 Census*.  
<https://www.stats.govt.nz/methods/data-quality-assurance-for-2018-census#background>
- (2019d). *Creating the 2018 Census dataset by combining administrative data and data from census forms. Our privacy impact assessment*.  
<https://cdm20045.contentdm.oclc.org/digital/collection/p20045coll1/id/2698/rec/1>
- (2019e). *Creating the 2018 Census dataset by combining administrative data and census forms data: Our privacy impact assessment (second edition)*. <https://www.stats.govt.nz/privacy-impact-assessments/creating-the-2018-census-dataset-by-combining-administrative-data-and-census-forms-data-our-privacy-impact-assessment>
- (2019f). *New Zealand's next Census of Population and Dwellings. A business case to support a funding bid for Budget 2020*.  
<https://www.stats.govt.nz/corporate/funding-and-delivering-the-next-census-2023-census-of-population-and-dwellings>
- (2020a). *New Zealand's population passes 5 million*.  
<https://www.stats.govt.nz/news/new-zealands-population-passes-5-million>
- (2020b). *Customer update on 2018 Census data and activities – May 2020*.  
<https://www.stats.govt.nz/reports/customer-update-on-2018-census-data-and-activities-may-2020#data>
- (2020c). *Families and households in the 2018 Census: Data sources, family coding, and data quality*.  
<https://www.stats.govt.nz/methods/families-and-households-in-the-2018-census-data-sources-family-coding-and-data-quality>
- (2020d). *Census data allows users to dive deep into New Zealand's diversity*. <https://www.stats.govt.nz/news/2018-census-data-allows-users-to-dive-deep-into-new-zealands-diversity>
- (2020e). *2018 Census – NZ.Stat tables*.  
<https://www.stats.govt.nz/information-releases/2018-census-nz-stat-tables>
- (2020f). *Stats NZ welcomes new advisory board*.  
<https://www.stats.govt.nz/news/stats-nz-welcomes-new-advisory-board>
- Te Mana Raraunga (2018). *Te Mana Raraunga statement on 2018 New Zealand Census of Population and Dwellings: A call for action on Māori census data*. <https://www.temanararaunga.maori.nz/>

## Appendix: The 2018 Census External Data Quality Panel

Stats NZ constituted the 2018 Census External Data Quality Panel (EDQP) in August 2018. Panel members included:

- Richard Bedford, Emeritus Professor, retired Professor of Population Geography, Auckland University of Technology and University of Waikato (co-Chair)
- Alison Reid, Team Manager, Economic and Social Research and Evaluation, Auckland Council (co-Chair)
- Associate Professor Barry Milne, Director, COMPASS Research Centre, University of Auckland
- Dr Donna Cormack, Senior Lecturer, Te Kupenga Hauora Māori, University of Auckland; Senior Research Fellow, Te Rōpū Rangahau Hauora a Eru Pomare, University of Otago, Wellington
- Ian Cope, international census expert, ex-Office of National Statistics (ONS), United Kingdom
- Len Cook, former New Zealand Government Statistician and former National Statistician of the United Kingdom
- Tahu Kukutai, Professor of Demography, National Institute of Demographic and Economic Analysis, University of Waikato
- Thomas Lumley, Professor of Biostatistics, University of Auckland.

As set out in the Terms of Reference,<sup>†</sup> the objectives of the panel were to provide independent advice to the Government Statistician about:

- whether the methodologies used to produce quality information from the census are based on sound research and a strong evidence base
- approaches to data processing and methodology, and increased use of administrative sources that affect the quality of the data
- data issues that may affect the usefulness of the data for Māori and iwi as Treaty partners
- any quality issues people will need to consider when using 2018 Census and related population data, and any further work required to assist customers.

---

<sup>†</sup> The Panel's Terms of Reference (version 3, 23 August 2019) can be accessed at: <https://www.stats.govt.nz/reports/2018-census-external-data-quality-panel-terms-of-reference>. The EDQP's three substantive reports and its interactive facility for exploring the sources of data for 2018 Census variables, can all be accessed at: <https://www.stats.govt.nz/Reports/?filters=Census&start=0>.