To the Justice Committee

From: Leonard Cook

15 April 2025

Submission on The Data and Statistics (Census) Amendment Bill, and the Electoral (District Boundaries) Amendment Bill

Leonard Cook Qualifications:

I was the NZ Government Statistician (1992-2000), then UK National Statistician (2000-2005). I remain actively engaged in the analysis of population statistics in several policy areas. I have been a member of the NZ Statistical Association, the NZ Association of Economists and the UK Royal Statistical Society for over 50 years, and the Population Association of NZ soon after its inception. I am a former Vice President of the International Statistics Institute (2005-2009), and Companion of the Royal Society Te Apaarangi. Over the past 50 years I have been involved in many censuses of population, written widely[1] about the uses of population statistics, and attended national and international professional forums advancing population statistics.

RECOMMENDATIONS

1) **I strongly recommend to the Justice Committee that it uses its own powers to commission an independent expert review of the Administrative Data First model and consequent cessation of the periodic Census of Population and Dwellings.**

   Trustworthy statistics are those where the Government Statistician can assure those who depend on them of their statistical integrity, relevance and fitness for purpose.

   An independent expert review would examine the validity of each of the key assumptions on which rest statistical and operational integrity. For example, although the per respondent cost of the 2023 Census of Population and Dwellings was grossly excessive, it has never been investigated.

[1] The quality and qualities of population statistics, and the place of the census. L Cook **Area** (2004) 36.2, 111–123

2) **I recommend a review of the validity of the amendments to definitions in the Electoral Act.**

Because the year-to-year volatility of any detailed analyses produced for small areas from the Administrative Data First model, it is unlikely that the annual statistics produced will meet current expectations of reliability, necessitating comprehensive validation. All counts will be continually revised whenever departmental databases are updated.

For each specific Census date, the annual statistics will be revised at each update of the source information for the Administrative Data First model. The reference to the most recent census is replaced by a new ad hoc concept - "***the figure most recently reported by the Government Statistician".***

3) **I recommend that the Justice Committee seek advice from those who are affected by the Bill, in local government, the community and those sectors of government that may not have understood their extent.**

I have identified important several critical inadequately explored issues.

4) **I recommend that the Minister of Statistics to be asked to request the Government Statistician to begin preparations for a Census of Population and Dwellings in October 2028**.

It is possible to organise a Census of Population and Dwellings in October 2028 that is consistent with existing legislation, and with those Censuses that existed in 2013 and before, subject to content and adoption of now contemporary practices from elsewhere. Census experts from the Australian Bureau of Statistics should be consulted. A proper independent external review of plans must be expedited. Section 3 of this submission (THE PATHWAY TO AN OCTOBER 2028 CENSUS OF POPULATION AND DWELLING) outlines what could be begun immediately to enable the system of population statistics to continue.

The preamble to the Bill states that it will "*move the next census to 2030 to allow the time needed for all-of-government efforts to identify, sequence, and deliver the necessary improvements to admin data"*. This delay provides no guarantee of eventual delivery. The shift to 2030 from 2028 occurred after the proposals were sent for Ministerial endorsement, so could not have resulted from statistical or managerial concerns.

Questions about the Administrative Data First model

Table of Contents

1. GENERAL POSITION:

1.1 Delaying the next Boundaries Commission from 2028 to 2030 is neither justified nor necessary

The Electoral (District Boundaries) Amendment Bill and the Data and Statistics (Census) Amendment Bill will delay the next Boundaries Commission by two years, from that which is currently in legislation. Despite endorsing the proposals for the legislatively determined Census of Population and Dwellings, Statistics NZ has since engaged in activity that was not compliant with the Data and Statistics Act 2022. They have ignored the significance of the constitutional implications in Electoral legislation. The delay from 2028 to 2030 is caused not by an unexpected event, such as depression, war or earthquake, but as a result of government officials changing their minds once the new legislation was passed in 2022. The rate of population changes and its variability among places around New Zealand is changing, as measured at a Local Authority level, which

will test the ability of the Boundaries Commission to recognise communities of interest within its 5% margin. The rate of change in the Papakura Local Board area between 2013 and 2023 seen in Table 1 in Section 2.6 is the greatest seen in any local authority or local board over that period. It highlights the potential for electorate boundaries to become less representative between Commissions. The Boundaries Commission's flexibility to manage this will be further constrained when the unit of adjustment will shift from meshblock to Area Unit level (200 to 2,000 persons approximately) which will make the practicalities of designating boundaries with natural or man-made landmarks. The comparatively rapid growth of the Māori electoral population and the implications for the distribution of Parliamentary seats are an even more complex element that should be well known to officials. Officials must inform Parliament how they will ensure that a future date can be met, and the various stages between now and then.

1.2 The quality of statistics will be undermined when methods are limited to the "administrative data first" model

The quality of future population, social and economic statistics will be undermined when their methods are limited to the "administrative data first" model. The Data and Statistics Act Amendment Act rigidly limits the statistical properties of all future population statistics to the "administrative data first" regime. If adopted, the Bill deliberately removes from the Government Statistician any flexibility in the implementation of statistical methods for future censuses. It may have constitutional implications as it is a methodological handcuff on the independence of the Statistician that is made explicit in both Section 16 of the Data and Statistics Act 2022 and the Statistics Act 1975. New Zealand statistics legislation will be considerably out of line with that of other countries. The effect on electoral boundary setting processes is not transparent.

1.3 One core piece of New Zealand's national infrastructure will cease

The five yearly Census of Population and Dwellings anchors New Zealand's long-term series of integrated enumeration-based censuses and surveys. It is the most precisely designed and quality assured set of statistical information of government. It is the gold standard by which other government and commercial statistical sources are evaluated. It enables valid comparisons of population characteristics across age, ethnic, occupation and other groups, for small places, our past, and with those nations that like NZ comply with UN definitions and standards. The administrative records gathered by a variety of agencies with disparate data management standards will lack the common standards or definitions which underpin the integration with other statistical sources that enable NZ to produce official statistics of knowable reliability. By fragmenting the

integrating elements of official statistics, this will significantly undermine the quality of information we have to make policy decisions in the future, in a very wide range of social and economic areas that have implications not only for now but future generations.

1.4 The Administrative Data First model brings more variability for counts

The year-to-year volatility of the proposed counts of any period produced by the Administrative Data First model may reduce confidence in them. New Zealand already has a rich array of population estimates produced for every intervening year. These annual estimates can readily take advantage of where administrative records can indicate shifts in trends. For any more complex analyses proposed to be provided annually, the smaller the locality that they refer to the more likely that the variation in the year-to-year change will make the counts of less use than what has existed before. This would reduce trust in any counts produced in this way. Indeed, for local authorities the integration of wider economic, climate and other information through a geospatial frame will enable a more rapid and reliable statistical response to local change than the Administrative Data First model.

1.5 The Administrative Data First model will most likely need to be underpinned by a population register

The Bill presents and overly optimistic picture of the Administrative Data First model, whereas its formulation in this way makes a move to a population register a natural step when problems of coverage become visible. Statistics NZ note[2] of the administrative data first approach, that:

> "*A major limitation of this approach is that the range and quality of information produced would be limited to what already exists in administrative systems. Any administrative census approach is likely to require a coverage survey and a large sample survey to adjust for the limitations.*"

These necessary complementary new statistical activities have yet to be prescribed fully. Their feasibility should be known before the Administrative Data First model can be comprehensively assessed. The above reservations reflect the fact that no country whose population statistics system is anchored in government administrative records operates without a population register. In 2026, what the Statistics NZ 2012 study noted still remains the position, that:

> "*All 15 countries that produce census outputs from administrative sources use a national population register*."

[2] Statistics New Zealand (2012). Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy. Wellington: Statistics New Zealand

1.6 Each Census of Population and Dwellings is more than a dataset

The proposal to have a "census" every year indicates a confusion in what makes up a Census. A Census of Population and Dwellings is not only the specially designed statistical source of known reliability, but it is an integrating element and statistical survey frame of a series of statistical surveys with sufficient elements specifically designed to be in common so that comparisons across time and generation are enabled. Each Census comes with a comprehensive body of published statistical reports about the social and economic condition of the population, and of particular groups within it.

No statistical office can ever expect to do this full range of activities every year, to the extent that has been expected of past censuses. There needs to be clarity in what is offered if it is not to be a false promise. There are issues of international credibility in adopting such a pragmatic approach to defining in law what is a "census", where it is peculiar to New Zealand. The implications for national and international trust in major economic and demographic statistics involve international organisations, and ratings agencies which use NZ statistics for comparison purposes.

1.7 An unprecedented degree of discretionary authority is needed for electoral measures

The combined effect of the two Bills is to muddle the concepts behind key statistical definitions used in electoral legislation and ignore how consistency of definition enables the integration of diverse statistical sources and their comparability over time and place. A standard formulaic approach is to be replaced by a series of essentially ad hoc judgements dependent on the Government Statistician of the day, and what is defined as the country's population on that day. The Government Statistician is to be given an unprecedented degree of discretionary authority of measures that were previously defined in statute, yet there is no oversight body such as the UK Statistics Authority that oversees the application of judgements made. There is no requirement that either the role of Government Statistician or Deputy Government Statistician be filled by a person with qualifications and experience known to be appropriate for official statistics, yet the proposals place even more reliance on that than before.

2. DETAILED ARGUMENTS

2.1. The Bill's preliminary statements are based on selective assessments.

> ***Policy context and objectives for admin-data-first approach to census***
>
> - "*The transition to an admin-data-first approach to future censuses is required to respond to declines in affordability, efficiency, data quality, and timeliness under the full field enumeration survey approach. The full field enumeration survey approach is no longer financially sustainable. There is growing resistance to completing census forms and rising costs in maintaining survey response rates to preserve data quality. Increased investment for the 2023 census lifted response rates but, despite this, target response rates for some populations were not met.*
> - *Alongside this, a full field enumeration census model based on a 5-yearly survey is not delivering data at the frequency that users need to make well-informed decisions"*
>
> Data and Statistics (Census) Amendment Bill

1. *Will other countries be making similar changes?*
    - New Zealand is in fact alone in the path that this Bill will set.
    - Adopting the proposals in this Bill will mean that NZ will be adopting a solution that is unlike any other country with which we usually compare ourselves. We will lose the benefit of the intensive collaboration with Canada and Australia by Statistics NZ over some forty years.

2. *Will the full field enumeration survey approach be no longer accepted with the New Zealand public?*
    - Records of the balance between non-contact and refusals at enumeration are poor, and unknown as a result of management failures, for which after the event the public have become blamed.
    - The reduction in completed Census questionnaires involved a significantly faulty Census enumeration, contradicting the preamble.

3. *Will the Administrative Data First model provide information about critical issues?*
    - The rich body of ethnic and family information we have long gathered in New Zealand does not exist in register-based countries. In most of those countries, immigrants are designated foreigners, while indigenous populations have no history of recognition of the sort seen in NZ or USA.
    - The model will not improve how we site some $200 billion of infrastructure, plan emergency management or foresee where houses will be needed in the

next decade. This needs trustworthy locality-based population projections. So does responding effectively to climate change and the environment threats which are certainly not evenly spread around New Zealand.
- A wide range of household and family statistics will not be available, and those that will be available will not be provided for small areas that they now are (see Insert 1 below).

4. *Will the Administrative Data First model will reduce the cost of producing population statistics?*
    - The cost of aligning the way information is classified using statistical definitions as well as operations meanings is impossible to assess. While administrative data will always be available and has always had an important place in population statistics, its other value in providing alternative imputation options for missing responses and informing a propensity index has been poorly explored. The Statistics NZ 2012 strategy outlined several other uses.

    Why was the total cost of the 2023 Census of Population and Dwellings grossly excessive, but never investigated?

    - In real terms, the per respondent cost of the 2023 Census was over twice that of any previous Census. The cost of the 2023 Census is simply not credible and has never been scrutinised by Treasury or Finance Ministers.
    - Why such a significant cost overrun was not examined by any of the commissioned reviews has never been explained.

Rigorously evaluating the fitness for purpose of future population statistics of the Administrative Data First model would involve comparing that model with the continuing the long-standing statistical census approach, including the methodological and operational developments that ordinarily would have been implemented for a Census in 2028 or 2030, as is occurring in Canada, the UK and Australia.

Definitions in the Electoral Act is to be amended by the Electoral (District Boundaries) Amendment Bill as a consequence of the absence of simultaneity in the population counts compared to those from the electoral rolls.

2.2.This Committee is the first chance for some to challenge census plans

To date there has been no independent expert review of the fundamental methodology changes proposed or their fitness for purpose for current and future uses of population statistics, that also assesses the assumptions behind the advice. Despite this, Government Statisticians, Ministers of Statistics and Cabinet have endorsed the proposals. The duty of care expected at each stage of decision making before these proposals have been presented to Parliament has been less than necessary, given the

huge consequences of getting it wrong. “Modernising” requires coherence between managerial, operational and statistical dimensions, as well as the total fiscal impact and the retention of public trust.

By seeking independent expert advice on the necessity and viability of what is proposed, Members of Parliament do have the means to question and challenge the arguments and evidence behind these legislative changes currently before them.

Because four of the last five Census reviews[3][4] initiated by the Government Statistician did not seek submissions from outside the public sector, this is the first occasion that I and many others have been able to contribute to any public consultation process on the future of population censuses in New Zealand.

In the material available, the Government Statistician has not provided a proper comparison of the proposed but internationally untested administrative data first model with that which would result from continuing the long-proven series of statistical censuses. I have outlined in Section 3 below (The Pathway to a 2028 Census of Population and Dwellings) what an experienced Government Statistician would have expected to have been achieved during the past ten years as part of the natural evolution of the existing census model and which has occurred in other jurisdictions. The reports of the Government Statistician provide no account of any such examination, although it is an obligation which exists until these Bills are passed.

That the Census of Population and Dwellings for 2018 was a statistical, operational and managerial failure is well recognised, but the full extent was not identified in the external review that two persons conducted. The reviewers had no interaction with any NZ experts external to Statistics NZ. The Census of Population and Dwellings of 2023 was an attempt at replicating the approach that produced the final solution for 2018. For 2023, little attempt was made to examine what had been learnt in how other countries were continuously evolving their field enumeration systems and practices. In fact, while a poor public response has been now blamed for the failures of both 2018 and 2023, the agency has done little to explain how much of the poor response has been shared between non-contact and refusals.

A review commissioned by the Government Statistician was the only time[3] that the selective reasoning behind the automatic adoption of the Administrative Data First model was transparently contested. The approach was rejected. From then on, unlike the normal practice with changes to well established practice, no alternative was seen to be needed for comparison. A comprehensive expert review should involve a user

[3] Stats NZ (2025). *Evaluation of New Zealand’s future census options for 2028 and beyond*. Retrieved from www.stats.govt.nz.ISBN 978-1-991307-69-9 (online)

[4] Report of the Independent Review of New Zealand’s 2018 Census July 2019 Wellington, New Zealand ISBN: 978-1-98-858337-2 (online)

focused assessment of the fitness for purpose. It would necessitate recognising the heightened importance now for population estimates and projections. A review must also recognise the need for a greatly increased capacity to integrate a wide range of information sources about place, facilitated by a much-needed national geospatial frame. This has not been evaluated by Statistics NZ. The preamble to the Data and Statistics (Census) Amendment Bill, and earlier advice to cabinet are both flawed in several ways, as are the reasons given for assuming a lack of alternatives.

2.3.The quality of administrative records cannot be assumed to be sufficient

Trustworthy statistics are those where the Government Statistician can assure those who depend on them of their statistical integrity, relevance and fitness for purpose. Unlike statistical surveys and censuses, the departmental records held by government departments fail to meet common standards of definition, or of what they contain, or of what populations they represent. Those departments whose systems are now outdated may take many years to invest in conforming to any statistical classifications. We have never achieved that in any part of government. At least some of the information about anyone that these records contain at any particular date will often be months or years old in the Administrative Data First model.

To an official statistician, departmental records are equivalent to the fossils that archaeologist use to build up the skeletons of extinct animals. Because what is collected in departmental records can change only slowly, compared to statistical surveys and censuses, they can have the same effect as driving a car with the rear vision mirror – what is important is just not available. This is particularly important for information needed by local government or for electoral purposes, such as comparison across time in small areas.

The smallest area that will provide information of comparable quality to the statistical meshblock (averaging around 33 occupied dwellings and 100 persons) or the statistical area 1 introduced for the 2018 Census of Population and Dwellings of which about 60% are just one meshblock (averaging around 56 occupied dwellings and 150 persons) would now be the statistical area level 2 (averaging around 750 occupied dwellings and around 2,000 persons). Multidimensional cross tabulations of people, say by relating ethnicity to place, age and occupation, will be of significantly lower quality. For example, Statistics NZ in the report "Predicting usual residence address from admin data in the 2023 Census" stated that usual residence address was 92 percent consistent with the 2018 Census address at the statistical area 2. The consistency is lowest for young adults, particularly between ages 18 and 25. In these comparisons with the 2018 Census the unprecedented degree of imputation from the IDI that is already in the system must be recognised. This loss of detail will require a radical change in how the five yearly electoral boundary commission determines electoral boundaries.

We will have lower quality information about Māori and Pacific communities despite them providing nearly 40% of the people of workforce age under 40 years in twenty years' time. Language and religion will not be collected. The richness and diversity of migrant communities is not going be measured to the extent we will need for understanding future population change. Because income tax returns are no longer compulsory, the Government will not have any more a regular full coverage of the occupations of all in New Zealand. This is at a time of rapid transformation of jobs from globalisation, artificial intelligence and the demographic dynamics of the population.

Apart from reporting on demographic, geographic and socio-economic change, the Census data is used for the design of social and household economic surveys. The designs may involve model-assisted estimation of both probability-based samples with frame bias or high non-response, and non-probability-based samples which increasingly use online panels. They are often recruited from reward programs and offer rewards for participation in surveys. Census data underpins small area estimation where estimates from the underlying survey for small areas such the minor urban areas or rare population groups, such as Niuean, have low accuracy and modelling The Census data improves the accuracy.

All sources that produce statistical measures must be assumed to contain a mix of errors. The Australian Bureau of Statistics uses a total survey error (TSE) model to assess the possible errors in statistical sources and measures. Errors can be the result of one or more of:

a) Under-coverage or over-coverage
b) Duplication or missing data
c) Specification Errors and measurement Errors
d) Data Processing Errors
e) Linkage errors.

A recent Australian review noted that the key coverage issues with a merged set of key administrative records were:

- *Under-coverage (In Australia this was due to infants, recent migrants, international students, and those living in remote areas especially Indigenous people.)*
- *Over-coverage (e.g. persons who have emigrated, perhaps temporarily, but are no longer resident)*
- *Duplication Persons who are duplicated on the list, perhaps because they have different addresses in the administrative systems, or their name was specified differently across the administrative systems. (In Australia, this was predominantly middle-aged persons especially males).*

The study noted that

"The net impact may even be small at the national level (as it was in the Australian study), but it is unlikely to be the case at the regional and sub-population levels. Accuracy assessments should not just be at the national level".

Ironically, the more use that is made of anonymised administrative data for policy and evaluation, the more important the availability of survey and census data becomes. Survey and census data are crucial for validating specific measures where they have to be derived from administrative data, not only in assessing coverage and reliability, but being able to respond to the need for information about new topics requiring investigation.

2.4.The Indexes of Deprivation (NZDep) will no longer be prepared

The Indexes of Deprivation (NZDep) have been prepared after every Census since 1991, but 2023 will now be the last year for which they can be prepared. NZDep is just one of a variety of highly valued statistical measures which have been regularly derived from each Census of Population and Dwellings for use in public policy.

Insert 1: Taken from Submission to Justice Committee by Professor Peter Crampton and others

What is the NZDep index?

The NZDep index uses data from the NZ Census of Population and Dwellings to quantify the relative socioeconomic deprivation of small areas (generally comprising 100 to 200 persons) throughout New Zealand.

The latest index, NZDep2023, is a statistical combination of nine socioeconomic deprivation characteristics available from the individual and household Census returns in 2023. These variables are listed in the first column of the table below.

Components of the NZDep2023 index and their future availability

NZDep2023 socioeconomic deprivation variables	**Source**	**Available in the administrative files in Stats NZ's Integrated Data Infrastructure (IDI)?**
People aged 18 – 64 receiving a **Means-tested Benefit**	Person	Yes
People aged 18 – 64 **Unemployed**	Person	Yes
People aged 18 – 64 with **No Qualifications**	Person	Yes
People living in households with Jensen-equivalised income below an income threshold	Household	**NO**, but may become possible by 2030, although with likely quality issues

(Low Household Income)		
People aged < 75 living in a **Sole-parent Family**	Family	**NO**, but may be available by 2030, although with unknown reliability
People living in equivalised households below a bedroom-occupancy threshold **(Overcrowding)**	Household	**NO**, but some other information may be possible by 2030, although it is unlikely to be as good
People living in dwellings that are always damp and/or always have mould greater than A4 size **(Unhealthy Home)**	Dwelling	**NO, NEVER**
People with no access to the **Internet at Home**	Household	**NO, probably NEVER**
People **Not Living in Own Home**	Household	**NO, possibly NEVER**

The two main consequences for NZDep if an Administrative Population Census is introduced

Consequence 1: Household data availability.

> "deprived individuals in an area have different needs and priorities for effective economic, social and cultural engagement."

Consequence 2: Data quality issues in a person-based administrative census.

The quality issues for data in the administrative census include timeliness, coverage and accuracy.

Timeliness: Unlike previous household-based censuses which collected information from everyone at the same point in time (Census Day), a person-based administrative census will combine information collected at the last occasion in any particular data set when contact was made. That is, the information will no longer be collected at a consistent date but across unknown previous days, weeks, months, or possibly years.

Coverage: There is no guarantee that every relevant person will be captured in any of the data sets in the IDI. The absence of such persons will result in underestimation of the characteristics being ascertained by that data set.

Accuracy: Since the IDI is comprised of various data sets, collected in different circumstances (in person or electronically), by different organisations, and for different purposes, it is likely that levels of accuracy across the data sets will vary. Inconsistencies in how variables are coded by the different organisations will add further inaccuracies. Cumulatively, these inaccuracies are likely to be greater than those in a single data set, such as each of the previous census data sets.

All these data quality issues will negatively affect the production of any future NZDep-like index. They will be encountered by everyone using the new census data, for whatever purpose.

2.5.The methodological straitjacket of the Bill reduces initiatives

The Data and Statistics Act amendments will shift the central element on which population statistics are to be anchored, from dwellings to persons. Continuity with previous surveys and censuses including the most recent will be lost. This will create a methodological and operational strait jacket that will undermine the rapid integration of locality-based information. Placing a legislative restriction in this way on the statistical characteristics of this most fundamental element of the system of population censuses and surveys is to put the system in a straitjacket. This will narrow options available to future government statisticians for collecting information about the population, households and their economic activity. This could have consequences for the integrity of economic measures, most significantly GDP per capita, as well as the household sector of the National Accounts.

A proper understanding of the expanding needs of emergency management, planning housing need, and identifying where national infrastructure is to be placed among localities have become more significant key priorities for population statistics in the immediate decades to come. Low fertility and higher longevity are changing the form and dynamism of family and community structures, as well as reducing the future fiscal position, and complicating the fit between the need for public services and their availability. How much longstanding measures of how family and household structures are evolving is critical to understanding the inevitable policy changes this will involve. Also, tracking the local dynamism of all forms of economic and community enterprises has now greater importance in predicting future population change for the diversity of places that make up New Zealand. The access to administrative records and community information requires a geospatial frame, that the Administrative Data First model cannot provide. The population of the Papakura Local Board by 2023 could not have been predicted in 2012/13 by existing projections methods and existing methods of integrating local information from relevant sources. It exemplifies the range of change now experienced in smaller localities.

Table 1: Population change the Papakura Local Board (2013-2023)

Census Year	European	Māori	Pacific	Asian
2013	26,064	11,979	6,201	5,448
2018	28,305	15,438	9,750	13,497
2023	26,541	17,811	14,811	24,732

In the more volatile environment we are now in, there is a need to increase confidence in the reliability of population measures, through designing more confrontation as part of the measurement system, and more measures at all levels of reliability.

A survey-based system of population statistics provides more vehicles for this, compared to the Administrative Data First model, which is in itself unable to confirm which parts of the population it does not cover. We are behind in meeting these needs, because the approaches used in 2018 and 2023 made no advances in the ability to integrate population statistics for localities with the expanding range of local data that technological advances including artificial intelligence are making available. (The 2023 Census review noted that the address list was even worse than the 2018 Census.)

2.6.Constitutional issues are not transparent

The IDI breaks new ground in the linking of government and community information about the connections across government agencies identifiable with individuals. There is no specific legislative oversight of the way that the statistical system is now tracking persons, by their identity.

Despite this, the Administrative Data First model cannot meet the current obligations in the Electoral Act with certainty without being underpinned by a population register, as occurs in all other countries that have adopted this form of census model. The fundamental importance of a population register was correctly recognised by Statistics NZ in a 2012 paper[5] but public discussion since on the need for this has been dormant. However, that departmental officials remain committed to this led them to raise the topic with the external departmental ethics committee in 2025, where it was universally rejected. This 2012 paper noted

> "*Political and public acceptance of the need for a population register, a personal identity number, and accurate and timely capture of address changes are prerequisites for a register-based census. This approach also demands close cooperation between the statistical agency, register authorities, and the public administration, as well as strict legislative oversight. Generally, establishing and maintaining a high-quality register-based system requires significant resources and societal will*."

Up to now, all countries that base their population statistics on their government records have a population register. However, no country that uses a population register has been able to produce family and household statistics of the quality long used in New Zealand. Both the Electoral (District Boundaries) Amendment Bill and the Data and Statistics (Census) Amendment Bill ignore important constitutional issues.

[5] Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy StatsNZ 2012

The Administrative Data First model cannot provide the extent of detail associated with small area meshblock counts that the Census of Population and Dwellings have always provided. They have been fundamental to the work of the Surveyor-General in providing the five yearly Boundary Commission with the detail needed for determining electoral boundaries.

- The future need to aggregate population information into larger areas to achieve the same accuracy and detail as now will reduce the flexibility in how population measures are used for electoral boundary determination
- The changes in how populations are defined Increase the range of arbitrary decisions of significance to electoral boundary setting that will now be made without the possibility of prospective scrutiny except that of a judicial body. For example, enabling the *definition of a districts population to be "***as assessed by the Government Statistician**" is contrary to practices that ensure statistics are trusted.

The combined effect of the two Bills is to muddle key statistical definitions and ignore the vital need for consistency of definition for the integration of statistical sources and their comparability over time and place. Conformance with international standards and definitions is to be replaced by a series of essentially ad hoc judgements dependent on the Government Statistician of the day, and what is defined as the country's population on that day. All such figures produced for a specific date by the Administrative Data First model, will be subject to continual revisions with each update from new departmental connections, generating a new estimate, including that of "***the figure most recently reported by the Government Statistician***". While this ad hoc approach may appear to have been crafted to meet the needs of the Boundary Commission, there are other contractual uses of population censes figures where the ongoing revisions will generate issues in commercial and other contracts.

- The reference to the most recent census is replaced by a concept of "***the figure most recently reported by the Government Statistician***"
- General electoral population is "***figure most recently reported by the Statistician***"
- Rather than a concept of census date, a less precise specification of "the reference date of the most recently published census." The meticulousness with which meshblock level information recorded at the same single point in time for everyone has been the mainstay of the boundary setting process. There will be a reduction in the quality of the counts and other meshblock level detail contained from the Administrative Data First model, as it will have been gathered across a plethora of previous periods. Because the IDI overwrites past values at each iteration, it will not be possible to evaluate the effect of choosing a particular date, compared to the previous rule.
- Stats NZ is moving to Machine Learning methods to predict usual residence address. ("*Predicting usual residence address from-admin data in the 2023 census")*. The outputs from algorithms in such methods are not auditable as are "Rules Based" methods or traditional data models[
- The Bill will redefine counting day and census day.

The new definitions expand the scope for significant arbitrary decisions to be made without objective professional scrutiny by experts and major users of statistics and are potentially challengeable because of their ad hoc nature. This contrasts with the rigid specifications explicit in existing legislation.

In moving from a de facto basis for collecting information from which the de jure population can be derived, as currently done by a Census of Population and Dwellings, the characteristics and location of people not actually present or temporarily in New Zealand will no longer be identifiable. This will affect children at boarding school, for example.

The Government Statistician will now have an unprecedented degree of discretionary authority of measures that were previously precisely defined in statute. There is no oversight body such as the UK Statistics Authority that oversees the application of judgements made. Nor is there is any requirement that either the role of Government Statistician or Deputy Government Statistician be filled by a person with qualifications and experience known to be appropriate for official statistics, yet the Bill places even more reliance on that than before.

2.7. The independence in methodology of future Government Statisticians is constrained by The Data and Statistics Act Amendment Bill

The Data and Statistics Act Amendment Bill rigidly defines the statistical properties of all future population statistics. This is because the Bill deliberately removes from the Government Statistician any flexibility in the implementation of innovation in statistical methods for future censuses. This is a methodological constraint that is inconsistent with the independence of the Statistician that is explicit in both the Data and Statistics Act 2022 and the Statistics Act 1975, and not seen in earlier statistics acts. These acts simply require that the Government Statistician conduct a regular Census of Population and Dwellings.

> 16 Independence of Statistician
>
> (1) The Statistician has the sole responsibility for deciding on—
>
> **(a)** the production of statistics by the Statistician, including the selection of data sources, concepts, definitions, methods, and classifications to be used; and
>
> **(b)** The timing, manner, and content of the dissemination and communication of those statistics

The Bill provides an endorsement of the recent (2025) Ministerial determination on methodology. It inhibits the independence of future Government Statisticians on matters that have in the past been devolved to the Government Statistician. It prevents the statistical strengths of the longstanding census model from continuing to evolve

with any ongoing technical, methodological and process advances that have always driven change in how each census builds on its predecessor.

Statistical legislation in New Zealand up to now recognises that each census is part of a continuous series of integrated statistical surveys that cover everyone in New Zealand. This integrated series of censuses enables long term trends to be identified, timely population estimates made for the years between censuses, along with the information required for essential projections of future population trends. These projections underpin predications and forecasts in many sectors of society.

2.8.Eroding the value of analyses using the Integrated Data Infrastructure (IDI)

Statistics New Zealand's Integrated Data Infrastructure (IDI) has become central to increasing public sector efforts to use evidence to inform policy making and evaluate social sector interventions in Aotearoa. Around half of studies using the Integrated Data Infrastructure rely on information only found in a Census of Population and Dwellings. Survey and Census data underpin the use of anonymised administrative data in three important ways.

a) Survey and census data is important because it fills in the picture for information that government agencies cannot collect as part of their administrative processes.
b) They provide the frame for understanding the administrative data. Data collected through administrative processes is a function of people's need to interact with government agencies. People who interact with the government a lot are well represented in the data, while people who interact less may be under-represented in the administrative data. Survey and census data, because they use a different sampling frame, allows us to triangulate against the administrative data in the IDI to see who is missing from the administrative data picture or under-represented.
c) With survey information available to identify biases in the administrative data it is possible to make use of the strengths of administrative data (its longitudinal nature) to inform planning. Without the ability to validate the administrative data picture against another sample frame decisions will tend to be based on and reinforce the blind spots and measurement errors administrative data.

2.9.The Bill retains the statutory invisibility of the IDI

The Integrated Data Infrastructure (IDI) which underpins the proposed changes and has been around for a good decade, is anchored in a de facto population register. It enables the joining together of personal records that historically have required legislative change to link. Neither the Integrated Data Infrastructure nor the existence of a de facto population register are recognised in the proposed legislation changes. This is despite

being regarded as vital for public trust in the activities of Statistics NZ by many submissions concerning the Data and Statistics Act 2022.

Advancing the collection of official statistics by extending the information obtained by a government department that has specific enabling legislation will connect the collection of confidential information for the purpose of official statistics with the compliance processes and associated penalties of that agency. There is no process to check on the potential for the overreach of the punitive nature of some compliance obligations. There is no legislation in New Zealand comparable to the Australian Data Availability and Transparency Act 2022, this means that the distinction between uses of the IDI for research, official statistics, delivery of public services, assessment of entitlements, surveillance and enforcement may vary with the assessor.

Ministers, recent Government Statisticians, and the Public Service Commissioner have moved Statistics New Zealand into the role of public sector data warehouse, intended for implementing ideas of social investment. The formal reporting about official statistics by the Government Statistician has become subservient to the Chief Data Steward activities in recent years, obscuring this shift in what Statistics New Zealand does. The continued production of trustworthy official statistics and access to them has become of secondary importance to this.

There are plans to place the Integrated Data Infrastructure on a cloud computing environment. Given that all the likely vehicles for this are likely to American, this means that this highly confidential information about the lives of all New Zealanders will be subject to the US Patriots Act 2001, which obliges any American company to provide whatever records it is in custody of, wherever in the world, to US Federal authorities should they seek it.

2.10. What key tests might the Select Committee use

New Zealand has up to now adopted international standards for statistical classifications, concepts and frameworks. We also have obligations for transparency in the integrity of official statistics as a member of the United Nations, the International Monetary Fund and the OECD.

The proposed changes would anchor New Zealand population, social and household economic surveys in a statistical frame anchored on reports about individuals, which is a major shift from the longstanding practice of surveying people as residents of the dwellings and households which are the prime statistical unit. This change means that the connections that individuals have with families, households and other communities will not be directly reported, but derived from the information people have had to give at some time in the past to a government agency (**See Insert 1**). The change will lead to an immediate discontinuity when comparing statistics prepared before and after the change, and this will affect all measures that use population statistics to calculate prevalence and incidence of characteristics.

The implications of this fundamental shift in statistical frames on how the results of social and economic surveys are integrated or post-stratified have been poorly explored.

There is insufficient transparency about the extent to which in the future information published about a particular date will in fact have been gathered across a plethora of previous periods, rather than the same single point in time for everyone. Complex derived measures will be affected by this change, such as the widely used indexes of deprivation.

The methodological characteristics unique to NZ Censuses of Population and Dwellings that should occur in any replacement data source include:

a) ***Guaranteeing the commonality of definitions across statistical sources and across key collection methods (unlike administrative records)***
    - With the Administrative Data First model there will be a loss in the breath of family information that is able to reflect social change
    - There will be limits to the connection of persons to the families and households that they are part of what their role in them is.
    - The current extent and history of ethnic analysis will not be retained
    - Connecting where people usually work to where they live will be of lower quality
    - Connection between income and occupation and qualifications will not be possible for all people at the same time

b) ***Continuing the existing rich array of small area information***
    - The capability to integrate economic, environmental and climate information with population statistics for small areas as needed for emergency management has become more urgent.
    - There is a continuing decline in the integrity of regional population projections needed for infrastructure planning, housing and climate change responses, and emergency management. This will be worsened by the shift to a population-based frame.
    - Statisticians must urgently collaborate with LINZ and other land-based organisations to utilise the potential to integrate all forms of area-based information (environmental, economic, climate) with population statistics for small places

c) ***Maintaining and extending commonality of methods with Australia, Canada and the UK to enable collective benefits from continuing innovation***
    - The shift from a household-based frame to a person-based one will lead to discontinuities in series at a critical time and remove any chance of returning to any collaboration with Australia on the timing of censuses.
    - The existing Census of Population and Dwellings provides the prime means for assessing the coverage across the population of all statistical and administrative data sources, both government and commercial.

- The capacity for the rapid adaptability of content will be lost with the Administrative Data First model, compared to that provided by the Census of Population and Dwellings (e.g. 1996 Disability question and survey)
- The Administrative Data First model will not be providing a vehicle for post survey validation of coverage of the many other statistical surveys
- The effect on ensuring adaptability to adverse situations, such as COVID, earthquake, flooding is weakened because of the diversion from investment in a geospatial frame that enables integration of data from many sources.
- Administrative concepts in the Administrative Data First model have yet to adopt the statistical definitions and standards that enable national, international and local comparisons, such as determined for population measures, and the system of National Accounts.

d) Expanding the use of administrative records for enumeration, imputation and quality assurance

- There are a range of ways of making better use of administrative records, while recognising their inherent limitations:

1. Australia, Canada, The USA, Ireland, and the United Kingdom use administrative data in the Census to correct for non-response or measurement error e.g. under coverage of total income but not to replace the Census.
2. Administrative records should be a major information source in the development of Propensity scores
3. Administrative records are vulnerable to policy change, which can be quite dramatic, as seen with the removal of overseas transaction records in 1985, and the removal of the requirement to complete an annual tax return from 2024.
4. Practicalities mean that administrative records cannot be certain to meet statistical standards as is exemplified by the inability to separate capital from operating expenses in GST returns, limiting their use in economic monitoring and as a source for national accounts.

e) Ensuring that comparisons between censuses at different times are not affected by changes to methods or operational practices

- Intercensal comparisons have been damaged by the poor management of the methodology and enumeration changes in the Censuses of Population and Dwellings of 2018 and 2023.
- The Administrative Data First model will result in individuals forming the key statistical frame for population statistics rather than households. There will be a break in the continuity of statistical series, affecting critical statistics in many areas including health, where the census has been essential to establishing population rates.
- Measures about families and household composition have to be reverse engineered from thin shards of information about how people with a common address are related, and the roles that they have within any household.

- Official statistics would now have two statistical frames that create new uncertainties when connecting the results of surveys that use household as the sample unit and frame. This will affect any post stratification required for commercial or community surveys to validate results.
- All censuses up to now obtain information on a de jure basis and identify usual residents from the questionnaires. The statistical effects of the change to ordinarily resident from usually resident not clear.

2.11. Government statisticians have poor understanding of what population statistics are used for

Official statisticians are extremely poor at keeping abreast at what the statistics they produce are used for – while in New Zealand the users of statistics are noticeably poor at getting their needs known. In short, the value of the system of population statistics is poorly recognised by all parties. There used to be well-established mechanisms, such as Reviews of Statistics and Advisory Boards, to help overcome these problems, but they were all abolished in 2015 by the then Government Statistician. Evaluating the fitness for purpose of the proposed changes by Statistics NZ by others has not been given the importance it should have given. This is why we cannot really know the value of that part of population statistics that we are destroying on behalf of New Zealanders.

We are in the midst of a population storm. Just one element is the huge outflow of young New Zealanders. This means that we have to better understand the labour market transformation that is taking place, and the generational consequences that are staring us in the face.

The Treasury remind us that over the next two decades, our taxable capacity will be short of what we need to fund existing public services by over 30 percent. The political consequences will mean a rewriting of the welfare state we now have, with consequent political turmoil if that is to be driven by political instincts rather than sound statistics. Family and community support mechanisms will have to play a larger part in our welfare system, yet they are not contained or measured in departmental records, which do not have the granularity, provenance and synchronicity that a periodic census provides.

A properly resourced independent expert review would be able to challenge, and independently confirm or otherwise the necessity and viability of the changes proposed.

3. THE PATHWAY TO A 2028 CENSUS OF POPULATION AND DWELLINGS

Over the past decade, there are a range of initiatives have not been given the weight that they should have had, compared to in Australia and Canada. Recent Government Statisticians have been fixated on the way that administrative records can underpin the

continuing evolution of the system of population statistics. These overdue key initiatives are:

a) ***Driving an ongoing effective connection with users*** across the population to understand how their needs were evolving. Fitness for purpose can only be evaluated if the planners of the population statistics system are known to understand those needs by those for whom population statistics are critical.
b) ***Expanding the connections of the Government Statistician with users*** rather than continuing to remove them. This intimacy when effective provides rich, highly informed feedback on the quality of official statistics that can often signal in advance where major problems are emerging.
c) ***Ensuring transparency and awareness of the strengths and limitations of administrative records*** in enhancing the quality, scope, costs and frequency of official statistics, and their limited capacity to adapt as statistical requirements change and uncertainties with quality management.
   By overrating the value of administrative records, Statistics New Zealand then undervalued the contribution of well-motivated field interviewers and enumerators up to the 2013 Census. For the 2018 Census of Population and Dwellings enumerators were obtained from labour hire firms on a piecemeal basis. They were poorly directed and without the rigorous training and support structures used until 2013.
d) ***Artificial intelligence and epidemiology*** offer methodological solutions that could identify and rate respondent contact difficulties and response challenges for each locality and small area prior to deploying the usual field force. The Australian Bureau of Statistics Propensity Index is a model that could be expanded for planning the deployment of enumerators should the long-term series of Censuses of Population and Dwellings be continued. The expertise within local government to apply local knowledge to enhance how such a model were it to be applied has also been very much underutilised.
e) ***Collaborating intensively on census developments*** with Australia and Canada. This should begin again, because of the obvious benefits we gain. The next Australian Census will take place on 11 August 2026. Statistics Canada will conduct its next Census of Population in May 2026. There is an opportunity to have an on the spot watching brief on either of these events, especially since neither of these countries have experienced such poor results as seen in NZ in 2018 and 2023. Australia[6] and Canada are taking a measured and more analytical approach to examining the options for taking the maximum advantage of administrative records. They are also maintaining the dwelling as the central

[6] www.abs.gov.au/about/data-services/data-integration/integrated-data/administrative-data-snapshot-population-and-housing

statistical focus of the collection and integration of statistical sources, whether they be administrative records, or from a statistical survey or census.

f) ***Developing a geospatial frame for New Zealand.*** Statistics NZ and LINZ must be required to collaborate in this overdue investment. The Parliamentary Commissioner for the Environment has expressed deep concern about the absence of investment by LINZ and Statistics NZ in a geospatial frame over the last decade. This has made NZ vulnerable because of the limitations we currently have in integrating environmental, population and other place-based knowledge. This failure will lead to a continuing decline in the quality of local authority and any other sub-national population projections, as economic change is likely to be the most important driver in population change in parts of New Zealand, overshadowing projected changes in births and deaths. Canada and Australia are investing in ways to improve knowledge of places of habitation, and their geographic characteristics.

g) ***Reversing the diminishing reliability of population projections and estimates for Local Authorities***. Methodologies should have been continually advancing for this. Births and deaths are now less significant drivers of population changes locally compared international and internal migration, and the local dynamism of economic structures. We have not recognised the wide range and future criticality of integrating with population statistics a wide variety of locality information. This now influences where people live or chose to stay more than demographic change. Little of that information will be found in the administrative records of government. Already, the half-life of local authority population projections is much less than that during the baby boom years, and declining.

h) ***Providing genuine access to statistical series and reports through the web*** and other publishing forms, in a manner that enables wide public use and meets the needs of expert users is essential. The Statistics NZ website is not a credible example for an agency that needs to show that it recognises how much technology can continually improve the range and ease of accessibility year by year. Bridget Williams Books publishes more printed material in a month than has been visible from Statistics NZ in the last decade.

i) ***Maintaining a long-term strategy for the content, methodology and operations*** of each integrated Census of Population and Dwellings. Long term continuity that enables comparison with previous results is balanced against ensuring that the content and methods evolve with the need of users and operational practicalities. The strategy set out by Statistics NZ[7] in 2012 was cast aside and has been replaced by ad hockery, with consequent loss of some of the benefits gained from the longest running statistical survey in New Zealand.

[7] Statistics New Zealand (2012). Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy. Wellington: Statistics New Zealand

The integration by design of censuses and other surveys allows us to capture not only shifts in what happens to people in particularly important age groups over time; to measure intergenerational change over a century or more; but also become aware of what is happening at the present time compared to the past. The system of population statistics that we now have provides us with reliable forecasts of our population in the future, compared to what we can forecast about any other aspect of our society, economy or environment.

j) ***Ensure that the quality of statistics can be assessed by users.*** This includes the statistics from the Census of Population and Dwellings, estimates and projections and basic counts for the smallest areas. We need to be driven by the needs of the future, not at the backward approach of planning with admin records that can be months or years behind in tracking peoples' characteristics.

4. Appendix. The use of sample surveys from departmental records has serious problems.

1) There will be a variable and generally incomplete coverage of the total population. This will make it impossible to get an overall proper random sample as individuals from different subpopulations will have different selection probabilities. As always, some people are hard to contact. However, it is appreciated that this problem will be the same for a general census but can be handled in other ways.
2) If a national total is required, all the individual estimates must be added so that the individual margins of error from the individual surveys will accumulate, However, the final error tis a bit less than their sum, but it will still be large. This will lead an inaccurate totals
3) Opinion polls are often used to compare subsets of data, for example percentages for preferred prime minister. This will suffer more from the same problem mentioned in (2) above in looking at differences. An example of this kind of problem is given below.

Methodological explanation

The most important aspect of a sample is the size of the sample n say. Estimates given by the news media are usually based on samples that are too small. To illustrate this kind of problem, we do a little statistical theory. Often, we wish to estimate the proportion p of people from a sample in a population with a given characteristic such as the proportion with diabetes. The estimate has a margin of error (ME) of approximately 2 times the square root of (p(1-p)/n)

For example, if p=0.2 (20%), n=1000, and the proportion of the population sampled is less than 5%, ME is 0.0247 (or 2.48%) which compared with 20% is substantial. When making a comparison, say 0.2 with 0.15 i.e. a difference of

0.05, then using the theory of comparing proportions from the same population (Seber, 2013, Statistical

Models for Proportions and Probabilities. Springer) the ME is about 0.0365, which is nearly as big as the observed difference! This shows that using subsets of data not based on a large population like an overall census can be problematic.

Len Cook and George Seber

15/4/206

https://lencook.co.nz/all