Statistics New Zealand
TATAURANGA AOTEAROA

Integrating New Zealand Census Mortality Study and New Zealand Longitudinal Census:

Privacy impact assessment

New Zealand Government

Citation
Statistics New Zealand (2014). *Integrating New Zealand Census Mortality Study and New Zealand Longitudinal Census: Privacy impact assessment.* Available from www.stats.govt.nz.

ISBN 978-0-478-42936-7 (online)

Published in October 2014 by
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

Contact
Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4610
www.stats.govt.nz

Contents

1 Purpose of this privacy impact assessment

This privacy impact assessment (PIA) provides a systematic evaluation of the privacy risks associated with integrating datasets from the New Zealand Census Mortality Study (NZCMS) and the New Zealand Longitudinal Census (NZLC). It outlines the processes for managing these risks.

The NZCMS comprise datasets of mortality records linked to the 1981–2006 Censuses and the NZLC datasets comprise linked census pairs covering the same period.

Data integration is when two or more unit record datasets, which were originally collected for different purposes, are linked together. Data integration can lessen the burden on respondents and lower production costs, but it can also raise real and perceived privacy risks. We need to consider these when determining whether to integrate datasets.

This PIA assesses both actual and potential effects on an individual's privacy due to proposed integration, and identifies how this risk will be managed. It assesses these risks against the criteria set out in our Data Integration Policy.

How this PIA relates to the Data Integration Policy

The Data Integration Policy defines our policy on integrating personal data. It applies to all integration of personal data that we undertake for statistical or related research purposes.

The policy states that data integration should be considered in the production of official statistics when the following principles are met.

Principle 1: The public benefits of integration outweigh both privacy concerns about the use of data and risks to the integrity of the Official Statistics System, the original source data collections, and/or other government activities.

Principle 2: Integrated data will only be used for statistical or research purposes.

Principle 3: Data integration will be conducted in an open and transparent manner.

Principle 4: Data will not be integrated when an explicit commitment has been made to respondents that prevents such action.

The policy is based on the United Nations Economic Commission for Europe's *Principles and guidelines on confidentiality aspects of data integration undertaken for statistical or related research purposes* (2009).

2 Proposed data integration

We will use the NZLC and the NZCMS datasets in this data integration project, and both are sourced from Statistics NZ.

The NZLC datasets contain five pairs of historic census data covering the period 1981–2006 (Statistics NZ, 2013). This provides longitudinal datasets spanning up to six census years.

The NZCMS datasets contain death records linked to the nearest previous census, and also cover the period 1981–2006 (Department of Public Health (nd); Blakely, Woodward, & Salmond (2000); Hill, Atkinson & Blakely (2002)).

Purpose and objectives

The primary purpose for integrating the NZLC and NZCMS datasets is to add the mortality information already linked in the NZCMS to the NZLC. The continuing development of the NZLC includes adding deaths as part of the resource. Adding the NZCMS information provides the links for those death registrations already linked in NZCMS without duplicating previous work.

Enhancing NZLC with death registration information will allow us to study the relationship between life course and mortality.

Life-course socio-economic influence on mortality, and pathways for intervention in the New Zealand context, is a gap in our knowledge of the population.

Mortality records linked with the historic longitudinal census dataset (linked records across the 1981–2006 Censuses) will fulfil four research aims.

- To test which life-course hypotheses best explain associations between socio-economic status and mortality: accumulation, sensitive period, social mobility, or instability.
- To test whether social and cultural capital protects against socio-economic risk.
- To assess ethnic disparities in mortality and test whether these are explained by the greater experiences of long-term harsh and unstable environments among some ethnic groups (eg Māori, Pacific).
- To test life-course hypotheses among siblings discordant on socio-economic risk, and among siblings discordant for social and cultural capital (thus controlling for family background factors).

The analyses outlined above are not an exhaustive list of all analyses that will be completed using the integrated datasets. It is likely that the results of these first analyses will raise additional research questions, which will lead to future research. Future uses of the integrated data will be consistent with the overall purpose of the project as described in this PIA.

This research is methodologically innovative and is important for an improved understanding of ethnic inequalities in New Zealand. It will also apply to the analysis of health disparities, housing concerns, and geographic diversity.

Both the NZCMS and NZLC are anonymised unit record datasets available for approved statistical research purposes within Statistics NZ's Data Lab. As data custodian of these microdata resources, we also store files of linked identifiers separately. These files enable the integration of the datasets by direct linking on the identifiers used originally to create the NZCMS and NZLC datasets, respectively.

The additional privacy risks associated with the linking of the two data sources are small. The protocols in place for each source are congruent. The protection of confidential information and privacy of individuals is adequately covered.

The integrated resource will be available to Statistics NZ researchers and also, in the secure Data Lab environment, to accredited external researchers for approved research projects. This will be a permanent research information resource and will be managed according to existing and future protocols to ensure the security of the data.

We have prepared privacy impact assessments for the NZCMS updates and the NZLC, respectively. The NZCMS has also received ethical approval from the Central Regional Ethics Committee.

Expected benefits

The outcome of the integration will be of substantial benefit to both Statistics NZ and the wider research community in that it contributes directly to intended future developments of the NZLC. Future development of the NZLC includes adding links to death registrations and to cause of death. This project will provide links for those deaths that are already linked to the NZCMS and will thereby minimise duplication of work. This enhanced information source will be of major value to future research into health, mortality, and life course analysis.

Linking the NZCMS to the NZLC will immediately improve our ability to provide support to longitudinal research on mortality and social equity. An example of the value of this capability is the research project of the Centre of Methods and Policy Application in the Social Sciences (COMPASS) of Auckland University. This project has been funded recently by the Health Research Council of New Zealand in the 2014 funding round. The content of the research project titled ‘Life-course predictors of mortality inequalities’ was prepared by COMPASS in collaboration with the Department of Public Health, Wellington School of Medicine, University of Otago. One of the benefits that will come out of this example is that by using the data, we will have an improved understanding of health dynamics that will benefit public health policy and delivery.

Linking the 1981–2006 NZCMS datasets to the NZLC database creates a unique opportunity to undertake research on the life-course socio-economic influences on mortality in New Zealand. It will contribute directly to a better understanding of a range of Tier 1 statistics, such as life expectancy.

The suggested integration of NZCMS and NZLC contributes towards the broader strategies of Statistics 2020 Te Kāpehu Whetū by addressing the need for longitudinal datasets, making better use of existing data, and responding to identified information gaps. There are few longitudinal studies in New Zealand due to the high cost and administrative difficulty of maintaining stand-alone large-scale longitudinal surveys, the resource requirements, and the high mobility of New Zealanders. The NZLC is unique in that it contributes such a longitudinal source developed from existing data with small marginal cost. Linking the NZLC and the NZCMS further enhances this as a longitudinal data resource.

3 Conclusion and recommendation

The proposed linkage between the NZLC and NZCMS databases integrates existing NZCMS datasets with the NZLC datasets. The linkage provides immediate analytical capability to the NZLC in that it links death information from the NZCMS to the NZLC dataset. This will contribute to one of the approved enhancements of the NZLC, which is the inclusion of death registration information in the NZLC. Future developments may include the 2013 update of the NZCMS (covering deaths 2013–18) linked to the 2013–2006 census pair in the 2013 update of the NZLC.

This proposal will add value to the NZLC without any additional risk. The existing and the proposed future developments provide a major platform, with ongoing relevance, for extensive research across a number of key demographic, geographic, social science, and health fields. The proposed research by COMPASS represents an initial significant example of this resource.

Both the NZCMS and NZLC are anonymised unit record datasets available for approved statistical research purposes within Statistics NZ's Data Lab. As data custodian of these microdata resources, we also store files of linked identifiers separately. These files enable the integration of the datasets by direct linking on the identifiers used originally to create the NZCMS and NZLC datasets, respectively.

The security protocols in place for each source are consistent with each other and with standard Statistics NZ practices to protect the confidentiality, security, and privacy of respondents.

For this project we will apply the following practices.

- We will use the integrated data only for statistical or research purposes.
- The linking will be carried out by Statistics NZ employees working on Statistics NZ premises. Our employees are bound by the Statistics Act 1975 to never disclose details about respondent information. We also have security systems that allow only approved individuals to access the premises.
- Names and addresses are not available in the existing source datasets and will not be in the final integrated datasets.
- We will store the final integrated data in a secure environment. The principles of the Privacy Act 1993 outlined below are fully covered.
- We will not publish any information that could identify an individual. All outputs will be confidentialised in accordance with Statistics NZ practices.

With these mitigation strategies in place, the proposed data integration has a very low residual risk that is outweighed by the substantial benefits it is likely to bring. The major benefit will be a strong evidence base to make well-informed decisions on improvements to socio-economic policy areas and statistical advances in estimation and longitudinal analysis. Additional potential will contribute to a better understanding of factors that may improve data collection.

Recommendation for approval

In light of the assessment set out in this document, we recommend that the Government Statistician approve this proposal. The proposed integration will offer value for money through the ability to produce new information from existing data sources with minimal additional cost and no additional respondent burden. The cost of development is offset by the value to be derived from the enhancement of longitudinal census data and the new uses to which it will be put. Substantial investment exists in the existing NZCMS and NZLC datasets and this project adds further significant value to this investment.

We have assessed the additional privacy risks associated with the linking of the two data sources as small. The protocols in place for each source are consistent. The protection of confidential information and privacy of individuals is adequately covered.

4 Legislation relevant to source data collections

We must comply with legislation relevant to source data collections (such as the Tax Administration Act 1994 or the Health Information Privacy Code 1994).

Statistics Act 1975

Statistics NZ operates under the Statistics Act 1975. As such, all data integration activities we undertake must adhere to the requirements of this Act.

Sections of the Statistics Act 1975 that are of specific relevance to data integration activities include the following.

Section 3: Official statistics and coordination

Official statistics must be collected to provide the information that is required by the New Zealand Government, local authorities, and businesses to help make policy decisions. Official statistics must enable the government, government departments, local authorities, and the general public to understand economic, social, demographic, and other information of interest.

Section 4: Classes of official statistics

Statistics NZ can require any person to provide information for a set number of categories. This data integration proposal contributes to:

(a) population and dwellings, migration (internal and external), vitals, and other demographic and socio-economic matter areas

(b) health, welfare, and morbidity

(e) matters relating to the social and physical environment

(i) household (including family) characteristics, conditions, and activities.

Section 14: Duties of the Government Statistician

Subsection 14(c) requires that Statistics NZ keeps the Minister of Statistics informed of the department's statistical projects, by explaining their purpose, scheme, methodology, and usefulness.

Subsection 14(d) requires that Statistics NZ does not collect any information without the written permission of the minister.

Section 15: Independence of the Government Statistician

The Government Statistician has the sole responsibility for deciding the procedures and methods used to produce statistics. This includes data integration.

Section 37: Security of information

Data can only be accessed for statistical purposes and there are strict rules about how information is to be kept confidential.

Section 21: Declaration of secrecy

All personnel who access respondent information must keep this information secret for their lifetime.

Privacy Act 1993

The Privacy Act 1993 contains 12 privacy principles that set out how agencies may collect, store, use, and disclose personal information. Personal information relates to identifiable, living individuals.

All data integration activities we undertake must adhere to the requirements of this Act. The Act aims to uphold public trust in government by regulating the practice of data matching in the public sector.

Sections of the Privacy Act 1993 that are of specific relevance to data integration activities include the following.

Principle 1: Purpose of collection of personal information

Agencies must have a lawful purpose to collect information and it must be necessary to fulfil this purpose. Statistics NZ has a lawful purpose under the Statistics Act 1975 to integrate datasets to produce statistics.

Principle 2: Source of personal information

This principle requires information to be collected directly from the source. Statistics NZ meets the exceptions that allow the organisation to use data collected by other agencies (eg from Inland Revenue). This is because the data will be used for statistical or research purposes [Principle 10 (f)(i)] and it will not be published in a way that could identify the individual [Principle 10 (f) (ii)].

Principle 12: Unique identifiers

This principle requires that an agency does not assign a unique identifier that has been assigned to an individual by another agency (eg an IRD number assigned by Inland Revenue). Current practice at Statistics NZ is that unique identifiers assigned by other agencies are used at the time of construction of the integrated datasets, but are not retained permanently in the datasets. The original unique identifiers for persons represented in the original datasets are removed. The final datasets contain (anonymised) unit records identified only through a Statistics NZ unique reference.

Public Records Act 2005

The Public Records Act 2005 sets out the requirements for creating and maintaining adequate records of the business of public offices, including Statistics NZ.

All source data for integration and the resulting integrated datasets must be retained or disposed of according to the requirements of the agreement between Statistics NZ and the Chief Archivist, authorised under section 20 of the Public Records Act 2005. The following disposal schedules govern what data must be preserved, and what can be destroyed when no longer needed for statistical purposes:

- Statistical data, documentation, and metadata disposal schedule (DA379)
- 2006 Census of Population and Dwellings disposal schedule (DA439)
- 2011 Census of Population and Dwellings disposal schedule (DA440)
- Statistical schedules disposal schedule (DA271)
- Administrative Data disposal schedule (DA272)
- 2013 Census of Population and Dwellings disposal schedule (DA563).

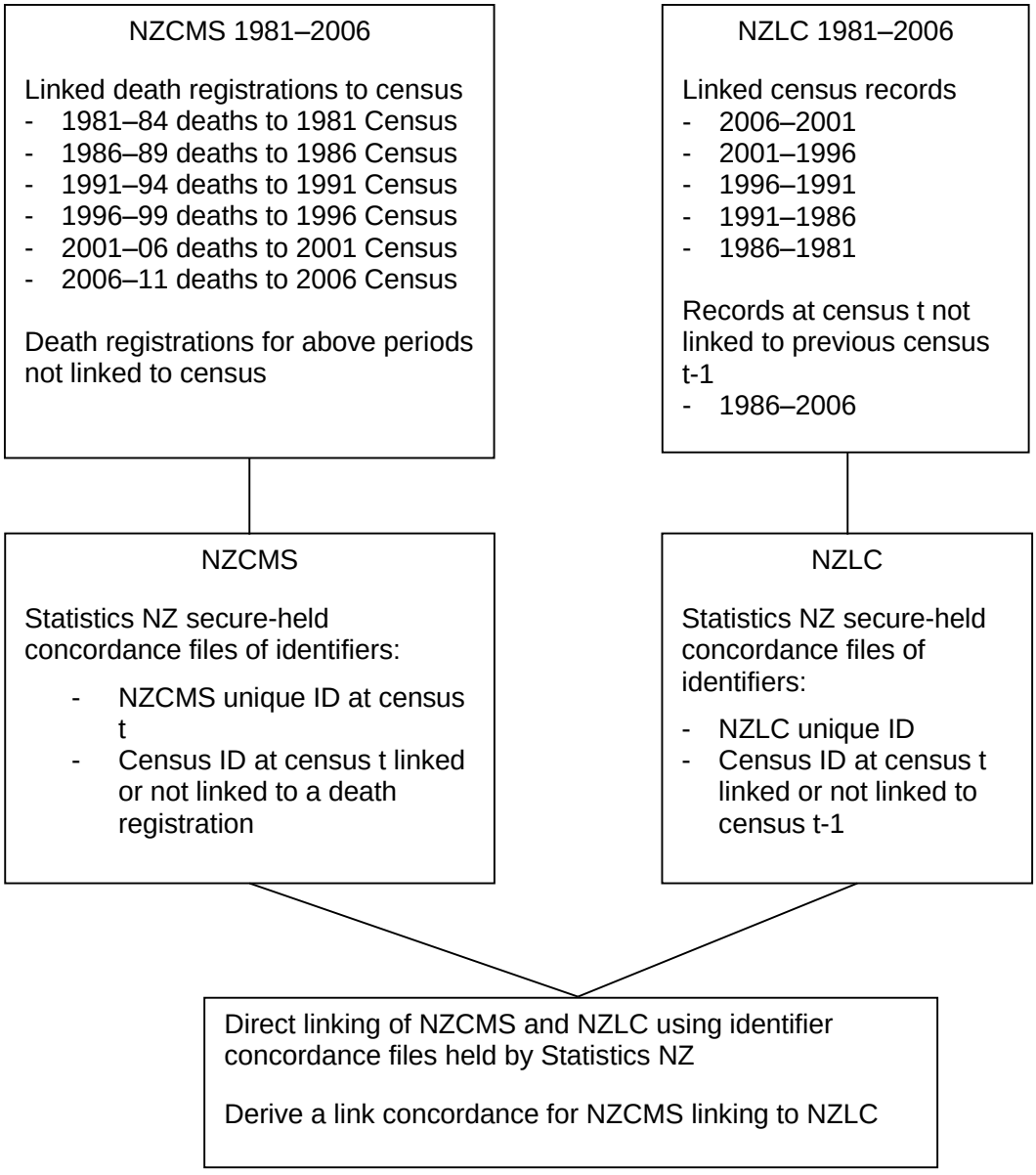
5 Data integration information flows

The proposed integration of the NZCMS and the linked census pairs in the NZLC does not require the collection of data from other agencies or the collection of any new data. Statistics NZ has custody of both data sources.

The proposed integration of the NZCMS datasets to the linked census pairs in the NZLC database will be a direct linking process using secure files of original census identifiers, and the corresponding unique identifiers created to anonymise the datasets for microdata access in the Data Lab. Figure 1 below describes the information held by the two databases and the identifiers needed for the direct linking process.

Figure 1

Overview of data flows for the integration process NZCMS and NZLC, 1981–2006]



Following the linking of the two datasets, we will remove all unique source identifiers. We will store data excluding unique identifiers in a secure area, which can be retrieved only by those with approved access. Access will only be provided for statistical analysis of the linked dataset.

Our standard security measures and protocols will govern the management of the data. We are required to comply with the confidentiality provisions of the Statistics Act 1975 and the Security in the Government Sector (SIGS) protocols.

We have well-established policies, procedures, and systems in place to ensure adequate measures of physical and electronic security.

6 Privacy analysis of proposed data integration

This chapter determines whether the proposed data integration complies with our Data Integration Policy.

The policy exists to ensure that impacts on privacy are minimised when we consider proposals to create statistical and research outputs using data integration.

In addition, compliance with this policy ensures that the proposed data integration is in-keeping with important legislation, including the Privacy Act 1993, the Statistics Act 1975, the Public Records Act 2005, and any other legislation relevant to the source datasets.

Stakeholders consulted

For this analysis we consulted the following stakeholders.

- Hamish James, Manager, Standards and Design, Statistics NZ
- Heather Jones, Senior Advisor, Strategy Performance and Privacy (Privacy Officer), Statistics NZ
- Gareth Meech, Manager, Census Statistics, Statistics NZ
- Vina Cullum, Manager, Population Statistics, Statistics NZ
- Paul Brown, Statistics NZ's Respondent Advocate, Statistics NZ
- Vince Galvin, Chief Methodologist, Statistics NZ
- June Atkinson, Researcher, Department of Public Health, University of Otago, Wellington
- Kate Sloan, Research Manager, Department of Public Health, University of Otago, Wellington
- Tony Blakely, Research Professor, Department of Public Health, University of Otago, Wellington

How the proposed integration complies with our data integration principles

We will consider integrating data to produce official statistics and related research only when all four of the following principles are met.

Principle 1: The public benefits of integration outweigh both privacy concerns about the use of data and risks to the integrity of the Official Statistics System, the original source data collections, and/or other government activities

a) How will the proposed data integration produce or improve official statistics and/or statistical research?

We will use the data from this data integration project for research and statistical purposes only, and will not use them for administrative purposes. Any amendments we make to the data during processing will be for statistical purposes only.

The proposed integration of the NZCMS and the NZLC will provide a new and rich evidence base for the analysis of changes in aspects of life course events over five intercensal periods (six censuses). This will include source information to derive measures of the consistency of the statistical measurement of social parameters over

time, and how this contributes to an understanding of socio-economic processes, including influences on mortality in New Zealand.

b) What are the expected public benefits of the proposed data integration?

The public benefit of the proposed data integration is the substantial improvement of the evidential base for a wide range of policy and research interests. These include public health research, for example in the area of mortality differentials and socio-economic inequalities. The proposed integrated data source enables, for the first time in New Zealand, a resource for analysing real cohorts over a long period to develop a scientific understanding of the drivers behind these differentials and inequalities.

c) What potential privacy concerns, both real and perceived, are associated with the data integration?

We have not received any adverse public reaction over privacy issues in any of our previous data integration projects, including the integration of census microdata across previous censuses (NZLC) and the integration of death registrations to census (NZCMS). Separate privacy impact assessments were prepared and approved for the data integration projects represented by the NZLC and NZCMS. Given that the proposed project involves similar integration to what we undertook previously, it is unlikely that this data integration will be perceived as an unreasonable invasion of privacy.

However, the proposed integration involves several potential privacy risks. These are discussed below.

Reuse of census data and death registration data. The proposed linking will involve reusing data that was provided as part of the census and as administrative registration of deaths. Respondents may believe this data is being used for a purpose other than what it was originally intended for.

Linked data may mean greater damage in the event of a breach. Integrating two or more datasets allows for greater information to be obtained about individuals than by using any single dataset. However, this also means that a breach (eg a leak, hack, intentional or unintentional misuse) of integrated data may be more harmful because more information may be released about individuals. Privacy breaches can harm those about whom the information is released, and may also harm our reputation and reduce trust in the ability of the public sector to protect personal data.

Some individuals may object to linking census data with NZCMS data. Specifically, their concerns may include:

- linking is privacy intrusive because it allows the government to know a lot of information about individuals
- security measures are not adequate to protect their personal information, and this information may be leaked or misused.

Adverse public perception of the proposed linking could result in individuals becoming distrustful of government data collection. As a result they may, over time, become less willing to provide information about themselves to the government.

d) What are the possible risks and potential benefits to the integrity of the Official Statistics System, the source data collection, and other government activities?

The proposed work will have several potential benefits for the Official Statistics System (OSS) and other government activities, including:

- greater efficiencies and reduced costs in developing the NZLC
- improved ability to provide accurate and timely information about the nature and timing of changes in New Zealand's population, with benefits for a wide range of components of the OSS, including population estimation and projection

- increased value of the census by demonstrating how it can be used to provide real cohort analysis of social and economic changes and for policy development
- strengthened Statistics NZ position as a responsive, flexible, national statistics office able to develop and adopt new methods in response to a changing environment
- reinforced Statistics NZ role as leader of the OSS, by advancing methods of data collection and estimation.

The major risk to the OSS and other government activities is that of adverse public perception. If individuals perceive that the proposed linking intrudes on their privacy, or that their information is not being kept secure and confidential, they may become distrustful of government data collection. As a result they may, over time, become less willing to provide information about themselves to the government. This will have a negative impact on the quality of information that we are able to produce about the New Zealand population.

However, if the risks discussed above are well managed and mitigated, the risks to the OSS, the census, and to the NZLC are minimal.

e) Will the data integration be reviewed? If so, what will the review period be?

The linking process of the NZCMS to NZLC will undergo review. The final linked datasets will go through a standard checking process for release to the Data Lab as outlined by our Methodological Standard for Microdata Access. We will also review the outputs we release from the integrated datasets to ensure they meet the requirements of this PIA. Ongoing review is part of the maintenance of the NZLC.

f) What information is required for linking the datasets together?

The information required to achieve the proposed linking comprises the census person identifiers, as these are common to both data sources, and key demographic variables, such as date of birth and sex to verify accuracy of linking. . In each case these datasets have been anonymised and records have been assigned a unique randomly generated reference number.

The data custodian (Population Statistics Unit) of the NZCMS and NZLC has custody of securely held files of the concordances between the original identifiers in census and the record numbers in the NZCMS and NZLC datasets. These concordances enable the linking of the NZLC and the NZCMS and are required to achieve the linking.

g) What policies and practices will be in place to protect the data from unauthorised access?

Handling data that identifies individuals and linking individuals across datasets is normal business for us, and we have standard processes and procedures to ensure data security and confidentiality. The proposed data integration will be conducted according to our standard policies and practices to protect respondents' privacy, confidentiality, and security.

- Linking will be carried out by Statistics NZ employees on Statistics NZ premises. We are bound by the Statistics Act 1975 to never disclose any details about respondents' information.
- We will restrict access to the concordance files identifying original personal identifiers and the records in the integrated datasets to approved Statistics NZ staff and will store these in a separate secure area. The NZLC data custodian will manage and audit access.
- We will remove all original personal identifiers in the census and death registration data from the integrated datasets after linking is complete and concordances stored on a separate secure server.
- Our premises have physical security systems that restrict entry to authorised individuals. Visitors are subject to strict regulations, including registration with

reception, supervision while on the premises, restricted access, and other procedures to ensure their activities are confined to legitimate business.
- Access to our information technology systems and databases is password-protected and all data stored on secure servers.
- We have a Security Office that actively audits and reviews these and other security processes, and identifies and addresses emerging security risks.

h) Do the outcomes of the data integration justify the impact on individuals' privacy?

The outcome of the data integration is a longitudinal dataset covering life courses from 1981 either to the death of the individual or to the latest census in which the individual is linked. This enables us to analyse the whole population or subgroups within the population (such as Māori, children of single parent families, immigrants) which we would otherwise have difficulty sampling or would be cost prohibitive.

We will use the findings from the proposed integration to inform socio-economic life course factors contributing to the patterns and timing of mortality.

Over the medium to long term, the integration will have a range of benefits.

- The dataset can be used to understand the impacts of social policies and real world events and in turn provide advice for improved social and economic policy.
- The information obtained will provide an evidential base for population change from perspectives previously not possible.
- Potential to provide rates of change with greater accuracy and reliability to population estimates and projections, leading to greater confidence population estimates and projections by external stakeholders.
- Stakeholders will have increased reassurance around the quality of analysis of subpopulations of high policy interest.
- The development of a longitudinal population database with no additional respondent burden.
- Increased value of the census by demonstrating how it can be used to provide cost savings and improve efficiencies across a range of new uses.
- Reinforced Statistics NZ role as a leader of the OSS, by maximising the value of data resources.

There are some privacy risks associated with this work, including:

- the re-use of census and death registration data for another purpose
- linked data means greater damage in the event of a breach
- adverse public perception about privacy, security, and confidentiality.

However, we can mitigate most of these risks to a large extent by using existing practices to protect respondent privacy, security, and confidentiality (these mitigations are discussed in detail in chapter 7). In addition, we will strictly control access to the integrated datasets as outlined in chapter 7.

Any risks that remain are outweighed by the substantial benefits this work is likely to bring.

Principle 2. Integrated data will only be used for statistical or research purposes

a) What are the statistical and related research purposes of the proposed data integration?

The data integration proposed in this project will be used for research and statistical purposes only, and data will not be used for administrative purposes. Any amendments to the data during processing will be for the statistical purposes of the project.

We will use the resulting dataset as a longitudinal information source for studying population change, and for analysing this change to develop strategies to improve well-being and for policy formulation. This includes, for example, investigating how life-course socio-economic influences mortality and what public health initiative might be developed to modify inequalities. This has not previously been possible in the New Zealand context, with the result that underlying scientific evidence has not been able to contribute this dimension to policy development.

This data integration project has four immediate research aims.

- To test which life-course hypotheses best explain associations between socio-economic status and mortality: accumulation, sensitive period, social mobility, or instability.
- To test whether social and cultural capital protects against socio-economic risk.
- To assess ethnic disparities in mortality and test whether these are explained by the greater experiences of long-term harsh and unstable environments among some ethnic groups (eg Māori, Pacific).
- To test life-course hypotheses among siblings discordant on socio-economic risk, and among siblings discordant for social and cultural capital (thus controlling for family background factors).

Further research questions are likely to follow from this work.

b) What is the expected long-term value of the data that will result from the integration?

The NZLC will have value as an enduring research tool. The NZCMS series has provided the basis for a significant number of theses on health, produced a large body of published research, and contributed to a range of health policies. The integration proposed here has greater potential in that it extends the cohort information available (the NZCMS links only to the previous census, whereas the NZLC links across censuses and has a much greater time-depth). This will enable health research in areas requiring this greater time depth. The NZLC is not limited though to health research, but will also have long-term value across OSS topics of population interest.

Developing the NZLC includes extending to future censuses and the ongoing linking of birth and death registrations. This proposed integration contributes to this objective by including existing links in the NZCMS.

We will retain the integrated data from this proposal and store them according to our data retention protocols. We will retain the integrated datasets of the linked census pairs and the NZCMS datasets, including the securely held link concordances of identifiers.

c) What policies and practices are in place to ensure the data is used only for appropriate research?

We will manage access to the integrated data through the microdata access protocols in force at the time of any application for access to the data. This application for access requires full disclosure of the research proposed. Access is also limited to approved researchers in approved environments, with specific constraints on how data is controlled for release. These protocols will ensure all access is for research purposes only.

Any research that will involve the integration of other sources not included in the NZLC would require a new PIA.

d) What confidentiality practices will be in place to ensure that data can only be used for statistical purposes?

Any statistical information released, whether in the form of official statistics or research outputs, should meet our confidentiality requirements, including the confidentiality rules for the release of aggregated longitudinal census data. These rules require that output

which might identify the characteristics of individuals is withheld, and in particular output in which there are both small geographies and small populations is restricted.

Longitudinal datasets introduces additional confidentiality risks in terms of identifying individuals. However, the probabilistic linking methods that we have used to create the census pairs and the NZCMS datasets of death records linked to census introduce additional matching uncertainty which further enhances confidentiality.

We will confidentialise all outputs from the proposed work (such as research reports, conference presentations) according to our policy to ensure that no information is published that could identify an individual either in fact or by implication.

Principle 3. Data integration will be conducted in an open and transparent manner

a) How will information about the proposed data integration be made publicly available?

We will inform the Office of the Privacy Commissioner about this project. We have also informed him of the NZCMS and NZLC development. In keeping with our protocols for data integration projects, information about the project will be placed on our website, including an explanation of the project and this PIA.

Our website also contains generic information on data integration.

b) How will information about the statistical results or research outcomes be made publicly available?

The findings from the proposed research may be made publicly available through:

- published research reports and articles, including reports on our website
- published policy analysis across government and non-government agencies
- conference presentations
- publication in peer-reviewed journals
- deposition as research theses in university repositories.

Principle 4: Data will not be integrated when an explicit commitment has been made to respondents that prevents such action

a) What commitments, both actual and implied, have been given to respondents about how their personal information will be used?

The Privacy Act 1993 (Information Privacy Principle 3) generally requires that agencies collecting personal information directly from individuals ensure that those individuals are aware of the purposes of collection. However, an exemption is made for information that will be used for statistical or research purposes which will not be published in a form that could identify the individual concerned. This project fulfils this exemption, so we do not consider it necessary to obtain consent directly from individuals. In the case of deceased people, consent cannot be obtained. However, it is still worth considering the issue of informed consent because the project potentially impinges on the privacy of individuals and the risks should be weighed up against the benefits of the research.

Subjects of the project may be aware that their information will be used for statistical purposes. Census forms state that data will be used only for statistical purposes. A related issue is whether people are aware that the term ‘statistical purposes’ includes data integration, as this is not explicitly stated on forms. Information about this project including this PIA will be available on our website.

7 Summary of privacy risks and mitigations

The significant privacy risks to Statistics NZ that have been identified are summarised in table 1.

Table 1

Privacy risk	**Risk mitigated by**	**Assessment of residual risk**
Breach of **security** involving death registrations and longitudinal census records under the custodianship of Statistics NZ or once the data is released.	Standard Statistics NZ physical and data security practices. Access to data by researchers according to: • microdata access protocols • existing release processes.	Low
Breach of **privacy** involving death registrations and longitudinal census records under the custodianship of Statistics NZ or once the data is released.	Names and addresses not used. Restricted access on 'need to know basis'. Integrating variables needed for statistical and research purposes, and using them for statistical purposes.	Low
Breach of **confidentiality** involving death registrations and longitudinal census records under the custodianship of Statistics NZ or once the data is released.	Applying Statistics NZ confidentiality rules for longitudinal census data.	Low
Adverse public perception or rejection of the legitimacy of the proposed use of this information. (Note: any adverse public reaction might not only impact on this project but could also impact on census data, the reputation of data providers, and the ability of government agencies in general to link data from unrelated sources in the future.)	Adherence to Cabinet directive [CAB (97) M31/14]. Office of the Privacy Commissioner informed about the project. Transparency about project objectives and processes. Consultation with key stakeholders. Ethics committee approval required for use of NZCMS and NZLC integrated data by researchers. Identifiable data not used during integration. Compliance with the Statistics Act 1975 and other relevant legislation.	Low

Privacy risk	Risk mitigated by	Assessment of residual risk
	Rigour in the privacy impact assessment.	
Questions raised on the use and retention of unique identifiers.	Adherence to Statistics NZ policies regarding use and retention of unique IDs.	Low
Use of the integrated data for other than statistical research purposes.	Access for specific genuine research purposes only. Minimum dataset provided to achieve the objectives of the research. All access subject to microdata access protocols and approval by Government Statistician. Additional use of the data subject to formal approval processes by Statistics NZ.	Low

8 Compliance mechanisms

We will manage the data according to our strict confidentiality and security protocols. We have audit procedures in place to monitor compliance.

All access to files that could enable the identification of individual records in the original resources to those in the resulting datasets will be restricted and stored separately. We will retain these files for future integration development, but researchers will have no access to this information. Only approved data custodians will be responsible for controlling access to this information.

References

References

Blakely, T, Woodward, A, & Salmond, C (2000). Anonymous linkage of New Zealand mortality and census data. *Australian and New Zealand Journal of Public Health, 24 (1):* 92-95.

Department of Public Health (nd). New Zealand Census Mortality Study (NZCMS), Overview of NZCMS. Available from www.otago.ac.nz.

Hill, S, Atkinson, J, & Blakely, T (2002). *Anonymous Record Linkage of Census and Mortality Records: 1981, 1986, 1991, 1996 Census Cohorts: NZCMS Technical Report No. 3.* Wellington: Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago.

Statistics NZ, (2013). Linking censuses: New Zealand longitudinal census 1981–2006 Available from www.stats.govt.nz.

Statistics NZ, (updated June 2014). Census. Available from www.stats.govt.nz/Census.aspx.

Further reading

Statistics NZ, (nd). Information about deaths data. Available from www.stats.govt.nz.

Statistics NZ, (nd 1). Microdata access protocols. Available from www.stats.govt.nz.

United Nations Economic Commission for Europe (2009). Principles and guidelines on confidentiality aspects of data integration undertaken for statistical or related research purposes. Available from http://live.unece.org.

Appendix 1: Variables to be used for statistical and research purposes

Variables for data integration

The data integration process will require access to each relevant unit record dataset prepared by the NZCMS and NZLC. The data integration is a direct linking process using existing link concordance files containing concordances of unique NZLC ID to Census ID at census, for each record linked or not linked to the previous census. The NZCMS contains links to deaths for those who died at ages less than 75 years for the three years following each of the 1981, 1986, 1991, and 1996 Censuses, together with links to deaths for all ages for the five years following the 2001 and 2006 Censuses.

Variables for analysis

The integrated datasets will comprise variables (Statistics NZ, updated June 2014) describing:

- dwelling, household, family, extended family, personal, and geographic information available at the 1981–2006 Censuses
- demographic information on the deceased available in death registrations for the periods 2006–01, 2001–1996, 1996–1991, 1991–1986, and 1986–1981
- cause of death information at an aggregated level as available in the NZCMS.

Appendix 2: Variables to be removed from statistical datasets

We will remove the original unique personal, dwelling, and family identifiers for the source census records and the original death registration identifiers from all statistical datasets.