

United Nations Statistical Institute for Asia and the Pacific

Seventh Management Seminar for the Heads of National Statistical Offices in Asia and the Pacific, 13-15 October, Shanghai, China

New Zealand Country Statement

1. Main uses of administrative data for statistical purposes

a. Established statistical series

New Zealand has a long history of producing demographic, social and economic statistics from administrative records.

Key demographic series derived from administrative data cover:

- births, deaths and marriages
- marriage dissolutions
- international migration
- abortions
- population estimates and projections

Regular social statistics outputs produced from administrative data span the areas of health, education, welfare, income and justice. Key series include:

- hospital discharges, cancer registrations, causes of death, workplace injuries
- educational enrolments, courses of study, qualifications
- recorded crime, convictions and sentences, prison inmates
- income sources and distribution, welfare benefit recipients

Administrative data provides the foundation for the production of a wide range of economic statistics including:

- building consents
- merchandise trade

b. Creation of integrated datasets

A number of recent initiatives in the use of administrative data for the production of official statistics in New Zealand have involved the integration of administrative data (with other administrative data or survey data) to fill a specific information need or increase the explanatory power of an administrative collection. Some examples include:

Linked Employer-Employee Database (LEED) - a longitudinal database created through linking together administrative data drawn from the taxation system with business data from Statistics New Zealand's Business Frame to measure labour market dynamics and provide insight into the operation of the labour market.

The success of LEED has led to initiatives to expand the database. One of these involves the addition of data on benefit type for individuals receiving an income-replacing main benefit. The expanded database will provide valuable insights into the labour market outcomes of different types of beneficiaries. A further initiative which is being investigated, involves the integration of tertiary students enrolments and completions data with LEED data to provide information on the employment outcomes of tertiary education and training. In the longer term, we plan to investigate the feasibility of adding key socio-demographic variables to the database to enhance its analytical potential.

Student Loans and Allowances Integrated Database - integrates administrative data from three agencies (Ministry of Social Development, Ministry of Education and the Department of Inland Revenue) to provide information about student loan borrowers. The longitudinal database provides insights into student loan borrowing, repayment of student loans, and the educational characteristics and post-study income of students who have participated in the Student Loan Scheme or received a student allowance.

New Zealand Census Mortality Study - links mortality records with Population Census records to provide information on the socio-economic correlates and trends in mortality. Following the success of this study, another study linking records from the cancer register to the census has been undertaken. The resultant database provides previously unavailable information on the socio-economic correlates of different types of cancer.

Prototype Longitudinal Business Database (LBD) - integrates longitudinal administrative and survey data at the firm level to support the analysis of business dynamics and firm performance without the need for new surveys or an increase in respondent load. The LBD is being used for microdata research about issues relating to business dynamics, economic transformation and the impact of government policies on business development. The prototype LBD is proving to be very useful for both microdata research and for improving our statistical processes. One of the issues we are facing is how to fund the development of a production model, which would be used to produce new and improved official statistics. Over the next few years we plan to maintain the prototype model, and update the data annually, while we work with relevant government agencies to secure funding for a production model. The development of a production model will benefit from the lessons we are continually learning as we use the prototype.

EFTPOS data - an experimental monthly Electronic Card Transaction (ECT) series has been developed for use as indicator of the change in the level of consumption expenditure and economic activity in general. This series is produced from aggregated administrative data generated in the process of administering New Zealand-based electronic transactions (EFTPOS) and covers all debit and credit card spending with New Zealand-based merchants. The data is currently received in a highly aggregated form. A future strategy for the series is currently under development.

Standard Business Reporting - a cross-government project has been initiated to reduce respondent load by providing businesses with the ability to report electronically to government using XBRL. The project will work with software providers to build facilities for reporting to government into standard accounting software. The first stage of the project will focus on an automated process for reporting financial information to five government departments.

c. Survey frames

Where administrative records provide good coverage, they are sometimes used to create, supplement or update survey frames. For example, Statistics New Zealand's Business Frame, which is a comprehensive list of private and public sector businesses and organisations in New Zealand, is continuously updated and maintained using tax registration and other administrative data sources. The frame is used as a statistical register for the various business surveys operated by Statistics NZ and underpins the quality and integrity of New Zealand's business statistics and national accounts programmes.

In the social statistics area, administrative records of migrant approvals for residence in New Zealand were used to create the frame for the Longitudinal Survey of Immigration in New Zealand (LisNZ).

d. Survey evaluation

Data from administrative records is often used in data confrontation studies to evaluate the quality of survey estimates (i.e. comparison of survey estimates with estimates from a related administrative program). For example, in the production of National Accounts, tax data is used to confront estimates from surveys and other administrative data sources.

e. Substitution or replacement of direct data collection

Statistics New Zealand is beginning to explore opportunities for using administrative records to replace data in surveys. Utilisation of administrative data in this way offers potential for reducing respondent burden and can also enhance the quality of the data on topics such as income where respondents may have difficulty or may be unwilling to report their income accurately. In the business statistics area, administrative data is already being used to replace direct collection of data in some surveys.

The Annual Enterprise Survey, which measures the financial performance and financial position of New Zealand businesses by industry, uses annual accounts summary information from New Zealand's tax agency (Inland Revenue) to provide data for all unincorporated businesses in all surveyed industries and for all businesses in the agriculture industries. Similarly, in the sub-annual financial statistics goods and services (GST) data is used to replace direct surveying of smaller businesses.

The statistical architecture that has been adopted for economic statistics will bring about a shift to a greater dependence on administrative tax data across all economic statistics. The current approach is to use administrative data only for less significant units. In the future, administrative data will be used wherever possible, with postal surveys being used to fill in gaps. Our comprehensive business register will be used to integrate survey and administrative data in a coherent way.

In the social statistics area, we are beginning to explore the feasibility of replacing survey data on incomes and benefits with data from administrative sources in some key sources.

2. Issues using administrative data for statistical purposes

Statistics New Zealand's experience has shown that there are real and significant benefits in utilising administrative data for statistical purposes. In particular,

administrative data allow the production of statistical outputs that could not otherwise be produced, both in terms of topics and fine levels of disaggregation. Administrative data also provide efficiencies for the statistics office because respondent contact and data collection costs are not incurred. A further advantage of using administrative data is the reduction in compliance costs on businesses and households.

While there are benefits in using administrative data for statistical purposes, there are some features of administrative data sources that may limit or undermine their usefulness. These arise from the fact administrative data systems are designed for operational rather than for statistical purposes. Often, the statistical uses of the data are unknown when the administrative system is designed and statistical agencies invariably have limited impact on the system. The key issues that Statistics New Zealand has faced in using administrative data for statistical purposes include:

Coverage – administrative records do not always represent the population of interest for statistical purposes, leading to issues similar to sampling and non-response bias.

Quality – the data often does not meet the standards for official statistics in terms of relevance, timeliness, accuracy, consistency and interpretability. Deficiencies in data quality can range from inconsistencies in the application of questions, to missing fields of data, errors/inconsistencies in the coding of responses, use of non-standard concepts/definitions, and the duplication of records/identities.

Continuity – the data is subject to change as a result of administrative, policy and technical decisions, often impairing the inter-temporal stability of the data and its usefulness for analysing trends over time.

Unit of measurement – the data is often event or case-based which can make the generation of statistics based on units of policy interest, such as individuals or families, problematic.

Explanatory/classificatory variables – administrative databases are typically limited to items needed to support a particular government program or administrative function, which means that they often contain few explanatory variables and few variables to classify the population into sub-groups of interest (e.g. ethnicity, country of birth, labour force status).

The up-front time taken to identify, and where possible resolve these issues before statistical production can commence is often quite lengthy. In some cases, the issues may undermine the usefulness of the data for monitoring trends or understanding outcomes, rendering the data unsuitable for the production of official statistics.

It is important to note that the relative significance of these issues and the extent to which they can be managed varies between social and economic statistics. For example, in the economics statistics area the existence of a comprehensive business frame makes coverage issues relatively straightforward to manage. Statisticians can choose a design that uses administrative data for the types of units where this has been shown to be appropriate. Replacing sampled units with very large weights that amplify changes with full coverage administrative data often improves the quality of our estimates.

Much of the use of administrative data for the production of economic statistics involves tax data. Respondents have strong incentives to supply the tax office with accurate financial information. Having now used tax data for several years, Statistics New Zealand has the confidence to expand its use of tax data to replace postal surveyed units.

3. Management of quality of official statistics

Statistics New Zealand has developed a Meta Information Template for the Description and Assessment of Administrative Data sources. The template has been used widely within Statistics New Zealand for assessing the suitability of an administrative data source for statistical purposes. It is based around the six dimensions of quality - *relevance, accuracy, timeliness, accessibility, interpretability and coherence*. Use of the template provides a structured format for the documentation of administrative data and the information needed to assess the statistical integrity of an administrative data source. A summary of the contents of the template is provided in Appendix 1.

As noted above, a number of recent initiatives by Statistics New Zealand in the use of administrative data have involved integrating the data with other administrative or survey data. One of the outcomes of this work has been the publication of a Data Integration Manual. The manual provides a guide to best practice in integrating data for statistical purposes and shares the insights gained from Statistics New Zealand's experience. In addition to discussion of the technical, legal and policy considerations associated with integrating data from different sources, the manual discusses the importance of understanding the quality of the source datasets.

Statistics New Zealand has established an Administrative and Integrated Data Network, comprising staff from methodological and subject matters in the agency. The network provides leadership in the use of administrative data within the statistics office and a mechanism for sharing information and building knowledge on issues relating to the use of administrative data for statistical purposes.

4. Management of the relationship with administrative departments collecting and compiling important indicators and other statistics

Statistics New Zealand has implemented several initiatives in recent years to improve the quality and co-ordination of statistics across the official statistics system (OSS), including official statistics derived from administrative data. These initiatives include the development of:

Principles and Protocols for Producers of Tier 1 Statistics – these provide guidelines to agencies on the production of Tier 1 statistics. Tier 1 statistics are official statistics that have been identified as important measures of New Zealand, and as such need to be produced, analysed and released to high statistical standards. They include statistical outputs derived from administrative data sources as well as from sample surveys. There are 10 principles which relate to relevance, integrity, quality, coherence, accessibility, efficiency, protecting respondent information, minimising respondent load, maximising existing data sources and international participation. The six protocols relate to:

- quality
- frameworks, standards and classifications

- respondent management
- confidentiality, privacy and security
- release practices
- management, documentation and preservation of statistical records.

Statistics New Zealand encourages agencies to apply these principles and protocols to the production of all official statistics.

Each year Statistics New Zealand reports on the health of Tier 1 statistics. The basis for this report is a self-assessment by providers of Tier 1 statistics on the quality of the statistics. The quality assessment covers the six protocols for official statistics.

Official Statistics Research Programme – facilitated by Statistics New Zealand, this programme commissions and funds statistical and methodological research. One of the research priorities for the programme since its inception in 2005 has been ‘to assess the viability and utility of administrative data and/or existing official statistics to enhance social and/or economic data collection and outputs, and produce new official statistics measures and constructs’. A number of research projects funded through the programme have been directly relevant to improving the quality of official statistics derived from administrative data.

Official Statistics System Seminar Series – provides a forum for presenting and exchanging information about the official statistics system, to build capability and co-operation across agencies involved in the official statistics system. Seminars have included presentations on developments relating to the use of administrative data in the production of official statistics.

Domain Plans – these provide a mechanism for establishing a shared understanding of statistical priorities in a particular field of statistics, and agreement between major users and data custodians on the activities required to address the statistical needs. The intention is that the domain plan will guide statistical development work, including collection activity, use of administrative datasets and data analysis. An important aim is to maximise the use, and ensure the quality of current data (including administrative data) and minimise the need for additional collection activity.

Memorandum of Understanding – are an important tool for managing the relationship with providers of administrative data, especially where disruptions to supply of data would impact the quality of official statistics. They ensure that Statistics New Zealand receives early warning of changes to administrative procedures that could affect the quality data used for the production of statistics.

These agreements will usually specify:

- conditions for the supply of administrative data;
- timetable for the supply of data;
- confidentiality and security of data ;
- consultation before making changes that will affect the variables supplied;
- consultation before changing administrative forms;

5. Major challenges in the use of administrative data for statistical purposes

A major challenge in using administrative data for statistical purposes in New Zealand is managing the quality of administrative data. There is typically a gap between the statistical data requirements and the administrative data that is available. As a result, appropriate methodologies and processes must be developed by the statistical office to transform, model or adapt the data to meet the statistical requirements. These can range from using statistical techniques such as estimation or forecasting to address timeliness issues, to using administrative data that is correlated to a required statistical variable that is not available from the administrative data source to model the required variable, to accepting the administrative data as it is and being very clear with users on defining the statistical outputs produced.

Much of the new potential of administrative data for statistical purposes is likely to be realised through data integration or through replacing survey data with administrative data. In New Zealand, integration of administrative data within a specific sector such as justice is possible because of the existence sector-specific unique identifiers. However, the lack of a unique identifier for individual transactions with all government agencies is a major impediment to the integration of social administrative datasets across sectors, such as between health and justice.

Another challenge associated with the use of administrative data arises from privacy concerns connected with the use of administrative data for purposes that differ from those for which the information was collected. These concerns are particularly important when administrative data records of individuals or other units are linked to other sources of data. Although Statistics New Zealand has developed principles and protocols that provide strict conditions on how data integration is undertaken, privacy issues and public acceptability of integrating data will ultimately determine how far the statistical office can push the boundaries in terms of creating an integrated system of official statistics, using administrative and survey data, particularly in the area of social statistics.

References

Statistics New Zealand (2006). *Data Integration Manual*. Statistics New Zealand Wellington.

Appendix One

Meta Information Template for Description and Assessment of Administrative Data

Summary of Contents

Data Source	Name of agency Application of data High level summary of time period, variables, file structure
Population	Target and actual population Reporting units Coverage
Variables	Name Definition Values
Glossary	Definitions of terms
File Structure	Conceptual IT
Data collection and data entry	Application Collection method Frequency and timing of collection Consistent approach (over collection centres) Question adequacy and respondent understanding Contextual or methodological issues Data capture, coding and editing Updating procedures
Data Accuracy	Missing data Imputation Duplication Rounding Internal consistency
Changes over time in:	Concepts Coverage Data collection and data entry Data accuracy
Accessibility	Privacy/security/confidentiality Documentation available Ease of access Forms of dissemination Timeliness of publications Storage of historical information
Comparison with other data sources	Consistency at aggregate level Comparison of variable concepts and values with Statistics NZ standards
Summary of Statistical Integrity	Is the population well defined? Do reporting units map to statistical units Are variables well defined Do data collection and entry systems have good quality controls? Is data accuracy reasonable (bias, reliability, internal consistency)? Is there a consistent time series?