

**UNITED NATIONS STATISTICAL INSTITUTE
FOR ASIA AND THE PACIFIC (SIAP)
CHIBA, JAPAN**

Seventh Management Seminar
For the Heads of National Statistical Officers in Asia and the Pacific
13-15 October 2008,
Shanghai, China

Country Paper



**MANAGING ISSUES ADDRESSING THE CHALLENGES OF USING
ADMINISTRATIVE DATA FOR STATISTICAL PURPOSES**

This paper is prepared by Mr. Ismail Yusoff, Deputy Chief Statistician (Social and Demographic Program), Department of Statistics, Federal Government Administrative Centre, Putrajaya, Malaysia.

DEPARTMENT OF STATISTICS, MALAYSIA

**Seventh Management Seminar
For the Heads of National Statistical Officers in Asia and the Pacific
13 – 15 October 2008, Shanghai, China**

**MANAGING ISSUES ADDRESSING THE CHALLENGES OF USING
ADMINISTRATIVE DATA FOR STATISTICAL PURPOSES**

1. Introduction

- 1.1 Administrative data is the set of units and data derived from an administrative source. Administrative sources are sources containing information that is not primarily collected for statistical purposes.
- 1.2 Administrative data collections are one of the three main types of statistical data collection. They provide a high level of detail and data are collected on an on-going basis, allowing the trend analysis. In conducting their day to day business, many organizations collect data about their clients or respondents. The data collected are often referred to as 'administrative by-product data' and such collections can be valuable source of information.
- 1.3 Official records of government departments and organizations can prove to be very important sources of statistics, although they tend sometimes to be overlooked. The statistics compiled from these sources have a fundamental difference from other statistics, in that they are derived from a system set up specifically for some other purposes. The need for such statistics may not have been foreseen and provided for when the system was put into operation. Therefore statistics have to be drawn from whatever data happen to be available from the system, and their utilization by the users may necessitate some adaptation of needs. This is the reverse of the approach which a statistician takes in designing statistical operation and in which he makes every effort to design the system to produce the data which meets the needs of the users.

2. Why use administrative data?

- 2.1 The main advantage of administrative data over survey data is usually cost. Surveys are very expensive, particularly if they are conducted as censuses or involve the use of personal interviews. Administrative sources are often 'free' particularly if they originate from the public sector. If there is a charge for the administrative

data, it is often still cheaper than collecting the same information via a survey. Less staff is usually needed to process the data and there is no need for response chasing. Using data from administrative sources also helps to reduce the response burden. Whilst businesses usually understand the reason for supplying data for registration and taxation purposes, even if they do not like doing so, they may see statistical data requests as an extra, less necessary, burden. If they have already provided details to other government departments, they may become annoyed at receiving similar requests from the national statistical institution. A related advantage is that the use of administrative data may in some cases allow statistics to be produced more frequently, with no extra cost to businesses.

- 2.2 Administrative data often give complete or almost complete coverage of the target population, whereas sample surveys only directly cover a relatively small proportion. The use of administrative data therefore eliminates survey errors, removes (or significantly reduces) non-response and provide more accurate and detailed estimates for various sub-population.

3. Administrative data sources

Administrative data are an important source of information for statistical data collection. These data are created by some process of government administration that require a document or record about an event or activity concerning an individual or organization. Malaysia uses the administrative data for statistical purpose and following are some examples of the areas in the Department that uses the administrative data.

3.1 Vital Statistics

Vital statistics refers to events that contribute to the enhancement or change in the population of a country. These events include population births, education, occupation, marriages, migration, divorces, ageing and deaths. These components are further used to generate population estimates, projections and life expectancies. Administrative data used to generate these statistics are based on Birth Registration Records; Death Registration Records; Stillbirth Registration Records; and Marriage and Divorce Registration Records. These records consist of detailed individual information, such as, name, age, ethnic group, place of birth/event etc.

3.2 Social Statistics

The data on social statistics i.e. housing, education, health, welfare services, public safety , communication & recreation are collected from various government agencies on a regular basis.

3.3 Environment Statistics

Environment statistics are collected and compiled by various agencies in accordance with their respective functions and responsibilities. An increasing awareness of the government on the need to ensure a clean, safe and healthy environment as envisaged in the National Policy on the Environment, has resulted in the urgent need for a co-coordinated approach in providing environmental data for the country. Hence Department of Statistics Malaysia (DOSM) has been entrusted with a wider role of being the central depository for environment statistics.

Since DOSM acts as a central depository for environment statistics, most of the statistics published in the Compendium of Environment Statistics (CES) were obtained from administrative data /records from existing line agencies and other sources. DOSM only compile, present and analyze these administrative data.

3.4 International Trade Merchandise Statistics

The compilation of International Trade Merchandise Statistics is one of the most dependent of administrative data i.e. Customs Import Export Declaration. Customs Declaration Data comprise of more than 95 percent of International Merchandise Trade Statistics compilation by DOSM. This increases the scrutiny of quality of the administrative data. This includes the coverage, accuracy and timeliness of compilations of International Merchandise Trade Statistics.

4. Issues and Challenges

Issues

4.1 Do concepts and definitions not conform to international standards?

Concepts and definitions used in DOSM generally conform to international standards. For example, The International Statistical Classification of Diseases and related Health Problems (ICD) comply with the international standards (WHO). The International Merchandise Trade Statistics compilations conform to the recommendations of UNSD, International Merchandise Trade Statistics (IMTS), Concept and Definition, Revision 2 and IMTS Compilers Manual, 2004.

4.2 Are incomplete or inaccurate records a serious problem?

Incomplete or inaccurate data maybe a serious problem because it does not reflect the actual situation when analytical report is produced. The report can be under-stated or over-stated.

Examples:

- a) The analysis based on death records received in 2001 and 2006 found the percentage of incomplete cases for addresses at 6.4% in 2001, and 3.9% in 2006. It requires an additional 19 and 13 days respectively to complete coding for addresses prior to this problem.
- b) The percentage of incomplete information for Causes of Death was 19.8% in 2001 and 14.5% in 2006. Additional 59 and 47 days respectively are needed to complete coding of causes of death to resolve this problem.
- c) By using the Census Survival Ratio Method to measure the completeness of the coverage, it was found that the completeness of data on population aged below one year collected by the Census in Sabah is better with 91.6% compared with data from the National Registration Department (NRD) which is only 80.7%.

In the case of International Merchandise Trade Statistics, the incomplete and inaccurate administrative data does exist but

DOSM have made pre-emptive measures; coverage: customs confirmation of forms/electronic data were acquired; accuracy: confirmations from importer and exporter were established and timeliness: consistent monitoring of forms/electronic data received from customs were done via regular communications.

4.3 Can the data be disaggregated to provide data for regions and other sub-populations?

Aggregation depends on the method of collection and level of disaggregation. Majority of the data can be disaggregated by state and some data such as population, number of schools and some health related statistics can be disaggregated up to district level.

The vast contents of administrative data available for International Merchandise Trade Statistics enable DOSM to compile the regional and international trade statistics for Peninsular Malaysia, State of Sabah and Sarawak. The compilation of statistics by the entry and exit points is also available.

4.4 How do you evaluate the quality of statistics from administrative data used for statistical purposes in the NSO?

The quality of administrative data needs to be improved in order for it to be used for statistical purposes. For example, validation International Merchandise Trade Statistics is done through range check of average unit value, commodity codes and country codes. DOSM and Royal Malaysian Customs (RMC) does have different objective in compilation of International Merchandise Trade Statistics i.e. DOSM, National Statistics Compilers whereas Royal Malaysian Customs, Revenue Collection. The level of checks was also different between agencies, where RMC focus on Imports Information while DOSM focuses 100 percent checks of Import and Export activities.

4.5 Are there any cross checks available from survey data?

Some of the administrative data is the one and only source and therefore, cross checks cannot be made. However, some of the data are comparable to other agencies. For example, data from the Population and Housing Censuses (by using the Census Survival Ratio Method to measure the completeness of data) can be compared with the data from the Ministry of Health. Some data such as the number of doctors, hospitals etc. can be checked through results obtained from related survey. International

Merchandise Trade Statistics can be compared with the Monthly Manufacturing Survey.

4.6 What training is already available in your country for line ministry personnel?

Training is opened to statistical cadre posts in line ministries. Customised courses such as for compilation of MDGs indicators is opened to the respective line ministries. If training is not available, then arrangement is made to do field work such as visiting the site. In the case of air quality statistics, these statistics are usually obtained through monitoring for example the Air Pollution Index (API) compiled by the Department of Environment. Study tour is organized to one of the stations where the data are collected. Discussions are held as and when the need arises to further understand the compilation of the index and ensure the quality of the data published.

In the case of International Merchandise Trade Statistics (IMTS), the gaps of priorities between DOSM and Royal Malaysian Customs (RMC) were bridged by conducting joint trainings of which DOSM participated in RMC Training i.e. product classification and DOSM conducts training to RMC officers on uses and importance of IMTS. This arrangement did value-add towards the quality of data and improve both parties understanding in IMTS compilation.

4.7 Is the available training well suited to ensure the quality of statistics generated?

Discussions are held with the relevant agencies/ ministries on the data requirement of the Department of Statistics. Training on the subject matter is often conducted for the officers concerned to ensure that the quality of statistics generated are acceptable for both the agencies/ ministries.

4.8 Do you have formal agreements with the administrative department?

There are cases where cooperation with the agencies is based on mutual understanding build-up through-out the years. For example, formal agreement between DOSM and Royal Malaysian Customs has been established for more than 5 decades through regular meetings, consultations and informal discussions at desk officer's level. The cooperation between two agencies was further strengthened by visits of Head of Departments at each others

office. The exchange of ideas between Heads of Departments, clarify the understanding between the two agencies and this hope to be passed down the chain of command.

Memorandum of Understanding (MoU) with relevant Agencies:

Statistics Department has developed agreements with relevant government agencies indicating the roles and responsibilities of each party, the processes to be followed and the expected outcomes. It is envisaged that the agreements will cover the broad range of services to be provided by the department as well as the responsibilities of individual agencies or issues such as implementation of best practice guidelines and the provision of data. The agreements will also provide an opportunity to set out benefits to all parties of a coordinated set of activities aimed at improving the national statistical system. In order to achieve this, it is recommended to have some sort of formal agreement with administrative data suppliers, such as 'memorandum of understanding (MoU)' within the public sector. To date, Department has drafted 'MoU' with five relevant agencies and is still pending approval from the legal advisor of Prime Minister's Department. The five relevant agencies are The Central Bank of Malaysia, Inland Revenue Board Malaysia, Companies Commission of Malaysia, Employees Provident Fund and Malaysian Communications and Multimedia Commission.

4.9 What are major obstacles / weaknesses faced in your country on the use of administrative data for statistical purposes?

Timeliness in getting the data and to ensure the data fits statistical requirement. Data coordination at small area levels are not undertaken by the agencies' headquarters. The agencies are unable to provide data at state level. Obstacles in accessing data due to confidentiality issues and getting approval for release of data.

Challenges

- 4.10 Demands for statistics on many aspects of our lives, society, economy and environment continue to grow. This may be due to our increased ability to handle and manipulate large sets of data as we move into the so-called information age. It may also be a reflection of the increasing complexity of our social and economic systems and our desire to understand them better. Whatever their cause, we face these demands in a climate of tight budgetary constraints for government statistical agencies. At the same time, statistical agencies are sensitive to the increased burden that would be imposed on respondents by an increase in survey-taking activity to meet these demands.
- 4.11 These factors have led to the exploration of other means of satisfying these statistical demands. Prominent among these alternative means is the increased use of existing administrative systems as sources of statistical data. For many years, statistical data have been a by-product of administrative process in domains such as vital statistics, imports and exports, health care and education.
- 4.12 Interest in administrative records use is growing in national statistical organizations that do not currently use these resources. The increasing cost to collect data, combined with shrinking budgets, demands new innovative methods to generate the detailed geographic data needed to plan and monitor social programs. Despite the promise, many technical and policy issues exist and must be addressed for these data to fulfill our high expectations.

5. Steps taken by Statistics Department to increase the usage of administrative data for statistical purposes

5.1 Creating 'one-stop-data centre' / data portal

A centralised and ongoing repository of information on administrative data should be established. The Department has initiated the existence of a 'one-stop-data centre' or data portal or in other words, creating an integrated data-infrastructure within which all available data and metadata are stored in such a way that the information in the statistical databases can be interlinked. All department published data should be stored in this data portal so that it can be readily accessed at any time. If this can be established, then Statistics Department will be the main centre of focused for data users. Statistics Department can be considered as the data provider as well as data users. Of course, this data portal will be updated from time to time so that it will always have the latest data as required by data users purposes.

5.2 Standardisation of Codes, Classifications and Definitions

One major problem often encountered when using administrative data is that the units in those sources do not correspond directly to the definition of the required statistical units. The process of converting from administrative units (which may often be equivalent to legal units) to statistical units (i.e enterprises and local units) is known as profiling. As well as differences in the definitions of units, there are also likely to be differences in the definitions of variables between administrative and statistical systems. The data in the administrative sources have generally been collected for a specific administrative purpose and the needs and priorities relating to that purpose are likely to be different to those of the statistical system.

The Department has already set up a technical working group to standardise the codes, classifications and definitions. This working group will be responsible in terms of standardisation within department and inter-departments.

5.3 Placing Cadres in all Government Agencies

It is good for the department to study the posts of cadres in the government agencies. Most departments conduct their own surveys for data collection and the existence of our own statistics cadres in the different government agencies will make it easier to get excess to their data if these data are useful for our purposes. Currently, our cadres are placed in selected government agencies but more posts should be created and if possible have our own statisticians as cadres to be placed in all government agencies.

5.4 Using administrative records in business registers

In the early 90's, major efforts were made at the Department of Statistics Malaysia (DOSM) to establish a business register. Hence, a Central Register System, commonly known as SIDAP was developed to provide a profile of active business establishments and enterprises which can be used as a sampling frame for establishment surveys. In order to provide the most up-to-date information for conducting business surveys, SIDAP to be regularly updated. Maintenance depends entirely on field operations, is a very costly and time consuming affair. DOSM is currently looking into the possibility of obtaining data directly from relevant administrative sources to update the SIDAP system.

Data for the business register are obtained mainly from the Registrar of Businesses (ROB) and Registrar of Companies (ROC) from the Companies Commission of Malaysia (CCM). All businesses and companies are required to register with the CCM except the professionals, the institutional and government agencies. Other data contributors include Employment Provident Fund (EPF), Construction Industries Development Board (CIDB), other government agencies, statutory bodies, professional associations, internet, newspapers and etc.

5.5 Tapping data from other Government Agencies

One of the easiest ways to obtain administrative data will be to tap data from other government agencies which has the data readily available for usage. The Statistics Department has to review the following issues so that tapping of data can be done smoothly.

a) *Set-up Technical Working Group*

The existence of technical group to undertake the responsibility of tapping the administrative data is important. This technical working group will comprise of representatives from all relevant divisions in the department and the responsibility lies on their shoulder to make sure that the data from the respective divisions will be made available for inter-division users. On top of that, this group of people will also be responsible to get hold of administrative data from other government agencies in relation to their work or division. The administrative data will then be updated in the data portal for easy access by the users.

b) *Revise Statistics Act 1965 (revised 1989)*

The Department is currently taking action to revise the Statistics Act by adding the clause in the act to give the authority to officers to have access to administrative records or data of the Government Department, Ministry or the statutory body for the purposes of compilation of statistics in order to make the usage of administrative data be effective and important.

c) *Provide technical assistance to other Government Agencies*

The Department provides technical assistance to other government agencies if it is needed. A good example of this will be assisting in the sampling design when conducting a survey. This is important so that the same concept of sampling design will be used every time when the survey is conducted.

d) *Mutual understanding with data provider / respondents*

Alongside formal arrangements it is also strongly recommended to build a good working relationship with administrative data suppliers. It is usually worth investing some time to visit suppliers to gain a greater knowledge of their work, systems and constraints. This leads to a better appreciation of quality issues and can help to build goodwill and mutual understanding, which in turn helps identification and resolution of problems as they arise.

6. Conclusion

Administrative records are and will continue to be increasingly important source of statistical data. The relative strength and weakness of data derived from administrative systems, in terms of cost, coverage, quality, relevance and timeliness, in comparison to census or survey-based data, dictate the manner in which these sources of data are most effectively used. Current uses of administrative records include direct tabulation, indirect estimation, substitution for survey responses, frame construction and maintenance and data evaluation. These uses now permeate most statistical programs and can be expected to extend even further in the future. With the growing use of administrative records, statistical agencies are becoming increasingly dependent upon other agencies for the uninterrupted flow of input data to their statistical programs. The ability of the statistical agency to influence the design or redesign of administrative systems rests on a mutual understanding of the requirements of the two agencies. Establishment of a government-wide policy or principle that the statistical agency should have a voice in decision regarding the design of administrative systems, or more generally in proposals for meeting the statistical needs of new programs, can help the statistical agency in this regards, but is no substitute for the fostering of close cooperation with administrative agencies.

References:

1. Issues in the use of administrative records for statistical purposes by G.J.Bracstone, Statistics Canada.
2. Australian Bureau of Statistics Web Site
3. OECD Glossary of Statistical Terms
4. Eurostat web Site
5. Paper on Implementation of DOSM'S Strengthening Plan under Main User Committee
6. Statistics Act 1965 (Revised 1989), Department of Statistics Malaysia
7. Statistics Canada Web Site
8. Statistics Finland Web Site

**Department of Statistics
Malaysia
October 2008**