

SINGAPORE REGISTER-BASED CENSUS

- Lessons Learnt and Challenges Ahead

INTRODUCTION

1 Internationally, censuses have been used as the source of benchmark data for the total population and its social and economic attributes. This is due to the unique characteristics of a census: universal coverage, wide scope of enquiry and specific time reference. The periodic taking of a census allows the monitoring of the changes in a population and its characteristics. In all countries, the taking of the census is perhaps the most important statistical undertaking for the national statistical offices.

2 The traditional approach in conducting a census is to engage fieldworkers to enumerate the entire population physically present in the country during Census day. This is a labour-intensive and logistically massive exercise. Over the years, Singapore has moved away from this fieldwork approach. In 2000, Singapore adopted the register-based approach for its Census of Population. In the register-based approach, the Household Registration Database (HRD) developed by the Singapore Department of Statistics formed the core database for the conduct of the Census 2000. It provided the basic count of individuals and overall profile of the population. Additional data pertaining to socio-economic characteristics of the population not available in the database were collected from a 20 per cent sample of the population.

SINGAPORE CENSUS 2000

- A REGISTER-BASED CENSUS

3 The Singapore Department of Statistics studied carefully the conceptual framework and practical issues before switching from the traditional approach to the register-based one. The key issues considered included :

- (a) Whether the quality of administrative data in Singapore is sufficiently high to produce an accurate count of the population and its basic characteristics.
- (b) Whether the legal environment and data confidentiality practices in Singapore permit the sharing of non-sensitive administrative information.
- (c) Whether the cost savings in adopting this approach are substantial.

4 Key developments over a ten-year period from the mid-1980s to mid-1990s progressively laid the foundation for a register-based Census in 2000. These developments have greatly enhanced the feasibility of a register-based census in Singapore. The key development milestones were :

- (a) the setting up of People Hub and the development of the Unique Identification Number (UIN) for record linking in the mid-1980s;
- (b) the 1990 Census capitalised on the UIN, using it as a key identifier to merge information captured in People Hub with a few other administrative databases, pre-printing this information onto census forms for verification with the respondents;
- (c) the establishment of the Household Registration Database (HRD) in 1996. The HRD, akin to a population register, was set up using information sourced from administrative registers and the 1990 Census. The HRD was updated regularly by administrative data from various sources, using the UIN for each individual as the linking key.
- (d) the database on dwellings maintained since 1980 was upgraded and renamed National Database on Dwellings (NDD) in 1996. The NDD and HRD together gave a physical location for every person and household in Singapore.

5 With the basic or core items on individuals and houses being available from the HRD and NDD, it was deemed possible to conduct a “register-based census” in the year 2000. Data on the basic characteristics of the population were extracted from the HRD and NDD. Only additional data items not available in the HRD and NDD were collected from a 20 per cent sample of the population. These data items include household structure, languages spoken, previous employment, transport mode and religion.

6 Another breakthrough achieved by Census 2000 was the implementation of the multi-modal data collection strategy. Respondents were given the opportunity to self-enumerate via Internet, to provide information through our CATI (Computer Assisted Telephone Interview) system or to the fieldworkers. Compared to the traditional model of canvassing information using a large field force, the multi-modal data collection strategy made full use of the high internet penetration rate, the availability of a good database of telephone numbers and dwelling addresses as well as the advanced computer technology in system integration.

LESSONS LEARNT

7 The advanced developments in areas such as database linkage, sharing of non-sensitive data, technologies in data storage capability and data collection (Internet & CATI) systems have contributed significantly to the successful conduct of the register-based Census 2000 in Singapore.

8 The following section describes the lessons learnt from the Register-based Census and outlines the challenges ahead.

Close Liaison with Data Source Agencies

9 The use of administrative data from multiple sources requires co-operation from various agencies which own the data. The Department started preparatory work with various government agencies on the provision of data for data-merging as early as 1996 when the HRD was set up. Concerns such as database linkages among different database platforms, timing and legislation on provision of data, confidentiality of data provided were addressed during this early stage of development.

10 As part of its on-going process to enrich its databases, we continue to work with various agencies to resolve issues related to data provision, data file format, data coverage and item specifications for new/additional data items. Discussion also includes data privacy issues.

Standardisation of Data Coverage, Quality and Definition

11 Since data from other registers are collected as part of administrative mechanisms, data coverage and definition may not conform to statistical requirements. Some examples of how such issues were resolved are described below :

Data Coverage

12 Administrative registers generally maintain good records. However, some of the administrative data may not be up-to-date as households and population may not provide updates of changes. All citizens, permanent residents of Singapore and foreigners holding valid passports are generally accounted for, in terms of the number and basic characteristics. However, it was found that the number of elderly population in the administrative records was different than that indicated in the official population estimates. To rectify the records, a 'focused' census on the elderly persons was conducted in 1998/99. This census covered

aged institutions/old folks' homes as well as all elderly persons aged 75 years and over, staying in residential dwellings. Results showed that most elderly aged 90 years and over had left Singapore or passed away overseas, while at least half of those aged 85–89 years were still found in Singapore. The results from this census were used to improve the quality of data in the HRD and to provide a more accurate count of the elderly population.

13 In general, administrative databases include only data items which are of direct relevance to the subjects under the agencies' purview. It is therefore necessary for the Department to identify the data gaps and to approach other administrative sources to obtain useful and relevant data items. Where necessary, data items were included in the 20 per cent sample survey when they were not found in any administrative databases.

Data Definition/Classification

14 Definition and classification of items used for administrative purposes may be different from statistical definitions and classifications. This gives rise to difficulties in maintaining consistent definitions across different government agencies and over time. Various measures were adopted to standardise data items. For example, to resolve the issue of different address format maintained by different agencies, national standards were developed through inter-Ministry consultations and there were concerted efforts to promote the use of these standards by the different agencies.

15 The classification of economic activity by various agencies is another problem commonly faced by statistical agencies when harmonising data on industry. We are responsible for developing the national standard for industry classification. Many agencies use the Singapore Standard Industrial Classification (SSIC) to meet their administrative requirements. However, there are a few agencies which do not impose strict validity checks or perform any timely updating on the codes. When merging data from such administrative sources, we have to resolve the inconsistent codes with diligence and care.

16 To facilitate standardisation of data items available from various administrative sources, we implemented a Code Conversion System (CCS). The CCS is a meta-data repository that translates the codes and definitions of items from different data sources into a standardised set of codes and definitions, specifically for the Department's use. The CCS is designed for quick and easy updating of codes and definition for newly acquired administrative data. In addition, the CCS also serves as a meta-data reference about the data items.

Maintain Specific Time Reference

17 A traditional census measures the characteristics of the population at a certain point in time, usually referred to as a snapshot of the population. The use of multiple data sources from various administrative registers presents the problem of reconciling the reference periods. Taking into account the timing of update and dependency of data from different sources, a logical system on the appropriate sequence of stepwise merging is developed and complied with strictly in order to produce meaningful and up-to-date data. Using this approach, we were able to use the HRD to produce the top-line data comprising the basic count and profile of Singapore's population for Singapore's Census Day, 30 June 2000.

Confidentiality and Security of Data

18 The move towards an integrated database through more extensive merging of data from multiple sources also implies that any compromise or breaches on data security would have severe consequences. It could be an illegal access to the database, intentional or accidental alteration or destruction of data. Therefore, we have to put in place a data protection regime to ensure security and confidentiality of individual data in our databases. The Department employs a three-prong approach in data protection, comprising legal, administrative and technical measures. These measures are institutionalised and strictly adhered to in order to ensure that data are protected against illegal access and use, accidental or intentional alteration, destruction, or disclosure. The confidence of respondents was maintained to the highest level possible at all times.

Data Collection Strategy

19 Singapore was among the first few countries in the world to attempt to collect census information via the Internet. The Internet is a convenient mode for the respondents to complete their survey return. It allows the respondents greater privacy as their personal particulars are not revealed to an interviewer but transmitted directly to the Census enumeration database.

20 To achieve a significant response rate, critical success factors such as design of the on-line form, incentives and publicity were taken into account in the 2000 Census. In addition, form-filling on the internet was designed to be user-friendly and interactive. Upon successful logging in at the Census website using a user-id and password, basic data already available in the pre-Census database would be displayed. The respondent would then proceed to fill up the rest of the questionnaire on-line. The respondent could also save the partially-completed form and retrieve it later to provide the rest of the information

required. Simple on-line checks were included and respondents would be prompted to re-enter the data if the information is incorrect or inconsistent.

21 Issues such as security and confidentiality were also addressed. A series of security features were incorporated to verify the identity of the respondent, monitor any unauthorised attempts and to put in place an early detection system. Data sent through the internet were also encrypted during transmission to the database, which was protected by appropriate firewalls.

DEVELOPMENTS AFTER 2000

22 After the successful completion of Census 2000 using the Register-Based approach, the Singapore Department of Statistics embarked on two key strategic directives – further development of database system and upgrading data processing and tabulation capabilities.

Upgrading the Technical Capabilities of HRD and NDD

23 The Department completed the migration of the HRD and NDD from mainframe to client-server platform in the third quarter 2003. This was a strategic move aimed to enhance the capability in the extraction, processing and analysis of household and population data. With the much enhanced capability, we are able to create more databases, using HRD and NDD as the core databases and tapping other new administrative data sources.

Improved Data Processing and Tabulation

24 Technology has advanced so much in recent years that it is possible to perform data-merging at lower costs and within a shorter span of time. We are now developing a new system called “merge-on-demand”, which is capable of merging a few large datasets within a short span of time. This new initiative will facilitate quick data processing and compilation to meet the data needs for a variety of statistical studies.

CHALLENGES AHEAD

25 The traditional census may no longer be the most suitable approach in the face of a fast-changing environment. The main disadvantage of a traditional census is its high cost. The register-based approach, together with new technological innovations that reduced manpower costs, helped to reduce the cost of conducting the 2000 census by 75 per cent. The combination of register-based approach with sample survey for additional data items optimizes the use of available data sources. It represents a shift from a traditional census system to a census information system where the data are enriched by regular updates from administrative sources. In a traditional census, the data items represent snapshots of the population in intervals of 10 years. In this new approach, the census information system reflects a dynamic and 'continuously-updated' statistical system, which is able to track the changes and generate snapshots of the population at intervals shorter than the 10-year cycle. As a result, inter-censal year estimates on basic profile of the population such as age, ethnic group, sex, house-type and other information, can be made readily available from the HRD/NDD.

26 Against this backdrop, the challenges ahead for the Department lie in how to meet users' demand for more information, in-depth analysis, with faster turnaround, without unduly increasing respondent burden.

Data Collection and Processing Methodology

27 Given the success of Internet submission and CATI in the Census 2000, these two modes are likely to remain the key modes of data collection strategies for future censuses. In Census 2000, a more controlled environment was adopted where only one mode of enumeration was open to the respondents at any point in time. The challenge for future censuses is to improve the workflow and provide a fully-integrated multi-mode data collection system at the same time. This would provide greater convenience to the respondent since he is able to select the preferred mode of submission anytime within the survey period.

28 We are also looking into both workflow and technology improvements to further improve the data processing operation. Data captured in the Census 2000 would be used to build a knowledge base system to facilitate the coding and editing process. This smart coding system not only allows us to build on past experiences, it also ensures the consistency and quality of data as well as improves the turnaround time.

New Approach to Sourcing More Data

29 The register-based approach, together with the availability of a new technology system of ‘merge-on-demand’, put a premium on sourcing data from administrative databases. Whenever new data items are required, the first step is to look for possible administrative sources within the public sector. If the required data items are available from an administrative source, with reasonable coverage and quality, we would obtain the information from this source. This approach allows new data requirements to be met, with a faster turnaround time and usually, with better coverage than the alternative of collecting data from the respondents directly.

30 Using the newly developed “merge-on-demand” technology, all newly acquired data, after being put through validation checks to ensure data consistency, can be merged with the basic data in HRD, to provide new statistical tables, within a short period of time. Only when the data required are not available from any administrative sources, we would then make an assessment on the suitability of including these data items in the Census or surveys.

Manage Users’ Demands and Expectations

31 With better and more advanced IT techniques and improved procedures, data can be generated within a short span of time. This may help raise data demand from key users and generate an expectation for quick, fast and accurate data supply. To ensure high productivity in our data production, the Department would need to ensure that the procedures for merging, data-cleaning/translation (to ensure consistency in definition) are robust. User expectations will also need to be managed so that we could continue to provide service at the high standards that have been achieved so far.

Respondent Management

32 Although only a small proportion of the population is selected for the sample survey under the register-based approach, those selected will need to respond to more “in-depth” questions, where data are not readily available from any source. As respondents become increasingly more aware that most data can be obtained by data merging among different government agencies, they may resist further attempts to collect any data directly from them. Therefore, special attention should be paid to managing the relationship with respondents to assure them that only required data not available elsewhere are collected directly from them. They have to be reassured of the confidentiality of the statistical data collection process as well. All data, inclusive of those from administrative records or those obtained from surveys or a Census, are protected under the Statistics or Census Act.

CONCLUSION

33 This paper highlighted the key lessons learnt from the Census 2000, as the Singapore Department of Statistics moves from a traditional census approach towards a register-based one. The register-based approach, together with further development of database system and the “merge-on-demand” technique in recent years, enables the Singapore Department of Statistics to provide timely information for a variety of statistical analyses and studies on population and households.

34 To meet the challenges ahead, the Department would be adopting a more holistic approach of data management. By establishing a sound methodology for database merging, using advanced IT technology as well as adopting an integrated and streamlined workflow, we will be able to achieve the maximum output within the resource constraints.

SINGAPORE DEPARTMENT OF STATISTICS
28 OCTOBER 2003