Conference on Chinese Population and Socioeconomic Studies:
Utilizing the 2000/2001 round Census Data
Hong Kong University of Science and Technology
19-21 June 2002, Hong Kong SAR

Singapore's New Approach to Census Taking

Introduction

1. Internationally and in Singapore, Censuses have been used as a source of benchmark data for all population and related social statistics. This is due to its universal coverage, wide scope of enquiry, and specific time reference. In the traditional approach, fieldworkers are engaged to enumerate the entire population physically present in the country during census day.

2. Singapore adopted the register-based approach in its latest Census of Population 2000, and changed its concept of population coverage to a de jure count. With globalisation leading to the frequent travelling of Singaporeans and more foreigners staying long term in Singapore, the de jure concept would more accurately reflect Singapore's population size.

Past Censuses and the De Facto Population Count

3. A de facto population includes all persons physically found within the geographical boundaries of a country at a designated reference time known as the “census day”. The total population comprises all persons present in the country on Census day and enumerated at the place where they are at that moment, regardless of their usual place of residence. This was the approach used by Singapore for Population Censuses in 1990 and before.

4. With the acceleration of the computerisation programmes in the public sector in the 1980s, there was a significant increase in the use of administrative databases in Singapore within the public sector, e.g. for school registrations and government flat applications. Basic information of citizens and permanent residents (PRs) were available in databases to facilitate administrative procedures when public made use of a wide range of services from government Ministries, Departments and Statutory Boards.

5. Data from selected public registers were first used on a large-scale basis in the Population Census 1990. Enumerators visiting each household had pre-filled basic information on every household member from the

merged database printed onto questionnaires, thereby reducing interviewing time and enumeration costs. Census 1990 was the first time when a database was used to conduct the Census, respondents were asked to verify the pre-printed information and to make amendments if need be. Subsequent analyses showed that the general characteristics of the population did not differ significantly between the database and field-collected census data. The experience gained gave positive indication that a register-based Census could be conducted in Singapore.

The Household Registration Database and the De Jure Population Count

6. By adopting a new approach such as the register-based Census, it was necessary to align the coverage of the Census population with the coverage of population available from the register. The de jure concept for population coverage was a good alternative that satisfied and supported the new approach.

7. A de jure population comprises all usual residents in a country. All persons at their places of usual residence will be enumerated, as well as those who may be absent from their places of usual residence, irrespective of where they are on Census day. Singapore adopted a de jure concept for its population estimates upon the use of a register-based approach for its Population Census 2000.

8. The Singapore Department of Statistics studied carefully into the conceptual framework and practical issues before the switch. Three key issues were considered carefully before the new approach was implemented:

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

9. Full examination of data quality issues and the differences of historical data produced from using the two approaches were undertaken. This helped to determine the impact on data resulting from the change in approach. The technical details on database merging, data updating as well as data tabulation and dissemination were also studied. As the database platforms used by various government agencies were different,

technical solutions were adopted to resolve them and to ensure compatibility between the databases.

10. The register-based Census was built on an integrated database system known as the Household Registration Database (HRD), developed in 1996 by the Singapore Department of Statistics. It provided the basic count of individuals and overall profile of the population. Some of the aggregate information obtained from the HRD included age group, sex, ethnic group, citizenship, house-type etc. This was possible because every resident in Singapore (i.e. Singapore citizens and PRs) had been issued with a personal unique identification number (UIN), and every foreigner staying or working in Singapore held a pass with a unique foreign identification number (FIN). This UIN/FIN was used as a key to match various government databases containing administrative information on individuals. The following frequency of updates to the HRD were made with respect to individual records:

Data Item	*Frequency*
• Live Births	Quarterly
• Deaths	Quarterly
• Immigration	Quarterly
• Emigration	Quarterly
• Marriages	Quarterly
• Change of Address	Quarterly
• Formation of Households	Quarterly
• Divorces	Annual
• Education Attending / Qualifications	Annual

11. In addition to basic demographic data of each individual in HRD, information on his/her registered address was available. Each address was checked against the National Database on Dwellings (NDD) to ensure its validity. The NDD contained an inventory of residential addresses in Singapore, and included type of dwelling, census district and address in the National Coded Address (NCA) format.

Total Population Count

12. The total population count was defined to include all residents with valid local addresses as well as all foreigners who had been issued with permits allowing them to stay for at least one year. The implicit

assumption was that citizens and PRs usually "reside" in Singapore if they had a valid local address, even though some may be out of the country for purpose of work, study or otherwise. Foreigners, by virtue of having applied for a working or non-working pass for Singapore, would be "resident" in Singapore as long as the pass remained valid.

13. Persons with foreign and invalid/blank addresses were excluded from the total population count. These were usually complete households who through the registration of their foreign address had indicated that they would be staying abroad for an indefinite length of time. They also included PRs who had not indicated a place of residence in Singapore.

14. Foreigners who had passes lasting for less than a year, were excluded in the total population count. Their place of "usual residence" was assumed to be elsewhere since their stay in Singapore was short.

Verification of Records and Data from HRD

15. As the HRD formed the core database for Census 2000, it was therefore of paramount importance for it to be as clean as possible. The following checks were carried out systematically to ensure that the quality of records and data from HRD were of a very high standard.

Systematic Error Corrections for Database

16. Comprehensive validation and verification rules were built into the database's ongoing updating cycles. Individual records that failed the validation checks were verified and corrected. The majority of the error corrections pertained to items on date of birth, sex, address and educational codes. An independent check on the basic personal information between the updated HRD and administrative stock figures was carried out with effect from the June 1998 data to ensure data accuracy and consistency.

Address Checks

17. A special team of officers was formed to deal with addresses in monthly updating cycles. Invalid addresses were investigated and checked with other published information, e.g. postal department's addresses, or through thorough fieldwork to obtain up-to-date information. Most of the invalid addresses were due to houses and flats being demolished as a result of upgrading programme in the public housing estates. Some new addresses of the rebuilt units were not updated to the database in time. There was also a small number of erroneous data recorded at source.

Comparisons with Official Population Estimates

18. Population figures based on HRD records for the years 1995 to 1999 were generated. These figures were compared with those obtained from the General Household Survey in 1995 and the official population estimates. The latter had been obtained by adding aggregate yearly figures of natural increase and net migration to the Census 1990 base population count. The magnitude and possible reasons for the differences were studied in detail. Follow-up actions were taken to check addresses of school going children, and conduct special surveys on the elderly.

Use of 20 Percent Sample Survey in Census 2000

19. Since the register-based Census provided only the total count and basic profile of the population, it would be necessary to collect the detailed socio-economic characteristics of the population from a sample of the population. The information would be essential for planning and policy making purposes. For Census 2000, a survey of 20 percent of the population was conducted.

20. A total of 54 items were included in Singapore Census 2000 for collection from the population. Of these 54 items, 8 items were obtained directly from the HRD (marked D next to numerical order, listed below). The remaining items that were not available from administrative source were collected from the 20 percent sample survey as part of the Census 2000. Updates to these items were allowed during the data collection phase of the Census.

Personal particulars and demographic characteristics

1 Name
2D Singapore NRIC No. or Foreign Identification No.
3D Sex
4D Ethnic group
5D Date of birth
6 Marital status
7D Place of birth
8 Year of first arrival in Singapore for purposes of taking up residence (if relevant)
9D Citizenship
10D Residential status
11 Religion
12 Year of first marriage (for women only)

13 Number of children born alive (for women only)

14 Number of members —

(*a*) living in Singapore;

(*b*) living abroad and details of the following:

(i) country of residence

(ii) reason for living abroad

(iii) expected duration of employment/course of study abroad (if applicable)

Household characteristics

15 Relationship to head of household

16 Spouse linkage (if applicable)

17 Parent-child linkage (if applicable)

Housing and home upgrading

18D Type of present dwelling

19 Type of proprietary interest in present dwelling

20 Year moved into present dwelling

21 Type of previous dwelling (if any)

22 Type of proprietary interest in previous dwelling

Education

23 Name of educational institution (including vocational institution) attending (for students only)

24 Level of education in educational institution (including vocational institution) (for students only)

25 Major field of study (for polytechnic/university graduates)

26 Highest qualifications obtained

27 Country where highest qualifications were obtained

28 Year when highest qualifications were obtained

29 First degree obtained (for persons who have obtained postgraduate qualifications)

30 Field of study for first degree

31 Language(s) literate in

32 Language(s)/dialect(s) most frequently spoken at home

Educational upgrading

33 Whether vocational qualifications obtained (for persons who have not attended a polytechnic or university)

34 Name of vocational or other educational institution where vocational qualifications were obtained (if applicable)

35 Type of vocational qualifications obtained (if applicable)

36 Major field of study for vocational qualifications (if applicable)

Employment

37 Economic status (whether active or economically inactive)

38 Occupational status (whether employed, self-employed or serving in National Service)

39 Occupation

40 Type of industry

41 Gross monthly pay for June 2000

42 Bonuses received in the previous 12 months

43 Actual number of hours worked per week

Job mobility

44 Length of service in present job

45 Previous occupation (if any)

46 Previous type(s) of industry worked in (if any)

Persons who are not working

47 Previous employment (if any)

48 Whether any action taken to look for work

49 Reason(s) for not working

Transport

50 Usual mode of transport to school (for students)

51 Usual mode of transport to workplace

Overseas travel

52 Number of overseas trips made for business/leisure in the previous 12 months to —

(*a*) Malaysia

(*b*) other countries (please specify)

Elderly persons aged 65 years and above

53 Main source of financial support

54 Ambulant status

21. The data collected from the Census sample survey were expanded using the overall control totals from the HRD to obtain the population value. Detailed characteristics of the population therefore came from the sample survey, while broad characteristics were from the HRD.

The Tri-Modal Data Collection Strategy

22. The Census sample survey made full use of the possibilities offered by the IT revolution by adopting a tri-modal data collection strategy. It comprised Internet enumeration, Computer-Assisted Telephone Interviewing (CATI) and fieldwork.

Internet

23. Singapore was among the first few countries in the world to collect Census information from households via the Internet. Several issues and concerns were addressed before making this bold step. It was recognised that the onus was on the respondents to self-enumerate via the Census 2000 Internet website. To achieve a significant response rate, the design of the online Census form, incentives and publicity were critical success factors. Furthermore, since a sample of the household database was accessible by selected households, security and confidentiality issues had to be addressed and measures put in place to prevent unauthorized access, hacking or denial of service attacks.

24. It was assessed that the advantages of Internet enumeration outweighed the potential risk factors, which were minimised through rigorous security measures. Respondents enjoyed greater privacy, as their information were not revealed to an interviewer, but transmitted directly to the Department's database. Furthermore, the form-filling experience over the Internet was a positive and interactive one. When the respondent logged into the Census website, some basic data already available in the pre-census database were displayed. The respondent proceeded to fill up the rest of the census questionnaire on-line. User-friendly help features and explanatory notes were provided instantaneously when required. The system also performed simple on-line checks, and prompted the respondent to re-enter data that were clearly wrong or inconsistent. For further convenience, partially completed questionnaires could be saved and retrieved at a later time for completion.

25. From an operational perspective, Internet enumeration had many advantages. Most of the data collected from the Internet were already electronically coded, thus reducing data entry and coding at the back end. There were also substantial manpower savings since interviewers were not required to "canvass" information from the population.

CATI

26. CATI operations commenced with a suitable time lapse after the launch of Internet enumeration. Unlike Internet, CATI was a tried and tested data collection strategy, having been deployed for the mid-decade General Household Survey (GHS) in 1995 where some 40,000 households were successfully enumerated by CATI.

27. Households selected were distributed evenly by postal districts (PDs). For each PD, households that had not submitted their returns by Internet were automatically scheduled and dialed up for CATI interview.

Fieldwork

28. Households were scheduled for fieldwork if they could not be contacted by CATI after a number of telephone attempts. These were grouped by PDs and passed to regional census offices.

29. Fieldworkers visited these remaining households to conduct face-to-face interviews. When they failed to get in touch with these households, they left a contact number for these households to call and arrange for a convenient time to complete the interview.

Confidentiality and Security of Information

30. The tri-modal data collection strategy involved the collection of personal and household information via the Internet and CATI where face-to-face contact with respondents was minimised. As personal data were collected electronically, it was important to maintain high standard of system security to protect the data as well as to uphold confidentiality.

31. To ensure confidentiality, all selected households received a notification letter with a house Identification Number (ID) and unique, randomly generated password. Using the house ID, password and the unique identification numbers (UINs) of two members, respondents were able to log-on and retrieve their household record in the database via the Census website. The checking of the password was performed in a secure manner with Privylink. The password was used as a key to generate a random sequence at the respondent's computer. This random sequence was transmitted over the Internet. As the respondent's password was not sent across the Internet, the password could not be intercepted and read. At the server end, the random sequence received was decrypted with a key server. When the decrypted sequence matched, the respondent was authenticated and granted access.

32. All personal information provided by respondents was 128-bit encrypted before transmission over the Internet. This protected the information from unauthorised interception. To protect the information from hacking, a Demilitarised Zone (DMZ) utilising two layers of computer firewalls was set up to protect the on-line database in which the information was stored. These security measures were subjected to the most stringent tests and conformed to Infocomm Development Authority of Singapore (IDA)’s computer security requirements.

33. For CATI, census interviewers quoted the respondent’s unique house ID over the telephone to identify themselves as genuine census officers

before proceeding with the interview. When in doubt over the identity of the CATI interviewers, the public could call the census hotline to verify their identity.

Benefits of the Register-Based Census Approach and Tri-Modal Data Collection for Sample Survey

Timeliness of Data Release

34. The register-based approach and the integrated use of the various technologies in Census 2000 provided a holistic solution to the entire workflow in data collection, processing and publication seamlessly. It provided data of high quality and reduced the turnaround time in the delivery of output to public tremendously. A quick count of the population and basic profile was available in August 2000, within two months from the Census reference date. The first advance data release for Census 2000 was ready in Dec 2000, as compared to May 1991 for the 1990 Census.

Reduced Respondent Burden

35. With the use of a register-based approach, basic information was obtained from database, while detailed information was canvassed from only 20 percent of the population. Four-fifths of the population were left undisturbed for the Census 2000. A traditional approach would have required a hands-on count and visit of everyone in Singapore at the time of Census. The tri-modal approach made it more convenient for respondents to provide information to the Census office.

Variety of Data and Data Quality

36. The register-based approach also provided the facilities for a variety of administrative databases to be merged to the Census database, thereby increasing the volume of data on the population without interviewing the population. These additional data expanded and enhanced the value of the Census 2000 data significantly.

Cost and Manpower Reduction

37. It would have cost $70 million for Singapore to do a traditional Census. This register-based approach coupled with new technological innovations that reduced manpower costs shrunk the entire Census bill to only $22 million, or less than one-third the cost of a traditional Census. Moreover, only 600 persons were employed to conduct the Singapore Census 2000, compared with the estimated 6,000 needed if we were to use the traditional approach.

Population Estimates and Censuses for Future Years

38. Prior to year 2000, Census population counts provided the basis for Singapore's inter-censal population estimates. Annual Singapore residents estimates comprising citizens and those granted Singapore permanent residence were obtained by adding aggregate yearly figures of natural increase and net migration to the Census base population count. The total population count was obtained by adding the number of foreigners to the Singapore residents.

39. From 2001, instead of deriving these population estimates by monitoring the individual components of population change, the HRD will capture these changes after regular data merging with other official sources. The inter-censal population estimates will be derived from the HRD which provides estimates on the Singapore residents (citizen and PRs) count, and administrative records which provide the foreign population count.

40. This would be similar to the experiences of countries that had been using the register-based approach to Census, such as Finland, Netherlands, Norway and Sweden. Their basic population counts during Census and inter-censal years were derived from national population registers. Comprehensive population and housing censuses were selectively undertaken to provide vital detailed information about the population. They served as valuable supplements to basic population statistics.

41. The decision to change to a register-based approach for Census 2000 was not only because benefits outweighed costs, but was also a strategic move. This change would eliminate the inaccuracies of inter-censal estimates as population estimates from the traditional approach would only get updated once every ten years when the next Census is conducted. The inter-censal estimates would be able to provide basic profile of the population such as age, ethnic group, sex, house-type and other information, previously not possible with the traditional approach.

42. The scale and size of future censuses in Singapore would be contained within a comfortable limit, thereby eliminating large shocks to the statistical system in terms of planning, logistics and manpower needs. More importantly, it minimises respondent burden significantly and improves timeliness of census results. The Singapore Department of Statistics envisages that the use of the HRD for future Censuses coupled with smaller scale surveys would continue to bring about greater benefits to data producers and users as well as the population.

Singapore Department of Statistics
June 2002