

**Combining Survey and Administrative Data for  
Singapore's Census of Population 2000.**

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, moving its concept of population coverage towards a de jure count. With globalisation leading to the frequent travelling of Singaporeans and more foreigners staying long term in Singapore, this concept would more accurately reflect Singapore's population size. Three key issues were considered carefully before the Singapore Department of Statistics finally implemented this new approach:
  - 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.
3. The Singapore Department of Statistics studied carefully into the conceptual framework and practical issues before the switch. It undertook full examination of data quality issues and studied the differences of historical data produced from using the two approaches. 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 thoroughly 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.
4. With Singapore's shift in approach for Census 2000 to a register-based one, the traditional method of employing a large number of fieldworkers to enumerate every household and person was abandoned. Administrative records were used to obtain the complete population count, based on a person's place of usual residence.

5. This paper presents the rationale for the conceptual change in the definition of population, explains steps taken to improve the quality of statistics from the administrative records, and summarises the difficulties and benefits of using a register-based approach combined with a survey to provide benchmark Census data.

### **Past Censuses and the De Facto Population Count**

6. 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. The de facto approach is used in about 160 of 237 countries (68 percent) which conduct censuses. Some of these countries include Australia, UK and most of the African countries. This was also the approach used by Singapore for Population Censuses in 1990 and before.
7. 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 residents 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.
8. 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. Since Census 1990 was the first time that 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

9. 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.
10. 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.
11. Strictly speaking, a de jure population count will include whole households that are temporarily out of the country and any citizen or Permanent Resident (PR) who is away on Census Day, but with their usual place of residence in the country. Temporary or transitory visitors including tourists would be excluded. About 50 of the 237 countries (21 percent) adopt the de jure concept. Some of the countries include Germany, Switzerland, Canada and most of the Scandinavian countries. In year 2000, Singapore adopted a de jure concept upon the use of a register-based approach for its Population Census.
12. 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.
13. Technically, the HRD formed the core database on population for the conduct of the register-based Census 2000. 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 Citizen and Singapore Permanent Resident) 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:

<i>Data Item</i>	<i>Frequency</i>
• 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

14. 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.
15. 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. It was set up using the records of all dwelling units listed during the 1980 Census of Population. Thereafter, the records were updated monthly from administrative sources and through field checks, and used as the basis for many surveys conducted by the Government.
16. The HRD maintained the following categories of addresses for residents (citizens and permanent residents):
- i. Valid local addresses,
  - ii. Valid foreign addresses without local address,
  - iii. Last known addresses, which could be invalid or blank.
17. The definition of “usual residence” was applied for all citizens and PRs with valid local addresses, as matched with the NDD. The information about the local addresses were further supplemented with the 6 digit postal codes provided by the Singapore Postal Authority.

### **Total Population Count**

18. 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.

19. 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.
20. Foreigners, with 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**

21. 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

22. 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

23. 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

24. 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. The following actions were taken for each of the problem area identified resulting from the above study:
- i. *Age Group 5-14 Years*  
Many in this age group had blank, invalid or unknown addresses due to movement of houses. Discussions with Ministry of Education (MOE) started in May 98 to obtain up-to-date addresses of school going children. MOE data had then been obtained and updated to the HRD. Consequently, those records with blank, invalid or unknown addresses had been significantly reduced from 7 percent for 5-9 years and 5 percent for 10-14 years to 3 percent and 1 percent respectively.
  - ii. *Age Group 0-4 Years*  
There were blank addresses for children aged 0-4 years as their parents moved home and there were no means to update their children's addresses. Programmes had been written to link the address records of the newborn to their mothers' addresses. Furthermore, MOE data would be used to update their addresses when these children reached school-going age.
  - iii. *Age Group 75+ Years*  
The elderly population in the HRD was found to be significantly larger than that indicated in the official population estimates. Information were obtained from aged institutions and old folks' homes and merged into the HRD to update their latest addresses. A specific survey was also conducted to verify the addresses of the elderly in Aug 98 - Apr 99. Results showed that among the elderly aged 90+ years, most had returned to home countries or passed away overseas. Among the elderly aged 85-89 years, about half were still found in Singapore. The data for the 75-84 years age group showed that the majority was still in Singapore. All results were used to improve the quality of data stored in the HRD and to provide a true count of the elderly population.

### **Use of 20 Percent Sample Survey in Census 2000**

25. 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. The sample frame was obtained from the NDD.

26. 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 field. 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

- 27. 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.

## **Problems of Matching Survey and Administrative Data**

### Possible Problem of Continuity with Change in Concept

- 28. The change from a traditional Census to a register-based Census entailed a change in concept from a de facto to a de jure count. This was not an acute problem for Singapore. In 1990 and earlier Censuses, Singapore's population was based on modified de facto, as all persons found in Singapore on Census day as well as those who were Singapore citizens or Permanent Residents (PRs) but were overseas on a temporary basis, were included in the population count. This was because a purely de facto count would exclude Singaporeans who were away on Census day but were normally members of a household.
- 29. Singapore's Census 2000 made use of the register-based approach of HRD to obtain a purely de jure count. It was imperative to go fully de jure in 2000 because of more Singaporeans working or studying overseas, and more foreigners working or staying long term in Singapore. The distortions resulting from excluding citizens and PRs who were overseas in 2000 and beyond would be significant for a small country.
- 30. The total population count from 1990 Census based on the de facto concept was 3,016,400. After adding 30,800 overseas Singapore citizens and permanent residents who had their usual residence in Singapore, the total population in 1990 based on the de jure concept was 3,047,100.

The impact of the change in concept on the population data was 1 percent (Table 1). The population estimates for the period 1991-1999 were updated to take into account the "error of closure" arising from the difference between the population estimate for year 2000 and the actual census count for the same year. The following table showed that the revised concept did not result in significant change to the population estimates.

*Table 1 - De Facto & De Jure Inter-censal Population Estimates*

Year	Population (De Facto)	Population (De Jure)	Difference	
			No.	%
1990	3,016,400	3,047,100	30,800	1.02
1991	3,089,900	3,135,800	45,900	1.49
1992	3,178,000	3,232,100	54,100	1.70
1993	3,259,400	3,315,400	56,100	1.72
1994	3,363,500	3,421,100	57,500	1.71
1995	3,467,500	3,525,600	58,100	1.68
1996	3,612,000	3,670,400	58,400	1.62
1997	3,736,700	3,793,700	57,000	1.53
1998	3,865,600	3,922,000	56,300	1.46
1999	3,893,700	3,950,900	57,300	1.47
2000	-	4,017,700	-	-

### Database linkages

31. Different government agencies had different software database platforms. Co-operation amongst government agencies played an important role in database merging. Much preparatory work was carried out to overcome the differences when HRD was first set up in 1996. Singapore Department of Statistics continues to work with various agencies to resolve technical issues in data file format, data coverage and item specifications.

### Erroneous Administrative Records / Coverage Errors

32. Administrative records also contained errors. The Singapore Department of Statistics endeavoured to identify and resolve all erroneous records as

far as possible prior to the start of Census 2000. The more significant steps taken to improve the HRD are provided in the earlier section.

### Classifications - Occupation, Industry and Education

33. Changed administrative definitions or classifications might cause difficulties in maintaining consistent definitions across different government agencies, and across time. The difference in address formats used by different agencies is an example. It was therefore necessary to develop national standards through inter-Ministry consultations and the promotion of the adoption of these national standards by the different agencies. For housing records, the address standard key is the national coded address (NCA) format. The Department of Statistics had taken the lead as the national statistical coordinator in completing the year 2000 revision of the occupation, industry and education codes which were the latest in the series of ongoing updates to the classifications.

### Confidentiality and Security

34. The move towards an integrated database through merging of particulars also meant that breaches in data security would have more widespread consequences. It could be an illegal access to the database or the accidental or intentional alteration or destruction of partial data. As such, much care and attention was paid to safeguarding the data integrity and confidentiality as provided under the Census Act. The use of stringent administrative and technical measures to protect the information obtained was institutionalised and strictly adhered to. The confidence of respondents was maintained to the highest level possible at all times.

### Specific Reference Time Point

35. The administrative database HRD was able to capture the most up-to-date data precisely as at 30 Jun 2000, the exact Census Day for Singapore, to provide a basic count and profile of Singapore's population. The additional 20 percent sample survey conducted within the Census, which relied on the database's records with an earlier cut-off date would require survey respondents to update their information between the cut-off date and the actual Census day. Some information such as marital status, address etc. may change. All other additional information referred to the Census day.

## **Benefits of the Register-Based Approach**

### Timeliness of Data Release

36. 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 in Dec 2000, as compared to May 1991 for the 1990 Census.

### Reduced Respondent Burden

37. 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.

### Variety of Data and Data Quality

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

### Cost and Manpower Reduction

39. 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

40. 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.
41. 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.
42. 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.
43. 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.
44. With this change, 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. In time to come, with continued checks on the HRD and its improvement, more information on the population such as marital status, education etc. would be available.
45. 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, and minimising respondent burden. 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.

4 July 2001