

**18th Population Census Conference
26 - 29 August 1998, Hawaii, USA**

**Planning for the
Hong Kong 2001 Population Census**

**by Alvin W K LI
Assistant Commissioner for Census & Statistics
Census and Statistics Department
Hong Kong, China**

18th Population Census Conference
26 - 29 August 1998, Hawaii, USA

Planning for the Hong Kong 2001 Population Census

Introduction

1. Hong Kong, with an area of 1096 square kilometres, has a population of 6.6 million people at the end of 1997. Over the years, Hong Kong (HK) has undergone remarkable social and economic developments to become one of the world's main financial and trading centres. On 1 July 1997, Hong Kong reunited with China¹. The transition, which took place very smoothly, signifies the start of a new era for Hong Kong.

2. Starting from 1961, it has been established practice for HK to conduct a population census once every ten years. Population censuses were conducted in 1961, 1971, 1981 and 1991. In order that more up-to-date information on the population could be made available, a population by-census is also conducted in between two censuses. Thus, in 1966, 1976, 1986 and 1996 there were population by-censuses conducted. A by-census differs from a full census in not having a complete headcount of the population but simply enquiring on the detailed characteristics of the population on the basis of a large sample. The size and characteristics of the entire population are inferred from the sample results in accordance with appropriate statistical theory. As a result, the statistics obtained from a by-census may be slightly less precise than those from a census. As regards data content, it would be quite similar between a census and a by-census.

3. The aim of conducting censuses and by-censuses is to obtain up-to-date information on the population of HK, which is vital to government planning and policy formulation, particularly in such fields as education, housing, transport, medical and social services. The data are also important to the private sector for formulating business strategies and to academic researchers for conducting social and economic studies.

¹ The Government of the People's Republic of China resumed its exercise of sovereignty over Hong Kong on 1 July 1997. On this date, the Hong Kong Special Administrative Region (HKSAR) was formed under the "One Country, Two Systems" principle. As stipulated in the Basic Law of the HKSAR, the Region would operate with a high degree of autonomy. It would maintain its way of living after the Handover and would preserve a full range of elements underpinning Hong Kong's achievements in terms of prosperity and stability. In respect of official statistics, the HKSAR would maintain a separate statistical system from that of the mainland of China and continue to compile and disseminate statistical data about the HKSAR. As regards the Census and Statistics Department, it would be responsible to the Government of the HKSAR and there would be no administrative relationship between it and counterpart departments in the mainland of China.

4. Statistics compiled from information collected in censuses and by-censuses have very good precision, even when the data are classified by detailed population sub-groups and small geographical areas. They therefore differ from other sample surveys on the population, which normally employ much smaller samples and tend to provide only limited statistics on detailed population sub-groups and small geographical areas.

5. In line with established practice, a population census will be conducted in March 2001. This paper provides an overview of the current planning for the Hong Kong 2001 Population Census (01C), with focus on the development of required computer systems and the use of information technology. Reference is made to the techniques and technology used in previous rounds of population census and by-census, wherever deemed appropriate.

Coverage

6. The 01C will cover the entire population of HK. Information will be solicited from all the residents, including those temporarily away from HK. For persons who are not residents, such as tourists, only count figures will be obtained.

7. Like the 1996 Population By-census, the 01C will adopt the *de jure* enumeration approach. Under this approach, HK residents will be enumerated and asked to provide their detailed characteristics at their usual place of residence. During the field operation, the household head will be asked to give a list of household members, each of whom will be asked a series of questions to ascertain whether he/she is a usual member of a household in HK.

Enumeration method

8. In the 01C operation, enumerators will visit all quarters (or vessels for the marine population) in HK. Six-sevenths of the population will be subject to a simple enumeration and enquiry on basic characteristics such as age and sex. The remaining one-seventh of the population will be selected for detailed enquiry on a broad range of socio-economic characteristics. There will hence be two types of questionnaire: short form for the simple enumeration and long form for the detailed enquiry.

9. The 'self-enumeration' approach will be adopted in the simple enumeration. Short forms will be mailed out about one week before the start of the 01C to all selected quarters with clear postal addresses for the householders

to complete. Enumerators will then visit the quarters during the field operation period and collect the short forms. For households which have completed the short forms on their own, enumerators will only check the information during the field visit. For households which have difficulties in completing them, enumerators will conduct the enumeration through face-to-face interview. Based on past experience, about 50% of the householders would complete the short forms themselves. For quarters without clear postal addresses, short forms will not be mailed to the householders beforehand but will be completed by enumerators during the visit.

10. The traditional face-to-face interviewing method will be adopted in data collection for long forms. Enumerators will visit the sampled quarters and interview each occupant therein to obtain the information. Enumerators have to make repeated visits to each household until all members of the household are interviewed. They will be instructed to visit households at different times of the day and on different days so as to enhance the chance of contacting the households.

11. Despite all these measures, there will still be some non-contact cases at the end of the operation. Therefore, on the last day of the operation, a self-administered questionnaire will be left with each of the non-contact households. They will be asked to complete and return the self-administered questionnaires by mail before a specified date. The self-administered questionnaire will be similar to the short form.

Information technology-related operations

12. To facilitate the planning and execution of the 01C and also the dissemination of census results, a suite of computer systems will be developed. A summary of the computer systems is given in the Annex. The major information technology-related procedures are described below:

Updating the frame of quarters

13. The Frame of Quarters records the addresses and the characteristics (e.g. type and usage) of quarters in HK. It is used as a sampling frame for population censuses and household surveys. The frame is currently divided into 2 parts:

- (a) Register of Quarters - A completely computerized list of addresses of permanent quarters in built-up areas; and

- (b) Register of Segments - A list of area segments in non-built-up areas with each segment containing about 10-15 temporary structures.

14. Both textual and spatial data of permanent quarters are updated continuously based on information collected from government departments, gazettes and visits by field staff. For the 01C, site visits will be conducted to update the records of the register of quarters and the register of segments. Aerial photographs of non-built-up areas will also be used in the re-segmentation process - regrouping the temporary structures such that the number of temporary structures in each segment will be around 8 -12. The existing computer system for the frame of quarters is considered inadequate for the 01C operation in areas such as map production and frame updating.

Sample selection

15. The 01C will comprise a complete but simple enumeration of all persons on their basic characteristics such as age and sex, and a detailed sample enquiry on a broad range of socio-economic characteristics of the population of HK. Sampling is hence required to select a sample of the population for the long form enquiry.

16. A stratified systematic sampling design with 2 stratification factors will be adopted. The sampling units are first stratified by quarters in built-up areas and area segments in non-built-up areas. They are then stratified by geographical area. In the 1996 Population By-census, the whole of HK was divided into 346 Constituency Areas and they were used as geographical strata. It is likely that Constituency Areas will again be used as geographical strata for the 01C.

17. Systematic sampling will be applied to each stratum. To ensure that the level of precision of estimates for small areas is acceptable, a disproportionate sampling scheme will be used such that small areas (in terms of the number of people) will have a greater sampling fraction (and hence a smaller selection interval). For each stratum, records of sampling units are first sorted according to pre-defined fields. Sampling units will then be drawn systematically following the selection of a random start and using a pre-defined selection interval.

Allocation of assignments

18. Enumerators will be asked to visit households according to the assignment list of addresses provided. There will be two types of assignment

list : one for the short form enumeration and the other for the long form enquiry. The assignment lists will be sorted into assignment sequences. The assignment sequences are generated based on the most desirable route such that the quarters assigned to each enumerator will not be widely scattered.

19. The sorted assignment lists in the assignment sequences will be divided into a number of divisions. A workload index, which reflects the amount of efforts required, will be assigned to each assignment. Once the workload index for each assignment is calculated, divisions will be formed systematically such that each division has about the same workload index value. The assignments for each division are divided into enumerator blocks such that the workload index for each enumerator block would be about the same.

Recruitment and training of temporary field workers

20. Over 22 000 teachers and students will be recruited as temporary field workers in the 01C. Allowing for drop-outs, some 30 000 teachers and students will be selected for training. To facilitate recruitment of temporary field workers, a data file containing relevant information of schools in HK will be established. This file will provide information for organizing recruitment, for procurement of field centres/training centres, and for organizing training classes.

21. Each applicant will submit an application form containing particulars such as name, sex, age and school. These data items may be in Chinese and English. Information on the forms has to be input into computer. In the 1996 Population By-census, the key-punching approach was used for input. Other input methods such as optical character recognition or optical mark reading will be considered for 01C. An application number will be assigned to each applicant for identification purpose.

22. There are different rounds of training for different ranks of temporary field workers. Applicants will be allocated to training classes based on some pre-defined criteria (such as district of residence). The availability of classrooms in each training centre will be input to control the number of training classes in each centre.

23. Payment to the temporary field workers consists of teaching bonus, training bonus and honorarium. Chief enumerators responsible for teaching enumerators are eligible for teaching bonus. Applicants who have attended training classes are eligible for training bonus. Temporary field workers having completed the enumeration satisfactorily are eligible for honorarium. The

payment is calculated based on training attendance record, number of days worked and so on, which are input into the system after the training classes and census operation.

Monitoring of fieldwork progress

24. To monitor the field operation, a network system with micro-computers will be installed in each field centre and linked up with a host mainframe computer at the headquarters through public electronic communication service. The basic functions of this system are outlined below:

(a) Monitoring progress of enumeration

It is necessary to closely monitor the enumeration progress such that remedial measures can be taken once slippage is identified. Daily reports will be produced on the enumeration progress of individual enumerators, divisions, districts, regions and the whole territory.

(b) Producing preliminary control statistics

Irregularities in enumeration results (e.g. unusually high rate of unoccupied quarters) should be identified as early as possible so that special checking or remedial measures can be taken. Control statistics such as occupancy rate, number of persons per occupied quarters and non-contact rate will be produced for individual enumerators, divisions, districts, regions and the whole territory on a daily basis.

(c) Supporting quality check measures

There will be 1 to 2 checkers in each field team to re-visit completed assignments and perform quality checks. A sample of completed assignments will be randomly selected for quality check and the results of checking will be produced for analysis and evaluation.

(d) Providing information for producing preliminary population estimates

Information on enumeration results and number of persons in occupied quarters will be used to compile preliminary population estimates by small geographical area shortly after the field operation.

(e) Facilitating communication among Field Centres and Headquarters

An efficient communication network among field centres and headquarters is essential for the field operation. In the 1996 Population By-census, communication was largely via facsimile. For 01C, in view of the large number of field centres, electronic communication network and fax will both be required.

(f) Controlling the despatch of self-administered questionnaires

At the end of the field operation, self-administered questionnaires will be issued to all non-contact cases. The assignments completed will be matched with those in the master assignment list to ensure that non-contact quarters have been issued the self-administered questionnaires.

Input of census data

25. In the 01C, it is necessary to input data pertaining to some 2.65 million households and 7.0 million people. The input of such a large volume of raw data poses a problem. The conventional key-to-disc method will not be totally adequate for inputting data collected through the short forms, long forms and self-administered questionnaires. Different methods may be used for inputting data collected from these forms.

Processing of census data

26. All input data have to be edited and validated to remove discrepancies and inconsistencies. The data are repeatedly checked by computer and manually corrected until all errors are cleared. Validated data will go through an imputation process to ensure that erroneous and missing data items of non-response cases are imputed by similar live data from the data base. After the imputed data have passed through the validation and imputation processes, they will be taken as error free and there should be no missing value. Appropriate weights will be assigned to individual records for tabulation.

Dissemination of census results

27. The 01C will collect a huge amount of socio-economic data on the population of HK. Tabulations will be produced and end-user computing supported.

28. The following media were used for disseminating the 1996 Population By-census results:

(a) Printed reports

For 1996 Population By-census, 16 publications (including 2 map packs) were produced.

(b) CD-ROM products

Two CD-ROM products were produced by the Census and Statistics Department, containing a set of 324 standard tabulations and a set of digital maps of HK for geographical presentation. More advanced products (called Value-added CD-ROM) which bundled the 1996 Population By-census data, digital map and tailor-made applications were developed by contractors through open tendering.

(c) Sample data sets

Two sample data sets of individual records were produced for use by academic researchers.

(d) Self-help tabulation service

Facility was provided to outsiders (from both public and private sectors) to compile statistical tables themselves on site in the census office.

(e) Tailor-made tabulation service

Tailor-made statistical tabulations were produced upon user's request. The tabulations were largely compiled by end-user computing using SAS.

29. These means will also be used for disseminating 01C results. In addition, the following will also be considered:

(a) Internet

A set of statistical tables on the main findings of 01C will be prepared and disseminated via the Internet free of charge.

(b) Data warehousing

An efficient and convenient data retrieval system will be set up to enable users to choose data items and then retrieve the statistical data required.

Changes and improvements under consideration for the 01C

30. The 01C is the single most extensive, complicated and costly statistical operation to be conducted in HK. It is complex and intricate, with a multi-year planning cycle and a lengthy data dissemination schedule. Experience gained from and statistical infrastructure developed for previous rounds of census and by-census are most valuable for the planning and execution of the 01C. There will be changes in the 01C, some of which exploit advance in information technology. Discussed below are some of the changes and improvement being considered for the 01C.

Data topics

Consultation

31. To meet the data requirements of users, two rounds of consultation have been conducted, one with government departments and the other with non-government organizations, including the Universities. There were suggestions on the addition of a variety of data topics. In deciding on the topics for the 01C, the following factors have to be considered:

- (a) Usefulness of data
- (b) Willingness and ability of the respondents to answer
- (c) Ability of the enumerators to comprehend
- (d) Existence of other data source/data collection channel
- (e) Processing efforts
- (f) Wide applicability
- (g) Comparability of data with previous censuses/by-censuses
- (h) Keeping interviewing time reasonable
- (i) Recommendations of international organizations

Proposed addition of topics for 01C

32. Respondents were asked in the consultation to suggest new data topics for 01C. A broad range of topics on the socio-demographic characteristics of the population have been suggested. They are assessed using the factors deliberated above.

33. Some topics cannot be asked in the 01C for technical and other reasons. For example, the suggested topics on ‘value of assets’ and ‘detailed breakdown of cash income’ may be sensitive to some respondents and accurate answers cannot be collected through a population census. The suggested topics on ‘previous occupation’ and ‘previous industry’ require enumerators to decide on-the-spot whether the respondents have changed ‘industry’ or ‘occupation’. This is clearly beyond the capability of the enumerators who are temporary staff.

34. There are many suggested topics which, though feasible for inquiry in a household survey, would better be asked in a survey of a smaller scale because very detailed breakdown of results is not required. Examples of these topics include ‘smoking habits’, ‘religion background’, etc. In fact, the Census and Statistics Department operates a continuous General Household Survey (GHS) which, in addition to collecting regularly information on the labour force, unemployment and underemployment, solicits information on an ad hoc basis on a wide range of special topics required by government departments and policy bureaux. As such, those topics would better fit into the GHS.

35. Some suggested topics require many questions in order that the information collected will be useful. These topics should better be asked in a dedicated survey. Topics on transport characteristics would fall into this category and they would better be consolidated into a special survey on transport characteristics.

36. After deliberation, a study is being carried out to add the following topics to the long form:

(a) Mortgage re-payment

This topic is the most commonly suggested in the consultation. The topic should help study the affordability of housing in HK.

(b) Mode of transport to place of work/study

This is the second most common topic suggested in the consultation. Since place of work/study will be asked in the 01C, the additional efforts required for asking the “mode of transport to place of work/study” will not be too great. To keep the interviewing time reasonable, only the mode used for home to place of work/study (and not the return trip) will be asked. Whether asking all modes used in a single trip or just the principal mode used (e.g. the mode used for the greater part of the trip) will be further explored.

(c) Highest level of education completed

In past censuses/by-censuses, highest level of education attained rather than completed was asked. The additional effort required for soliciting information on “highest level of education completed” will be small.

(d) Number of rooms, excluding kitchen and bathroom

This indicator aims to measure the level of overcrowding and privacy and thus the quality of life of the residents.

Proposed deletion of topics for 01C

37. In order to keep the interviewing time to a reasonable limit, it is proposed to delete some less essential topics, which were included in the 1996 Population By-census. The topics proposed to be deleted include : (a) ‘previous district of residence’ and ‘duration of residence in present district’. Half of the policy bureaux/government departments rated these topics as neither ‘essential’ nor ‘desirable’ in the consultation. Indeed, such data topics are considered not very useful as they are quite similar to the data topic ‘district of residence five years ago’, which will be included in the 01C.

Map-based Register of Quarters System (MRQS)

Existing system

38. For sampling and field operation purposes, the 01C requires a complete and accurate frame of quarters in HK. The existing frame of quarters was established in the late 1960’s. The part of the frame for built-up areas, i.e. Register of Quarters, was computerized in 1985. The existing computer system for the Register of Quarters was re-developed in 1995 for conducting the 1996 Population By-census.

39. The other part of the Frame of Quarters, viz. Register of Segments, is not completely computerized. Details of the segments can only be retrieved manually. Each area segment is identified by a description sheet and location map. Currently, there are about 16 000 segments containing about 160 000 quarters.

40. To support the Frame of Quarters, a total of 12 000 maps of HK are maintained, including 3 000 survey maps of scale 1:1000, 6 000 street block

maps and 3 000 segment maps. Information like segment boundaries and special annotations (to facilitate field operation) is added on the paper maps manually. The maps are quite old and updating is quite tedious and time consuming.

Basic requirements of MRQS

41. With the availability of the complete set of 1:1000 digital maps of HK, it is intended to incorporate digital maps in the Frame of Quarters. The MRQS is planned to be set up for storing and manipulating the digital maps and individual records of quarters in both the Register of Quarters and Register of Segments in an integrated manner. The existing Frame of Quarters will then be replaced by the proposed MRQS.

42. The requirements of MRQS broadly consist of 2 main parts, viz.: data conversion and purposely-built applications. The data conversion requirements are:

- (a) The MRQS will use the 1:1000 digital maps of Lands Department as base maps. The information added on the existing paper maps has to be digitized as separate layers.
- (b) Building polygons of the digital base maps will be used as the key for linking the existing Register of Quarters records with the digital maps. The existing Register of Quarters records have a building key which can link up the building polygon on a 1:1 basis in most cases but many-to-one and one-to-many linkages between Register of Quarters records and building polygons may exist.
- (c) The case of Register of Segments is more complicated. Records of the Register of Segments contain information of individual quarters in the segments. The quarters should be linked to building polygons of the digital base maps, but Register of Segments may contain temporary structures not shown on the digital map. In these cases, the outline of the structures on the paper maps has to be digitized first. Like records of the Register of Quarters, many-to-one and one-to-many linkages may exist.

43. MRQS should include purposely-built applications to facilitate the use of the digital maps and records of the Register of Quarters and Register of Segments. Specifically, the functions below are required:

- (a) Address searching

The address of a building can be searched by pointing to the building on the digital map. Alternatively, the location of a building on the digital map can be searched by entering the address.

(b) Map production

Tailor-made maps can be printed. Users may specify the map scale, paper size and boundary (e.g. by address or coordinates), and add titles to the maps.

(c) Line drawing

User-friendly facilities for drawing lines and polygons, and adding text to different layers of the digital maps are required.

(d) Online enquiry of buildings/quarters information

Users can obtain detailed information of a building or quarters by entering its address or pointing it on the digital maps.

(e) Information updating

User-friendly facility for updating and amending textual and map data should be provided.

(f) Compilation of quarters statistics reports

Statistics on the records kept in the system should be produced regularly.

(g) Compilation of management statistics reports

Management statistics of the system (e.g. no. of amendments) should be produced regularly.

(h) Sample selection

Samples can be selected by specifying criteria like type of quarters covered and number of sampling units. Facility to print sampled addresses in specified formats (e.g. on gummed labels or as control lists) should be provided.

(i) Storage and compilation of images

The system should be able to store and manipulate images (e.g. photos or scanned sheets) and relate them to individual buildings and/or quarters.

Data capture

44. Data capture will be required for data collected from the following:

- (a) short form;
- (b) long forms;
- (c) self-administered questionnaires; and
- (d) application forms of the temporary field workers.

45. In the 1991 Population Census, data collected from the short forms and self-administered questionnaires were transcribed onto optical character recognition (*OCR*) forms for scanning. The electronic images thus captured were converted into numeric codes and these became the data file for short forms. Similar input method will be used for 01C. However, the feasibility of direct scanning of short forms will be studied. Direct scanning without transcription onto OCR forms avoids transcription errors and reduces enumerators' workload.

46. The traditional key-to-disc method was adopted to key in data recorded in the long forms and the application forms for temporary field workers. This approach may not be cost-effective and efficient when compared with the imaging approach. Imaging techniques embrace machine-print OCR, hand-print OCR, OMR, bar code and key-from-image (KFI). These techniques enable information on long forms to be captured without substantial additional manual efforts.

47. In particular, if imaging technique is used, the whole coding/editing process can be performed on-line with reference to the images retrieved on screen; thus manual retrieval of questionnaires for coding, data input and on-line editing is not required. This will greatly improve the efficiency and shorten the processing time. Specifically, tracking of questionnaires will no longer be required.

Concluding Remarks

48. The 01C is a large-scale, complex statistical operation. It poses a challenge to those involved in its planning and execution. As planning of the 01C is still in its early stage, it is not yet possible to finalize the approach to be adopted. Nevertheless, opportunity is being taken to make use of advanced information technology for the conduct of 01C. The changes discussed above serve to illustrate the initial thinking of the planning team. Other changes may be brought in as the planning progresses.

Annex

Data Processing Systems for the 2001 Population Census

A suite of data processing systems will be required to support various activities of the 2001 Population Census, ranging from development of the quarters frame to data dissemination. The computer systems and their major functions are shown below:

Systems	Major functions
(a) Map-based Register of Quarters System (<i>MRQS</i>)	<ul style="list-style-type: none"> • To store and manipulate quarters records and digital maps in an integrated manner.
(b) Sampling System (<i>SAMS</i>)	<ul style="list-style-type: none"> • To select samples from MRQS.
(c) Blockcutting System (<i>BCS</i>)	<ul style="list-style-type: none"> • To allocate assignments to enumerators. • To produce maps to support field operation.
(d) Enumerator Information System (<i>EIS</i>)	<ul style="list-style-type: none"> • To store and handle individual temporary field workers records. • To support recruitment, selection, deployment, training and payment of TFWs. • To produce householder letters.
(e) Hotline Enquiry System (<i>HES</i>)	<ul style="list-style-type: none"> • To handle public enquiry on field operation. • To check identity of enumerators online.
(f) Fieldwork Control System (<i>FCS</i>)	<ul style="list-style-type: none"> • To monitor field operation progress. • To support quality check during field operation. • To compile preliminary population estimates. • To facilitate communication among field centres and headquarters.
(g) Questionnaire Tracking System (<i>QTS</i>)	<ul style="list-style-type: none"> • To track questionnaire movements between different processing sites, if required.
(h) Data Input System (<i>DIS</i>)	<ul style="list-style-type: none"> • To input the data collected.
(i) Self-Administered Questionnaire System (<i>SAQS</i>)	<ul style="list-style-type: none"> • To process the returned SAQs.
(j) Data Validation and Editing System (<i>DVES</i>)	<ul style="list-style-type: none"> • To validate and correct erroneous data.
(k) Imputation System (<i>IMS</i>)	<ul style="list-style-type: none"> • To perform imputation.
(l) Grossing Up System (<i>GUS</i>)	<ul style="list-style-type: none"> • To gross up the data.
(m) Statistics Dissemination System (<i>SDS</i>)	<ul style="list-style-type: none"> • To prepare statistical tables. • To facilitate statistics dissemination by other means. • To perform statistical analyses and desk-top publishing.