香港中文大學

心中究研會社

# Sampling in the Kwun Tong Industrial Community Research Programme

Y. K. Chan

SOCIAL RESEARCH CENTRE

THE CHINESE UNIVERSITY OF HONG KONG

**Suggested citation:**

Chan, Y. K. 1972. *Sampling in the Kwun Tong Industrial Community Research Programme*. Hong Kong: Occasional Paper No. 3, Social Research Centre, The Chinese University of Hong Kong.

The Chinese University of Hong Kong
Social Research Centre

# SAMPLING IN THE KWUN TONG INDUSTRIAL COMMUNITY RESEARCH PROGRAMME

by

Y. K. Chan

February, 1972.

CONTENTS

## I. Introduction

A series of surveys have been carried out in the Kwun Tong Industrial Community Research Programme*. For both the Kwun Tong Life Quality and the Kwun Tong Health System Studies it was necessary to interview residents living in different types of housing in Kwun Tong, and a probability sample survey covering the whole population was required. Taking interviewing expenditure into account in our budget estimates, we calculated that 1000 and 700 sample units would be reasonable for the Life Quality Survey and Health System Survey respectively.

The purpose of this paper is to relate our experience in the Kwun Tong studies to standard techniques of sampling in the hope that it may contribute to the improvement of our future projects.

## II. The Sample Frame

In constructing the sample frame, we used the stratification technique[1]. The population was stratified using type of housing

---

* This paper is produced for the Kwun Tong Industrial Community Research Programme which was financially supported by the Harvard-Yenching Institute and was under the auspices of the Social Research Centre, the Chinese University of Hong Kong.

1 Since precise data about certain subdivisions of the population were required, and stratification also yields better estimates of characteristics of the whole population.

and location as criteria, because ancillary information about housing and population was obtainable from every Resettlement Estate, Low Cost Housing Estate, Cottage Area and Squatter Area[2]. Accommodation standards vary greatly from one type of housing to another, and among various sub-districts. In addition, type of housing is a very important variable in studies on the residents' living standard, income, socio-economic status, level of education and life style in general.

In our sample, housing in Kwun Tong was divided into four main categories (Appendix I) according to their characteristics — rent, space, location etc. Different sample designs were made for Private Housing (multi-storey buildings), Public Housing (including Resettlement Estates and Low Cost Housing Estates), and Cottage/ Squatter Areas, one for each. Data were collected from the above types of housing which constitute 96% of the total population of Kwun Tong District. We excluded watchmen and caretakers, etc., living in the industrial area, and those living in squatter boats.

1. Private Housing

Sample units were taken randomly from a total of 10,817 flats (residential units)[3]. The sample fraction used here was the

---

2 However, to adopt housing as a criterion, we had to prelist units in private housing.

3 We define residential unit as:
-- a room in a Resettlement Estate
-- a room/flat in a Low Cost Housing Estate
-- a flat in a multi-storey building (private housing)
-- a structure in a Cottage/Squatter Area.

same as in other type of housing. Based on our Landuse Survey carried out in April 1971, a complete list of flats for domestic use (excluding dormitories) in multi-storey buildings was established according to street names and in numerical order. Flats to be interviewed were drawn at random.

We could not do the prelisting down to the level of the household, for calling at ten thousand flats and listing all households would have required much more manpower and time than was available. It needed 25 mandays to prelist all the flats of private housing in Kwun Tong, to prelist the actual households would have tripled expenditure of money and time. "Multi-stage" sampling method, as employed by the Institute for Social Research (University of Michigan)[4], is very suitable for a nation-wide sample survey, and can reduce the difficulties of prelisting. But Kwun Tong District is only part of the Hong Kong metropolitan area, and the residents of private domestic buildings comprise only 13% of Kwun Tong's population. Obviously our surveys are different from nation-wide surveys in other countries. Since we could afford the manpower to prelist all domestic flats in private domestic buildings in Kwun Tong, we used the simple random method to draw sample units in private housing. In this way we also simplified the working out of estimates.

---

4 Interviewer's manual - Section C. Survey Research Centre, Institute for Social Research, The University of Michigan. Ann Arbor, Michigan, 1969.

2. Public Housing

In Resettlement Estates and Low Cost Housing Estates, most blocks are of uniform design. Data on population and number of households are obtainable from the Estate Office, and there is only one household in each residential unit. This facilitates the drawing of a complete list of units for sampling.

There are different types of accommodation in Resettlement Estates (Marks I, II, III, IV, V etc.) and various sub-types in each Mark.[5] We can list and stratify residential units in estates of different locations according to their type and sub-type. Sample units of every sub-type can then be drawn by random sampling. Government Low Cost Housing Estates, Housing Authority or Housing Society Estates can be treated in the same way. We can list residential units according to their size and location, and then take a random sample.

The sample frame for Public Housing is shown in Appendix IIa and IIb. We used location, type and sub-type as criteria for stratification because communication and interdependency between an estate and town centre or another estate are variable. Different estates of different types may be built in different years; their accommodation standards, population size and service facilities differ[6].

---

5 "Annual Department Reports - Commissioner for Resettlement" Hong Kong Government Printer.
Also "Main Facts on Resettlement" Resettlement Department, Hong Kong Government (Quarterly).

6 In Mark I type Resettlement Estates (e.g. Jordan Valley Resettlement Estate), one resident has only 24 sq. feet living space, and toilet and kitchen are communal. But in Mark V type Resettlement Estates (Ngau Tau Kok Resettlement Estate), the living space of each resident in 35 sq. feet, as is the case in Low Cost Housing Estates, and every unit has its own toilet and kitchen. In Ngau Tau Kok, Ham Tin, Sau Mau Ping Resettlement Estates, there are building for schools and Social Welfare Agencies, but these are lacking in Jordan Valley or Kwun Tong Resettlement Estate.

3. Cottage/Squatter Areas

About 9% of the population of Kwun Tong District are living in this type of housing, which includes Licensed Area, Resite Area, Resettlement Area and Squatter Villages. Sample units were arranged by area, and the required number of structures (residential units) in each area were chosen by random.

We used different numbers of criteria in the procedure of stratification for various kinds of housing - for Public Housing, we used location, type and sub-type; for Cottage Areas we used location; for Private Housing, we employed only simple random. This was because of the availability of ancillary data for different types of housing, and also because of the heterogeneity within the types of housing, particularly in the category of Public Housing as explained above.

III. Sample Fraction

The sample was designed to be a proportionate stratified random sample. The same sample fraction (1.48% for the Life Quality Survey and 0.78% for the Kwun Tong Health System Survey) was applied to each stratum; in other words, this was a self-weighted sample with unified sample fraction. Each housing unit in every type of housing had equal probability of being taken into the sample. This reduced the work required to weight data when making estimates.

Among the four types of housing, Cottage Areas have the smallest number of residential units; even so there are more than seven thousand structures in all Cottage Areas and though a sample fraction of about 1% was taken, the sample size was not too small —— 75 in total. In the Public Housing category when a sub-stratum contained too small a population, instead of raising the sample fraction, we grouped these small sub-strata together according to their characteristics, so the sample remained a self-weighted one.

## IV. Sample Units

In our sample design, as we have mentioned above, the elementary units are residential units[7]. But the ultimate units were the households, and the household head or his wife was taken to represent the household. For Public Housing, this design posed no trouble at all, because in every residential unit there is only one household. But in flats of multi-storey buildings and in cottages, there may be two or more households living in the same unit (though this is quite rare according to the results of our surveys and the census data)[8]. In this case, residential units must be treated as a small cluster[9], each consisting of one or more households. We must bear this in mind when we analyze our data.

---

7 Elementary unit is the small unit which could conceivably be sampled. The sample unit which is completely surveyed is the ultimate unit. See also - A Short Manual on Sampling, Vol.1. Statistics Office of the United Nations, United Nations, New York, 1960.

8 According to the crude result of the Life Quality Survey, the average number of households living in each flat is 1.04.

9 We will discuss this in detail in the section under the heading "Problems and solutions".

V. Advantage of the Sample Design

Theoretically speaking, when precise information is required about certain sub-divisions of the population, it is necessary to adopt the technique of stratification. In both the above-mentioned surveys, we had no trouble in employing this technique. For administrative convenience and because type of housing is an important variable in both studies, stratification by area and by housing type was time-saving and was believed to be most suitable. We therefore designed our sample frame accordingly.

In the sample, thought certain types of housing were broken into many small sub-strata and the procedure of working out estimates was made more complicated, it was still worthwhile doing this as it would make comparisons easier in studies of life quality, socio-economic status, education level or even of the income of different sub-divisions of the population living in Kwun Tong in various ways. (As we could see from the ancillary information, the characteristics of residents living in different kinds of housing are not alike).

With our sample design, data can be grouped by housing type, or by location (by sub-district). Furthermore, for the category of Low Cost Housing, data can be grouped by sub-type —— Government Low Cost Housing Estates, Housing Authority Estates, and Housing Society Estates, etc., or according to the location of the estate; then, data can also be

regrouped by type of accommodation (e.g., flats for 4,5,6, ..... persons). For Resettlement Estates, data may be grouped by location, by Mark (the accommodations in different Marks vary greatly), or in greater detail, by sub-type (½D, D, E, A, Double, Large, Standard, Medium, Small etc.)[10]. Consequently, variables may be easily compared between different groups as required, and higher precision of data on every sub-division of the population is possible.

Since characteristics of the population vary greatly from area to area and from one type of housing to another, so instead of using only one criterion for stratification, we used more than one variable as criteria for stratification. In both Life Quality Survey and Health System Survey we used location, housing type and sub-type. This we consider to be the best way to choose a sample in Hong Kong, and it is particularly appropriate in Kwun Tong.

## VI. Problems and solutions

### 1. Household and sample unit

In stratified sampling, we may probably face various kinds of difficulties in different sub-divisions of the population. There was no exception for our surveys in Kwun Tong. In the procedure of drawing sample units, we did not meet any problem in Low Cost Housing Estates, because there was only one household in every residential unit

---

10 See note 5 above

(or Room). But in Resettlement Estates, the situation was quite different. We found that some families were living in more than one residential unit. But these units might not be adjacent to each other. They might be on separate floors, or even in separate blocks. In this case, we can treat that family as two or more separate households, or combine them into one. The latter method is more reasonable but the former one causes less statistical problems. Of course, how to make the choice is up to the requirements of the project.

In multi-storey buildings or in cottages the situation differed again. One residential unit might be shared by two or more households (quite common in Hong Kong). Therefore, when making estimations, we should treat the residential units in Private Housing and in Cottage Areas as small clusters, and corresponding adjustments must be made. In Life Quality Survey, where more than one household was found in a flat being drawn into the sample, all these households were interviewed; so when analysing our data, we should use the process technique of Cluster Sampling - with equal probability and unequal clusters[11]. But in Health System Survey, only one household (the one who answers the door) in a flat was interviewed. In other words, flats

---

11 United Nations, A Short Manual On Sampling, Vol.1, Dept. of Economic and Social Affairs, United Nations, New York, 1960. pp. 120-129. Also L. Kish, Survey Sampling, John Wiley and Sons, Inc., New York, 1965. pp. 182-216.

and cottages were treated as primary sample units and households were second-stage units. Or we can say flats and cottages are small but unequal clusters, selecting one household out of a flat or a cottage produces a sample of elements with unequal probability, and the probability of a household being drawn into the sample remains unknown until the residential unit is interviewed. Thus different weights must be given to each small cluster when working out estimates.[12]

2. Non-response

Besides the above-mentioned problems, we faced problems of non-response too.[13] According to the crude results of our surveys, the non-response rates[14] were rather high, 40.4% for Life Quality Survey and 27% for Health System Survey. Among those non-responses many were refusals (145 out of 259 in the Health System Survey). This is partly due to the fact that Hong Kong people are not used to

---

12 L. Kish, op. cit., pp. 154-161, 396-404, 424-429.

13 When we encounter a deliberate reject by a respondent, we give up the case and treat it as a refusal. If no one answers the door, or some one answers the door but no qualified respondent is present, more successive visits (2 in Life Quality Survey and 3 in Health System Survey) will be made to the same unit at different time of the day. If still no qualified respondent is present the case is classified as not-at-home. If appointment can be made at the last call, one more visit must be paid to that unit.

14 $$\frac{\text{No. of non-responses}}{\text{No. of non-response + No. of responses}} \times 100\%$$

this kind of Social Survey interviews, and to the deterioration of law and order in Hong Kong recently, so people are unwilling to open their doors to strangers. Sometimes when the husband is not at home, the housewife may not dare to open the door to interviewers. In many cases, both spouses go to work during the day time, and the interviewer could not find a qualified respondent in the household. In addition, we were not allowed to visit residents after 8 p.m. in certain Low Cost Housing Estates, which increased the number of non-response cases too.

The non-response rate was higher in Life Quality Survey than in Health System Survey, because for the latter interviewers were requested to make at least four successive visits to any residential unit in case no respondent was at home, instead of three successive visits in Life Quality Survey, thus the "not-at-home" rates were 30.6% and 11.8% respectively.[15] Also, based on past experiences of Life Quality Survey, fieldwork supervision and control methods had been improved, thus yielding better results for Health System Survey. Besides, different attitudes of eligible respondents towards various kinds of questionnaires might also be an important factor.

For both Life Quality and Health System Surveys, a high non-response rate was anticipated. So a reasonably large initial sample was first established. Then it was separated into our initial sample, plus

---

15 The percentages are calculated by: $\frac{\text{No. of not-at-homes}}{\text{No. of responses + No. of non-responses}} \times 100\%$

reserve for supplement. After the initial sample had been released and the result was known, we then released the required supplement to the field. Since we placed strict intructions and control on field procedure, our desired sample size was fulfilled, and problem of the reduction of sample size due to non-response was overcome.

The proportion of non-response was certainly not small, and its effects must be considered. The reported size of non-response can help to estimate the effects. In case information about difference between responses and non-responses is obtainable from the sample, relative bias can be measured.[16] Even when information about characteristics of non-responses is practically nil, we can still sub-sample the non-responses or accumulate substantial information from other sources. Or we can treat the problem by combining the bias of non-responses with the sampling variance into a total error, as suggested by Cochran.[17] The method may be over-simplified, but it may be most suitable for our studies. Substantial information on non-responses was available from Resettlement Estates and Low Cost Housing Estates residents, but to collect information from cottages or multi-storey buildings would be too time-consuming and would mean enormous manpower.

---

16 L. Kish, op. cit., pp. 532-562.

17 Cochran W. G., Sampling Techniques, John Wiley and Sons, Inc., New York, 1953, pp. 292-304.

3. Non-coverage

Non-coverage of the sample frame is also another source of errors. Luckily, good records and complete lists of residents in Public Housing were available from estate offices, and the number of structures in Cottage Areas could be obtained from the Resettlement Department. For Private Housing, we had prelisted all the flats with precision. This information enabled us to construct our sample where cases of non-coverage were rare and their effects were negligible.

VII. Resume

In Kwun Tong, the population is now reaching 460,000. But the district is small in area - around 12 sq. km., and residences are rather concentrated; travelling within the district is not so time-consuming, and the cost is low. Thus we did not need to apply the techniques of cluster sampling or multi-stage sampling for the purpose of reducing the travelling of fieldworkers; another reason is that these methods would make statistical analysis more complicated and yield larger bias. For Life Quality Survey and Health System Survey, large sized representative samples were required, as well as precision of sub-divisions of the population. Therefore stratified sampling method was most preferable and was actually employed since ancillary information was accessible.

We used housing type and location as criteria for stratification because they were important variables for both studies. But for other surveys in other districts, it may not be the same. In fact, when a sample design is requested, the project investigator should provide the person responsible for sample designing information about the project, especially the purposes and variables considered as most important. The project investigator should notice also the amount of ancillary data which are currently available and can help the designer to work out a best-fit sample frame.

In case supplement is required so as to match the desired sample size as has been done in Kwun Tong, new assignments to fieldworkers should be taken from the list prepared in advance, following the sequence and not exceeding the required number which can be determined by computation of reporting of the initial sample. This is very important when stratified sample is designed to be a self-weighted one.

Strict control methods cannot eliminate the possibility of cheating by fieldworkers; anyway, good control is necessary when processing our fieldwork. Supplement must be released carefully if they are required. Cases of refusal, not-at-home, or ineligible should be reported in detail as well as number of calls and interviewing time. These not only can reduce errors in the procedure of fieldwork but also can provide adequate information for estimating the effects of non-response and non-coverage in the sample.

Specific statistical problems have not been discussed in detail in the paper. Readers can refer to specialised literature on that field. Here we just intend to discuss problems in practising sample survey, particularly "sampling" in the Kwun Tong District, to explain how we explored alternatives to solve these problems, and hope to provide reference for future sample surveys in other communities which may or may not be similar to Kwun Tong.

Appendix I. Number of residential units and sample size, of different categories of housing

| | No. of residential units (n) | Sample fraction (f) | Sample size (s = nf) |
|---|---|---|---|
| 1. Private Housing | 10,816 | $f = \frac{S}{N} \times 100\%$ | 161 |
| 2. Resettlement Estates | 56,818 | | 845 |
| 3. Low Cost Housing Estates | 16,927 | | 252 |
| 4. Cottage Area | 7,405 | | 110 |
| Total | 91,966 | f = 1.487% | s = 1,368 |

N.B. The Sample fraction (f) in the above table is that for Life Quality Survey.

Appendix IIa. Sample Design for Resettlement Estates

| Estate | Number of residential units by type | | | | | Total | Population (31/12/70) |
|---|---|---|---|---|---|---|---|
| Jordan Valley | MI/½D | MI/C | MI/D | MI/E | MI/A & MI/A+ | | |
| | 80 | 904 | 1615 | 941 | 189 | 3771 | 20,339 |
| Kwun Tong | MI/½D | MI/C | MI/D | MI/E | | | |
| | 808 | 2700 | 4771 | 2721 | | 11000 | 57,196 |
| Yau Tong | MIII/Small | MIII/Standard | MIII/Large | MIII/Double | | | |
| | 761 | 1855 | 702 | 39 | | 3357 | 19,313 |
| Sau Mau Ping | MIII/Small | MIII/Standard | MIII/Large | MIII/Double | | | |
| | 1149 | 2033 | 698 | 44 | | 3924 | |
| | MV/Small | MV/Standard | MV/Medium | MV/Large | | | |
| | 3537 | 5099 | 3579 | 1306 | | 13521 | 86,015 |
| Lam Tin | MIV/Small | MIV/Standard | MIV/Large | | | | |
| | 2864 | 3968 | 3862 | | | 10694 | 55,187 |
| Ngau Tau Kok | MIV/Small | MIV/Standard | MIV/Large | | | | |
| | 869 | 990 | 949 | | | 2808 | |
| | MV/Small | MV/Standard | MV/Medium | MV/Large | MV/Special A & B | | |
| | 1459 | 3145 | 2200 | 870 | 69 | 7743 | 55,147 |
| | | | | | Grand Total | 56,818 | 293,197 |

N.B. (1) Each box in the Table (excluding the last 2 columns) of the above table represents on stratum in the sample.

(2) Sample size of each stratum is obtained by multiping the No. of units in the stratum with the over-all sample fraction (f).

(3) Only occupied residential units are counted.

Appendix IIb. Sample Design for Low Cost Housing Estates under the control of Housing Authority and Housing Society

| Estate | No. of persons per room | | | | | | | | Total Flats | Population (Capacity) |
|---|---|---|---|---|---|---|---|---|---|---|
| | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | | |
| Garden Estate (phase I & phase II) | | 1008 | 555 | 1486 | 310 | | 80 | 15 | 3454 | 22217 |
| Wo Lok | | 954 | | 994 | | | | | 1948 | 11728 |
| Ping Shek | | | 3780 | | 816 | | | | 4596 | 29208 |
| Kwun Tong | 409 | 559 | | 22 | 112 | 24 | 2 | | 1128 | 5717 |
| Ngau Tau Kok | 730 | 2723 | | 1693 | 297 | 358 | | | 5801 | 33984 |
| | | | | | | | | Grand Total | 16927 | 102854 |

N.B. (1) Each box in the Table (excluding the last 2 columns) of the above table represents one stratum in the sample.

(2) Sample size of each stratum is obtained by multipling the No. of units in the stratum with the over-all sample fraction (f).

(3) Only occupied residential units are counted.

(4) The No. of residents in the Low Cost Housing Estates is the population capacity of the estates; but we must note that some flats in these estates are overcrowded and some are under-populated.

(5) Garden Estate is under the control of Housing Society, Wo Lok & Ping Shek Estates belong to Housing Authority, Kwun Tong & Ngau Tau Kok Estates are Government Low Cost Housing (with a relatively lower monthly rent), but they are under the control of Housing Authority.

Appendix III. Sample Design for Cottage Areas

| Name | No. of domestic structures (n) | Sample size (s) | Population (in March, 1971.) |
|---|---|---|---|
| Lyemun Village (including Ling Nam Sun Tsuen) | 634 | 11 | 2743 |
| Cha Kwo Ling Village | 584 | 10 | 3614 |
| Ma Yau Tong Village | 233 | 4 | 605 |
| Yau Tong Squatter | 162 | 3 | 467 |
| Sam Ka Chuen Class II Area | 1883 | 31 | 6331 |
| Kowloon Bay Licensed/Resite Area | 628 | 10 | 1823 |
| Ngok Yue Shan Class II Area | 408 | 7 | 1029 |
| Hong Ning Road Class II Area | 866 | 14 | 2812 |
| Ngau Tau Kok Cottage Area | 1209 | 20 | 8762 |
| | 6607 | 110 | 28186 |

N.B. There are more than 20 villages in Kwun Tong District, but sample units are drawn from 9 large villages, small villages having less than 2% of the total number of structures in the district are exempted. Totally, there are 7405 structures within these villages; with an over-all sample fraction of (f). The sample size will be of 7405 x f. These sample units will be taken from the above 9 villages by random and in proportion to the number of structures in each village. In the above table, the sample size (s) is that of the Life Quality Survey.

The Population living in Cottage Area attained 31486 in March, 1971.

Appendix IVa. Result of calls in Life Quality Survey

| Housing Type | Desired Sample Size | Cases Completed | Non-responses | | Ineligibles |
|---|---|---|---|---|---|
| | | | Refusals | Not-at-home | |
| Private Housing | 161 | 144 | 23 | 86 | 15 |
| Low Cost Housing | 252 | 183 | 31 | 77 | 5 |
| Resettlement Estates | 845 | 648 | 121 | 386 | 43 |
| Cottages | 110 | 99 | 2 | 2 | -- |
| Total | 1386 | 1074 | 177 | 551 | 63 |

Appendix IVb. Result of calls in Health System Survey

| Housing Type | Desired Sample Size | Cases Completed | Non-responses | | Ineligibles |
|---|---|---|---|---|---|
| | | | Refusals | Not-at-home | |
| Private Housing | 84 | 85 | 19 | 11 | 11 |
| Low Cost Housing | 132 | 135 | 20 | 6 | 4 |
| Resettlement Estates | 441 | 442 | 105 | 93 | 15 |
| Cottages | 40 | 40 | 1 | 4 | 3 |
| Total | 697 | 702 | 145 | 114 | 33 |

References:

Cochran, W. G., Sampling Techniques, John Wiley and Sons, Inc., New York, 1953.

Interviewer's Manual, Survey Research Centre, Institute for Social Research, the University of Michigan, Ann Arbor, Michigan, 1969, Section C.

Kish, L., "A Two-stage Sample of a City", Am. Social. Review, 17 (Dec., 1952), pp.761-769.

Kish, L., Survey Sampling, John Wiley and Sons, Inc., New York, 1965.

Kish, L., "Confidence Intervals for Clustered Samples", Am. Sociol. Review, 22 (April, 1957), pp.154-165.

Lazerwitz, B., "Sampling Theory & Procedures," in H. M. Blalock & A. B. Blalock (ed.), Methodology in Social Research, (McGraw-Hill, 1968), pp.278-326.

Sharp, H. & A. Feldt, "Some Factors in a Probability Sample Survey of a Metropolitan Community", Am. Sociol. Review, 24 (Oct., 1959), pp.650-661.

Stephan, F. F. & P. J. McCarthy, Sampling Opinions - An Analysis of Survey Procedure, John Wiley and Sons, Inc., New York, 1959.

United Nations, A Short Manual on Sampling, Vol.1 & 2, Department of Economic and Social Affairs, United Nations, New York, 1960, 1970.

Yates, F., Sampling Methods for Census and Surveys, Charles Greffin and Co., Ltd., London, 1953.