

Sample Design and Evaluation for an Occupational Mobility Study

C. A. Ó MUIRCHEARTAIGH
R. D. WIGGINS

London School of Economics and Political Science

I INTRODUCTION

THE aim of this paper is to describe the design and evaluation of the sample for an occupational mobility study in Northern Ireland and the Irish Republic undertaken by Jackson *et al.* (1973–74)¹. The objective of the study was to demonstrate the characteristics of occupational mobility in both parts of Ireland and also to develop an analysis of social structure, change and development. An important constraint was that of cost. In order to obtain a large enough sample in each part of Ireland to justify detailed breakdown by age cohort, while retaining sufficient numbers in each cell, some sacrifices had to be made. The most important of these was the decision not to include women in the study and to limit the target population to males aged 18–64.

The objective of the sample design therefore was to achieve an equal probability sample of males aged 18–64 in Northern Ireland and a similar sample in the Irish Republic. The sample size was fixed at 2,500 achieved interviews for both the North and the South to permit the desired subclass analysis. The two primary considerations were to minimise cost and maximise precision. The two are normally interdependent and an increase in precision can only be obtained by increasing the costs. Cost in this case should be viewed in its broadest sense and should include cost attributable to the complication of the analysis due to complexity in the sample design. The overall approach was to place the sample design and fieldwork execution in the context of the survey as a whole and to treat the overall survey design as a single unit. Thus the sample design should take into account the analytical methods to be used on the data.

1. The Irish mobility study was designed by Professors Jackson and Iutaka in association with Professor Hutchinson to study "Determinants of Occupational Mobility in Northern Ireland and the Irish Republic". The research, supported by the Social Science Research Council under grant number HR1430/1, involved two major field surveys; the first limited to Northern Ireland is reported in the paper by Boyle (1977), this issue; the second, involving Northern Ireland and the Irish Republic is referred to in the present paper. Further information concerning the project may be obtained from Professor J. A. Jackson, Trinity College, Dublin.

The analysis of social mobility data has been developed assuming simple random sampling. Therefore a basic aim of the design was to permit departures from simple random sampling only when either the gains in precision were sufficient to justify them or when cost considerations required them. Thus we considered a simple random sample from the Electoral Register as a starting point for the discussion on the effect of the sample design on the precision of our estimators.

The Electoral Registers were used as sampling frames for the two samples since they are the only lists readily available and convenient to use for such a sample design. The Electoral Register lists persons eligible to vote at elections, namely, persons 18 years and over, together with the address at which they lived at the time the register, which is revised annually, was last drawn up. The two major inadequacies of the registers are incomplete coverage and "out-datedness". If we are to achieve a probability sample of males aged 18-64 we must consider sampling designs which remedy these inadequacies.

As there is no exact list of males aged 18-64, in using the register as a frame one can either

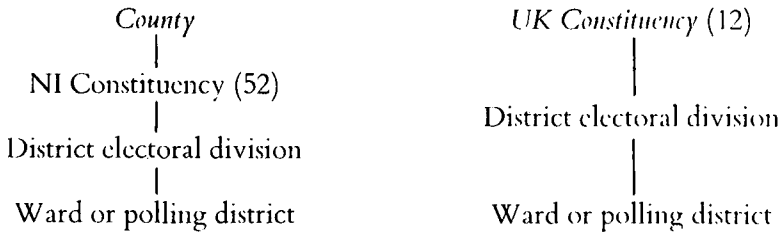
- (i) fix the number of sampling points selected at the final stage, *or*,
- (ii) fix the number of male names sampled at the final stage, *or*,
- (iii) fix the number of contacts with males aged 18-64.

(i) has the property of operational simplicity whereas (ii) and (iii) increase the degree of control over final sample size. It was decided to adopt strategy (i) in order to reduce the amount of interviewer decision-making in the field.

A two-phase sample design was therefore adopted. The first phase would consist of the selection of a number of sampling points which could be expected to generate fewer than 2,500 interviews. On the basis of the returns from the fieldwork a more reliable estimate could be made of the number of interviews generated by a given number of sampling points. Consequently, a second phase sample of appropriate size could be selected from the same sampling units in order to generate an equal probability sample with the desired number of respondents. In order to decide on the number of sampling points in the first phase three factors were taken into account—the proportion of male names on the register, the proportion in the age group 18-64, and the likely response rate.

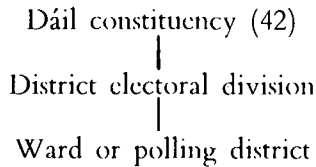
In the North the first phase sample generated less than 2,500 completed interviews but funds did not permit us to complete a second phase. In the South it looked as though the first phase was going to generate more than 2,500 completed interviews so in fact a second phase procedure was adopted to reject a random half sample of all outstanding contacts in December 1973. The elements included in this second phase then received a weight of "2" in addition to any reweighting implied by the rest of the design.

In Northern Ireland the qualification date for registration was 15 September, 1972 and the register was published the following February. The arrangement of the registers can be viewed in two ways:



In the Republic, electors are registered to vote at Dáil and/or local elections. There is no exact qualification date for registration but the register comes into operation for one year as from 15 April. Registers are compiled over a 5-month period beforehand and can be amended up to 1 April.

The registers are arranged as follows:



Unlike the North the majority of DED's do not correspond one-to-one to a polling district. Some polling districts belong to more than one DED and most DED's contain more than one polling district. County and county electoral area information is also given on the register.

Section II of this paper discusses briefly the principles on which the sample was designed and describes the overall sample design. In Section III the strategy adopted for the final stage of selection is discussed. Section IV presents the results of the fieldwork in terms of response rates and the design is evaluated in Section V.

II OVERALL SAMPLE DESIGN

1. THE DESIGN EFFECT

The basic element sample design is simple random sampling, where each combination of elements in the population has the same chance of selection. Departures from simple random sampling affect the inference which can be made from the sample statistics to the population parameters. In practice, departures from simple random sampling are either found necessary (on cost grounds) or desirable (on theoretical grounds). In order to discuss the effect of different departures from simple random sampling in the design in terms of the precision of estimators it is useful to introduce a single key concept, the design effect, D_{eff} :

Deff = Ratio of actual variance to variance of simple random sample of the same size.

Stratification implies the division of the population into subgroups or strata and the selection of independent samples within the strata. The principal advantages of stratification in this case were:

- (i) An increase in precision, i.e., $Deff < 1$. However, the extent of this increase will depend on the variables by which stratification is possible and their relationship with the variables under study. Since the only stratification factor available was a crude geographical one we only expect gains in precision of the order of 5–10 per cent.
- (ii) Freedom to use different sample designs in different parts of the population.

Whereas stratification leads to an increase in precision and is therefore generally desirable, the second type of departure from simple random sampling—*clustering*—typically produces a decrease in precision (an increase in the variance), and is generally necessary in order to keep costs within reasonable limits. The design effect for cluster samples is

$$Deff = 1 + \rho(b-1)$$

where ρ is the intracluster correlation coefficient, which is a measure of the internal homogeneity of the clusters, and b is the number of elements selected from each cluster. In practice ρ will be positive and thus $Deff$ will be greater than 1 and there will be a relative loss of precision.

2. URBAN AREAS

Cost considerations ruled out the use of simple random sampling or systematic sampling within all strata. However, in the largest urban areas in each of the populations—i.e., Belfast in the North and Dublin in the South—the size of the sample to be selected indicated that the cost of an element sample would be only marginally more expensive than a multi-stage sample. Consequently, it was decided to select a systematic sample from the 4 Belfast UK constituencies and the 10 Dublin Dáil constituencies. The “Belfast” sample consisted of 5 independent systematically selected samples and the “Dublin” sample consisted of 4 independent systematically selected samples. The ordering of polling districts and constituencies provided a further implicit geographical stratification for the systematic sample.

3. OTHER AREAS

For the areas outside the Belfast and Dublin constituencies three basic alternatives were considered, all of which give a probability sample, and are

listed in descending order of precision and in descending order of costs of preparing the sampling frame before selection.

For the North and the Republic polling districts were first arranged by UK and Dáil constituency respectively.

(i) Zoning with geographical stratification.

Polling districts would be arranged north to south moving east to west, west to east across each constituency. All primary sampling units (psu's) are of equal size electorate, i.e., zones, thus zone boundaries may cut across polling districts.

(ii) Probability proportional to electorate size with minimum size psu and geographical stratification.

Polling districts are grouped similarly to (i) and psu's are formed so that they exceed a minimum electorate size. Psu's are selected with probability proportional to electorate size; in (i) this reduced to equal probability as zones are equal size psu's.

(iii) Probability proportional to electorate size with minimum size psu.

Polling districts are not grouped as in (i) and the geographical stratification is not used.

The question of minimum size zone/psu arose since when sampling psu's with probability proportional to size we must have a constant number of sampling points per sampled psu if we are to achieve an equal probability sample of elements.

$$\frac{\text{psu size}}{\text{population size}} \times \frac{\text{no. of contacts}}{\text{psu size}} = \text{probability of selection (two-stage)}$$

Here size of electorate is used as a measure of size.

Two considerations were involved. First, for each zone/psu enough sampling points should be selected to generate sufficient contacts to provide a reasonable interviewer workload in that zone/psu. Second, the situation should be avoided where interviewers visited a high proportion of the dwellings in an area. Hence some minimum size psu was necessary. The minimum size of psu was estimated so as to allow for around 30 eligible contacts per sampled psu on the basis of 3 electors per household and the desire not to sample more than 1 in 10 households.

In the North as polling districts generally corresponded to a DED zoning was feasible, so design (i) was adopted for the constituencies outside the Belfast constituencies. UK constituencies were first grouped into the way described in (i) and then polling districts were similarly ordered within these. This meant, of course, that some zones cut across UK constituency boundaries.

In the Republic DED mappings lead only to an approximate guide to polling district location so (i) was not feasible and design (ii) was adopted for the areas outside the Dublin constituencies. All of the Dáil constituencies were represented in the design and psu's were allocated in proportion to Dáil electorate, the only published electorate figures available at the time of preparing the sample frame. Psu's were selected with probability proportional to electorate within each Dáil constituency.

4. WEIGHTING FOR THE ALL-IRELAND ESTIMATOR

In obtaining estimates for parameters of the All-Ireland population account has to be taken of the fact that there are different sampling fractions in the two strata. However, the total number of eligible males in the population is not known for either Northern Ireland or the Republic of Ireland. Hence, the overall population figures must be used to provide the weights for the estimator.

$$\begin{aligned} (1) \text{ Total population (TE)} &\times \text{proportion of men in electorate (PM}_p\text{)} \\ &\times \text{proportion of men in the age group 18-64 (AM}_p\text{)} \\ &= \text{Eligibles in survey population (E}_p\text{)} \end{aligned}$$

$$\begin{aligned} (2) \text{ Sampling points (SP)} &\times \text{proportion men in sample (PM}_s\text{)} \\ &\times \text{proportion of sample men in age group 18-64 (AM}_s\text{)} \\ &= \text{Eligibles in sample (E}_s\text{)} \end{aligned}$$

We use (TE/SP) as an estimate of (E_p/E_s) for each stratum.

III FINAL STAGE OF SELECTION

Using the electoral register as a sampling frame for a sample of males aged 18-64 presented us with two main problems:

- (i) what to do about non-registered males who are aged 18-64 i.e., males who for one reason or another did not register.
- (ii) what to do about males aged 18-64 who moved since the registration qualification date.

Some of the addresses on the electoral register contain more than one dwelling unit or household. It was therefore necessary to develop a definition of household so that we could refer to non-registered and moving eligibles belonging to the

sampled elector's household rather than to the whole address. The basic household definition adopted was as follows:

"A household is a group of people who all live regularly at the address given, and who are all catered for, at least one meal a day, by the same person. A household may contain no more than 6 males aged between 18 and 64."

Literally any dwelling unit falling outside this definition is taken to be an "institution". As there were no reliable estimates as to the likely proportion of eligibles falling into this category it was decided to wait until the extent of the problem was known before contacting the *individuals* in any institution. This strategy applies to all the sampling procedures considered below. The procedure proposed is given in Appendix I. Lack of funds prevented us from carrying out this procedure.

1. POSSIBLE STRATEGIES

The five sampling procedures considered for the final stage of selection were:

- A. Male names on the register are selected with equal probability. The selected male name serves as a "tag" to identify the household at an address. If the male selected has died or moved away we are interested in the household remaining or replacing the selected male's household. If there is one male aged 18-64 living in the household then he is the interviewee, if there is more than one male aged 18-64 living in the household then the interviewer adopts a Kish selection procedure to select one of them. In this way households are selected with probability proportional to the number of male electors and the respondent selected from within the household by the Kish selection procedure. (See Kish, (1965), Section 11.3).

- B. "Nuffield". This is the procedure adopted by Nuffield College in their 1970 study of occupational mobility (excluding their procedure for dealing with institutionalised males):

Male names are selected with equal probability from the register.

 - (i) If the selected male is aged 18-64 and still living at the address given then he is interviewed.
 - (ii) If the selected male is aged 18-64 and has moved, then he is followed up and interviewed at his new address.
 - (iii) If there are males aged 18-64 living in the household and not registered for the address then all such males are listed in descending order of age and the Kish selection procedure used.

C. Male names are selected with equal probability from the register. If the selected male is aged 18–64 and still living at the address at the time of contact then he is interviewed as long as there are:

- (i) no males aged 18–64 living at the selected male's household since before registration date and not registered;
- (ii) no males aged 18–64 who moved to the household since registration date.

If the answer is “yes” to either (i) or (ii) then the interviewer lists all males aged 18–64 in descending order of age for the household and uses a Kish selection procedure. In this way any males who have moved have a probability of selection at the place they move to. As in A if the selected male elector has died or moved away we are interested in the household remaining or replacing the selected male's household.

D. In procedures A–C female names are treated as blanks in the sampling frame. This procedure merely reconsiders A–C by taking both male and female names as sampling points.

E. Here we ignore the non-registration and movers problem by excluding such individuals and simply selecting male names with equal probability and interviewing them if they are still living at the address given and are aged 18–64.

2. STRATEGY ADOPTED

The target population can be viewed as consisting of two components:

- (i) males aged 18–64 whose names appear on the electoral register for their current address;
- (ii) males aged 18–64 whose names do not appear on the electoral register for their current address; these may be.
 - (a) living in a household with a registered male
 - (b) living in a household with a registered elector
 - (c) living in a household without a registered elector

This provides a useful starting point to examine the coverage problem and the solutions which the five alternatives provide to this problem were judged in the light of coverage, cost, contamination and simplicity.

Group (i) above are picked up by any of the five strategies.

Under B if the selected male is eligible and has not moved he is picked up perfectly, otherwise a follow-up is required; but there is the possibility of more than one interview in a household. Under A and C there will never be more than one interview per household, which is an obvious advantage if one wishes to avoid duplication of information or contamination. For A, the probability of selection is proportional to (no. registered)/(no. eligible).

The problem of weighting disappears only if the correlation between registered and eligible males = 1. In both A and C eligible movers are equivalent to "non-registered" males. B has the advantage of minimising the weighting involved, but the follow-up of movers must be balanced against the cost of reweighting in the analysis. In the North the register was 12 months out of date by the time interviewers were in the field and the follow-up of movers was likely to be problematic; in the South the register was 5 months out of date and there were no reliable estimates of the extent of movement. Thus, given the desire to have only one interview per household, the choice of an alternative fell between A and C with D as a possible amendment.

Strategy C compared to A reduces the work involved since it has the attractive feature (also present in B) that the interviewer uses a Kish selection procedure only if there is a "non-registered" eligible male at the household for which he/she has a sampled male contact. Strategy D improves the coverage in all cases since eligible non-registered males living in households where the registered elector(s) are female will then have a non-zero probability of selection (i.e., group (ii)(b) above). However, the available evidence suggested that the cost per eligible contact would be increased substantially by this strategy since many more households with no eligible males would be contacted. Consequently, strategy C was adopted.

IV COVERAGE AND RESPONSE

1. COVERAGE

Failure to include some parts of the defined target population in the sampling frame constitutes noncoverage. Such elements have zero probability of inclusion in the sample. The sample design described above excludes two subclasses:

- (i) Males aged 18–64 not on the electoral register who live in either
 - (a) households where females only are registered or
 - (b) households with no registered electors.

We expect these two classes to be very small but have no way of checking within the survey itself.

- (ii) Eligible males living in institutions.

The strategy for sampling from institutions is described in Appendix I although, on cost grounds, this was not put into effect. In effect the survey population was defined as males aged 18–64 living in households.

Table 1: *Outcome of fieldwork*

<i>Category</i>	<i>North</i>		<i>South</i>		<i>Overall</i>	
	<i>Frequency</i>	<i>Relative freq.</i>	<i>Frequency</i>	<i>Relative freq.</i>	<i>Frequency</i>	<i>Relative freq.</i>
1. Institution	44	1.0	106	2.7	150	1.8
2. House demolished	24	0.5	13	0.3	37	0.4
3. House vacant	122	2.8	54	1.4	176	2.1
4. Refusal of any information	105	2.4	31	0.8	136	1.6
5. Person female	23	0.5	11	0.3	34	0.4
6. Person dead (No to IIA and IIB)	78	1.8	72	1.8	150	1.8
7. Person moved (No to IIA and IIB)	302	6.8	299	7.6	601	7.2
8. Person wrong age (No to IIA and IIB)	589	13.3	520	13.2	1,109	13.3
9. Interview completed	2,404	54.2	2,269	57.7	4,673	55.9
10. Refusal to complete—schedule used	9	0.2	20	0.5	29	0.3
11. Incomplete interview (used)	3	0.1	2	0.1	5	0.1
12. Not at home	266	6.0	204	5.2	470	5.6
13. Interview refused	350	7.9	202	5.1	552	6.6
14. Refusal to continue interview	4	0.1	6	0.2	10	0.1
15. Incomplete—other	94	2.1	109	2.8	203	2.4
16. Wasted interview	6	0.1	12	0.3	18	0.2
17. Not contacted	10	0.2	20	0.5	29	0.3
<i>Total</i>	4,433	100.0	3,930	100.0	8,363	100.0

2. RESPONSE

In surveys, whether samples or complete enumerations, where the respondents are under no obligation to co-operate, it will normally be the case that data will not be collected for all the designated units. Data will be missing for some part of the population and the problem arises as to what conclusions are legitimate from such surveys or alternatively what means are available to gain insight into the part of the population not covered. In each survey, particular sources of non-response will be dominant. In Table 1 the outcome of the fieldwork is presented both for the North and the South. An examination of the categories shows that they fell into two conceptually different classes. The first class consists of 1 through 8 and includes all those contacts which did not generate eligible respondents. Category 1 is a special case in which those living in institutions were defined as not belonging to the survey population. Those who moved from a household where the selected individual had died were covered by another aspect of the sample design. The response rate should be calculated excluding these categories and should deal with responses and non-responses among the eligibles only, i.e., categories 9 through 17.

The effect of non-response on the survey estimates is caused by differences between the respondents and the non-respondents. For that reason the categories in the second class do not all have the same significance. Refusals are more likely to differ from the respondents than are the not-at-homes for some variables, and the effect of non-response will vary according to the parameter of the population being estimated. Categories 4 and 17 represent a special problem since in the absence of any information we do not know whether an eligible respondent could have been generated by that contact point. Categories 9 through 11 are the respondents, although in categories 10 and 11 information on some questions is missing. Table 2 gives the overall response rates broken down into two components.

Table 2: *Response rates*

	<i>Number of contacts set</i>	<i>Eligible elements</i>	<i>% Eligible</i>	<i>Interviews</i>	<i>Effective response rate</i>
North	4,433	3,227	73	2,416	75%
South	3,930	2,867	73	2,291	80%
Overall	8,363	6,094	73	4,707	77%

The overall response rates of 75 per cent in the North and 80 per cent in the South were very satisfactory for a survey with a questionnaire of this length where the demands on the respondent were considerable.

V CONCLUSION

The success of the sample design and execution reported here can be evaluated using four criteria. First, the degree to which the design achieved adequate

coverage of the population, which includes a consideration of response rates, second, the efficiency of the design in terms of the cost of obtaining an interview, third, the simplicity and practicality of the design² and finally, the precision of the estimators obtained from the sample.

It is impossible to assess from the survey itself the degree to which some elements in the target population may not be covered by the sample design. However, the experience in the use of the Kish selection technique to extend the coverage beyond those included in the register of electors seems to indicate that the strategy was successful. The only individuals not included in the coverage were those living in households which had no registered male elector. Some of those could have been included by using both male and female names as sampling points but the pilot work suggested that the additional cost involved in such a strategy would not have been justified by the returns. The effect of the non-response, which is the other area of non-coverage, can, however, be assessed within the survey itself. The overall response rates were satisfactory and the information collected on the contact sheets can be used to estimate the likely effect of the non-response on different types of estimators. Three procedures are possible. The first is to calculate the estimators simply on the basis of the unweighted responses. This is equivalent to assuming that the respondents are representative of the selected sample. The second is to weight the respondents according to the differential response rates in different parts of the sample. The third and most promising is to use a *successive stages model*. This assumes the existence of a functional relationship between the difficulty of obtaining a response, measured by the call on which the response was obtained, and the survey variable. Extrapolation from the respondents to the non-respondents based on a least-squares estimation procedure is then possible. The data as recorded permit the use and comparison of the three procedures.

The number of sampling points required to identify an eligible member of the population was approximately 1.3. Since the eligible population comprised less than 40 per cent of the sampling frame this was very satisfactory. The strategy of dealing with movers at their current rather than registered address also reduced the costs considerably.

The principal advantages of the design were the close approximation to equal probabilities of selection achieved together with the fact that no more than one interview was carried out in any household. Duplication would have been liable to produce two undesirable effects—first, the possible contamination of the second response within the household, and second, the cost of obtaining some of the same information in two separate interviews.

It is our intention in further publications to present sampling errors computed, using the balanced repeated replication method for complex statistics (Kish, Frankel and Van Eck (1972).) The sample design facilitates the use of simple

2. As evidence of the practicality of the stratified design for the selection of psu's described in Section II, we may note that it has now been adopted as the standard sample design for national surveys by The Economic and Social Research Institute Survey Unit.

methods of variance estimation and should provide valuable estimates of the likely design effects for similar surveys in both parts of Ireland. The two phase design was prevented from being used to the full by a shortage of funds in the latter stages of the fieldwork, but the information in Table 1, together with information on the number of electors per household and the extent of non-registration should provide better estimators of the population parameters for use in designing future surveys.

REFERENCES

- BOYLE, J., 1977. "Educational Attainment, Occupational Achievement and Religion in Northern Ireland", *The Economic and Social Review* Vol 8, No. 2.
- COCHRAN, W. G., 1963. *Sampling Techniques*, New York: John Wiley & Sons.
- KISH, L., 1965. *Survey Sampling*, New York: John Wiley & Sons.
- KISH, L., M. FRANKEL and N. VAN ECK, 1972. *Sampling Error Program Package*, Ann Arbor: Institute for Social Research, University of Michigan.

APPENDIX: PROCEDURE FOR SAMPLING FROM INSTITUTIONS

1 OFFICE PREPARATION

- (i) All contacts that fall into the "institution" category, final result code 1, should be filed separately from all other contact sheets.

Covering letters go out to "institutions" informing them of the purpose of the survey etc., stressing that we are only interested in males aged 18-64 and asking them when it would be convenient for someone to call so as to arrange to select a sample of *not more than X* males aged 18-64. The number X will be the number of contact points for that "institution" and will therefore be known in advance.

2 INTERVIEWER MAKES FIRST VISIT

Requirements are:

- (i) the address of the "institution"
- (ii) the maximum number X of interviews she/he may have to conduct at that "institution"
- (iii) a list of all registered males for that "institution". Either a photocopy from the register or the relevant copy of the register.

- (iv) On making contact with the "institution" the interviewer lists *all males* at the "institution" who have:
- (a) been living there since before 15 September 1972 and who are registered at the "institution" address
 - (b) been living there since before 15 September 1972 and who are *not* registered for this "institution"
 - (c) moved into the "institution" before 1 November 1973* and who are not registered for this "institution".
- (v) Interviewers then require the age of *every* male they have listed, paying particular attention to males with ages at the extremes of the eligibility range i.e., "18" and "64".

If interviewers construct a list as follows sampling in the office should be made easier:

<i>Males registered</i>	<i>comment</i>	<i>List all males</i>	<i>comment</i>	<i>Age</i>	<i>No.</i>
List all male names appearing on the Electoral Register for the institution	e.g., moved out Not a member of institution e.g., has caretaker's flat		e.g., moved in		Interviewer leave this column blank

By using "comment" columns interviewers can cross check the list of all males registered for the "institution" against her list and this should prevent the risk of omission.

If an interviewer discovers that some of the males listed as registered for the "institution" in the left hand column happen to live in a household e.g., caretaker's flat which is not strictly part of the institution then this *must* be noted.

3. SAMPLING IN THE OFFICE

- (i) the individuals sampled will not necessarily be the same as the named contact coded "institution" on the contact sheet.

*November 1 1973 was taken as the modal point when approximately half of the fieldwork would have been completed. Anyone aged 18-64 who moved in after 1 November 1973 would already have had a chance of selection.

- (ii) Check that the original named contact points are/were residents of the "institution". If so then $X = \text{No. of interviews to be set for the "institution"}$. Number the males *aged 18-64* listed in the right hand column in descending order of age. Make X selections at random using random number tables.
- (iii) If it happens that one of the original named contact points turns out to be the caretaker, say, living in a separate household (see definitions) then treat this contact as an ordinary household contact as in the main fieldwork.

Interviews to be set in the "institution" now become $(X-1)$, i.e., make $X-1$ selections at random from those males you have listed for the "institution" aged 18-64. Remember interviewers only list in the right hand column males at the "institution". For example, if in case (ii) there are 2 contacts at a particular "institution" the numbering of eligibles produces 10 and there are no "households" in the "institution"; then select 2 random numbers between 1 and 10. This gives the selections for interview, they retain the original contact serial numbers. In case (iii), e.g., if 1 of 2 original contacts is the caretaker, make 1 selection from 10. Treat the caretaker contact in the normal way.