No. 3

Manual on Sample Design

BASIC DOCUMENTATION

MARCH *1975*

INTERNATIONAL STATISTICAL INSTITUTE
Permanent Office . Director: E. Lunenberg
428 Prinses Beatrixlaan
The Hague – Voorburg
Netherlands

WORLD FERTILITY SURVEY
Project Director:
Sir Maurice Kendall, Sc. D., F.B.A.
35–37 Grosvenor Gardens
London SW1W OBS, U.K.

The World Fertility Survey is an international research programme whose purpose is to assess the current state of human fertility throughout the world. This is being done principally through promoting and supporting nationally representative, internationally comparable, and scientifically designed and conducted sample surveys of fertility behaviour in as many countries as possible.

The WFS is being undertaken, with the collaboration of the United Nations, by the International Statistical Institute in cooperation with the International Union for the Scientific Study of Population. Financial support is provided principally by the United Nations Fund for Population Activities and the United States Agency for International Development.

This publication is part of the WFS Publications Programme which includes the WFS Basic Documentation, Occasional Papers and auxiliary publications. For further information on the WFS, write to the Information Office, International Statistical Institute, 428 Prinses Beatrixlaan, The Hague – Voorburg, Netherlands.

Manual on Sample Design

Prepared by:
WFS CENTRAL STAFF
International Statistical Institute
35-37 Grosvenor Gardens
London SWIW OBS, U.K.

Acknowledgments

This manual was prepared by the WFS central staff.
Thanks are due to Professor Leslie Kish and
Dr Ivan Fellegi, who reviewed the manuscript and made
many helpful suggestions for its improvement.

Contents

1. Introduction

This manual is directed specifically towards the problem of sample design for the World Fertility Survey; it is not intended as a general manual on sampling. We assume that the reader has some knowledge of elementary statistical concepts (such as "expectation", "variance", etc.) but possesses only a passing acquaintance with sampling theory and practice; technicalities are avoided as far as possible.

The primary objective of the manual is to assist survey directors and managers to understand the main issues involved and hence to collaborate fruitfully with sampling experts in the design and selection of the sample. The manual gives great emphasis to practical problems and may be found useful by sampling experts who have limited experience in dealing with the practical issues which are involved in survey design in different types of countries. It should be emphasized, however, that the manual is not a substitute for a sampling expert. It does not attempt to give recipes for suitable WFS sample designs. The collaboration of a sampling expert at the planning stage remains essential in every country.

After some preliminary remarks on sampling theory and practice (Section 2), we first define the problem to be tackled, beginning with the population to be studied and the instruments for studying it (Section 3). We continue with the physical and psychological constraints on sample design (Section 4), and the problem of sampling frames (Section 5). At this stage we introduce a rough description of the type of sample proposed (Section 6) with the aim of clarifying the more detailed discussions on specific sampling problems in the subsequent sections. These cover the role of the household schedule (Section 7); stratification (Section 8); and varying probability sampling and multi-stage sampling (Section 9). Section 10 deals with non-response and other defects of sample implementation and a final section examines the sampling requirements for supplementary field operations – post-enumeration check and husbands survey (Section 11). The household schedule appears as an appendix.

While the World Fertility Survey involves, by its nature, an attempt to achieve a degree of standardization in the collection and reporting by different countries of data relating to fertility, there is no particular reason why the sample designs used by different countries should be standardized. Comparability depends on standardization of the concepts, the questions and the ultimate reporting units, not of sample designs. Thus the WFS organization will not attempt to press for any one uniform sample design. The hope is, rather, that each country will adopt whatever design is of optimal efficiency in its own special circumstances. Nevertheless, experience shows that many countries share the same problems and habitually arrive at similar solutions in the field of sampling. Thus it may be helpful to make available to participating countries a manual in which the more commonly met situations are described and a range of possible solutions is suggested, together with specific recommendations where appropriate. No country need take the manual as prescriptive, but it is hoped that most will find it a useful starting point in working out a sample design appropriate to their specific circumstances.

One point should be stressed at the outset: The World Fertility Survey will insist on the highest technical standards; in the field of sampling this implies, at least for the main survey operation, the rejection of such methods as quota sampling, purposive sampling, judgement sampling, etc., in which the probability of selection of each unit is uncontrolled. Rigorous probability sampling is the only approach which provides a scientific basis for generalization to the survey population and a procedure for estimating sampling error. This issue is treated in most sampling textbooks and will not be further discussed here; the need for strict probability sampling is taken for granted in the rest of the manual.

2. Sampling theory and sampling practice

2.1 Sampling: science or art?

Of all the aspects of sample surveys, sampling has the most highly developed theoretical basis. An extensive literature exists on sampling ranging from abstract mathematical investigations to handbooks for the practical sampler.

Newcomers to surveys are sometimes surprised to observe how little of this theory seems to find an application in actual sampling practice. Often the best sampling practitioners arrive at their sample designs by hunch and judgement (though later, in the sample selection itself, they naturally insist on strict scientific methods). In order to put the content of this manual into perspective, it may be useful to begin with a glance at this situation and to enquire whether a more rigorous approach to sample design than is usual would be desirable.

For the survey sampler, the basic problem may be stated simply: to find that sample design which will minimize the error for a given overall cost, subject to the various physical, psychological and organizational constraints which may be inherent in the specific situation. Sampling theory provides the means for, first, refining the statement of this problem and, second, solving it, assuming that the parameters involved can be specified. A rough sketch of the way this can be done follows – but note that this is a theoretical account which bears little relation to the way most samples are planned.

Refining the statement of the problem begins with an attempt to specify more exactly the "error" which has to be minimized. Survey results consist of numerous variables; we have to decide first which one of these, or what combination of these, is to have its error minimized. We then have to relate this error to the sample design as a mathematical function; normally this will involve treating separately the sampling and non-sampling components of error. Then costs have to be similarly expressed as a function of sample design. We can then proceed with computation of the optimal parameters of a given design and these optima can be compared for different possible designs to yield the absolute optimum. Finally, we check this against the supplementary "constraints" and, if necessary, make adjustments to the computed optimum to satisfy these.

This procedure evidently requires (1) a clear specification of a single overriding survey objective[1] and (2) some rather detailed knowledge of the parameters affecting the organization of the survey: variances between and within strata and clusters, non-sampling errors, cost functions. In practice it is quite rare to find a situation in which all this is feasible. *Specification of a single objective* (which may be a single variable or a combination, perhaps weighted, of several measured variables into a composite) is feasible in some demographic surveys, e.g., the crude birth rate, or the natural growth rate, at the national level; but in very few other cases. Even if we can specify one variable it may be difficult to specify the relative weight to be attached to the objectives of estimating the national total on the one hand, and the regional

[1] *Sub-objectives can, however, be brought in as supplementary constraints (e.g. "error on variable z must not exceed 5%").*

or other breakdowns on the other. *Knowledge of the error and cost parameters* is likely to be available only when a similar survey has been done before in the same or a similar country, and even then we obtain the parameters only for the particular design which has been used before (e.g., we obtain the variances within strata for the strata which were used, but these may not be optimal).

Not only are the data for optimizing the sample design difficult to obtain, but also the practical importance of such optimization is limited by two factors. Firstly, it has been found that the efficiency of sample design is (fortunately) rather insensitive to the exact values of the parameters near the optimum: that is to say, deviations from the optimal values can be quite large without greatly affecting the efficiency. This makes for more flexibility and encourages compromise solutions in designing the sample. Secondly, there are the "constraints" which we have already mentioned: options are limited by such factors as the availability of sampling frames, the need to keep a team of interviewers working full-time yet within reach of their supervisor, the total size of the field force which can be recruited and/or effectively controlled, and so on. It sometimes happens that constraints of this kind almost completely determine the sample design, leaving no significant room for manoeuvre with a view to design optimization.

Thus we find in practice that the great majority of samples are designed with little or no reference to the elaborate methodology of optimization which has been worked out by the sampling theoreticians. While some rough calculations along the lines described are often made as a guide, the final decisions are typically based on the "feel" of the situation, the past behaviour of the country or institution concerned, and the various organizational constraints which limit the options. In practice, then, survey sampling is partly a science, partly an art.

As long as we are concerned with *ad hoc*, or one-time, surveys no one is to blame for this situation. But when a survey is repeated, the opportunity exists for applying sampling theory with a view to improving sample design on the basis of the experience gained in the first survey. Here most survey organizers are indeed to blame – not those who organize the second survey but those who are responsible for the first one; very few of them undertake (and still fewer publish) the analytical studies of variances, costs and organizational problems which are necessary for improving the efficiency of future surveys. Thus, though experience is collected, it is not passed on. This is regrettable, for efficient sampling, like civilization, depends on the recording of experience for the benefit of posterity.

We therefore urge national organizers of surveys for the WFS to give attention not only to their current problems, but to the problems of their successors. A further WFS publication, supplementary to this manual, is under preparation in which some suggestions will be made as to the nature of the information which should be recorded to this end.

2.2 Sampling for the WFS

The preceding section indicated how samples ought to be designed in ideal conditions and how they generally are designed in practice. How does this apply to the WFS?

We look first at the *objectives*. Can any one variable be singled out as representing the crucial objective of the WFS?

The answer varies between two groups of countries. In one group it can reasonably be said that an important purpose of the WFS is to measure the level of fertility in the population. These are countries with defective birth registration which have not recently carried out any census or survey from which a reliable estimate of current fertility can be derived, or, rather, an estimate as reliable as any which the WFS can hope to provide. All such countries are developing, though not all developing countries are in this position.

In the second group, which includes all developed countries, good estimates of current (or almost current) fertility are available already. Here, the primary objective of the WFS is investigation of the correlates of fertility, and study of behaviour and attitudes affecting fertility and family size. (These objectives, of course, are also included in the surveys for the first group of countries.) In this second group of countries, as in the first, the questionnaire includes a detailed pregnancy history which leads to estimates of current and past fertility, but in the second group these estimates will serve primarily for comparisons between population sub-groups.

In the first group of countries we have a single variable which could perhaps serve as a key variable for optimization, namely, current fertility at the national level. A certain amount of data are available from surveys all over the world regarding the variance of the fertility rate and we shall draw on this information in designing the sample. As will be seen in later sections, however, there are two distinct questionnaires recommended for the WFS: a household schedule and an individual questionnaire. In the first group of countries the household schedule would be used for a large sample of households while the individual questionnaire will be limited to a sub-sample. For these countries the data on current fertility will come from the household schedule, that is, the large sample. Thus the use of current fertility as the criterion for optimizing the sample may help to fix the sample design for the large household sample but will not indicate the optimum design for the small, individual sample.

In the second group of countries there is no overriding variable, and indeed the situation is the same in the first group when we come to the sample for the individual questionnaire. This questionnaire is complex and the analysis involves multi-way tables. There is probably not a single variable whose total population value is of significance, without breakdowns; still less is there one which can be said to represent "the objective "of the survey. Nor is there any obvious weighting of variables which could be cited in this way. As to a knowledge of the variances, costs and non-sampling errors, we have no more than a smattering of information, at present, based on some far-from-typical surveys.

Thus the World Fertility Survey is no exception: here, as in so many surveys, most of the details of the sample design cannot be optimized in any scientifically rigorous way, but must depend largely on judgement and hunch and numerous organizational constraints. We shall indeed, in this manual, use such data as are available from published sources, but these provide no more than a few hints as to the optimal strategy for sample design.

3. The population of study

A sample is designed to represent a particular population. In this section we define the population of interest to the World Fertility Survey and attempt to fix the conditions which qualify a household, or a person, to be interviewed.

Like almost everything else in survey work, the fixing of definitions has to be a matter of compromise. We can state what population we would like to cover but it may not be practicable to cover it. We may have to survey a more inclusive group and eliminate the excess at the processing stage, or cover a slightly smaller population than we want, and try to estimate the missing fraction from some other source. These problems will have to be kept in mind in the discussion which follows.

The household schedule and the core questionnaire for individual women have been published in a document entitled *World Fertility Survey: Core Questionnaires*. For the present section we need to note only the basic arrangements. The approach to the household begins with the completion of a household schedule, in which each person is listed with a few basic demographic characteristics. Thereafter, either at the same interview or a later one, the "eligible" females (to be defined later), or a sub-group of them, are interviewed with the individual questionnaire. In some cases there may be supplementary interviews with men, or re-interviews with the sample of women, but these operations will be left out of account until the final section of this manual. A more detailed discussion of the role of the household schedule will be found in Section 7.

In this section, we suggest some rules regarding eligibility of households and individual women for inclusion in the survey population. For the sake of convenience, we have stated these as firm rules without qualifications. However, most of them are arbitrary and it is not certain that they will always be the best rules. Thus if good reasons are found for modifying them in a particular case, they should be modified, though only to the minimum extent necessary.

3.1 Eligibility for interview: individual women

It would be useful to interview all women who are subject to the risk of pregnancy, and this could be defined simply in terms of age limits, say 15–49. However, in many cultures it is unacceptable to put questions on contraceptive practice to young girls who have never been married. In a few, it may even be unacceptable to ask such questions of women who are not currently married but have been married in the past. Thus many countries may wish to restrict the questionnaire to *ever-married* women and some even to the *currently married*. Yet such a variation in coverage would threaten the international comparability of the surveys. The existence of the household schedule, which contains no questions on family planning or current pregnancy but does cover all children ever born, offers some opportunities for escaping this conflict. If the individual questionnaire has to be restricted to ever-married women, at least we can extend the coverage of the household schedule to include all women in the relevant age range, and this may give us some idea of how much fertility we are missing; and perhaps allow extrapolation on some of the variables to cover the missing fraction.

The WFS core individual questionnaire has been drafted on the assumption that the survey universe will be *ever-married women* under 50 years old, and this is likely to be the widest acceptable definition in most countries (the word "married" being taken in the broadest sense). Where it is feasible to include *all* women of childbearing age (say 15–49), this should be encouraged. A possible compromise would be all women over 20 plus ever-married women aged 15–19.

As to the age range itself, a lower limit at 15 would be inclusive enough for most countries, but a few countries may find it necessary to go lower. No upper limit is needed for the household schedule. An upper limit of 49 is recommended for the individual questionnaire though 44 may be allowable in some developed countries.

Finally, we have a choice between a *de jure* and a *de facto* coverage definition. The *de jure* population consists of the household members, whether present or absent; *de facto* coverage includes all those present, whether members or not, but excludes those absent. Definitions are required for household *membership* and for *present/absent*.

Since the individual questionnaire requires an interview with the woman herself, the *de facto* definition is generally recommended for this, though with several important reservations (see Section 7.7). However, presence at the time of the interviewer's visit cannot be accepted as a criterion: this would automatically exclude making callbacks, and women who go out to work would in many cases be omitted, with an obvious bias. It is suggested that a respondent be classified as "present" if she spent "last night" in the household. Even this will result in some bias (against short-term visitors) and it is recommended that the *household* schedule should cover both the *de facto* and the *de jure* populations, i.e., household members, whether present or absent, plus visitors who spent "last night" with the household. This will provide some evidence on non-response bias.

Certain modifications of the *de facto* concept are possible, and, in some circumstances, desirable. Firstly, note that the above definition will lead to non-response in the case of a person who spent "last night" in the household but who moved away "today" before the interviewer's arrival. It is possible to remedy this bias by interviewing the counterpart group: those who, though absent last night, have entered the household since then and who expect to stay "tonight". Most of these would be returning members who were temporarily away. This modification may be desirable where such movements are very frequent, but it is generally *not* recommended, because of the added complexity needed in the instructions. Much of the non-response of this kind can in any case be picked up by callbacks, or by following the absent person to her new household. The latter is generally recommended, but only if the respondent's new location is *known* and *nearby* (same village in the rural case, same district in the urban case).

Secondly, there may be a significant proportion of people absent in an institution; or the number may be significant within an important sub-group of the population even if negligible in the population as a whole. (See also Section 3.2 below). The most common example would be women who go to a hospital to have their baby. Even in some developing countries this may be significant in the urban sector. If no arrangements are made to cover this by a special sampling of the institutional population, then a modification will be needed to the *de facto* definition in order to pick up such people through their own household. In such a case, and in a country where the institutional population cannot be ignored, a column should be added to the household schedule asking the present whereabouts of any absent household member. If an otherwise eligible woman is then reported by the household as absent in an institution, she should be regarded as *eligible*. An attempt should then be made to follow her up and interview her, either in the institution itself, or by a callback at her household after her return.

Thirdly, a special problem arises in some countries because of the time interval between the household interview and the individual interview. We shall see in Section 7.7 that in certain countries these two questionnaires will be completed at two different visits while in others they will constitute a single interview. If they are separated, the question arises whether "*de facto*" for the second (individual) interview, relates to the same "last night" as the first (household) interview, or whether we should base it on the night before the second interview. This question is bound up with the decision as to how the household interview and the individual interview are to be related. This is a somewhat complex problem and there are several options, which are discussed more fully in Section 7.7. To avoid burdening the present discussion, we leave undefined in the present section whether "last night" for the individual interview is to be taken as the night before the *household* interview or the *individual* interview.

Finally, the principle of *de facto* enumeration, based on a moving reference date ("last night"), relies on the assumption that visits are made at random and are not specifically related to the presence of a respondent. Many well motivated interviewers find it difficult to conform to this principle: if they find that a potential respondent is absent (and was absent "last night") they feel duty bound to callback at a later date when she has returned. This introduces a new "last night" specially chosen so as to ensure the respondent's presence. If this is done consistently we get a bias of over-coverage: such a woman has had a double chance of inclusion in the survey. A possible solution would be to define "last night" strictly as the night preceding the first attempt to get the interview, even if the attempt was unsuccessful.

However, this method is not generally recommended. It complicates the interviewer's instructions – indeed it is difficult to persuade interviewers that the procedure is truly correct – and the upward bias is in any case likely to be small. Note also that nearly all surveys suffer from a bias of *under*-coverage; the contrary "bias" we are describing here may well lead to the inclusion of some eligible respondents who would otherwise have been omitted. It is thus recommended that *de facto* enumeration be based on the *night before the first successful call for interview.*

Turning now to the definition of household membership, this involves such questions as:

How long must the respondent have resided in the household to establish membership? How long must she have been absent to lose membership? Is her intention to stay taken into account? There is no need for standardization between countries here: any definition is acceptable which minimizes the chance of a person being omitted or regarded as a member of two households. (See also Section 3.2 below).

To summarize: the following criteria are recommended for determining a woman's eligibility for interview (i) on the household schedule's fertility items and (ii) on the individual questionnaire.

	Household Schedule Fertility questions	*Individual Questionnaire*	
		EITHER	OR
Age	15 or over (12, 13 or 14 acceptable)[1]	Up to 49 (44 acceptable)	15–49 (44 acceptable)
Marital Status	Any	Ever married	Any
Residential Status[2]	Household members present; Household members absent; Non-members present (night before first successful call for interview)	Any person present (night before first successful call for interview)[3]	

3.2 **Eligibility for interview: households**

As with household membership, the definition of a household can be left to individual countries. In this survey the household is significant not as a reporting unit but as an intermediate step towards the setting up of a sample of eligible women. Moreover, most countries have some experience with household surveys and have developed a definition of the household appropriate to their own circumstances. To ask them to modify this for the sake of international comparability would invite new problems for little purpose.

Nevertheless, these remarks should not be taken to imply that the choice of definition of a household is a matter of no importance. It is rather that the overriding consideration on this issue is not international comparability but *consistency within the survey.* In particular, if a list of households is used for sampling (whether a pre-existing list or one specially made) and if identification of eligible women is done at a later visit by the individual interviewer, it is essential that the household definition used for listing and for the interview itself be identical. Failure to ensure this can lead to serious coverage error. For example, suppose two groups living in the same household are counted as a single household by the lister and this household is selected. If the interviewer regards them as two households, she will interview only one of the groups without being aware of the discrepancy. In these circumstances, the other group is arbitrarily excluded from the survey.

[1] *No upper age limit is necessary for the household schedule, though some countries may wish to introduce one.*

[2] *The residential qualifications for the household schedule apply to all persons listed, not merely the eligible women.*

[3] *See Section 7.7 for discussion of whether this should relate to the first call for household interview or individual interview.*

Ensuring rigorous definitions is not the only measure we can take to avoid this kind of slippage. Perhaps still more important is to provide the lister and interviewer with a catalogue of all the different types of difficult cases which can arise, for example, boarders, lodgers, servants, women staying temporarily in their mother's home, wives of polygamous husbands living outside the husband's dwelling, groups of "households" living in the same compound, families away on vacation, migrant labourers away from their regular homes, and so on. Instructions should be given, for each of these categories, on (i) How to ensure that such cases are not overlooked and (ii) How to group them – with the main household, or as a separate household. However, the most effective way of ensuring that the lister and interviewer are operating on the same groups is to have the lister list the names of all persons he considers to be included in the household. In many countries this will be the required procedure (see Section 7).

The inclusion of households of non-nationals seems again a problem on which countries will want to take their own decision. In some countries these households constitute a significant proportion of the population and the country may wish to include them in the survey. In others, it may be appropriate to include nationals of certain foreign countries but not all. Provided the excluded group is small (say less than 3 per cent), countries should adopt whatever ruling they consider appropriate for their needs; but if the group concerned is larger than this, it should be included.

Coverage of the institutional population (persons not living in households) has already been mentioned. Many of these groups are clearly outside the scope of any fertility enquiry (prisoners, certain hospital patients), or are all-male (armed forces), or negligible in numbers (hotel guests), but a few may be significant in some countries (students, nurses in hostels). The most important group, already mentioned above, is that of women having their baby in hospital. Where these are believed to constitute a significant fraction of any important survey sub-group, they should either be covered by a special sampling in hospitals, or by a modification of the eligibility definitions (as already described), to enable them to be picked up through their own household. Since their average stay in hospital is quite brief, the latter will generally be the more convenient approach. It is suggested that, as a general rule, all other institutional population groups should be excluded unless there seem to be exceptionally strong reasons for retaining a category in a particular case, or unless the total of such exclusions is estimated to exceed 3 per cent of the population.

3.3 Geographical coverage

The WFS aims in principle to cover every participating country exhaustively. It is recognized that some regions are relatively inaccessible and cases must arise in which certain areas are deliberately omitted from the sampling. Nevertheless, it is expected that these will be reduced to a minimum and that countries will make every reasonable effort to achieve complete national coverage.

3.4 Time reference

The timing of the survey is in the hands of the individual country, which will wish to take account of the numerous organizational, climatic and other factors. In principle, the execution phase of the World Fertility Survey is scheduled for a 3-year period beginning in mid-1974. Within this period, individual countries will choose a date for the survey to suit their convenience.

4. Physical, psychological and other constraints

4.1 Total sample size

The sample size adopted for the World Fertility Survey may vary from country to country. Several factors influence the choice and these will be discussed in turn below.

(1) *Sampling error*

As a rough approximation, it is true to say that sampling error depends on the sample *size*, not on the sampling *fraction*. Thus, as long as we are considering only the national total estimates, it could be argued that all countries should have the same total sample. But several factors intervene to modify this basic principle.

(2) *Sub-groups*

When a sample becomes too small, the random fluctuations increase so much that the data are worthless. Obviously, this affects first the sub-groups of the population – the individual cells of a cross-tabulation – since these get only a fraction of the sample; if the sample in these sub-groups is too small, the tabulation will be of no value. More than any other factor, the choice of sample size is determined by the amount of detailed breakdown required by the tabulation plan. While the WFS tabulation plan is fairly standardized in its main structure, there will be considerable variations between countries as regards *regional* and *ethnic* breakdowns. In a very broad way, one can expect larger and more heterogeneous countries to require more detailed regional and ethnic classification of the survey findings, and hence to need larger samples.

(3) *Inter-country comparisons and syntheses*

Other things being equal, sampling error is minimized for comparison between groups when the sample size is the same in the groups being compared. For paired comparisons between countries, this implies a fixed sample size for all countries. But often comparisons will be made between two groups of countries, or between ethnic groups each covering more than one country; within such groups, the optimal sample would be proportional to population. The same would be true for estimation of features totalled or averaged over several countries.

(4) *Cost*

At a given level of *per capita* income, a larger country can afford a larger sample than a smaller one. On the other hand, it is probably incorrect to argue that a developed country can afford a larger sample than a developing one: in developed countries rates of pay are higher. In practice, there does not seem to have been much difference between the total sample sizes adopted for comparable surveys in developed and developing countries.

(5) *Survey capability*
This is a highly important factor, too often ignored in fixing sample size. Poor organization and control are a common feature of surveys, particularly in developing countries; the errors which result can easily outweigh sampling error for all practical purposes. Since over-optimism on this subject is also a common feature, it is wise to adopt an attitude of great caution in fixing the sample size, with a deliberate bias towards smaller samples.

(6) *Variance*
The more homogeneous the population within strata, the smaller the sample can be to achieve a given sampling error. In most countries there is not enough information on this factor to justify taking it into account at the stage when sample size is decided.

(7) *Related surveys*
In some countries the fertility survey is closely associated with another study; for example, one survey may be conducted in a subsample of the other. Obviously, this can affect costs and other design factors and hence the optimal sample size.
In fixing the sample size for a given country, all the above factors have to be considered. Some of them argue for a fixed sample in every country; others imply that the sample should be proportional to the country's population. But several other factors have to be considered as well. Clearly no simple formula can be given; this is a matter for careful judgement in each case. As a broad indication, however, we suggest the following limiting principles:

(i) A sample of less than 2000 women would be unacceptable: there would be insufficient cases in the sub-groups.
(ii) With one or two possible exceptions, no country should attempt a sample of more than 8000 women.
(iii) Other things being equal, a larger country should expect to use a larger sample, but the relationship should be much less than proportional to population.

4.2 Total duration of field work
In most countries there will be constraining factors affecting the duration of the survey field work. Perhaps the work has to be completed between two rainy seasons; or during the school vacations; or before the next project, which is scheduled to a strict timetable. The conditions will vary from one country to another. It is expected that a total duration of 3–4 months in the field would be typical, or somewhat longer, if extensive field operations are required for mapping and sampling purposes.
A country with limited administrative or supervisory capacity might find it easier to spread

the work over a longer period, using a relatively small number of field workers. While this is attractive from the point of view of field control, there is a danger in pushing such an idea too far. In many countries there are seasonal migrations: if the survey team "migrates" in step with these (or exactly out of step with them), migrants may have a multiple (or zero) probability of being selected in the sample. What is really a seasonal trend may then appear as a spurious difference between regions. If the number of field teams is very small and the field period is prolonged, it will be difficult to avoid such effects. Thus, in countries where seasonal migration is thought to be substantial, the number of teams should not be very small and their movements should be planned so that regions of out-migration and regions of in-migration are covered simultaneously and not successively. Note that not all seasonal migrations are regional in nature: in many countries there is a movement from the towns to the countryside at harvest time and a return movement later. The same principle should be applied here.

4.3 Organization of the field work

It is considered essential that the interviewing of women with the individual questionnaire should be conducted by female interviewers. If the household schedule is completed at a separate interview (which it may or may not be – see Section 7), the interviewer for that operation could be male or female. Field supervisors could, in principle, be male or female. In some countries it may be argued in favour of male supervisors that females work better under men, or that men are needed to chaperone the interviewers. On the other hand female supervisors would have the advantage of being able to repeat the interview where necessary. However this question is decided, there will be many countries in which female interviewers will not be able to travel alone in rural areas.

These considerations are relevant to sample design because they imply the need, in most countries, for *teams* of field workers to visit the sample villages. The team method of field work also offers advantages as regards supervision and control. The supervisor can travel with the team each day and check the questionnaires each evening; omissions and other errors can then be corrected by a return visit to the household. Thus, it is expected that in most surveys in the WFS programme, at least in developing countries, the field work will be organized by the team method. Such a team would consist, typically, of about five persons under a supervisor; at least three of these would be female interviewers working on the individual questionnaire. The sample design must, therefore, provide enough work in each rural location to keep three interviewers busy during the team's stay.

The typical individual interview is expected to last 45 minutes. The number that can be conducted per day per interviewer will depend much on the distance between selected sample households and the means of transport. In rural areas of developing countries it would probably be advisable to assume no more than three households per day per interviewer for the individual questionnaire. In urban areas and in developed countries a higher rate may be possible.

Since some provision must be made for callbacks with persons not found at home (see Section 10), we must plan for a stay by the team of more than one day in each village.

All these considerations imply for developing countries a total sample in each rural location of at least 15 eligible women and preferably 25 or more. In a few cases it may happen that neighbouring villages in the sample are close enough for field workers to move between them daily; the above considerations would then no longer apply. However, this could hardly be very common.

Finally, the team organization is relevant to sample design in that it offers the opportunity for sampling in the field. This is discussed in later sections.

4.4 Contamination and other interactions

In nearly all cultures the questions asked in the WFS core individual questionnaire will be regarded as intimate and personal and it will be necessary for the interviewer to press for an interview situation in which she can be alone with the respondent. Thus direct contamination, in the sense of one respondent hearing the interview of another, is unlikely. However, other analogous interactions can occur and these affect two particular issues: whether one should interview all eligible women in a household or only one, and how many households to select in each location. We consider these in turn.

In a household where there is more than one eligible woman these will commonly be of unequal status – often a mother and daughter. If both are to be interviewed then convention will often require that the senior woman be interviewed first. Having heard the interview, she may then "protect" her daughter by refusing to allow her to be interviewed. This reaction seems less likely where only ever-married women are being interviewed (i.e., where the daughter will have been married). On the other side it is arguable that, if we interview only one woman per household and happen to pick on the younger one, the senior one (who may well be the initial contact) may refuse on her behalf. As we shall see below, there are sampling advantages in interviewing all eligible women in each selected household. The WFS provisionally recommends this procedure for most countries, at least until further evidence is available as to whether the above adverse reactions do in fact occur. It is hoped that a few countries will, nevertheless, try the alternative procedure (interviewing one woman only per selected household) so that experience of both approaches can be accumulated. (For further details of this procedure, see Section 7.7, penultimate paragraph.)

In a few countries surveys on sensitive subjects have run into a local reaction which has built up explosively after a few days of interviewing, when news of the interview content has spread around the area. This danger argues for a sample design which requires no more than about three days of interviewing in each location, say, a maximum of 50 women selected in each area.

4.5 Reliability of field workers

Finally, any good sample design should be based on a realistic assessment of field workers' performance. An example of one option which is ruled out on these grounds is sampling "as you go", e.g., asking the interviewer to list households in the field and, while doing so, to stop at every fourth one for the full individual interview. In several countries this has given biased results. The temptation for the interviewer to select the more convenient cases for the sample is too strong and it would be unrealistic to assume that such a bias can be eliminated by firmly worded instructions or even by supervision. Sampling of households should therefore be carried out by someone other than the interviewer who will work on the households selected.

A more difficult case arises when one asks interviewers to report a characteristic which then determines the need for further work. For example, she asks for age at the household interview and then has to complete the full individual interview for all women aged 15–49. In many surveys this kind of situation has been shown to lead to biased age reporting: presumably the interviewer pushes borderline ages out of the crucial age-group in order to avoid the work of the long interview. This could be prevented by insisting that the household interview be conducted at a separate visit by a different field worker (hopefully not in collusion with the other). This issue is discussed more fully in Section 7. Where there is sub-sampling of households for the individual interview, (i.e., a larger sample for the household interview than the individual interview), such a separate visit will be necessary in any case. But where there is no such sub-sampling, we have to ask whether a separate visit would be justified for no other purpose than to avoid the distortion in the age distribution (and possibly also marital status). We suggest that the cost of separating these visits would *not* normally be justified. At the most, it may be worth extending the age limits for eligibility by a year at each end in order to push the zone of age-distortion away from the age-range of concern to us.

5. Sampling frames

5.1 General remarks

The choice of a sample design depends above all on the availability of suitable sampling frames, that is, lists of units from which a sample can be selected. Units may be large areas, small areas, households, persons, or anything else. Where such lists do not exist, or are out of date, they can, in principle, be created by a specially conducted listing operation; but sometimes this may be so time-consuming and expensive that it is better to seek some other unit for sampling.

The ultimate unit with which the survey is concerned is the individual person. Persons are found in households, households in dwellings, dwellings in area units of various sizes (e.g., villages within districts within provinces). At each of these levels, a sampling frame, or list, may or may not exist; and if it does not, it may be necessary to make one. We shall consider them in turn but, first, we note the characteristics which are required of a good sampling frame for any unit.

1. It should be exhaustive: no units omitted.
2. It should be non-repetitive: each unit listed once only.
3. It should be up to date.
4. The units should be clearly and unambiguously demarcated. For example, for area units the boundaries should be clearly given; for social units, such as households, the membership criterion should be clear.
5. The units in the list must be traceable in the field. For example, there should not be so many villages of the same name, or so many household heads of the same name, that it is impossible to identify on the ground the one selected on paper.

In addition, for efficient sampling, it is desirable that the units should be fairly constant in size or, if this is not so, that the approximate size of each unit should be known before sample selection. "Size", for this purpose, means the number of elements which the unit contains, e.g., the size of an area is the number of households or eligible women in it. Finally, it is useful if the frame or list also shows relevant characteristics of each unit: "relevant" here means *correlated with the main survey variables.* Where this is so there is opportunity for stratification, with a resulting reduction in sampling error.

In the various stages of a sample it often happens that at one point we wish to select *all* the units listed. This is perhaps not strictly "sampling", but it is convenient to use the same word to describe such "take-all sampling". The problems of obtaining a sampling frame still apply: in order to take all we must have a list, or something equivalent.

5.2 Sampling frames for individuals and households

In seeking sampling frames for individuals and households, the most serious problems are (3) and (5) above, i.e., *up-to-dateness* and *traceability*. Up-to-dateness is a special problem in a fertility survey because, in almost any society, both the movement of individuals and the formation of households are closely related to marriage, pregnancy and childbearing: the essential topics of our survey. For example, in many societies expectant women habitually move to their mother's home to have their baby and return again to their husband's household a few weeks after the birth; thus if women are listed and then visited a month or two later for interview, we will selectively miss those who are in the last stages of pregnancy as well as those who have a very young child (the latter group were listed at their mother's home but have left it again by the time of the interview). In developed countries a common defect of fertility surveys has been the use of an administrative household list for sampling. Such lists are almost inevitably several months out of date and thus yield a sample seriously biased against new households, causing distortions of data on fertility and family planning. For these reasons, it is advisable to avoid altogether any use of lists of individuals except those that are made by the survey field workers themselves, either at the same visit as the individual interview or very shortly beforehand when the household schedule is completed (see Section 7).

Households are, of course, less mobile than individuals but it is still not advisable to use any household list which is more than a few months old. This virtually rules out the use of any pre-existing list except, perhaps, in the rare case where the survey follows shortly after a census or another survey. In developing countries, even a census often does not provide a satisfactory household sampling frame: addresses are generally inadequate and the field work is often of poor quality so that the selected households cannot be identified. Another, earlier survey may supply an acceptable sampling frame if the time interval is very short, but this assumes that the sample designs are compatible. In practice, a fertility survey will nearly always have to depend on a household listing operation carried out specially for the survey, unless a dwelling sampling frame can be used as a substitute.

5.3 Sampling frames for dwellings

Units of housing are variously termed "dwellings", "structures", "addresses", "housing units", etc., in different countries and it is difficult to conduct a discussion about dwelling sampling without confusing some reader who uses the words differently from the author. In the present section, we use the term "dwelling" as broadly as possible to mean any identifiable, structural unit of habitation which might be used for listing and sampling. A dwelling *may* correspond exactly to a household – and it will certainly be helpful if it does, since households are what we want to interview. But in many circumstances, when lists are made of units of habitation they do not correspond one-to-one to households; we shall, nevertheless, use the term "dwelling" to include such units.

Dwelling sampling frames are considerably more durable than lists of households; moreover, where lists of dwellings exist, the individual selected dwellings are likely to be quite easily identified. In some cases, a dwelling sampling frame can substitute for a household frame: after selecting the sample of dwellings, we survey all households in the selected dwellings. This works well in the following circumstances:

a) if households are never split between two units listed as separate dwellings;
b) if one dwelling corresponds normally to one household, perhaps occasionally to two or three, but never to a large number;
c) if we are willing to send an interviewer to all households, irrespective of composition. In some situations (see Section 7), we wish to use the listing to identify households containing females eligible for interview: lists of dwellings are unlikely to provide such information.

Dwelling lists in developing countries, and in many developed countries, are likely to exist in urban areas only. Apart from this, dwelling lists in any country are likely to be incomplete in two ways. Firstly, they may not cover recent housing developments. This defect can usually be corrected by some careful checking in local or municipal records, with the help of local knowledge, followed by preparation of supplementary lists. Secondly, there are often omissions resulting from structural changes to existing dwellings, or changes from non-residential to dwelling use. These can usually be rectified by the "half open interval" method, which operates as follows. It is assumed that, barring errors, neighbouring buildings have consecutive numbers.[1] It is also assumed that the numbers appear in the sampling frame list. Then, if the interviewer is sent to dwelling number N she must interview all households living at dwelling N or anywhere between that dwelling and dwelling N + 1.[1]
In many countries, postal addresses fulfil the above conditions and these are listed in directories, but in other cases, such addresses do not satisfy the necessary conditions; for example, there may be a large number of households at each address.
Dwelling lists may substitute for household lists not only in the situation where the dwelling list already exists, but also where we have to make our own list. Dwelling lists can generally be made much more easily and quickly since it is not necessary to interview anyone. If the listed dwellings cannot be identified by means of an existing address system, one alternative is for the lister to attach numbers on buildings. This can be done with paint, chalk, number discs or paper stickers. Each method has advantages and drawbacks but perhaps stickers win by a short margin on grounds of convenience and adaptability to different wall materials. Another alternative is to list the dwellings in terms of the name of the head of the household (or chief household) which occupies it. If this is done, two of the advantages of dwelling lists over household lists are lost: greater durability of the list and avoidance of the need for interviewing by the lister. But in many cases, there will still be one advantage remaining, namely, that dwellings are most easily defined than households so that listing of dwellings requires less probing enquiries and proceeds more quickly. Note that if dwellings are listed

[1] *If even numbers are on one side of the street and odd numbers on the other, then "consecutive numbers" above should be read as "consecutive even numbers or consecutive odd numbers" and "N + 1" should be read as "N + 2".*

by name of the chief householder, the list must be used within a few months at most. One should also warn the interviewer that the householder's name is used only for identifying the dwelling; the household to be interviewed is the one (or all the ones) inhabiting the selected dwelling at the time of the interview survey even if this is not the household listed. When the dwelling lists are made for the survey itself, it is nearly always desirable to make a sketch-map during the listing operation and to show the location of each dwelling on the map. If numbers are affixed to the dwellings then these numbers should also be shown on the sketch-map. If dwellings are identified by the name of their household head then these names should be listed and numbered and again the numbers should be shown on the map. It is also often useful to cross-reference the dwelling numbers to a careful description of particular features of the dwelling or its location.

"Home-made" dwelling lists used as substitutes for household lists must satisfy the same conditions, mentioned above, as pre-existing lists, that is, firstly, the units must always cover a whole number of households, usually one, and never very many. Secondly we must be content to send an interviewer to all households in the selected dwellings without knowing in advance whether each one contains a woman eligible for interview.[1] A further condition will emerge in Section 7: the data requirements must be such that we do not need to collect substantive information during the listing operation. Provided these conditions are all satisfied, it should be both cheaper and more convenient to make a dwelling list than a household list.

In this section, we have said much about the use of a dwelling list as a *substitute* for a household list – a type of design which presupposes that each household in the selected dwellings will be included in the final interview survey. There is also, however, the possibility of the dwelling being used as an intermediate sampling unit. This may be appropriate where the "dwelling" (often termed a "structure") is something much larger than the unit considered above. In this kind of design, after the sample of "dwellings" has been selected, a field operation is carried out to list all the households found in the selected dwellings. The final household sample is then selected from this list. This arrangement has been widely used in developing countries, though usually in urban areas only. Note that it does not require that conditions (b) and (c) above should hold, although (a) still applies, i.e., we must not have households split between more than one dwelling.[2] It often happens in developing countries, however, that such "dwellings" or "structures" vary greatly in size; for example, from 1 to 50 households within the same district. This will cause a large sampling error unless some way can be found of stratifying the dwellings in the list according to their size.

5.4 Area sampling frames: mapping versus listing

In order to limit the amount of time and money spent by interviewers in travelling, the organizer of any interview survey wishes to arrange that each interviewer's sample of households should be concentrated in a small area. Moreover, we have seen that a special listing operation, of households or of dwellings, will normally be required for a fertility survey and,

[1] *If we collect this information during the listing operation, we have a household list or an individual list, not a dwelling list.*

[2] *If, nevertheless, we can find out when this is occurring, we can make a weighting adjustment to allow for it and condition (a) is then no longer essential. However the risk of error with such a procedure is high and the method should not be considered unless such cases of split households are known in advance to be quite rare.*

here again, to keep costs within reasonable bounds it will be necessary to limit the listing to areas of fairly small size. Thus, both for the listing and for the interviewing itself, there is a requirement for a sampling frame of small area units. Such a frame may already exist in the form of lists or maps, but if it does not, we will have to create one by a special mapping operation within a sample of larger area units. Once again, in order to limit the cost of *this* operation we seek an existing sampling frame of area units, and the smaller the units for which a frame exists, the better.

Thus the first step in area sampling is to look for a sampling frame covering the smallest area units possible. If these are small enough, we can select a sample of such units (in one or more sampling stages) and then proceed direct to the listing of households or dwellings, within the selected areas. But if they are too large, there has to be a further stage of area sampling: in that case we first select a sample of the large units, then conduct a special mapping operation within the selected units in order to divide them into smaller area units. We then select one or more of these and carry out the household or dwelling listing within the selected area. What exactly is meant by "small enough" and "too large" in this formulation? In other words, how small an area should we seek, and when does it become more profitable to start listing households or dwellings than to create smaller area units? Several factors are involved, which will now be examined in turn.

Dividing an area into smaller area units by mapping is generally less laborious, and less costly, than listing every household or dwelling, and the difference becomes larger the larger the size of the unit to be divided. On the other hand, mapping only works well when physical features can be used as boundaries to delineate the areas – roads, rivers, fences, etc. – and this means that it is difficult to make clear maps of very small areas containing only a few houses. Thus, for large areas mapping is best, but for small areas listing is preferable. Where is the break-even point? As this question involves balancing cost against accuracy, and the latter depends on the skill of the field workers as well as the character of the terrain, it is not feasible to give any precise answer to this question. However, a reasonable and practicable rule, at least for rural areas, is to aim for areas corresponding to the smallest natural population groupings – *villages* in most places. This is a sensible policy, firstly, because it is above all the attempt to draw boundaries *within* villages that causes difficulties and, secondly, because the use of a natural social unit eases the task of the field workers in gaining the co-operation of the local people. In most of Africa and Latin America, a unit size from about 400 to 1,000 population (say 100–200 households) would be large enough to include most villages, but in much of Asia, the size would have to be considerably bigger than this.

Yet this conclusion needs important qualifications. The cost of listing depends crucially on whether it is going to be necessary to contact someone in each dwelling/household or whether the listing can be done without this. If there is a good address system satisfying the conditions cited in Section 5.3, or if dwellings are arranged in some very simple way (e.g. on each side of the highway in single file, or in a regular grid), or if the village chief has a list of all the households, then listing may require no more than a brief visit to the area to check that this is so.

In this case, listing costs are very small. In an intermediate case, it may be found necessary to affix stickers on every dwelling but the relationship with the authorities perhaps allows this to be done without consulting the householder and the dwelling units are so distinct that there can be no ambiguity in identifying them. Listing costs will then be somewhat higher but not excessive: perhaps one lister will cover 50–100 dwellings per day. But in other cases – and these may well be the majority of cases in developing countries – listing will require contact with a household member, either to obtain permission to affix the sticker, or to obtain clear identification of the unit listed. In this case, listing will be a costly operation – a single lister may not cover more than 20 households per day. Listing of a unit as large as 100 households could then take 5 man-days – perhaps 50% of the cost of the individual interviewing in the same area unit. Clearly in these circumstances one would wish to reduce, as far as possible, the size of the ultimate area units, and a unit as large as 100 households would appear barely acceptable.

Another important factor which affects the choice of size of area units is the population density. The size 100–200 households suggested above assumes that the households are clustered together so that travel between them (both for the lister and for the individual interviewer) will not be very time-consuming. However, in some parts of the world rural populations are widely dispersed and a "cluster" of this population size would be excessively large in area. Where this is so, the optimal size of area units will necessarily be reduced. In certain cases, the need for simplicity will argue in the opposite direction and in favour of a larger unit, particularly where dwellings can be listed without the need to contact any household member. The argument here is as follows.

Dividing larger area units into smaller units involves a special field operation. Instructions have to be prepared, field workers have to be trained and even in the best circumstances errors will sometimes occur. If this stage can be skipped, the procedures will be simplified and the risk of error reduced. This factor should be weighed against any cost saving. If, for example, we desire small area units of 100 households and we have an existing sampling frame of units which average 300 households, it is doubtful whether the additional operation of creating smaller area units would be worth the trouble. It would be much simpler, and perhaps also less costly, to carry out a dwelling listing in the complete large area without breaking it up. It is important to bear in mind that we have, at best, only a very rough idea of the true optimal size for the final area units. Generally, one would want to consider introducing a new stage of area sampling only if the available units were *much* too large: perhaps five or more times the estimated optimum, though this is a matter for individual judgement.

To summarize the issue of mapping versus listing, omitting the refinements: We begin by seeking the smallest type of area for which a sampling frame exists. The ideal size of rural area units is very roughly that of a village. If this ideal is *substantially* smaller than the smallest type for which we have an acceptable frame, then we do a mapping operation within the selected large area units to create smaller area units of roughly ideal size. We select a sample of such smaller area units and within these we list all households, or all dwellings.

If we can make a dwelling list, and without contacting household members, this is preferable. If not, listing will be costly and the ideal size for the ultimate area unit becomes smaller – almost certainly under 100 households.[1]

These procedures are suggested as a general rule but may not be ideal for every individual case. In particular, household listing may sometimes serve some additional function and this may influence the choice of the size of the ultimate area unit.

5.5 **Area sampling frames: demarcation of boundaries**

At the beginning of Section 5, we discussed the characteristics required of a good sampling frame. In the case of area units, condition 5.1 (4) is of particular importance, namely, *clear demarcation of the boundaries of the units.* Some degree of fuzziness about boundaries is almost inevitable. However, the more effort we are willing to put into a mapping operation, the more this uncertainty can be reduced. One of the most difficult problems for the practical sampler is to judge how much effort in this direction is justified, that is, to find a balance between the cost of the work and the pay-off in terms of bias reduction.

Before considering this, it will be helpful to decide the different types of area sampling frame that may be available in different countries. The ideal frame is a map on which small areas have been marked off, using natural boundaries, as far as possible, and on which the main localities have been entered. In some countries such maps have been made for administrative purposes; in others for the population census, the small areas being then *census enumeration areas*. In either case, there may well be data on the population of each area unit, probably out of date and inaccurate, but nevertheless worth using for the purpose of sampling with probability proportional to population. In many countries such maps are available but only for relatively *large* area units, so that a further mapping operation is needed for the survey in order to make suitably small units.

In some countries the maps are supplemented by verbal descriptions. This has been done in several censuses in West Africa, for example, because it is found that inexperienced map-makers often make errors which can be corrected by reference to the description.

In many countries, there are *lists of villages*, classified by province, district, or other administrative area, but without any maps showing the precise location of each village.

Finally, there are sometimes aerial photographs. Until these have been turned into maps they are unlikely to be of much use to the sampler except in urban areas.

In towns, there will usually be maps or aerial photographs from which it is quite easy to mark out identifiable area units (e.g., city blocks); in most cases the only serious problem is that of up-dating.

Turning now to the problems of clear demarcation of boundaries, there are two main types of demarcation error: the boundary may be given clearly but wrongly, or it may not be given clearly. We consider these in turn.

In a map, the boundaries are almost always clearly shown but they may be erroneous in one of the following senses:

[1] *The question of combining and splitting exceptional units is discussed in Section 9.1.*

(1) A place (village or town) may be entered at the wrong location on the map.
(2) In the case of a map of census enumeration areas, with accompanying census population figures for each area, the boundaries shown may not be identical with those used in the census, so that the population figures are wrong.
(3) There may be a conflict between the map and the accompanying verbal descriptions.

None of these problems are serious: we only have to decide in advance which source of information is to have priority. For example, the following rule has been used with success: use the map as a sampling frame and accept the boundaries as drawn on the map except where they conflict with the verbal description, in which case accept the latter. This implies ignoring errors (1) and (2) above. This will not cause any bias. (The use of the wrong population figure will cause an increase in sampling error – usually negligible in practice).

A further common error arises when the boundary is clearly and correctly defined but the field worker makes a mistake of interpretation – following the wrong path, or the wrong stream, for example. Most such errors are likely to occur at random and will, therefore, not cause any systematic bias.

Thus "wrong" boundaries may not be a very serious source of error in practice. As long as the units in the frame cover the whole country, without gaps and without overlapping, it does not matter if they are "wrong". Unclear boundaries, on the other hand, can cause serious bias because they allow the field worker a choice and the choice may be motivated, for example, by the desire to avoid additional work. The main problems of this type arise when a village list is used as a sampling frame, without any detailed maps. Two sources of error are especially common:

(1) The list covers main villages only, dependent hamlets not being individually listed, on the assumption that they are included with the main village. The field worker is motivated to omit outlying hamlets because of the work involved in visiting them. Result: a sample bias against people who live in the remotest areas.
(2) In many countries, the villages are regarded, by their inhabitants, as ethnic rather than geographical units. This may result in the omission of temporary settlers in the village who owe allegiance to another chief or headman, or of squatters from another area who settle near the sample village. Similarly, in Europe there is a likelihood of omitting a camping site.

Motivated omission can be reduced if the fixing of boundaries in the field is performed by a person other than the one who will do the interviewing. Some of the other errors can be reduced by careful supervision and by strict rules requiring reference back to the office in case of doubt.

It should be noted that the significance of a given degree of demarcation error will vary between different types of survey. In a dual system or multi-round survey, this source of error is of special importance because of the need to maintain constant coverage in the different operations. In the World Fertility Survey, which is a single round operation, this is a less crucial consideration. Note also that bias due to coverage error is more serious for estimation of totals than of rates, means and percentages. For example, an under-coverage of x per cent will cause (other things being equal) an under-estimate of x per cent in any total, but if we are estimating rates, means or percentages, with both numerator and denominator coming from the sample, coverage error will cause bias only if the value for the population omitted is different from that for the population covered, and generally this difference has to be very large for the bias to be appreciable. For example, suppose that the crude birth rate is 50 per 1000 but that this falls to 30 per 1000 among the people living in the remotest areas (an unusually extreme differential). Suppose that the latter group accounts for one-tenth of the population and suppose there is a 25 per cent under-coverage of this group (of which we are not aware). Then the resulting bias will cause us to estimate the birth rate at 50.5 instead of 50.

5.6 The WFS in the context of other surveys, past and future

The discussion above on sampling frames has assumed that the fertility survey is essentially an isolated operation. In some countries, however, it will be planned as part of a programme of surveys. If it precedes the other surveys, the mapping and listing operations of the WFS may provide an investment from which the other surveys can profit. If it follows other surveys, the WFS may itself make use of their mapping and listing work.

The possible interrelationships are too numerous to be examined individually here, but it may be useful to list the main problems that have to be considered when planning such combined operations.

(1) If *lists* compiled for one survey are to be used for another, account must be taken of the rate at which they go out of date. Lists of persons are virtually useless unless the two operations are almost simultaneous. Lists of households generally have to be used within a few months. Lists of dwellings may be useful for much longer, depending on the method used for identifying the dwellings. Lists of areas are valid for many years; the population figures for the area units may go out of date but are still likely to be useful for several years as supplementary sampling information.

(2) *Mapping operations* conducted for one survey may be useful for another, if the methods of sampling the area units are compatible (stratification, use of varying sampling probabilities, etc.). It may then be profitable to design an area master sample.

(3) *Timing of linked surveys* often creates problems. Some surveys, e.g., in agriculture, have to be conducted at a particular time of the year. Others, e.g., household budget surveys, often have to be spread over a long period. Since the WFS will use female interviewers, there may be no clash in the field work itself, but overloading of organizational capacity and office facilities is a familiar problem in developing countries when too many operations are taken on within a short period.

If these problems can be solved, the effect of sharing the sampling work between several surveys will be to justify the expenditure of greater resources on sampling, mapping and listing than would be reasonable for an isolated survey. This, then, has to be taken into account in evaluating the various options which have been considered earlier in this section. For example, if an area master sample is to be created as an investment, then a considerable effort for mapping will be justified.

6. Preliminary description of the sample design recommended

Although several features of the sample design remain to be considered, it may be useful at this stage to interrupt the discussion with a rough description of the kind of sample envisaged for the WFS. This should make the later, more detailed discussion easier to follow. We also give references to the sections in the manual where each point is more fully treated.

The first step is the search for a sampling frame of area units whose boundaries are reasonably well defined and whose individual population sizes are as small as possible. For convenience, we may call these "basic area units". In some countries these basic area units will be large, in others small; whatever their size it is convenient to take these units as the starting point when planning the sample because they mark the point beyond which the sampling process has to leave the office and move into the field. The basic area units may be primary sampling units, intermediate units, or ultimate stage area sampling units; this now has to be decided.

If the basic area units are small, and/or if the country is large, we may decide to introduce one or more area sampling stages above them; for example, if the basic area units are listed by *district*, we may first select a sample of districts and then, within the selected districts a sample of basic area units. The main purpose of this will be to save travel costs by clustering the field work. Sampling at these stages will be stratified and, in general, with probability proportional to size. (See Section 8 on stratification and Section 9 on sampling with probability proportional to size and multi-stage sampling.)

If the basic area units are small enough – perhaps a maximum of 200 households each – they will probably be acceptable as *ultimate* area sampling units. See Section 5.4 for discussion of the optimal size of such units: the crucial factor, but not the only one, is the cost of listing dwellings or households within the unit. If the basic area units are larger than this, then a further stage of area sampling should be introduced: each selected basic area unit will be divided up in the field into smaller units. Possibly two such sampling stages will be needed if the basic area units are very large. Again, sampling will generally be with probability proportional to size. We thus arrive again at a sample of small area units: this time the units are "home made" and should therefore be close to optimal size. In most cases the optimum will probably be under 100 households,[1] possibly as low as 50, but if *listing without interview* is possible it may go well above 100. These are now the ultimate area units.

In each selected ultimate area unit we conduct a listing operation. If possible we list *dwellings*, but in many cases clear identification of the dwelling would require an interview with the household and in this case we list *households*. (For other circumstances in which dwelling

[1] *The discrepancy between this figure and the 200 maximum suggested above is accounted for by the "need for simplicity", discussed towards the end of Section 5.4.*

sampling would be avoided, see Section 5.3). The household schedule may be used at this stage (see Section 7).
Finally, we select from this list in each ultimate area unit enough dwellings or households to yield an expected sample of between 20 and 50 eligible women. (Sections 4.3 and 4.4 give reasons for these limits). The probability of selection at this stage will be such as to give a self-weighting sample (see Section 9). In most surveys, in each selected dwelling or household there will be an interview with every eligible female (see Section 3.1 for criteria of eligibility).

In developing countries the number of eligible women per household generally works out around 1.0; in developed countries it is typically about 0.4. It will be seen that the maximum recommended "take" per ultimate area unit corresponds to 50 households in a developing country and that this is also the minimum recommended size of the ultimate area unit. While neither of these extremes would normally be a desirable figure it may happen, exceptionally, that the sample designer finds his estimates of the optima for both of these parameters approaching the common boundary of 50. In this case, there is clearly an advantage in "take-all sampling" at the household stage (compact cluster campling). In developed countries the optimal take is likely to be much less than 50 but on the other hand the frequency of women eligible for interview is much lower, so that again such an overlap can occur and compact cluster sampling may occasionally be desirable.

7. The role of the household schedule

7.1 Fertility questions in the household schedule

The household schedule is attached as an appendix to this manual. It will be seen that it contains a block of questions on fertility, addressed to each woman above a certain minimum age. These questions reappear in Section 2 of the individual questionnaire and it may be wondered why we ask them twice.

Before this question is answered, two general points should be noted. Firstly, the schedule shown in the appendix is a maximum version: in many countries the fertility questions would not be included. Secondly, the fertility questions would never be asked twice *at the same interview:* if the household schedule is completed at the same interview as the individual questionnaire, the basic fertility questions would be removed from one or the other to the extent necessary to avoid repetition.

In many countries there is a pressing need for information on the current level of fertility and this can come from the household schedule. Schedules of this kind have been widely used in surveys and censuses in developing countries. Their usefulness depends on the application of analytical techniques, devised by Brass and others, designed to correct and adjust the data for reporting errors.[1] Some of these methods involve the fitting of model distributions and experience has shown that reasonable confidence in the estimation of the relevant parameters requires a sample of the order of 20,000 households.[2] Moreover, large-sample data on children-ever-born also offer the opportunity of measuring *fertility differentials* on such items as educational level of husband or wife, religious or ethnic affiliation, location of residence, etc., with a smaller sampling error than would be possible in the individual sample.

Thus the purpose of these questions in the household schedule is to obtain some simple fertility data on a sample several times larger than we can afford to use for the individual interviews. The household interview is cheaper both because it is shorter than the individual interview and because any respondent will do – we do not have to obtain an interview with the woman concerned, thus saving time on callbacks. It is also cheaper because it can often be combined with the listing operation (see Section 7.3); where this is so the listing may justify a large part of the cost, and only the marginal cost of actually putting the fertility questions to the respondent needs to be attributed to the household schedule. The counterpart of these advantages is the lower quality of the data, but corrections are made to allow for this at the analysis stage. The question whether such large-sample data are required in a given country must be answered by demographers familiar with the available data for that country and with the methodology of this kind of analysis; in this manual, we merely need to note that in many developing countries there will be a demand for such data in the context of the WFS, and these will be countries lacking adequate estimates of current fertility.

1 *Methods of estimating basic demographic measures from incomplete data,* UNITED NATIONS, SALES NO. 67.XIII.2.

2 *Methodology of demographic sample surveys,* PP. 46–57, UNITED NATIONS, SALES NO. E71.XVII.11.

7.2 Estimation of fertility levels by two-phase sampling

The fertility information in the household schedule is obtained in a rough and ready way: we do not insist on interviewing the woman herself and in many countries we will not be using female interviewers at this stage. But the information does have the advantage of coming from a large sample. Thus we can expect relatively high reporting error but relatively low sampling error. The position is the reverse as regards the individual questionnaire: here the sample is small but the interview technique is more sophisticated. This situation offers the well known possibility of the *two-phase sampling estimate:* we use the large sample to reduce sampling error but we correct the data with an adjustment derived from the more reliable small-sample data. Specifically, we estimate from the fomula[1]:

$$\text{Large sample, household schedule} \times \frac{\text{Subsample, individual qre}}{\text{Subsample, household schedule}}$$

As is shown in most sampling textbooks, the gain in precision with this method is likely to be substantial in practical applications, but despite this it is rare that this gain in itself will justify the additional cost of the large survey. Only if the large survey is *much* cheaper to run and the correlation between the two sets of data is very high will the large survey be a good way of spending additional resources. In the WFS, the cost per household of the household interviews is unlikely to be low enough to satisfy this requirement. Thus, if a household interview on a large sample is not required for some *other* reason, it should not be introduced for the *present* reason. On the other hand, the marginal cost of *including the fertility items* in a household interview that is going to be required anyway should be relatively low. Thus, if a household interview on a large sample is going to be needed – as it may be in certain countries for listing purposes – then the present argument would be a strong one for including the fertility items in the interview. And if a full household schedule, with fertility items, is going to be used on a large sample in any case, for the reasons described in Section 7.1, then the two-phase sample estimate should certainly be used at least for the more crucial variables. For the two-phase estimate to be useful, two conditions must be satisfied:

(1) In the small sample (individual questionnaire) we must have results from *both* the household schedule *and* the individual questionnaire, otherwise the estimate cannot be computed.

(2) The data from the household schedule in the small sample must be obtained under strictly the same conditions as those in the large sample. In particular, they must not be modified afterwards to agree with corrections discovered during the individual interview.

7.3 Main roles of the household schedule

We are now in a position to state the three main possible roles of the household schedule. These are:

[1] *This is the ratio estimate – the most commonly used in two-phase sampling estimates. One could alternatively use a difference estimate or a regression estimate.*

(1) Listing of households to provide a sampling frame.
(2) Listing of persons to identify those eligible for individual interview.
(3) Obtaining fertility information on a large sample.

Only the second of these will apply in all cases. The listing of households, as we have seen, may sometimes be replaced by a listing of dwellings, while the third objective depends on country needs. Moreover, if the listing of households were done at one visit and the listing of persons at another, we would presumably not need the same kind of household schedule for both. Thus, for the moment, these are described as *possible* roles. However, in those countries where two or three of these operations are needed there appears at first sight to be a possibility of combining them into a single interview. Before examining exactly how this should be done, we will take a closer look at the two listing operations.

7.4 **Listing of households**[1]

We saw in Section 5 that a special field operation will usually be necessary for listing dwellings or households in order to obtain a sampling frame from which we can select the households to be interviewed. A crucial question is whether this listing can be done without the need to contact a resident of the dwelling or a member of the household. If such a contact is necessary, then we will have to contend with the time spent on these interviews besides the necessary call-backs. But if listing can be done without the need for such personal contacts the work will proceed very much more quickly, and hence cheaply.

In many cases this listing operation will be carried out at a separate visit from the listing of persons for the individual interview (see Section 7.7) and often by a different field worker. It is thus crucially important that the unit identified by the lister can be later accurately re-identified by the interviewer. Slippage at this point could cause serious bias (see Section 3.2). The minimum information which must be recorded by the lister is, first, a clear address pin-pointing the exact dwelling or household (where addresses do not exist, the lister should, if possible, affix numbered stickers to the dwellings and provide a sketch map and descriptions) and, secondly, "boundary" information to define the dividing line between one dwelling or household and the next. Sometimes dwellings are distinct and clearly defined; but in developing countries very often they are not and households hardly ever are. Where such uncertainties exist it may often be that the best way of identifying households is to list their members. Where this is done such a listing of persons may also serve the purpose of identifying eligible women and there may be no need to introduce a listing of persons just before the individual interview.

Thus, the lister of households may also perform the role of listing persons. One other task he may carry out – crucial in some countries but insignificant in others – is to identify the language or ethnic group of the household so that a suitably qualified interviewer can be sent for the individual interview.

At the end of Section 6 we mentioned the possibility that, exceptionally, it may be decided

[1] *The term "household listing" is avoided deliberately; it can be confused with the listing of household members, here called "listing of persons".*

not to sample dwellings or households at the last stage but to "take all" (compact cluster sampling). Will this avoid the need for listing of dwellings or households? Where dwellings are numbered systematically, correspond closely to households, and can be clearly distinguished one from another, it is probably acceptable to drop the listing operation. But survey specialists working in developing countries generally recommend pre-listing even for take-all sampling, because without it the risk of omissions by the interviewer is too great. We shall assume henceforth that listing will always be necessary in such cases, although in developed countries it may occasionally not be.

7.5 Listing of persons

For the purpose of identifying women eligible for the individual interview we need a list of household members by name, sex, age, marital status and presence "last night". To reduce interviewer bias it is advisable to require a full listing of *all* household members on these variables, not just the women. Note that these requirements, in themselves, do not imply a schedule anywhere near as elaborate as that reproduced in the appendix.

7.6 Relationship between the household interview and the individual interview

We have now seen the main roles which might be played by the household schedule. We next consider whether these different functions can indeed all be fulfilled at once and, more generally, how the necessary operations may be organized with maximum efficiency into a single plan. For brevity we shall use the term "eligible woman" to mean a woman eligible for the individual interview and we shall call a household which contains one or more eligible women an "eligible household".

There are five possible field operations to be considered:

(A) Listing of dwellings
(B) Listing of households
(C) Listing of persons to identify eligible women
(D) Collection of fertility data on a large sample (household interview)
(E) Individual interview.

Note that a list of dwellings may already exist, in which case (A) is not a *field* operation. If (A) *is* a field operation, then (A) and (B) would be alternatives; we would not have both. Note also that (D) will not always be required. Thus the above list represents a maximum. Obviously efficiency requires that we reduce the number of visits to a minimum. How far can the operations be combined?

First, can we combine everything into one visit? This would mean having the interviewer list households and stop at every n-th eligible one for an individual interview. The objection to this is the usual objection to sampling "as you go", mentioned in Section 4.5: experience in many countries shows that it is very difficult to get interviewers to perform such sampling reliably. This method is therefore not generally recommended. There is, however, one situation in which a single field operation can cover all that is needed. In some countries, notably in Europe, adequate lists of dwellings or households already exist and can be sampled. In such countries (D) is also unlikely to be needed. Thus (C) and (E) are the only field operations required and these can be performed in one visit. In all other situations, however, two visits seem to be the minimum.

If two visits are used, obviously the first would be the listing of dwellings or households (A or B), covering every dwelling or household in the ultimate area unit, and the second would include the individual interview (E), covering only a sub-sample. Where do we put (C) and, if required, (D)? The answer depends on the conditions affecting the listing of dwellings or households. As we have seen, it will often happen that dwelling listing will be judged too unreliable and that we have to opt for listing of households; and in such cases it may further be judged that the only clear way of identifying a household is to give a list of its members. In this situation, (B) and (C) are brought into one operation almost inevitably. Moreover, if (D) is required (fertility information on a large sample), this can be included in the same operation at a very small marginal cost since we are already collecting some details on each individual. This arrangement may well prove to be the commonest in developing countries. However, it will sometimes happen that dwelling listing is judged to be sufficiently reliable to obviate the need for the listing of households, or that a listing of households can be made without the need for an interview. In this case the listing operation would proceed much more rapidly and cheaply, and it would then represent a heavy expenditure if we decided to add a household interview to the listing operation for the purpose of obtaining (C) and (D). As regards (C), the listing of persons, clearly it would *not* be economic to add this into the listing of dwellings or households: it should go in the second visit, just before the individual interview. As for (D), the large-sample fertility enquiry, the issue is more difficult. It cannot go into the second visit because this is a small sample only. If it goes into the first visit the cost is not marginal; it is more like the cost of a complete additional demographic survey. Some demographers would argue that such large-sample data, of relatively poor quality, are more valuable than better quality small-sample data because the response errors can be corrected analytically (by "Brass techniques"); others argue that the money would be better spent on a larger sample for the individual questionnaire. This issue has to be settled in terms of each country's need for the large-sample fertility data. As we have seen, in most countries the issue will never arise; *either* the listing of households will require *in any case* a large-sample listing of persons so that the cost of the additional fertility questions would be marginal, *or*, at the other extreme (and this includes all developed and many developing countries), there will be no demand for (D) because the country has good civil registration or has recently completed a large-sample demographic survey, or a census, analogous to the WFS household schedule.

The above discussion is somewhat over-simplified in that it is suggested that the listing of dwellings or households is *either* very simple and quick *or* so elaborate that every individual has to be listed. In reality, as we saw in Section 5.4, there are intermediate situations; for example, conditions may be such that the lister has to contact some household member, whether for information or for permission to affix a sticker, but does not need to list every person in the household. Where this is so, there would be a case for putting (C) with (E) and the case against (D) would be intermediate in strength.

One further issue should be examined. In suggesting that (D) might be combined with (B), we have assumed that the listing of households, which covers the whole of each ultimate area unit, would involve a sample of about the same size as the "large sample" required for the fertility data (D). What if these two sample sizes cannot conveniently be brought into agreement?

As we saw in Section 7.1, the household schedule fertility data ideally need a minimum sample of about 20,000 households. Allowing some leeway, we may say that the household schedule sample should be from 3 to 6 times that of the individual interview. In most cases it should be possible to achieve this without deviating too far from the optimal parameters already mentioned, by adjusting either the size of the ultimate area unit or the "take" of households within each such unit, so that the former is from 3 to 6 times the latter. However, if such adjustments become too large we begin to reach, once again, the situation in which the cost of the large-sample fertility enquiry is no longer "marginal" and the issue has to be raised again of whether such an expensive enquiry is justified. If it *is* justified, one has to consider whether it may not be more economical to *sample* the listing for the purpose of the household interview, rather than accept an unnecessarily large household interview sample. This will generally imply a 3-visit arrangement.

The possible arrangements involving not more than two visits can now be listed.

Method	Pre-existing dwelling or household list used for sampling	Field Operations				
		Listing of dwellings or house-holds. Quick (A) or (B)	Listing of households Slow (B)	Listing of persons (C)	Quick information on fertility (D)	Individual interview (E)
(i)	Yes			1		1
(ii)		1		2		2
(iii)			1	1	1[1]	2
(iv)			1	1, and 2	1[1]	2

In the table above, 1 indicates the first visit and 2 the second, the latter nearly always covering

[1] *If desired.*

a sub-sample. The methods should be read *across*. For the purpose of the table, a "slow" method of listing households is taken as one which involves listing the members. The "quick" method is any other: this may be a fresh listing or it may be an up-dating of an existing list. Column (D) refers to the fertility items in the household schedule.

Developed countries can be expected to use methods (i) or (ii). Developing countries would use (ii), (iii) or (iv) – perhaps most commonly (iii) with the quick information on fertility. The distinction between methods (iii) and (iv) will be clarified in Section 7.7.

In methods (i) and (ii), if appropriate, the interviewer conducting the individual interview would correct the listing by using the half-open interval method (see Section 5.3).

In (iii), the document used for "slow" listing of households and for listing of persons will be the same and will be the household schedule as reproduced in the appendix, except that the fertility questions may be dropped if they are not needed.

It may be useful to repeat here in terms of the table an important point made earlier. The "slow" method of listing households is an expensive one. Comparing method (ii) on the one hand with methods (iii) or (iv) on the other, if it is found necessary to introduce the slow listing *because of the need for* (D), then (D) must be regarded as a costly operation; on the other hand if the slow listing is introduced because it is judged to be the only reliable method, then this cost should not be attributed to (D), and the cost of (D) is marginal. This must be considered in reaching the decision as to whether to include (D) or not.

7.7 De facto/de jure and relisting

This sub-section is concerned solely with methods (iii) and (iv). In these methods there are two visits and the first is essentially for listing of households. The reason we gave above for including the listing of persons in this first visit was that in some countries such a listing would be the only effective way of identifying households. This argument still leaves open the question whether we may wish to include a listing of persons *again* at the start of the second visit, for identifying eligibles. At first sight this may appear wasteful, but in certain circumstances it may be desirable.

After the first visit the results for one area will be collected and a sample of households will be selected to which interviewers will be sent for the individual interview. There are three ways in which this link between the two visits may be organized: one is method (iii), the other two are variants of method (iv). They are as follows:

(iii) Select the sample from the eligible households. In this sample, attempt an individual interview with the eligible women reported at the first visit.

(iv) a. Select the sample from the eligible households. Visit these, re-list to identify eligible women and interview these.

(iv) b. Select the sample from all households, whether eligible or not. Visit these, re-list to identify eligible women and interview these.

The lapse of time between the two visits may lead to a substantial loss through mobility and in certain circumstances this could justify the re-listing (or up-dating of the list). Let us look at the coverage loss and additional cost of each method before considering in what circumstances each might be justified.[1]

METHOD (iii)

Target coverage:		Women eligible at visit 1.
Coverage loss:		Omission of eligible women who move away between visits 1 and 2 (some may be traceable if they stay within the area).
Cost:	(1)	Wasted visits to households whose listed eligible women have moved away between visits 1 and 2.
	(2)	Cost of tracing eligible women who have moved away.

METHOD (iv) a.

Target coverage:		Women eligible at visit 2.
Coverage loss:		Omission of households which become eligible between visits 1 and 2.
Cost:	(1)	Wasted visits to households which become ineligible between visits 1 and 2.
	(2)	Interview time spent re-listing persons.

METHOD (iv) b.

Target coverage:		Women eligible at visit 2.
Coverage loss:		Nil.
Cost:	(1)	Wasted visits to ineligible households.
	(2)	Interview time spent re-listing persons.

Comparing (iii) with (iv) a., there is little to choose on coverage; the former offers the chance of tracing some of the movers – at an additional cost, of course. (iv) a. entails the additional cost and trouble of re-listing. A further significant advantage of (iii) is that the selected eligible women are explicitly identified at the sampling stage; this means that there is no need for a good household definition, distinguishing one household from another. As long as the lister covers everyone in his area, the exact way in which he divides the people up into households is of no importance. Thus (iii) should be favoured and (iv) a. is not recommended.

Comparing (iii) with (iv) b., while (iii) has a bias which increases with the time interval between visits (iv) b. has no bias at all. On the other hand, (iv) b. has a constant cost penalty, mainly dependent on the proportion of ineligible households in the population. In developed countries this proportion might exceed 50% and the cost of the resulting wasted visits would be quite inadmissible – but this method would never be used in a developed country in any case. In developing countries we do not know the proportion of ineligible households to expect but it would certainly be very much smaller. (In the WFS Fiji survey it was 10%.[2]) Is this small additional cost justified by the elimination of bias due to loss of eligibles who

[1] *We assume de facto coverage definition, ignoring the small slippage between "last night" and the moment of interview.*

[2] *Criterion of eligibility: ever-married women aged 15–49 present the night before the individual interview.*

move away between visits 1 and 2? Since the latter increases with the time interval there is presumably a break-even time interval in each country, but its length is very difficult to judge. In some countries, it is customary for pregnant women to go to their mother's house for the period around their confinement: this would introduce an unusually specific form of bias.[1] Where this practice is known to be common we suggest that the average time interval should not be allowed to run much beyond one week in method (iii); if the average is expected to be much longer than this then re-listing should be used (method (iv) b.). In countries or regions where women do *not* habitually move away to have their baby, it is suggested that a time interval of 3 weeks would be acceptable without re-listing. In practice these suggestions might amount to the following simple principle:

> If the household listing, household sampling and individual interviews are carried out in an unbroken sequence while the team is in the area, use method (iii). If the household lists are returned to headquarters (or regional headquarters) for sampling, use (iv) b.

Of course this principle relates only to the choice between methods (iii) and (iv); it is not relevant where methods (i) or (ii) are available.

If method (iii) is chosen, we have the option of following up absent eligibles to their new location. It is suggested that this should be done only where this location is *known* and *nearby*.

Finally, it may be noted that, in case (iii), eligibility is defined at the time of the first visit and it is not certain that a *de facto* coverage, based on this moment, will lead to fewer failures at the time of the individual interview than a *de jure* coverage. The issue depends on the definition of household membership and on the length of the time interval between the visits: the longer the interval the more favourable would be the *de jure* coverage compared with *de facto*. Once again, the break-even point is not easy to estimate. In this manual we have recommended *de facto* coverage for the individual interview because we have assumed that this break-even interval will be longer than the break-even interval already discussed, beyond which method (iii) would not be used. Nevertheless, this assumption may well be wrong and some countries may prefer to try *de jure* coverage in method (iii). Note that for the first visit (household schedule) we have in any case recommended coverage of *both* the *de facto and* the *de jure* populations. This offers a third alternative for method (iii): we could cover both populations also in the individual interview and then select for analysis the one which yields a higher coverage. However, this would mean discarding interviews already obtained.

One further alternative may be mentioned. In Section 4.4 we considered the possibility that some countries may prefer not to interview every eligible woman in each household but, after the household listing, to sample *women* instead of *households*. If systematic sampling (see Section 8.4) is used at this stage, this will almost always yield one interview per household visited. This method may be appropriate where there are known to be households containing considerable numbers of eligible women. The best way of arranging it would be to modify

[1] *If the average time a woman spends away from home when having a baby is 10 weeks for example, then a 2-week interval between the two interviews would mean that the sample would miss 1/5th of all women around the period of their confinement.*

method (iii). The first visit would identify *de facto* (possibly *de jure*?) eligible women and a sample of these would be selected for individual interview. The pros and cons of this were discussed in Section 4.4. The coverage loss and cost penalty are approximately the same as those given for method (iii) above except for an additional cost due to the increased number of households to be visited to obtain a given number of interviews. (There may also be a small effect on the sampling error for a given number of interviews. For some variables sampling error is likely to be reduced, for others perhaps increased.)

Finally, as already mentioned, it may happen that a pre-existing list, or a quick listing, is available but that we need the large-sample fertility data; to obtain the latter we would have to reject (i) or (ii) in favour of (iii) or (iv).[1] This would imply a substantial additional cost and one would want to examine carefully whether the need for the large-sample fertility data justifies the cost penalty.

7.8 Additional uses of the household schedule

In addition to the uses already mentioned, the household schedule serves several other purposes.

Most important, it supplies the base data about the population needed for computation of population rates. Here the items on age, education and affiliation are particularly important. Whatever sampling method is used, these should be retained for the *whole* population and not limited to households containing eligible women. While such information is often available from censuses, the household schedules makes it possible to ensure the strict comparability, as to time reference and coverage, of these data with the data from the individual questionnaire.

Secondly, it offers the possibility of obtaining some additional information about the household (see the front page of the household schedule).

Finally, the possibility of using the household schedule to obtain some additional information on borderline cases of special interest has already been mentioned in Section 3.1. Specifically, we recommend three such applications:

1. The household schedule should cover both the *de jure* and the *de facto* population. This will yield some information on absent members of the household. Comparison of this group with the "visitors" (theoretically the same group, if absence abroad is negligible[2]) should yield some information on possible coverage error due to confusion about residential status.
2. The fertility items in the household schedule should cover women of all ages. This should throw some light on any age bias, or age-boundary error, in the individual questionnaire.
3. The fertility items in the household schedule should, if feasible, cover all women above the minimum age, irrespective of marital status. This may throw light on marital status bias in fertility reporting and, specifically, on births to never-married women.

[1] *In certain cases one might prefer a composite: select a dwelling sample as in (i) then, within these dwellings, proceed as in (iii) or (iv). This would meet the possible difficulty that the number of households in the selected ultimate area units might be larger than the "large sample" needed for the quick fertility data.*

[2] *Where it is believed not to be negligible, the absence question (col. 4) should be supplemented to identify those absent abroad.*

Note that none of these three considerations is of crucial importance. They would not in themselves justify the introduction of a household schedule, nor even perhaps the introduction of the fertility items into a household schedule already adopted for other reasons. But they are likely to be useful devices where the appropriate schedule has already been decided upon.

7.9 Link between listing and interview: summary

Countries are recommended to use one of the following methods:

(i) If a pre-existing dwelling list or household list exists which is adequate for sampling, and if large-sample quick information on fertility is not needed, use this sampling frame. Interviewers should then visit the selected dwellings or households, using the half-open interval correction if appropriate, list their members (including visitors present "last night") and interview eligible women among the *de facto* population so listed. For this listing of persons use the household schedule (omitting the fertility questions).

(ii) If a quick but reliable dwelling or household listing can be done in the field, and if large-sample quick information on fertility is not needed, use this listing as the first step, then proceed as in (i) above.

(iii) If the conditions for (i) or (ii) do not apply, and if it is practicable to conduct the individual interviews shortly after the listing, then use the household schedule for listing households, with or without the fertility items according to need. Identify eligible women and select a sample of households from those containing one or more eligible women, preferably counting *de facto* only[1], possibly *de jure* only. Send interviewers to these households with the names of the women to be interviewed. Follow up any who have moved, provided their new location is both known and nearby.

(iv) If the conditions for (i) or (ii) do not apply, and if the interval between listing of households and the individual interview cannot be kept short, use the household schedule for listing households, with or without the fertility items according to need. Select a sample of households *without* regard to eligibility. Send interviewers to the selected households with instructions to up-date the list of persons and to identify and interview eligible women (*de facto* only[2]).

Exceptionally, a modification of (iii) would be acceptable. Instead of sampling households from those containing one or more eligible women, sample women directly from those listed as eligible – once again, preferably *de facto*[1], possibly *de jure*.

One other exceptional option is described in footnote 1 at the end of Section 7.7.

[1] *Reference date: the night before the first successful call for the household interview.*

[2] *Reference date: the night before the first successful call for this up-dating visit (which is normally also the individual interview visit).*

8. Stratification

8.1 Definition and purpose

Stratification means dividing up the population into sub-groups or "strata" and selecting the sample separately in each stratum. This guarantees that each stratum receives the sample planned for it: without stratification the overall sample is planned but the way the sample distributes between sub-groups is left to chance. Strata are very often *regions*, but this need not be so. They might be, for example, size groups of localities, size groups of households, income groups, ethnic categories, etc.

Stratification has two distinct purposes. Firstly, it allows one to sample different sections of the population with differing sampling fractions. Secondly, whether or not unequal sampling fractions are used, stratification will usually reduce sampling error. These two purposes are discussed in turn in the present section.

Stratification may also arise naturally, from the practical necessity, which one sometimes finds, of sampling different kinds of areas in different ways, or of drawing different parts of the sample separately when the relevant sampling frames are located in different places.

It is important to distinguish *domains of study* from *strata*. A study domain is a population sub-group for which it is desired to report separate data, for example, "the urban population", or "Moslems", or "women aged 15–19". A stratum is a sub-group in which the sample is selected separately. Strata *may* be study domains, but they need not be. However, as far as possible strata should not be designed to cut across domains – that is, a domain should consist of a stratum or a group of strata. But it is virtually only in the case of geographical domains that we have any choice about this.

8.2 Stratification with unequal sampling fractions in the strata

Deliberate introduction of unequal sampling fractions in different strata may have one of two motives.

1. It is shown in sampling textbooks that optimum sampling design involves a *higher* sampling density in strata where the variance is greater[1] and a *lower* sampling density where unit costs are greater. Thus sampling fractions may be manipulated in order to reach an optimum design.
2. The survey planner may wish to report findings for a group of people – a domain of study – which is only a small percentage of the whole population. If a fixed sampling fraction is used, this small group will get only a small sample and, as a result, the sampling error for this domain may be unacceptably high. We can reduce the sampling error if we can make the special domain into a stratum and use a specially high sampling fraction for this stratum. To avoid bias in the overall estimate for the whole

[1] *This refers to the actual variability among units of the population and not to the error variance.*

population, we have to introduce a weighting at the data-processing stage which downgrades the special domain to the same degree as it was upgraded at the sampling stage. The total estimate suffers some loss of sampling precision (computed for a given total sample size) as a result, but this is the price we have to pay for getting a better estimate for the special domain.

The first of these techniques – varying the stratum sampling fractions to maximize cost efficiency – is of limited importance in the WFS. One reason is that the survey involves many different variables and an increase in the sampling efficiency for one variable may mean a decrease for another. Moreover, in most cases costs and variances do not vary much between strata and, even where they do, very little information about such variation is available in advance of the survey so that big improvements in sampling efficiency are unlikely. Finally, variation in sampling fractions involves some additional complexity in data processing. For all these reasons the strategy of optimal sampling in the strata is unlikely to prove worthwhile in the context of the WFS. A possible exception might be a strategy of raising the sampling fraction in urban areas, where costs per interview tend to be lower, but even here, any gain in sampling efficiency is likely to be small.

The second technique – oversampling a small domain of study to give it a more substantial sample and hence reduce its sampling error – may sometimes be of value in the WFS but some important reservations should be noted. First, as we have already mentioned, any gain of this kind means a loss somewhere else. Many of the survey estimates are going to concern the whole sample and all of these will suffer in sampling efficiency if the special domain is oversampled (except in the unlikely case where the previous argument and the present one lead to similar choices of unequal sampling fractions). There is also a data-processing problem. If the oversampled domain is a geographical region there is no great difficulty, but if the domain is defined at a level below the first stage of sampling, for example, "small villages" or "Moslem households" or "widows", there will be complications at the data-processing stage. Thus suppose, for example, that one wishes to oversample an ethnic or religious group. Processing will be simpler if one can achieve this by identifying as a stratum a particular zone or zones where such people predominate and then oversampling in this stratum, rather than introducing a special sampling fraction for such groups at the household sampling stage. Finally, it should be frankly faced that a *small* group, whether ethnic, religious or regional, is, on a straightforward human basis, of correspondingly *small* importance: any argument that claims to justify achieving equal precision for all groups whether small or large should be scrutinized with considerable scepticism. Once again, the most tempting case for use of a special sampling fraction is that of the urban population; in many countries this includes the only substantial sub-group using modern means of contraception and this may be offered as an argument for increasing the sampling fraction in the urban sector. But on the other side, it can be argued that we are primarily interested in what *most* people are doing, not what any special group is doing. Special groups should be investigated by special studies. Moreover, rural populations, no matter how remote, are not – anywhere – reproducing at a biological maximal rate: somehow their fertility is being restrained. It is important to know why, and how. In summary, it is suggested that survey planning for the WFS should start from the position that the sampling fractions will be equal in all strata. Any departure from this position in a particular case would need very careful justification, though it would not necessarily be excluded.

8.3 Stratification with equal sampling fractions in the strata

The purpose of such stratification (also called "proportional allocation") is to reduce sampling error. The technique achieves this, essentially, by ensuring a good spread in the sample: we are *certain* to get an appropriate ration in each stratum whereas without stratification the whole sample might, by bad luck, be seriously distorted. The effectiveness of stratification depends on the strength of the correlation between the criterion used in stratifying and the survey variables.

Stratification nearly always improves sampling efficiency, (i.e., reduces sampling error for a given sample size), if only marginally, and generally the more strata the greater the gain: right down to the extreme case in which only one unit is selected in each stratum.[1]

However, the gains from stratification with equal sampling fractions are usually quite modest. Moreover, each additional stratum based on a given stratifying variable gives a smaller advantage than the last, that is, two strata give some gain, three give a little more, four give hardly any more, and so on. Rather than bring in more strata on the same variable, it is more profitable to switch to a new stratifying variable. For example, suppose we use census data to stratify localities and suppose we create strata based on "per cent employed in agriculture" (a rough measure of rural/urban character). We might put into Stratum 1 all localities with 90% or more employed in agriculture, into Stratum 2 those with 50% to less than 90%, and into Stratum 3 those with less than 50% (the "urban" stratum). Now it would be possible to create a 4th stratum on the same basis, say "below 25%", leaving only 25–50% for stratum 3. But it would be more likely to yield gains if, instead, we switched to a new basis. For example, we might look at the three main regions of the country, say North, Centre and South. Cross-stratifying on this variable and the last one simultaneously, we would have up to 9 strata, from

90% or more in agriculture; North

to

less than 50% in agriculture; South.

These can be treated just like 9 separate strata. If any of them is too small, we can group it with another. For example, there may be very few towns in the North and we might therefore decide to group together *Stratum 2 – North* and *Stratum 3 – North*, reducing the total number of strata from 9 to 8.

[1] *Strictly speaking this is not the extreme case. Kish's method of controlled selection uses less than one unit per stratum. This will not be discussed here.* (KISH, L., 1965, *Survey Sampling*, SECTION 12.8, WILEY, NEW YORK.)

The same principle can be continued, using yet another criterion of stratification, crossed with the 8 or 9 strata already created.

How many strata should be made in all? Suppose we have already decided to select a total sample of n primary (first-stage) sampling units (PSUs). If we made n strata we could select 1 PSU in each. Or we might create ½n strata and select 2 PSUs in each, which would be slightly less effective in reducing sampling error. However, unless we select *at least 2* PSUs in each stratum we cannot make an unbiased estimate of the sampling error. If we select only 1, then we have to group strata together ("collapsing" strata) before computing the sampling error, and this will result in an overestimate of the error. It is in fact common to select exactly two units per stratum, though this is certainly not essential. (The sampling error estimation formula takes a specially simple form in the case of two selections per stratum.) This implies creating just half as many strata as the number of PSUs it is intended to select in the whole sample.

If a fixed number of area units is to be selected from every stratum, it will be desirable to create the strata in such a way that they all contain roughly equal numbers of units; otherwise, the larger strata will inevitably get a smaller sampling fraction, which will not normally be an optimal design. (This situation will be modified by multi-stage sampling but the general principle remains broadly true.) We have already seen that excessively small strata can be grouped together to make larger ones; similarly, excessively large ones can be split by introducing a new stratifying variable. For example, if the urban sector in the South is excessively large because it includes the national capital, one could divide this South-urban stratum into East and West, or perhaps into the capital city versus the rest. Thus the pattern of stratification can be completely flexible: a new stratifying variable (such as East-West) can be introduced in one or two isolated strata without necessarily using it in all the strata. In this way, one can generally arrive at approximately equal sized strata.

Summing up: stratification with equal sampling fractions in the strata brings only modest gains in reducing sampling error. The more strata created on any one stratifying variable, the greater should be the reduction in sampling error; but the gains fall off rapidly after the first two or three strata. Few strata on each of several stratifying variables is better than many strata on one variable. Unless we select at least 2 PSUs per stratum we cannot get an unbiased estimate of the sampling error. Most commonly, sample designs provide for 2 selections per stratum but this is not essential. If a fixed number of units are selected in every stratum, one will normally want the strata to be approximately equal in size. Within this limitation, the structure of the stratification and cross-stratification can be as complex or arbitrary as one likes.

8.4 Replicated sampling

Replicated, or interpenetrating, samples are samples so selected that they divide naturally into two or more subsamples each of which represents the whole population. For example, if we select two area units A and B in each stratum, then the set of selected As is a sample

of the whole population and so is the set of selected Bs.
There are several advantages in this procedure, particularly where the number of subsamples is just two. Firstly, formulae for the sampling error are simplified.[1] Secondly, it often pays to arrange to complete all the field work for one of the subsamples before starting on the other; then if a hitch occurs during the second half of the field work period (whether a hurricane, a revolution or a shortage of funds) one can still salvage something useful. Thirdly, one can vary the survey technique, or personnel, randomly, using the replicated sample as a factorial design, in order to experiment with the methodology (details will not be given in this manual). Finally, it is sometimes argued that the two samples of data should be not only collected separately but also processed separately, and then published side by side, as a demonstration of sampling reliability which is more convincing to the layman than the citing of sampling errors.

Any of these effects can be achieved with the kind of design already suggested, with two PSUs per stratum, by randomly allocating one PSU of each pair in each stratum to the first subsample and the other to the second. Note, however, that we cannot have all these advantages at once. If we want to compare two survey techniques, A and B, we will need 2 PSUs per stratum for A and 2 for B in order to estimate the within-stratum variance, thus making 4 per stratum in all. If we complete one subsample "early" and one "late" and yet have only 2 PSUs per stratum, the variance estimate will be based on the "early minus late" difference in each stratum so that we will have no way of excluding the time interval effect from the error variance.

8.5 Systematic sampling

Systematic sampling means selection from a list at a fixed interval, from a randomly selected starting point. This spreads the sample evenly through the list. If the list is in some significant order, which many existing lists are, this technique has a similar effect to stratified sampling with equal sampling fractions in the strata: the stratifying variable being now the variable in terms of which the list is ordered. Systematic sampling also has the advantage that we do not have to bother with defining the boundaries of the strata. Suppose, for example, that we wish to stratify according to "urbanism", that is, the variable defining the urban/rural continuum. A possible measure of this would be the census "size of locality". An ideal solution would be to list localities in decreasing size order and sample systematically. This gives excellent urban/rural stratification without the need to define the borderline between urban and rural. This advantage of systematic sampling – that it does not require *explicit* strata with defined boundaries – has one limitation: it cannot apply to more than one of the stratifying variables at a time. If two or more stratifying variables are crossed, all but one of them must be explicit strata.

One disadvantage of systematic selection is that it does not allow strict computation of the error variance. We can get around this by making assumptions about the relation between the study variable and the variable in terms of which the list is ordered, but a more usual

[1] *But it is important to base these on the set of PSU pairs in the set of all strata, summing the individual stratum variance estimates. An estimate based on the single difference between the two subsamples after aggregation to the national level would be subject to a very large sampling error and should never be considered.*

approach is simply to group the sample in pairs, according to the order of selection from the list, and regard each pair as coming from one stratum. This will slightly overestimate the variance.

If more than 2 units are to be selected from any larger unit or stratum, systematic selection is nearly always used: it is both more efficient and simpler than random sampling. Systematic sampling is also likely to be used in selecting dwellings or households at the last stage of sampling.

9. Varying probability sampling and multi-stage sampling

9.1 Sampling with probability proportional to size

A common sampling plan is to select the area units with a probability proportional to the estimated population of each unit. Thus, if unit A is estimated to be 10 times as large as unit B, it is given 10 times as many chances of being selected. Of course this gives a sample biased in favour of large units, but this is corrected later. This method is called "sampling with probability proportional to size", or "PPS sampling".

One way of correcting the bias is to use the exactly opposite system at the household sampling stage, i.e., sampling with probability *inversely* proportional to the measure of size used at the area sampling stage. This means that the sampling fraction *for households* in area unit A will be 10 times *smaller* than in unit B, thus cancelling the bias. A given household in unit A now has exactly the same probability of selection as a household in unit B. Since no weighting is now necessary to remove the bias, this is called a self-weighting sample. The great advantage of a self-weighting sample is that, for purposes of computing means, percentages and rates, we can treat the sample data just as if they came from a census.

With this sample design, if the estimate of size used at the first stage were always exactly equal, (or proportional), to the number of households in the unit, then it is easily seen that at the second stage one would be selecting a *fixed number of households* in every selected unit.[1] In practice, the estimate of size is inaccurate, to varying degrees in different situations. To emphasize this, one sometimes refers to PPES sampling: probability proportional to *estimated* size. If the above sampling plan is followed out with PPES sampling one gets a household sample in each area which is approximately constant.

It is the approximate constancy of the work load in each area that constitutes the main attraction of PPES self-weighting sampling, together with the self-weighting property itself. The field work is a good deal easier to organize if we do not have very large work loads in some areas and very small ones in others. This advantage is particularly significant where we are sampling ordinary administrative units, which commonly vary widely in population size. However, if the area units used for sampling are fairly constant in size (as is often the case with census enumeration areas) the gain in this respect will be relatively small and then fixed probability sampling may be preferred as simpler. This is a matter for judgement in relation to the situation in a given country.

It sometimes happens that the area units are reasonably constant in size except for a few outliers: a small number of very large or very small units. In these circumstances, the large ones can be dealt with by splitting and the small ones by combining, in such a way as to yield a more nearly constant size distribution; and this will be particularly worth doing if

[1] *In a unit 10 times as large, one would be selecting a fraction 10 times as small, i.e., the same actual number. More generally: if n households are selected from a total of N and the sampling fraction $f = n/N$ varies inversely with N, then $f = k/N$ where k is constant. Therefore $n = k =$ constant.*

we are opting for fixed probability sampling. Note that the splitting need not be done before sampling: that is, we need not decide until the particular unit is selected exactly how the split is to be made, though we must decide in advance *which* units are to be split and *into how many sub-units*. For example, if we know before selecting the sample that unit No. 73 in the frame is 4 times as large as we would wish, we list it as 73.1, 73.2, 73.3 and 73.4, then proceed with the sampling as if these were 4 units. If any one of them is selected, we will have to map out the exact split, then select 1 of the 4 sub-units; typically, this latter selection would be made at random if the sample is being selected with equal probability, otherwise with PPS. As for the combining of units that are too small, we must decide in advance exactly which groupings to make and revise the list before sampling so that each group appears as one unit.[1] Of course field work is more convenient if the units combined together are contiguous, but in case no maps are available this is not absolutely essential.

A further advantage of PPS (or PPES) sampling is that, in general, it greatly reduces the sampling error for estimates of totals. It will also, in general, improve sampling efficiency for means, rates, ratios or percentages.

A common error in PPS designs may be worth describing here. Arrangements are sometimes made to sample area units with probability proportional to census population and then to select a fixed *number*, as opposed to a predetermined *proportion*, of households (or women) in each selected area unit. This does not give a self-weighting sample because the census population is not exactly proportional to the number of households in the area, for at least 4 reasons:

1. The census is out of date, and attempts to up-date by extrapolation can never be reliable at the level of small areas.
2. The census may have been erroneous in some areas.
3. The census may have used a different residence criterion from the survey.
4. Households vary in size, so that population may not be proportional to the number of households. This source of error might be eliminated if the census count of households is available for PPS selection, provided that the definitions of a household are the same in the census and the survey listing.

Thus, if this sample is treated as self-weighting, there will be a bias.[2]

We return to the problem of PPES sampling at the end of this section, where we discuss the question of organizing a special field operation designed to estimate "size" before sampling.

9.2 Multi-stage sampling

Up to now, we have described all the procedures as if there were only one stage of area sampling, followed by one stage of household sampling. This assumption has been made only to simplify the description; in practice the area sampling may often be organized in two stages, possibly even three. A decision on the choice of the number of area sampling stages is difficult: it involves a number of factors on which precise data are almost certainly lacking

[1] *Strictly, we need not decide in advance as long as we decide independently of the actual selection.*

[2] *It is possible to compute the correct weighting for each area unit, and thus remove the bias, provided that the number of households in each selected unit has been counted at the listing stage, but in some surveys no such count has been made and there is then no way of eliminating the bias.*

in any given case, and account has to be taken of various characteristics of the country and the way in which the survey is organized. This is an issue on which both local knowledge and wide experience are likely to play a role.

There are two ways in which the introduction of an additional area sampling stage may save costs. Firstly, it may reduce the amount of field work (or, occasionally, office work) involved in preparing sampling frames, that is, creation of sub-area units, listing operations, and field or office operations to obtain measures of size for PPS selection. Secondly, it may have the effect of clustering the area units in which the interviewers are working, with a possible reduction in travel and easing of the work of supervision. We consider these advantages in turn.

9.2.1 *Reduction in field work for creation of sampling frames*

In many surveys the sampling frame of ultimate area units (UAUs) is not given but has to be created by a mapping operation. The cost of such work carried out all over the country would be prohibitive. In such circumstances, it is usual to select a first stage consisting of a sample of large area units for which maps *are* available and to confine the task of mapping smaller units to the selected sample of large units.

Such a scheme may be extended to yet another stage. For example, in the National Fertility Survey of Peru (1969) a first stage rural sample of *districts* was selected. The selected districts were then visited and mapped and each was divided into a number of *zones*, from which a sample of zones was selected. Similarly, the selected zones were mapped and divided into *segments*, within which a sample of households was selected. The segment was the UAU. The purpose of this multi-stage design was essentially to reduce the amount of map-making and household listing.

A decision has to be made on the number of stages to use and the size of the units. This depends partly on the cost of the map-making operations. It is difficult to make any useful generalization about this. However, the options can first be usefully narrowed down by the following type of argument. We referred in Section 6 to the "basic area units" (BAUs): these are the smallest units for which we have a satisfactory sampling frame (districts, in the Peru survey). Any effort to reduce mapping must come in *below* this level in the sampling hierarchy since maps are already available for the BAUs and any higher stages. We have also referred several times to the "ultimate area units" (UAUs): these are the smallest type of area units to be used for sampling in the survey concerned (segments, in the Peru survey). As we saw in Section 5, there are a number of constraints affecting the size of UAUs in the WFS and in many cases we may regard the UAU-size as virtually fixed in advance by these constraints. Thus, in WFS surveys, it would be a reasonable strategy to start by fixing the BAU-size and the UAU-size and to build the sample design around these. If we now compare the size of the BAUs with that of the UAUs, we may ask whether it would be worth the trouble to bridge the gap by introducing a further stage of sampling between the two. Any new sampling stage, though it may save mapping effort, increases the complexity of the design, implying more complex instructions to the sampling field workers. More seriously, it is likely in practice to increase the sampling error because it will introduce a less accurate size estimate into the estimation process (this is discussed further in Section 9.3). For these reasons, most sampling specialists hesitate to introduce an additional stage of area sampling unless each unit of the higher stage contains at least 5 – many would say 10 – of the lower stage. Given the size of the BAU and the UAU, this rule of thumb enables us to judge whether an intermediate stage of sampling would be worth considering. For example, if the UAUs average 200 households and the BAUs 2500 there is, on this rule, no room for an intermediate stage. Only if the BAU reaches about 5000 households or more does the introduction of this intermediate stage begin to look worth contemplating as a means of saving work on mapping.

9.2.2 *Clustering the ultimate area units*

Higher stages of sampling above the UAU may be introduced not merely to reduce mapping costs but to group the selected UAUs into clusters, thus reducing travel and easing supervision. However, the apparent gain here should be examined very carefully: in many cases it will be found that the saving is zero or negligible and in any case it has to be set off against the increased sampling error which results almost inevitably from any clustering of the sample.
A useful approach to this problem is to start at the bottom – the UAU – and to see how scattered the rural sample would look if the selected UAUs were spread evenly over the whole country. The number of UAUs in the sample is known in advance since we already have the total sample size and the average size of the household sample which it is intended to select per UAU. For example, if the total rural sample is 3000 households (for individual interview) and the sample per UAU is to be 30 households, then there will be 100 rural UAUs in the sample. If 100 UAUs were scattered evenly over a country of size 1 million sq. kms., the distance between nearest neighbour UAUs would be 100 kms. Of course, in practice, they would not be scattered evenly but distributed more or less as the population and this would greatly reduce the amount of between-unit travel, perhaps by 50 per cent or more. (An idea of the actual distance to be expected can sometimes be obtained by looking at the distribution of field work locations in an earlier survey in the country.) A rough study of the problem in these terms can show how much the average distance travelled by interviewers between UAUs will be reduced by any particular clustering plan. If this saving is costed it will usually be found negligible except in large countries. As for supervision, if the supervisor moves with the interviewing team, so that the whole team deals with one UAU at a time, this presents no special problem. Even the potential saving in travel for higher level supervisors visiting one team after another is unlikely to materialize because, with clustering of UAUs in higher units, one would still be unlikely to place two teams in the same higher unit.
Thus, in practice, in most countries the introduction of a second or third stage of area sampling could only be justified in terms of the saving in mapping or listing, or other preparatory work on sampling frames. Whether it is so justified will depend much on an assessment of the adequacy of existing area sampling frames. If a satisfactory frame exists of units which are

small enough to serve as ultimate area units (for example, census enumeration areas), then a single stage of area sampling may often be enough.

9.3 Multi-stage sampling with PPS or PPES

The principle of PPS sampling with self-weighting can be readily applied to the case of a multi-stage sample. In practical applications it is necessary to work out the actual selection probabilities at each stage. The following treatment therefore inevitably uses some mathematical notation.

If unit i is of (estimated) size A_i, and if we wish to select it with probability proportional to size, then the selection probability p_i must take the form

$$p_i = k\, A_i .$$

The constant k is easily seen to be equal to $m/\Sigma A_i$, where m is the number of units to be selected and ΣA_i is the sum of all the "sizes" in the population concerned. The formula

$$p_i = \frac{m\, A_i}{\Sigma A_i} \qquad (1)$$

is crucial in practical applications.

A common way of PPS sampling is to write down the size A_i against each unit, then cumulate these values, then select at a fixed interval in the cumulative column (systematic sampling). If this is done, the sampling interval I will be equal to $\Sigma A_i/m$, so that formula (1) may be written

$$p_i = \frac{A_i}{I} \qquad (1')$$

Now consider two stages, indicated by the suffixes 1 and 2. At the second stage, it is important to distinguish the *conditional probability* from the *overall probability*. Thus, if village j (second stage) is contained in district i (first stage), the "conditional probability" of selecting this village means the probability of doing so *once district i has been selected.* The overall probability of selecting village j is the product of this with the probability of selecting district i. We shall denote the conditional probability by p_{2ij}. The overall probability for village j in district i will then be:

$$P_{ij} = p_{1i}\, p_{2ij} \qquad (2)$$

Similarly, for 3 stages the overall probability of selecting the final stage unit ijk will be:

$$P_{ijk} = p_{1i}\, p_{2ij}\, p_{3ijk} \qquad (3)$$

Now the essential requirement is self-weighting, which means that the final overall probability has to be constant throughout the whole sample. This constant is known in advance: it is the target overall sampling fraction, f. For example, if we wish to sample 5000 eligible women in all and if we estimate the number of such women in the whole population at 5,000,000, then $f = 1/1000$. Thus, we want (3-stage case):

$$P_{ijk} = f \tag{4}$$

This constraint, together with other parameters, determines the selection procedure in each stratum, in the following manner. P_{ijk}, as shown in (3), is the product of the selection probabilities used all the way down the sampling family-tree along the pathway leading to unit ijk. The first of these is the primary stage selection probability, p_{1j}, given by formula (1). Once we have decided on m, the number of selections to be made in the stratum, and when we have obtained all the estimated PSU-sizes (that is, the A_i) in the stratum, then all the elements in (1) are known. Passing to the second sampling stage, which takes place within each selected PSU i, the conditional selection probability for PPS selection is given by a formula analogous to (1), namely:

$$p_{2ij} = \frac{m_i A_{ij}}{\sum_j A_{ij}} \tag{5}$$

where

m_i is the number of second stage units (SSUs) to be selected in the i-th PSU;
A_{ij} is the (estimated) size of the ij-th SSU;
$\sum_j A_{ij}$ is the sum of the sizes of all the SSUs in the i-th PSU.

Once again, when we have decided on m_i, the number of SSUs to be selected in the i-th PSU (typically, this would be a fixed number, such as 2, in every PSU), and when we have estimated the sizes of all the SSUs in the i-th PSU, all the elements in (5) are known. Finally, as to the third stage probability p_{3ijk}, this now has to be chosen to satisfy the constraint (4), in order to obtain self-weighting. At the third stage, we will be selecting some number m_{ij} of units (households) from a total of, say, M_{ij}, and this means a selection probability of m_{ij}/M_{ij}. (We assume that all households in a given SSU are to be allotted an equal chance of selection; thus suffix k is not brought in.) Note that M_{ij} is the actual number of households found during the listing, not on estimate. Hence, to satisfy (4), we must have:

$$\frac{m A_i}{\sum_i A_i} \cdot \frac{m_i A_{ij}}{\sum_j A_{ij}} \cdot \frac{m_{ij}}{M_{ij}} = f \tag{6}$$

This can be solved for m_{ij}, telling us how many households to select in the ij-th SSU to ensure self-weighting.
Note also the following:

(i) If the measures of size are consistent, then $A_i = \sum A_{ij}$ (in other words, the "size" of a PSU is the sum of the sizes of the SSUs which make it up). In this case, the A_i in the numerator of the first term in (6) will cancel the ΣA_{ij} in the denominator of the second.

(ii) If the measure of size for SSUs is the *number of households* in the SSU, then $A_{ij} = M_{ij}$ and the A_{ij} in the numerator of the second term in (6) will cancel the M_{ij} in the denominator of the third term.

Hence, if we use throughout, as the measure of size, the *estimated number of households contained*, then in so far as this size estimate is correct we will get:

$$\frac{m}{\Sigma A_i} \cdot m_i \cdot m_{ij} = f$$

or (7)

$$m_{ij} = \frac{f \Sigma A_i}{m\, m_i}$$

If we make m_i fixed for all PSUs (i.e., if we always select a fixed number of SSUs per PSU) then the expression on the right of (7) is a constant for all i, j, k. Thus, this procedure gives us *a constant workload in every ultimate area unit*, in so far as the estimated sizes correctly represent the number of households contained. If the size estimates are inaccurate, this workload will be found to vary but the self-weighting property will still be strictly satisfied, provided we compute m_{ij} from (6) and not (7). Note that self-weighting does *not* depend on the consistency of the size estimates: if the sum of the SSU sizes in a given PSU does not equal the size assumed for that PSU in selecting it (i.e., if $\Sigma A_{ij} \neq A_i$), self-weighting still applies. Thus it is fully permissible to use different methods for estimating PSU sizes and SSU sizes.

Summing up: for a 3-stage self-weighting sample with PPES at the first two stages, a recommended procedure would be as follows:

(i) Compute the target overall sampling fraction f.

(ii) Decide on the number m of PSUs to be selected per stratum and the number of SSUs to be selected per PSU. (The latter number appears as m_i in formula (6) above; however we now wish to fix this as a constant, say $m^{\dagger}$.)

(iii) Select m PSUs in each stratum with probability proportional to estimated number of households A_i. Record the selection probability p_{1i} for each such PSU selected:

$$p_{1i} = \frac{m A_i}{\Sigma A_i}$$

the summation being taken over all PSUs in the stratum.

(iv) Select $m^{\dagger}$ SSUs in each selected PSU, with probability proportional to estimated number of households A_{ij}. Record the conditional selection probability p_{2ij} for each SSU selected:

$$p_{2ij} = \frac{m^{I} A_{ij}}{\sum_{j} A_{ij}}$$

the summation being taken over all SSUs in the i-th PSU.

(v) For each selected SSU ij, compute the third stage conditional selection probability for self-weighting:

$$p_{3ij} = f / (p_{1i}\, p_{2ij})$$

(vi) After listing the households in the ij-th SSU, apply this selection probability to draw the sample of households (use systematic sampling with interval $1/p_{3ij}$).

The modifications which would be required for 2-stage or 4-stage sampling are obvious.

The above procedure is most likely to be used where there is no adequate sampling frame of suitably small area units and where, in consequence, the sampling frame has to be created stage by stage "as we go". We have already described the Peru survey which made use of this approach. First, a sample of rather large PSUs is selected. A team visits a selected PSU, maps it and divides it into SSUs, selects a predetermined number of these with probability proportional to estimated size, maps these, and so on until, ultimately, suitably small area units are reached – the UAUs. This process requires estimates of "size" to be made for the various units. At the first stage the size estimate may come from the census. At subsequent sampling stages we may use "quick counts". The method of counting may depend on the size of the units concerned. For small areas a house count can be used. If we are trying to cover a large area then the cost of a house count (even a rough one) might be excessive. We might then use estimates obtained from local officials, or else, while touring the area for mapping purposes, we might simply classify villages as containing approximately 500, 1000, 1500 or 2000 households, on the basis of rough eye estimates where the village is seen, or local enquiries where it is not.

Note that we only need to make size estimates for those units which lie in the larger units *selected* during the previous stage. Note also that, with PPES sampling, we can start afresh with our size estimates at each stage: there is no need to make them consistent from stage to stage. Moreover, we could use population at one stage and households at another: this would not affect the self-weighting, though if households varied systematically in size in different areas, it would cause some variation in interviewing workloads among different UAUs.

9.4 Self-weighting and the household schedule

A self-weighting sample for the individual interview is highly desirable for the sake of smooth data-processing. Despite advances in computer hardware and software, the experience of recent years has in no way changed this situation. If complete self-weighting is abandoned there should be, at most, 2 or 3 different weights only, applicable to large, geographically defined strata.

The method of sample selection described in the last section assures self-weighting for the individual interview sample. However, in Section 7 we saw that in many countries the household schedule will be producing substantive data on a much larger sample, namely that covered by the listing of households in UAUs. In the sample design we have described, this latter sample would not normally be self-weighting. Does this matter, and if so can it be avoided? The data collected on the household schedule are so much less extensive than those of the individual interview that the self-weighting requirement must be regarded as much less pressing – a convenience but certainly not a necessity.

We could only achieve self-weighting on both samples simultaneously if the final stage sampling fraction were constant in all UAUs. In the description in Section 9.3, we would have to stop short one stage earlier, achieving self-weighting for the sample of UAUs, and then select a fixed fraction of households in each selected UAU. The workloads in different UAUs would then be subject to the same relative variation as the UAU-sizes themselves. It would thus be a matter of some concern to have the UAUs approximately constant in size. If the UAUs are "home-made" units, as in the method described at the end of Section 9.3, it should be possible to achieve this. If they are pre-existing units, such as census enumeration areas, they are likely to be somewhat more variable in size than one would wish. In some cases this problem might be met by grouping small units and splitting large ones to reduce their range of variation (see Section 9.1). If this is not practicable, one could consider the following alternative, which makes a limited sacrifice of the self-weighting feature. Divide the UAUs into two strata, "larger than average" and "smaller than average" before selection. Use the type of design described in Section 9.3 but *double* the selection probability for the large UAUs and *halve* the selection probability for households within these large UAUs. This leaves the sample still self-weighting for the individual interview but introduces a very simple weighting into the household schedule sample: the large UAUs will need weighting by half in the household schedule data-processing. The range of variation in workloads among different UAUs will be roughly halved. It may well be considered that the limited inconvenience caused by the introduction of just two different weights into the data-processing of the household schedule is worth paying for the advantage in field work organization.

These devices for ensuring self-weighting, or near-self-weighting, in the household schedule sample are worth considering only if it is planned to use the large-sample household schedule data for a significant volume of work in demographic analysis.

10. Errors in implementing the sample design

10.1 Introduction

It is relatively easy to make a good sample design; it is much more difficult to have it precisely implemented. Indeed, even in the best of circumstances, perfect implementation of the plan will never be achieved: there will always be some refusals or non-contacts at the interview stage even if there are no other errors.

In this section, we discuss the different sources of implementation error and possible preventive and corrective action.

10.2 Sources of error in sample implementation

Firstly, the sampling frame may be defective. Such defects may be of various kinds:

(i) Under-coverage. Parts of the target population are omitted.
(ii) Duplication. Some units are included more than once.
(iii) Incorrect or inadequate identification. A unit once selected cannot be confidently identified in the field.
(iv) Inaccurate supplementary information in the sampling frame. For example, there may be errors in the data used for stratification or in the size data used for PPS sampling. This category of defect is not, however, a source of bias.

There is also the problem of out-of-dateness in the frame, but this is not so much a separate defect as a cause of one or more of the defects listed above.

Secondly, there may be errors in sample selection: the sampling instructions may be wrong or misunderstood, or mistakenly applied, leading to wrong selections or selection with wrong probabilities.

Thirdly, there may be coverage errors in the field: part of an area unit may be omitted in error, or a wrong area included, by the interviewer. A household may be omitted during listing or an eligible woman may be overlooked. An important source of error in this category is imperfect identification of the household. The lister counts two groups as a single household but the interviewer who comes later for the individual interview considers them as separate and covers only one of the groups. Also in this category may be counted the accidental omission of certain questions in the interview.

Fourthly, failures of coverage in the field may arise not through any error by field workers but because of actual field circumstances. For example, an area may be inaccessible – perhaps due to floods or to civil disturbance – or a household may be absent even after repeated call-backs, or an individual may refuse the interview or part of it.

Fifthly, completed questionnaires (or punched cards or tapes) may be lost at any stage after the interview.

Most of these types or errors amount to failure to obtain or transmit desired information. A few involve obtaining the wrong information: either information for the wrong unit or wrong information for the right unit. When we come to discuss corrective measures we shall deal with these two classes separately, but first we consider the problem of error prevention.

10.3 **Preventing errors in sample implementation**

Everyone knows that "prevention is better than cure". While the errors listed above cannot all be prevented, much can be done to reduce their frequency.

The first essential is to foresee the possible errors; the above list is intended to help in this respect. Once a source of error is foreseen, some simple counter-measures can usually be devised. Many of the sources listed above can best be tackled by a careful policy based on the following principles:

(1) Provide clear instructions for every step in the procedures.
(2) Train the workers concerned in the understanding of these instructions.
(3) Check that the instructions are understood.
(4) Supervise their implementation. Provide rapid feedback on errors to those responsible.
(5) Check the results – wholly if practicable, otherwise in part.

Defects in existing area sampling frames can sometimes be anticipated by examining the experience of other recent surveys in which the same frame was used. Indeed, except where the frame has been often used, and convincingly, it is always advisable to carry out a number of spot checks in the field before a final decision is taken to make use of the particular frame. If it is found that maps of small areas are grossly unreliable, a sampling frame of larger areas may have to be used instead and a special sample mapping operation introduced to create small area units in the field.

Errors in identifying the household can be minimized by careful attention to the drafting of listing forms and instructions. For example, in societies where a single household may be divided between separate dwellings, the listing form should provide for information on the number of dwellings occupied by each listed household. In other cases the problem may be met by a decision to use the dwelling, rather than the household, as the basic unit. If the dwelling is used as the sampling unit, then the interview will cover whoever is living in the selected dwelling. (But this is not *always* the best solution; the question requires study in each individual country.) Our choice of the *de facto* population for the survey is similarly motivated – in this case to minimize the incidence of non-contacts at the stage of the individual interview.

Field troubles, in general, including refusal and partial refusal, should be tested in advance through a pilot study. In some cases this may suggest improvements leading to better response. (See *WFS Training Manual.*)

10.4 Detecting errors in sample implementation

However many precautions are taken, some errors will still occur. An important part of the policy towards errors of implementation is to detect these errors as early as possible.

In general, this is done by means of independent checks. Firstly, a regular system of spot-checking by supervisors should be applied to all field operations. Secondly, one should check against existing sources of information where possible: for example, the household count can often be checked against the census count for the same area. While exact agreement is not to be expected, discrepancies which appear excessive can be singled out for a re-count. One may also be able to check the *mean household size* against that found in the census. A particularly good check on bias due to omission of eligible women is to compare the *mean number of eligible women per household* found in the individual survey against the mean number reported with the same eligibility qualifications in the household schedule.[1]

Similarly, the number of interviews achieved should be checked against the number expected (based on the listing). Interviewers and supervisors should report on the work completed and the number of interviews they report should be checked against the number of questionnaires received at headquarters. This in turn should be checked against the number punched and, finally, against the number in the computer print-out.

The number of interviews achieved, refusals and non-contacts should always be stated in the survey report, separately for the household schedule and the individual questionnaire. A simple way of ensuring that this information is collected is to insist that a household schedule be initiated for *every household in the sample*, even if no interview was obtained, and, similarly, for at least the cover page of the individual questionnaire, for *every identifiable eligible woman*.

10.5 How to deal with errors of sample implementation once they have been detected: (A) Incomplete coverage

There are several different kinds of action which may be taken once an error of incomplete coverage is detected.

(1) We may go back and collect the missing data.
(2) We may use some other data as a substitute.
(3) We may make some kind of re-weighting adjustment.
(4) We may modify the survey definitions or specifications, after the event, to conform to what actually happened.
(5) We may take no action – beyond mentioning the error in the survey report.

It is not practicable to specify exactly which policy should be followed in what circumstances, but we consider here some of the factors involved.

Serious omissions at the listing stage should probably be dealt with by method (1): going back and doing the job properly. Apart from this, method (1) would usually be too costly unless the error comes to light before the field team leaves the area.

[1] *Of course such a check is relevant only where the two questionnaires are administered at separate interviews and the household membership list in the household schedule does not determine the sampling frame for the individual questionnaire. This implies method (iv) b of Section 7.7.*

Method (2), replacement of missed areas or households by others, is often thought to be essential by those unfamiliar with sampling theory but in reality it may not be worth-while. The need to maintain the initially planned sample size is *not* an important consideration in this respect: the bias due to omission (non-response) is not corrected by adding in substitutes, and the loss in sampling precision due to the slightly smaller sample is generally negligible and not worth the trouble of a special replacement operation.[1]

If replacements *are* made they must be selected by the supervisor according to a procedure devised by the head office. Choice of the replacement should never be left in the hands of the interviewer.

A common source of sample defect in developing countries is the complete failure to gain access to an area, whether through floods, epidemic, civil disturbance, wild animals or whatever. Often this wipes out a whole area from the sample. Of course the best remedy is to return later when access can be obtained, but this may be ruled out by the time factor. Failing this, the best policy is replacement of the area by another chosen to be as similar as possible. The choice should be made by the head office.[2]

Before leaving method (2), it may be worth mentioning that many samples have been completely invalidated by attempts to modify them after selection. For example, in one survey, clusters once selected were found to be unexpectedly small, so a neighbouring cluster was added to each selected cluster; this led to a bias whose size could not be estimated. This type of modification is dangerous and should never be attempted without consulting a sampling specialist.

Method (3), re-weighting adjustment for losses, is often used to correct for non-response. The assumption is made that the non-respondents are like respondents (as in the duplication-of-data method). This assumption may be qualified if some kind of information is already available about the non-respondent; we then assume that she is like respondents who have the same known characteristic. The main objection to mathematical adjustments is the slight increase in complexity at the processing stage: a special weighting has to be introduced and if the sample was specially designed to be self-weighting this feature is lost. Duplication of data is simple and, once done, can be forgotten in the rest of the processing (unless so frequent that the variance is affected).

Method (4), modifying the definitions or specifications to conform to what actually happened is, obviously, an undesirable policy but may occasionally seem the best way out. For example, if a substantial number of areas have proved inaccessible and if most of these were in a particular region of the country, it may be advisable to redefine the survey coverage and regard the whole project as representing only that part of the country which excludes the region concerned.[3] Similarly, if an attempt is made to include nomads in the survey but the

[1] *On the other hand, it is reasonable to take account of the expected rate of non-response at the planning stage and arrange for a somewhat larger sample than needed, to cover this. If this is done, the additional households should be regarded as part of the sample, like any others, rather than set aside as "substitutes" to be used if the occasion demands. The latter tends to encourage field workers to accept non-response.*

[2] *If we are planning on two units per stratum and one is omitted, a possible replacement policy is to duplicate the data for the other one. Note, however, that this will give zero variance for the stratum concerned. Whether this is acceptable will depend on the cost of the alternative policies and on a decision as to whether we can reasonably do without that stratum in estimating variances.*

[3] *Re-defining the survey coverage before sample selection is simple. But if the re-definition is undertaken in order to fit the particular set of sample failures after they have occurred, considerable care is needed. In principle, one should attempt to exclude from the universe whatever would have been excluded had the complete pattern of inaccessibility been known before sampling.*

attempt is unsuccessful, one might reasonably modify the description of the project and say that it covers the sedentary population only.

Method (5), taking no action, may be a reasonable policy if the omission rate is very small, say up to 2 or 3 per cent. Note, however, that this would be sensible only for means, rates and percentages, and even then only if both numerator and denominator come from the sample. If the denominator comes from a separate source, or if we are estimating a total, then obviously one should inflate the sample *pro rata* to allow for the omission rate.

10.6 How to deal with errors of sample implementation once they have been detected: (B) Incorrect information

A few types of errors listed in Section 10.2 involve the collection of data for the wrong unit. In principle, of course, such data should simply be thrown away. However, this will leave a gap, or omission, and we have already seen that one policy for dealing with omissions is to substitute other data. Could we not simply use the data collected for the wrong unit as the substitute? This means, in effect, leaving the error uncorrected.

To decide whether this is a reasonable policy in any given case, one should look very carefully at the cause of the error in order to determine whether there is any likelihood of bias. Here are two hypothetical examples.

(i) The listing in one area produced 120 households. Instructions were to select a systematic sample of 1/6, which should have yielded 20 households. The sampler misunderstood the method and, in counting the interval of 6, always included the one he had just selected. This meant that he really selected 1/5, which yielded 24 households. Here the error is random. We only have to reject the data for 4 of the selected households. (Select a random number between 1 and 6, say 3. Reject the 3rd, 9th, 15th and 21st households selected.) The remaining 20 households, though not the 20 which should have been selected, can be retained as an unbiased sample.

(ii) The interviewer reports that she could not find a group of selected households so she made her own selection to replace them. Reference to a map reveals that the households she missed were all in a remote hamlet far from the main village, while her replacements came from the main village. There is a clear bias here.

Once again, whatever methods are adopted to deal with coverage of wrong units a full account should be included in the survey report and the number of cases involved should be stated.

Finally, there are cases where, although the right units are covered, wrong supplementary information is obtained, leading to an error in stratification, to use of the wrong "size" in PPS sampling, or to an unexpectedly large or small sample size in some or all areas. In nearly all cases *no correction should be made*. Mis-stratification is difficult to correct after the event and would seldom be worth attempting. Use of the wrong "size" will not lead to a biased sample in the PPS self-weighting design, provided the procedures described in Section 9 are followed: the relevant probabilities are those that *were in fact used* in sampling, not those that should have been used. If the "size" used is wrong, the only effect will be an abnormally large or small work load of households in the area concerned; generally, this can be tolerated.[1] One further type of "error" that is not a true error and does not require correction is the appearance in the sample of persons who are outside the domain of the survey. In some surveys foreigners are supposed to be excluded, and in all surveys a listed person who dies before the individual interview is an exclusion. These exclusions are in no sense to be corrected or replaced; they should simply be dropped.

[1] *If not, one can adjust the last stage sampling fraction and introduce a weight to allow for this – but this abandons the self-weighting feature. Alternatively, in a multi-stage sample, if the discrepancy is discovered in time one can adjust the number of selections at an earlier sampling stage, maintaining both self-weighting and the fixed work load (but introducing other complexities).*

11. Sampling for supplementary operations

11.1 Introduction

Two kinds of supplementary operations have been recommended for use with WFS projects, not necessarily in every country but in a selection of countries where they appear feasible and appropriate. These are the post-enumeration quality check (QC) and the husbands survey (HS). These operations each have more than one possible use and it is important to have a clear view of their purpose before fixing the sample design.

Firstly, a checking operation based on partial repetition of a survey may be intended merely to give an idea of the reliability of the responses, or it may be expected to provide correction factors. In the latter case it must be organized in such a way as to be markedly more accurate than the main survey, moreover, it must be based on a substantial sample.[1] If we know how to obtain much more accurate data from a substantial sample it would seem sensible to use our resources in doing just that – for the main survey. Thus we assume, in the WFS, that the checking operations are intended primarily to throw light on response reliability (or "stability") rather than to provide correction factors. It follows that a modest sample will be acceptable.[2]

Secondly, a repeat survey may be intended as a case-by-case check or as a statistical check. Although several surveys have in fact been repeated with the intention of providing a statistical check, using a separate sample, such a check is almost worthless. If conditions are the same, such a check must give the same result, within the limits of sampling error. Sampling error is defined as the variation occurring when an identical operation is repeated on another sample of the same population; thus, such a check is checking nothing but sampling theory – and our ability to hold conditions constant. If it gives the same result (as it normally will), we still do not know anything about the stability of individual response. For these reasons, the WFS checks are intended for analysis on a case-by-case basis. They will result in tables showing how many people gave a different response in the check operation: a cross-tabulation of first response by second response for each question would seem appropriate. To make this possible, the QC sample must be a subsample of the main survey respondents, not an independent sample.

In some checking surveys tabulations of the above kind have been made and presented in a "blind" way straight from the computer, without even eliminating cases in which the first and second responses relate to different people, and without considering what changes would be expected to arise simply with the passage of time between the main survey and the check. To get the full benefit from such checks a more probing analysis is desirable: the results should be confined to cases in which the two responses relate to the same person and some attempt should be made to estimate the incidence of "legitimate" response changes (in age

[1] *Most errors do not take the form of small biases affecting nearly every response but rather of large mistakes occurring only rarely. This means that their variance is large and a fair-sized sample is needed to estimate any correction factor.*

[2] *An alternative approach is the "dual record system", in which estimates are based on the assumed independence of two operations without assuming that either one is more accurate. However, this technique assumes that the only errors are errors of omission; in the QC and HS we are concerned with response errors in general and this assumption is clearly inappropriate for most items.*

and marital status, for example). The time interval between the survey and the check should also be minimized. Finally, each discrepancy found should be re-checked to determine whether it is an error of processing.[1] Such precautions are an essential part of any serious attempt to pin-point response variance.

The husbands survey (HS) plays the role of a check, like the QC, but may also serve other purposes. For some of the items, such as husband's occupation, we obviously expect to get a more accurate response from the husband. Moreover, the HS provides the opportunity to obtain some additional information of interest. However, we shall assume that the HS, like the QC, should cover a relatively small subsample of the households selected in the main survey and that the husband's and wife's responses to the same item will be cross-tabulated. Countries are encouraged to include either the QC or the HS in their WFS programme, but since both are essentially checking surveys the same country would not normally use both.

11.2 Sample design for the quality check and husbands survey

Before fixing the sample design, one has to settle the crucial question of the time interval between the main survey and the QC or HS. The QC involves repetition of many of the original questions to the same respondent. If the interval is very short, many respondents will recall the questions asked and the answers previously given: this both reduces the interest of the check and risks annoying the respondents. Thus the QC should normally follow the main survey after a substantial interval, say at least a week. Probably the same argument applies to the HS: although a different respondent is involved, we cannot normally guarantee that neither of them is present at the other's interview; hence, to avoid causing annoyance, it seems desirable to space the two interviews by at least a week.

On the other hand, an excessively long spacing is also undesirable: as the period is lengthened we loose increasing numbers of respondents through mobility, and the incidence of real changes in the couple's circumstances also steadily increases. Both of these factors would reduce the interest of the comparisons between responses. Thus no more than a few weeks should be allowed to pass between the main interview and the QC or HS.

These constraints affect the sample design in two ways.

(1) The QC or HS will require a separate visit to the survey locations; we cannot take the check "as we go", for example by interviewing the husband in every tenth household in the course of the main survey. It follows that there will be an economic argument in favour of clustering the QC or HS sample. If, for example, we fix an overall subsampling fraction of 1/10 for the QC or HS, it would be uneconomical to visit all the initial sample clusters (UAUs) and select 1/10th of the previously interviewed households in each; the amount of travelling would be disproportionate to the amount of interviewing.

(2) The sample design for the QC or HS should be such that we can select the subsample of areas without awaiting nationwide completion of the initial survey.

[1] *Gross error, including processing error, is of interest, but so is response error in the narrower sense. Both should be assessed.*

A very simple design which would satisfy these conditions would be to select a fixed fraction of the UAUs selected in the main survey sample and then a fixed subsampling fraction of the households interviewed in the subsample UAU in the main survey. The QC is particularly a check on the field workers themselves and should therefore cover a substantital fraction of all the field workers. Thus there is a case for a relatively high sampling fraction at the area stage for the QC.[1] This argument hardly applies to the HS. The following subsampling fractions are suggested as roughly appropriate for the two operations:

	Subsampling Fraction		
	Quality Check	Husbands Survey (a)	or (b)
Ultimate area units	1/2	1/3	1/4
Households	1/5	1/3	1/2
Overall	1/10	1/9	1/8

Alternative (b) would be cheaper than (a) but should not be considered unless the total number of UAUs in the main sample is large, say over 150. Sampling should be by systematic selection both at the area stage (but with one reservation: see below) and the household stage. The reservation is as follows: if the initial sample had provided for two UAUs per stratum then systematic subsampling of 1/2 would lead to selecting always the first (or always the second) member of the pair in every stratum. If the list was ordered in the same way in each stratum, this could distort the sample. In such a case, it would be better to select *randomly* one of each pair, instead of systematically. Clearly the same kind of trouble could arise in other ways – e.g., initially two UAUs per stratum with systematic sampling of 1/4, or initially three UAUs per stratum with systematic selection of 1/3. When there is any danger of this kind, random selection within strata should be used for the subsampling at the area stage. It is suggested that a household should be included as eligible for the subsample only if an individual interview was successfully conducted in it during the main survey. The QC interview should then be conducted with all, and only, women in the selected household who were successfully interviewed before, and the HS with the husbands of such women (if available). The reasons for this limitation of the coverage are discussed in Section 11.3.
An alternative plan would be to select the subsample from the list of *women*, rather than *households*, successfully interviewed. In practice, there is little to choose between these two schemes.

11.3 Check on coverage and non-response

In addition to the information collected at the interview itself, the QC and the HS provide

[1] *Specifically designing the subsample so as to represent each interviewer's, or each team's, work would create considerable organizational problems in a survey in which interviewers work in teams, and this does not seem worth attempting.*

an opportunity for field checks on the coverage achieved in the main survey and on the accuracy of the listing. In general it will be worth taking advantage of the presence of field workers to arrange such an independent check.

The QC or HS field workers should independently determine the boundaries of the area unit and carry out an independent listing operation. A good plan is to give the original list to the QC or HS supervisor, who then makes the comparison after the interviewers have completed their new listing. In case of discrepancy he then re-checks to determine which listing (if either) is at fault. If an error is demonstrated in the original listing, it is not likely to be worth correcting at this late stage (in any case, the check is limited to a subsample) but the finding should at least be reported, together with any available indication of the reason for the error, so that future surveys may benefit from the experience.

It might be suggested that the QC should cover not merely the households (or women) successfully interviewed in the main survey but also the main-survey non-respondents in order to obtain information about these. However, with an overall subsampling fraction of 1/10 it seems unlikely that anything useful would come out of this. Most of the original non-respondents will (presumably) still be non-respondents; those who are not will be such a small group that it will not be possible to draw statistically reliable conclusions. It is for this reason that we have suggested the simpler, and less costly, procedure of subsampling only the successful interviews.

11.4 Subsampling for modules

Some countries may wish to include one or more of the WFS "modules" in their interview, but only in a subsample of cases.

The most efficient way of handling this would normally be to designate one or more interviewers in each team to carry out the augmented interviews (core plus module). In every ultimate area unit, a systematic subsample of households would be selected from the main sample immediately after the latter is selected, using a fixed subsampling fraction throughout, and the specialized interviewers would be sent to this subsample. This arrangement differs from that suggested for the QC and HS in that the subsample is treated simultaneously with the main sample.

Cluster H-h

APPENDIX

CONFIDENTIAL

Information to be used for research purposes only

World Fertility Survey
(International Statistical Institute)

HOUSEHOLD SCHEDULE

[NAME OF COUNTRY]

[NAME OF ORGANIZATION]

IDENTIFICATION	
PLACE NAME ______________________	
CLUSTER NUMBER ______________	HOUSEHOLD NUMBER ______________

Interviewer Calls	1	2	3	4
Date Interviewer name Result*				

***Result codes**	1. Completed 2. No competent R at home 3. Deferred 4. Refused	5. Dwelling vacant 6. Address not a dwelling 7. Address not found or non-existent Other (SPECIFY)

NOTE TO READER:

Countries are invited to add questions on characteristics of the dwelling or the household such as:

Materials used in its construction
Number of rooms
Presence of WC
Sewage facilities
Water supply
Form of lighting
Tenancy status
Possession of "modern" objects such as bicycle, motorized vehicle, sewing machine, gas or electric cooking stove, refrigerator, radio, clock or watch, etc. (RECOMMENDED FOR ALL COUNTRIES)
Affiliation of the household in terms of language, race, ethnic group, religion, etc.

Now we would like some information about the people who ordinarily live in your household, or are staying

	NAMES OF USUAL RESIDENTS AND VISITORS	RELATION-SHIP	RESIDENCE		SEX	AGE	EDUCATION		MARITAL STATUS: FOR THOSE AGED — AND OVER	
	Please give me the names of the persons who usually live in your household.	What is the relationship of this person to the head of the household?	Does this person usually live here? Y/N	Did this person sleep here last night? Y/N	Is this person male or female? M/F	How old is (he/she)?	Has (he/she) ever been to school? Y/N	IF YES: What was the highest level and year of schooling (he/she) completed?	Has (he/she) ever been married? Y/N	IF YES: Is (he/she) now married (M widowed (V divorced (D or separated (S)?
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
01										
02										
03										
04										
05										
06										
07										
08										
09										
10										
11										
12										

Just to make sure I have a complete listing

IF CONTINUATION SHEET USED, TICK HERE: ☐

[*Double page reduced from 42×30 cms.*]

th you now.

FERTILITY: FOR ALL WOMEN AGED____YEARS AND OVER												ELIGIBILITY	
NUMBER OF LIVE BIRTHS						SUM	PARTICULARS OF HER MOST RECENT LIVE BIRTH				FERTILITY RESPON-DENT:		
)oes she ave any hildren of er own ving with er? F YES: Iow many ons and ow many aughters?		Does she have any children of her own who do not live with her? IF YES: How many sons and how many daughters?		Has she ever given birth to a child who later died? IF YES: How many sons and how many daughters have died ?		Just to make sure I have this right, she has had ___(SUM) births. Is that correct? IF NO: CORRECT RESPON-SES.	In what month and year did her last birth occur?		Was that a boy or a girl? B/G	Is that child still living? Y/N	GIVE LINE NUMBER OF PERSON ANSWER-ING COLUMNS 11 - 21	TICK ALL WOMEN ELIGIBLE FOR INDIVI-DUAL INTERVIEW	
S (11)	D (12)	S (13)	D (14)	S (15)	D (16)	(17)	MONTH (18)	YEAR (19)	(20)	(21)	(22)	(23)	
													01
													02
													03
													04
													05
													06
													07
													08
													09
													10
													11
													12

Are there any other persons, such as small children or infants, that we have not listed?

YES ☐ (ENTER EACH IN TABLE) NO ☐

In addition, are there any other people who may not be members of your family, such as domestic servants, friends or lodgers who usually live here ?

YES ☐ (ENTER EACH IN TABLE) NO ☐

Do you have any guests or visitors temporarily staying with you ?

YES ☐ (ENTER EACH IN TABLE) NO ☐