10.4: Enumeration

A census of the population may be conducted after, or at the same time as, mapping. The census will involve the collection of information on the composition of each household and demographic, and possibly other, data on each household member.

4.1 Organization of enumeration of households

A combination of speed and accuracy is required in the conduct of a census. It is useful to draw a flow chart of the data collection and processing operations. Simple examples of such charts are shown in Figure 10.2 for collection of data on paper or on an electronic device.

(a) Data collected on paper
A field manual is essential and should include a checklist of equipment that the interviewers will need to take with them each day (see Chapter 16).

4.2 Definition of dwelling units

The definition of a village and a household (or compound) within a village will vary, depending on the location of the trial. Villages generally share the same leaders, although the inhabitants may be dispersed over a wide area. In parts of Africa, for example, in the Sahel zone of West Africa, a compound is a cluster of households fenced or partitioned off from other compounds and may have features, such as a well or latrine, which all the households of a particular compound share. In parts of Asia, such as in parts of Borneo and Indonesia, several households live together in a single building called a ‘longhouse’.

A household is usually defined as a nuclear or extended family group, whose members usually eat together (the ‘from the same cooking pot’ definition of a household). The exact definition of a household should be decided before mapping and enumeration begin and clearly defined in the field manual. Households can be spread over several buildings, or several households may share the same building. There are no uniquely correct ways of defining households,
compounds, or dwellings, but, in any particular study, it is important that clear definitions are agreed for all of the different terms to be used in describing people’s living arrangements. New investigators in an area should find out what systems others have used who have worked in the same area, and whether or not these worked satisfactorily.

### 4.3 De facto and de jure populations

Before conducting a census, it is necessary to decide which individuals will be registered as members of the study population. The two commonest options are the so-called *de facto* and *de jure* populations. The *de jure* population comprises the ‘normal residents’ and includes individuals who usually live in a particular household but who may be absent during the enumeration. The *de facto* population consists of those who slept in the household the night before the census. In national censuses, it is usual to enumerate the *de facto* population, but, for the purpose of most intervention trials, the *de jure* population is the most appropriate. In some cultures, the definition of household membership may be difficult to specify. Some individuals may live in one household but spend a significant amount of time in another household either within or outside the study area. These individuals may be incorrectly enumerated twice, unless care is taken to assess the unique ‘normal’ domicile of each person. When using a *de jure* enumeration, each resident’s status can be recorded as ‘absent’ or ‘present’. This will give some indication of the degree of temporary migration and would allow the calculation of the *de facto* population from the *de jure* census. Similarly, fieldworkers will have to distinguish between ‘temporary’ visitors and those who will remain for a long time. It may be difficult to obtain such information reliably, as respondents may inform a fieldworker that a temporary visitor is ‘permanent’ if it is thought that some benefit may derive from this. The definition of who is a normal resident will depend upon the objectives of the trial. It is important to decide upon a period of time that a person should have been in or out of a community to be considered as having migrated in or out. In general, a clear and full definition is required as to who should be considered as a resident, especially in long-term studies that may involve multiple census updates.

The definitions should be clearly stated in the field manual.

### 4.4 Ensuring completeness of the census

As houses may be empty at the time the interviewer calls or some residents may be away, the interviewer may have to rely upon proxy reporting in some instances. If a house is empty, arrangements should be made to call back at a time when someone is likely to be there. Whenever possible, all households reported as being empty should be revisited, ideally later the same day, by a supervisor or another interviewer. This helps to avoid interviewers reporting remote households as being empty to reduce their workload.

Information about the composition of the household is best elicited if there is a standard order in which information is sought about individuals (discussed in Section 4.5). In a simple census, it is not necessary that the information on all members of a household should be given by a single respondent, nor that the interviews are held privately, unless sensitive information is also being collected. Whenever there is some lack of certainty, respondents can be encouraged to consult others in the household or compound to provide information. It is useful to specify in advance who would be regarded as an acceptable informant in a household. For example, for information on young children, the list, in order of preference, is often first the mother, second another adult female relative living in the household, and third the father.

Whether or not a respondent is willing to co-operate in the study may depend on the initial impression an interviewer
makes and on the respondent’s understanding of the reasons for the census. Co-operation may be poor if the study subjects suspect that the information collected may be used to their disadvantage (for example, for tax collection). Involvement of local leaders and the CAB, if one is set up, may be critically important in obtaining co-operation (see Chapter 9). The interviewers should introduce themselves properly to the respondents, explain the purpose of the study, and assure them that any information given will be regarded as confidential. It may be necessary to reassure them, specifically, if appropriate, that the information will not be made available to the local administration for compiling lists of taxable adults. If those in a household refuse to participate, the field supervisor should be informed, and, with input from the CAB, the reasons for their refusal investigated as soon as possible. An initial refusal should not be taken as final. Individuals may be unwilling to collaborate merely because they have not properly understood the objectives of the trial or have not appreciated the potential benefits to them. However, the right of an individual not to participate in a survey should always be respected. If more than a small proportion of individuals refuse to participate, the generalizability of the trial findings may be compromised. Discussions should be held with village leaders if it appears that such problems are developing, in order to ascertain the reasons and to seek suitable remedies.

If data are collected using mobile phones or PDAs, these should be synchronized with computers, and the data uploaded each day. Whether the data have been collected on paper or electronically, at the end of each day, all completed forms should be carefully checked by the interviewers and, whenever possible, also by a supervisor for errors or omissions, so that these may be corrected either immediately or on the following day, before the team moves on to another area. Plans should be made to revisit any household that could not be enumerated, because of the absence of eligible informants or because the house was empty.

4.5 Numbering and identifying individuals

One purpose of a census is to allocate a unique identification number to each member of the population. This number will remain assigned to the individual for the duration of the trial, since it may be used to link information on an individual from different sources, such as from interviews, clinical examinations, and laboratory studies, and also on different occasions such as baseline, interim, and final surveys. Therefore, the person’s identification number must never be changed or reallocated to any other individual, even if they die or move either within or outside the study area. There are several different ways that are commonly used to allocate identification numbers. As an example of one such system, suppose, in village B, the first compound is numbered 01. Within compound 01, the first household is numbered 01, and the household head is given the number 01 within the household. Thus, this individual has the unique identification number B010101, plus a check digit (see later within this section and Box 10.1). (Note that such a numbering system assumes there are fewer than 27 villages in the study, fewer than 100 compounds in every village, fewer than 100 households in every compound, and fewer than 100 persons in every household (see also Chapter 20, Section 5).) If this identification system is used, a separate record should also be kept of the location of this individual at each study visit. If, for example, this same individual is currently within their original household, their current household will be B0101 plus a check digit. However, if they have moved to household 28 within compound 17 of village K, then their identification number will still be B010101 plus a check digit, but their current household will be K1728 plus a check digit.

Box 10.1 Method of assigning check digit to six-digit number
Suppose the trial number consists of a six-digit number, and it is desired to add a one-digit check number that will guard against transcription errors (such as reversing the order of two digits or recording one digit incorrectly). The number will take the form of:

<table>
<thead>
<tr>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prime</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>c</td>
</tr>
</tbody>
</table>

The first six prime numbers are shown below the digits of the trial number. The check digit c is calculated by multiplying each digit by the corresponding prime, summing the results, and the last digit of the result is taken as the check digit. Thus, for example, we would have:

<table>
<thead>
<tr>
<th>Trial number</th>
<th>Trial number with check digit</th>
</tr>
</thead>
<tbody>
<tr>
<td>467913</td>
<td>4679133</td>
</tr>
<tr>
<td>476913</td>
<td>4769135</td>
</tr>
<tr>
<td>567913</td>
<td>5679134</td>
</tr>
</tbody>
</table>

Source: based on methods supplied W. Meade Morgan (personal communication).

Alternatively, numbers might be allocated in a simple continuous sequence to each member of the trial population, without building codes for village or household into the number. An advantage of this system is that forms can be pre-numbered before they are taken to the field, and the number allocated to an individual is simply that on the form that is filled in for them.

Whichever system is used, it is important to supplement the number with a check digit or character to aid the detection of transcription errors. These work by using a formula whereby any number can only correspond to one character or digit. If the number is transcribed wrongly, then the check digit or character will not match. One source for check digit systems is available at [http://code.google.com/p/checkdigits](http://code.google.com/p/checkdigits). A simple method of generating check digits to guard against common transcription errors (such as reversing the order of two digits or recording a digit incorrectly) is given in Box 10.1.

In addition to, or instead of, the check-digit system, the practice in some trials is to record, for data linkage purposes, both an individual’s identification number and the first few, say five, letters of their name. Checks are made that both of these items match, before any linkage procedures are undertaken. However, this system does require that an individual does have a name with an explicit spelling. Sometimes, people use several different names and are not consistent about how they are spelt, so we recommend using a check digit.

### 4.6 Household or individual forms within a census?

After mapping the study area and assigning numbers or codes to villages, compounds, and households, the household and/or individual census survey forms can be marked with household identification numbers. Whether all members of a household should be recorded on one form or on separate individual forms will depend on the way in which the survey is organized, the amount and degree of standardization of the data collected on each individual, and the design of the data collected.
processing system. Sometimes, both a household form and individual forms will be required—the former to collect basic demographic information on all members of a household, and the latter to record more detailed information on some, or all, members of the household.

If the census is being conducted at the same time that other procedures are being undertaken on the study subjects, it may be best to use individual forms, in addition to household forms, as otherwise it may be necessary to wait until a complete household has been registered before other procedures can start. If household sizes are large, this may lead to significant delays for those following the interviewers, especially at the start of each day.

Figure 10.3 is an example of a simple household form to collect basic demographic information. General issues related to production and coding of questionnaires and forms are considered in Chapters 14 and 20.

### 4.7 Coding relationships

Interviewers should be instructed regarding the order in which individual household members should be registered, as a systematic approach is less likely to lead to omissions.

In polygamous households, it would be usual to begin recording with the male household head (if there is one), followed by his first wife, and all her children living in the household; his second wife and her children; and so on.
might be any brothers of the household head, each followed by their wives and children, as for the head. Unrelated individuals, such as lodgers and employees, might be recorded last. Relationships between different household members may be coded so that, in so far as possible, everyone is linked to one or two others in the household in a simple way, using as close a relationship as is possible. Codes for brother, sister, mother, and so on should only be used when wife, son, or daughter cannot be used to describe a relationship. To the extent possible, terms, such as granddaughter, grandson, grandmother, grandfather, niece, nephew, uncle, aunt, and cousin, should be avoided. An example of a coding system for relationships that was used in a vaccine trial in Uganda (Smith et al., 1976) is given in Box 10.2. Two alternatives to this procedure may be better in some circumstances—either everyone is related to the household head or detailed records are made of the name of each individual’s mother and father (even if they are dead or do not live in the household).

In some societies, it may be very difficult to ascertain the precise relationship between individuals. For example, no apparent distinction may be made between children and nephews/nieces—both the father and uncle might refer to them as his children. So long as this is appreciated it may cause little confusion, but it may be very important if, say, genetic studies are being conducted.

Box 10.2 Example of instructions for coding relationships

Coding of relationships

In this column, write down the relationship of the individual to the other persons in the household. Since each person will be entered against a person number (the second item in the columns), the relationship can conveniently be expressed by reference to these numbers, for example, ‘Wife of 01’ or ‘Son of 01 and 02’.

The following abbreviations may be used:

<table>
<thead>
<tr>
<th>Head of household</th>
<th>H</th>
<th>Sister</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wife</td>
<td>W</td>
<td>Grandson</td>
<td>GS</td>
</tr>
<tr>
<td>Son</td>
<td>S</td>
<td>Granddaughter</td>
<td>GD</td>
</tr>
<tr>
<td>Daughter</td>
<td>D</td>
<td>Grandfather</td>
<td>GF</td>
</tr>
<tr>
<td>Mother</td>
<td>M</td>
<td>Grandmother</td>
<td>GM</td>
</tr>
<tr>
<td>Father</td>
<td>F</td>
<td>Other blood relative</td>
<td>R</td>
</tr>
<tr>
<td>Brother</td>
<td>BR</td>
<td>Unrelated</td>
<td>X</td>
</tr>
</tbody>
</table>

Example: A household consists of the head, his two wives, and five children, three by his first wife and two by his second, and also his mother, and an unrelated visitor and her child. These would be coded as follows.
4.8 Names and addresses

The most important way of identifying an individual will be through his or her names, and these must be recorded with special care. Interviewers must be instructed how to spell names, including those given by semi-literate individuals. It is important to try to record all of the names of a person, including nicknames, as it is not uncommon, in some cultures, for individuals, and especially children, to employ different names in different situations. The most frequently used names should be recorded first. In some areas, confusion may arise, as many people have the same names, especially in cultures in which the first-born males or females are always given a set name or in which they are always named after their grandmother or grandfather. In some societies, very young children are not named until some time after birth, and, until this time, they may have to be recorded as 'unnamed'. In some cultures, young infants are not thought to be part of society, and specific questioning may be necessary to elicit information, even about their existence.

In addition to the names, the complete addresses of study participants should be recorded. In some instances, this will be just the name of a village, but, if there is some system of subunits within a village, then this also should be recorded. Often, it will be useful to record the name of the local leaders or elder who have some responsibility in the area in which a participant resides, though it should be remembered that this person may change during the course of a study.

4.9 Ages

In some societies, it takes only a few seconds to elicit an individual’s age or date of birth through a simple question, but, in others, these are very difficult to obtain, as individuals do not know their age or date of birth and this information has no special significance to them. The importance of collecting accurate information on ages or dates of birth will depend upon the objectives of the trial.

Accurate dates of birth may not be necessary for all age groups, and those in age groups not pertinent to the trial may not have their specific age recorded at all (for example, but just be recorded as ≥50 years). In some trials, however, accurate estimates of dates of birth may be needed for all age groups. It is generally better to record the date of birth, rather than the age at last birthday, as the latter will change during the course of a trial. During the census, field staff can convert ages to dates of birth, using a simple application on a PDA or mobile phone or transcription tables (relating ages to years of birth), which should be included in their manual. Protocols and methods of estimating dates of birth, such as
those described in this section, should be an integral part of the interviewers’ training and be included in their field manual. Even if the study area does not have universal civil registration of births, various other sources of information may be available. For children, health cards and the mother’s antenatal card may be a good source of information. However, one should remember that, for children who were born at home and not taken to a health facility immediately after birth, they may be less accurate. Mothers can be questioned as to how many days or weeks old the child was when taken to the health facility. Antenatal cards should have dates of delivery or, if not, when the mother was seen and the estimated gestational age. In the absence of any documentation, various other methods of estimating dates of birth of a child have to be employed.

Developmental characteristics, such as the ability of the child to place the right arm over the head to touch the left ear (roughly possible from age 5 years onwards), the ability to sit upright unaided, walking, talking, and so on, can all be used to estimate the developmental age, and hence the approximate date of birth of young children.

Older children are more difficult to age by means of physical and developmental characteristics, due to variations in growth patterns. Age may be inferred from their grade in school or the grade in which they would be if they went to school. However, some educational systems make pupils repeat grades if they are thought unsuitable for higher grades, or a child may start school late.

If the interviewers can accurately age one child, the ‘index child’ method can be used. The mother is asked about her other children in relation to this child. For example, the fieldworker might ask questions such as: ‘Before Ebrima, did you deliver a live birth? Is that child here? How many rainy seasons passed before you became pregnant again?’ With such information on the birth interval, the preceding child’s date of birth may be estimated. Similarly, procedures can be used for the following child’s date of birth and all her other children.

To estimate the month of birth, calendars can be constructed. The calendar will list the months of rains, dry season, and so on. Religious or cultural festivals, such as Ramadan, Easter, or Christmas, can be included for recent years. For example, a mother might be asked if her child was born in the rains and, if so, whether at the beginning, middle, or end of the rains. At set times of the year, members of the village will be ploughing, planting, sowing, weeding, or harvesting different crops. An example of part of a monthly event calendar that was used in a study in Ghana is given in Box 10.3 (D. A. Ross, personal communication).

Children whose dates of birth are accurately known can be used as index children to estimate the dates of birth of children in other neighbouring households.

Having estimated the ages of all members of a household, the fieldworker should look at all the family together to assess if the ages are plausible, bearing in mind any infant or childhood deaths, stillbirths, or abortions.

The age of adult women can be estimated in several ways. Although age at menarche varies between women, a question about whether the woman had reached menarche before a certain event of known date can give a rough estimate of their date of birth (though, in many cultures, it may be difficult to discuss). Similarly, age at marriage may be, or may have been, fairly uniform for women in some societies, and women can also be asked if they married early or late, compared to their contemporaries. But when ‘marriage’ is deemed to have occurred must also be elicited, as, in some societies, the marriage process involves numerous stages.

Given an estimated age at first marriage, birth histories can be elicited to estimate a woman’s current age. Under
conditions of natural fertility, on average, approximately 2.5, 1.5, 1.0, and 1.0 years elapse between births, which are respectively, a live birth that was weaned, a live birth that died in infancy, a stillbirth, and an abortion. This method assumes no infertility, spouse separation, or use of contraception. In areas where these conditions are common, different assumptions have to be made.

Historical event calendars are one of the most commonly used methods to estimate ages. This method is especially useful where societies have a predominantly oral tradition. Historical event calendars require much effort to develop, and, before doing this, it is worth finding out if they already exist in government census departments or elsewhere. If they do not, a calendar can be created, with the assistance of local members of staff, teachers, and community leaders. The calendar should include all the major national historical events, and their dates, and all outstanding local events such as major bush fires, murders, drownings, deaths of religious and political leaders, wars, droughts, floods, famines, and so on. If an individual can remember an event and can estimate how old he or she was (for example, just married, just started school) at the time of that event, their date of birth can be estimated. This method is time-consuming and should be pilot-tested before use. It may be decided that it is too slow and cumbersome to be of use or there may be too few significant events that can be dated that individuals will remember for the method to be used. To be most useful, it is necessary to construct calendars which focus on local, rather than national, events and which are particular to a relatively small geographical area.

An example of an event calendar that was used in the same vitamin A trial in northern Ghana, that was referred to in Box 10.3, is given in Box 10.4 (D. A. Ross, personal communication).