

NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

Smith PG, Morrow RH, Ross DA, editors. *Field Trials of Health Interventions: A Toolbox*. 3rd edition. Oxford (UK): OUP Oxford; 2015 Jun 1.

Chapter 10 Censuses and mapping

1. Introduction to conducting censuses and mapping

In nearly all intervention trials, it will be necessary to compile a register of individuals included in the trial. The register should include sufficient identification information on each person to enable participants to be followed over time, with minimal possibility of confusing one individual with another. To assemble a suitable group for inclusion in a trial, it may be necessary to enumerate (i.e. count and identify) all the members of a geographically, or otherwise, defined population or a specific subgroup of it (for example, children aged less than 5 years). Such a population enumeration (census) may serve as a sampling frame to select a representative subset of the population or may be used to assess how representative the study group is of the whole population, if some individuals refuse to participate, or are not included, in the trial for other reasons.

Identification and follow-up of the members of a population and selecting a sample of them will usually be easier if a map is drawn of the area, marking individual homes and prominent topographical features. Mapping may also be valuable in planning the logistics of fieldwork and in studying the epidemiology of a disease, for example, to determine if cases of a disease tend to occur near water courses or in some other non-random fashion geographically.

Mapping and enumeration of a population are not always necessary, but often such information collected at the start of a trial is vital to its successful conduct. For example, in a leprosy vaccine trial in Venezuela, the trial group was defined as the household and other close contacts of prevalent leprosy cases (Gupte, 1999). The prevalent cases were distributed over a very wide area, in which most of the population were not included in the trial. It was necessary to enumerate the household and other contacts of prevalent cases, but it would have been inappropriate to enumerate the entire population or to map the locations of all households, other than was necessary to be able to find the contacts during the course of the trial. Conversely, in a malaria chemoprophylaxis study in The Gambia, an attempt was made to include all children in a defined area, and detailed mapping and enumeration were undertaken to facilitate the conduct of the study (Jukes et al., 2006).

In this chapter, guidelines are given on mapping and on ways of compiling a population register to facilitate long-term follow-up of the participants in a trial. Resources, including tools and advice on doing this in LMICs, are available from INDEPTH (<<http://www.indepth-network.org>>).

2. Uses of maps and censuses in intervention trials

A map of the trial area and a population enumeration (census) provide:

- ◆ a sampling frame for the selection of those in the target population who will be included in a trial
- ◆ denominators for the computation of morbidity and mortality rates
- ◆ baseline population characteristics, which may affect the impact of an intervention and which can also be monitored for changes during the study
- ◆ a basis for planning the logistics of the fieldwork, for example, which households should be visited by one fieldworker and in which order or to demarcate clusters within a cluster randomized trial
- ◆ a means for studying factors that affect disease rates. Age, sex, and place of residence affect the risk of many diseases, and information on these and other factors that may influence exposure or susceptibility to disease, or which may influence its outcome, should be recorded at the start of a trial.

3. Preparations for a census

3.1. Planning

Early in the planning of a census, it is important to ascertain what information already exists about the population, either in national censuses or from local or national surveys that may have been conducted previously. In planning a

census, it is important to seek the active collaboration of the community, generally through the community advisory board (CAB) and the local health services, using the special knowledge these groups are likely to have on the local population (see Chapter 9). This will also enable them to get to know their area better, and they may wish to use the information collected in the census for the benefit of the population, while the trial is in progress and after it has finished. Indeed, if local health workers and community leaders are not involved in the planning, they may be antagonistic to the study and may transmit these feelings to the study population.

In some populations, local administrative offices maintain up-to-date lists of tax payers that may give a good indication of the size of a population (or may not if large numbers of people avoid registering for tax collection!). Lists of voters or of residents may also be available through local administrative offices. Health or other surveys may also have been conducted previously. Gathering this information will entail visits to the study area, to government statistical offices, and possibly to universities or other institutions that may have organized specific surveys.

Useful data are usually available from national censuses, generally undertaken every 10 years. In the planning of a trial, census data may be used to select a suitable area such as a group of contiguous villages whose population is of adequate size for the trial. Often, however, the information in a national census is out of date or may be inaccurate. For example, a population census for a trial in Ghana found that the study census numbers matched those of the recent national census very well, except in one area that had applied to become a separate district where the national census numbers were roughly 50% higher than those in the study census! From a national census, it will usually be possible to obtain data for an area regarding the distribution of the population, with respect to age, sex, ethnic group, household size, and population density, though this may require a specific request to the census bureau. Estimates of mortality, fertility, and migration rates may also be available. Migration rates may be especially useful to estimate potential losses to follow-up in a longitudinal study.

For detailed planning and conduct of a trial, a special enumeration will usually be necessary. The population may be enumerated at the same time as the intervention is being started or as a separate exercise in advance. The decision regarding which to use will depend on the specific circumstances of the trial. In the rest of this chapter, the census is assumed to take place shortly prior to the start of the intervention, but the basic principles of enumeration are similar whenever it is conducted.

The initial census may be the first formal contact that most members of a population have with the trial team, though it should have been preceded by liaison of the trial organizers with local officials and local leaders (discussed in Chapter 9). The enumeration exercise provides an opportunity to explain the aims, objectives, and procedures to be used in the trial. For example, an information sheet or newsletter might be left with each household explaining key issues, announcing community meetings where the trial will be explained in more detail, and giving contact details for further information.

Although adequate time needs to be allocated for enumeration and mapping, these tasks should be conducted fairly rapidly to minimize the amount of migration, including from one house to another house within the study area, during the course of the census. The aim of a census is to enumerate the resident population as completely as possible, so the timing of the census is often very important. In areas where there is seasonal migration, the census might be planned for a period when most people are at their normal residence, and, in some populations, trading seasons and market days should be avoided. It may also be important to avoid the rainy season when areas may be inaccessible or the harvest season when people may spend most of the day away from their homes working in their fields. In urban areas, weekends may be the best time for surveys, since, during the week, a high proportion of people may be at work. The time of day may also be important. In some areas, it has been found best to conduct a census after dark, when people have returned from work, but this may not be acceptable or safe in other settings.

It is tempting to try to collect as much information as possible about the study population during the initial census such as information on education or fertility histories. In the interests of speed, however, it is usually preferable to collect such information in a separate round of interviews after the initial census.

Once they have been entered into a computer, data from the census may be used for printing questionnaires, lists of children, and so on, which will aid subsequent surveys (see, for example, Schellenberg et al., 2001).

To conduct a census, a house-to-house enumeration is necessary in most populations. In densely populated villages, with only a few items of data being collected for each individual, a fieldworker going from house to house might be expected to complete census schedules for about 200 people in a day. The number of households this will comprise

will depend upon the population structure. In less densely populated areas or with a longer census schedule, 50 persons a day might be a realistic target (see also Chapters [14](#) and [20](#)).

3.2. Pre-testing

The design and testing of questionnaires, including their pre-testing and pilot testing, whether developed for use with pen and paper or on mobile electronic devices such as mobile phones, tablet computers, or PDAs, are discussed in Chapters [13](#) and [14](#). This process will involve several steps, from initial drafting and pre-testing to pilot testing under field conditions on, say, between 50 and 200 households. Field testing will provide an opportunity to train and evaluate the performance of staff and may assist in the identification of those suitable to become supervisors for the main enumeration.

3.3. Recruitment and training of field staff

Guidelines for the recruitment of staff are given in Chapter [16](#). Training in census techniques is a good way of introducing staff to field research methods. Following instructional ‘classroom’ sessions, trainees should practise conducting a small census themselves.

3.4. Mapping

While a population census can be conducted without a detailed map being drawn, for many trials, especially large ones, or when the trial will last several years, they will greatly benefit from maps being drawn of the study area. These can be used for planning and conducting an initial census, for subsequent house-to-house surveys, and/or for following up participants, but also for displaying trial results and for spatial analyses. The type and accuracy of mapping will depend on how maps are to be used, but there are two main types: paper maps (either official or hand-drawn) and digital maps.

The simplest mapping is the use of existing maps from Departments of Lands and Surveys (or their equivalents) and from special sources such as the Army, Agriculture Departments, Tourist Offices, and the Central Statistics Office (for example, maps that were specially drawn for a national census). These maps may provide enough information for the trial to be carried out without the need for further mapping. More likely, they will form the initial starting point for additional mapping.

While existing paper maps and hand-drawn maps may supply all the information required, digital maps provide far better functionality. For example, if looking at the relationship between a population and a water source or access to services, distances can be calculated quickly and easily, using digital mapping software.

Digital maps do not need to be expensive or complicated, and modern Internet mapping sites (for example, <https://maps.google.com> or <http://www.openstreetmap.org>) may provide maps of sufficient resolution to identify individual houses, streams, and tracks. Where data are missing, a global positioning system (GPS) device can be used to record the location of each household, uploading this information to a computer. These Internet mapping packages allow simple maps to be produced but have very limited scope for spatial analysis. If any spatial analysis is going to be carried out or in order to provide more flexibility with the mapping, dedicated mapping software is required. There is an increasing amount of both commercial and open source or freeware mapping software available such as ARCGIS (<http://www.esri.com>), MapInfo (<http://www.mapinfo.com>), and Quantum GIS (<http://www.qgis.org>).

GPS devices use signals from at least three satellites orbiting the earth to give the longitude and latitude of the hand-held device. The accuracy of the positioning depends upon the number of satellites from which a signal can be received and the strength of their signals. Usually, the accuracy is to within 20 metres, but, in open areas, with single-storey buildings or huts, it can be to within less than 10 metres, while, in areas with poor satellite coverage or where it is heavily forested, it can be worse than 50 metres. There are many ways to collect GPS data, including specific GPS receivers, data loggers, and modern mobile phones. The choice of which to use depends on how the GPS data are to be collected and used. If GPS data will be collected at the same time as other survey data, a GPS-enabled data logger may be most efficient. However, if the mapping is to be done as a separate exercise, dedicated GPS receivers are more cost-effective. Most GPS receivers can store several hundred ‘waypoints’ (for example, households or other points of interest for the map), which can be uploaded into computers at the end of each day’s work. The cost of a simple GPS receiver is around \$100. Commonly used systems are produced by Garmin (<http://www.garmin.com>), Magellan (<http://www.magellangps.com>), and Trimble (<http://www.trimble.com>).

When either paper or digital maps are obtained, the information recorded may be incomplete or inaccurate, and it should be checked in the field. Names of villages may have changed or they may be known by different names locally, and villages and households may have been abandoned or been newly formed if the maps are not recent. Checks, and alterations as necessary, should be made on the positions of roads and tracks, health facilities, schools, official offices, markets, churches, mosques, bars, shops, hotels, boreholes, and other locally important features.

In field trials, the first time that a map is likely to be needed is for planning the baseline survey. In longer-term field trials where houses will be revisited, individual houses are usually mapped. It is good practice to assign a code number to each house on the map. This may consist of a location (for example, village), code (for example, village BS), and a number to indicate the house within that location (for example, BS374). If it is locally acceptable, the number can be painted on the house or fixed to a board (take care: numbers painted on mud walls may be washed off in the rains or painted over, and boards with numbers can be taken down and moved by the residents to a new house!). This helps to ensure that each house is only mapped once and as a quick check on arrival at a house.

The numbering system should be designed to take account of the local family structures and their living arrangements. For example, in studies in some parts of Africa, the same number might be assigned to all houses that comprise a 'compound' where extended family members live. This is not always straightforward to do and is discussed further in Section 4.2.

Figure 10.1 shows part of four trial clusters of a large vitamin A trial in the Kintampo area of central Ghana (Kirkwood et al., 2010). The map was produced using ARCGIS software and shows roads, paths, schools, a hospital, a market, a refuse site, and two communal latrines, along with the location of each compound (identified with a 4-digit number).

Once each house or compound has been mapped and assigned a code, fieldworkers can use either a printed or digital map to locate the households that they need to visit. If small numbers of fieldworkers are involved, the list of households to be visited can be uploaded into a GPS receiver, and a 'GO TO' function used to direct the fieldworker to the location of the house. While these methods may not be exact, they can save large amounts of time.

Once a census has been carried out, the combination of the map and household population data can be used to delineate trial clusters or fieldwork areas. If the households have been mapped digitally, there are functions to allow this to be done manually or using an automated method. Users simply specify the number of people required in each cluster, and either the user or the computer will group houses together to form clusters or groups of the appropriate size. Once fieldwork starts, maps can be printed out, as required, or displayed on a hand-held computer to report on progress.

Maps are also very useful for dissemination of trial results and for community engagement. Because they can display data in a visually striking way, maps, if used well, can have a much bigger impact than other methods of displaying results such as tables or text. They can also be used at routine staff meetings during the trial, such as to display which areas still need to be surveyed or to highlight where unusual results have been recorded.

In many field trials, only simple mapping is required, but the more data that are available, the more spatial analyses can be carried out. The two commonest ways maps are used in analysis are spatial overlays and for calculation of distances. Many health outcomes have a spatial relationship to a risk factor, for example, schistosomiasis to water sources, or malaria to swamps, elevation, and climate. Here, the 'exposure' to these risk factors can be calculated, using Geographical Information System (GIS) software. This requires two geographical datasets: one for the population data and one for the risk factor data. Often, the risk factors, such as rivers and lakes, are collected as part of the mapping process. In other cases, datasets of vegetation type, rainfall, and elevation are available online or from satellite images. In simple cases, these 'layers' can be overlaid to link the population to the risk factor, for example, what the elevation, mean daily temperature, or annual rainfall at the location of each house is. If useful, the results from a regression analysis of such overlays can then be fed back into the computer mapping software to produce risk maps. There are very good examples of such risk maps for infectious diseases such as soil-transmitted helminths, trachoma, or malaria, for which global atlases have been produced (<<http://www.thiswormyworld.org>>, <<http://www.trachomaatlas.org>>, <<http://www.mara-database.org>>).

The other spatial analysis that is commonly used is for calculating distances, for example, from a house to the nearest river or to the nearest health facility. This type of analysis is widely used to investigate access to services. An example of this in a multisite community-based social mobilization trial related to HIV counselling and testing in South Africa is given in (Chirowodza et al. 2009).

Computer mapping and spatial analysis are increasingly being used in trials, and the methods available are constantly being improved and refined. For example, satellite imagery is increasingly being used to plan surveys, as this does not require someone to physically visit and locate each house, in order to create the map. It is possible to use the images provided by sites, such as Google Earth, to mark the location of each structure in the trial area. Once all structures are marked, these can either form the basis for a full survey or a random selection of structures can be selected and surveyed. In some cases, the approximate population can be estimated by multiplying the number of structures by a population per structure estimate. These methods currently tend to be used by research groups with relatively advanced GIS expertise but will increasingly be used more widely, as user-friendly software packages are developed.

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 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.

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/>. 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 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. Next 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.

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).

The age of adult men can also be estimated using event calendars, but there are fewer cross-checks, such as menarche or parity, to confirm the approximate date of birth. Even in traditional societies living in rural areas, adult males may have dated documentation, such as voting cards, military service papers, and other official papers, which may include age information. As for children, if the age of some adults can be determined accurately, that of others may be estimated in relation to those of known age by asking if any of them attended circumcision ceremonies together, grew up together, played together, or went to school together. If not, perhaps they did so with the older brothers or sisters of the index individuals, and so on. Interviewers must review their data to check that the age information derived is plausible. If data are collected digitally, such checks can be made automatically on entry and time-consuming errors avoided. For example, a woman born in 1935 could not have had a child in 1942; in many societies, it is very unlikely that a woman would be 10 years older than her husband, and birth intervals of less than 9 months are uncommon and those of less than 7 months are not possible. Interviewers and supervisors must be trained to check for such inconsistencies.

Well-known problems with age reporting are age 'heaping' and 'shifting'. Age heaping refers to terminal digit preference—the tendency for ages to be recorded as 10, 20, 30, 40 years or 25, 35, 45 years, and so on. Interviewers should be made aware of this during their training. However, there are examples where such training resulted, at the end of the census, in there being too few individuals with ages ending in 0 or 5! Such effects may be of no great

consequence for older adults, for whom precise age estimation is rarely necessary. Age shifting is more difficult to detect and is commonest where age is a criterion of social status. Individuals may falsify their ages, so that they will appear to interviewers to have higher 'age status'. Conversely, women who have not yet married may falsify their age downwards, as may young men who wish to avoid taxes or military draft. In Muslim communities, the ages of women may be especially difficult to estimate, as women may be secluded and the men may respond on their wives' behalf.

4.10. Other identifying information

In some countries, a full name and date of birth are usually sufficient to identify a person. In many others, this will not be enough, but the addition of the individual's parents' names, place of residence, and their relationships to other household members is likely to be sufficient. Individuals may be issued with an identity number by the state or with a social security number that they keep throughout life, and these should be recorded, whenever possible. For trials involving adults, it may be worthwhile to take photographs of all those registered and give each person a laminated photo-ID card that also includes their trial number, with a copy of their photo kept in the trial records. In some trials, involving long-term follow-up of large populations, hand-, foot-, or fingerprints may be used to check identities. This last method was used in a large BCG trial against TB in South India (*Tuberculosis Prevention Trial Madras, 1979*), and also in a large study to assess whether vaccination with hepatitis B vaccine shortly after birth protects against liver cancer in adult life (*The Gambia Hepatitis Study Group, 1987*). There are now several commercially available digital print scanning and reading devices that can be used for this; some are combined with hand-held digital data entry devices that can be used for the completion of census and other forms in the field.

5. Processing of census data

Most censuses involve the collection of substantial amounts of data. It is important to plan how these data will be processed, before the study is started. Usually, it will be desirable that the information is either entered electronically on collection into a PDA, tablet computer, or mobile phone or is entered into a computer shortly after it is collected, so that a large backlog of work does not accumulate. Rapid data entry and checks for transcription errors are especially important if the information collected at the census is to be used to produce forms for the recording of additional procedures to be performed on the trial participants shortly after the census. Furthermore, once the information is in a computer, consistency checks can be conducted, and errors or queries referred back to the relevant fieldworkers. Such feedback should occur as soon as possible after the original information has been collected.

In recent years, there have been major advances in computer systems and transmission of digital information, enabling the collection, processing, and checking of data virtually anywhere, eliminating many of the bottlenecks that, until quite recently, so slowed analysis of trial data in LMICs. Nevertheless, data management is a major task in all field trials and requires a well-defined data management strategy, as discussed in Chapter 20. The design of the recording system may need to allow for changes in the composition of the study population over time, due to in-and-out migration, and for movement between households within the population. It is usually desirable to seek help and guidance from an experienced statistician or data analyst for these aspects. This should be done at the start, rather than in the middle, of a study.

6. Post-enumeration checks and quality control

As discussed in Chapter 20, SOPs should be drawn up for QC at all stages of a trial, and, since the mapping and census are usually the first major field data collection stages of a trial, it is usually these steps that require the most preliminary pre-testing and pilot testing of all procedures, including post-enumeration checks and QC. After the census, the list of the population can be checked against other sources of information on the population. For example, school attendance records can be compared to the eligible age bands in the census, and the information collected can also be compared with other census data from the Central Statistics Office or elsewhere. Population pyramids can be drawn to see if there are any unusual features such as age heaping or disproportionate numbers of individuals at certain ages. Sex ratios can be checked, though one must allow for selective migration of certain age groups and of males or females.

7. Keeping the census up to date: demographic surveillance

In some trials, the enumeration of the population at the start of the study is all that is required, and there is no reason to monitor the population 'continuously' for births, deaths, and migration. In other trials, however, a system of

registration of vital events may be required. This is usually known as demographic surveillance. A good source of advice on how to do this is available at <<http://www.indepth-network.org>>.

After the initial census and when the intervention has started or been applied, follow-up surveys may be required to assess the effects of the intervention. For diarrhoeal or respiratory illness episodes, weekly or twice-weekly visits may be required, whereas, for deaths, annual or quarterly surveys may be adequate. These visits provide an opportunity to update the census by ascertaining births, deaths, address changes, and migration into or out of the study area.

Maintaining an up-to-date population database in this way is a major undertaking. It requires good organization, especially in areas with substantial migration such as in peri-urban slums. For example, in a study carried out in southern Brazil, one half of the families with young children changed address within 2 years (Barros et al., 1990). It may be difficult to conduct long-term follow-up studies in such populations.

A census is relatively easy to update if a computer listing is available, either on paper or on a digital device, which gives the names of the residents in each household at the previous survey, with appropriate spaces for updating information (for example, see Stephens et al., 1989). Pregnant women should be noted, so that, in the next survey, enquiries may be made about the outcome of that pregnancy. Maps should be updated, marking any new or abandoned houses. To obtain reasonable information on births and deaths, the maximum interval between surveys should not exceed a year and preferably will be less—ideally every 3–6 months.

The recording of deaths occurring in the population is usually of special interest. Information on these may be obtained by employing ‘village informants’ to notify the trial investigators when deaths occur. Information may also be available through health facilities, religious institutions, or cemetery records. Usually, it will be necessary to supplement this information with periodic re-surveying of the population if complete ascertainment of such events is required. Deaths tend to be missed, unless specific questions are asked about each individual who was registered in the last round of the fieldwork, and stillbirths, neonatal, and infant deaths may well be missed, unless full demographic surveillance with frequent survey rounds is employed. Such questioning must be done with sensitivity, and the responses may need to be interpreted in the light of any local taboos against speaking of the dead.

References

- Barros, F. C., Victora, C. G., and Vaughan, J. P. 1990. The Pelotas (Brazil) birth cohort study 1982–7: strategies for following up 6,000 children in a developing country. *Paediatric and Perinatal Epidemiology*, **4**, 205–20.10.1111/j.1365-3016.1990.tb00639.x [PubMed: 2362877] [CrossRef]
- Chirowodza, A., Van Rooyen, H., Joseph, P., Sikotoyi, S., Richter, L., and Coates, T. 2009. Using participatory methods and geographic information systems (GIS) to prepare for an HIV community-based trial in Vulindlela, South Africa (Project Accept-HPTN 403). *Journal of Community Psychology*, **37**, 41–57.10.1002/jcop.20294 [PMC free article: PMC2747527] [PubMed: 19774224] [CrossRef]
- Gupte, M. D. 1999. Vaccine trials in leprosy—Venezuela, Malawi and India. *International Journal of Leprosy and Other Mycobacterial Diseases*, **67**, S32–7. [PubMed: 10700935]
- Jukes, M. C., Pinder, M., Grigorenko, E. L., et al. 2006. Long-term impact of malaria chemoprophylaxis on cognitive abilities and educational attainment: follow-up of a controlled trial. *PLoS Clinical Trials*, **1**, e19.10.1371/journal.pctr.0010019 [PMC free article: PMC1851720] [PubMed: 17013430] [CrossRef]
- Kirkwood, B. R., Hurt, L., Amenga-Etego, S., et al. 2010. Effect of vitamin A supplementation in women of reproductive age on maternal survival in Ghana (ObaapaVitA): a cluster-randomised, placebo-controlled trial. *Lancet*, **375**, 1640–9.10.1016/S0140-6736(10)60311-X [PubMed: 20435345] [CrossRef]
- Schellenberg, J. R., Abdulla, S., Nathan, R., et al. 2001. Effect of large-scale social marketing of insecticide-treated nets on child survival in rural Tanzania. *Lancet*, **357**, 1241–7.10.1016/S0140-6736(00)04404-4 [PubMed: 11418148] [CrossRef]
- Smith, P. G., Revill, W. D., Lukwago, E., and Rykushin, Y. P. 1976. The protective effect of BCG against *Mycobacterium ulcerans* disease: a controlled trial in an endemic area of Uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **70**, 449–57.10.1016/0035-9203(76)90128-0 [PubMed: 841647] [CrossRef]
- Stephens, J., Alonso, P. L., Byass, P., and Snow, R. W. 1989. Tropical epidemiology: a system for continuous demographic monitoring of a study population. *Methods of Information in Medicine*, **28**, 155–9. [PubMed: 2796757]

- The Gambia Hepatitis Study Group. 1987. The Gambia Hepatitis Intervention Study. *Cancer Research*, **47**, 5782–7. [PubMed: 2822233]
- Tuberculosis Prevention Trial Madras. 1979. Trial of BCG vaccines in south India for tuberculosis prevention. *Indian Journal of Medical Research*, **70**, 349–63. [PubMed: 535944]

Figures

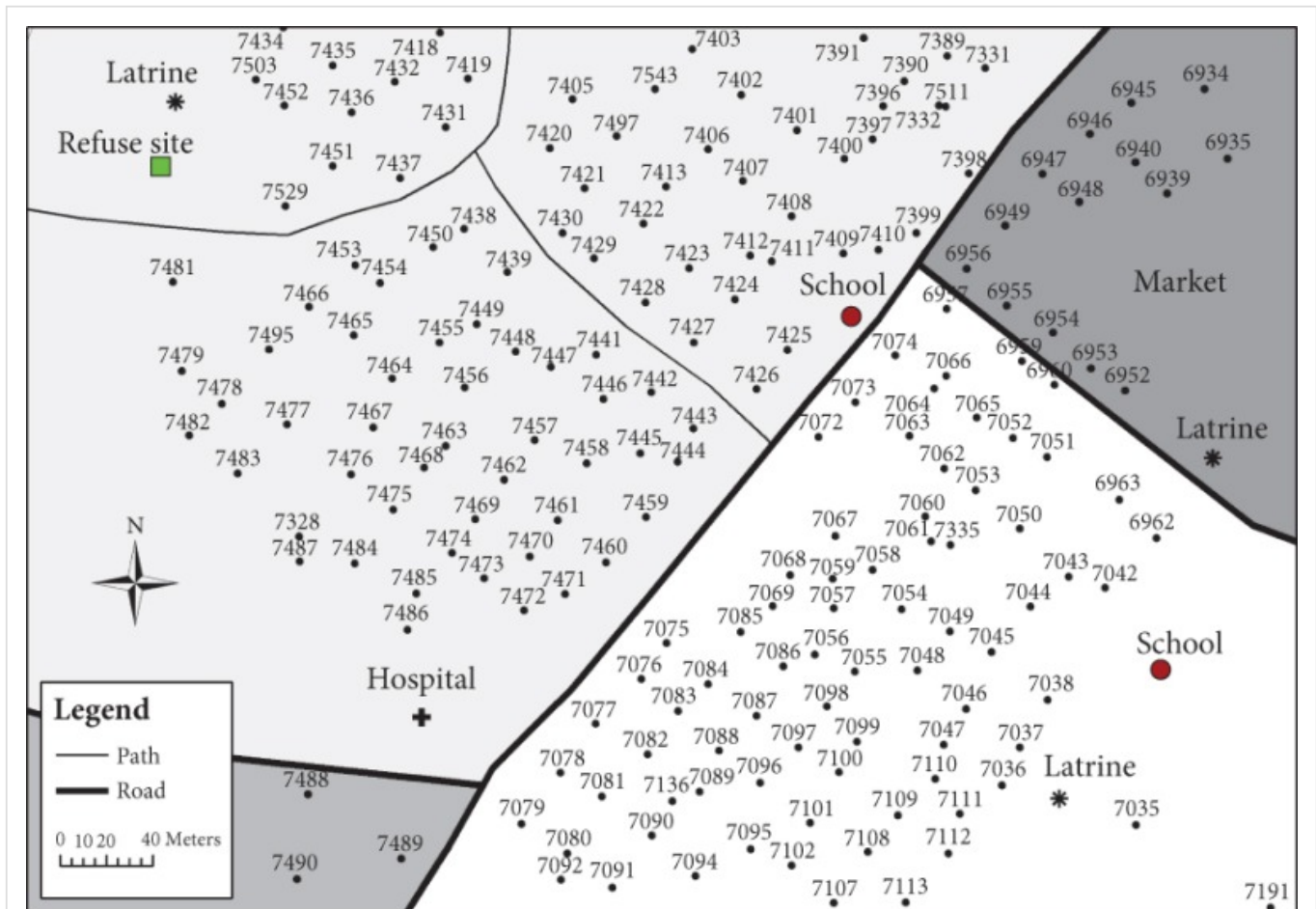
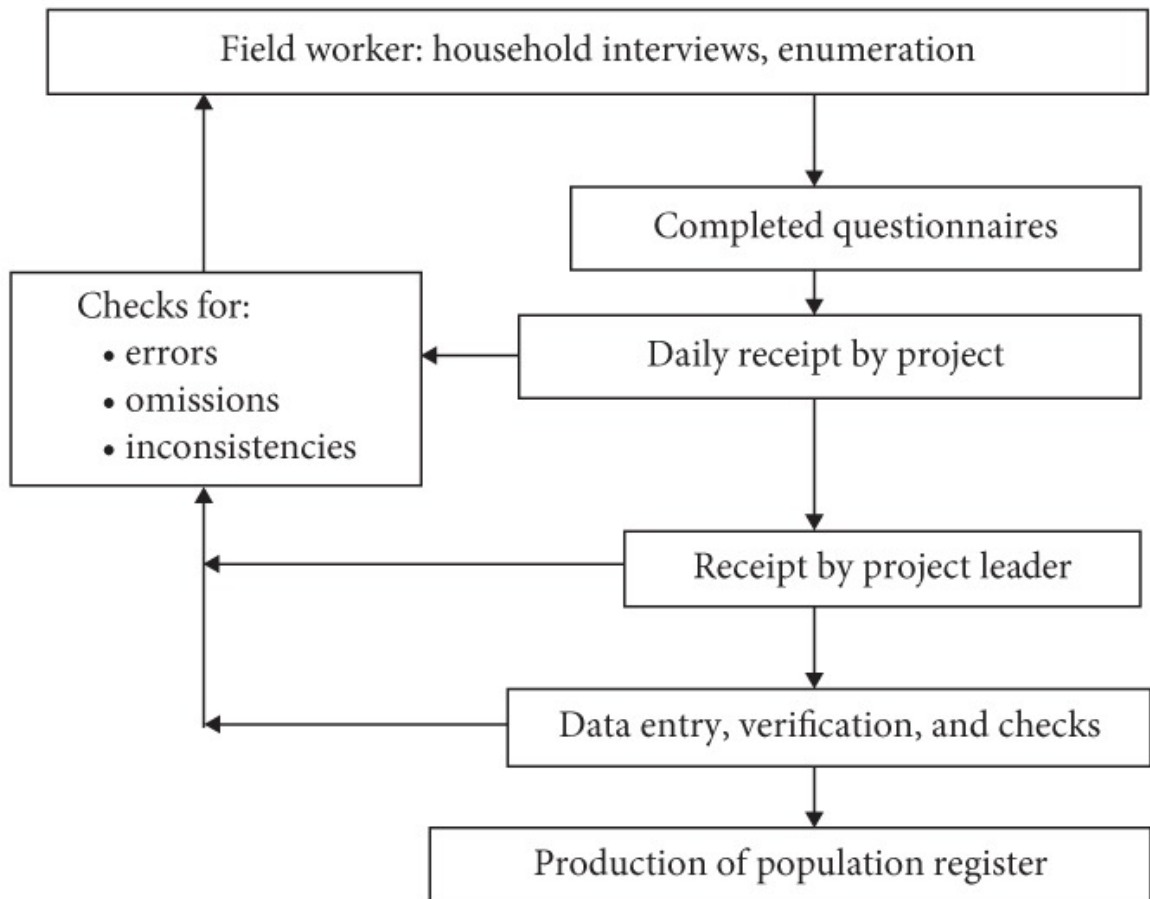
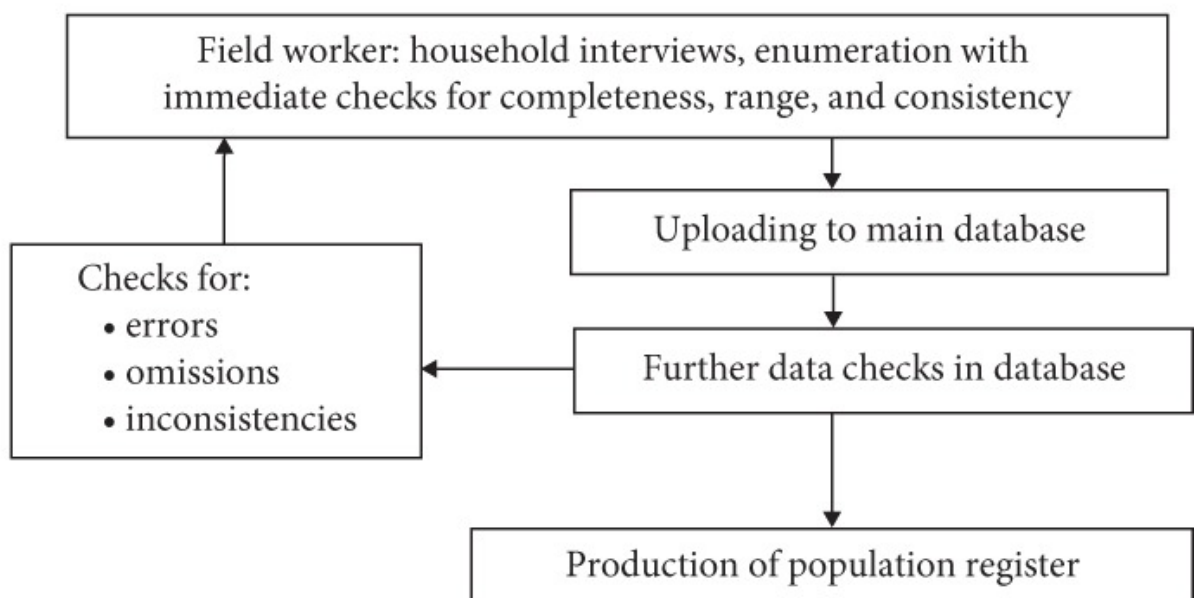


Figure 10.1

Part of a trial map.

Reproduced courtesy of C. Grundy, B. Kirkwood, and S. Owusu-Agyei. This image is distributed under the terms of the Creative Commons Attribution Non Commercial 4.0 International licence (CC-BY-NC), a copy of which is available at <http://creativecommons.org/licenses/by-nc/4.0/>.

(a) Data collected on paper**(b) Data collected on a digital device****Figure 10.2**

Flow chart of census data collection.

Boxes

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:

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:

	d1	d2	d3	d4	d5	d6	c
(Prime	11	7	5	3	2	1)	

Trial number		Trial number with check digit
467913	$4 \times 11 = 6 \times 7 = 7 \times 5 = 9 \times 3 = 1 \times 2 = 3 \times 1 = 153$	4679133
476913	$4 \times 11 = 7 \times 7 = 6 \times 5 = 9 \times 3 = 1 \times 2 = 3 \times 1 = 155$	4769135
567913	$5 \times 11 = 6 \times 7 = 7 \times 5 = 9 \times 3 = 1 \times 2 = 3 \times 1 = 164$	5679134

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

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:

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.

Head of household	H
Sister	SR
Wife	W
Grandson	GS
Son	S
Granddaughter	GD
Daughter	D
Grandfather	GF
Mother	M
Grandmother	GM
Father	F
Other blood relative	R

Brother	BR
Unrelated	X

Person number	Person	Code
01	Head	H
02	Wife 1 of household head	W01
03	Child 1 (M) of household head and his wife 1	S01,02
04	Child 2 (F) of household head and his wife 1	D01,02
05	Child 3 (M) of household head and his wife 1	S01,02
06	Wife 2 of household head	W01
07	Child 1 (F) of household head and his wife 2	D01,06
08	Child 2 (M) of household head and his wife 2	S01,06
09	Mother of household head	M01
10	Visitor	X
11	Child 1 (M) of visitor	S10

Box 10.3 Example of part of a monthly calendar of local events used in the Ghana Vitamin A Supplementation Trial (VAST)

September

Specific dates: none

Farming:	Harvesting of groundnuts, cowpeas, and maize starts
	Harvesting of sweet potatoes, peas, and wet-season rice starts
General:	Heavy rains continue (<i>Dului</i>)
	Drumming and other loud noises banned
	Beginning of the school year

October

Specific dates: 1 to 31 'Rosary'

Farming:	Harvesting of groundnuts, cowpeas, maize, sweet potatoes, <i>pesa</i> , and wet-season rice ends
General:	Rains slackening off
	Season of abundant food (<i>Womodaabu Ch'ana</i>)
	Ban on drumming and other loud noises is lifted

November

Specific dates: 1 All Saints; 2 All Souls

Farming:	Late millet harvest
	Construction of dry-season gardens starts
	Coccidiosis disease (<i>Choguru</i>) tends to start in fowls

General:	Harmattan (dust-laden wind from the north) starts
	Cutting of 'sange' and grass starts
	Firewood collecting season starts
	Frog hunting season starts
	School Nov/Dec exams start

December

Specific dates: 19 *Feok* Festival in Sandema; 25 Christmas Day; 26 Boxing Day; 27–29 *Fao* Festival in Navrongo; 31 Anniversary of the 31 December Revolution

Farming:	Dry-season tomatoes and other vegetables start to become available
	Collection of kapok starts
	Harvest of ebony fruits starts
	Gathering of millet stalks starts
	Storing of grain
	Domestic animals allowed to move about freely again
General:	Harmattan continues
	Making of bricks and repairing of houses start
	Bush fire season starts
	Hunting season starts
	Many Northerners start to move South, looking for farm work
	Christmas school holidays start

Source: data courtesy of D. A. Ross (personal communication).

Box 10.4 Example of part of a calendar of local events used in Ghana Vitamin A Supplementation Trial (VAST)

1900 (approx.)	◆ War with Zabog people from Burkina Faso
1906	◆ Founding of the Catholic Mission in Navrongo (by Father Oscar Morin)
1908	◆ First Kassenas enrolled for Catechism training
1913	◆ Baptism of the first Kassenas in Navrongo
1916	◆ First conscription of local people into the British army for the First World War
1918	◆ Collection of mats from each household for roofing of houses for British people to stay in
1919	◆ Bad plague of locusts
1977	◆ Introduction of First Phase of Junior Secondary Schools
1978	◆ Achaempong overthrown by General Akuffo (5 July)
	◆ Change of currency notes (50 cedi note added)
1979	◆ J. J. Rawlings first came to power (4 June)
	◆ Shooting of Colonel Felli and others by firing squad
	◆ Elections for the Third Republic (Hilla Liman elected)

1981	◆ PNDC revolution (31 December)
1983	◆ Year of drought, bush fires, caterpillars, and food shortages
1984	◆ Bumper harvest
	◆ Cancellation of all O- and A-level results throughout West Africa
1985	◆ Major dust storm, when it was dark all day (13 March)
	◆ 25th anniversary celebrations of Navrongo Secondary School (Navasco) and Notre Dame Secondary School
	◆ Start of the Mamprusi/Kusasi War in Bawku
	◆ Fighting between Saboro and Wusungu started (Nov–Dec)
1986	◆ Good harvest
	◆ Heavy rain storm which destroyed part of the Bolgatanga–Navrongo road
1987	◆ Introduction of Second Phase of Junior Secondary Schools (September)
	◆ Ritual murder of an old man in Navrongo
1988	◆ Start of armyworm invasion (June)
	◆ Ordination of three local men to the RC priesthood in Navrongo (23 July)
	◆ The bodies of three Kassenas who had been killed in a road traffic accident in Nigeria were brought back (August)
	◆ Very heavy rain storm and floods with many houses destroyed
	◆ President J. J. Rawlings' visit to Sandema (15 December)
	◆ First <i>Navro Fao</i> Festival celebrations for many years (27–29 December)

Source: data courtesy of D. A. Ross (personal communication).

© London School of Hygiene and Tropical Medicine 2015.

This is an open access publication. Except where otherwise noted, this work is distributed under the terms of the Creative Commons Attribution NonCommercial 4.0 International licence (CC BY-NC), a copy of which is available at <http://creativecommons.org/licenses/by-nc/4.0/>. Enquiries concerning use outside the scope of the licence terms should be sent to the Rights Department, Oxford University Press, at the address above.

Monographs, or book chapters, which are outputs of Wellcome Trust funding have been made freely available as part of the Wellcome Trust's open access policy

Bookshelf ID: NBK305507