Chapter 4   Trial design

1. Introduction to trial design

Trials should be designed to produce unambiguous estimates of the effects of interventions, which are precise enough for public health planning. A common goal of all intervention studies, including trials, is to evaluate the effect of a specific intervention (or a specific package of interventions) applied in a specific manner to a well-defined population. In the trial design, the major issues will be: (1) the nature of the intervention, the strategy for its implementation, and the natural size of the unit at which the intervention is applied (for example, individual, household, school, village, district); (2) the likely effects, including possible adverse effects, and how they should be measured; and (3) the comparisons that need to be made with other interventions.

In most LMICs, disease control is the responsibility of the Ministry of Health (MOH). Therefore, wherever possible, the Ministry should be involved in the planning and monitoring of trials, and the results must be made available in such a way that they are of direct relevance to national disease control activities (see Chapter 23). As the Ministry is often the implementing agency for interventions in public health programmes, it is generally desirable that independent investigators actually conduct the trials of interventions.

This chapter gives an overview of the main factors to consider in the development and implementation of health intervention trials in LMICs.

1.1. Planning a trial

The trial planning process is a major exercise which starts, and which should be largely completed, before any field activities have taken place, other than initial feasibility studies and small-scale pilot investigations (see Chapter 13). The planning process should encompass all aspects of the trial, from formulation of detailed objectives, based on the initial idea, through preparation for all field activities, collection of data, and analysis of results, to their publication, dissemination, and potential use in disease control. The plan should also try to anticipate the form of any studies that will follow, depending on the possible different outcomes of the trial.

Detailed planning is necessary for several purposes. First, information on the trial will be required by local and national administrations for them to review as part of the trial approval process. A similar description will be required by any agency that is going to review the proposal for funding. The detail required in such grant applications varies greatly from agency to agency. Some require a comprehensive document with full details of all trial procedures, while others put quite a small upper limit on the size of any application they are prepared to review. It is usually more time-consuming to prepare the former kind of application, but the latter kind may present a more formidable challenge, because, in relatively few words, the investigators have to present convincing evidence that they have considered and worked out all issues that would have been included in the longer type of application. Advice on the preparation of grant applications is given in Chapter 8.

A second reason for detailed planning at the start of an investigation is that possible problems must be anticipated in advance and solutions thought through, in order to reduce the likelihood of the trial falling behind schedule or having to be radically changed or abandoned, due to problems that could have been foreseen and avoided. Commonly, funding agencies require a section on potential risks to the trial, in which the investigators are asked to specify what could go wrong and the consequences this would have for the trial. It is rare to be able to predict all potential problems, but the more that have been considered in advance, the smaller the chance of catastrophe.

Realistic estimates must be made of the resources needed (for example, for transport, staff salaries, allowances, items of equipment) and the likely trial duration, including the time to analyse and report the trial, in order to be able to calculate the required budget for the trial. Underestimating the support needed may jeopardize some of the objectives, which may have to be revised or abandoned in the middle of the trial, whereas overestimating the cost may prejudice the funding agency against agreeing to support the trial. It is tempting to underestimate costs in the hope of increasing the chance of funding, but this may be self-defeating and, in any case, will often be picked up by the experienced investigators asked to review the trial proposal by the funding agency. The time it will take to conduct and analyse a

trial is also often underestimated, particularly for trials where implementation of the intervention, or package of interventions, is not directly under the control of the evaluators but depends instead on the MOH or other partners. Advice on the preparation of budgets is given in Chapter 18.

In the present chapter, the steps to be included in the trial plan are discussed in the approximate order that they would arise, from the formulation of objectives through to the eventual publication, dissemination, and use of the findings. In the remaining chapters, specific issues relevant to the planning process are reviewed in greater detail, and cross-references are given in this chapter, where appropriate.

1.2. Ethical considerations in designing a trial

Ethical considerations impinge on many aspects of the design and conduct of trials and are discussed fully in Chapter 6. Briefly, any research investigation that involves human subjects should be submitted for ethics committee review. Intervention trials in some communities in LMICs may pose specific ethical dilemmas. The dogma that an investigator 'should treat everyone in the trial as though they were a member of his or her own family' is both difficult to apply and often inappropriate in situations of extreme poverty, in which some trials in LMICs will take place. Related issues concern the responsibility that an investigator has to those who live in the same community as the trial subjects but who, for whatever reason, are not included in the trial, and what happens regarding the public health use of an intervention after a trial has shown an intervention to be efficacious. Very commonly, an investigator must walk a tightrope, balancing his or her responsibilities to the individuals in the trial with those related to the potential of the interventions being evaluated to improve public health. The MOH knows these problems well, as they are implicit in any allocation of the health budget between the various potential preventive and curative services, but, commonly, the officials allocating the routine health budget are several steps removed from the individuals and communities that their decisions will affect. The field trial researcher usually has to face these issues directly. There are no simple solutions to these problems. It is important that each research study is subject to strict ethical review, with due attention to the specific conditions in and under which it will be conducted.

1.3. Trial governance

Since the first edition of this book was published, there has been a much greater emphasis on trial governance and quality control (QC) in trials. There are now extensive international guidelines on the governance of clinical trials, in which the roles of bodies, such as the trial 'sponsor', the principal investigator (PI), the trial Steering Committee, and the Data and Safety Monitoring Committee (DSMC), are discussed and defined. These aspects are considered in more detail in Chapter 7.

2. Definition of trial objectives

Once an idea for a trial has been formulated, it will be necessary to detail the specific objectives of the trial. To do this, the researcher will need to find out what has already been done regarding the evaluation of the intervention or interventions of a similar kind. This may involve meeting or corresponding with those undertaking similar studies, and it will almost invariably involve conducting a systematic literature review to find out what has been published that is relevant (see Chapter 3).

With this background information, the objectives of the trial can be formulated. These should include the overall aim or purpose of the trial, such as 'to evaluate the efficacy of a specific microbicide gel for the prevention of HIV infection in women' or 'to measure the impact of a breastfeeding promotion strategy on the incidence of diarrhoeal diseases in infants'. The specific objectives give more detailed statements of the particular questions that the trial is designed to answer, or the hypotheses that it will test. Finally, a list of subsidiary objectives may be given which relate to issues which are not central to the overall objectives but about which information will also be gathered while the trial is in progress.

2.1. The idea for a trial

One of the most creative phases of the planning of a trial is the selection of the subject area of the research and the formulation of the specific questions that will be addressed. A major motivation for most successful researchers is that they are doing something that they really enjoy and are researching questions about which they feel passionate. Their motivation may come from scientific curiosity about the causes or treatment or control of a particular disease, or about the effects of a specific intervention, or their concern may be to explore different ways that health or social

systems can improve the public health. The field researcher may be motivated by working directly with people in their communities and be stimulated by the challenges posed by working in remote or difficult situations, outside of the hierarchy that may exist, for example, in a hospital environment.

The development or refinement of an idea for a field trial should take place in interaction with others at local, national, and possibly international levels. The research activity must not only be acceptable to the population in which it will be undertaken, but also to those who will authorize it nationally and to those who will fund it. Most good ideas for field research on the control of a disease that is of public health importance are likely to attract support.

Field research likely to receive the highest priority, both nationally and internationally, is that directed at control of diseases of greatest public health importance. An important preliminary to the development of a research proposal on a specific disease or condition may be a survey in the local community to determine the importance of the disease of interest. Such local data might be presented side by side with estimates of the global burden of disease attributable to the condition being studied.

The progress of science (and of public health) is not only dependent on groundbreaking first trials that show that a new intervention can be effective in one context. Progress also requires the replication of such trials in different settings to determine whether the findings from the original trial may be generally applicable. Replications of trials of bacille Calmette–Guérin (BCG) vaccination against TB and leprosy and of rotavirus vaccines, for example, have shown substantial variations in the efficacy of the vaccines in different parts of the world. This is even more important for effectiveness trials of interventions that are delivered through routine services where results may show important variations from one location to another, due to contextual differences. Although sometimes disparagingly called 'me too!' trials, such confirmatory (or otherwise!) trials are very important for the assessment of the public health usefulness of an intervention in a specific context.

A trial may either test for superiority or for equivalence. The choice will depend on the nature and effectiveness of the comparison intervention and has important implications for the choice of trial size (see Chapter 5). For example, if the aim is to test whether a new drug for the treatment of visceral leishmaniasis is more effective than the standard drug treatment, this will require what is called a 'superiority' trial. However, it could be that the new drug is much cheaper or is thought to have fewer side effects. If this was confirmed in a field trial, it would be likely to be adopted even if it was no more effective than the standard drug, so a trial that is designed to test for 'non-inferiority' or 'equivalence' would be appropriate.

2.2. Trial purpose

The statement of the purpose of a trial (termed 'goal' by some agencies) should convey to the reader the type of intervention, or package of interventions, to be evaluated (without details of how it will be applied, dose, and so on) and the endpoints against which the impact will be measured, without necessarily specifying the magnitude or precise nature of the impact expected or which the trial will be designed to detect. It may also include a description of the ways in which the results of the trial may influence public health policy and contribute to scientific knowledge. For example, in a trial of the use of the drug ivermectin against onchocerciasis, the statement of the purpose might be 'to assess the impact of mass treatment with ivermectin on the transmission of onchocerciasis and to measure any side effects in those treated with the drug'. For a trial of a new vaccine against the blood stages of the malaria parasite, the purpose may be 'to measure whether a *Plasmodium falciparum* asexual blood stage vaccine reduces episodes of clinical malaria'. For a trial to test the effect of cash payments conditional on girls either staying in, or returning to, secondary school on their risk of HIV infection, the purpose might be 'to assess whether educational conditional cash transfers reduce acquisition of HIV infection in girls'. Finally, for the example of the equivalence trial of a new drug for visceral leishmaniasis treatment, the purpose might be 'to test whether the new drug is at least as effective as the standard treatment for treatment of visceral leishmaniasis'.

2.3. Specific objectives of the trial

In the specific objectives (called specific aims by some agencies), a quantitative statement should be made regarding the size of the effect of an intervention that a trial is designed to detect and the precision with which the effect will be measured. Such specifications are necessary in order to calculate how large a trial should be, using the methods described in Chapter 5. The nature of the intervention should be given in more detail than in the statement of purpose (for example, dose and frequency of administration), and the endpoints of the trial clearly stated. They should also include a specification of the size of the trial and detail the population in which the intervention will be applied. For

the example of the trial of ivermectin against onchocerciasis, the specific objectives would include a statement of the size of the impact on transmission which the trial would have a reasonable chance of detecting and the frequency with which adverse reactions of different kinds would have to occur to be detected in the trial, while, for a malaria vaccine, a more detailed description of the formulation of the vaccine would be required and statements included on the magnitude of the true effects on the incidence of malaria that the trial would be very likely to detect as being statistically significant. Finally, for the conditional cash transfer trial (see Section 2.2), the specific objectives should state the size of payment, to whom it will be given (for example, to the girl herself, her parents, or some combination of the two), the age range of the girls in the trial, and the size of effect on HIV incidence that the trial would have a reasonable chance of detecting.

The proper specification of the specific objectives is crucial to a successful trial. They should include a concise, but detailed, description of the intervention to be evaluated, the outcome(s) of interest, and the population in which the trial will be conducted. The more specific and detailed the objectives are, the clearer it will be how to design a study to meet them. It is crucial to set appropriate objectives, and it is worth spending time to get these both correct and unambiguous.

2.4. Subsidiary objectives of the trial

In the context of many trials, there will be secondary endpoints which will be measured in the trial but which are not the prime purpose for which the trial is conducted. Also substudies may be included, having subsidiary objectives, such as the comparison of various serological tests or the analysis of genetic markers and their correlation with disease. It may be decided to add other objectives on to an intervention trial which do not relate to the main objectives. In the trial of ivermectin against onchocerciasis, for example, the impact on some other parasitic diseases might be assessed.

To increase the plausibility of trial findings, it is important to document changes in intermediate outcomes, which are directly related to the outcomes of principal interest, whenever this is possible. This requires laying out an 'impact model' (see also Chapter 15), describing how the intervention is expected to lead to the major outcome being studied. For stand-alone biological interventions, these models tend to be quite simple. For example, a trial of the effect of periodic vitamin A supplementation on child mortality should document that the vitamin A status improved in children receiving the supplement, but not in the comparison group. Impact models for non-biological interventions are often more complex. For example, in the conditional cash transfer trial, the impact on retention in secondary school and school achievement grades or the impact on reported sexual risk behaviours or on the incidence of other STDs or of pregnancy could also be studied, as well as the primary endpoint of HIV incidence. Impact models are essential for deciding which intermediate indicators must be measured.

The introduction of an intervention may also provide a special opportunity for determining particular key factors in the pathogenesis of disease. For example, trials of ivermectin, a microfiliaricide, against *Wuchereria bancrofti* may provide evidence for the role of microfilaria, as compared to that of adult worms, in the pathogenesis of lymphatic filariasis disease. Decisions to add on studies of this kind should not be taken lightly, as they will invariably need additional commitment of resources and may involve the trial population in additional inconvenience. They may thus have a negative impact on the primary objectives, perhaps by overstretching the trial team's technical or managerial resources, and the final 'cost' to the trial may be much greater than it appeared to be in purely monetary terms.

Once a large field trial is successfully under way, it is not unusual for the trial organizers to be approached by other investigators who wish to graft on additional procedures to answer questions of interest to them. There may be considerable value in utilizing the same trial for multiple purposes, but full consideration should be given to the extra work that this will entail, especially for key members of the research team, and to other possible harmful effects such as upsetting the rapport between the trial team and the trial population.

3. Selection of interventions

3.1. Intervention characteristics required

Several criteria should guide the suitability of candidate interventions to be evaluated in a large-scale field trial. The intervention, or package of interventions, should usually be one that could be introduced into a national or regional disease control programme (though this criterion might not apply for 'explanatory' or 'proof of principle' trials—see Chapter 2, Section 3.3). The dose (when applicable) should be 'optimal'. Evidence would usually be required from

smaller preliminary studies (sometimes called Phase I and II trials, particularly with respect to trials of drugs and vaccines) that the intervention is relatively safe and produces a convincing intermediate response, such as a good antibody response to a vaccine or a change in self-reported sexual behaviour for an intervention to prevent unwanted pregnancies.

When an intervention has to be repeated several times to be effective (for example, micronutrient supplements), there should be evidence that the interval between each intervention is appropriate. For some interventions, the concept of dose is meaningless, such as the application of a diagnostic or screening test. Corresponding relevant evidence would then be required that the test is adequate (for example, previous studies indicating that it had good sensitivity, specificity, and predictive values). For continuous or repeated treatments, similar considerations apply to the duration of treatment. For example, with vitamin supplementation, the duration required will depend on whether the outcome of interest is the reversal of the acute effects of severe deficiency or of the chronic effects of more moderate deficiency. In addition to being safe and giving promise of being efficacious, the intervention must be acceptable to those to whom it is directed, relatively easy to deliver, and, at least eventually, of sufficiently low cost that it could be incorporated into the national disease control strategy if it is proved to be effective within the field trial.

3.2. Number of interventions compared

The choice of the number of different interventions to compare in a field trial is likely to be determined not only by the number of competing alternatives, but also by the implications the choice has on the size of the trial. This, in turn, is dependent on the frequency with which the outcome of interest occurs. 'Rare' outcomes require large trials (as discussed in Chapter 5). For example, in a trial of leprosy vaccines in South India, it was planned that each 'arm' (one of the alternative intervention assignments) included in the trial would require around 65 000 trial participants, in order for the trial to have the desired statistical power to detect effects that would be of public health importance (Gupte et al., 1998). Clearly, in this situation, a decision to add another arm would have had enormous cost and logistic consequences.

If the outcome is common, however, trials to compare more than two interventions may be undertaken more readily. For example, if seroconversion following vaccination is the outcome of interest, it may be straightforward to compare multiple vaccines or vaccination strategies in a single trial.

It is important to note, however, that many researchers try to build too many comparisons into a trial. There is often a tendency to divide groups after the sample size has been calculated or to plan comparisons within groups, without going through the appropriate computations (as given in Chapter 5).

Comparisons within a single trial can always be made with much greater confidence than those between trials. Thus, if drug A is found to be 50% more effective than a placebo in one trial and drug B is found to be 50% more effective than a placebo in another trial, it will not necessarily be possible to conclude that A and B are equally effective, as the circumstances in which the two trials were conducted will not have been identical. A further trial may be necessary for a direct comparison of A and B. If the need for this trial could have been anticipated in advance, it would have been more efficient to conduct one trial involving both drugs A and B and a placebo. A trial like this may be more complex to organize and would probably have to be substantially larger than either of the '2-arm' trials but would still tend to be smaller than the sum of the two trials.

When two interventions are being compared to a control intervention, and in situations where it would be possibly appropriate to apply both interventions to the same individual (or community), an efficient way of comparing both interventions with the control arm in the same trial is to design it as a 'factorial' trial. In such trials, some individuals receive the control intervention, others receive one or other of the new interventions, and some receive both interventions (typically 25% in each of four groups) (Montgomery et al., 2003). Although not commonly used, this design is very efficient, unless there is 'interaction' between the two interventions, i.e. the effect of both interventions applied at the same time is different from the simple sum of the separate effects of each of the interventions. Ayles et al. (2008), Awasthi et al. (2013a), and Awasthi et al. (2013b) are examples of the design of such trials.

3.3. Combined interventions

For some diseases, there are several possible interventions that may reduce the disease impact on a population. For example, interventions against malaria include destruction of mosquito breeding sites, spraying of residual insecticide, personal protection measures (for example, use of bed-nets and repellents), drug prophylaxis, and drug treatment, and trials might be designed to evaluate each of these interventions individually. A malaria control programme may

choose to use more than one intervention at the same time and may wish to evaluate the impact of the 'package' of interventions, rather than the individual components of it. In such a case, the trial might compare an integrated strategy incorporating several different interventions applied simultaneously with a control group in which only the routine interventions that were previously available would be applied.

Several trials of this kind have been conducted for the prevention of HIV. For example, a recent trial in Tanzania tested the effectiveness of a package of interventions targeted to young people. Those in the intervention group received HIV prevention education in school; health workers in their local health facilities were given special training and support to try to make their facilities more 'youth friendly'; new suppliers who were thought to be particularly attractive to young people were trained and supported to sell condoms, and annual 'youth health weeks' were organized in their local communities (Ross et al., 2007). The advantage of this kind of trial is that it allows the testing of a package on interventions that might reasonably be expected to have a greater impact than any single component of the package. However, if no effect is seen, then although it may be reasonable to conclude that no one of the components of the intervention (at least, as applied in the trial) would have been effective on its own, it is necessary to think carefully about whether the existence of several concurrent interventions might have diluted the effect of one component on its own, or even that one component might have counteracted the effect of another. Another disadvantage is that, if an effect is demonstrated, it is not possible to be sure of the contribution to the overall result of each of the various components of the intervention.

3.4. Choice of comparison intervention

The best way to evaluate an intervention is to compare its effect with that of another intervention in the same population at the same time. Whenever possible, the allocation of individuals or groups of individuals to the different interventions should be 'at random' (see Section 4.1 and Chapter 11). In general, the intervention that is the current 'best' should be used as the comparison, but the choice of the 'control' intervention is not always straightforward and may involve difficult ethical considerations (see Chapter 6). When no effective intervention is known, the comparison must be with a group in which 'no intervention' is made; ideally, a placebo should be administered in order to preserve 'blinding' (see Section 4.1). For example, before the development of ivermectin no effective and safe treatment for onchocerciasis existed. Thus, placebo-controlled trials of the drug were ethically acceptable, at least until the beneficial effects of ivermectin had been established. For most tropical diseases, however, some kinds of intervention already exist and may already be deployed by the health services or by a control programme in the area where a trial is planned. Only in very rare circumstances would it be ethical to withdraw these existing interventions for the purposes of a trial. A more complex issue is with respect to the extent to which they should be introduced in the context of a trial. It is known that regular prophylaxis with anti-malarial drugs reduces morbidity from malaria, for example, so would it be necessary to give this intervention to all those in the 'control' arm of a malaria vaccine trial, even though, in normal circumstances, very few, if any, of them would otherwise have been on prophylaxis? Indeed, would it even be ethical to withhold prophylaxis from those who would be receiving a malaria vaccine whose efficacy was unknown? The optimistic reader will seek a definitive answer to these questions in Chapter 6! Unfortunately, the search will be in vain, as there are no general definitive solutions to problems such as this; each situation has to be considered on its own merits, taking full account of the circumstances in which a particular investigation is planned. However, in Chapter 6, key principles are outlined that should be used when making such judgements.

In a leprosy vaccine trial in Venezuela, the new leprosy vaccine consisted of a mixture of BCG and killed *Mycobacterium (M.) leprae* bacilli. When the trial was designed, a choice had to be made between using BCG for the control arm (the efficacy of BCG alone against leprosy in Venezuela was unknown at the time) or using a placebo. BCG was chosen, even though doing this might reduce the chance of showing a protective effect (as BCG alone may have been protective). The inclusion of a third, placebo, arm would have allowed the protective effect of BCG alone to be evaluated, but the incidence of leprosy was too small for a third arm to be feasible within the trial. The major purpose of the trial that was conducted was therefore to evaluate whether a leprosy-specific vaccine (i.e. one which included *M. leprae* bacilli as well as BCG) was more effective than a non-specific vaccine (in this case, BCG). If the comparison had been with a placebo instead of BCG, any effect due to BCG could not have been distinguished from that due to the addition of *M. leprae* bacilli to the vaccine. In a larger trial of the same vaccine that was conducted in India, it was possible to include a placebo arm (Gupte et al., 1998).

The use of a placebo may be very important to derive an unbiased measure of effect (see Section 4.1 and Chapter 11, Section 4), but it requires careful ethical justification, and thought must be given to whether particular circumstances might lead to treatment being offered to participants, irrespective of their trial arm. In a placebo-controlled trial of

vitamin A supplementation in Ghana, for example, the objective was to determine if a reduction of child mortality was produced by supplementation. As eye signs of vitamin A deficiency are effectively treated by vitamin A supplements, all in the trial were monitored for such signs and treated immediately if such signs were detected, even though this was likely to reduce the power of the trial to detect an impact of vitamin A supplementation on mortality.

A related issue concerns trials which do not test new interventions as such but evaluate new ways of delivering existing interventions. In a cluster randomized trial in Bangladesh, the Integrated Management of Childhood Illness (IMCI) strategy promoted improved ways of delivering interventions such as antibiotics for pneumonia, oral rehydration therapy, and vaccines; these interventions were also available from routine services in comparison areas. It was judged ethical not to change routine practices in the comparison areas, because these reflected what was already in place in the country as a whole (Arifeen et al., 2009).

3.5. Complex interventions

The design of a trial to evaluate the efficacy of a new vaccine or drug is relatively straightforward, in the sense that there are many past examples of such evaluations to draw upon when planning a new trial. However, the evaluation of some interventions, such as the deployment of a new procedure in the health service or public health practice, may involve consideration of several interacting components, including, for example, educational components and behavioural change. Such interventions pose special problems for evaluation, and these kinds of intervention have been called 'complex'. Many of the extra problems relate to the difficulty of standardizing the design and delivery of the interventions, their sensitivity to features of the local context, the organizational and logistical difficulty of applying experimental methods to service or policy change, and the length and complexity of the causal chains linking intervention with outcome. See Chapter 2, Section 2.3.4 and the associated Box 2.1 for further discussion.

4. Allocation of interventions within the trial

4.1. Randomization and 'blindness'

Once a potential intervention has been shown to be safe and acceptable for use in humans and the dose schedule established, trials should be conducted to evaluate quantitatively the benefit attributable specifically to the intervention under trial, compared to some other intervention, while attempting to exclude the confounding effect of other variables. The best way to exclude the potential effects of other factors—both those already known to be confounders and also those that are confounders but are not known to be so—is to base allocation decisions as to which intervention is applied to a particular individual, or group, on a random process. Incorporation of randomization into the trial is an extremely important design issue (see Chapter 11).

The randomized intervention trial is as close to a rigorous scientific experimental study involving human beings as it is possible to achieve ethically. The main study design features of a randomized trial are:

1. to avoid bias in assignment to the alternative interventions, all eligible trial participants should be assigned at random to the alternative treatment groups. This involves two steps; the first is selecting participants on the basis of the pre-established criteria for eligibility, and the second is the randomization procedures should ensure that each eligible participant has the same chance of receiving a particular intervention procedure

2. to avoid bias in the assessment of the trial endpoints, whenever possible, the person(s) assessing the outcome measures should not know to which intervention group the participant was assigned (i.e. the assessor should be 'blind' to the intervention group)

3. to avoid bias in the behaviour or reporting by the participant, whenever possible, the participant should also be 'blind' (i.e. the intervention group assignment should not be known by the participant).

If neither the assessor nor the participant is aware of the intervention allocations, the trial is said to be 'double-blind'. If only the assessors (or, more rarely, only the participants) are aware of the allocations, the trial is called 'single-blind'. For situations in which there is no known effective treatment or preventive method, a placebo of some sort must be used if double-blinding is to be assured. The 'double-blind' approach is the key to the elimination of bias in the assessment of the impact of an intervention, and, wherever possible, a 'double-blind' design should be used. Sometimes it is not possible because of the nature of the intervention procedure, for example, where participation in health education sessions is being compared to no intervention, or where cervical surgery is being compared to drug treatment for cervical cancer. But even if the providers of the intervention must know the assignments, the person who

assesses the trial outcome should be kept 'blinded', if feasible. The more clearly defined and objective the outcome to be measured, the less critical it becomes to ensure blinding of the assessor. For example, as long as there is complete ascertainment of all deaths in all arms of the trial, blindness is unlikely to be important in a trial with mortality as the endpoint. Similarly, the less likely a patient is to be influenced by knowledge of which intervention they have received, the less important their blinding is.

4.2. Unit of application of the interventions

Different interventions can be applied either to an individual or groups of individuals, such as everyone in a family or household, everyone working in a particular company, or everyone in the community. The unit for randomization should usually vary in parallel with this. The choice of the unit for application of the intervention depends upon the nature of the intervention, the administrative method for its application, and the purpose for which the intervention is being applied. In statistical terms, the most efficient design, in most circumstances, is to use the individual as the unit of application, and this should be the design of choice, unless there is good reason for household or community (group) application and randomization. There are four main reasons for applying an intervention to a group, rather than by individual.

First, group allocation is appropriate when, by its nature, the intervention must be applied to everyone in the group such as all those living in a geographical area, workplace, school, or community. Examples include most environmental alterations and many vector control interventions. It also applies to many educational or health promotion interventions which, although they can be delivered at individual level, are likely to spill over or 'contaminate' other individuals living in the same community.

Second, it may be logistically easier to administer the interventions to groups, rather than on an individual basis. Sometimes it is administratively simpler and/or more acceptable to randomize by household or village, rather than by individual. Furthermore, with individual randomization of medications, for example, there may be a risk of individuals sharing medications within households or villages.

Third, if the purpose of applying the intervention is to reduce transmission of infection by a parasite, for example, the appropriate unit of application would be the 'transmission zone', i.e. the area in which people (and, where appropriate, vectors and intermediate hosts) may be interacting and sharing a common pool of parasites. Factors of importance in defining such zones may include the flight range of vectors and the movements of people, vectors, and intermediate hosts. To reduce interchange ('contamination') among transmission zones, it may be useful to have intervening buffer zones that are not involved in the trial. For many diseases, however, the size of the transmission zones may be difficult to determine and may vary over time.

Some interventions may be applied to individuals, but with the expectation that there may be an effect on transmission, through applying them to a high proportion of individuals in the community, that goes beyond the effect that would be achieved directly within the individuals who received the intervention (for example, through 'herd immunity'). The extent of coverage required to produce such effects depends upon the epidemiological circumstances, the presence of other control measures, and the type of intervention being introduced. For example, the use of a malaria vaccine to reduce the transmission of malaria in parts of Africa where the disease is 'holoendemic' may require so near to complete coverage that such a purpose would not be seriously considered. However, in other parts of Africa where the disease is much less prevalent, achieving high coverage with a highly effective vaccine might be sufficient to interrupt transmission.

For some types of intervention procedures, when the procedure itself provides individual benefit, such as ivermectin in the treatment of onchocerciasis, a further important issue is whether reduction of transmission provides a benefit, in addition to the individual reductions of morbidity/mortality. Trial designs to demonstrate this additional benefit are likely to be complex.

A fourth reason for applying interventions to a group or community as a unit would be for trials involving an intervention of already proven efficacy in individuals, but for which the delivery may be more effectively carried out on a group or community basis. The trial might consist of a comparison of different delivery systems. Generally, the end result desired in this type of trial is based upon cost-effectiveness criteria. Here the question would be whether it is possible to achieve a greater disease reduction for a given expenditure (or alternatively the same disease reduction for less expenditure) by use of a community-based distribution system than by the usual individual distribution methods. Many types of community-based distribution systems require community participation studies. The basic

principles involved in community participation studies and in cost-effectiveness studies are described in Chapters 9 and 19, respectively.

When group randomization is adopted, the efficiency of the design can be improved by ensuring that the groups allocated to the different intervention arms are as similar as possible with respect to risk factors for the outcomes of interest, in the absence of the intervention. In other words, there is 'balance' between the risks of the outcomes of interest between the trial arms. When there are large numbers of units to be allocated, randomization itself will ensure comparability, but usually when communities or other groups are the units to be randomized, the number of units is relatively small, and randomization may leave considerable differences between the groups in the different arms. Attempts can be made in the analysis to allow for these differences, but the persuasiveness of the results may be reduced if the conclusions depend upon extensive statistical manipulation of the trial results. A more efficient approach to increase the comparability of the groups in the different arms is to stratify the groups into 'blocks' having similar underlying pre-intervention risks of the disease outcome in question and to randomize within each block. Stratification should be either in terms of variables which are strongly related to the risk of the outcome under study or in terms of this risk itself. For example, in trials of interventions against malaria in which villages are to be randomized, the villages might be stratified according to their pre-trial malaria prevalence or incidence rates, if such information is available, and the randomization done within each of these strata. An extreme type of stratification is when each 'block' includes the same number of groups (for example, villages) as there are arms of the trial, with each village within each 'block' having similar malaria rates. One village in each block is then randomly allocated to each intervention (see also Chapter 11, Section 3).

An alternative to stratification, when the number of available units for simple randomization, or even for stratification, is too small, is known as 'constrained' or 'restricted' randomization. Assume there are 20 villages to be randomized. All possible combinations of ten versus ten villages are evaluated, and only those combinations with good baseline comparability between the two sets of villages are selected. Next, one of the shortlisted combinations is chosen at random, and one of the two sets of ten villages is randomly selected to become the intervention group (Moulton, 2004). An example of the use of this approach is given in Sismanidis et al. (2008). See also Chapter 11, Section 3.3.

Often, good information on the distribution of the outcome measures will not be available in the trial population. In such circumstances, baseline studies to obtain the required information should be considered. Sometimes, as an alternative, surrogate measures must be used (i.e. measures which are thought to correlate closely with the outcome measures of principal interest). In the absence of detailed data on the population, geographical proximity and socio-economic level may be used as stratification characteristics. Thus, if a small geographical area is chosen as the randomization unit, the total trial area would be divided into regions containing a small number of relatively homogeneous units and, within each region, an equal number of units allocated to each treatment arm.

4.3. 'Stepped wedge' design

The issue of the ethics of randomization is presented in acute form in situations where previous studies, perhaps using short-term endpoints or a more intensive intervention than is feasible on a population basis, indicate that the intervention is likely to be beneficial. Withholding the intervention from those in one of the treatment arms for the duration of the trial may then be argued to be unacceptable. Also, some individuals or organizations have an inherent, if irrational, distrust of randomization, worrying that it is 'experimentation' (which of course it is!) or even 'treating humans like laboratory animals'. Such positions can make it impossible for a straightforward RCT design to be accepted. An approach that can be adopted in this situation is the phased introduction of the intervention on a group-by-group basis, until the entire target population is covered. In order to avoid bias, the order in which the groups are given the intervention should be randomized and the number of groups should not be too small—at least six, preferably many more. This approach was first used in The Gambia to evaluate the long-term effects of vaccination against the hepatitis B virus (HBV) (The Gambia Hepatitis Study Group, 1987). A recent example of this design was a trial in Ghana to evaluate the impact on child mortality of treating fever using anti-malarials, with or without also treating with antibiotics (Chinbuah et al., 2012). Other examples are given in Brown and Lilford (2006).

The trial design is illustrated in Figure 4.1. This type of design has been called a 'stepped wedge' design. The power of this approach, compared to a simple allocation of groups to one or other treatment arms, is of the order of 75–80%, depending on the number of groups. The same considerations apply to stratification and blocking, as in the static allocation designs.

In the trial in The Gambia, hepatitis B vaccine was introduced into the routine child vaccination programme over a period of 4 years. The order in which the different vaccination teams (there were 17 at the time the trial was planned) began to use the vaccine was random. At the end of 4 years, there was a cohort of children who had received the vaccine and a cohort who had not. These cohorts are being followed to compare the incidence rates of liver cancer and chronic liver disease. At the end of the 4 years, all vaccination teams had started vaccinating children, so subsequent cohorts of children were vaccinated. This phased introduction of the intervention mimicked the way in which many public interventions are introduced, but the key feature of the random order of the introduction of the intervention across the 'clusters' (in this case, vaccination teams) brought the crucial benefit of reducing the potential for the trial producing biased results.

4.4. Other approaches to allocation of the interventions

The allocation of interventions to individuals based on a random scheme is the best approach to rigorously exclude the potential biasing effects of other factors. However, non-randomized designs are often used. For example, a common approach is the 'before–after' or 'pre–post' design, in which the incidence or prevalence of the disease under study is compared before and after the intervention has been applied, and an attempt is made to attribute any difference to the effect of the intervention. This approach has important limitations as it may be wrong to assume that, in the absence of the intervention, the disease rate would have remained the same. Many diseases, and especially those of parasitic or infectious origin, vary greatly in incidence and severity from year to year and place to place, for reasons that are incompletely understood. Certainly variations in climate (for example, temperature and rainfall) can have profound effects. Some diseases show marked declines (or increases) over time in some communities (for example, TB and malaria), and sometimes these cannot be predicted in advance, or even related to any obvious specific factor. 'Before and after' evaluations of interventions in such situations may be very misleading. Also, it is not uncommon that the methods used to ascertain the trial outcomes change over time, either in terms of the actual data collection method or the person or organization doing the data collection changes, and the two produce systematically different results.

Another commonly employed approach is to apply an intervention in one community, and not in another, and to attribute any difference in disease rates between the two communities as being due to the intervention. This also may be very misleading, as a change may have occurred in one community, but not in the other, for reasons that had nothing to do with the intervention. Random, rather than purposive, allocation of the intervention to one of the two communities does not make any difference to this.

The commonest reason that is advanced for using a non-random allocation between intervention groups is for simplicity of design and administrative ease. Approaches like these also seem easier to explain to officials and to gain public acceptance. The rationale for randomization is difficult to communicate, even to other scientists, but the arguments in favour of randomization, as outlined in Section 4.1, are extremely strong, and failure to accept this approach has frequently led to studies from which erroneous conclusions have been drawn.

There are, however, situations in which allocation cannot be made on a randomized basis. There are occasions when the benefits of an intervention appear so clear that a properly randomized trial cannot be contemplated, or when the intervention or package has already been subjected to randomized trials and is being scaled up under routine conditions. The value of the intervention then has to be assessed by comparison of the situation before and after its introduction, or by the use of case-control studies after the intervention has been introduced (Smith, 1987). Although before vs after studies suffer from the major limitations described earlier in this section, the plausibility of the trial's conclusions can be increased by trying to rule out alternative reasons why the changes might have occurred (Bonell et al., 2011; Victora et al., 2004). First, if possible, data should be collected on more than one occasion, both before and after the intervention is introduced (sometimes called a time-series study). This allows checking that the outcome of interest was not already declining at the same rate prior to the start of the intervention, and that any decline after the intervention was introduced was consistently present, rather than only there at one time point. Second, a comparison should be made with time trends in disease rates in neighbouring populations where the intervention or package of interventions has not been delivered, and/or in the country or region of the country as a whole. Third, the sharpness with which changes in disease rates take place should be consistent with what might be reasonable to expect from the intervention and related to the speed with which the intervention is introduced over the entire population. Fourth, knowing and recording possible confounding variables in the before and after periods or in the populations being compared in a non-randomized study may also aid interpretation of differences. For example, in a study in which an

objective is to reduce transmission of lymphatic filariasis by treating the human population with antifilarial drugs, monitoring the vector population for changes in density and infectivity might be undertaken.

While acknowledging these exceptions to the use of randomization as the basis of allocation, such studies do not have the rigour of a randomized design, and any conclusions drawn from them must be viewed with some caution. It is reasonable to think of there being a hierarchy of evidence from intervention studies, with (1) well-designed and well-conducted RCTs providing the strongest evidence, followed by (2) quasi-experimental studies, in which there is a similar contemporaneous comparison group, but the receipt of the intervention has not been allocated randomly, and then (3) non-experimental designs, in which there is no similar, contemporaneous comparison group such as the before–after, time-series, or after-only designs outlined earlier in this section. Formal guidelines have been developed by the GRADE working group (Guyatt et al., 2008) (<http://www.gradeworkinggroup.org>) to rank the quality of evidence on the effect of an intervention, based on different kinds of study, ranging from the RCTs, which are judged to provide the highest quality of evidence (if properly conducted), through to other kinds of study, providing lower-quality evidence, including observational studies. The WHO has now adopted these guidelines and attempts to undertake a formal grading of the quality of the evidence, with respect to policy recommendations they make regarding specific interventions. The main focus of this book is on RCTs.

5. Choice of outcome measures and trial duration

For many interventions, there will be a range of outcomes that could be affected and which might be of interest to study (see also Chapter 12). Nutritional supplements, for example, might affect any or all of the following:

1. biochemical measures

2. short-term acute consequences of deficiency

3. the consequences of chronic deficiency

4. mortality due to the specific causes of death that the intervention is intended to rectify

5. total (all-cause) mortality.

In determining which outcome is of the greatest importance for the trial, consideration must be given to whether:

1. the outcome is of clinical or public health importance

2. the probable effect on that outcome is large enough to be of clinical or public health interest

3. it can be accurately measured.

A substantial impact on total and age-specific mortality rates is always of public health importance, and systems can usually be set up to ensure that they are well recorded (even though such systems often require considerable input if they are not already in place), but they are unlikely to be sufficiently affected by most interventions to enable effects to be detected with studies of manageable size. Mortality from the specific causes that the intervention is designed to reduce should be more greatly affected, of course, but is usually much more difficult to measure accurately. In most low-income settings, routine reporting of births and deaths by medically certified cause of death is not available or is very incomplete and therefore potentially misleading. In these circumstances, measuring cause-specific mortality rates will require interviews with close relatives or friends of the deceased to try to ascertain the signs and symptoms preceding death, so that an attempt can be made to assign a likely cause of death. Such interviews are known as 'verbal autopsies'. The International Network for the Demographic Evaluation of Populations and Their Health in Developing Countries (INDEPTH) has produced model verbal autopsy questionnaires (<http://www.indepth-network.org>). Using total mortality as the trial outcome, however, will dilute the effect that might be seen if specific causes were examined, since the variation in deaths due to the unaffected causes is included. The choice may have to be made between setting up special mechanisms to collect high-quality information on the cause of each death or to allow for a dilution of the observed effect by increasing the size of the trial. It should be stressed that, for conditions that are life-threatening, mortality is an important outcome to evaluate and, wherever possible, should be a primary trial outcome, but this generally has substantial implications, with respect to the size of the trial.

Short-term outcomes are clearly attractive in that, if used as the outcome on which the design is based, the trial size will be smaller and the duration shorter than if mortality were to be used. The danger is that the short-term measure in

itself may not be of principal public health importance, and the effect of the intervention on that outcome may not correlate well with the effect on more serious conditions. There is, for example, little point in measuring an antibody response to infection if it bears no or little relation to the risk of disease. Conversely, in the relatively rare situations where it is known that a short-term outcome is highly correlated with an outcome of greater public health consequence (and is effectively a surrogate measure of the more important outcome), it will be more efficient to focus the trial on the surrogate outcome.

In most circumstances, the appropriate outcome for determining the duration and size of the trial would be the most serious consequence of the specific condition at which the intervention is aimed. However, it is not always feasible to use such outcomes in a trial. For example, in a trial of a new measles vaccine in a HIC where death in someone who has measles illness is rare, the onset of measles illness might be a sensible trial endpoint, rather than death from the disease or total mortality. In contrast, in a country where a relatively high proportion of children with measles die, death from measles might well be the outcome of choice. If mechanisms for establishing accurate diagnosis were inadequate, total mortality might even be considered (especially as measles vaccine may reduce the risk of death attributable to diseases other than measles).

Even in trials where total or cause-specific mortality are the primary trial endpoint, short-term 'intermediate' outcomes should also be collected as valuable secondary monitoring and explanatory outcomes, as laid out in the impact model. They provide information, as the trial progresses, as to whether the trial is on target to meet its primary goals and, if it is not on target, should help to identify what remedial action might be required. Also, if the trial does not find a significant impact on its primary outcome, the 'upstream' outcomes may help provide an explanation for why. For example, in a trial of the impact of insecticide-treated nets on malaria mortality, it would be important to also measure net coverage and use, and data on the incidence of malaria illness and age-specific prevalence of malaria parasitaemia by trial arm. When short-term outcomes are used in this way, any assumptions about the natural history of the disease should be clearly thought through and stated in the trial protocol.

Definition of the primary trial outcome will have consequences for the duration of the trial. Prior information should be available on the time needed for the intervention to affect the outcome. In some situations, such as the prevention of liver cancer in adult life by hepatitis B vaccination in the first year of life, the final outcome measure may not be observed for several decades. The need for monitoring of intermediate outcomes (such as the hepatitis B carriage rate) then becomes even more important.

The choice of trial duration is critical for interventions whose impact does not increase linearly over time. For example, the impact of a health education programme in schools to reduce sexual risk taking might be relatively small, until a high proportion of the students have become sexually active. But even then, the impact might be small, until both the students and their sexual partners (who might be several years older or younger) had been through the programme. And finally, the impact may reach a 'tipping point' when enough people had been exposed to the programme to change general social and sexual norms in the population as a whole. However, the choice of trial duration is complicated by the fact that few funding agencies are keen to fund research projects that last more than 3–5 years. A common strategy is to apply for initial funding for a 3- to 5-year trial that will be able to measure the intervention's impact on important intermediate outcomes but is large enough to measure the impact on the primary trial outcomes if continued into a second trial follow-up phase, with the application for further funding based on the results of the first phase.

A final and important point to stress in this section is that it is essential that attention is given to monitoring the severity and frequency of adverse effects of an intervention. In their desire to assess the effectiveness of an intervention, investigators often do not pay sufficient attention to finding and documenting adverse effects, which may require additional effort and resources. In most situations, the future applicability of the conclusions drawn from a trial will involve an assessment of the balance between positive and negative (adverse) effects.

6. Trial population

6.1. Criteria for selection of trial population

The criteria for selection of the population to be included in the trial depends primarily upon what condition the intervention is directed against and upon the purpose of the trial. In general, the population will be chosen from an area in which there is high incidence of the condition of interest, because the higher the incidence of the primary trial outcome, the smaller the study population for the trial has to be. Exceptions are when the purpose of the trial is to

determine the efficacy under special epidemiological circumstances or in special population groups such as in pregnant women.

Good community and governmental co-operation and participation are also key factors in the successful conduct of a trial. The trial area should be accessible at the times surveys are to be conducted (for example, during the rainy season). Well-qualified and experienced field teams should be available or be able to be recruited. In addition, access to high-quality clinical and laboratory facilities may be necessary for the trial. If required, entomological, behavioural science, economic, and other appropriate disciplinary expertise should be available. Planning the trial will be much simplified if baseline data are already available in the trial area.

If the trial design involves the repeated follow-up of members of the study population over several years, as will be the case for many intervention trials, it is important to select a location for the trial in which substantial migration into, or especially from, the area is unlikely to occur. Migration rates in excess of 10% per year are not uncommon in many rural areas and may be considerably higher in urban or peri-urban settings. Unless the trial is conducted within a demographic surveillance population, migration rates may well not be known in advance, so a rapid survey of a sample of the proposed trial population may be useful to determine if a reasonable proportion of the population have been resident in the area for several years.

The choice of trial population may affect the external validity of the trial results. For example, many micronutrient trials are carried out in areas with high prevalence of the specific deficiency. The health impact from supplementation in such areas is likely greater than what would be expected in areas where micronutrient deficits are less frequent, which may represent the majority of areas where supplements will be used in the future.

6.2. Inclusion and exclusion criteria

In general, the trial population should be chosen to represent the group that would be the target for the intervention in a potential future public health programme, if the intervention is found to be effective within the trial. Care should be taken to define the target population. To the extent feasible, those included should be the persons for whom benefit is likely to be the greatest, and those excluded should be the persons for whom benefit is likely to be minimal or indeed who may be harmed. Specific inclusion and exclusion criteria should be developed for the trial. For example, because the major morbidity and mortality associated with malaria in a holoendemic area are seen in infants and young children, these groups are likely to be the focus of a major field trial of a malaria vaccine in such an area, though older children and adults might be used in preliminary studies to test the safety of the vaccine in those who already have some immunity or may be the focus of a vaccine trial where malaria transmission is much less intense.

In early trials of an explanatory nature, special groups at high risk may form the trial population, either to maximize the potential effect, to ensure good compliance, or to facilitate the logistics. Valuable information concerning the potential of the intervention can result, but the extent to which the results can be extrapolated to the general population may be limited.

Exclusion criteria need to be carefully considered so as to eliminate subjects who may be put at greater risk by the intervention or who have underlying conditions that may interfere with the assessment. Exclusion criteria should be stated explicitly and unambiguously, before the trial begins. It is usual to exclude from trials those who are seriously ill, those who are very old, those who are very young, and pregnant women, unless any of these are the specific target group for the intervention. These groups are excluded either because it is considered that they are unlikely to derive benefit from the intervention, or if they are thought to be more likely to be susceptible to possible adverse effects of the intervention, or they are likely to suffer adverse events (AEs) which might incorrectly be associated with the intervention if they are included. Ascertaining pregnancy is difficult, especially in its early stages, without specific testing, and, in some trials, this may not be feasible. Sometimes all women of childbearing age are excluded from trials, if it is thought that damage may be caused by the intervention to the fetus. Against this must be balanced the potential benefit that the excluded groups may receive from the intervention. Also, if pregnant women or children, for example, have been excluded from a trial that shows the intervention to be effective, resulting public health programmes may consider it is inappropriate for them to receive the intervention, in case there are unforeseen risks to them or because the safe and optimal dosage of any drugs involved are not known. As a result, it may be appropriate to include them in later 'bridging' trials, with careful monitoring of pregnancy outcomes.

6.3. The size of the trial population

Attention needs to be given to the required size of the trial, in terms of the precision of the effect estimates and of the power to detect important differences. These aspects are discussed in detail in Chapter 5. It is important to allow for the loss of power that results from group randomization if such a design is adopted (see Chapter 5, Section 6).

For interventions that are likely to be given to large numbers of individuals, if they are subsequently introduced into disease control programmes, there are strong arguments in favour of designing trials of the interventions to also be large not only to pick up any rare side effects, but also to obtain a relatively precise measure of their expected impact.

6.4. Compliance

Conclusions from a trial will be based on a comparison of the outcome measures adopted for the trial in those allocated to the alternative intervention arms of the trial. Only a certain proportion of those allocated to a particular intervention will receive that intervention effectively. Effective delivery of an intervention requires both that the provider carries out the intervention procedure correctly and that the trial participants co-operate in the desired fashion. In field trials, the provision of the intervention will usually be under the control of the investigator, but a successful trial also requires the compliance of the participants, who are not under the control of the investigator, and will depend on the understanding and co-operation of the community involved. Hence, the strong emphasis in this manual on the importance of communication and feedback between the investigating team and the participating communities has a pragmatic, as well as an ethical, basis.

In most trials, however, some participants will not fully comply, and the intervention procedure either will not be carried out or it will not be done in an effective manner. For trials to determine the public health value of an intervention (pragmatic trials), some degree of non-compliance may give a more realistic measure of effectiveness than a tightly controlled trial in which every effort is made to ensure that the intervention is effectively delivered, but for explanatory studies, in which an important objective may be to determine the maximum effect possible, every effort should be made to keep compliance high. Wherever possible, the degree of compliance should be continually monitored, at least on a sample basis. This might be done, for example, by doing urine or blood analyses to check that the expected drug or nutritional supplement has actually been ingested. For intervention measures that are administered sequentially over time or on a continuing ongoing basis, repeated specimens should be taken. In a trial to measure the impact of introducing improved water supplies, for example, it will be important to measure the proportion of the target population who actually access the improved water source. This is particularly relevant in trials in which a health effect is mediated through a change in behaviour, as is the case in a breastfeeding promotion trial with morbidity or mortality as endpoints. Documenting compliance with counselling—assessed through changes in feeding practices—is essential.

A further aspect of compliance that is sometimes overlooked is that those in the 'control' arm of a trial, who are allocated to routine care or placebo, may adopt the test treatment under study. For example, if health centres in some villages are allocated to receive an intervention, such as offering voluntary medical male circumcision or improved STD treatment, while those in other villages serve as controls, people in the control villages may go to the health centres in the intervention villages to obtain the intervention. Monitoring for the possible occurrence of this latter form of non-compliance (sometimes called 'contamination') is important. Care should also be taken in the construction of the different treatment groups to minimize the opportunity for such contamination. In the circumcision example, ensuring there is clear geographical separation of villages in the different arms of this trial by leaving a 'buffer zone' would be one means of minimizing contamination.

7. Implementation

7.1. Community acceptance

Critical to the conduct of a successful trial is that the trial population co-operates during the conduct of the trial and takes up the intervention offered. They must feel a part of the trial and perceive it to be for the benefit of their community. To ensure these aspects will require careful planning and investigation before the trial starts, including appropriate discussion with, and explanation to, community leaders and potential participants. Feedback and interaction should be continued throughout the course of the trial. These aspects are discussed in several chapters, and especially in Chapters 6 and 9 and part of Chapter 15.

7.2. Feasibility studies and pilot testing

Unless the acceptability and feasibility of implementing the intervention and the evaluation procedures that will be used in the trial have already been tested locally, it is usually wise to conduct a smaller feasibility study in advance of the main trial. The feasibility study may only include some aspects of the trial, such as the acceptability and feasibility of delivering the intervention, or the feasibility of enrolling trial participants or of administering a questionnaire or collection and testing of laboratory specimens. Whether or not such a feasibility study has been conducted, it is essential that all the trial procedures are tested together in a pilot study, exactly as they will be applied in the actual large-scale field trial. However, the pilot study should be conducted on a much smaller number of participants and with enough time for the trial procedures to be modified in the light of the findings. Feasibility studies and pilot studies are discussed in detail in Chapter 13.

7.3. Staff recruitment, training, and retention

The dedication and commitment of the staff employed to conduct a field research project are essential. This will involve their careful selection, training, and then support. They must understand the importance of their role in the trial and how it relates to that of others. The importance of high-quality work must be emphasized, and this must be monitored throughout the trial (see Section 9 and Chapter 16). Trials of long duration present the additional challenge of keeping staff motivated and performing at adequate levels of quality and avoiding excessive turnover. Open and frank discussions with staff are essential, and benefits, such as regular increases in salaries over time, may help motivation and retention.

7.4. Field organization

All aspects of field procedures should be planned in advance, and potential problems and solutions anticipated (for example, in case of staff sickness or vehicle, computer, or laboratory equipment failure). The trial design must reflect not only what is ideal, but also what can be done, given the constraints under which the trial must be conducted. These aspects are considered in detail in Chapters 16 and 17. Issues relating to mapping and conducting a census of the trial area are covered in Chapter 10.

8. Data handling

8.1. Data collection

A necessary part of most trials will be the collection of baseline (pre-intervention) data. These will include identification information on participants, such as name, age, sex, place of residence, and information on other factors that may influence the risk of occurrence of the outcome measures under study in the trial. Although randomization of a large enough number of individuals, or clusters of individuals, should result in an approximately equal distribution of all the important characteristics between trial arms, such baseline data, which should ideally include all known confounders, can be used to check that this balance has actually occurred in practice. And if it has not, then it can also be used to adjust for such imbalances in the trial analyses.

In addition, it may be important to collect general baseline data on the population where the trial is being carried out. These may include not only the epidemiological characteristics of the population, but also the socio-economic, cultural, political, health services, nutritional, and other relevant characteristics. Such contextual factors may be essential to interpreting whether the trial's results can be generalized to another setting.

Additional data will be collected during the course of the trial to monitor the application of the interventions and to record information on the outcomes of interest. The conduct of a population census is described in Chapter 10, and methods to obtain high-quality data at the start of a trial and during its course are described in Chapter 14. Obtaining data using social or behavioural methods is outlined in Chapter 15, and for measuring the costs of the interventions is outlined in Chapter 19. Of crucial importance in any trial is the proper measurement of the incidence of endpoints against which the intervention is designed to protect, and these aspects are discussed in Chapter 12.

8.2. Data processing

Methods of coding, entering, and then managing computerized data collected in a trial are described in Chapter 20.

9. Quality control

In most intervention studies, members of the population are invited to participate, the intervention is applied, perhaps repeatedly, and the population is kept under surveillance, until the final trial outcomes are recorded. The quality of each step in this process must be monitored. The two major reasons, which hardly need stating, are first to ensure that each operation is being performed to an acceptable standard, and second to identify areas where attention is required. A third reason is to be able to ascertain, at the end of a trial that failed to show anticipated effects, the possible reasons for failure. The damage done by a misleading 'negative' result can be serious. The following are major aspects of quality control (QC) that need attention.

9.1. The intervention

Regular monitoring of the delivery of the intervention should be an integral part of the design to ensure that there is no change in the quality, as a trial goes on. For example, in a vaccination trial, continual review would be needed of the quality of the vaccination techniques being used by fieldworkers and of the quality of the vaccine(s) used in the intervention. For example, the potency of each batch of vaccine used should be assayed, together with monitoring of the maintenance of any required cold chain. Particularly relevant for trials where the intervention includes case management or counselling is monitoring the quality of these procedures through regular observation of a sample of provider–client interactions.

Short-term endpoints may be used for monitoring the quality of the intervention. At the individual level, repeated surveys of physiological measures of response to the intervention will provide an assessment of whether an effective intervention agent has been delivered. Examples would be antibody levels against a vaccine or levels of a micronutrient in serum. In trials including provider–client interactions, exit interviews with clients can be used to monitor their understanding of the advice that was provided. Such evaluations may have to be done or be evaluated by an independent trial monitor to ensure that those who will assess the main endpoints in the trial are kept blind— whenever possible—to the identity of those in intervention and control groups.

9.2. Follow-up

For many intervention studies, the endpoints of interest may not emerge until a lengthy period after the start of the intervention. It may not be necessary to keep the entire trial population under active observation, and this is often not feasible (for example, cases might be detected, as they report to clinics, rather than by conducting periodic surveys of the trial population), but it is essential that the trial is designed in such a way that losses to the trial population (for example, cases who do not go to clinics) will not distort the conclusions. The follow-up rate should be monitored, in order to identify potential problems at an early stage (for example, disgruntlement in a particular village or to identify a fieldworker whose work quality is declining). If possible, the reasons that individuals are lost to follow-up should be ascertained. Some losses may be inevitable, such as participants who die or who move out of the trial area, while it may be possible to take remedial action to prevent others such as participants who withdraw their participation or who are temporarily absent but could be found by repeated visits to their homes. The baseline characteristics of those who are lost to follow-up should be compared with those of participants who remain in the trial, and this information should be analysed to assess any effect that the losses might have on the interpretation of the results of the trial.

9.3. Assessment of trial outcomes

Mechanisms have to be established to ensure that the quality of information on all the trial outcomes is acceptable. Ongoing monitoring is required to establish that the data on trial outcomes are maintaining acceptable quality and that no biases are present in the way outcomes are recorded in different treatment arms. Attention needs to be paid to inter-observer variation in the assessment of the outcomes and changes that may occur in this variation, as the trial progresses.

9.4. Other field and laboratory procedures

QC should pervade all field activities, and the question as to how high quality is to be achieved and maintained should be addressed specifically for all activities. This is discussed in most of the chapters that follow, and specifically in Chapter 16, Section 7.

Laboratory procedures should be subject to constant scrutiny, and 'blind'-coded duplicate samples or known positives or negatives should be introduced into the workload regularly to monitor performance.

In interview surveys, a proportion of respondents should be re-interviewed by a second interviewer, blind to the results of the first interviewer, to check on the repeatability of the responses. If the questionnaire is long, the re-interviews might focus on a subset of key questions, rather than repeating the full questionnaire, in order to avoid undue demands on participants.

It is important that all involved in the trial accept and understand the need for constant checking and re-checking. This is both so that any sanctions that are taken for repeated poor performance do not come 'out of the blue', but, more importantly, as a way of encouraging all trial staff to maintain high quality at all times, because they know that errors will be spotted reasonably quickly. On the other hand, errors are bound to occur, and their detection should usually result in support and, where necessary, additional training, with reprimands being reserved for where there is evidence of dishonesty or continual carelessness. Incentives or rewards to encourage high-quality work may be worthwhile.

All members of the field team are, and must be made to feel, important contributors to the research project. Feedback of results and progress should be continuous and frequent, so that they can appreciate where their contribution fits into the overall project. Neglect is a great stimulus to poor-quality work.

10. Analysis, monitoring, and reporting

10.1. Planning the main analyses

The main analyses that are expected to result from the trial should be developed in some detail, with the use of dummy tables. Such an exercise is a great help when planning the trial, as it helps clarify exactly what data are actually needed and highlights redundant data. All specific objectives should be tied to planned analyses.

10.2. Analyses during the trial

Analysing relevant data from a trial, as they accumulate during the trial, is an important way of monitoring the satisfactory progress of a trial. Administrative analyses of the numbers of participants recruited each day or week and of the data collected by different fieldworkers are important for QC. A running tally should be kept of the numbers of participants experiencing the various trial endpoints to verify that the estimates of incidence rates used to plan the size of the trial were appropriate. Ideally, the investigators will be blind with respect to which interventions have been allocated to which participants, but differences between the different interventions might be analysed by a data and safety monitoring committee (as discussed in Section 10.3). Other aspects of interim analyses are discussed in Chapter 5, Section 7.1 and Chapter 7, Section 4.1.3.

Increased reliance on the use of smart phones or personal data assistants (PDAs) to record data when interviewing participants facilitates real-time data quality checks and analyses. Considerable ahead-of-time preparation and planning, however, are necessary, in order to programme devices to be able to produce such analyses regularly.

Interim reports, based on such ongoing analyses, may be required during the course of a trial by national authorities and by the trial's funding agency, in order to check that the original proposal is being adhered to and that the assumptions underlying the trial design were correct.

10.3. Data and Safety Monitoring Committee

For large trials, it is advisable for the investigators to set up an independent DSMC. Such a committee generally has access to selected unblinded data during the course of a trial and, for example, will conduct analyses to monitor whether there are an unacceptable number of adverse events (AEs) associated with an intervention. In such circumstances, the committee may recommend changes to the design of the trial or, in more extreme cases, that the trial be stopped, either temporarily or permanently.

The DSMC might also be charged with conducting interim analyses of the trial with respect to the primary endpoint, so that if the efficacy of intervention is substantially lower or substantially higher than expected, changes to the trial design, including early stopping, might be recommended.

The roles and functioning of DSMCs are discussed in Chapter 7.

The most important function is usually to hold the randomization code for the trial and to monitor the results of the trial, both in terms of effectiveness and safety, as they accumulate. If there is evidence of a substantially increased risk of adverse reactions associated with any of the interventions under study, the committee would have the power to

advise the Trial Steering Committee to stop further recruitment. Similarly, if evidence accumulates that one intervention is substantially better than the others (or one is substantially worse), the committee would usually recommend that the trial be ended or that at least one of the trial arms is discontinued. In blinded trials, a major advantage of these functions being undertaken by an independent committee is that the investigators can remain blind to the randomization codes, which is an important way of ensuring unbiased assessment of the trial endpoints. But, even where the trial is not blinded, it still has the considerable advantage of ensuring that the recommendation of stopping or continuing a trial is as objective as possible, because stopping a trial early usually has considerable logistic implications and may not be popular with the investigators, staff (who may even need to be laid off early), or participants.

The circumstances in which a trial will be prematurely ended should be carefully considered when the trial is being designed, and the DSMC should be party to such discussions. It will not be possible to predict all possible situations that may cause a decision to be taken to end a trial, but this should be done to the extent possible. In particular, there should be consideration as to how large a difference may be apparent between the interventions, with respect to their impact on specific endpoints, before it is decided to end the trial. In some circumstances, it may be important to go on beyond the point where statistical significance is reached. These issues are discussed in Chapter 5, and there are also ethical considerations which are discussed in Chapter 6.

The DSMC might also set up independent QC checks on trial procedures and, for example, may arrange to review the diagnoses of all cases of the diseases of interest arising in the trial (which should be done, of course, 'blind' to knowledge of the randomization codes).

The committee usually works on a pro bono basis and does not have auxiliary staff. If its activities will require QC checks or diagnostic reviews, it may be necessary to budget for these activities when preparing the protocol.

In some trials, the DSMC may consist of one person, sometimes called the 'clinical monitor'.

10.4. Analysis methods

The analysis of a large field trial will usually be a complex undertaking and will usually require the involvement of a professional statistician, sometimes under supervision of a senior statistician or epidemiologist. It is not feasible in a manual of this kind to detail all of the analysis methods that it might be appropriate to employ in different trials. However, in Chapter 21 an outline is given of the main methods of analysis that are likely to be employed. It is included as it summarizes relevant methods that are not covered as comprehensively in the most basic epidemiological texts or books on medical statistics.

10.5. Reporting results

Once a field trial has been completed and the results analysed, it is essential that the results and their implications are made available to the scientific community, to those who participated in the trial, and to those responsible for designing and implementing regional and national disease control strategies. These aspects are discussed in Chapter 23.

10.6. Further studies

Many trials will provoke questions amenable to further research. One example might be if a trial of a hookworm vaccine shows that it provokes good specific antibody- and cell-mediated immune responses and reduces the incidence of infection by 80% but is associated with prohibitive adverse reactions, further studies may well be needed to explore which antigens are causing the adverse reactions and whether removing these will also reduce the vaccine's effectiveness against hookworm.

Alternatively, if a trial of traffic-calming measures in one city shows that they are highly effective in reducing road traffic accidents, questions may well arise on how best to implement similar measures in other settings and/or to monitor the effectiveness of such interventions when implemented on a wide scale and over a long period of time. Such studies are often called Phase IV studies, as they evaluate interventions in real-world settings after the Phase III trial has been completed. These are discussed in Chapter 22.

11. The 'SPIRIT' checklist for standard protocol items for clinical trials

Nearly all intervention trials will need to have a protocol developed, which serves as the basis for trial planning, conduct, and reporting. Before a trial starts, it is recommended or, in many cases, required that the protocol is deposited in a trial register (see Chapter 7, Section 5). Until recently, there has not been specific guidance as to exactly what items should be included in such a protocol. However, such guidance has recently been published (Chan et al., 2013a; Chan et al., 2013b) as a component of the EQUATOR project (Enhancing the QUAlity and Transparency Of health Research) (<http://www.equator-network.org/>). The publications include a 33-item checklist, the so-called SPIRIT (Standard Protocol Items: Recommendations for Intervention Trials) 2013 checklist, which is reproduced in Table 4.1. This gives a useful outline of how a trial protocol might be organized, bearing in mind the issues we have discussed in this chapter. Readers should refer to the SPIRIT website (<http://www.spirit-statement.org/>) for the most recent version.

References

Arifeen, S. E., Hoque, D. M., Akter, T., *et al.*2009. Effect of the Integrated Management of Childhood Illness strategy on childhood mortality and nutrition in a rural area in Bangladesh: a cluster randomised trial. *Lancet*, **374**, 393–403.10.1016/S0140-6736(09)60828-X [PubMed: 19647607] [CrossRef]

Awasthi, S., Peto, R., Read, S., *et al.* 2013a. Vitamin A supplementation every 6 months with retinol in 1 million pre-school children in north India: DEVTA, a cluster-randomised trial. *Lancet*, **381**, 1469–77. [PMC free article: PMC3647148] [PubMed: 23498849]

Awasthi, S., Peto, R., Read, S., *et al.* 2013b. Population deworming every 6 months with albendazole in 1 million pre-school children in north India: DEVTA, a cluster-randomised trial. *Lancet*, **381**, 1478–86. [PMC free article: PMC3647147] [PubMed: 23498850]

Ayles, H. M., Sismanidis, C., Beyers, N., Hayes, R. J., and Godfrey-Faussett, P. 2008. ZAMSTAR, The Zambia South Africa TB and HIV Reduction Study: design of a 2 × 2 factorial community randomized trial. *Trials*, **9**, 63.10.1186/1745-6215-9-63 [PMC free article: PMC2585552] [PubMed: 18992133] [CrossRef]

Bonell, C. P., Hargreaves, J., Cousens, S., *et al.*2011. Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *Journal of Epidemiology & Community Health*, **65**, 582–7.10.1136/jech.2008.082602 [PubMed: 19213758] [CrossRef]

Brown, C. A. and Lilford, R. J. 2006. The stepped wedge trial design: a systematic review. *BMC Medical Research Methodology*, **6**, 54.10.1186/1471-2288-6-54 [PMC free article: PMC1636652] [PubMed: 17092344] [CrossRef]

Chan, A. W., Tetzlaff, J. M., Altman, D. G., *et al.* 2013a. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of Internal Medicine*, **158**, 200–7.10.7326/0003-4819-158-3-201302050-00583 [PMC free article: PMC5114123] [PubMed: 23295957] [CrossRef]

Chan, A. W., Tetzlaff, J. M., Gotzsche, P. C., *et al.* 2013b. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*, **346**, e7586.10.1136/bmj.e7586 [PMC free article: PMC3541470] [PubMed: 23303884] [CrossRef]

Chinbuah, M. A., Kager, P. A., Abbey, M., *et al.*2012. Impact of community management of fever (using antimalarials with or without antibiotics) on childhood mortality: a cluster-randomized controlled trial in Ghana. *American Journal of Tropical Medicine and Hygiene*, **87**, 11–20.10.4269/ajtmh.2012.12-0078 [PMC free article: PMC3748510] [PubMed: 23136273] [CrossRef]

Gupte, M. D., Vallishayee, R. S., Anantharaman, D. S., *et al.*1998. Comparative leprosy vaccine trial in South India. *Indian Journal of Leprosy*, **70**, 369–88. [PubMed: 10189587]

Guyatt, G. H., Oxman, A. D., Vist, G. E., *et al.*2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, **336**, 924–6.10.1136/bmj.39489.470347.AD [PMC free article: PMC2335261] [PubMed: 18436948] [CrossRef]

Montgomery, A. A., Peters, T. J., and Little, P. 2003. Design, analysis and presentation of factorial randomised controlled trials. *BMC Medical Research Methodology*, **3**, 26.10.1186/1471-2288-3-26 [PMC free article: PMC305359] [PubMed: 14633287] [CrossRef]

Moulton, L. H. 2004. Covariate-based constrained randomization of group-randomized trials. *Clinical Trials*, **1**, 297–305.10.1191/1740774504cn024oa [PubMed: 16279255] [CrossRef]

Ross, D. A., Changalucha, J., Obasi, A. I., *et al.*2007. Biological and behavioural impact of an adolescent sexual health intervention in Tanzania: a community-randomized trial. *AIDS*, **21**, 1943–55.10.1097/QAD.0b013e3282ed3cf5 [PubMed: 17721102] [CrossRef]

Sismanidis, C., Moulton, L. H., Ayles, H., *et al.*2008. Restricted randomization of ZAMSTAR: a 2 × 2 factorial cluster randomized trial. *Clinical Trials*, **5**, 316–27.10.1177/1740774508094747 [PubMed: 18697846] [CrossRef]

Smith, P. G. 1987. Evaluating interventions against tropical diseases. *International Journal of Epidemiology*, **16**, 159–66.10.1093/ije/16.2.159 [PubMed: 3301705] [CrossRef]

The Gambia Hepatitis Study Group 1987. The Gambia Hepatitis Intervention Study. *Cancer Research*, **47**, 5782–7. [PubMed: 2822233]

Victora, C. G., Habicht, J. P., and Bryce, J. 2004. Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health*, **94**, 400–5.10.2105/AJPH.94.3.400 [PMC free article: PMC1448265] [PubMed: 14998803] [CrossRef]
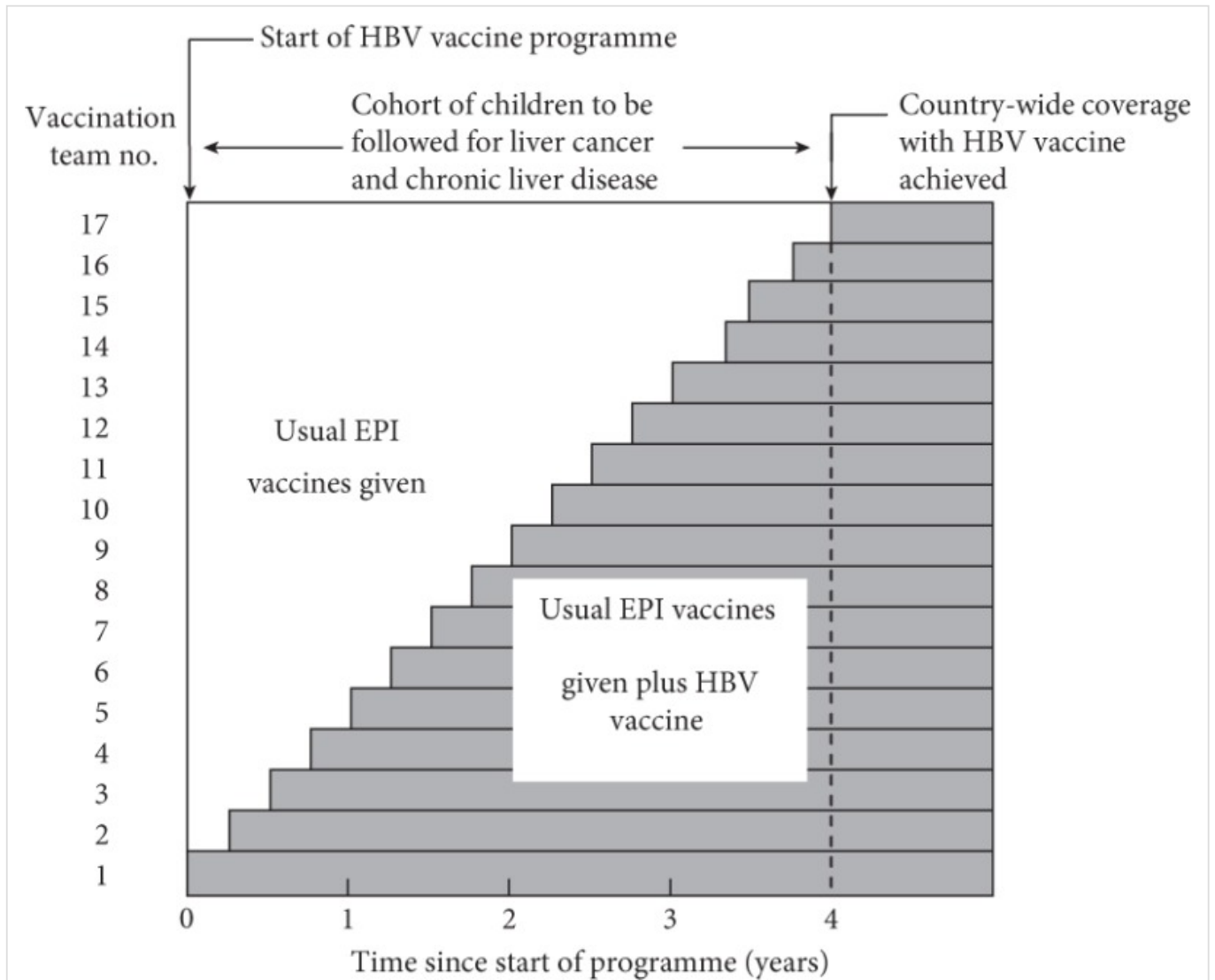
Figures

Figure 4.1

The 'stepped wedge' trial design used to evaluate the impact of hepatitis B vaccination on liver cancer rates in The Gambia.

Tables

Table 4.1   The SPIRIT 2013 checklist: recommended items to address in a clinical trial protocol and related documents[*]

Section/item	Item no.	Description
ADMINISTRATIVE INFORMATION		
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry
	2b	All items from the World Health Organization Trial Registration Data Set
Protocol version	3	Date and version identifier
Funding	4	Sources and types of financial, material, and other support
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors
	5b	Name and contact information for the trial sponsor
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see item 21a for data monitoring committee)
INTRODUCTION		
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention
	6b	Explanation for choice of comparators
Objectives	7	Specific objectives or hypotheses
Trial design	8	Description of trial design, including type of trial (for example, parallel group, crossover, factorial, single group), allocation ratio, and framework (for example, superiority, equivalence, non-inferiority, exploratory)
METHODS: PARTICIPANTS, INTERVENTIONS, AND OUTCOMES		
Study setting	9	Description of study settings (for example, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (for example, surgeons, psychotherapists)
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (for example, drug dose change in response to harms, participant request, or improving/worsening disease)
	11c	Strategies to improve adherence to intervention protocols and any procedures for monitoring adherence (for example, drug tablet return, laboratory tests)
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial

Section/item	Item no.	Description
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (for example, systolic blood pressure), analysis metric (for example, change from baseline, final value, time to event), method of aggregation (for example, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended
Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see figure at <http://annals.org/article.aspx?articleid=1556168>)
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size
METHODS: ASSIGNMENT OF INTERVENTIONS (FOR CONTROLLED TRIALS)		
Allocation:		
Sequence generation	16a	Method of generating the allocation sequence (for example, computer-generated random numbers) and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (for example, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (for example, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence, until interventions are assigned
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions
Blinding (masking)	17a	Who will be blinded after assignment to interventions (for example, trial participants, care providers, outcome assessors, data analysts), and how
	17b	If blinded, circumstances under which unblinding is permissible and procedure for revealing a participant's allocated intervention during the trial
METHODS: DATA COLLECTION, MANAGEMENT, AND ANALYSIS		
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (for example, duplicate measurements, training of assessors) and a description of study instruments (for example, questionnaires, laboratory tests), along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (for example, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol
Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol
	20b	Methods for any additional analyses (for example, subgroup and adjusted analyses)
	20c	Definition of analysis population relating to protocol non-adherence (for example, as randomized analysis) and any statistical methods to handle missing data (for example, multiple imputation)
METHODS: MONITORING		

Section/item	Item no.	Description
Data monitoring	21a	Composition of Data Monitoring Committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported AEs and other unintended effects of trial interventions or trial conduct
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor
ETHICS AND DISSEMINATION		
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval
Protocol amendments	25	Plans for communicating important protocol modifications (for example, changes to eligibility criteria, outcomes, analyses) to relevant parties (for example, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorized surrogates, and how (see item 32)
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained, in order to protect confidentiality before, during, and after the trial
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site
Access to data	29	Statement of who will have access to the final trial dataset and disclosure of contractual agreements that limit such access for investigators
Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, health care professionals, the public, and other relevant groups (for example, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions
	31b	Authorship eligibility guidelines and any intended use of professional writers
	31c	Plans, if any, for granting public access to the full protocol, participant level dataset, and statistical code
APPENDICES		
Informed consent materials	32	Model consent form and other related documentation given to participants and authorized surrogates
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable

* Reproduced with permission of the SPIRIT group. It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Statement (Chan et al., 2013a), in order to fully understand the scope and context of the checklist. It is important to note that this is a minimum list of items, and certain trial protocols may warrant the inclusion of additional items. This table 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/.

Bookshelf ID: NBK305500