4.5: 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.