16.7: Ensuring data of high quality

To be able to derive reliable and accurate conclusions from a health intervention trial, it is important to ensure that all processes and procedures, at all stages in the conduct of the trial, are performed at high quality. The many steps involved in planning and carrying out trials are described in the other chapters of this book. Here, we focus on the actions needed to ensure that all data collected are of high quality and that this high quality can be demonstrated both to those directly involved in the trial and to all those external to the trial but who have responsibilities or interests in relation to the trial. The general principles and some of the terminology that is commonly used related to what is called ‘Good Clinical Practice’ or ‘GCP’ will be described, but this chapter is not a GCP manual. Investigators who require formal training in GCP should contact a local internationally accredited institution that offers such training or one of the many internationally accredited online courses that are available from groups such as the Clinical Research Network of the United Kingdom (UK)’s National Institute for Health Research (<http://www.crncc.nihr.ac.uk>) or the OnlineGCP Group (<http://www.onlinegcp.com>).

If high-quality data are to be achieved, the investigator and all of the trial team must accept the need for rigour in the collection of all data and in the checks built into every step of the trial.

Figure 16.1 Example of an organizational timetable for a field trial.
7.1 Regulatory requirements and good clinical practice

The ICH (1996) (<http://www.ich.org>) is an internationally accepted set of standards that are intended to apply to all research on human subjects. It is mainly applied in the context of trials of medicinal products, but increasingly there is an expectation that observational epidemiological studies will be conducted to a similar standard. The aim of the guidelines is to ensure the safety and rights of all participants in the research study, while, at the same time, ensuring that the study is likely to achieve valid and reproducible results.

Based on the ICH–GCP guidelines, regulatory bodies, such as the US Food and Drug Administration (<http://www.fda.gov>) and the European Medicines Agency (<http://www.ema.europa.eu/ema>) have set out a rigorous series of procedures and checks that must be followed in clinical trials of new drugs and vaccines to provide the standard of evidence necessary for the licensing of a new product. There is a widespread misconception that all trials (including field trials of social or public health interventions or of alternative delivery mechanisms for licensed drugs or other medical products) must meet all requirements of such regulatory bodies—often called being ‘fully GCP-compliant’. This is not the case, and some flexibility is appropriate as to exactly how closely the GCP guidelines are implemented for non-licensing trials (i.e. of an intervention for which a licence is not being sought from a regulatory agency). However, all trials should comply with the basic principles contained within the ICH–GCP guidelines. The basic principles are that all studies involving human subjects should be conducted ethically (including that the interests of participants should be central to the trial design and implementation) and that all data collected should be of high quality and be likely to be valid. Furthermore, the investigators must be able to demonstrate that both these fundamental principles have been met. A trial can comply with the principles of GCP without meeting all the regulatory requirements for the licensing of a new product. This is important, since the full regulatory requirements are very demanding and will greatly increase the cost and human resources required. At an early stage in the planning of any trial, and certainly before any proposal is submitted to a funding agency or ethics committee, the PI and sponsor must make a clear decision as to whether their proposed trial needs to be ‘fully GCP-compliant’. As discussed in Chapter 2, many field trials in LMICs do not test investigational products but test the effectiveness of alternative delivery strategies for licensed products, or test
interventions that do not include any medicinal products at all, such as trials of health promotion or other public health interventions.

The key individuals and institutions that have responsibility for ensuring that a trial is complying with the principles of GCP have been defined in Chapter 7. The trial sponsor has overall responsibility for all aspects of the trial; the PI has primary responsibility for ensuring that the trial is carried out according to protocol; and the ethics committee (sometimes called the Institutional Review Board (IRB)) has primary responsibility for monitoring the ethical aspects of the trial. To ensure trial data are of high quality, the sponsor ‘should determine the appropriate extent and nature of monitoring which should be based on considerations such as the objective, purpose, design, complexity, blinding, size, and endpoints of the trial’ (International Conference on Harmonisation, 1996).

### 7.2 Supervision and data checks

Although, in some large trials, someone is designated to be the overall quality manager, the entire trial team should have data quality at the forefront of their minds. From the start, the investigator should assign quality assurance tasks to the team and build quality assurance (QA) processes into the trial procedures.

The two key principles for obtaining high-quality data are to plan ahead and to check everything. Nothing should be taken on trust, and, while remaining optimistic, it should be assumed that anything that could go wrong might go wrong!

Key issues in data quality are covered in other chapters: clear case definitions and valid measures of all trial outcomes in Chapter 12; preliminary studies and pilot tests in Chapter 13; questionnaire design, selection, and training of fieldworkers in Chapter 14; and some of the checks that can be done on the quality of the data collected are given in Chapter 20. In this chapter, we focus on steps that can be taken once fieldworkers have been deployed at the end of their initial training to ensure both that the data that they collected is of high quality and that this high quality can be demonstrated to external trial auditors. Most of these activities fall under field supervision.

In successful field supervision, prevention is better than cure. A supervision system should be designed not only to detect problems and provoke responses to solve them, but also to prevent problems from occurring in the first place. For example, if fieldworkers know that every piece of data they collect might be checked but have no way to know which pieces will actually be checked, they are more likely to always be careful. Conversely, if the fieldworker knows that only data collected on a Tuesday will be checked, then he or she may be less conscientious on other days.

Also, it is important to institute a system for checking all data collected, and especially data critical for identifying and linking data on the same individual throughout the trial. Examples of checks that should be built into field supervision include checks of completed forms, observation of work, replicated collection of a sample of data, checks without repeated data collection, review of errors detected after data collection, and checks with participants and community representatives.

The record forms that each fieldworker completes should be checked for accuracy and completeness. Because of delays between data collection and entering the data into a computer, if paper forms are used, some preliminary checks should be done before the forms are submitted, while the fieldworker is still in the vicinity of the participant, and before the participant’s situation may have changed. When the data are directly entered electronically, checks for data completeness, range, and consistency can be incorporated into the data capture program.
Each fieldworker should receive regular scheduled visits from their supervisor, during which the supervisor observes them carrying out their routine data collection tasks and gives them constructive feedback and a chance to discuss any issues that they have faced. These scheduled visits should be particularly frequent during the early phases of the trial. Observation tests whether the fieldworker knows how to carry out their tasks (competence) and can do so when being observed, and provides an opportunity for the supervisor to identify and correct any problems with their understanding of how the data should be collected. For example, they may have misunderstood how to measure a child’s height or may not be asking questions exactly as they are written in the questionnaire. However, it does not show whether they actually do so when they are not being observed (performance) (see later in this section).

Throughout a survey, it is important to monitor the performance of each interviewer and to institute corrective training, if required. One means of quality control (QC) is to organize for a proportion of respondents (selected at random) to be re-interviewed by another interviewer. Discrepancies in the two interviews may identify deficiencies in the interview methods of one or other interviewer. It is not an uncommon experience in large surveys that some interviewers complete some questionnaires without ever having seen the ‘respondents’. A good system of checking, supervision, and QC is necessary to prevent this, or at least to detect it soon after it occurs, so that remedial action can be taken.

To check whether the fieldworker actually collects the data correctly when not being observed, unscheduled checks need to be implemented. For example, in a field trial of vitamin A supplementation in northern Ghana, each fieldworker received unscheduled, as well as the scheduled, visits from the supervisor, who would ask permission to sit in on any interview that was happening or about to happen when they arrived and then collect the forms that the interviewer had completed earlier that day. These would be sealed in an envelope in front of the fieldworker. The supervisor would then go back to the previous five households that had been visited that day. In two households, they would merely ask the household whether the fieldworker had actually visited them, conducted an interview with an appropriate person, and taken the appropriate biological specimens from them. This checked that the fieldworker was not fabricating the data. In the other three households, they would request an independent partial re-interview of the trial participant. When they returned to the trial office, the supervisor would then submit their own forms and the original forms collected by the fieldworker, and the data centre would generate a comparison of the two.

Various checks can, and should, be carried out after data collection. For example, all data incompleteness (for example, missing items on the questionnaire) or variables that are out of range (for example, an infant’s weight being recorded as 100 grams) or inconsistent (for example, a woman recorded with penile warts or a person recorded with fever in one part of the questionnaire but afebrile in another) should be identified. It is useful for all such errors to be tabulated on a regular basis by the fieldworker and the field team. Such tabulations will show which fieldworkers or field teams have more errors detected. The reasons for these can then be investigated, and steps put in place to rectify them. One method that has been used for this has been to send data queries back to the fieldworker. For example, if a check shows that a participant’s height is lower than it was in a previous study round, the fieldworker can be asked to go back to collect that participant’s height again. Ideally, the fieldworker should not be told why they are being asked to re-collect the height, let alone what they had entered the height as during their recent visit or during the previous round. In the field trial of vitamin A, this method was extended, so that each fieldworker received some such requests, even when there was no reason to suspect an error in the data. These checks occasionally identified errors that were not detectable by routine range or consistency checks.

Finally, it is important that the trial team has periodic meetings with participants or their representatives and with other members of the trial communities to check that they are happy with the activities of the fieldworkers and their
An important principle is that every error or problem that is detected should provoke a response. This is for two reasons. First, if errors are not investigated and acted on, the effort of detecting them is wasted, and also the field staff may interpret this to imply that the importance of data quality is being neglected. Second, since it is never possible to check every piece of data collected, any errors that are detected are likely to be the ‘tip of the iceberg’.

The actions taken when errors are detected should not generally be punitive but should include support and further training to help the fieldworker improve. However, if this fails to correct the problem, or if the errors have come about through data fabrication, disciplinary mechanisms should be in place, and ultimately these may need to include termination of employment.

It is important that field staff are aware of all the types of checks that will be conducted on the data they collect. This is partly to avoid their feeling that they have been spied on behind their backs, but also so that they will be encouraged to ensure that all data are collected as well as possible.

It is a good plan to have weekly fieldwork meetings which include reports from individuals, on progress, work accomplished, identified problems and how they were solved, queries, etc. This also provides an opportunity for systematic feedback from the central administration on fieldworker performance, including results of repeat interviews for quality checks. Meetings of this kind may greatly assist in maintaining staff morale and improving the quality of the data collected.