17.7: Coding and linkage of results

In order to remove the possibility of bias, staff working in the laboratory should not know which trial arm any sample is from, and it should not be possible for this to be deduced from the labelling system employed. Specimens must be labelled in such a way, however, that each is identified uniquely, and any test results can be linked back to other records of the individual from whom the specimen was taken. While this seems to be stating the obvious, the problems that arise with these aspects of large studies are often substantial. Special care is necessary in longitudinal studies where individuals may be followed for many years, in studies involving many different research groups or laboratories, and in studies where results need to be linked with census information that may be updated over time (for example, individuals may move house, and this may cause problems if the coding system for individuals is too closely linked to a house code). Pre-printed labels are highly advantageous.

Laboratory results will usually be recorded in laboratory books or on specially prepared forms for data entry. Where the machine used for a particular test prints out the results, these should be carefully transcribed on to data forms, preferably using double entry (see Chapter 20, Section 5.1), and the printed output stored. Some machines generate printouts on heat-sensitive paper. In this case, a heat-stable photocopy must be made and stored. Increasingly, electronic record keeping will render these particular storage methods obsolete. Result codes that identify particular problems or features, such as lost and broken samples, technical problems with batches of samples (for example, staining, storage, transport), and the identification of the technicians involved with each test (to check variations between observers) should be used. Errors in readings on some automated machines (for example, values outside the normal range) will be reported or ‘flagged’ immediately, so that the assay can be repeated, if necessary.

If the study uses laboratory numbers, in addition to individual identification numbers, as is often the case, both numbers should be entered on a computer form for data entry, so that cross-checks and data linkage can be done in the computer.
If multiple laboratory tests are being performed on samples from the trial population, it may be best to wait until all the results have been assembled and collated before entering them into the computer, so that the checking and linkage back to other data on each individual can be done in relatively few steps. This will depend on how the data entry system is organized, but repeated processing of many small sets of data is liable to lead to confusion and may be unnecessarily time-consuming. However, a compromise may be necessary if results are needed in a timely manner for selection for QC or QA checks so that they can be used for the clinical care of the participant.