CHAPTER 10

Data Collection and Quality Control

No study is better than the quality of its data. In clinical trials data are collated from several sources – interviews, questionnaires, subject examinations or laboratory determinations. Also, data that have been collected and evaluated by someone outside the study can be used in a trial; for example, diagnoses obtained from death certificates or hospital records.

Quality control starts with clear definitions of response variables and procedures and with training; it ends with data collection, editing and assessment. Results from any experiment, either based on poorly standardized procedures that use ambiguous definitions or conducted by insufficiently trained staff with limited knowledge about the study protocol, can lead to erroneous results and conclusions. This chapter will review why problems in data collection arise, and provide some general solutions. The section on quality monitoring emphasizes issues that need to be considered in drug trials.

FUNDAMENTAL POINT

During all phases of a study, sufficient effort should be spent to ensure that all key data are of high quality.

PROBLEMS IN DATA COLLECTION

Problems in data collection can be of several sorts and can apply to the initial acquisition of data (eg, physical examination) as well as to the recording of the data on a form or data entry into a remote computer terminal or microcomputer. First, incomplete and irretrievable data can haunt the investigator. Such a situation arises, for example, from the inability of subjects to provide necessary information, from inadequate physical examinations, from laboratory mishaps or from carelessness in study form completion or data entry. The percent of missing data in a

study can be considered as one indicator of the quality of the data and, therefore, the quality of the trial.

Second, erroneous data may not be recognized and, therefore, can be even more troublesome than incomplete data. For study purposes, a specified condition may be defined in a particular manner. A clinic staff member may unwittingly use a clinically acceptable definition, but one that is different from the study definition. Specimens may be mislabeled. In one clinical trial, the investigators suspected a mislabeling when, in a glucose tolerance test, the fasting glucose levels were higher than the one-hour glucose levels in some subjects. Badly calibrated equipment can be a source of error. In addition, the wrong data may be entered on a form. A blood pressure of 84/142 mm Hg, rather than 142/84 mm Hg, is easy to identify as wrong. However, while 124/84 mm Hg may be incorrect, it is perfectly reasonable, and the error would not necessarily be recognized.

The third problem is variability in the observed characteristics. Variability reduces the opportunity to detect real changes. The variability between repeated assessments can be unsystematic (or random), systematic, or a combination of both. Variability can be intrinsic to the characteristic being measured, the instrument used for the measurement, or the observer responsible for obtaining the data. The problem of variability is not unique to any specific field of investigation.[1,2] Reports of studies of repeat chemical determinations, determinations of blood pressure, physical examinations, and interpretations of x-rays, electrocardiograms and histological slides[3-14] indicate the difficulty in obtaining reproducible data. People perform tasks differently and may vary in knowledge and experience. These factors can lead to inter-observer variability. In addition, inconsistent behavior of the same observer between repeated measurements may also be much greater than expected. While less than inter-observer variability, intra-observer inconsistency nevertheless can be appreciable.

In 1947, Belk and Sunderman[3] reviewed the performance of 59 hospital laboratories on several common chemical determinations. Using prepared samples, they found that "unsatisfactory results outnumbered the satisfactory." In 1969, Lewis and Burgess[4] assessed interlaboratory measures of red blood cell count using two methods (visual and electronic). The ranges for both methods were extremely broad. Results for the visual method varied from 2.2×10^6 *RBC*/mm^3 to 5.1×10^6 *RBC*/mm^3 and for the electronic method from 0.7×10^6 *RBC*/mm^3 to 4.7×10^6 *RBC*/mm^3. In 1978, others[5] looked at six selected laboratories. The overall performance was reasonably good. However, performance on simulated patient specimens was worse than when designated quality-control specimens were analyzed. This indicates that when special attention is given to the analyses, laboratories perform better. Classification of histologic specimens can also be highly variable. Feinstein and col-

ues[6] reviewed interpretation of cellular types of lung cancer by five erienced pathologists. On second readings, pathologists disagreed h their first diagnoses up to 20% of the time. When Davies[7] looked at ding of electrocardiograms by nine experienced cardiologists, he nd large inter-observer and intra-observer disagreement. On the rage, on re-reading electrocardiograms, the cardiologists disagreed h one in eight of their original interpretations. A study that assessed ical findings of the respiratory system[8] concluded that "inter-erver repeatability of respiratory signs falls midway between chance d total agreement."

Regardless of the source of variability, several factors may have an pact on the magnitude of variability. Vagueness in definitions, inade-uate methodology, lack of training of personnel and carelessness can all crease the variation inherent in any measurement.

MINIMIZING POOR QUALITY DATA

General approaches for minimizing potential problems in data collection are summarized below. Most of these should be considered during the planning phase of the trial. Examples in the cardiovascular field are provided by Rose and Blackburn.[15]

Clear definitions of entry and diagnostic criteria and methodology are essential. These should be written so that all investigators and staff can apply them in a consistent manner throughout the trial. The same question can be interpreted in many ways. Even the same investigator may forget how he previously interpreted a question unless he can readily refer to instructions and definitions. Accessibility of these definitions is also important. Ideally, the study forms should contain all necessary information. If that is not possible, the forms should outline the key information and refer the investigator to the appropriate page in a manual of procedures. Such a manual of procedures should be prepared in every clinical trial. This document provides detailed answers to all conceivable "how to" questions. Although it may contain information about study background, design, and organization, the manual of procedures is not simply an expanded protocol. In addition to listing eligibility criteria and response variable definitions, it should indicate how the criteria and variables are determined. Most important, the manual needs to describe the subject visits—their scheduling and content—in detail. Instructions for filling out forms, performing tasks such as laboratory determinations, drug ordering, storing and dispensing, and compliance monitoring must be accurate and complete. Finally, recruitment techniques, informed consent, subject safety, emergency unblinding, use of concomitant therapy and other issues need to be addressed. Updates and

clarifications usually occur during the course of a study. These revisions should be made available to every staff person involved in data collection.

Descriptions of laboratory methodology and the ways the results are to be reported also need to be stated in advance. In one study, plasma levels of the drug propranolol were determined by using four methods. Only after the study ended was it discovered that two laboratories routinely were measuring free propranolol and two other laboratories were measuring propranolol hydrochloride. A conversion factor allowed investigators to make simple adjustments and arrive at legitimate comparisons. Such adjustments are not always possible.

Training sessions for investigators and staff to promote standardization of procedures are crucial to the success of any large study. Whenever more than one person is filling out forms or examining subjects, training sessions help to minimize error. There may be more than one correct way of doing something in clinical practice, but for study purposes, there is only one way. Similarly, the questions on a form should always be asked in the same way. The answer to, "Have you had any chest pain in the last three months?" may be different from the answer to, "You haven't had any chest pain in the last three months, have you?" Training laboratory personnel is equally important. Two technicians may use slightly different techniques. These differences can lead to confusing results. Kahn and colleagues[16] reviewed the impact of training procedures instituted in the Framingham Eye Study. The two days of formal training included duplicate examinations and discussions about differences, and the use of a reference set of fundus photographs.

Mechanisms to verify that all clinic staff do things the same way should be developed. These could include instituting certification procedures for specified types of data collection. If blood pressure, electrocardiograms, pulmonary function tests or laboratory tests are important, the people performing these determinations should not only be trained, but also be tested and certified as competent. Periodic retraining and certification are especially useful in long-term studies since people tend to forget, and there is likely to be personnel turnover. For situations where staff must conduct clinical interviews, special training procedures to standardize the approach have been used.[17]

Well-designed forms will minimize errors and variability. Forms should be as short and as well organized as possible, with a logical sequence to the questions. Forms should be clear, with few "write-in" answers. As little as possible should be left to the imagination of the person completing the form. This means, in general, no essay questions. The questions should elicit the necessary information and little else. Questions which are tacked on because the answers would be "nice to know" are rarely analyzed and may distract attention from pertinent

questions. In several studies where death is the primary response variable, investigators have expressed interest in learning about the circumstances surrounding the death. In particular, the occurrence of symptoms before death, the time lapse from the occurrence of such symptoms until death, and the activity and location of the subject at the time of death have been considered important. While this may be true, focusing on it has led to the creation of extraordinarily complex forms which take considerable time to complete. Moreover, questions arise concerning the accuracy of the information, because much of it is obtained from sources who may not have been with the subject when he died. Unless investigators clearly understand how these data will be used, simpler forms are preferable. The experience regarding forms and recommendations from the Coronary Drug Project have been reported.[18] Wright and Haybittle have also provided general guidelines to forms design for clinical trials.[19]

Pretesting of forms and procedures is useful. Several people similar to the intended subjects should participate in a simulated interview and examination to make sure procedures are properly performed and questions on the forms flow well and provide the desired information. Furthermore, by pretesting, the investigator grows familiar and comfortable with the form. Fictional case histories can be used to check form design and the care with which forms are completed. When developing forms, most investigators cannot conceive of the numerous ways questions can be misinterpreted until several people have been given the same information and asked to fill out the same form. Part of the reason for different answers is undoubtedly due to carelessness. Misinterpretation may not be detected when forms are filled out on real subjects. Anyone editing a form has no way of identifying errors which are not completely unreasonable. Inadequacies in form structure and logic can also be uncovered by use of pretesting. In conclusion, pretesting reveals areas where forms might be improved and where additional training might be worthwhile.

There is little point in constructing fictional case histories unless there is an opportunity for follow-up discussion. This helps people completing the forms to understand how the forms are meant to be completed and what interpretations are wanted. Discussion also alerts them to carelessness. When done before the start of the study, follow-up discussion allows the investigator to modify inadequate items on forms. These case history exercises might be profitably repeated several times during the course of a long-term study to indicate when education and retraining are needed. Ideally, forms should not be changed after the study has started. Inevitably, though, modifications are made. Pretesting can help to minimize this.

Both variability and bias in the assessment of response variables

should be minimized through repeat assessment, blinded assessment, or (ideally) both. At the time of the examination of a subject, for example, an investigator may determine blood pressure two or more times and record the average. Performing the measurement without knowing the group assignment helps to minimize bias. In unblinded or single-blinded studies, the examination might be performed by someone other than the investigator. For blood pressure, another method of minimizing bias might be the use of a random zero device.[13] In assessing slides, x-rays, or electrocardiograms, two individuals can make independent, blinded evaluations, and the results can be averaged or adjudicated in cases of disagreement. Independent evaluations are particularly important when the assessment requires an element of judgment. Classification of response variables such as cause of death or nonfatal events can be performed in a similar manner.

The introduction of microcomputers into clinical trials has the potential for improving data quality. Programs have been developed that identify missing, extreme or inconsistent values and which prohibit further data entry until a correction has been made. In cases where an investigator must go back to the subject to check the information, this aspect is particularly valuable, because the error is identified rapidly. Double entry of data is also used to reduce the error rate. Adequate training of staff is essential for fully realizing the advantages of this technology.

An issue being debated is whether forms can be eliminated. Typically, a paper form is completed and the data transferred to the microcomputer. Thus, a record trail is available for data verification and audit. If only the final entered data are available, there is no assurance that the data have not been altered inappropriately.

Programs can be developed which will ensure that both original and revised data are saved. Thus, a computerized audit trail can be developed. In such a case, it is conceivable that an investigator can dispense with paper forms.

In general, there has been a favorable experience with entering clinical trial data into microcomputers. Error rates have been low and corrections have been minimal.[20]

QUALITY MONITORING

Even though every effort is made to obtain high quality data, a monitoring or surveillance system is crucial. When errors are found, a monitoring system enables the investigator to take corrective action. In order to accomplish this, monitoring needs to be current. What is more, monitoring allows an assessment of data quality when interpreting study

results. Numerous forms and procedures can be monitored, but doing so is usually not feasible. Rather, monitoring those areas most important to the trial is recommended. Form completion, procedures and drug handling also need to be monitored.

During the study, all forms should be checked for completeness, internal consistency and consistency with other forms. On a follow-up visit to evaluate a subject's progress, the investigator might want to know whether the subject has had a myocardial infarction since the previous follow-up visit. If the subject has had such an event, then more information about the infarction can be collected by completing a special event form. The number of nonfatal myocardial infarctions listed on follow-up visit forms and special event forms should agree. When the forms disagree, the person or group responsible for ensuring consistent and accurate forms should question the person filling out the forms. Consistency within a given form can also be evaluated. Dates and times are particularly prone to error.

It may be important to look at consistency of data over time. A subject with a missing leg on one examination was reported to have palpable pedal pulses on a subsequent examination. Cataracts which did not allow for a valid eye examination at one visit were not present at the next visit, without surgery having been performed. The data forms may indicate extreme changes in body weight from one visit to the next. In such a case, changing the data after the fact is likely to be inappropriate because the correct weights may be unknown. However, the investigator can take corrective action for future visits by more carefully training his staff. Sometimes, mistakes can be corrected. In one trial, comparison of successive electrocardiographic readings disclosed gross discrepancies in the coding of abnormalities. The investigator discovered that one of the technicians responsible for coding the electrocardiograms was fabricating his readings. In this instance, correcting the data was possible.

Someone needs constantly to monitor completed forms to find evidence of missing subject visits, or visits that are off schedule in order to correct any problems. Frequency of missing or late visits may be associated with the intervention. Differences between groups in missed visits may bias the study results. Monitoring and editing of forms is often insufficient and, to improve upon results, it may be necessary to observe actual clinic procedures. Observing clinic procedures is particularly important in multicenter trials.[21]

Extreme laboratory values should also be checked. Values incompatible with life (eg, potassium of 10 mEq/1) are obviously incorrect. Other, less extreme values (ie, cholesterol of 125 mg/dl in male adults in the United States) should be questioned. They may be correct, but it is unlikely. Finally, values should be compared with previous ones from the same subject. Certain levels of variability should be present, but when

these levels are exceeded, the value should be flagged. For example, unless the study involves administering a lipid-lowering therapy, any determination which shows a change in serum cholesterol of perhaps 20% or more from one visit to the next should be repeated. Repetition would require saving samples of serum until the analysis has been checked. As well as checking results, a helpful procedure is to monitor submission of laboratory specimens to ensure that missing data are kept to a minimum.

Investigators doing special procedures (laboratory work, electrocardiogram reading) need to have an internal quality control system. Such a system should include re-analysis of duplicate specimens or materials at different times in a blinded fashion. A system of resubmitting specimens from outside the laboratory or reading center might also be instituted. As noted by McCormick and colleagues,[5] these specimens need to be indistinguishable from actual study specimens. An external surveillance system, which should be established in the planning phase of a trial, can pick up errors at many stages (specimen collection, preparation, transportation, and reporting of results), not just at the analysis stage. Thus, it provides an overall estimate of quality. Unfortunately, the system most often cannot indicate at which step in the process errors may have occurred. The external quality control programs implemented in the Coronary Drug Project have been described by Canner et al.[22] Another example is provided by the National Cooperative Gallstone Study.[23]

All recording equipment should be checked periodically. Even though initially calibrated, the machines can break down or require adjustment. Scales can be checked by means of standard weights. If aneroid sphygmomanometers are used, they should be compared regularly with a mercury sphygmomanometer. Factors such as linearity, frequency response, paper speed and time constant should be checked on electrocardiograph machines. In one long-term trial, the prevalence of specific electrocardiographic abnormalities was monitored. The sudden appearance of a three-fold increase in one abnormality, without any obvious medical cause, led the investigator to correctly suspect electrocardiograph machine malfunction.

In a drug study, the quality of the drug preparations should be monitored throughout the trial. Monitoring, as such, includes periodically examining containers for possible mislabeling and for proper contents (both quality and quantity). It has been reported[24] that in one trial, "half of the study group received the wrong medication" due to errors at the pharmacy. Investigators should carefully look for discoloration and breaking or crumbling of capsules or tablets. When the agents are being prepared in several batches, samples from each batch should be examined and analyzed. Occasionally, monitoring the number of pills or capsules per bottle is useful. The actual bottle content of pills should not

vary by more than one or two percent. The number of pills in a bottle is important to know because pill count may be used to measure compliance of subjects.

Another aspect to consider is the storage shelf life of the preparations and whether they deteriorate over time. Even if they retain their potency, do changes in odor (as with aspirin) or color occur? If shelf life is long, preparing all agents at one time will minimize variability. Of course, in the event that the study ends prematurely, there may be a large supply of unusable drugs. Products having a short shelf life require frequent production of small batches. Complete records should be maintained for all drugs prepared, examined and used. Ideally, a sample from each batch should be saved. After the study is over, questions about drug identity or purity may arise and samples will be useful.

The dispensing of medication should also be monitored. Checking has two aspects. First, were the proper drugs sent from the pharmacy or pharmaceutical company to the clinic? If the study is double-blind, the clinic staff will be unable to check on this. They must assume that the medication has been properly coded. However, in unblinded studies, staff should check to assure that the proper drugs and dosage strengths have been received. In one case, the wrong strength of potassium chloride was sent to the clinic. The clinic personnel failed to notice the error. An alert subject to whom the drug was issued brought the mistake to the attention of the investigator. Had the subject been less alert, serious consequences could have arisen. An investigator has the obligation to be as careful about dispensing drugs as is a licensed pharmacist. Close reading of labels is essential, as well as documentation of all drugs that are handed out to subjects.

Second, when the study is blinded, the clinic personnel need to be absolutely sure that the code number on the container is the proper one. Labels and drugs should be identical except for the code; therefore, extra care is essential. If bottles of coded medication are lined up on a shelf, it is relatively easy to pick up the wrong bottle accidentally. Unless the subject notices the different code, such errors may not be recognized. Even if he is observant, he may assume that he was meant to receive a different code number. The clinic staff should be asked to note on a study form the code number of the bottle dispensed and the code number of bottles that are returned by the subject. Theoretically, that should enable investigators to spot errors. In the end, however, investigators must rely on the care and diligence of the staff person dispensing the drugs.

It may be worthwhile periodically to send study drug samples to a laboratory for analysis. Although the center responsible for packaging and labeling drugs should have a "foolproof" scheme, independent laboratory analysis serves as an additional check on the labeling process.

The drug manufacturer assigns lot, or batch, numbers to each batch

of drugs that are prepared. If contamination or problems in preparati
are detected, then only those drugs from the problem batch need to
recalled. This is especially important in clinical trials, since the recall
all drugs can severely delay, or even ruin, the study. When only so
drugs are recalled, the study can usually manage to continue. Therefo
the lot number of the drug as well as the name or code number should
listed in the subject's study record.

Finally, monitoring of data quality proves most valuable when ther
is feedback to the clinic staff and technicians. Once weaknesses an
errors have been identified, performance can be improved. Chapter 1
contains several tables illustrating quality control reports. With carefu
planning, reports can be provided and improvement can be accomplished
without unblinding the staff. All quality control measures take time an
money; it is thus impossible to be compulsive about the quality of every
piece of datum and every procedure. Investigators need to focus their
efforts on those procedures which yield key data; those on which the conclusions of the study critically depend.

REFERENCES

1. Koran, L.M. The reliability of clinical methods, data and judgments. Part 1. *N Engl J Med.* 293:642–646, 1975.
2. Koran, L.M. The reliability of clinical methods, data and judgments. Part 2. *N Engl J Med.* 293:695–701, 1975.
3. Belk, W.P., and Sunderman, F.W. A survey of the accuracy of chemical analyses in clinical laboratories. *Am J Clin Pathol.* 17:853–861, 1947.
4. Lewis, S.M., and Burgess, B.J. Quality control in haematology: report of interlaboratory trials in Britain. *Br Med J.* 4:253–256, 1969.
5. McCormick, W., Ingelfinger, J.A., Isakson, G. et al. Errors in measuring drug concentrations. *N Engl J Med.* 299:1118–1121, 1978.
6. Feinstein, A.R., Gelfman, N.A., Yesner, R. et al. Observer variability in the histopathologic diagnosis of lung cancer. *Am Rev Respir Dis.* 101:671–684, 1970.
7. Davies, L.G. Observer variation in reports on electrocardiograms. *Br Heart J.* 20:153–161, 1958.
8. Smyllie, H.C., Blendis, L.M., and Armitage, P. Observer disagreement in physical signs of the respiratory system. *Lancet* 2:412–413, 1965.
9. Assaad, F.A., and Maxwell-Lyons, F. Systematic observer variation in trachoma studies. *Bull WHO* 36:885–900, 1967.
10. Birkelo, C.C., Chamberlain, W.E., Phelps, P.S. et al. Tuberculosis case finding: a comparison of the effectiveness of various roentgenographic and photofluorographic methods. *JAMA* 133:359–366, 1947.
11. Detre, K.M., Wright, E., Murphy, M.L. et al. Observer agreement in evaluating coronary angiograms. *Circulation* 52:979–986, 1975.
12. Frieden, J., Shapiro, J.H. and Feinstein, A.R. Radiologic evaluation of heart size in rheumatic heart disease. *Arch Intern Med.* 111:90–96, 1963.
13. Rose, G.A., Holland, W.W., and Crowley, E.A. A sphygmomanometer for epidemiologists. *Lancet* 1:296–300, 1964.

14. Wilcox, J. Observer factors in the measurement of blood pressure. *Nurs Res.* 10:4–20, 1961.

15. Rose, G.A. and Blackburn, H. *Cardiovascular Survey Methods.* Geneva: World Health Organization, Monograph Series No. 56, 1968.

16. Kahn, H.A., Leibowitz, H., Ganley, J.P. et al. Standardizing diagnostic procedures. *Am J Ophthalmol.* 79:768–775, 1975.

17. Russell, M.L., Ghee, K.L., Probstfield, J.L., et al. Development of standardized simulated patients for quality control of the clinical interview. *Controlled Clin Trials* 4:197–208, 1983.

18. Knatterud, G.L., Forman, S.A., and Canner, P.L. Design of data forms. *Controlled Clin Trials* 4:429–440, 1983.

19. Wright, P., and Haybittle, J. Design of forms for clinical trials. *Br Med J.* 2:529–530, 590–592, 650–651, 1979.

20. Bagniewska, A., Black, D., Curtis, C., et al. Data quality control in a distributed data processing system: nature and method. (Abstract). *Controlled Clin Trials* 4:148, 1983.

21. Ferris, F.L., and Ederer, F. External monitoring in multiclinic trials: applications from ophthalmologic studies. *Clin Pharmacol Ther.* 25:720–723, 1979.

22. Canner, P.L., Krol, W.F., and Forman, S.A. External quality control programs. *Controlled Clin Trials* 4:441–466, 1983.

23. Habig, R.L., Thomas, P., Lippel, K., et al. Central laboratory quality control in the National Cooperative Gallstone Study. *Controlled Clin Trials* 4:101–123, 1983.

24. Quality assurance of clinical data: Discussion. *Clin Pharmacol Ther.* 25:726–727, 1979.