

Ensuring trial validity by data quality assurance and diversification of monitoring methods

Colin Baigent, Frank E Harrell, Marc Buyse, Jonathan R Emberson and Douglas G Altman

Clinical Trials published in 2008
DOI: 10.1177/1740774507087554

The online version of this article can be found at:
<http://ctj.sagepub.com/cgi/content/abstract/5/1/49>

Ensuring trial validity by data quality assurance and diversification of monitoring methods

Colin Baigent^a, Frank E Harrell^b, Marc Buyse^c, Jonathan R Emberson^a and Douglas G Altman^d

Errors in the design, the conduct, the data collection process, and the analysis of a randomized trial have the potential to affect not only the safety of the patients in the trial, but also, through the introduction of bias, the safety of future patients. Trial monitoring, defined broadly to include methods of oversight which begin when the study is designed and continue until it is reported in a publication, has a role to play in eliminating such errors. On-site monitoring can be extremely inefficient for the identification of errors most likely to compromise patient safety or bias study results. However, a variety of other monitoring strategies offer alternatives to on-site monitoring. Each new trial should conduct a risk assessment to identify the optimal means of monitoring, taking into account the likely sources of error, their consequences for patients, the study's validity, and the available resources. Trial management committees should consider central statistical monitoring a key aspect such monitoring. The systematic application of this approach would be likely to lead to tangible benefits, and resources that are currently wasted on inefficient on-site monitoring could be diverted to increasing trial sample sizes or conducting more trials.

Clinical Trials 2008; 5: 49–55. <http://ctj.sagepub.com>

Introduction

The main purpose of quality assurance methods applied to randomized trials should be to protect the rights and safety of trial participants and to reduce the likelihood that the trial results are affected by bias (thereby affecting the safety of future patients). Trial monitoring, defined broadly to include methods of oversight which begin when the study is designed and continue until it is reported in a publication, has a role to play in achieving these aims. The purpose of this paper, which resulted from discussions at the Sensible Guidelines workshop held in Washington, D.C. during January 2007, is to take a fresh look at the ways that different types of monitoring can improve the quality of inferences from data collected in a randomized trial. First, we consider the types of errors arising in randomized trials, and the circumstances under which they might introduce bias or adversely affect patient safety during the trial. Second, we discuss the types of monitoring available and their efficacy for detecting the types of errors we have identified. Third, we make suggestions about how risk assessment of each individual trial can be used to identify the likelihood of those errors, thereby guiding the formulation of an appropriate monitoring scheme for the trial. Finally, we consider where further research would help to improve study monitoring methods.

A taxonomy of errors affecting trials

We distinguish between four types of error which might affect the safety of trial participants or introduce the potential for bias in trial results (see Table 1).

^a Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Richard Doll Building, Old Road Campus, Roosevelt Drive, Oxford OX3 7LF

^b Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee,

^c International Drug Development Institute (IDDI), 30 Avenue Provinciale, 1340, Ottignies, Louvain-la-Neuve, Belgium,

^d Centre for Statistics in Medicine, Wolfson College Annexe, Linton Road Oxford, OX2 6UD

Author for correspondence: Marc Buyse, ScD IDDI Inc., 363 N. Sam Houston Parkway East, Suite 1100, Houston, TX 77060, USA, Email: marc.buyse@iddi.com

Table 1 Error types, their potential for adversely affecting patient safety or introducing bias, and suggested methods of monitoring for their detection in a large-scale multi-center randomized trial

Error type	Adverse effect on safety during trial	Bias in study results	Monitoring method
Design error	+++	+++	Peer review, and oversight by trial committees
Procedural error	+++	+ (blinded)	Avoidance through initial training and subsequent mentoring during site visits, preferably through direct observation
		+++ (unblinded)	
Recording error			
Random	–	+	Central statistical monitoring. On-site monitoring for sites with poor performance only
Fraud	+	–	Central statistical monitoring. On-site monitoring directed at sites with unusual data patterns
Analytical error	–	+++	Peer review, and oversight by trial committees

Design errors

Serious design errors in a randomized trial can adversely affect patient safety, compromise interpretation of study results (which may endanger future patients), or both. An example of a trial design error which might adversely affect patient safety within a trial would be a failure to specify the systematic recording of safety outcomes. This error could result in the Data Monitoring Committee being unable to recognize any large differences in safety outcomes that might emerge before the end of the trial. Design errors which introduce serious bias to the study findings may make a trial difficult or impossible to interpret. For example, a trial may be unable to address the intended hypothesis because of errors in the basic study design, such as an inadequate sample size or failure to ensure follow-up of all patients originally randomized. Similarly, a flawed method of randomization or potential for unblinding could introduce serious bias, which, unless discovered promptly, may also render interpretation of a trial’s findings impossible.

Prevention of such critical errors is essential, so it is important to ensure before a trial begins that the protocol and any relevant standard operating procedures are reviewed carefully by experienced investigators. Peer review is often provided in the course of obtaining financial support, but such review may have been directed at an early or summary protocol, and may not identify subtle yet serious problems. We suggest that one of the first tasks of the Trial Steering Committee and of the Data Monitoring Committee (see below) ought to be to provide a careful assessment of the adequacy of the trial

design through a review of study documentation and materials. The avoidance of critical design errors is at least as important as any form of monitoring during the recruitment and follow-up phase.

Procedural errors

We define a procedural error as a systematic failure to follow the study protocol or standard operating procedures. Such errors vary in their potential to affect patient safety or to introduce bias. Patient safety within the trial would be compromised, for example, if patients who are ineligible because of a previous adverse reaction to one of the drugs under study were to be randomized inappropriately. By contrast, inclusion of some patients who are technically ineligible because they are slightly too old or too young, or because some clinical parameter, such as blood pressure or tumor size, is marginally outside the specified range, is not generally likely to materially affect the safety or quality of the trial.

Although minor procedural errors (such as failing to maintain current versions of trial paperwork on file) are often the focus of site monitoring, it is unlikely that such errors will truly compromise patient safety within the trial, or bias study results. Procedural errors would be expected to introduce serious bias into the study findings only in specific instances. For example, a frequently repeated procedural failure that resulted in extensive misclassification of study outcomes might attenuate any measured treatment effect. A potentially more serious bias might arise, however, if a frequent procedural error was

non-random with respect to treatment allocation. Although such an error is highly unlikely in a properly randomized multicentre study in which there is concealment of treatment allocation, large-scale trials with an open design are potentially vulnerable to procedural errors if the outcome is subjective.

Data recording errors

Errors introduced when investigators provide inaccurate (including falsified) data in case report forms are called data recording errors. Such errors may result secondarily from procedural errors: for example, a patient's blood pressure may be recorded inaccurately in one site if an investigator does not calibrate a sphygmomanometer correctly.

The most important consideration for data recording errors is their potential to create serious bias in the trial findings. As we have already noted in the context of procedural errors, the potential for bias is greatest when such errors are non-random with respect to treatment allocation. It follows that there is a particular risk of bias when local staff have knowledge of the treatment allocation. Purely random errors, such as misreading a digit when transcribing a blood result, or rounding off the result, are unlikely to introduce serious bias but may still damage the trial if they are very extensive (possibly causing an attenuation of the measured treatment effect), or if subgroup results are an important part of the trial (in which case misclassification of baseline characteristics will tend to blunt any real trends in treatment effects that might occur with an increasing or decreasing level of some baseline covariate).

It is important to note that falsified data, while they might affect the safety of individual trial participants when treatments have serious toxicity (e.g., cancer chemotherapy), generally do not introduce serious bias into the overall trial findings when treatment allocation is properly concealed and follow-up is systematic. Nevertheless, even in blinded trials, study monitoring should aim to detect such transgressions in order to identify the individuals concerned, but—depending on the

design of the trial— this might be done most efficiently through central statistical review (see below). As we discuss below, central statistical monitoring of case report forms is remarkably efficient in the detection of fabricated data, whereas on-site monitoring may fail to uncover fraud.

Analytical errors

Errors introduced during the analysis and interpretation of trial results are called analytical errors. Although such errors are not generally regarded as a subject for monitoring, their detection should be part of quality assurance since they may affect the future use of an intervention, possibly resulting in the inappropriate use of an ineffective or hazardous treatment, or a failure to deploy an effective one. Although statistical errors do occasionally come to light, they are less common than simple errors of interpretation. For example, the lack of a statistically significant benefit of a treatment may sometimes be mistaken for evidence that a treatment is ineffective, when in truth the trial was too small to identify plausibly moderate effects (i.e., a Type II error)[1]. Overemphasis on subgroup analyses in preference to the main treatment effect [2] is another example of a statistical error.

Such overemphasis may be misleading even if the subgroup analyses had been pre-specified, since if a large enough number is planned then at least one is likely to be extreme. Statistical errors may be identified and corrected when the main publication is peer reviewed for a journal, but the Data Monitoring Committee, which is (or should be) entirely independent of the study investigators and sponsors, may also be able to help identify such errors before a paper is submitted [3,4].

Types of study monitoring

There are three basic categories of study monitoring [5]: oversight through trial committees; central monitoring by the data management center; and on-site monitoring.

Trial oversight committees

The appointment of experienced trial oversight committees before studies begin is the appropriate starting point for a secure system of trial monitoring. The exact arrangements for trial oversight may vary depending on trial size and complexity, but three types of committee are generally necessary: a Trial Management Committee; a Trial Steering Committee; and a Data Monitoring Committee [5].

Trial Management Committee (TMC)

The membership of the Trial Management Committee should generally include those responsible for the day-to-day running of the trial, such as the chief investigator, statistician, trial coordinator, nurse coordinator, and data manager.

This key committee bears the primary responsibility for monitoring the progress and conduct of the trial, and for ensuring that the protocol and standard operating procedures are followed. Most importantly, the TMC is responsible for ensuring that the protocol is accurately translated into appropriate study systems. These systems include recruitment and training of study sites and staff, implementation and maintenance of IT systems, arrangements for treatment supply and randomization (to ensure that both allocation concealment and proper randomization are maintained), laboratory procedures and quality assurance, and preparing information for other oversight committees and outside agencies.

Monitoring of data received by the TMC at the coordinating center should include checks that the data are complete, valid, and consistent with adherence to the trial protocol. The TMC should also ensure that a system is set up to enable tracking of recruitment rates, withdrawals, and losses to follow-up, so that early action can be taken if this information indicates that the trial may not be completed successfully. The TMC is therefore responsible for ensuring that systems are in place to identify all types of errors, and that a suitable strategy for monitoring is devised before the trial begins.

Trial Steering Committee (TSC)

The Trial Steering Committee should consist of the key members of the TMC and several other experienced investigators who have no direct involvement in the daily running of the trial.

Like the TMC, the TSC's role is to supervise the conduct of the trial, but it generally takes a more strategic and advisory role, thereby helping to ensure that the study is being conducted in accordance with the protocol and regulatory requirements. The TSC plays an important part in ensuring that the design of the trial is appropriate before the trial starts; this not only helps to avoid design errors, but may also involve identifying ways in which modifications to trial design can assist with improving the quality of trial data.

The TSC should also play a key role in ensuring that analytical errors are avoided when the study is published.

Data Monitoring Committee (DMC)

The purpose of a DMC is to protect the safety of trial participants (ensuring that any emerging hazards of an intervention do not clearly outweigh any benefits), to safeguard the credibility of the study, and to ensure the validity of study results [3,4,6]. The Committee reviews unblended interim results and is responsible for advising the TSC whether there are any reasons for modifying the protocol or discontinuing the trial in some or all participants. Although in practice many DMCs are not assembled until after the study design is complete, a DMC should ideally be able to comment on the study protocol and proposed methods before it begins. This may help to avoid lengthy delays should serious design errors later come to light.

Central monitoring

Central monitoring serves several purposes. First, it allows assessment of whether trial entry procedures and other aspects of the protocol are being followed (i.e., the avoidance of procedural errors). For example, if permitted by local laws, a requirement that trial consent forms are faxed to the central coordinating center may in some

circumstances allow a check of the consent process.

Second, central monitoring of data using statistical techniques may help to identify departures from expected patterns which might suggest incorrect procedures, or even data fraud, thereby identifying study sites requiring further investigation (which could include a site visit) [5]. Examples of central statistical monitoring techniques include [5,7]:

Checks for missing or invalid data

Range checks can be used to identify unlikely or implausible values in a trial participant's data, such as measured height below 1m or systolic below diastolic blood pressure. Electronic methods of data capture allow such checks to be programmed into the case report form.

Calendar checks

The day of the week on which a trial participant attended a study visit could be examined. For example, study visits may not be expected to occur on weekends, and this could provide an early indication of error.

Checks for unusual data patterns

Data from one site can be compared with data for the whole trial to identify patterns, such as digit preference, rounding, or unusual features of a frequency distribution (e.g., outliers, inliers, or atypical degrees of skewness or kurtosis). Such checks can be applied to either a single variable (e.g., systolic blood pressure) or to the joint distribution of several variables (e.g., systolic blood pressure and weight). More elaborate methods (such as checking for consistency with Benford's Law, which states that in lists of numbers from many real-life situations the leading digit is 1 almost one third of the time and larger numbers occur with steadily lower frequency) may also be applicable to some types of data and could be used to supplement simple techniques when there is a high level of suspicion about a particular site or sites.

Assessment of rates of reporting

The frequency and nature of reported adverse events and of missing data can be compared between study sites.

Checks of performance indicators

The possibility that there is error in procedure can sometimes be assessed by central monitoring of performance indicators, such as number of appointments per day, duration of visits, and delays in entering results or transferring data. Electronic data capture methods make all of these indicators easier to assess, although some (such as delays in returning results) may be monitored even if paper systems are in place.

Comparison with external sources

In some situations, it may be possible to validate information using external sources. For example, checks with birth and death registries or with disease-specific registries (e.g., cancer or diabetes registries) could be used to confirm that particular patients exist and that incident events have occurred.

There is potential for central monitoring to provide a more efficient and less costly means of monitoring than traditional methods of intensive on-site monitoring with 100% source data verification [7]. Central monitoring provides an enormously powerful technique for identifying fraud. It is very difficult to invent data with realistic distributional properties, so statistical review can provide a quantitative estimate of the likelihood that a particular dataset was generated fraudulently [8]. For example, in the Second European Stroke Prevention Study, data on 438 patients were fabricated at one site. This was first detected by statistical anomalies in the data and later confirmed by a central review of blood results: a visit to the site had failed to identify any transgression [9]. Central monitoring should ideally involve routine surveillance of site data as they accrue, with more detailed and specialized statistical testing if anomalies arise. There is considerable scope for using central monitoring more widely as a means of directing

visits to those sites where the a priori probability of irregularities is highest.

Site monitoring

Intensive site monitoring is often believed to be an absolute requirement for compliance with the International Conference on Harmonisation-Good Clinical Practice (ICH-GCP) guidelines, but in fact Section 5.18.3 of Guideline E6 acknowledges—though seriously undervalues—the potential for central statistical monitoring to play a prominent role in any monitoring plan [10]. It states:

“The sponsor should ensure that trials are adequately monitored. The sponsor should determine the appropriate extent and nature of monitoring. The determination of the extent and nature of monitoring should be based on considerations such as the objective, purpose, design, complexity, blinding, size and endpoints of the trial. In general there is a need for on-site monitoring, before, during, and after the trial; however, in exceptional circumstances, the sponsor may determine that central monitoring in conjunction with procedures such as investigators’ training and meetings, and extensive written guidance can assure appropriate conduct of the trial in accordance with GCP. Statistically controlled sampling may be an acceptable method for selecting the data to be verified.”

On-site monitoring is expensive, so it is important that its value to the trial is considered carefully. Central statistical monitoring should guide the frequency and content of on-site visits. Even in a trial with no on-site monitoring planned, if central monitoring suggested a high level of concern at a particular site, the appropriate action would be to visit that site.

Site monitoring should perhaps be regarded as ‘mentoring’, providing an opportunity for training and supporting study staff. On-site monitoring should generally involve:

- Ongoing training to ensure that knowledge and understanding of the study protocol and procedures is maintained, that problems are

resolved, and that staff remain enthusiastic and committed

- Verification that resources remain adequate within a site (e.g., that a room is available, that the pharmacy is able to store drugs, or that a laboratory can perform an assay that has to be done on-site)
- Checking adherence to the study protocol (for example, by reviewing a random subsample of patient consent forms or comparing trial information on serious adverse events with what has been recorded in medical notes)

Risk assessment should guide the monitoring plan

When designing a trial, the trial management committee should formulate the monitoring plan with due regard to the risk of the types of errors we have identified, their consequences for trial participants and future patients, the potential for them to be detected with central and on-site monitoring, and the available resources. Wherever possible, the study design should attempt to obviate the need for expensive on-site monitoring. The type and intensity of monitoring that is appropriate for a given trial will depend on the scope for bias arising from recording errors. In a trial in which the randomization process is secure, allocated treatments cannot be distinguished, and outcomes are assessed reliably, there is little scope for the main trial result to be seriously biased by data errors at individual trial sites, so it may be possible to limit on-site monitoring to those sites where central monitoring suggests that there might be a problem.

If the study design lacks either secure randomization, reliable outcome assessment, or if the treatment allocation is known, the consequences of procedural errors could be more serious, and on-site monitoring may be desirable.

In an unblinded comparison of two drugs, for example, it would be particularly important to ensure that sites record all study outcomes uniformly, irrespective of allocated treatment. In this case, site visits might help to assess whether

there is any potential for biased recording of outcomes.

Outstanding uncertainties requiring new research

In many large-scale trials, particularly those with blinded treatments and endpoints involving little risk of misdiagnosis, it may be possible to design procedures which allow central review of data to be the main or only form of monitoring. However, if data are to be reviewed frequently in order to identify unusual data patterns (see above), then consideration must be given to the training of trial statisticians in appropriate analytical methods. Since all of the methods involve repeated assessment of data patterns at many sites, these methods involve multiple looks at emerging data, and hence there is a need to control for type I errors (i.e., falsely concluding that a site has a problem when it does not). More empirical work is needed to develop a toolkit for statisticians wishing to employ central monitoring methods in their own trials.

Conclusions

Quality assurance in trials does not automatically equate to intensive on-site monitoring. In fact, on-site monitoring methods may be extremely inefficient for the identification of errors most likely to compromise patient safety or bias study results. A variety of monitoring strategies is available, and each new trial should conduct a risk assessment to identify the optimal means of monitoring with regard to the likely sources of error, their consequences for patients and the study's validity, and the available resources. Trial management committees should consider central statistical monitoring a key aspect of trial monitoring. The systematic application of this approach would be likely to lead to tangible benefits, and resources that are currently wasted on inefficient on-site monitoring could be diverted to increasing sample sizes or conducting more trials.

References

1. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *Br Med J* 1995; 311: 485.
2. Sterne JAC, Davey Smith G. Sifting the evidence—what's wrong with significance tests? *Br Med J* 2001; 322: 226–31.
3. The DAMOCLES Study Group. A proposed charter for clinical trial data monitoring committees: helping them do their job well. *Lancet* 2005; 365: 711–22.
4. Grant AM, Sydes M, McLeer S, et al. Issues in data monitoring and interim analysis of trials (the DAMOCLES study). *Health Technol Assess* 2005; 9: 1–238.
5. Buyse M, George SL, Evans S, et al. for the ISCB Subcommittee on Fraud. *Statist Med* 1999; 18: 3435–51.
6. Ellenberg SS, Fleming TR, DeMets DL. *Data Monitoring Committees in Clinical Trials. A Practical Perspective*, John Wiley, London, 2002.
7. MRC/DH joint project to codify good practice in publicly funded, UK clinical trials with medicines. Workstream 4: Trial Management and Monitoring (http://www.ct-toolkit.ac.uk/_db/_documents/trial_MP.pdf (accessed 4th April 2007)).
8. Al-Marzouki S, Evans S, Marshall T, et al. Are these data real? Statistical methods for detection of data fabrication in clinical trials. *Br Med J* 2005; 331: 267–70.
9. The ESPS2 Group. European Stroke Prevention Study 2. Efficacy and safety data. *J Neurol Sciences* 1997 (Suppl); 151: S1–S77.
10. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonized tripartite guidelines. Guideline for good clinical practice E6. <http://www.ich.org/LOB/media/MEDIA482.pdf> (accessed 26th April 2007).