Maintaining Trial Integrity for Randomized Open-Label Trials
A Statistical Programmer's Perspective

Wenyun Ji
Amgen, Inc., One Amgen Center Drive, Thousand Oaks, CA 91320

Abstract
Many oncology clinical trials are randomized open-label trials. In order to minimize operational bias while not hindering critical activities in order to ensure subject safety, data quality and protocol compliance, measures of trial integrity need to be put in place by restricting access to treatment information from certain personnel who conduct the clinical trials. This paper discusses three different ways that may be used by statistical programmers during the conduct of randomized open-label trials in order to maintain trial integrity. Statistical programmers must recognize the advantages and potential issues of each methodology and choose one that fits the purpose for their randomized open-label trials.
Key Words: oncology, clinical trials, statistical, programmer, CDISC, data analysis

1. Introduction

Many oncology trials are randomized trials comparing a novel treatment against the Standard of Care treatment (SOC). In practice, it is not always possible to mask the treatment assignment for reasons such as different dosing schedules between treatment arms, different intervention methods (drugs vs surgeries), etc. Therefore, the trials are often designed as open-label trials, meaning that the study investigators, site administrators and the study subjects are all aware of the treatment assignment. Operational personnel for the study such as clinical operation, data management and their corresponding CROs could be 'unblinded'[1] to individual treatment assignments during the conduct of the trial. The clinical database contains treatment assignments that are viewable to operational personnel.

For these randomized open-label trials, measures need to be taken by the sponsor to minimize operational bias and maintain trial integrity while not hindering critical activities in order to ensure subject safety, data quality and protocol compliance. The focus is usually on the control of access to and sharing of restricted data (see section 2 for definition) by trial sponsors and their service providers during the conduct of the study.

2. Restricted Data

2.1 Types of Restricted Data

In this paper, restricted data refer to the clinical trial data from a study of one or more individual subjects, or summarized by group or for the whole study, collected or created

[1] Throughout the paper, words ''blinded', 'unblinded', 'blindness', 'unblinding', and 'blinding' are used not to describe a blinded study, but to represent the situation of the knowledge of the restricted data.

prior to the interim and/or the final database lock where limited access must be maintained and controlled in order to reduce bias and prevent unauthorized release of study level information (e.g., treatment assignments, potentially unblinding data, and data that implicitly or explicitly reveal endpoint results, data summaries, and interim analysis results) .

Restricted data are classified into the following three categories:

- Restricted data elements such as treatment assignments, study drug administration, and potentially unblinding data such as adverse events and lab data;
- Data summaries and analyses by treatment assignments such as unblinded interim safety reports, unblinded interim efficacy results;
- Blinded data summaries and analyses such as pooled event rate modeling which could lead to speculation of treatment effect of an active arm.

2.2 Sources of Restricted Data

2.2.1 Central Randomization Information

The randomization code from the central randomization system is the source for treatment assignment information. However, in cases of open-label studies, the treatment assignments are often transferred to the clinical database on a regular basis.

2.2.2 Clinical Data Repository

A clinical data repository is a database that consolidates data from a variety of clinical sources to enable a unified view for individual subjects. It is an important tool for collecting and cleaning of clinical data as well as for preparing data for statistical analysis. If the treatment is unmasked, the subject's randomization code may reveal the treatment assignment or treatment effect to anyone with access to the database.

Examples of such databases include:

- Clinical database
- Safety database
- Central laboratory database
- SAS datasets (e.g., raw datasets, SDTM datasets, and ADaM datasets)

2.2.3 Interim Data Summaries and Study Results

Individuals or groups, internal or external, may conduct 'blinded' and 'unblinded' analyses to generate interim data summaries or study results before study 'unblinding'. Should the study continue as planned, knowledge of such interim data may affect recruitment of subjects, choice of treatment options, assessment of endpoints, and statistical analysis and ultimately jeopardize the study objectives. Therefore, access to interim data must be controlled.

Examples of such interim summaries and study results include:

- Unblinded interim efficacy analyses by Data Monitoring Committee (DMC)
- DMC safety reviews

2.2.4 Study Sites

The study investigator knows which treatment a study subject is receiving in an open-label study. The investigator may reveal such information when interacting with

individuals from the sponsor or its service providers involved in the conduct of a trial. There may be discussions on treatment options with sponsor's medical monitor, and discussions on data issues with Clinical Research Associates (CRAs).

3. Authorized Access to Restricted Data

For randomized open-label studies, the treatment assignment information could be automatically transferred to the clinical database as subjects enroll into a study. The fields on the Randomization electronic Case Report Form (eCRF) are usually pre-populated with the randomization information.

Measures must be taken to control or restrict access to the treatment information so that the conduct of study can continue and the study remains 'blinded'. There are different approaches to this issue.

One approach is to prohibit any 'unblinded' analysis using real treatment codes before database lock. Dummy treatment codes (often done by scrambling of data) can be used by statistical programmers to create Tables, Figures and Listings (TFLs).

Another approach is to separate programmers into two groups. One works on the data from the clinical trials repository where real treatment assignments are visible to the programmers. The other group (usually is just one person who is the co-lead for programming function) is blinded and goes to the study team meetings to participate in the conduct of the study.

The third approach is to restrict summary tables to such a minimum that no conclusion can be drawn from the TFLs.

The following sections will discuss pros and cons of these methods.

3.1 Scrambling Treatment Assignment

Scrambled treatment assignment can be done at the SDTM or ADaM level.

If scrambled at the SDTM level, no programmers are unblinded. Summary tables can be produced using dummy treatment. Listings are restricted so that the potential unblinding data won't be seen by any study team members. However, there are issues associated with the SDTM-level scrambling. When data issues are observed by programmers, they usually cannot tell if it's a real issue or due to the fake treatment assignment. Because it is an open-label study, the CRF designs usually have no desire to mask the treatment. Data points such as AE relations to 'active drug' and SOC become entries on the CRF. The dummy data may confuse the logic of the CRF flow.

IF scrambled at the ADaM level, the SDTM programmers are unblinded. They cannot attend any study team meetings and will need to rely on the communication from the blinded lead programmer for the study needs. There needs to be clear separation between SDTM and ADaM programmers. Some summary tables may not make sense because the CRF entries may apply to only one treatment group. Summarizing on these data could potentially unblind everyone on the team. At the planning stage, teams need to identify these tables to be held off until unblinding occurs.

Cost for scrambling treatment is usually very high because of the extra effort to maintain dummy treatment assignment and real treatment. During the TFL delivery period, time and effort usually have to be doubled in order to deliver several rounds of blinded packages and then one final unblinded package. Often new data issues arise when teams unblind the treatment for the first time. More time should be budgeted for this kind of approach.

3.2 Maintaining Two Groups of Programmers

Maintaining separate groups of blinded and unblinded programmers seems to work well if study teams could hold off the review of draft TFLs to a much later time. Statisticians need to be blinded in order to develop unbiased statistical analysis methodologies.

Blinded Team: One blinded programmer and one blinded statistician would attend the study team meetings and participate in the decision-making discussions for the conduct of the study. They will bring back analysis requests and any communications needed to help develop the TFLs. The blinded statistician creates the table shells. The blinded programmer creates the programming specification such as analysis data definition tables.

Unblinded Team: The unblinded team would carry out the analysis on the open-label study data using real randomization code.

The communication between blinded and unblinded teams should ensure the protection of treatment information. Summary tables over real treatment assignment should be limited to the extent that no study conclusion can be derived. Teams need to decide at the study start-up and planning stage which tables should be held-off until the database lock.

With this approach there are still issues. Without access to data, the blinded programmer may not be able to create a perfect data definition table for the unblinded programmers to follow. This requires certain level of communication between the two groups of programmers. People need to keep treatment information and other potential unblinding data confidential to the blinded team members.

The resource burden to this approach is relatively high because of the need of dual programming groups.

3.3 Restrict Summary Analysis that Uses Real Treatment Assignment

Often the dosing route and the administration schedules are distinctly different between treatment arms, so the case report forms of the investigational product are different between the two treatment groups. In addition, there is a risk that subjects randomized to the controlled arm will drop-out prior to the first dose of study drug once they learn they will not be receiving the investigational drug. In order to monitor dosing and protocol compliance, the clinical study team will need to have unrestricted access to the treatment assignment of individual subjects. There will be restrictions, prior to database lock for the final analysis, on data that can be summarized and pooled over treatment groups.

This is probably the most straight-forward way of maintaining a good level of trial integrity. The resource burden is the smallest among the three approaches.

3.4 Discussion

People who have access to data should not make decisions based on treatment assignments or study results. In some cases, people are even taken off the decision team to protect the 'blindness'. However, the sites and the subjects all have the knowledge of treatment assignments and often they are the ones who make decisions. For example, some patients may choose to leave a study when they know they are randomized to the control group. When choosing the programming method that best suited to maintaining trial integrity, programmers need to consult with their study statisticians and take into account these factors.

4. Conclusion

Programmers need to recognize the importance of maintaining trial integrity during the conduct of randomized open-label clinical trials. Programmers will need to decide at the start of the study which method works the best for maintaining trial integrity for their study. A Trial Integrity Document (TID) may be needed to provide specific guidance to the cross-functional study team on handling the restricted data for the randomized open-label clinical trials. In some cases, the guidance is provided at a product-level strategy document. Programmers need to work closely with the statisticians and other cross-functional team members to make sure the programming strategy is well documented and strictly followed.

References

None