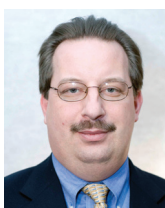


Improving Ratings Precision

Designing and implementing an optimal ratings surveillance system, explain Oto Markovic and Rusty Bealer of United BioSource Corporation, will help sponsors get higher quality data



As Regional Manager of United BioSource Corporation's Prague office, Dr Oto Markovic, MD, is responsible for European investigator services. In this capacity, he directs clinical and project delivery teams that design and implement UBC's global training programmes. Previously, Oto was the European Medical Communication Lead – EMEA – Global Medical Affairs for Bristol-Myers Squibb. Before that, he worked in clinical development at Eli Lilly on CNS, anti-infective and cardiovascular compounds and served as Deputy Chief of Psychiatry at Horní Berkovice in the Czech Republic. A board-certified psychiatrist at Charles University in Prague, Oto attended medical school at University Komeniana, Slovak Republic.



Rusty Bealer is Director for Information Technology of Investigator Services for UBC. With over 20 years' experience in this area, he leads the development of applications that track rater training, provide web-based/self-paced training and monitor rater performance throughout clinical trials. Heading a team of software developers and system administrators, he is also responsible for providing the technical infrastructure to support UBC's internal and external user base. Prior to joining UBC, Rusty held various IT leadership, consulting and development positions with CIGNA Healthcare, HealthAxis and Shared Medical Systems.

Clinical trials frequently rely on subjective clinician-rated outcomes as pivotal endpoints. The scoring of these assessments is often based on clinical subjectivity and individual interpretation, which can increase variability across sites and raters – potentially contributing to failed or inconclusive studies.

The use of electronic tools can assist in reinforcing the continuing reliability of data collected using clinician-rated outcomes. Real-time ratings surveillance systems are an important new tool for monitoring these data. Moreover, there is some indication that the FDA and EMEA are considering guidance to the industry addressing how clinical trial sponsors should incorporate subjective endpoints in their studies.

This article will examine how these systems can be designed and deployed in support of clinical trials that will rely on these subjective endpoints. Specifically, we will examine techniques for:

- ◆ Improving the competency of the clinician-rater
- ◆ Ensuring greater consistency of data collection
- ◆ Setting up surveillance systems for optimal results

IMPROVING THE COMPETENCY OF CLINICIAN-RATERS

Subjective endpoints are included in virtually all central nervous system (CNS) clinical trials, as well as in many trials in other disease areas such as oncology, pulmonology, gastroenterology, dermatology, urology and gynaecology.

As a result, it is important to improve the competency of the data collection 'vehicle' – in this case, the clinician-rater. Certainly, rater training and certification (RTC) is an essential component in developing the 'best raters' for a study, since it provides the basis for obtaining initial inter-rater reliability across raters and intra-rater reliability within raters (1). However, rater training prior to study initiation does not guarantee sustained ratings quality and reliability in-study.

Thus, while optimising inter-rater reliability at the outset of the study is vital, a method for determining that consistency within and between raters continues over the course of the trial is also essential. According to regulatory agency guidance, careful trial monitoring can ensure that potential problems are noticed early, thereby minimising their occurrence or recurrence. Clearly, the benefits of trial monitoring are leveraged exponentially when applied in real-time.

Moreover, the seamless integration of real-time surveillance with corresponding corrective interventions establishes a system for ensuring the acceptability and quality of data being accrued.

ENSURING GREATER CONSISTENCY OF DATA COLLECTION

While ongoing assessment of data quality is a complex topic, it is further complicated by the subjective nature of clinician-rated scales. Developing a general approach to optimising data quality across subjective scales is difficult. Not only are cost and benefit important considerations, but implementation



feasibility is also a relevant factor in defining an optimal and pragmatic data quality strategy to support ratings reliability, especially in large multi-national studies.

Accordingly, a data quality strategy must be constructed that is both easily and consistently applied across all raters without the risk of introducing bias or multiplicity concerns. The goal is to avoid affecting the robustness of the treatment effect and primary study conclusions. Ideally, it will also accommodate the nuances and specific design of a study and easily support multi-national programmes.

At the core of a surveillance system that is designed to improve ratings precision are a series of protocol-specific data edits used as a flagging device which are applied as near-real-time as possible to raters. The edits should be designed to highlight differences from expected thresholds of normative activity so as to detect potentially ‘deviant’ ratings or data.

Immediately following the identification of potentially erroneous ratings techniques, customised tutorials should be conducted that target rating behaviours associated with the ‘flagged’ poor technique. To eliminate execution bias, data edits should be developed and set during protocol formation and prior to initiating the surveillance programme – and then uniformly applied to all raters and all patient visits. The ratings surveillance methodology should include:

- ◆ A description of the procedures for flagging ‘aberrant’ ratings
- ◆ The timing of evaluation relative to the patient visit
- ◆ The procedure for handling aberrant ratings relative to programmed data edits

Staff who are responsible for implementing the surveillance system for a given trial are blinded to the treatment condition throughout the duration of the trial.

SETTING UP SURVEILLANCE SYSTEMS FOR OPTIMAL RESULTS

There are limits to what traditional site monitors can do – even though they play a variety of roles in multi-centre clinical trials. For instance, they ensure: adherence to GCPs and protocol design; study drug accountability; proper documentation and reporting of SAE/AEs; and maintenance of required regulatory documents. In addition, site monitors also review source documentation to confirm completeness and accuracy. However, with many factors affecting the role of the monitors in a clinical trial – more centres, global sites and resource pressures – it is very difficult for them to be able to evaluate data on a case-by-case basis in a timely fashion. These challenges, combined with the fact that clinician-rated scales are inherently subjective and open to clinical interpretation,

can limit a monitor’s role to tasks such as identifying missing and incomplete data.

Then there is the issue of time lag. In a traditional monitoring approach, site visits occur every four to eight weeks. Surveillance systems allow for near-real-time evaluation of ratings data and thus identify rater problems before it is too late.

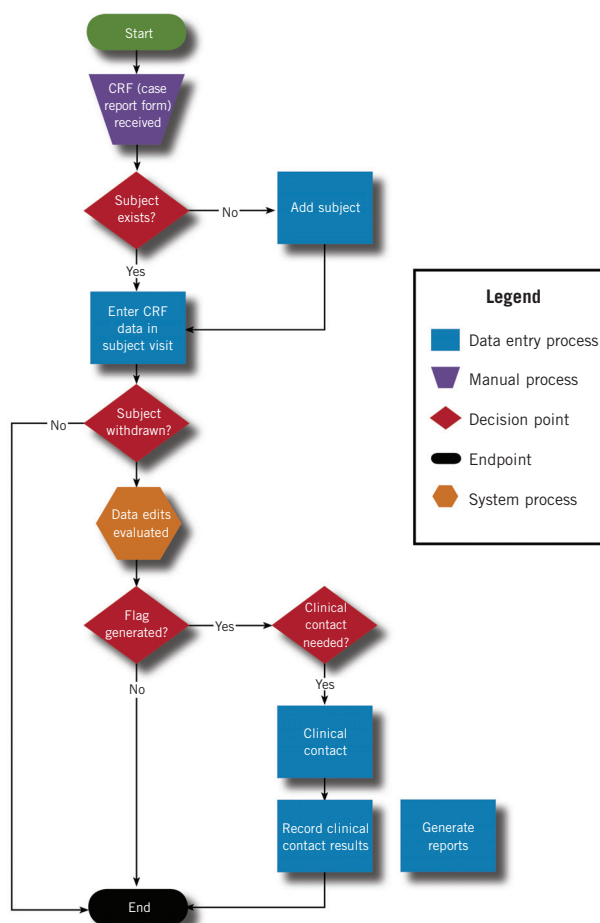
Step 1: Establish Your Requirements

There are a number of factors to consider when designing a ratings surveillance system, including:

- ◆ The number and type of clinician-rated scales used in a study
- ◆ The visit frequency of the scales being administered
- ◆ The clinical background of the raters participating in the study
- ◆ The length of the enrolment and treatment phases of a study
- ◆ The data collection processes

Ratings surveillance systems rely heavily on protocol-specific conditions to guide the design of data edits to flag potential ratings errors; using a protocol or disease expert to identify

Figure 1: Ratings surveillance programme



these 'flags' is a critical step in ensuring that a surveillance system is effective. Many depression studies, for instance, include two primary endpoints:

- ◆ A severity measure such as the Hamilton Depression Rating Scale (HAM-D), typically administered by a study rater
- ◆ A global assessment measure such as the Clinician Global Impression assessment (CGI), scored by a site clinician

In this example, a ratings surveillance system may include data edits that compare HAM-D patient ratings with CGI scores across visits in order to identify discrepancies between the two measures. So, if the patient appears to be improving based on HAM-D ratings, but deteriorating according to CGI scores, the rater would be flagged for evaluation by an appropriately trained clinician and for a possible customised tutorial directed at the flagged issue.

Step 2: Design the System

Surveillance systems are developed in accordance with the same processes and procedures utilised by other electronic systems in clinical trials. Identifying how data will be collected and managed in the clinical trial should be determined at the outset.

Surveillance systems can be designed to accommodate any data collection system. Today, such systems can include electronic data capture (EDC) using computer systems or hand-held devices, interactive voice response (IVR) using telephone-based systems, or paper forms transmitted by fax. Surveillance systems operate by 'listening in' on the data as they are transmitted from the site to the primary data management source.

For EDC and IVR systems, this can mean reports or data transfers following patient visits. For paper-based programmes, this can mean delivery by fax of paper case report form (CRF) pages to the surveillance provider. After the mechanism for observing the data is established, the business logic of the data surveillance must be programmed. This involves designing database queries that review the data as they are collected and identifying potential ratings errors. These are broadcast as potential data checks and routed to a dedicated clinician who evaluates the data flags when they are triggered. These reports can be issued as paper printouts or, more frequently, as an electronic notification of a flag.

Step 3: Implement and Execute

Dedicated, trained clinicians are responsible for evaluating the flags as they are triggered. By centralising the observation of these data, a smaller team of clinicians is able to review many more cases than a traditional monitor could handle. If a clinician reviews a flag and is concerned about how a scale is being utilised, the site is contacted and a discussion about the scale and the possible error takes place. In multinational studies, local-language clinician reviewers

are identified in advance and are utilised for outreach to sites on a case-by-case basis.

By monitoring patient enrolment, a surveillance system proactively tracks new subjects and ensures that clinical trial sites submit their data in a timely fashion. Immediate review of data is a significant advantage in a surveillance programme – it allows for the correction of errors that may have compounded for months between monitoring visits in a traditional programme.

CONCLUSION

In data presented at the 2007 Mid-Year Conference of the International Society for CNS Clinical Trials and Methodology, it was demonstrated that use of a ratings surveillance system resulted in a statistically significant decrease in ratings errors between the initial and final study visits (2). The surveillance system was carefully designed to apply protocol-specific clinical surveillance of ratings in near-real-time.

By utilising existing CRF data capture systems – and designing customised tools to view, process and flag the data – observing clinicians were able to identify raters exhibiting differences from expected thresholds of normative activity.

Clinicians were then able to provide tutorials to raters before the study subject returned for a consecutive visit, preventing an existing ratings problem from being compounded. These errors would probably not have been detected by traditional monitoring because most monitors simply do not have the advanced clinical background required to evaluate ratings and remediate raters.

Moreover, errors detected during traditional monitoring would likely have been discovered weeks or even months after the occurrence based on typical monitoring schedules. A ratings surveillance system is able to systematically locate and remedy problematic raters in real-time, while remaining both cost-effective and geographically scaleable, making it an invaluable methodology for improving ratings precision. ◆

*The authors can be contacted at
oto.markovic@unitedbiosource.com
and rusty.bealer@unitedbiosource.com*

References

1. Targum SD, Evaluating rater competency for CNS clinical trials, *J Clin Psychopharmacol* 26(3): pp308-310, June 2006
2. Busner *et al*, Quality assurance methodology for site based clinical trial ratings, 2007 Mid-Year Conference of the International Society for CNS Clinical Trials and Methodology, Brussels, Belgium, 2007