

# Supplement to APPLIED CLINICAL TRIALS



## Clinical Trial Data Integrity

### Using Internet Technology to Collect Reliable Data

Jules T. Mitchel, Joon You, Yong Joong Kim, Amy Lau,  
Linda Cheng, Seymour Fein, and Ronald Nardi

Electronic data capture can be invaluable to speed clinical development processes. A recent study demonstrates the reliability of data collected by this method.

The development of timesaving processes that safeguard data integrity continues to be a primary goal of pharmaceutical companies. Remote data entry allowed the industry to improve the integrity of its data collection in the 1980s. Additional timesaving pro-

cesses evolved during the 1990s via the advent of the Internet. Today, electronic methods allow high integrity data collection. In the current study, we were able to demonstrate that data entered on electronic case report forms accurately matched data entered on paper forms, thus providing qualification of data integrity.

#### History

During the 1980s, the personal computer changed the business world. One of the most important applications for pharmaceutical companies came in the form of remote data entry (RDE). By using RDE techniques, pharmaceutical companies could provide computers to clinical sites, allowing staff at the sites to collect data in paper source documents and then enter

the data into computer-generated data entry screens. The computerized data were usually returned to the sponsors on floppy disks via overnight delivery services. These tasks occurred repeatedly during clinical trials. Data lock occurred after completion of source document verification by the study monitors.

Following the development of the Internet, anyone who had a phone line and a browser could carry out data collection and transmission. Now it is no longer necessary to suffer the laborious processes of classical RDE. Instead, RDE has evolved into a process we describe as Internet-based remote data entry (I-RDE), more commonly called electronic data capture (EDC). Although the technology is available, many companies have not yet made the changeover.

## The benefits of I-RDE

I-RDE use can be invaluable for companies seeking to expedite their clinical development processes. Potential benefits include

- more rapid access to trial data and trial progress.
- more efficient monitoring.
- higher-quality data entry due to edit and logic checks at the time of data entry.
- elimination of double-key entry and related tasks.
- decreased number of queries.
- elimination of commercial paper case report forms (CRF).
- instantaneous alerts to safety and enrollment issues at the time of data entry.
- faster time to data lock.

To date, there have been few published papers supporting the attributes of I-RDE. One previous study showed that when compared to paper CRF systems, the query rate in I-RDE clinical trials is reduced by approximately two-thirds.<sup>1</sup> Another study evaluated the potential cost savings in a Phase 1 study in bone marrow transplant patients.<sup>2</sup>

This study addresses the impact of I-RDE on data integrity.

## Entering the data

As part of quality assurance to demonstrate that data entered via the Internet using electronic case report forms were accurately captured in Statistical Analysis System (SAS) datasets, data entry clerks entered data from the paper CRFs, via the

Internet, into a database using Web-based data entry screens. This primary database was to be used for the final data analyses. The data entry screens were equivalent to paper CRFs in both format and content.

**A second database.** To evaluate the number of errors solely due to data transcription, all data were entered a second time into a second, but identical, database by a second group of data entry clerks. At the time of the second data entry, the entered values were immediately compared, electronically, to the data already in the primary database. If the double-key entered value matched the initial value, the double-key entered value was accepted and an electronic message was generated that the data compared correctly. If, however, the computers found a discrepancy between the content of the primary database and the secondary database, data entry was stopped and the data entry clerk made a decision, based on the information present in the paper CRF, as to whether the initial or second data entry value was correct.

**Comparing entries.** After confirming that the initial data entry value was correct, the second data entry value was corrected. If, however, the second data entry was confirmed correct, the primary data entry value was corrected and an automatic audit trail of the change was documented. Then the CRF page was printed, showing either the confirmation of agreement between the two databases, or an audit trail of the correction. After changing the original database, a revised data listing was generated in SAS, and a data entry

clerk verified that each corrected value had been incorporated into the primary database.

After correcting all data entry discrepancies as a result of the double-key data entry exercise, we generated a new set of SAS data listings. Then we manually compared a select group of demographic, efficacy, and safety CRF entries to the data printed in the SAS data listings.

## Evaluating the errors

The clinical trial used in this study involved 124 subjects, at 15 clinical research sites, in the area of reproductive endocrinology. The paper CRF consisted of 120 pages with an average of 15 variables per CRF. As shown in Table 1, our analysis involved approximately 15,000 CRF pages with approximately 230,000 variables.

Among the 229,152 variables entered into the database, we observed 950 errors, reflecting an overall error rate of 0.41% in the initial and double-key data entry. The primary errors occurred in text fields, where either the field could not be properly read or unique symbols or abbreviations were used. Table 2 shows that data entry for medications, medical history, and adverse events had the highest numbers of errors. Out of 23,628 variables that were verified against SAS data listings, only 13 (0.06%) corrections needed to be made to the final data listings. The results of an analysis of these 13 errors can be found in Table 3. None of these errors would have had any impact on the final data analysis,

**Table 1** Double-key entry and data listing for error rates

| Number of subjects | Number of CRF pages | Number of variables entered via double-key entry | Number of variables analyzed against SAS data listings |
|--------------------|---------------------|--------------------------------------------------|--------------------------------------------------------|
| 124                | 14,913              | 229,152                                          | 23,628                                                 |

**Table 2** The most common errors during double-key data entry

| Number of errors | Adverse events | Vital signs | Medications | Medical history | Physical exam |
|------------------|----------------|-------------|-------------|-----------------|---------------|
| 950 (0.41%)      | 5.7%           | 1.1%        | 17.7%       | 12.2%           | 2.1%          |

**Table 3** Error analysis between paper CRF and SAS data listings

| Event                                                 | n          |
|-------------------------------------------------------|------------|
| Missing word                                          | 1          |
| Spelling                                              | 3          |
| Query not correctly modified                          | 2          |
| Initial and double-key entry with same error          | 3          |
| Data not entered                                      | 4          |
| Total data listing variables requiring correction (%) | 13 (0.06%) |

had they not been detected.

### Making the case

One of I-RDE's main advantages is that the people who know the data the best, the clinical study site personnel, enter the data directly into the data entry screens using paper source documents. They need not question the investigator's intent, there are no illegible fields, and nobody needs to interpret unique abbreviations or symbols, because the staff has worked together throughout the study. In addition, all missing data and out-of-range data are flagged as requiring explanations at the time of data entry. Data discrepancies can be confirmed either before or during the monitoring and/or the query processes.

**Initial qualification study.** For the present study, because I-RDE would change the way this sponsor conducted its clinical research and data management procedures, we performed an initial qualification study to evaluate whether data collected via I-RDE would be accurately captured into SAS datasets. To accomplish the goal, we entered data via I-RDE from a study where personnel at the clinical research site also entered data initially into traditional paper CRFs. All paper CRF data entries were also verified by performing source-document review.

By doing this, in the event I-RDE did not meet expectations, the sponsor could easily have recovered the efficacy and safety data from the paper CRFs without returning to the source documents. Comparison to SAS data listing showed that the data entered via I-RDE did reflect accurately the data derived from paper CRFs.

**FDA guidances.** In 1997, the Food and Drug Administration issued a regulation within 21 CFR 11 addressing the use of electronic signatures and electronic records.<sup>3</sup> In 1999, the FDA issued another guidance to address issues pertaining to

computerized systems used to create, modify, maintain, archive, retrieve, or transmit clinical data intended for submission to the FDA.<sup>4</sup> Although the primary focus of this guidance was on computerized systems used at clinical sites to collect data, the principles in the guidance are also appropriate for computerized systems used at contract research organizations, data management centers, and sponsors. The goal of the guidance was to assure that individuals using the data from computerized systems would have confidence in the integrity of their data, and that data entered electronically were no less reliable than data collected in paper form.

To assist companies in addressing FDA requirements and expectations, the FDA has recently issued two additional guidances, one regarding electronic records and signatures<sup>5</sup> and another regarding software validation requirements.<sup>6</sup> In order to comply with 21 CFR 11, all changes to the database are to be automatically tracked including an electronic stamp of who changed the data, the date and time of the change, and the reason for the change, if required. If electronic signatures are used, the standard operating procedure for signoff must conform to the requirements of 21 CFR 11.

### To err is human

The current study demonstrates that overall data capture using Internet-based remote data entry has a very low error rate. The errors observed are related to human activities—not the underlying technology.

### References

1. Jules Mitchel, Joon You, Amy Lau, and Yong Joong Kim, "Paper versus Web: A Tale of Three Trials," *Applied Clinical Trials*, August 2000, 34–35.
2. Jules Mitchel and Otto Mills, "Using an Internet Tool to Maximize the Quality and Timeliness of Pharmacokinetic and Pharma-

dynamic Data Retrieval in Patients Undergoing Bone Marrow Transplantation," *Clinical Pharmacology and Therapeutics*, 71(2) 88 (2002).

3. Department of Health and Human Services, "21 CFR Part 11, Electronic Records; Electronic Signatures Final Rule," *Federal Register* 62 (54), 13429 (20 March 1997).
4. Food and Drug Administration, *Guidance for Industry—Computerized Systems Used in Clinical Trials* (FDA, Rockville, MD, April 1999).
5. Food and Drug Administration, *Guidance for Industry—21 CFR Part 11; Electronic Records; Electronic Signatures Validation* (FDA, Rockville, MD, August/September 2001).
6. Food and Drug Administration, *General Principles of Software Validation; Final Guidance for Industry and FDA Staff* (FDA, Rockville, MD, 11 January 2002).

**Jules T. Mitchel**, \* MBA, PhD, is president of Target Health Inc., 261 Madison Avenue, 24th Floor, New York, NY 10016, (212) 681-2100, fax (212) 681-2105, email: [julesmitchel@targethealth.com](mailto:julesmitchel@targethealth.com).

**Joon You**, BS, is the company's chief technology officer; **Yong Joong Kim**, MS, is director, Application Development; and **Amy Lau**, MPH, is director of Biostatistics.

**Linda Cheng**, MS, is the president of Markus Research, Inc., 4 Dara Court, West Nyack, NY 10994. **Seymour Fein**, MD, is the president of Fein Associates, Inc., 476 Canoe Hill Road, New Canaan, CT 06840. **Ronald Nardi**, PhD, is the executive vice president, Research and Development, of Ferring Pharmaceuticals, Inc., 120 White Plains Road, Tarrytown, NY 10591. Their study used Target Health's Target E\*CRF as its I-RDE system.

\*To whom correspondence should be addressed.