

# Evaluation of Data Entry Errors and Data Changes to an Electronic Data Capture Clinical Trial Database

**Jules T. Mitchel, MBA, PhD**  
President, Target Health Inc,  
New York

**Yong Joong Kim, MS**  
Senior Director of  
Application Development  
and Data Management,  
Target Health Inc, New York

**Joonhyuk Choi, BS**  
Director of Application  
Development, Target Health  
Inc, New York

**Glen Park, PharmD**  
Senior Director Clinical and  
Regulatory Affairs, Target  
Health Inc, New York

**Silvana Cappi, MSc, MBA**  
Executive Director,  
Biometrics, Ferring  
International PharmaScience  
Centre, Copenhagen,  
Denmark

**David Horn, MA**  
Energy Policy Analyst,  
Energetics Incorporated,  
Washington, DC

**Morgan Kist**  
Haverford College,  
Philadelphia, Pennsylvania

**Ralph B. D'Agostino Jr., PhD**  
Professor, Public Health  
Sciences; Director,  
Biostatistics Core at the  
Comprehensive Cancer  
Center, Wake Forest  
University School of  
Medicine, Winston-Salem,  
North Carolina

**Key Words**  
EDC; Data management;  
Risk-based monitoring

**Correspondence Address**  
Jules T. Mitchel, Target  
Health Inc, 261 Madison  
Avenue, 24th Floor,  
New York, NY 10016  
(email: julesmitchel@  
targethealth.com).

Monitoring of clinical trials includes several disciplines, stakeholders, and skill sets. The aim of the present study was to identify database changes and data entry errors to an electronic data capture (EDC) clinical trial database, and to assess the impact of the changes. To accomplish the aim, Target e\*CRF was used as the EDC tool for a multinational, dose-finding, multicenter, double-blind, randomized, parallel, placebo-controlled trial to investigate efficacy and safety of a new treatment in men with lower urinary tract symptoms associated with benign prostatic hyperplasia. The main errors observed were simple transcription errors from

the paper source documents to the EDC database. This observation was to be expected, since every transaction has an inherent error rate. What and how to monitor must be assessed within the risk-based monitoring section of the comprehensive data monitoring plan. With the advent of direct data entry, and the elimination of the requirement to transcribe from a paper source record to an EDC system, error rates should go down dramatically. In addition, protocol violations and data outside the normal range can be identified at the time of data entry and not days, weeks, and months after the fact.

## INTRODUCTION

Monitoring of clinical trials includes several disciplines, stakeholders, and skill sets. The aim of the study was to identify database changes and data entry errors to an electronic data capture (EDC) clinical trial database and to assess the impact of the changes.

Williams (1) clearly identified the various monitoring processes including statistical methodologies, interim data analysis, site monitoring for quality control and quality assurance, and safety reporting to regulatory agencies. Williams also stated that through careful attention to the design and conduct of a clinical trial, the expense of monitoring can be markedly reduced and that the determination of the extent and nature of monitoring should be based on considerations such as the objective, design, and complexity of the trial. Williams added that statistical sampling may be an acceptable method for selecting the data to be verified, and that study data should be monitored on an ongoing basis to ensure patient safety. Khosla et al. (2) recommended the identification of critical and noncritical data and to focus the source docu-

ment verification (SDV) process on critical variables. Critical variables are the data that are vital to the interpretation of safety and efficacy data, and which must be accurate. Eisenstein et al. (3) suggested that noncritical data collection should be kept to a minimum (ie, monitored on a reduced basis, perhaps 25% instead of 100% as performed for critical data), and that this measure saves time, both when the data are initially collected (patient visit) and when the monitor verifies it.

Califf et al. (4) suggested that sampling methods for source documentation can be used to eliminate costs incurred by reviewing every record. Califf added that these measures, coupled with prospective clinical judgment about areas of concern in the conduct of trials, can reduce costs without sacrificing quality. Eisenstein et al. (5) suggested that efficiencies could be realized when study sponsors switched to a centralized monitoring approach, and that with the advent of EDC, central monitoring can retrieve source data in real time and then use predetermined statistical models to monitor the data. Helms (6) suggested that EDC could be used to

perform independent double initial EDC. Helms suggested that the interview questions could be asked twice by different interviewers, duplicate measurements could be made independently, and samples could be split (or two samples taken) and analyzed by different instruments or by the same instrument on different days.

A recent article by Mitchel et al. (7) discussed the need for, and feasibility of, the integration of the electronic medical record (EMR) and EDC. The article addressed the paradigm shift once data are entered directly into electronic medical systems without the use of traditional paper source documents.

According to FDA (8), the extent of documentation necessary to support clinical effectiveness from a scientific perspective depends on the particular study, the types of data involved, and the other evidence available to support the claim. The FDA concluded that it is able to accept different levels of documentation of data quality, as long as the adequacy of the scientific evidence can be ensured. An FDA concept paper (9) suggested looking for alternative models for site monitoring. FDA acknowledged that for commercial studies to develop pharmaceutical products, on-site monitoring of performance has become the rule, with sponsors visiting essentially all sites every 4 to 8 weeks to ensure performance. FDA stated that while this contributes to aspects of study quality, it is not really practical for large outcome trials. FDA added that as interest in large trials grows, it is becoming apparent that the industry model is not feasible and that sponsors should consider implementing risk-based approaches to data collection, analysis, and monitoring.

ICH E9 (10) suggested that the form and content of the information collected should focus on the data necessary to implement the planned analysis. The Good Clinical Data Management Practices (11) guidance suggested focusing the data quality checks on the variables that are critical to the analysis. The guidance adds that it is not practical to design a quality check for every possible error and that quality checks

performed as part of data processing, such as data validation checks or edit checks, should target fields critical to the analysis where errors are expected to be frequent, and where it is reasonable to expect a high percentage of error resolution.

## METHODS

Target e\*CRF was used as the EDC tool for a multinational, dose-finding, multicenter, double-blind, randomized, parallel, placebo-controlled trial to investigate the efficacy and safety of a new treatment in men with lower urinary tract symptoms associated with benign prostatic hyperplasia. Data were collected from 566 subjects who signed informed consent and from 492 of these same subjects who were subsequently randomized. The study was performed under good clinical practices (GCP) and was compliant with local regulations. A decision was made to cut off the analysis for this article after at least 40,000 forms were entered.

Electronic case report form (eCRF) pages were monitored both prior to on-site monitoring and at the time of SDV. Queries were generated based on edit checks that were transmitted at the time of data entry and online edit checks run in a batch mode within the EDC system. Each data element change was subject to an electronic audit trail and for each modification, a reason for change was required. Reasons included additional information, entry error, and other. Changes could include one change or multiple changes per page as well as multiple changes on the same page. Site coordinators were trained in the use of the functionalities of the EDC system and were provided CRF completion guidelines.

To evaluate the impact of data changes on the analysis, an assessment was made of 331 data changes of five numeric variables contained within 1,287 uroflowmetry forms. A data change could include one or more variables. Descriptive statistics were calculated to assess the differences between the original data entered and the final clean data. In addition, two sample paired *t* tests were performed to compare the

**Changes to eCRF Pages and Reasons for Change**

TABLE 1

Total Pages Entered	Total Number of Changes	Data Entry Errors	Additional Information	Other Reasons
41,568	2,584 (6.2%)	1,836 (71.1%)	486 (18.8%)	262 (10.1%)

means between the comparison groups for each of the five variables.

**RESULTS**

**CHANGES TO THE DATABASE**

There were a total of 2,584 (6.2%) changes to 41,568 eCRF pages that were entered into the EDC database (Table 1). Of the 2,584 changes, 1,836 (71.1%) were designated due to “data entry errors,” 486 (18.8%) due to “additional information,” and 262 (10.1%) due to “other reasons.”

Table 2 shows that of the 2,584 form changes, 2,192 (85%) occurred in 10 forms; 72% occurred in six forms; and, remarkably, 47% in just three forms. More specifically, 20.8% of the changes occurred in the micturition diary log, 14.0% in the medication form, and 12.6% in the medical history form.

For most pages, the majority of the reasons for change was entry error. Many of these entry errors were due to transcription errors from the paper source document to the EDC system or lack of sufficient training on how to fill out a form. However, this was not consistent between CRFs (Table 3). For example, for adverse events, not unexpectedly, additional information was the main reason for change.

To better understand the nature of the changes to the database, rather than summarizing the detailed results from all pages, several pages were selected to provide more granularity for the reasons for change as well as examples of the specific changes.

**ELECTROCARDIOGRAM**

Most changes in the ECG forms were to the numerical variables of ventricular rate, PR interval, QRS duration, and QTcF interval. There were

**Summary of eCRF Pages With Data Entry Changes of Over 2%**

TABLE 2

Form	Total Page Changes (%)	Cumulative Percentage
Micturition diary log	538 (20.8)	84.8
Medications	361 (14.0)	64.0
Medical history	325 (12.6)	50.1
EQ5D questionnaire	229 (8.9)	37.5
Micturition diary question	211 (8.2)	28.6
Uroflowmetry	184 (7.1)	20.4
ECG	115 (4.5)	13.3
Adverse events	96 (3.7)	8.9
Physical examination	78 (3.0)	5.2
Vital signs	55 (2.1)	2.1
Total	2,192 (84.8)	

**Summary of Reasons for Change for eCRF Pages With Over 2% Changes**

TABLE 3

	Additional Information (%)	Entry Error (%)	Other (%)
Vital signs	3.6	94.5	1.8
Physical examination	16.7	80.8	2.6
Adverse events	62.5	33.3	4.2
ECG	32.2	55.7	12.2
Uroflowmetry	1.6	93.5	4.9
Micturition diary	29.9	67.3	2.8
EQ5D questionnaire/VAS score	24.0	59.8	16.2
Medical history	21.2	62.8	16.0
Medication	33.2	50.7	16.1
Micturition diary log	1.9	91.1	7.1

TABLE 4

Details of Reasons for Database Changes for ECG					
	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	115		32.2	55.7	12.2
Change in evaluation	20	Choice	25.0	75.0	0.0
Change in ECG results	72	Text	20.8	62.5	16.7
Changes in both results and evaluation	20	Choice/text	80.0	20.0	0.0
Other <sup>a</sup>	3		33.3	0.0	66.7

<sup>a</sup>For completeness, miscellaneous changes were combined into the "Other" category.

virtually no changes to the yes/no fields (Table 4). The QTcF interval in particular was changed in a large number of cases. Below are some examples of changes in the "abnormal specify" variable of the Overall Evaluation field:

1. "Sinus bradycardia" changed to "sinus bradycardia right bundle branch block inferior infarct" (additional information).
2. Ventricular rate, QRS duration, and QTcF interval changed from abnormal to normal and "specify" changed from "incomplete right bundle branch block" to blank (entry error).
3. Sinusal arrhythmia NCS was changed from normal to abnormal and added to "specify" (entry error).

#### UROFLOWMETRY

In this form with multiple numerical variables, most changes were to a single value (Table 5). However, when changes were to several or all values, they were categorized as multiple. Most of the changes were transcription errors. Some examples of changes to values:

1. Average flow rate
  - a. 9 to 4.8
  - b. 8 to 4
2. Voiding time
  - a. 25.2 to 25.3
3. Flow time including time to maximum flow
  - a. 16 to 20
  - b. 6 to 76
  - c. 39.3 to 57.3
  - d. 87.7 to 126.9

#### ADVERSE EVENTS

Examples of typical changes in the adverse event form are summarized in Table 6.

1. Adverse event description changed from mastitis to gynecomastia (additional information).
2. Serious changed from yes to no (entry error).
3. Intensity changed from mild to moderate (entry error).
4. "Other action specify" changed from blank to "Cipro" (entry error).
5. "Causal relationship to study drug" changed from unrelated to possible (additional information).

Examples of changes to the adverse event description field that tended to be refinements of what had already been entered are as follows:

1. Adverse event description changed from "soreness at injection site" to "soreness at injection site, right upper quadrant" (additional information).
2. Adverse event description changed from "tenderness at injection site" to "tenderness at injection site, upper left quadrant" (additional information).
3. Adverse event description changed from "depression—headache" to "depression (other)."

#### VAS SCORE

The VAS Health Scale was part of the EQ5D Questionnaire data entry form. With only six exceptions, all changes to this form were made to the Health Scale field (Table 7). The category Sponsor Request or Error includes changes that

Details of Reasons for Database Changes for Uroflowmetry

TABLE 5

	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	184		1.6	94.5	4.9
Added decimal place or changed value <1.00	49	Number	0.0	100.0	0.0
Change in single value	95	Number	1.1	93.7	5.3
Other <sup>a</sup>	2	Number	50.0	0.0	50.0
Multiple <sup>b</sup>	38	Number	2.6	89.5	7.9

<sup>a</sup>For completeness, miscellaneous changes were combined into the "Other" category.  
<sup>b</sup>Changes to multiple fields.

Details of Reasons for Database Changes for Adverse Events

TABLE 6

	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	96		62.5	33.3	4.2
Date	9	Date	0.0	100.0	0.0
Outcome	47	Choice	89.4	6.4	4.3
Adverse event	22	Text	36.4	54.5	9.1
Other <sup>a</sup>	9		33.3	66.7	0.0
Multiple <sup>b</sup>	9		77.8	22.2	0.0

<sup>a</sup>For completeness, miscellaneous changes were combined into the "Other" category.  
<sup>b</sup>Changes to multiple fields.

Details of Reasons for Database Changes for the EQ5D Questionnaire

TABLE 7

	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	229		24.0	59.8	16.2
Change in health scale	223	Number	23.8	59.6	16.6
Change in patient answers	2	Choice	0.0	100.0	0.0
No change evident	3	Choice	66.7	33.3	0.0
Multiple	1		0.0	100.0	0.0

were coded 19 different ways, all referring to a request from the sponsor. For this form, the data entry errors were caused when the sites photocopied the VAS form, which resulted in a physical shrinkage of the form.

#### MEDICAL HISTORY

Most of the changes to the Ongoing field were from yes to blank (Table 8). The following are some examples of changes to diagnosis/symptoms:

TABLE 8

Details of Reasons for Changes for Medical History					
	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	325		21.2	62.8	16.0
Date	79	Date	34.2	59.5	6.3
Ongoing	129	Choice	10.9	58.9	30.2
Medication	27	Choice	22.2	74.1	3.7
Diagnosis/symptoms	47	Text	27.7	72.3	0.0
Other <sup>a</sup>	11		9.1	45.5	45.5
Multiple <sup>b</sup>	32		25.0	68.8	6.3

<sup>a</sup>For completeness, miscellaneous changes were combined into the "Other" category.  
<sup>b</sup>Changes to multiple fields.

1. Prostatitis changed to acute prostatitis (additional information).
2. "Heart bypass with valve" to "coronary artery bypass with valve repair"; "medication to treat diagnosis/symptoms" from yes to no; "coronary artery bypass with valve repair due to coronary artery disease" (additional information).
3. ESWL to left ESWL (additional information).
4. "Skin cancer removed, nose and forehead" to "basal cell carcinoma removed, bilateral hands, nose, and forehead" (entry error).
5. CABG changed to "coronary artery bypass graft surgery."

### MEDICATIONS

Reasons for page changes are shown in Table 9. When reasons for change were entered, instead of selecting "additional information" or "entry error," a large number of changes to medications were entered under "other" as some variant of "sponsor request." All of these changes in the "other" column involved deselecting the "ongoing at end of study" field and leaving it blank.

Many of the changes were related to how doses were described and are illustrated below:

1. Trade name changed from Timolol to Timolol 0.5% (entered as sponsor request).
2. Dose per administration changed from 1,000 to 81 (mg in unit; fish oil in trade name) (additional information).

3. Dose per administration changed from 2 to 600/800; unit changed from tablets to mg (additional information).
4. Dose per administration changed from 75 to .075 (entry error).

Some examples of changes in the Indication field:

1. "Runny nose" to "AE upper respiratory infection" (additional information).
2. "Hoarseness" to "GERD" (entry error).
3. "Right shoulder pain" to "right shoulder pain and arthritis" (entry error).
4. "Arthritis" to "arthritis and herpes zoster" (additional information).
5. "Arthritic pain management" to "arthritis" (additional information).

### DIARY FIELDS

While this was a relatively simple page, and the page itself was changed only 3.5% of the time, because the page was entered over 15,000 times, it represented 20.3% of all database changes (Table 10). Many of the changes took place at the end or beginning of the month and around the New Year. Some changes also appeared to be transcription errors that occurred during monitoring.

Other changes involving various combinations of the urgency, leakage, and sleep disruption fields were also common. Changes in volume varied—some were small (174 to 175) and

TABLE 9

Details of Reasons for Changes for Medications					
	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	361		33.2	50.7	16.1
Date	64	Date	64.1	34.4	1.6
Ongoing	126	Choice	16.7	50.0	33.3
Indication	31	Text	54.8	45.2	0.0
Trade name	25	Text	60.0	40.0	0.0
Dose	22	Number	27.3	50.0	22.7
Frequency	7	Choice	57.1	42.9	0.0
Other <sup>a</sup>	9		0.0	22.2	77.8
Multiple <sup>b</sup>	77		20.8	75.3	3.9

<sup>a</sup>For completeness, miscellaneous changes were combined into the "Other" category.  
<sup>b</sup>Multiple reasons for changes of a single page.

TABLE 10

Summary of Reasons for Changes for Micturition Diary Log					
	Total	Field Type	Additional Information (%)	Entry Error (%)	Other (%)
All	538		1.9	91.1	7.1
Date/time	336	Date/Time	0.9	92.0	7.1
Volume	15	Number	6.7	86.7	6.7
Urgency/leakage/sleep disruption	75	Number	6.7	82.7	10.7
Other <sup>a</sup>	16		0.0	100.0	0.0
Multiple reasons <sup>b</sup>	96		1.0	93.8	5.2

<sup>a</sup>For completeness, miscellaneous changes were combined into the "Other" category.  
<sup>b</sup>Changes to multiple fields.

others were large (150 to 450, 240 to 40, 355 to 118).

### IMPACT OF DATA CHANGES

To assess the impact on data interpretation due to changes to the database, an analysis was made of the means and standard deviations (SD), before and after data cleaning, of five variables in the uroflowmetry form (Table 11).

When each of these variables was compared using paired *t* tests, only voiding volume indicated a statistically significant change ( $P =$

0.049) but not a clinically meaningful difference of less than 1 unit.

When examining the data from the impact of data changes, one can see that the mean values of each of the five measurements were nearly identical for the initial and final values. Therefore, in this illustration, the clinical impact on estimating the overall level of the outcome is minimal. In addition, in all cases, the estimate of the SD was smaller in the final cleaned data than in the initial data entry. This points to the area sponsors of clinical research may need to

TABLE 11

Initial and Final Values of Five Uroflowmetry Variables Before and After Data Cleaning		
	Initial	Final
<b>Maximum flow</b>		
Mean	10.81	10.77
SD	4.71	4.55
Range	(0–56.0)	
<b>Average flow</b>		
Mean	5.56	5.54
SD	2.21	2.16
Range	(0–20.6)	
<b>Flow time</b>		
Mean	38.16	38.21
SD	22.59	22.05
Range	(0–246.0)	
<b>Voiding time</b>		
Mean	46.07	46.30
SD	25.39	25.06
Range	0–344.0	
<b>Voiding volume</b>		
Mean	216.96	217.56
SD	97.12	96.70
Range	(0–812.8)	

consider most when determining the impact that errors may have on results when there is a reduced level of SDV in clinical trials where data are transcribed from paper source documents to EDC systems. However, these error rates should be lower when data are entered directly into EDC systems, using direct data entry, since there will be no transaction errors and a reconciliation of all out-of-range values will occur at the time of data entry.

The fact that the variability is slightly larger in the initial data suggest that less precise estimates of treatment effects will be detected than would have been detected had the correct variability estimates (from the clean data) been

used. To illustrate this, we examine the variable “voiding time,” which had an initial standard deviation estimate of 22.59 and a final clean estimate of 22.05. This amounts to approximately a 3% reduction in the size of the SD.

For most standard two-arm randomized placebo-controlled clinical trials, the primary efficacy comparison is a two-sample *t* test comparing the treated and control groups at a fixed point in time. For power and sample size calculations in such settings, the protocol needs to prespecify the alpha level (normally 0.05 for a two-sided test), the power (normally 80%), and the expected level of variability for the outcome of interest. Based on these inputs, a detectable difference can be described in terms of an effect size, where effect size represents the standardized difference in the mean value between two groups.

For instance, for a study with 100 participants each in two groups ( $n = 200$  total), with  $\alpha = 0.05$  (two-sided) and 80% power, the detectable effect size is 0.398. This means that 39.8% of an SD difference can be detected between groups for any outcome measure of interest. Using the “voiding time” data from above, the impact of the errors is such that in the initial data a difference of  $22.59 \times 0.398 = 8.99$  could be detected, while in the final clean data a difference of  $22.05 \times 0.398 = 8.78$  could be detected.

The impact on detectable differences, however, is a direct function of the sample size, so in the same example if there were 1,000 participants each in two groups (rather than 100), the detectable effect size is 0.125, and therefore the detectable differences in the initial and final cleaned data would be 2.82 and 2.76, respectively. Here the impact on the final inference would be much less than when the sample size is smaller.

What is important to recognize in all of these calculations is that the relative impact of the errors on the final inference is the same, namely that there is about a 3% reduction in the size of the detectable difference when the correct data are used when compared to the original data.

The absolute impact of a 3% reduction is dependent on the overall sample size.

In both examples with different sample sizes, it is clear that when the change in SD is small the impact on the detectable difference is also small. Therefore, an important diagnostic to perform when examining data for the impact of errors is to conservatively estimate how much the variability in the outcome measures is increased in the presence of the errors when compared to the clean data. If this variability difference is in the range of 5% or less, then the effect size estimates will also be impacted by 5% or less. Therefore, when performing power/sample size calculations, researchers may wish to consider designing their studies to have power to detect differences slightly smaller than they would have originally planned (eg, 5% or smaller) so that they can have the flexibility to be able to detect meaningful differences even if the variability is increased by 5% during the trial.

## DISCUSSION

The main errors observed in the present study were simple transcription errors from the paper source documents to the EDC database. This observation was to be expected, since every transaction has an inherent error rate.

Clearly, early training on how to complete the date fields in the micturition diary log would have made an impact on the frequency of data changes. Once these types of errors and their magnitude are identified by the clinical research associates and by data management, the clinical sites could easily have been retrained. Changes to the text fields for medications and adverse events could easily be monitored remotely and rules for data entry established. For example, for medications, based on the observations from a simple database summary table, the sites could be instructed just to enter the drug name and not to enter units. For adverse events, the sites could be instructed not to enter any abbreviations or just to indicate the adverse event as “injection site reaction.” These types of data errors must be assessed early in

the clinical trial as they could make the results from the form invalid.

There is no question that while monitoring a clinical trial under GCP, the trial sponsor must ensure that (a) clinical research subjects are fully informed about the risks and benefits in participating in a research study, (b) the clinical research sites fully understand and are following the protocol, and (c) the drug supply is being maintained and distributed properly. However, when using EDC, monitoring clinical sites every 4–8 weeks just to perform SDV appears to be inefficient and wasteful unless there is proactive coordination with clinical research, data management, and biostatistics.

With the advent of advanced data management tools, statistical monitoring, and direct data entry, there is now more than ever the need for a comprehensive data monitoring plan (CDMoP) based on acceptable error rates and empirical or risk-based approaches to clinical trial data monitoring. This protocol-based document is an expansion of the traditional monitoring plans as it integrates both on-site clinical monitoring and data monitoring by data management.

The stakeholders of the CDMoP are clinical research, data management, biostatistics, regulatory affairs, and quality assurance. The CDMoP needs to address the methodology and rationale for (a) ongoing site/staff protocol training, (b) schedule and purpose of monitoring visits to the clinical sites, (c) early and ongoing online review of eCRFs, and (d) review of trends of online and batch edit check hits. Aspects of monitoring can be performed remotely by observing and then resolving systemic errors such as how to describe an injection site reaction. By reviewing the trends of online and batch edit check hits early in the clinical trial, field monitoring and data entry errors could be dramatically reduced. Training and retraining of the sites and assurance of protocol compliance then become the focus of monitoring, rather than SDV of original data. In addition, weaknesses within the sponsor's clinical research personnel can also be identified and corrected.

While compiling errors by form is not ideal, as the forms contain several different field types of data entry (eg, free text, radio buttons, check boxes), an initial decision was made to simplify the analysis by analyzing field types within forms. Since the main source of database changes was transcription errors, it was decided that a further analysis by the mechanism by which the error was discovered (eg, field monitoring, in-house monitoring, batch edit checks) would not have added to interpretation of the data.

## CONCLUSION

With the advent of direct data entry and the integration of EDC with EMR, and the elimination of the requirement to transcribe from a paper source record to an EDC system, error rates should decrease dramatically with a corresponding improvement of data quality. In addition, protocol violations and data outside the normal range can be identified at the time of data entry and not days, weeks, and months after the fact.

When the clinical research sites are able to bypass the use of traditional paper source documents as original data, and the pharmaceutical industry performs risk-based monitoring together with data management and statistical tools, there is the real possibility to:

1. Increase the quality of clinical trial data.
2. Reduce the time to database lock.
3. Stop development of ineffective or unsafe drugs early in the cycle.
4. Reduce unnecessary work.
5. Reduce the costs of clinical trials.
6. Accelerate the time to market.

---

Ralph B. D'Agostino Jr. and David Horn have disclosed that they served as consultants to Target Health, Inc. The other authors report no relevant relationships to disclose.

## REFERENCES

1. Williams GC. The other side of clinical trial monitoring: assuring data quality and procedural adherence. *Clin Trials*. 2006;3:530–537.
2. Khosla R, Verma DD, Kapur A, Khosla S. Efficient source data verification. *Ind J Pharmacol*. 2000;32:180–186.
3. Eisenstein EL, Lemons PW, Tardiff BE, et al. Reducing the costs of phase II cardiovascular clinical trials. *Am Heart J*. 2005;149:482–488.
4. Califf RM, Karnash SL, Woodlief LH. Developing systems for cost-effective auditing of clinical trials. *Control Clin Trials*. 1997;18:651–660.
5. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials*. 2008;5:75–84.
6. Helms RW. Data quality issues in electronic data capture. *Drug Inf J*. 2001;827–837.
7. Mitchel JT, Kim YJ, Choi JH, et al. The final efrontier. *Appl Clin Trials Online*. 2010;(May).
8. Food and Drug Administration. Guidance for industry: providing clinical evidence of effectiveness for human drug and biological products. May 1998.
9. Food and Drug Administration. Concept paper: quality in FDA-regulated clinical research; background to HSP/BIMO Workshop 5/10–5/11/07. 2007.
10. International Conference on Harmonization. ICH topic E9: statistical principles for clinical trials. September 1998.
11. Society for Clinical Data Management. *Good Clinical Data Management Practices (GCDMP)*. October 2010.