Evaluation of a Teleform-based data collection system: A multi-center obesity research case study

Todd M. Jenkins, PhD, MPH[*],
3333 Burnet Avenue, MLC 7000, Cincinnati, OH, USA 45229-3039

Tawny Wilson Boyce, MS, MPH[*],
3333 Burnet Avenue, MLC 7000, Cincinnati, OH, USA 45229-3039

Rachel Akers, MPH[*],
3333 Burnet Avenue, MLC 5041, Cincinnati, OH, USA 45229-3039

Jennifer Andringa, BA[*],
3333 Burnet Avenue, MLC 5041, Cincinnati, OH, USA 45229-3039

Yanhong Liu, MS[*],
3333 Burnet Avenue, MLC 5041, Cincinnati, OH, USA 45229-3039

Rosemary Miller, RN, CCRC[*],
3333 Burnet Avenue, MLC 7000, Cincinnati, OH, USA 45229-3039

Carolyn Powers, RD[*], and
3333 Burnet Avenue, MLC 5041, Cincinnati, OH, USA 45229-3039

C. Ralph Buncher, ScD[†]
University of Cincinnati, PO Box 670056, Cincinnati, OH, USA 45267-0056

Tawny Wilson Boyce: Tawny.Wilson@cchmc.org; Rachel Akers: Rachel.Akers@cchmc.org; Jennifer Andringa: Jennifer.Andringa@cchmc.org; Yanhong Liu: Yanhong.Liu@cchmc.org; Rosemary Miller: Rosemary.Miller@cchmc.org; Carolyn Powers: Carolyn.Powers@cchmc.org; C. Ralph Buncher: Ralph.Buncher@uc.edu

Abstract

Utilizing electronic data capture (EDC) systems in data collection and management allows automated validation programs to preemptively identify and correct data errors. For our multi-center, prospective study we chose to use TeleForm, a paper-based data capture software that uses

Corresponding Author: Ph. +1 (513) 803-2181, Fax. +1 (513) 803-5050, todd.jenkins@cchmc.org.
[*]Cincinnati Children's Hospital Medical Center
[†]University of Cincinnati

recognition technology to create case report forms (CRFs) with similar functionality to EDC, including custom scripts to identify entry errors. We quantified the accuracy of the optimized system through a data audit of CRFs and the study database, examining selected critical variables for all subjects in the study, as well as an audit of all variables for 25 randomly selected subjects. Overall we found 6.7 errors per 10,000 fields, with similar estimates for critical (6.9/10,000) and non-critical (6.5/10,000) variables – values that fall below the acceptable quality threshold of 50 errors per 10,000 established by the Society for Clinical Data Management. However, error rates were found to widely vary by type of data field, with the highest rate observed with open text fields.

Keywords

data collection; Teleform; data quality

Introduction

High quality data are essential to any research endeavor, making it important to proactively identify and correct errors prior to entering values into the database. Recently, there has been increased adoption of electronic data capture (EDC) systems given their reported performance in error prevention [1]. However, the cost of developing or purchasing an EDC system may often exceed the means of many researchers. One less expensive alternative is paper-based data capture software with recognition technology. Recognition technology includes intelligent character recognition (ICR), optical character recognition (OCR), optical mark recognition (OMR), and bar code recognition (BCR). These recognition technologies interpret machine and hand-printed marks and convert them into data. However, a challenge to using recognition software is identifying and removing errors produced by stray marks, respondent corrections, or improperly completed fields. While many errors can be uncovered by carefully defining fields through built-in field property settings, we have found that the error reducing capacity of recognition software is greatly expanded by use of customizable programs referred to as "scripts". Through customizable scripts, the form designer is able to add robust cross-field edit checks and other complex checks not available via field property settings. Others have investigated the data fidelity resulting from recognition software [2–8]. These studies have identified factors such as illegible handwriting, incomplete or faint markings, and responses placed outside of data entry fields as common issues leading to data inaccuracies [2, 4–6]. However, none of these investigations to our knowledge have reported use of customizable scripts. In this report, we describe our approach using optimized recognition software and provide evidence of its accuracy.

Methods and Materials

We selected TeleForm as our primary tool for the data collection and management based on the customizable features and the extensive experience of the Data Management Center, who have standardized many of the processes and workflows. TeleForm is a paper based data capture system that uses ICR/OCR/OMR/BCR software, called RecoFlex, to convert marks on scanned forms into data which then can be evaluated and entered into a database. The

mechanics of Teleform processing have been extensively described elsewhere [3, 7]. The software package is comprised of several applications, the core being: Designer, Scan Station, Reader, and Verifier.

Similar to an EDC system, case report form (CRF) creation involves defining data fields by configuring many built-in property settings (e.g., data type, entry requirements, and ranges). Our standard data types and validation settings are presented in Table 1.

In addition to the settings listed in Table 1, we use other variable-specific settings such as entry required, database lookup, restricting expected characters, and applying formatting. Furthermore, we have raised the default reader confidence level of constrained print fields and image zones from 80 to the maximum of 100 for our entire system - this alteration greatly reduces the chance that a misidentified character will be passed to the database.

These configuration settings are able to identify many CRF completion errors and mark fields for review. However, use of customizable scripts linked to each CRF in the system is key in catching a wider spectrum of potential errors prior to entering the data into the database. For example, these scripts make logical comparisons between fields to ensure skip pattern logic is correct.

There are several points at which the customized scripts may be implemented. The first point applies status flags to data that may have failed skip pattern logic or other logical comparisons. If RecoFlex could not interpret a mark as data or a status flag was applied because the evaluated data either violates a property setting or the scripting logic, the operator can then either correct the data to match the hand writing on the CRF or generate a query to be rectified from the study site submitting the CRF. We have found that scripted edit checks are extremely useful in identifying stray marks that were misinterpreted as data and identifying required data that were not read in by the software because they were not marked within the data entry field. Without the use of scripting, fields with conditional requirements are unable to be evaluated for proper completion until all data were exported to the database.

Following this initial review, another more comprehensive block of scripting is run to re-evaluate the edited data. This scripting repeats the original edit checks with the addition of more complex logical checks. These include: required field completion based on skip patterns, date comparisons to verify they occur in logical order, ensuring a response of "none" is not selected with other responses for "select all that apply" type variables, and yearly visits reported as months can be checked to verify they are multiples of 12 – something a basic range check would not be able to restrict. Because the scripting triggers in a recursive manner, item responses which violate the form's logic will continue to be flagged after subsequent review. After all problematic data are identified, the secondary operator sends the generated queries to the reporting site and rectifies issues while the data remain held in the TeleForm system. Once data are corrected, they are then entered into the database.

Case study - methodology

To quantify the accuracy level of our optimized recognition system, in Spring 2010 we conducted an in-depth data evaluation from Teen-LABS, a multi-center, prospective study investigating bariatric surgery in adolescents [9]. Seventy-four unique forms that contained a mix of selection, text, and numeric data fields were used by trained coordinators to collect data. A total of 255,964 data fields were evaluated, of which 61.4% (n=157,101) were selection fields, 34.2% (n=87,659) were numeric fields, and 4.4% (n=11,204) were text fields. CRFs were developed from completed forms using Teleform v10.0. The system was consistently upgraded to remain up to date. The assessment included a review of all variables from five randomly selected subjects at each of the five study sites, as well as a complete review of all collected critical outcome variables for all study participants. A data manager conducted the assessment by comparing the CRFs against the study database. Each error received a field type designation (selection, text, numeric), description of who committed the error (i.e. site coordinator, data coordinating center (DCC), N/A, unknown), and corrective action steps. This information on potential errors was then independently reviewed by a different data manager.

Findings in this report were restricted to items designated as data processing errors attributed to the DCC (e.g., Teleform read errors). Each of these items was categorized as a critical or non-critical error. Critical variables were defined as fields that directly related to primary study endpoints. Error rates, expressed as number of errors per 10,000 fields assessed, were calculated by dividing the number of errors detected by the number of fields inspected, multiplied by 10,000. Approximate ninety-five percent confidence intervals were also generated. Further, errors were calculated by critical or non-critical and field type designations.

Case study - results

Of the 255,964 fields evaluated, 171 errors were identified, corresponding to 6.7 errors per 10,000 fields (95% CI: 5.7, 7.7; Table 2). Among the 119,998 critical fields reviewed, 83 errors were identified, or 6.9 per 10,000 fields (95% CI: 5.4, 8.4). Of the 135,966 non-critical values assessed, 88 errors were discovered, or 6.5 per 10,000 fields (95% CI: 5.1, 7.8).

Among selection field variables, 34 errors were identified among 157,101 fields, (2.2 per 10,000 fields). There were 46 errors among 11,204 text fields (41.1 per 10,000 fields). Within numeric fields, 91 errors were detected among 87,659 numeric fields, representing 10.4 errors per 10,000 numeric fields. Field type-specific error rates differed little across critical/non-critical variable classifications.

Conclusion/Discussion

Overall, we found 6.7 errors per 10,000 fields – an amount that falls well within the commonly utilized acceptable quality threshold of 50 errors per 10,000 [10]. In contrast to the larger CRF-to-database audit literature, our rate was markedly lower than the average of 14 errors per 10,000 fields [11]. Additionally, our error estimates for critical (6.9/10,000)

and non-critical (6.5/10,000) variables also met or exceeded common quality metrics of 0 to 10 errors per 10,000 fields and 20 to 100 errors per 10,000 fields, respectively [10]. We found that the frequency of errors widely varied by type of data field, a conclusion noted by several others [2, 4–7]. Our results indicated text fields were associated with the highest rate of errors, while the lowest rate was observed with selection-based fields. Compared to previous reports evaluating the accuracy of Teleform without the use of customizable scripts, our overall error rate compares favorably to some [2, 7] and is similar [5] to or marginally higher [4] than estimates reported previously.

Quan et al. reported an error rate of 0.4% (or 40 per 10,000 fields) and concluded that data field type had a great effect on data quality [7]. However, they did not provide field type-specific estimates. Guerette et al. compared Teleform accuracy against manual single data entry [2]. They observed a 1.4% (or 140 per 10,000 fields) rate of error for Teleform and noted that Teleform error rates associated with print fields were nearly three times that with manually entered data. Again, data type-level error estimates were not reported. Our overall, selection, and numeric field error rates were found to be similar to those reported by Jorgensen [5]. However, text field-specific error estimates were not described in their assessment.

In contrast, Jinks et al. reported an accuracy level that slightly exceeded our findings – 0.041% error rate (or 4.1 per 10,000 fields)[4]. However, their evaluation was comprised of selection and numeric fields, but did not include text fields. To create a more accurate comparison, we calculated the error rate combining our selection and numeric fields. This combined selection/numeric error rate (5.1 per 10,000; 95% CI: 4.2, 6.0) was found to be similar to the estimate reported by Jinks.

Our use of customizable scripts allowed us to achieve accuracy levels that were predominantly superior to other published findings. Additionally, we were able to replicate the previously reported variation in error rates across data field types. As might be expected, we found that text fields had the highest levels of inaccuracy, with our text field-rich, patient-reported medications form accounting for the largest percentage of errors among all forms. Similarly, Quan and colleagues noted that free-text fields were labor-intensive for data managers and specifically singled out patients' medications as a primary contributor[6]. They suggested that collection of medication data could be improved by utilizing a check-box approach that includes commonly encountered medications, potentially incorporating the World Health Organization Drug Dictionary classification system. We, too, agree that alternatives for potentially problematic free-text and other data field types should be explored in the planning phases of every project.

By optimizing TeleForm through customized scripts, we were able to achieve high quality data for a multi-center clinical study. We found that this technique yielded error rates that exceeded industry standards, with some variation observed by type of data field. This approach to data capture/management should be considered as a viable, high-quality, and potentially lower cost alternative to EDC systems.

Acknowledgments

References

1. Pavlovic I, Kern T, Miklavcic D. Contemporary clinical trials. 2009; 30:300–316. [PubMed: 19345286]

2. Guerette, P.; Robinson, B.; Moran, WP.; Messick, C.; Wright, M.; Wofford, J.; Velez, R. Proceedings / the … Annual Symposium on Computer Application [sic]. Medical Care. Symposium on Computer Applications in Medical Care; 1995; p. 86-90.

3. Hardin JM, Woodby LL, Crawford MA, Windsor RA, Miller TM. Public health nursing. 2005; 22:366–370. [PubMed: 16150019]

4. Jinks C, Jordan K, Croft P. Computers in biology and medicine. 2003; 33:425–437. [PubMed: 12860466]

5. Jorgensen CK, Karlsmose B. Computers in biology and medicine. 1998; 28:659–667. [PubMed: 9878977]

6. Quan H, Biondo PD, Stiles C, Moulin DE, Hagen NA. Contemporary clinical trials. 2011; 32:173–177. [PubMed: 21147264]

7. Quan KH, Vigano A, Fainsinger RL. Journal of palliative medicine. 2003; 6:401–408. [PubMed: 14509485]

8. Wahi MM, Parks DV, Skeate RC, Goldin SB. Journal of the American Medical Informatics Association : JAMIA. 2008; 15:386–389. [PubMed: 18308994]

9. Inge TH, Zeller M, Harmon C, Helmrath M, Bean J, Modi A, Horlick M, Kalra M, Xanthakos S, Miller R, Akers R, Courcoulas A. Journal of pediatric surgery. 2007; 42:1969–1971. [PubMed: 18022459]

10. Good Clinical Data Management Practices. Society for Clinical Data Management; Apr. 2011 Measuring Quality Data.

11. Nahm ML, Pieper CF, Cunningham MM. PloS one. 2008; 3:e3049. [PubMed: 18725958]

Summary

High quality data are essential to any research endeavor, making it important to identify and correct errors prior to entering values into the database. There has been increased adoption of electronic data capture (EDC) systems given their reported performance in error prevention. However, the cost of developing or purchasing an EDC system may exceed the means of many researchers. One less expensive alternative is paper-based data capture software with recognition technology. These recognition technologies interpret machine and hand-printed marks and convert them into data. However, a challenge to using recognition software is identifying and removing errors produced by stray marks, respondent corrections, or improperly completed fields. While many errors can be uncovered by carefully defining fields through built-in field property settings, we have found that the error reducing capacity of recognition software is greatly expanded by use of customizable programs (i.e., "scripts"). In this report, we describe our approach using optimized recognition software and provide evidence of its accuracy.

We selected TeleForm, paper-based data capture system with recognition technology, as our primary tool for the data collection and management. Case report form (CRF) creation in Teleform involves defining data fields by configuring many built-in property settings (e.g., data type, entry requirements, and ranges). We also use other variable-specific settings such as database lookup, restricting expected characters, and applying formatting. These configuration settings help identify many CRF completion errors and mark fields for review. However, use of customizable scripts linked to each CRF is key to identifying a wider spectrum of potential errors.

To quantify the accuracy level of our optimized recognition system, we conducted an in-depth data evaluation from prospective study investigating bariatric surgery in adolescents. The assessment included a review of all variables from five randomly selected subjects at each of the five study sites, as well as a complete review of all critical outcomes for all study participants. A data manager conducted the assessment by comparing the CRFs against the study database. Each error received a field type designation (selection, text, numeric), description of who committed the error (i.e. site coordinator, data coordinating center (DCC), N/A, unknown), and corrective action steps.

Findings were restricted to items designated as data processing errors attributed to the DCC (e.g., Teleform read errors). Each error categorized as critical or non-critical. Critical variables were defined as fields that directly relate to primary study endpoints. Error rates were expressed as number of errors per 10,000 fields.

Of the 255,964 fields evaluated, 171 errors were identified, corresponding to 6.7 errors per 10,000 fields – an amount that falls below the commonly utilized acceptable quality threshold of 50 errors per 10,000. Error estimates widely varied by type of data field. Our results indicated text fields were associated with the highest rate of errors, while the lowest rate was observed with selection-based fields.

By optimizing TeleForm through customized scripts, we were able to achieve high quality data for a multi-center clinical study. We found that this technique yielded error rates that exceeded industry standards.

Table 1

Standard Settings

Data type	Field type	Validations
Numeric/Date	Constrained print	Ranges
Text	Image zone	Always review & Data review
Multiple choice – select one	Rectangle choice fields	None
Multiple choice – select all that apply	Rectangle choice fields	Always review & Data review

Table 2

Error Rates per 10,000 fields (95% CI) by Field Type and Critical/NonCritical Classification.

	Critical	Non-Critical	Total
Selection	2.2 (1.1, 3.2)	2.2 (1.2, 3.2)	**2.2 (1.4, 2.9)**
Text	39.0 (23.4, 54.6)	43.5 (25.3, 61.7)	**41.1 (29.2, 52.9)**
Numeric	10.8 (7.6, 14.1)	10.0 (7.2, 12.9)	**10.4 (8.3, 12.5)**
Total	**6.9 (5.4, 8.4)**	**6.5 (5.1, 7.8)**	**6.7 (5.7, 7.7)**