

The “Measuring Outcomes of Clinical Connectivity” (MOCC) Trial: Investigating Data Entry Errors in the Electronic Primary Care Research Network (ePCRn)

Patricia Fontaine, MD, MS, Tai J. Mendenhall, PhD, LMFT, Kevin Peterson, MD, MPH, and Stuart M. Speedie, PhD

Introduction: The electronic Primary Care Research Network (ePCRn) enrolled PBRN researchers in a feasibility trial to test the functionality of the network’s electronic architecture and investigate error rates associated with two data entry strategies used in clinical trials.

Methods: PBRN physicians and research assistants who registered with the ePCRn were eligible to participate. After online consent and randomization, participants viewed simulated patient records, presented as either abstracted data (short form) or progress notes (long form). Participants transcribed 50 data elements onto electronic case report forms (CRFs) without integrated field restrictions. Data errors were analyzed.

Results: Ten geographically dispersed PBRNs enrolled 100 members and completed the study in less than 7 weeks. The estimated overall error rate if field restrictions had been applied was 2.3%. Participants entering data from the short form had a higher rate of correctly entered data fields (94.5% vs 90.8%, $P = .004$) and significantly more error-free records ($P = .003$).

Conclusions: Feasibility outcomes integral to completion of an Internet-based, multisite study were successfully achieved. Further development of programmable electronic safeguards is indicated. The error analysis conducted in this study will aid design of specific field restrictions for electronic CRFs, an important component of clinical trial management systems. (J Am Board Fam Med 2007;20:151–159.)

Background

In 2002, the Director of the National Institutes of Health (NIH) convened a series of meetings to chart a “Roadmap” for medical research in the 21st century. Developed with input from nationally recognized leaders in academia, industry, government,

and the public, the NIH Roadmap presented an urgent call for a more efficient and productive system of medical research. One of the major Roadmap themes, “Re-Engineering the Clinical Research Enterprise,” is intended to promote the rapid translation of basic research findings into treatments and prevention strategies that will improve health in the United States. The Roadmap seeks to expand capacity and improve communication within existing clinical research networks and to develop clinical research protocols that capitalize on modern information technology platforms with improved features for collecting and recording research data.

This article was externally peer-reviewed.

Submitted 4 May 2006; revised 6 November 2006; accepted 9 November 2006.

From the Department of Family Medicine and Community Health, University of Minnesota Medical School, Minneapolis, MN (PF, TJM, KP); and Division of Health Informatics, Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis, MN (SMS).

Funding: This work was supported by National Institutes of Health contract no. HHS268N200425212C, “Re-Engineering the Clinical Research Enterprise.”

Prior presentations: This article is based on a presentation made at the American Academy of Family Physicians National Research Network 2006 Convocation of Practices and Networks, Dallas, TX, February 23–26, 2006. Minnesota Academy of Family Physicians Annual Research Forum, Maple Grove, MN, March 2006.

Conflict of interest: none declared.

Corresponding author: Patricia Fontaine, MD, MS, Department of Family Medicine and Community Health, Suite 220 Dinnaken, 925 Delaware Street Southeast, Minneapolis, MN 55414 (E-mail: pfontaine@um Physicians.umn.edu).

The ePCRn

Funded through the NIH Roadmap Initiative, the ePCRn is a state-of-the-art, Internet-based electronic architecture being developed to allow PBRNs to enroll subjects and pool data for large randomized controlled trials.¹ The ePCRn features a secure web portal for online recruitment and consent, real-time computerized randomization, and capability for direct data entry into a centralized database.

For the ePCRN to become a successful tool that is widely adopted by researchers, it is important to demonstrate the functionality of the electronic system and validity of the procedures used. The ePCRN must develop a data collection system that promotes accuracy of data entry and is in compliance with the standards set by the US Department of Health and Human Services for computerized systems used in clinical trials (see www.fda.gov/ora/compliance_ref/bimo/ffinalcct.htm).² As a tool for multisite, multiple network trials, the ePCRN must develop appropriate contractual relationships, address Institutional Review Board (IRB) concerns, and overcome significant challenges in communication and training for research protocols. The ePCRN's initial feasibility trial, the Measuring Outcomes of Clinical Connectivity (MOCC) Trial was designed as a practical study with 2 goals. The first goal was to lay the groundwork for conducting randomized controlled trials (RCTs) in PBRNs by introducing PBRN researchers to the ePCRN's Internet-based research portal. This included the establishment of contracts with participating PBRNs, the development of shared materials for recruitment and training, and the creation of prototypes for electronic case report forms (CRFs) that would be compatible with the ePCRN's clinical trial management software. The second goal of the MOCC trial was to evaluate two alternative data entry strategies commonly used in clinical trials and develop recommendations for refining future versions of electronic CRFs.

Rationale for Data Entry Strategies Tested in the MOCC Trial

Planned and systematic procedures to optimize the quality of entered data are important in any clinical trial and may be particularly important in PBRNs, where practices are distant from each other and training and oversight are logistically difficult. One approach to reducing data entry errors is chart abstraction, where data are copied onto a CRF. All desired data points are listed on the CRF, and a trained individual examines the medical record, systematically elicits the data, and enters it on the form. Thus, the common practice of using a CRF requires 2 steps: 1) transcribing data from the patient record to the CRF, and 2) transcribing data from the CRF to the electronic database. As an alternative to creating a CRF, entering data directly from a medical record into a study's database has

the theoretical advantage of requiring only one step that could be completed by the clinician during a patient visit, thus allowing point-of-care data entry. Although Web-based entry of personal demographic information has become common in all walks of life, the medical literature lacks evidence regarding optimal field configurations for electronic CRFs and the types of electronic validations best suited to minimize errors from data transfer. Specifically, the literature does not currently provide evidence on the accuracy of data entered from CRFs versus data transcribed directly from the medical record. Thus, the MOCC trial sought to characterize data entry errors and compare error rates when data were transcribed from short forms resembling CRFs versus directly from a simulated medical record.

Methods

Participants

The MOCC Trial was approved by the University of Minnesota IRB. Ten PBRNs representing diverse geographic locations across the United States participated in the MOCC Trial (see Table 1). Each PBRN director was asked to recruit 10 participants from among their members. An e-mail message describing the MOCC trial was provided to PBRN directors for this purpose, and an electronic announcement was also visible on the ePCRN desktop after logging on to the Web portal.

Family medicine physicians and research assistants were eligible to participate in the MOCC trial if they had previously completed the ePCRN authentication process through their participating PBRN. The authentication process required several steps, beginning with nomination by the PBRN Director. PBRN directors certified each new user's research competency, including training in the protection of human subjects and privacy rules (Health Insurance Portability and Accountability Act), and appropriate Federalwide Assurance (IRB) coverage. The user then completed a notarized identification form to obtain a secure logon device (RSA SecurID fob). Finally, the user registered with the ePCRN by logging in to the secure system with the RSA fob and entering individual and practice demographic information into the ePCRN central database. After all steps in the authentication process were completed, the user was

Table 1. Practice-Based Research Networks Participating in the ePCRN

APBRN	Alabama Practice-Based Research Network Myra Crawford, PhD; T. Michael Harrington, MD; Jean Marie White	Alabama
BIGHORN	Building InvestiGative practices for better Health Outcomes Research Network Peter Smith, MD; Linda Niebauer	Colorado
CaReNet	Colorado Research Network Bennett Parnes, MD; Doug Fernald, MA; Linda Niebauer	Colorado
HPRN	High Plains Research Network John Westfall, MD; Linda Zittleman, MSPH	Colorado
INET	Indiana Family Practice Research Network Deborah Allen, MD; Carolyn Muegge	Indiana
MAFPRN	Minnesota Academy of Family Physicians Research Network Kevin Peterson, MD; Patricia Fontaine, MD; Tai Mendenhall, PhD; Jacky Hansen	Minnesota
OKPRN	Oklahoma Physicians Resource Research Network James W. Mold, MD; Zsolt Nagykalai, PhD	Oklahoma
PSARN	Penn State Ambulatory Research Network Alan Adelman, MD; Marie Graybill, RN	Pennsylvania
SoFla-PBRN	South Florida Primary Care Practice-Based Research Network John G. Ryan, DrPh; Fulton Vasquez, MD; Andrea Escobar	Florida
STARNET	South Texas Ambulatory Research Network Walter Calmbach, MD; Michael Parchman, MD	Texas
UNYNET	Upstate New York Practice-Based Research Network Chester Fox, MD; Christine Kudla	New York
NRN	National Research Network–American Academy of Family Physicians Wilson Pace, MD; Mindy Spano	National

invited by either e-mail or desktop message to participate in the MOCC Trial.

Interventions

After logging onto the secure ePCRN portal and consenting to participate in the MOCC Trial, participants were randomized to receive simulated health information for 5 fictional patients in 1 of 2 formats: 1) short form, an abstracted form very similar to a CRF; or 2) long form, a dictated progress note in a Subjective-Objective-Assessment-Plan (SOAP) format. The short and long forms for each simulated patient contained identical information. Participants were asked to transfer 50 data elements, including patient demographics, date of visit, weight, blood pressure, and hemoglobin A1c and serum creatinine values, from the short or long form into an electronic CRF. The electronic CRF was identical for the 2 groups and was presented on the electronic desktop next to the short or long form (see Figures 1 and 2).

Because the numbers and types of data errors were the subject of the investigation, the electronic CRF contained no integrated field restrictions such as alphanumeric or range validations. Participants

were able to open the records in any order and to log off and return to the trial later. Participants were queried after each case, “Do you want to go back and check your data?” Once each participant was finished, the data were submitted and no further access was available.

Outcomes

Overall accuracy of data entry was the primary outcome, and it was evaluated in two ways: 1) as the total percentage of correctly entered data fields, and 2) as the number of patient records where 100% of fields were correctly entered. Researchers developed a priori rules for distinguishing correct from incorrect responses, and answers were re-coded into a bimodal format (“correct” or “incorrect”) for analysis.

Feasibility outcomes considered indicative of ease of use for the participant included the time to complete the study, the total number of logins required, and the number of disconnections from the system due to technical problems. In addition, the time and date of each logon and the number of concurrent users (peak usage) were tracked electronically.

Directions Exit

The MOCC Trial uses simulated health information from mock patients to evaluate the ePCRN's technical capabilities as well as optimal procedures for data entry.

Step 1:
Click on any one of the 5 icons along the right hand side.

Step 2:
Fill out the requested information on the data entry form.

Step 3:
Click on the "Submit" button when finished.

Step 4:
Repeat steps 1 through 3 for the remaining 4 patients.

If at any point you would like to go back and make changes to one of the documents you may do so by reselecting that document.

Data Entry Form - Patient Record 1

Cancel

Patient Name: Lashanna S. Jenkins

Date of Birth: 5/21/1960 mm/dd/yyyy

Date of Visit: mm/dd/yyyy

Gender: ☐ Male ☒ Female

Pregnant: ☐ Yes ☒ No

Age: years

Weight: lbs

Blood Pressure: mmHg

Hemoglobin A1C: %

Serum Creatinine: mg/dl

Submit Patient Data

Electronic Case Report Form

Patient Record 1

DATE OF VISIT: 5-30-05
NAME: Lashanna S. Jenkins
Ht: 4'11" Wt: refused BP: 132/78 mmHg Pulse: 86
Tobacco: Y
Allergies: None known

S: Lashanna is a pleasant 45 year-old African American female who was diagnosed in Jan 2001 with Type 2 diabetes, here for routine check. She checks her blood sugars 2x/day. She did forget her glucose log at home, but states that typical am values range from 180-200 and pm values range from 210-230. Lashanna refused to let us weigh her today stating that she knows she has gained weight since her last visit despite her dietary and exercise efforts.

PMH: Positive for hypertension, hyperlipidemia, depression. Immunizations are up to date.

MEDICATIONS: Metformin 500mg QD, Glyburide 10mg QD, Avandia 4mg QD, Atorvastatin 20mg QD, HCTZ 25QD, Metoprolol 50mg QD, ASA 325mg QD, Wellbutrin 150mg QD.

HABITS: 1-2 drinks/week, does not smoke.

ROS: She states that her depression has been under good control for the last 6 months. All other ROS negative.

O: At her last visit, March 12, she weighed 245lbs. Her last two BP's in March and Feb were 110/78 mmHg and 128/82 mmHg.

HEENT: normal. Eyes: PERLLA, EOM full, fundi benign. Lungs: Clear to auscultation throughout. Heart: S1S2 normal. Extremities: foot exam showed intact light touch sensation throughout to monofilament.

LABS: A1C and Lipid Panel were done earlier this week. A1C value 8.4 %, Total Cholesterol 200, Triglycerides 94, HDL 40, LDL 140. Creatinine 1.2 mg/dl.

A: 1. Type 2 diabetes and hyperlipidemia -- need improved control.
2. Hypertension -- today's 132/78 mmHg BP was slightly elevated, however, previous values have been good and we will recheck next visit.

P: 1. We will increase metformin to 1000mg QD and atorvastatin to 40mg QD.
2. In addition I have recommended that she undergo nutrition and diabetes education with Sarah Lee our diabetic educator. An appt has been set up for Jun 1st.
3. Her last dilated eye exam was in 2003. I let her know she needs yearly exams and referred her to the ophthalmology group at N.W. Eye.
4. RCT in 1-2 months.

Amy D. Bennett, MD

Long Form

Figure 1. MOCC Trial desktop for the Long-form group.

Sample Size

Using the total percentage of correct fields as the primary outcome variable, we hypothesized that the short-form group would have 98% correct, compared with 90% for those entering data using the long form. For moderate effect size ($d = 5.0$) with a power of 0.80, 49 participants were needed in each group.³

Randomization and Blinding

Group assignment to either short or long form was determined by a programmed block randomization formula that was implemented in real time following the online consent process. Participants, investigators, and PBRN directors were blind to group assignments throughout the trial and the analysis.

Statistical Methods

Descriptive and frequency data for all variables were analyzed using SPSS version 14.0 for Windows. Comparisons of group means were conducted through independent samples t tests, and categorical variables were evaluated via χ^2 analyses.

Types of errors were subjected to a two-way, repeated-measures analysis of variance (patients within doctors) with one between group factor (groups) and one within group factor (patients).

Results

Recruitment

The first 100 ePCRN registrants to access the Web portal and agree to participate were randomized into the 2 study groups. The study began on November 14, 2005, and ended on December 29, 2005; thus recruitment, data entry, and data download were completed in less than 7 weeks.

Participants

Of 100 individuals who consented to be in the study, 98 completed it. Two participants logged out of the study without entering any data, thus data from 98 participants were available for analysis. Demographic characteristics of the participants are presented in Table 2. Nearly all participants (89%) were physicians, including MDs, DOs, and physicians with more than one degree (eg, MD,

The MOCC Trial uses simulated health information from mock patients to evaluate the ePCRN's technical capabilities as well as optimal procedures for data entry.

Step 1:
Click on any one of the 5 icons along the right hand side.

Step 2:
Fill out the requested information on the data entry form.

Step 3:
Click on the "Submit" button when finished.

Step 4:
Repeat steps 1 through 3 for the remaining 4 patients.

If at any point you would like to go back and make changes to one of the documents you may do so by reselecting that document.

Data Entry Form - Patient Record 1

Cancel

Patient Name: Lashanna S. Jenkins

Date of Birth: 5-21-1960 mm/dd/yyyy

Date of Visit: mm/dd/yyyy

Gender: ☐ Male ☒ Female

Pregnant: ☐ Yes ☒ No

Age: years

Weight: lbs

Blood Pressure: mmHg

Hemoglobin A1C: %

Serum Creatinine: mg/dl

Submit Patient Data

Electronic Case Report Form

Patient Record 1

Name: Lashanna S. Jenkins

Date of birth: 5-21-1960

Date of visit: 5-30-05

Gender: female

Pregnant?: No

Age: 45 years

Weight: refused

Blood pressure: 132/78 mmHg

Tobacco use? Never

Hemoglobin A1C: 8.4 %

Creatinine: 1.2 mg/dl

Eye exam documented in record? Yes

Date: 5-21-03

Short Form

Patient Record 1

Patient Record 2

Patient Record 3

Patient Record 4

Patient Record 5

Figure 2. MOCC Trial desktop for the short-form group.

MPH). There were no significant differences between the study groups in age, gender, degree, or number of years in practice. In addition, analysis of group assignment by PBRN of participant confirmed a random distribution of PBRNs in each group ($\chi^2 = 91$, $df = 9$, $P = .92$).

Outcomes

Accuracy of Data Entry

The overall error rate, defined as the combined number of incorrect and missing data fields of 4900 total fields, was 7.3% (see Table 3). Participants entering data from the short form had a higher

Table 2. Demographic Characteristics of Participants by Group Assignment (Long- or Short-form Group)

	Short-form Group n = 48	Long-form Group n = 50	P Value
Gender (%male)	28	35	.23
Age (mean in years)	49.2	46.9	.89
Professional degree			.60
MD/DO*	44	43	
NP/PA	2	1	
MS, MPH, PhD	2	4	
Ethnicity			.33
Non-Hispanic/Latino	46	46	
Hispanic/Latino	1	3	
Years in practice (mean)	21.7	20.8	.66

* MD, doctor of medicine; DO, doctor of osteopathy; NP, nurse practitioner; PA, physician's assistant; MS, master of science; MPH, master of public health; PhD, doctoral degree.

Table 3. Errors (Incorrect and Missing Data) by Field Type* and by Group Assignment

Field Types (N = 4900)	Data Elements	Errors		Totals (%)
		Short-form Group†	Long-form Group‡	
Numeric (n = 2450)	Age	2	15	
	Weight	12	34	
	Blood Pressure	0	5	
	A1C	23	29	
	Creatinine	26	37	
	Numeric subtotal			183 (7.5%)
Date (n = 980)	Birth date	15	9	24
	Visit date	13	6	19
	Date subtotal			43 (4.4%)
Text (n = 490)	Name	28	56	84
	Text subtotal			84 (17.1%)
Select Option (n = 980)	Pregnant	12	34	45
	Gender	1	1	2
	Select subtotal			47 (4.9%)
	Total			357 (7.3%)

* Field types are groups of data elements with common characteristics. The MOCC Trial case report forms contained numeric fields (numbers only), date fields (numbers in specific date formatting), text fields (alphabetic characters), and select option fields (choice of options from a menu).

† Forty-eight participants entered 10 data items for 5 cases for a total of 2400 data fields.

‡ Fifty participants entered 10 data items for 5 cases for a total of 2500 data fields.

overall accuracy rate (94.5% vs 90.8%, $t = 2.95$, $P = .004$). Of the simulated patient records processed by each group, 142 (59%) were error-free in the short-form group, compared with 106 (42%) in the long-form group ($t = 3.05$, $P = .003$).

Accuracy was further analyzed according to the type of field in which errors occurred (number, date, text, or select option). As seen in Table 4, the short-form group performed significantly better on 3 of the 4 data types, with a minimum of 94% correct data entry in text fields and a maximum of 95% correct in fields where options could be selected. The long-form group's performance ranged from 86.4% to 97% correct, and only the date field

type failed to show a difference in accuracy between groups.

Errors were further characterized as correctable if one of the following programmable field validations could apply: 1) making all fields required (ie, inability to proceed to the next field until preceding fields are filled in); 2) using predetermined ranges and formats for dates and laboratory values; 3) not allowing text entries in numeric fields; and 4) employing select options such as check boxes or drop-down menus whenever possible, including situations where data were not available.

When data were defined in this way and programming errors were not included, the rate of unavoidable errors was 111 per 4900, or 2.3% (see Table 5).

The most common correctable error occurred when information was not available (eg, lab test was not ordered or weight was "refused"). This posed a unique problem for numeric fields where 100 text errors occurred. Nonstandard text responses were frequently entered, such as "na," "N/A," "?," and "unk."

Of the noncorrectable errors, misspelling of the patient name filed was the most common, occurring 49 times and accounting for 44%. Uncom-

Table 4. Percentage Correct Data Entry by Field Type and Group

Field Type	Percent Correct		F(1,96)	P Value
	Short Form	Long Form		
Numeric	94.8	90.4	5.813	.018
Date	94.2	97.0	1.386	.242
Text	94.0	88.2	6.671	.011
Select option	95.0	86.4	6.805	.011
Total	94.5	90.8	—*	.004

* $t = 2.95$.

Table 5. Correctable and Noncorrectable Errors by Field Type*

Field Type	Error Type	Frequency
Numeric	<u>Wrong numeral (within range)</u>	<u>37</u>
	Wrong numeral (outside range)	3
	Text error	100
	Missing (no data entered in field)	43
Date	<u>Wrong numeral (within range)</u>	<u>20</u>
	Wrong numeral (outside range)	2
	Text error	11
	Missing (no data entered in field)	10
Text	<u>Misspelling</u>	<u>49</u>
	Incorrect format or punctuation	35
	Missing (no data entered in field)	0
Select option—gender	Incorrect selection	0
	Missing (no data entered in field)	2
Select option—pregnant	<u>Incorrect selection</u>	<u>5</u>
	Missing (no data entered in field)	22
	Programming error	18
<u>Total noncorrectable error rate</u>		<u>111 (2.3%)</u>
Total error rate		357 (7.3%)

* A total of 4900 fields. Correctable errors are those amenable to programmable validations. Noncorrectable errors (underlined) would have escaped programmed range checks, forced data entry (inability to proceed past an empty data field), and formatting.

monly spelled names were particularly prone to inaccuracy, with 1 name entered 13 different ways.

Wrong numerals in numeric fields were frequently not correctable by range checks; for example, a blood pressure of 132/78 was entered as 110/78. However, in date fields, a program to correct formatting errors made significant improvements in accuracy. By applying a standard formatting transformation (mm/dd/yyyy) to the raw data, the percentage of correct dates improved from 48% to 94% in the short-form group and from 65% to 97% in the long-form group. The corrected date format was used for analysis, since the program is standard, widely available, and did not need to be created for the ePCRN. There was no difference between groups in the accuracy of date entry whether uncorrected or corrected data were used.

Feasibility Outcomes

The short form took an average of 7.0 minutes to complete, compared with 13.6 minutes for the long form. No inadvertent disconnections from the system, security breaches, or adverse events were recorded. Ninety-seven percent of participants completed the trial in one sitting. Seventy-three percent of participants logged on between 7 AM and 5 PM, and the hours between 9 AM and 11 AM

showed the most frequent usage. The system log showed a maximum of 4 concurrent logins over the short course of the trial.

Discussion

In the MOCC Trial, feasibility outcomes integral to completion of an Internet-based, multisite study were successfully achieved, including the following: 1) establishment of contracts with participating PBRNs; 2) development of shared materials for recruitment and training; 3) IRB approval for the online human subjects consent form; and 4) creation of programs for online randomization and electronic case report forms.

Although the 7.3% rate of incorrect entries in this study is higher than the 0.1% to 2% range reported for data entry errors in the literature,^{4,5} this was not unexpected due to the lack of pre-programmed field restrictions and stringent a priori definitions for correct responses. For example, if the correct response for hemoglobin A1c was “8.4,” the response “8.4%” was considered incorrect. The lack of preprogrammed field restrictions and internal validations was intentional, as the study was designed to cast a wide net for all types of errors. However, preprogrammed protocols for field validation could

have prevented many of the errors encountered in the MOCC trial and would have decreased the error rate to 2.3%, a number more consistent with previous literature reports.

The types of errors found in this study were consistent with those described in the literature. For example, programs may systematically generate errors during the data entry process (eg, by incorrectly mapping the contents of a drop down list to an incorrect data value), and human errors can occur when patient measurements are transferred to study forms and subsequently to the database used for analysis.⁶⁻⁸ Types of errors included letter and number reversal (e.g., “sh” entered as “hs;” unintentional repeats or deletions of numbers, letters, or decimal points; extraneous characters; simple transcription and reading errors; data entered into the incorrect field; and skipping fields when data were available).^{9,10} In this study, the name fields were most prone to noncorrectable errors. Long alphabetic fields such as names and addresses have previously been noted to have error rates 10 to 15 times higher than numeric fields.¹¹ Such errors may be reduced by requiring duplicate data entry and verification.^{11,12} Although some have questioned the need for duplicate data entry,⁷ our findings support the practice for complex fields such as names. Even using duplicate entry may not have completely eliminated all of the human errors that we observed. For example, with unusual names, it appeared that the name was not perceived correctly and that a more familiar spelling was inserted (eg, “Jenkins” for “Jenkies”).

Finally, although computer programming can be used to reduce data errors, it can also introduce them.¹¹ This was the case with the select option field (“Pregnant? Y/N”). Invalid responses occurred when a programming error characterized some male patients as pregnant.

Limitations

As a test of a point-of-care data entry strategy for RCTs, the MOCC trial is limited by the fact that the simulated patient data were relatively simple and physicians could enter the data at any time. Nevertheless, the electronic security safeguards built into the ePCR authentication process were sufficient to handle confidential health information from real patients, and the forms represented reasonable electronic versions of paper forms commonly used in PBRN research. For example, the

long form closely simulated a transcribed office visit in standard SOAP format. Although the electronic system performed well in terms of allowing participants to complete the study without malfunctions, the small numbers of concurrent logins did not generate a load test of the system.

As a test of two strategies for data entry, the short form’s advantages would likely be less than seen in our study. In addition to the time it took to transfer data from the short form to the electronic CRF, the total time for data entry in a RCT involving real patients would include the time needed for an initial step of abstracting the data from the medical record to the short form. Additional errors could be introduced during the additional step. Thus, the time required would be increased and accuracy of the short form could be diminished, compared with our findings.

Although the findings regarding data entry errors may apply to data entry by physicians and research assistants in PBRNs, they may not generalize to research settings that utilize other data entry personnel. Trained data entry personnel have substantially different training and may have a different error rate.

Recommendations

As Internet-based data collection systems become more common in PBRNs, and as Web access extends to more primary care examination rooms, point-of-care data collection will become an increasingly utilized methodology in PBRN research. Results of the MOCC Trial indicate that electronic CRFs used in such settings need to be programmed to address the most common types of data entry errors and that a number of simple constraints on data fields can substantially improve data quality. Specifically, numeric data fields should be restricted to acceptable ranges with no text entries allowed. Date fields should be restricted to predefined date formats as well as restricted to acceptable ranges. Select option fields (ie, check boxes) contributed few data entry errors in this study but were the source of a programming error. For complex text fields such as patient names, safeguards against formatting errors and misspellings are required. Such safeguards include 1) using separate text boxes for first, middle, and last names to eliminate punctuation and ordering errors; and 2) making no restrictions on use of uppercase versus lowercase letters. Because of the substantial error rate

due to misspellings, our findings support the use of additional field validation such as duplicate entry for names. Finally, since missing data often cannot be distinguished from inadvertent skipping, fields should be allowed to remain blank only when the data are electronically confirmed as missing.

Future studies should aim to determine whether, as this study suggests, such electronic safeguards can minimize data entry errors so that rates compare favorably with accepted standards. Comparing the performance of specially trained data entry personnel to physicians and research assistants in PBRNs could provide further information about the quality of data PBRNs can be expected to provide in RCTs.

Conclusion

The ePCRN's MOCC Trial demonstrates that large numbers of primary care researchers can be rapidly recruited and use secure Internet-based technology to enter data from geographically dispersed practice sites in a simulation of a multisite clinical trial. The comparison of two data entry methodologies indicates that data transcribed from a short, abstracted form is more accurate than data transcribed directly from a longer medical SOAP note. With either approach, electronic CRFs can provide an important function in maintaining data quality, and further development of programmable electronic safeguards is indicated. The error analysis conducted in this study will aid in designing specific range restrictions, validations, and other field constraints for electronic CRFs, an important component of clinical trial management systems for PBRN research.

We acknowledge the PBRN network directors and coordinators who participated in the MOCC trial; Brendan Delaney, MD, and Theo Arventis, PhD, of the University of Birmingham, UK, for conceptual work; Bruce Center, PhD, for statistical consultation; Joseph Stone, Mark Janoweic, and Adam Wolff for tech-

nical support; Carol Lange and Gillian Lawrence for assistance with recruitment; and Jacky Hansen for administrative support.

References

1. Peterson K, Fontaine P, Speedie S. The electronic Primary Care Research Network (ePCRN): a new era in practice-based research. *J Am Board Fam Med* 2006;19:93–7.
2. United States Department of Health and Human Services. Computerized systems used in clinical trials. US Department of Health and Human Services, Food and Drug Administration 1999.
3. Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
4. Kinkman MS, Green LA. Using ICPC in a computer-based primary care information system. *Fam Med* 1995;27:449–56.
5. Horbar JD, Leahy KA. An assessment of data quality in the Vermont-Oxford Trials Network database. *Control Clin Trials* 1995;16:51–61.
6. Goldberg S, Ancukiewicz M, Niemierko A. The impact of data entry errors on the findings in radiation oncology research. *Int J Radiat Oncol Biol Phys* 2005;63:S191–S192.
7. Gibson D, Harvey AJ, Everett V, Parmar HKB. Is double data entry necessary? The CHART trials. *Control Clin Trials* 1994;15:482–8.
8. Kreher NE, Sesney JW. Accuracy of on-site data entry in a rural primary care research network. *Fam Med* 1995;27:457–9.
9. Samuels RC, Appel L, Reddy SI, Tilson RS. Improving accuracy in a computerized immunization registry. *Ambul Pediatr* 2002;2:187–2.
10. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing and editing data abnormalities. *PloS Med* 2005;2:966–70.
11. Neaton JD, Duchene AG, Svendsen KH, Wentworth D. An examination of the efficiency of some quality assurance methods commonly employed in clinical trials. *Stat Med* 1990;9:115–24.
12. Kawado M, Hinotsu S, Matasuyama Y, Yamaguchi T, Hashimoto S, Ohashi Y. A comparison of error detection rates between the reading aloud method and the double data entry method. *Control Clin Trials* 2003;24:560–9.