Research Data Collection Methods
From Paper to Tablet Computers

Adam B. Wilcox, PhD, Kathleen D. Gallagher, MPH,* Bernadette Boden-Albala, DrPH,†
and Suzanne R. Bakken, RN, DNSc**

Background: Primary data collection is a critical activity in clinical research. Even with significant advances in technical capabilities, clear benefits of use, and even user preferences for using electronic systems for collecting primary data, paper-based data collection is still common in clinical research settings. However, with recent developments in both clinical research and tablet computer technology, the comparative advantages and disadvantages of data collection methods should be determined.

Objective: To describe case studies using multiple methods of data collection, including next-generation tablets, and consider their various advantages and disadvantages.

Materials and Methods: We reviewed 5 modern case studies using primary data collection, using methods ranging from paper to next-generation tablet computers. We performed semistructured telephone interviews with each project, which considered factors relevant to data collection. We address specific issues with workflow, implementation and security for these different methods, and identify differences in implementation that led to different technology considerations for each case study.

Results and Discussion: There remain multiple methods for primary data collection, each with its own strengths and weaknesses. Two recent methods are electronic health record templates and next-generation tablet computers. Electronic health record templates can link data directly to medical records, but are notably difficult to use. Current tablet computers are substantially different from previous technologies with regard to user familiarity and software cost. The use of cloud-based storage for tablet computers, however, creates a specific challenge for clinical research that must be considered but can be overcome.

Key Words: primary data collection, tablet computers, iPad, clinical research

(*Med Care* 2012;50: S68–S73)

Primary data collection is a principal component of the clinical research process. Unless analysis is performed only through secondary use of existing data, at some point all clinical research studies will require primary data collection. For years, researchers have sought and used electronic data systems for primary data as electronic systems have demonstrated benefits over paper-based methods. These benefits have been most clearly shown in the storage and management of data.[1–3] Sufficient benefits in data collection have lagged because data collection involves a specific interactive workflow where electronic devices can become cumbersome. However, recent changes in consumer electronic devices, both in functionality and portability, have increased the potential utility of mobile technologies for research data collection.[2,4,5] In this paper, we discuss these changes and their potential impact on the clinical research process, including specific case studies highlighting their use.

BACKGROUND AND SIGNIFICANCE

Data collection is critical to clinical research, and often is a prominent factor in determining the cost and success of a research project. How data are collected has a sizeable impact on how data are managed, and ultimately how the research is performed. Many technologies exist for data collection, ranging from simple paper forms to portable electronic devices. As yet, no data collection method is perfect and each has its own benefits, costs, and risks. A challenge for researchers is matching the capabilities of the different data collection methods to the data collection priorities of the research project.

Shapiro et al[6] discussed issues related to research data collection, and identified factors to consider regarding different data collection methods.[6] They addressed these factors with both paper-based and computer-based data collection methods. Other researchers have also compared paper-based data collection with computer-based approaches; a literature review of controlled trials comparing handheld computers with paper methods found improvements in storage, management, and collection of data, and computers were preferred by users.[1] Although the use of computers for clinical research is increasing, paper-based methods still remain common for clinical research data collection, because of remaining advantages of paper over computer-based approaches.[6,7] Paper is especially useful because of its simplicity, with low initial costs for implementation, expertise, support, equipment, and training. In certain implementations and settings, these benefits

can far outweigh the advantages of immediate feedback and monitoring, incorporated logic, and decreased duplicate documentation inherent in computer-based approaches.

Recent developments have made considerations of electronic data collection tools for clinical research more compelling. First, the value of electronic data collection is increasing. Prospective research studies leveraging data from electronic health records (EHRs) have intensified the need for data collected in electronic form, so they can be merged with other electronic data sources.[8–10] Translational research initiatives also require data that can easily be integrated with EHRs. The use of multicenter trials is expanding, and they strongly benefit from electronic data systems that manage the collection and transfer of data. A second significant development is the introduction of a new generation of mobile computing devices, including smartphones and tablets. Although handheld devices and tablet PCs have been used for years, current devices have shown greater adoption, such that the end-user needs for training and the disadvantages to computerized data collection may be significantly reduced. These next-generation tablet devices already represent over 90% of tablet computer use,[11] with estimates of overall consumer use as high as 13%.[12–16] These tablets also have redesigned operating systems focused on smaller devices that use small distributed applications, or "apps," and increasingly leverage cloud-based storage. These 2 differences, the high acceptance rate and the small apps using cloud storage, represent a significant difference in design and potential implementation from previous tablet technology. To date, while researchers have suggested ways newer tablets could be used[17–21] studies have not addressed the important differences in data storage approaches *and* connectivity, *or even* the combination of functions (eg, capture of still and video images in addition to textual and numeric data) of these next-generation tablets that are relevant to primary data collection in clinical research. Although Shapiro et al[6] studied handheld computer forms, these next-generation tablets differ significantly from the evaluated handheld devices in the factors considered.

In this paper, we review 5 projects as case studies using different forms of primary research data collection. Each project uses data extracted from EHRs in addition to primary research data collection. With these projects we examine the different methods of data collection, using the factors considered by Shapiro et al,[6] and specifically consider methods most affected by recent developments in clinical research.

METHODS

The 5 case studies are the Comparative Outcomes Management with Electronic Data Technology (COMET) Study at Stanford University, the Indiana PROSPECT project as part of the Indiana Network for Patient Care (INPC), the Pediatric Enhanced Registry project at Cincinnati Children's Hospital Medical Center, the Scalable Architecture for Federated Translational Inquiries Network (SAFTINet) project at the University of Colorado School of Medicine, and the Washington Heights/Inwood Infrastructure for Comparative Effectiveness Research (WICER) project at Columbia University. Each project was recently funded through the Agency for Healthcare Research and Quality as part of its investment to build infrastructure to conduct comparative effectiveness research with electronic and prospective data. The projects therefore represented research studies that required primary data collection, had high value for electronic data, and are recent enough to have considered newer technologies for data collection. In addition, each project collected data from multiple sites or research centers.

For each project, we performed a semistructured interview with a project representative (either the principal investigator or the team expert identified by the principal investigator), specifically related to their primary data collection tasks. The goal of the interview was to create a general qualitative description of their project, while concurrently identifying specific factors relevant to data collection. After acquiring a general description of the primary data collection process and the method used, we asked specific questions about the method with regard to workflow, connectivity, security, and data integration (Fig. 1). Interviews lasted approximately 30–45 minutes. Using the results of the interviews, we then assessed each data collection method using the factors considered by Shapiro et al,[6] specifically ease of use, the experience required to develop the data collection forms, the end-user training needed to use the forms, instrument and distribution cost, instrument flexibility, speed of data entry, accuracy of data entry, potential for data loss, need for technical support, and hardware/software requirements. For each method, we assessed whether that factor was a strength of the method or a weakness, as noted by the project representatives for their specific use case.

RESULTS

Table 1 shows the general characteristics of a primary data collection activity for each project. In each case, data collection was done by a research coordinator, rather than directly by subjects or patients. A description of each project and the context of the primary data collection activity is given below.

The COMET Study at Stanford University is developing an electronic network infrastructure to collect and link prospective data from multiple clinical centers and from various patient and research participant populations. Specifically they are using the network to integrate data from clinical and research centers at 4 academic institutions across the United States to support cross-institution analysis.[22] The COMET project mainly uses web-based forms for primary data collection, using standard desktop or laptop computers during an interview performed with each research participant. Occasionally data are collected using paper-based forms, and then entered directly into the web-based form. A main advantage of the forms is data collection validation, to ensure that all questions were asked and answered appropriately. Other advantages were that the researchers can do rapid quality assurance on the data capture process, because the data are entered into electronic form during or immediately after collection.

- Why did you choose the specific technology that you used?
- How did the data collection approach fit with the workflow of the individuals collecting the data?
- What were the specific benefits you saw over other approaches?
- What was the setting (research site, clinical setting, field) of the interview?
- If an electronic device was used, what was the connectivity required during data collection?
- If paper was used, how did the information eventually become stored as electronic information?
- How was data security maintained?
- Who actually documented the data (research coordinator or subject)?
- How were the data integrated with other electronic research data?

FIGURE 1. Interview questions. The first three questions focused on factors leading to the technology choice, while the other questions considered its effects.

The Indiana PROSPECT project expands the INPC by capturing more extensive clinical data including patient outcomes, patient study enrollment, and genomic information.[23] INPC is a state-wide health information exchange network that contains structured and text data for approximately 12 million patients. The Indiana PROSPECT project uses scannable forms and a barcode scanner tool (caTrack) as part of data collection for biological samples to enhance the data already in the INPC. When a sample is drawn, information about subjects and the biological sample are documented in structured form in paper forms that are later scanned by a computer to extract the data. The paper forms also include barcodes so they can be automatically linked with the biological sample. The goal of the scannable forms and barcode scanner was to link the subject identifier to the sample at multiple points of data processing to avoid human errors. Samples are first scanned when blood is drawn, then again when the samples arrive to the lab, and then when the sample is spun and scanned. Data are then transferred from the device to the central server, where data are integrated.

The Pediatric Enhanced Registry project at Cincinnati Children's Hospital Medical Center creates a patient research registry that links to EHRs in different types of delivery sites.[24] The Pediatric Enhanced Registry involves 30 geographic sites across the United States that are all populating data into the registry. Investigators are using the registry for studies demonstrating the impact of a quality improvement initiative using both the EHR data and registry. Primary data collection for the evaluation is mainly in qualitative interviews, with some questionnaires. This project uses paper forms and audio recording devices for data collection, and information is entered directly or transcribed. Paper was preferred because of the low costs for initial creation of forms, and no technology was seen as flexible enough to fit the workflow of the qualitative interview process. Paper was also portable, which was necessary as interviews were done at the most convenient location for the subject (usually a clinical site or research center, or home). Data collected from the qualitative interviews on paper forms are eventually transferred to electronic form during the analysis stage.

The SAFTINet project at the University of Colorado School of Medicine is creating a distributed health data network to support CER and quality improvement efforts targeted at safety net populations.[25] SAFTINet has created a research partnership with organizations throughout the Intermountain West and United States with diverse clinical and administrative data, and uses this partnership and data to create a learning community for research in safety net populations. Much of the data are from existing clinical sources, using inpatient, outpatient and claims data for research purposes. SAFTINet also collects primary data for patient surveys regarding their current disease status (specifically, asthma control). As most of these surveys are connected to the clinical encounter, a template was created that can be implemented within an EHR, so that data can be entered and automatically connected to the patient/subject. This approach is similar to a web-based form, except that subject and user context is already established through the EHR. Some sites use paper forms for initial data collection and data are then retrospectively entered into the EHR, whereas other sites document directly into the EHR template.

TABLE 1. General Characteristics of Primary Data Collection for Each Case Study

	COMET	Indiana PROSPECT	Pediatric Enhanced Registry	SAFTINet	WICER
Purpose	Case report forms during clinical visits	Collecting information on tissue samples that will be linked to clinical data	Qualitative interviews with patients and family members	Administering validated instrument as part of a patient assessment	Survey about health-related issues for individuals within a community
Data collection site	Clinical research site; direct patient entry could be done at clinic or home	Clinical research site	Clinical site, research center, or patient's home	Clinical site	Patient's home
Data collection workflow	Collected and entered by research coordinator; pursuing direct patient data entry	Collected and entered by research coordinator, both for mobile devices and scannable forms	Collected and entered by research coordinator during qualitative interviews	Collected and entered by clinical research coordinator as part of clinical visit	Collected and entered by clinical research coordinator during interview
Data entry approach	Web-based forms during interview	Scannable paper forms (with attached barcodes)	Paper forms and audio recordings	Form template within EHR (similar to web-based form)	Tablet computer form template

COMET indicates Comparative Outcomes Management with Electronic Data Technology; EHR, electronic health record; SAFTINet, Scalable Architecture for Federated Translational Inquiries Network; WICER, Washington Heights/Inwood Infrastructure for Comparative Effectiveness Research.

The WICER project at Columbia University contains a research data warehouse that integrates patient-level data, including clinical data from multiple facilities, settings and sites of care, with person-level self-reported information collected through a community survey.[26] WICER uses primary data collection with the community survey, where most of the data are being collected in residents' homes by community health workers. Individual survey data are linked to data of others in the household, as well as linked with any available individual EHR data. Because of the scale and complexity of the community survey (8000 individuals in 3500 homes interviewed annually; ~200 discrete response questions with branching logic; 45–75 minutes per survey; 8–10 community health workers), a next-generation tablet computer (Apple iPad2) was selected to support community survey administration. As the survey administrators would be traveling to residents' homes, portability was especially important and the iPad was judged to be as portable as paper. Other advantages specific to the iPad were relatively modest requirements for user training and technical support, and the ability to purchase third-party software for survey administration and processing. In addition, using the camera or GPS in the device to collect data was being investigated, and was seen as a potential, but unrealized, advantage.

Table 2 contains a comparison of the different data collection approaches used in the projects, highlighting the advantages and disadvantages of the methods. For each factor, we assessed whether the data collection approach was a strength (+) or weakness (−) for the project. Paper is the simplest and most commonly used method, and also was best understood in terms of its strengths and weaknesses. The most significant benefits of paper were its simplicity in development and implementation, but it has disadvantages in obtaining and managing data. Scannable forms maintain some benefits of paper in ease of use, while mitigating some of its disadvantages. Web-based forms are the most flexible method, and forms can be easily adjusted with moderate effort by the form designer. However, they are not as easy to use as paper-based methods, and are generally not portable.

EHR form templates are the most secure with data, by linking the information within the clinical record and storing in the EHR database. However, they must be developed and used within the constraints of an EHR, and EHRs are not generally lauded for their ease of use. Although previous versions of tablet computers were difficult to use and required the use of specialized software with significant developer and end-user training, next-generation, consumer-focused tablet computers are much easier to use and have less disadvantages in end-user experience or available software. They have the potential to gain most of the advantages of direct electronic data entry without incurring the user experience disadvantages normally associated with new technologies.

DISCUSSION

A review of 5 modern research projects has demonstrated that there remain multiple methods of primary data collection, with varying advantages and disadvantages. Paper remains the easiest to use, and in studies with a small number of subjects that require complex data collection, it remains a preferred method. Other common methods attempt to reduce the disadvantages of paper (scannable forms) or leverage the benefits of computers (web-based forms). Recent developments in the need to integrate EHR data or the emergence of consumer-focused devices have produced additional data collection strategies (EHR templates, tablet computers).

Although the main data entry method with the project was the focus of the case study and showed the breadth of approaches, the researchers have also pursuing other data collection approaches. For example, COMET is pursuing direct patient entry for a specific questionnaire, which could be entered either online or at a clinical site with kiosk machines. However, some data collection was seen as not amenable to direct patient entry, especially for the initial interviews. The Pediatric Enhanced Registry project reviewed mobile tools that could integrate notes, patient consent, and audio recordings. These were seen as potentially useful, but the researchers felt during initial implementation

TABLE 2. Comparison of Data Collection Approaches

Category	Paper Forms (Pediatric Enhanced Registry)	Scannable Forms (Indiana PROSPECT)	Web-based Forms (COMET)	Form Template in EHR (SAFTINet)	Tablet Computer (WICER)
Ease of use	+	+	—	—	+
Experience required of designer	+	+	—	—	—
End user training	+	+	—	—	—
Cost	+	—	—	—	—
Flexibility	—	—	+	—	+
Speed of entry	—	—	—	—	—
Accuracy/error rate	—	—	+	—	+
Potential for data loss	—	—	—	+	—
Need for technical support	+	+	—	—	—
Equipment/software requirements	+	—	—	—	—

COMET indicates Comparative Outcomes Management with Electronic Data Technology; SAFTINet, Scalable Architecture for Federated Translational Inquiries Network; WICER, Washington Heights/Inwood Infrastructure for Comparative Effectiveness Research; +, strengths for the data collection method; −, weaknesses.

that no technology was sufficiently capable or robust to justify replacing paper. With the SAFTINet project, initially considered technologies such as kiosks and handheld devices, but prior experience through other projects identified cost barriers to their implementation.

Next-generation tablet computers are particularly interesting, because they most closely achieve the experience of paper-based methods while remaining computer-based. These devices leverage portable applications and smaller-device operating systems, and can potentially introduce new opportunities in clinical research settings. They also introduce challenges, especially around data security and connectivity with cloud-based storage where the collected data are stored and managed by an external company, and outside the direct control of the researchers. With the WICER project, this was a significant barrier, because the research data included protected health information and could not be stored outside our institutional control—control of the data is both an institutional requirement and regulatory recommendation for research data.[27] We addressed this challenge with a data encryption process controlled by the software. Data were encrypted on the tablet computer before transmission to the cloud server, and data were only decrypted after being retrieved by the research institution. To address connectivity, we used a local application on the device rather than using the tablet computer as a browser for web-based forms. This reduced workflow issues because of connectivity, but also carried the risk of storing information on the device. However, this risk was similar to the risk of losing paper forms. In addition, the device was password protected, and the application had functionality to automatically delete data if the device was identified as missing or stolen.

Although effective mobile technologies are emerging for use in research data collection, the technologies are still nascent, paper-based methods are still preferred in many instances, and few implementations successfully achieve the benefits of electronic data collection while also minimizing its disadvantages. With the WICER study, the number of surveys to be administered by community health workers was sufficiently large that electronic systems were seen as preferable, but previous generation electronic tools were considered too difficult, complicated, and expensive to replace paper. The lower cost of hardware, software, and data transfer for next-generation tablets made them viable. Issues around portability, user training and experience, and ease of use that have made paper preferred over electronic tools were significantly reduced with this newer hardware.

However, there are areas where the device portability would not achieve the same advantages, and even be a disadvantage. Most data entry was done by research coordinators, and in only one instance the patients entered data directly. Direct patient data entry has different considerations, and a mobile solution would likely not be appropriate because of device control. Theft risk was a specific consideration against pursuing tablets for the SAFTINet project. In addition, having no or minimal training required is even more critical for patient entry than coordinator entry. This usually requires purpose-specific rather than consumer-focused applications.[8] As a result, patient data entry approaches usually have more complicated software developed, that is more expansive than just questions and answers (eg, including clarifying instructions to help users understand and answer questions).[9]

Another potential benefit of next-generation tablets, beyond user familiarity and software costs, is that the tools can combine functions. Many tablets include capabilities for data entry, audio recorders, cameras, and global positioning system receivers. These are all capabilities that have been used for data collection in some form in research studies. By combining the functions together, research coordinators could more easily collect disparate data that could be internally linked. The WICER project links some data with geographic information systems, and the Indiana PROSPECT considered using smartphones for barcode scanning. Although the capability to combine tools is a new benefit without many demonstrated examples, it should be considered a potential advantage.

There were limitations to our case study approach. Although controlled trials and even systematic reviews have been performed studying previous generation tablets,[1,2,4] the use of the next-generation tablets is early. Case studies may be more informative to define the important characteristics for a controlled trial. Another limitation is that each project is in the initial stages of data collection, and the final outcome of each approach was not measured. It will be important to note whether the findings at implementation are still valid after the studies are completed, and to consider if there are differences why they may have occurred. However, there is a need for literature describing main considerations in implementing the newer technologies. With the expanding growth and interest in tablet computers, there is an expected high interest in using these tablets for clinical research studies.

CONCLUSIONS

These case studies demonstrated various approaches for data collection, leveraging technology ranging from paper to current EHRs to next-generation tablet computers. Next-generation tablet computers, which have not been considered in previous studies of data collection methods, most closely approach the ease of use simplicity of paper, while gaining the benefits of computer-based approaches. Although they can introduce risks with data security and connectivity, these risks have been successfully mitigated. As more experience in these tools is gained, even greater benefits from combining device functions may also be significant.

REFERENCES

1. Lane SJ, Heddle NM, Arnold E, et al. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. *BMC Med Inform Decis Mak.* 2006;6:23. 10 pages.
2. Cole E, Pisano ED, Clary GJ, et al. A comparative study of mobile electronic data entry systems for clinical trials data collection. *Int J Med Inform.* 2006;75(10–11):722–729.
3. Curioso WH, Mechael PN. Enhancing 'M-health' with south-to-south collaborations. *Health Aff (Millwood).* 2010;29:264–267.
4. Haller G, Haller DM, Courvoisier DS, et al. Handheld vs. laptop computers for electronic data collection in clinical research: a crossover randomized trial. *J Am Med Inform Assoc.* 2009;16:651–659.
5. Pundt H. Field data collection with mobile GIS: dependencies between semantics and data quality. *Geoinformatica.* 2002;6:363–380.

6. Shapiro JS, Bessette MJ, Baumlin KM, et al. Automating research data collection. *Acad Emerg Med*. 2004;11:1223–1228.

7. Shelby-James TM, Abernethy AP, McAlindon A, et al. Handheld computers for data entry: high tech has its problems too. *Trials*. 2007;8:5. 2 pages.

8. Libby AM, Pace W, Bryan C, et al. Comparative effectiveness research in DARTNet primary care practices: point of care data collection on hypoglycemia and over-the-counter and herbal use among patients diagnosed with diabetes. *Med Care*. 2010;48(suppl):S39–S44.

9. Abernethy AP, Ahmad A, Zafar SY, et al. Electronic patient-reported data capture as a foundation of rapid learning cancer care. *Med Care*. 2010;48(suppl):S32–S38.

10. Bernstam EV, Hersh WR, Johnson SB, et al. Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical andTranslational Science Award programs. *Acad Med*. 2009;84:964–970.

11. MacNews. Nielsen: iPad has 82% of the tablet market. 2011. Available at: http://www.macnews.com/2011/05/05/nielsen-ipad-has-82-tablet-market. Accessed August 31, 2011.

12. DiscountCoder.com. A brief history of the iPad. 2011. Available at: http://www.discountcoder.com/images/infographics/iPad_infographic.jpg. Accessed August 31, 2011.

13. Brown B. Gartner: nearly 20 million tablets to be sold in 2010. 2010. Available at: http://www.networkworld.com/news/2010/101510-gartner-tablets-ipad.html. Accessed August 31, 2011.

14. Elmer-DeWitt P. How many tablets in 2011? 2010. Available at: http://tech.fortune.cnn.com/2010/11/27/how-many-tablets-in-2011/. Accessed September 2, 2011.

15. Keizer G. iPad lead shrinks as tablet sales grow. 2011. Available at: http://www.computerworld.com/s/article/9207319/iPad_lead_shrinks_as_tablet_sales_grow. Accessed September 2, 2011.

16. Long M. Tablet mania: 50 million to ship in 2011. 2011. Available at: http://www.tabletsmania.com/technology-news/tablet-mania-50-million-to-ship-in-2011/. Accessed September 2, 2011.

17. Choi JS, Yi B, Park JH, et al. The uses of the smartphone for doctors: an empirical study from samsung medical center. *Healthc Inform Res*. 2011;17:131–138.

18. Kiser K. 25 ways to use your smartphone. Physicians share their favorite uses and apps. *Minn Med*. 2011;94:22–29.

19. Berger E. The iPad: gadget or medical godsend? *Ann Emerg Med*. 2010;56:A21–A22.

20. Mashman W. The iPad in cardiology: tool or toy? *JACC Cardiovasc Interv*. 2011;4:258–259.

21. Putzer GJ, Park Y. The effects of innovation factors on smartphone adoption among nurses in community hospitals. *Perspect Health Inf Manag*. 2010;7:1b. 20 pages.

22. Kushida CA. Comparative Outcomes Management with Electronic Data Technology (COMET) Study. 2010. Available at: http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=R01+HS19738-01. Accessed June 8, 2011.

23. Overhage JM. Indiana PROSPECT. 2010. Available at: http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=R01+HS19818-01. Accessed June 8, 2011.

24. Hutton JJ. Building modular pediatric chronic disease registries for QI and CE research. 2010. Available at: http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=R01+HS20024-01. Accessed June 8, 2011.

25. Schilling LM. Scalable Architecture for Federated Translational Inquiries Network (SAFTINet). 2010. Available at: http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=R01+HS19908-01. Accessed June 8, 2011.

26. Wilcox AB. Washington heights initiative community-based comparative effectiveness research. 2010. Available at: http://gold.ahrq.gov/projectsearch/grant_summary.jsp?grant=R01+HS19853-01. Accessed June 8, 2011.

27. Food and Drug Administration (2010). Guidance for Industry: Electronic Source Documentation in Clinical Investigations—Draft Guidance. Food and Drug Administration US Department of Health and Human Services.