2-13-2013

Security Approaches in Using Tablet Computers for Primary Data Collection in Clinical Research

Adam B. Wilcox
Columbia University, wilcox@dbmi.columbia.edu

Kathleen Gallagher
Columbia University, kathleen.riordan@dbmi.columbia.edu

Suzanne Bakken
Columbia University, suzanne.bakken@dbmi.columbia.edu

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Security Approaches in Using Tablet Computers for Primary Data Collection in Clinical Research

Abstract
Next-generation tablets (iPads and Android tablets) may potentially improve the collection and management of clinical research data. The widespread adoption of tablets, coupled with decreased software and hardware costs, has led to increased consideration of tablets for primary research data collection. When using tablets for the Washington Heights/Inwood Infrastructure for Comparative Effectiveness Research (WICER) project, we found that the devices give rise to inherent security issues associated with the potential use of cloud-based data storage approaches. This paper identifies and describes major security considerations for primary data collection with tablets; proposes a set of architectural strategies for implementing data collection forms with tablet computers; and discusses the security, cost, and workflow of each strategy. The paper briefly reviews the strategies with respect to their implementation for three primary data collection activities for the WICER project.

Acknowledgements
Funding was provided by R01 HS019853 from the Agency for Healthcare Research and Quality; Title: Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research (WICER).

Keywords
informatics, security, primary data collection, tablet computers, iPad, clinical research, WICER

Disciplines
Health Services Research

Security Approaches in Using Tablet Computers for Primary Data Collection in Clinical Research

Adam B. Wilcox, PhD; Kathleen Gallagher, MPH; Suzanne Bakken, RN, DNSc[i]

Abstract

Next-generation tablets (iPads and Android tablets) may potentially improve the collection and management of clinical research data. The widespread adoption of tablets, coupled with decreased software and hardware costs, has led to increased consideration of tablets for primary research data collection. When using tablets for the Washington Heights/Inwood Infrastructure for Comparative Effectiveness Research (WICER) project, we found that the devices give rise to inherent security issues associated with the potential use of cloud-based data storage approaches. This paper identifies and describes major security considerations for primary data collection with tablets; proposes a set of architectural strategies for implementing data collection forms with tablet computers; and discusses the security, cost, and workflow of each strategy. The paper briefly reviews the strategies with respect to their implementation for three primary data collection activities for the WICER project.

Introduction

Primary data collection, whereby researchers collect information on subjects for purposes of analysis within a study, remains a core and necessary requirement for many clinical research studies, even with the expanded secondary use of electronic health information for research. The costs of collecting and managing research data can be significant. A study by Emanuel et al. on the costs of conducting clinical research found that the two most expensive components of a trial are the conduct of research visits with subjects and data management and analysis, both of which require extensive work with data.[1] To be managed effectively, data need to be in electronic form, leading many investigators to seek ways to collect data through electronic systems, including handheld tablet computers used by research data collection staff. Studies have documented improvements in data management with handheld computers and that users of the devices often prefer them to paper methods.[2–4] However, handheld data collection technology still poses challenges in fitting the research data collection workflow such that paper-based data collection remains common even when the costs of data transcription are substantial.

Recent changes in tablet computer technology have been significant and can improve direct electronic data collection in clinical research studies. These changes are the result of next-generation tablet devices, such as iPads and Android-based tablets, which were introduced in 2010 and already account for over 95 percent of the tablet technology market.[5] Not only are these tablets less costly than previous tablet computers, but they also support an application distribution model that has led to far more applications available at a much lower cost. The devices also demonstrate much higher adoption rates; in 2011, 11 percent of consumers owned a tablet, with 47 percent of consumers expected to own one by 2013.[6] Lower costs and greater consumer experience with tablets have overcome some of the main challenges to their use in clinical research.

However, significant challenges remain in using next-generation tablets for clinical research, especially as related to data security. Security issues such as local data storage, data transmission, and server data storage must be addressed when tablets (or any computer) rather than paper-based methods are used for primary data collection. The issues are complicated by the fact that tablet applications generally host data "in the cloud" or on servers owned by software developers and outside the direct control of a research institution. Research institutions that want to use tablet applications when exchanging protected health information (PHI)—which includes most data collected for clinical research—need to have a business associates agreement (BAA) in place. BAAs impose obligations and penalties for data security breaches on software developers similar to those imposed on research institutions. With increasing penalties for data security breaches of electronic health information, software companies have generally been unwilling to assume the full risks of PHI, especially when the software is not explicitly designed for health care applications. As a result, it is important to consider strategies for primary data collection with next-generation tablets that maximize their potential and demonstrate to institutional review boards and privacy boards that these strategies can effectively limit the exchange of protected health information.

[i]Columbia University

We have direct experience in navigating security issues that involve the use of tablets for primary data collection in the Washington Heights/Inwood Infrastructure for Comparative Effectiveness Research (WICER) project at Columbia University in New York City.[7] The project's overall goal is to advance comparative effectiveness research designed to improve hypertension care delivery and population outcomes. The WICER project uses an existing institution-focused data infrastructure to create a robust community-focused data infrastructure that will support the types of innovative studies needed for effectively tackling seemingly intractable public health problems. WICER contains a research data warehouse that integrates patient-level data, including clinical data from several facilities, settings, and sites of care, with person-level self-reported information. In addition, the project will map the linked data to variables that support prospective comparative effectiveness research studies. The most significant parts of the WICER project using primary data collection are its surveys of the community and its primary data collection from patients participating in a care management comparative effectiveness study.

Even though guidelines and regulations already govern the management of data security, practical details are lacking for implementing data security with the use of next-generation tablets for primary data collection. Without clear examples of how new technologies can satisfy data security guidelines, researchers may choose more established technologies that are less effective but more standard in their implementation. In this paper, we describe considerations and approaches for effectively using tablets when PHI is collected for research. We also describe applications within the WICER project.

Methods

To assess feasible security strategies for the use of next-generation tablets for collecting community-level health data, the WICER investigators applied a set of security considerations to the project's goals, as described below. The project reviewed and evaluated four alternative security approaches with respect to a set of considerations, including cost (assuming a three- to five-year project) and impact on data collection workflow.

Specific Application: WICER

Longitudinal community surveys in the WICER project are collected at various locations for over 8,000 individuals: in a community outreach center, in individuals' homes, and in an ambulatory care clinic. The care management study recruits 50 patients into a care management intervention for treating uncontrolled hypertension and, as with the community surveys, collects information directly from the patient. Survey administration takes approximately 20 to 45 minutes per individual, with the survey administered directly by a research coordinator.

For each instance of primary data collection, we evaluated how tablet computers could be used in lieu of paper-based data collection, considering each architectural approach as a potential application. Once we defined an approach, we applied for approval to conduct the study with the Columbia University Institutional Review Board (IRB).

Defining the Desirable Security Attributes of Primacy Data Collection Systems that Preserve PHI

For primary research data collection, three significant data security issues must be addressed when using tablets or any computer versus paper-based methods. The issues necessitate strategies to preserve the privacy of PHI at three stages of work with primary data: (1) local data collection and storage; (2) data transmission; and (3) central server data storage for analysis. The following sections outline major privacy-preserving considerations for primary data collection and storage.

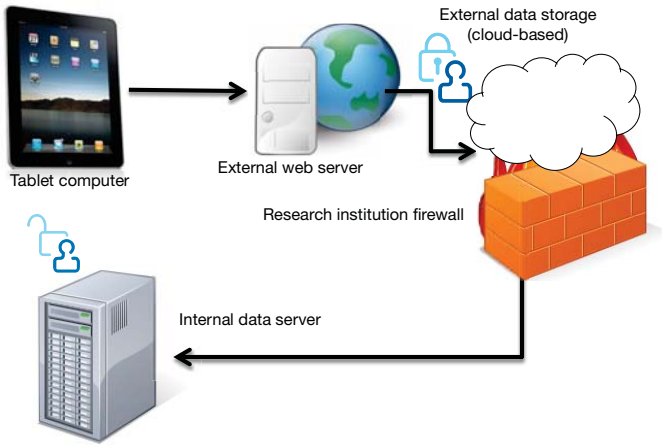
Local data are the data that are stored on an electronic device. They are at risk if a device is lost or stolen. Local data storage may be enhanced and security risks mitigated by using passwords to protect access to the device applications or even by encrypting the device storage, making the data more secure than if paper-based data were lost or stolen. Still, electronic devices store much more data than is reasonably captured on paper, thereby increasing the potential damage associated with data loss. However, to limit the data stored on a device, the data may be periodically uploaded to central servers and deleted from the tablets. In general, data stored on and periodically uploaded from tablet computers (and then deleted from the tablets) are more secure than paper-based approaches because the data are held locally for a shorter period.

Data transmission occurs when data stored on a tablet computer are transferred to a central data server. In the process of transmission, data may be vulnerable to attack. Unlike the case of the transmission of paper forms, which is performed manually in a single location, data transmission from devices generally occurs over data networks that are accessible to individuals throughout the world. Therefore, all PHI transmitted from tablets should be encrypted, especially if the data are transmitted outside an institutional firewall. Fortunately, data encryption approaches are advanced and have found widespread adoption in many sectors and fields. The risk of data loss during transmission with the use of standard secured networks or when data are encrypted while "at rest" is minimal and lower than the risk associated with physical transmission and storage.

Server data storage refers to the ultimate storage location for the data; it is where the data are usually gathered for analysis. While data may be stored in encrypted form when not being used, data used for analysis must at some point be decrypted and therefore are often stored or made accessible in unencrypted form. The cost of unencrypted data breaches in servers is significantly higher than a breach for a single device because all the data are stored at the server. Servers are also accessible via networks and may potentially be accessed by an individual at any location, whereas paper-based data are stored in a single location. Even with paper-based surveys, though, data are often transcribed into electronic form and stored in databases managed by servers, thereby

making the risk equivalent. However, if the server is controlled by a separate organization (e.g., the "cloud"), a business associate's agreement is required, which is difficult to execute.

Figure 1. Web-based data collection with an external hosting service that encrypts data to store it



Notes: Data are not stored outside the institution in unencrypted form—data can be decrypted once the data are downloaded to a server within the institution's firewall.

Given WICER's goals to maximize the security of local data as well as that of data transmission and server data storage, the following section describes a set of alternative approaches for achieving secure primary data collection with next-generation tablets, along with the inherent tradeoffs in each approach.

Results

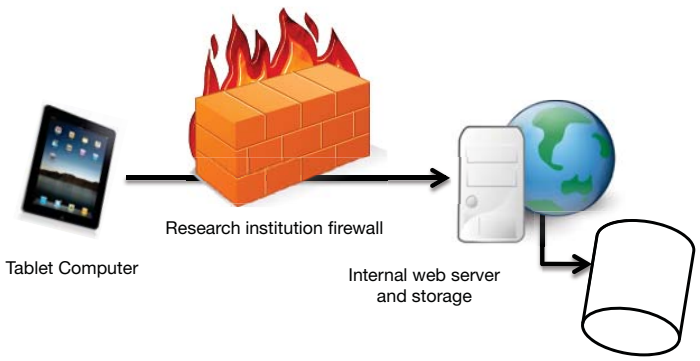
The first and perhaps simplest approach to data collection efforts that rely on tablet computers is to use the tablet to access a network-connected web-based data collection form. As shown in Figure 1, web-based data collection forms may be hosted by an external company (e.g., Survey Monkey or Survey Gizmo) or hosted within an institution by using available survey software (e.g., RedCAP or Lime Survey). External hosting is the less expensive and more common option and allows for some data analysis or reporting from the hosting application. However, external hosting places the data outside the control of the research institution and is therefore unsuited to PHI without a business associate's agreement. Some external hosting services store data in an encrypted form that may be decrypted only by a research institution; such an arrangement would be permissible for PHI, though data analysis or reporting tools from the hosting application would not be available (Figure 1). We found external hosting costs to be less than $500 per year.

Internal hosting is a second alternative (Figure 2). Data may be secured within an institution and its firewall, provided that the network connection between the tablet and server is secure. Hosting web forms internally is more expensive than external hosting, mainly because it requires the institution to manage the application. Free or low-cost software is available, but internal maintenance costs are significant.

The use of web-based data collection forms either for external hosting (Figure 1) or internal hosting (Figure 2) requires the data to be collected with a network connection: wired, wireless local area networks (WiFi), or mobile broadband (3G or 4G). This approach may work for data collected at a research site but may be difficult when data are collected in the field, where network connectivity is not as robust. Mobile broadband service has increased the availability of network connectivity in the field but can still be subject to connectivity issues and "dead spots" beyond researchers' control. Some solutions under development include local storage of data collection forms for uploading after a network connection is re-established.[8]

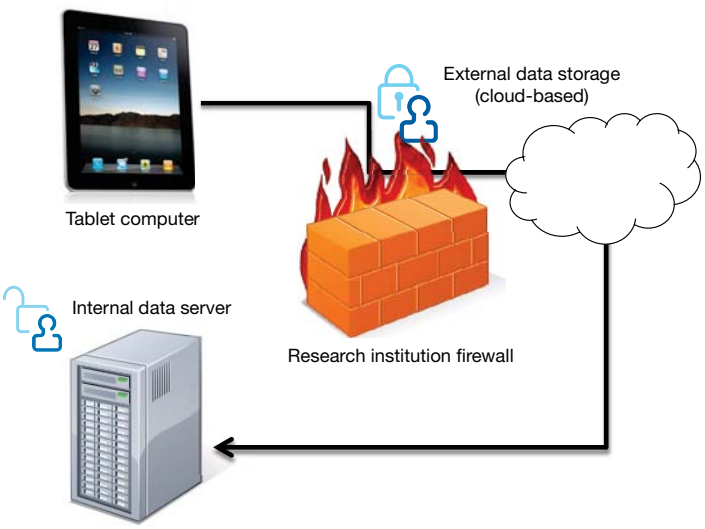
A third option, which reduces dependence on an active network connection, uses the capability of the tablet application infrastructure rather than relying on the tablet as a portal to web-based forms (Figure 3). Tablet applications, or apps, may provide some functionality beyond that offered by web-based survey software and may be easier to use for data collection. For example, if web-based applications require users to authenticate directly over the web, such a requirement may interfere with the data collection workflow. Apps, however, may be opened directly and may even link to other information collected by the tablet, such as location (from GPS) or pictures. The security issues with apps are similar to those with externally hosted web-based surveys, except that the data are expected to be stored on the device such that the device data must be secured. Externally hosted data are stored outside the institution by the application and under the control of the app software company. Therefore, a software requirement of the app is that data must be encrypted before uploading and may not be decrypted except by the institution. While most apps that collect survey information do not have encryption sufficient to handle PHI, some currently do;[9] in fact, the capability becomes increasingly available. Apps can be inexpensive; the apps for the devices may be free, but the form creation and hosting services often cost from $1,000 to $5,000.

Figure 2. Web-based data collection with internal hosting



Notes: A web-server is hosted by the institution, either within the firewall, or that stores data within the firewall. Data are not stored outside the institution and never are outside the institution's control.

Figure 3. Tablet app-based data collection



Notes: Data are stored in the cloud under the control of the software company, and therefore must first be encrypted on the device. The data can be downloaded to an internal server where it is decrypted.

A fourth option, which some institutions have implemented, is to use tablets and apps to collect data and host the applications locally. In this case, app software companies maintain a data server within the institution's firewall and under the institution's control (Figure 4). The benefit of such an approach is that the application's full functionality, including the data aggregation capabilities performed by the data server, is available and may include editing or retrieving data from the server that were previously uploaded from the tablets—a function not available if the data are encrypted at the server. The disadvantage is the increased costs – if this option is even available at all. Licensing with internal servers may cost 10 times more (e.g., $50,000) than hosting directly with the software company.

WICER Case Example

In the WICER project, we considered the four potential security strategies discussed in the previous section with respect to three instances of primary data collection needed for the WICER study: (1) community outreach surveys administered in the home; (2) community surveys administered in ambulatory care clinics; and (3) research surveys for the care management study, administered in the patient's home. Each survey type is considered separately, though the community outreach surveys were administered on paper. The community outreach survey was the first survey to be administered and was completed before a tablet computer solution could be defined.

For the community outreach survey administered in the home, the primary consideration was that several individuals would be surveyed within a single household, requiring the presence of research staff in one household for a considerable number of hours. With connectivity through mobile broadband networks variable within the apartment buildings in the community, we concluded that we could not rely on continuous connectivity. We therefore planned to use tablet app-based data collection instead of a web-based approach in order to minimize potential information loss while preserving privacy.

We found that the use of tablet apps (discussed third in the prior section) was the most feasible (Figure 3) approach. We identified a commercial software application that used public-private key encryption to encrypt data from the tablet before uploading to the vendor's site and then decrypted the data after downloading to our servers. Using this approach, which was approved by our IRB, we demonstrated successful data collection, transmission, and storage of the survey data. Eventually, we were not able to use the applications because the survey was much larger than the application was designed to support. In addition, we were not able to store enough versions of the survey on one tablet within the application before the data had to be uploaded. We saw the problem as a temporary capability limitation and are pursuing the use of tablets with app-based storage for data collection for the second round of surveys, with a change to both the application and device capacity.

For the survey administered in the ambulatory clinics, we faced the same software and device limitations. However, with the surveys administered at a single location, we were able to consider a network-dependent solution. We tested connectivity and determined that the network coverage was sufficient either through WiFi or mobile broadband networks. We therefore elected to use the second option (Figure 2), a web-based approach with web servers installed in our institution. Our IRB approved the approach after we passed security scanning checks for our site and demonstrated that the data would be transmitted in encrypted form.

For the care management survey, we initially planned to use app-based data storage with the tablets. However, once we identified the software and device limitations and discovered the need for several changes to the survey during the testing stage, we eventually decided to rely on paper-based collection. As a practical matter, the number of surveys was small enough that a large up-front investment was not justified both to develop and implement the survey with tablets. However, our IRB approved both an app-based and web-based approach (Figures 3 and 2) upon initial consideration.

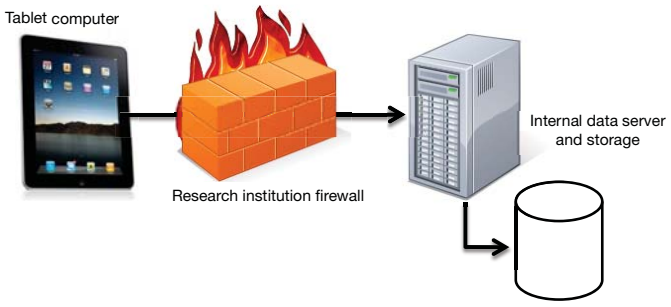
Discussion

Next-generation tablets offer great potential for improving the collection and management of clinical research data. However, innovations in data storage with tablet apps can create security issues. In this paper, we have described various architectural approaches to addressing security. The WICER project provided an incentive to decrease the overall costs of data entry and management with the use of tablets. To some degree, we investigated each option discussed and received IRB approval both for the web-based and app-based data collection.

Experience to date with the WICER project demonstrates that, in addition to implementing an appropriately secure data collection strategy, the project has benefited from the use of next-generation tablets for data collection. With respect to the web-based approach, we have been successful in implementing this strategy when network connectivity was robust. The app-based approach

was successful in meeting security requirements and in initial implementation as defined by the Columbia University Medical Center (CUMC) IRB; its limitations are largely a function of the survey's unusual length.

Figure 4. Tablet app-based data collection, with an internally hosted application server



Notes: An app software company installs an application server within an institution's firewall, and data remain under the control of the institution.

It is also useful to consider that one ongoing challenge to the emerging technology of tablets and apps is that some capabilities needed in survey software may not be available and may be particularly costly to develop. App development has grown far faster than other software development and distribution models, but it is still incomplete. The existing suite of applications does not meet all research data collection requirements. Fortunately, app development has grown substantially, and services have emerged for contracting with app developers for customized apps. The advantage of a customized app is that it can meet a research project's highly specific requirements. However, app development is far more expensive than the purchase of apps, where the costs of development are distributed over a large number of customers. While some studies have estimated the average app development cost at around $6,500,[10] research data collection apps with appropriate security are more complex. Our cost estimate, which is based on various sources,[11–13] is $25,000 to $50,000, depending on the complexity of the survey. With custom development, app data servers would be hosted locally such that the security diagram would be the same as Figure 4.

Despite the ongoing challenges, we have seen significant improvements over the course of the WICER project in the available options for using next-generation tablets. As a result, the WICER team is now actively developing its second-round surveys for use with tablets. Our experience implementing the discussed iPad security strategy for community-level data collection has made us optimistic about the potential to use tablets for future studies, and we anticipate that the acceptability of the current approach with the CUMC IRB may facilitate future studies using the same technology. Even though some approaches may be costly, we have found that the costs of data entry can exceed the costs of these approaches.[14] Therefore, we recommend that new research studies seriously consider the benefits of tablets.

Acknowledgements

Funding was provided by R01 HS019853 from the Agency for Healthcare Research and Quality; Title: Washington Heights/Inwood Informatics Infrastructure for Community-Centered Comparative Effectiveness Research (WICER).

References

1. Emanuel EJ et al. The Costs of Conducting Clinical Research. Journal of Clinical Oncology. 2003; 21(22): 4145-50.
2. Lane SJ 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.
3. Shapiro JS et al. Automating Research Data Collection. Acad Emerg Med. 2004; 11(11):1223-8.
4. Shelby-James TM et al. Handheld Computers for Data Entry: High Tech Has Its Problems Too. Trials. 2007; 8(5):2.
5. Strategy Analytics. Android Captures Record 39 Percent Share of Global Tablet Shipments in Q4 2011. http://www.strategyanalytics.com/default.aspx?mod=pressreleaseviewer&a0=5167. Accessed November 2012.
6. Lunden I. OPA: iOS and Android Level in U.S. Tablet Market, Penetration 47% by 2013, $2.6B Spent on Apps in 2012. http://techcrunch.com/2012/06/18/opa-ios-and-android-level-in-u-s-tablet-market-penetration-47-by-2013-2-6b-spent-on-apps-in-2012. Accessed November 2012.
7. Wilcox AB et al. Research Data Collection Methods: From Paper to Tablet Computers. Med Care. 2012; 50(Suppl):S68-78.
8. SurveyGizmo Offline Surveys. http://www.surveygizmo.com/survey-software-features/offline-surveys. Accessed January 2013.
9. iFormBuilder, Maximum Security. https://www.iformbuilder.com/features/maximum-security. Accessed November 2012.
10. Stetler M. How Much Does It Cost to Develop a Mobile App? http://appmuse.com/appmusing/how-much-does-it-cost-to-develop-a-mobile-app/. Accessed November 2012.
11. Ibid.
12. Impiger Mobile. How Much to Pay for My iPhone App Development? http://www.impigermobile.com/how-much-to-pay-for-my-iphone-app-development. Accessed January 2013.
13. PadGadget. The Cost of Building an iPad App. http://www.padgadget.com/2010/10/17/the-cost-of-building-an-ipad-app. Accessed January 2013.
14. Wilcox, op. cit.