Original Paper

How Strong are Passwords Used to Protect Personal Health Information in Clinical Trials?

Khaled El Emam[1,2], BEng PhD; Katherine Moreau[1], BA BEd MA; Elizabeth Jonker[1], BA(Hons)

[1]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada

[2]Department of Pediatrics, Faculty of Medicine, University Of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Khaled El Emam, BEng PhD
Children's Hospital of Eastern Ontario Research Institute
401 Smyth Road
Ottawa, ON, K1H 8L1
Canada
Phone: 1 613 738 4181
Fax: 1 613 731 1374
Email: kelemam@uottawa.ca

Abstract

Background: Findings and statements about how securely personal health information is managed in clinical research are mixed.

Objective: The objective of our study was to evaluate the security of practices used to transfer and share sensitive files in clinical trials.

Methods: Two studies were performed. First, 15 password-protected files that were transmitted by email during regulated Canadian clinical trials were obtained. Commercial password recovery tools were used on these files to try to crack their passwords. Second, interviews with 20 study coordinators were conducted to understand file-sharing practices in clinical trials for files containing personal health information.

Results: We were able to crack the passwords for 93% of the files (14/15). Among these, 13 files contained thousands of records with sensitive health information on trial participants. The passwords tended to be relatively weak, using common names of locations, animals, car brands, and obvious numeric sequences. Patient information is commonly shared by email in the context of query resolution. Files containing personal health information are shared by email and, by posting them on shared drives with common passwords, to facilitate collaboration.

Conclusion: If files containing sensitive patient information must be transferred by email, mechanisms to encrypt them and to ensure that password strength is high are necessary. More sophisticated collaboration tools are required to allow file sharing without password sharing. We provide recommendations to implement these practices.

(J Med Internet Res 2011;13(1):e18) doi: 10.2196/jmir.1335

KEYWORDS

Privacy; security; passwords

Introduction

Information technology is being increasingly used in clinical trials. One recent study estimated that 41% of Canadian clinical trials are using an electronic data capture (EDC) system [1]. Researchers are also turning more to electronic medical records as a source of clinically relevant patient data, and this is fueled by their growing adoption in practice [2-6].

The data collected during clinical trials consist of sensitive personal health information (PHI). Most clinical trial data sets contain fields such as participant initials, date of birth, and gender; information about the location of the participant's residence; and the clinical trial site where the participant is receiving treatment. This kind of information can be used to reidentify individuals [7-10]. In some cases, clinical trial data contain detailed contact information (eg, email addresses, residence address, or telephone numbers) for participants to

receive reminders of upcoming visits or reminders to complete specific data collection forms.

Despite strong assurances about the safety of PHI entrusted with researchers [11] and arguments about the paucity of publicly known privacy violations in medical research [12], there have been recent publicized cases of data breaches from clinical trials [13]. Risky behaviors that can result in data breaches when handling data in clinical trials have been reported [14]:

- Engineering and mathematics graduate students were participating in a study that involved the analysis of medical images. These students did not receive sufficient education on privacy issues and how to handle PHI. Consequently, they were exchanging the personal data of subjects among themselves by email without any encryption.
- There were reported cases of study coordinators taking data home to finish some work off by saving it on to a memory stick or emailing the information to public accounts that they can access from home (eg, Gmail, Sympatico, or Rogers accounts). The data that were taken home were not encrypted.
- In one study progress notes had to be completed in an EDC system during a patient visit. There were cases where the physician or nurse completing the clinical notes mentioned the patient's name, family physician name, sibling or parent name, or other identifying information in what they wrote. Therefore, even if the structured questionnaires used to collect data in a clinical research study exclude any identifying or potentially identifying information, patients can potentially be identified from the clinical notes that were submitted as part of the study.
- Another example involved the audit trails. If, for example, a nurse saved identifying information in the notes or comments section in an EDC form and then subsequently deletes that information, the information remains in the audit trail. In this scenario patients were reidentifiable through data that were available in the audit trails.
- In one study where an EDC was used, there were examples of password sharing (to avoid having to re-log in every time an individual was to work on a shared computer), and passwords written on notes posted on monitors were common.

Computer users are known to use email quite often to share files, and frequently as their primary file-sharing mechanism [15-18]. One survey of US enterprises found that approximately one-quarter reported that personal information (including PHI) was included in outbound emails in breach of regulations, and one-third had investigated a violation of data-protection regulations related to email within the previous year [19].

An earlier qualitative study indicated that email was often used to transfer information during Canadian clinical trials [14]. It has been noted that email is the most widely used communication mechanism in clinical trials [20]. One survey found that 50% of professionals working on clinical trials use email as their predominant method for sharing information [21], and two-thirds of clinical trials professionals responded that documents and files are exchanged with investigative sites via email [22]. Unfortunately, there are many ways for an adversary to access information sent by email, either during transmission or at its destination (see Multimedia Appendix 1).

In the United States the Health Insurance Portability and Accountability Act (HIPAA) permits the electronic transmission of PHI without encryption if the risk is deemed reasonable [23]. However, under many state breach notification laws the transmission of unprotected personal information by email may be considered a breach (see Multimedia Appendix 1). Many health care providers admit that they do not encrypt patient data when they are transmitted electronically [24]. On the other hand, some states, notably Nevada and Massachusetts, have mandated the encryption of electronic personal information in transit over public networks [25,26]. Noncompliance can subject data custodians to significant fines and penalties. It is likely that more states will follow with similar laws. Furthermore, recognizing the potential for a breach, various health systems have mandated the encryption of data transferred by email for delivering care and for research purposes [27-29].

Trials using an EDC system will have raw data available in electronic form throughout the study. Regulated trials need to comply with the US Food and Drug Administration's (FDA's) 21 Code of Federal Regulations (CFR) Part 11 regulations where electronic systems are used [30-35], and these include provisions for securing data to avoid tampering and ensure data integrity. Regulated trials have a higher likelihood (than unregulated trials) of being audited, and the FDA has publicized its intention of increased audits [36]. Failure to address FDA concerns expressed in warning letters could result in delays in drug and device submissions. The out-of-pocket clinical development costs for a self-originated new drug are estimated to be on average $282 million (US $467 million for capitalized costs) [37], making any delays in submissions to the FDA quite costly. Therefore, there are strong incentives by sponsors to implement reasonable security practices for such trials.

Trial participants have the expectation that their PHI will be protected by the sponsors and sites collecting data. There are also potential financial and social harms to participants if their PHI is inadvertently disclosed (see Multimedia Appendix 1).

To investigate the extent to which research staff actually protect PHI, in this paper we report on two studies: (1) a direct evaluation of one behavioral indicator of secure information management practices: the strength of passwords used to transfer encrypted electronic health information among the stakeholders in regulated Canadian clinical trials, and (2) a series of interviews of study coordinators to understand their file-sharing practices and how files are protected when shared.

Methods

We performed two studies to investigate password strength and file-sharing practices in the context of clinical trials. Each is described below. Both study protocols were reviewed and approved by the research ethics board of the Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada.

Study 1: Password Strength Analysis

Over a period of 6 months the first author contacted stakeholders in 15 clinical trials known to him to determine whether they were interested in participating in this study. All of these trials used a form of EDC system for data collection and management. Stakeholders in four clinical trials were willing to participate in this study. Stakeholders were site coordinators, statisticians, monitors, and study project managers. Three studies had at least one commercial sponsor and were consequently expected to follow FDA regulations. The fourth trial did not have a commercial sponsor but was sufficiently high in profile that it received strong regulatory oversight by Health Canada.

The clinical trials that participated were not representative of all clinical trials in Canada. They were, however, likely examples of trials where the stakeholders were sufficiently comfortable with their security practices that they agreed to participate.

The stakeholders identified password-protected electronic files that were generated or created during these trials and that were sent or received by email. All files met the following criteria:

- Their format was either Microsoft Office (Microsoft Corporation, Redmond, WA, USA) or ZIP (eg, WinZip Computing, Mansfield, CT, USA) (compressed archive; the contents of the compressed files may be any other data file type, such as Word, Excel, SAS, or XML data files). All Microsoft Office files were version 2003 or earlier (for example, with the .doc or .xls file extension).
- They were encrypted or protected using a password.
- The files were sent by email between sites, data management groups, statistical analysis groups, external consultants, or central labs with at least one party in the communication within Canada.
- They were suspected or known to have PHI of the participants.

We chose these file formats because they are the most commonly used based on their market penetration. Focusing on these document types provided us with an indicator of password strengths used by PHI custodians when they are free to select whatever password they want.

Even if the EDC system used in the trial supported some form of secure file sharing, the email exchanges we obtained the files from were with individuals involved in the trial but who did not have an account on the EDC system (eg, external statisticians and information technology specialists).

In total we examined 15 files from the four clinical trials. Nine were ZIP files and the remainder were Microsoft Office documents.

We purchased two commercial password recovery tools (Visual Zip Password Recovery Master version 6.2, Rixler Software, and Accent Office Password Recovery version 2.6, AccentSoft Utilities, St Petersburg, Russian Federation) and attempted to recover the passwords. We selected those tools based on listings at the openwall.com site, usability, and recommendations from security administrators at our institutions. Using commercial tools allowed us to assess the risk from an unsophisticated adversary.

One tool would attempt to recover the password for the Word document, and the second tool would attempt to recover the password for the whole of the ZIP archive (ie, there is one password for the whole archive). The tools use a number of techniques, including a dictionary attack, common password patterns, heuristics, brute force to recover the password, and by taking advantage of known vulnerabilities.

For dictionary attacks, we enhanced the dictionaries used to include Canada-specific terms (such as city and province names and famous personality names) and other commonly used terms and passwords (see Multimedia Appendices 2 and 3).

There are known vulnerabilities in some of the encryption methods that are used for these file types. Up to and including Word 2003, the default encryption was “97/2000 compatible.” This was an RC4 stream cipher with a 40-bit key. Because of the small key size, it would be possible to try all binary keys until one that works is found. This would not recover the password itself but would allow an adversary to access the contents of the password-protected file. Similarly, older versions of WinZip used the ZIP 2.0 encryption standard, which was considered weak. Only versions 9 and above of WinZip provide stronger encryption algorithms, such as Advanced Encryption Standard.

We used a computer running a 2 GHz dual processor with 2 GB of memory to execute the tool. The password recovery tools were allowed to run for 24 hours on each file before they were stopped.

The password recovery process was performed under the auspices of or by the stakeholder(s) themselves. Therefore, no files were transferred to any entity outside the data custodian to perform this study. The password recovery software was installed on a virtual machine and the software was run within the virtual machine on the data custodian’s equipment. The first author participated in running and monitoring the execution of the software. Each virtual machine instance, including all of the data files within it, was deleted after the analysis. We determined how many files had their password recovered during the 24-hour period.

Study 2: Study Coordinator Interviews

We identified 121 study coordinators who responded to a previous survey [1] and were located within the Toronto-Ottawa-Montreal corridor. We randomly selected a subset of 80 coordinators and sent each an email request to participate in a 1-hour interview. Assuming that we would not be able to reach 25% of the group due to a change in contact information following the previous study (eg, change of employment, relocation), we expected our email invitation to be received by approximately 60 coordinators in total. We expected a response rate of 33% from those 60 [38]. We therefore planned for a group of 20 interviewees. The purpose of the interviews was to understand the file-sharing practices used within a recent clinical trial in which each coordinator had been involved.

The 80 selected individuals were invited by email to participate (Multimedia Appendix 4 contains the text of the invitation email). As an incentive to participate, we organized a raffle for an iPod shuffle (Apple, Cupertino, CA, USA) that took place after the interviews had been completed. All interviewees were entered in the raffle.

Depending on the location and timing, some interviews were conducted face-to-face and some were conducted by telephone. The interviews were recorded and then transcribed verbatim. The open-ended interview questions are presented in Multimedia Appendix 4. The interview guide included a series of questions on the electronic file-sharing practices used during the conduct of clinical trials. Specifically, the questions elicited information related to how research coordinators addressed security and privacy issues and why they made certain file-sharing choices during clinical trials.

We used a general qualitative thematic approach to analyze the interview transcripts [39]. NVivo software version 8 (QRS International, Cambridge, MA, USA) facilitated the management and analysis of the data. We analyzed the data by developing a "start list" of codes based on the interview guide for the study, as well as the issues and themes that we expected to see in the data. However, recognizing that some codes would emerge or disappear during the analysis, we only used these predefined codes as starting points and embraced any new or revised issues or themes that emerged from the data.

Results

Password Strength Analysis

The ZIP files contained more than 2000 data files in their archive. In all cases the tools were able to recover the password, except for one file where the password could not be cracked within the 24-hour period. One of the recovered files contained coding information and dictionaries, and therefore did not have any PHI.

In all cases the recovered passwords were poorly constructed [40], with names of local locations (eg, "ottawa"), names of animals (eg, "cobra"), car brands (eg, "nissan"), and common number sequences (eg, "123"). This makes it easier for password recovery tools to guess them.

The files with recovered passwords that had PHI included Microsoft Word, Microsoft Excel, SAS, and XML (Clinical Data Interchange Standards Consortium Operational Data Model format files). They contained raw data from the clinical trials. In total, more than 10,000 patient records were in these files, and many with PHI on the subjects. For example, fields included name of study site, dates of screening and randomization, date of birth, initials, gender, and medical history.

For Microsoft Office document files, password-protecting a document is not the same as encrypting its contents [41]. Password protection controls the actions that can be performed on the document, such as who can modify a document, but the contents themselves are not encrypted. It may not always be obvious to an end user that such document protection does not protect the document contents themselves. A different program that ignores the document protections can be used to read the unencrypted contents, or they can be examined through a binary file viewer. All of the files in our sample were encrypted, but all used the default "97/2000 compatible" encryption.

Passwords on older versions of Word and Excel files are relatively straightforward to recover under certain conditions [42]. Word and Excel 2003 also have an option to use an RC4 stream cipher with a key length of up to 128 bits. A weakness in the implementation of the encryption module makes it possible for an adversary to compare two versions of a password-protected file to recover its plaintext contents [42,43]. In such a case password strength would not have affected the ability to extract the PHI. However, in our study we had only one version of each document and therefore our files were not vulnerable to this attack.

All of the ZIP files in our data set used the ZIP 2.0 encryption standard. All of the recovered passwords from the ZIP files were poor choices, and most of them were in our dictionaries or derived from words in the dictionaries (eg, ottawa followed by a digit).

Study Coordinator Interviews

We interviewed 20 study coordinators in the Toronto-Ottawa-Montreal corridor.

There was a marked difference between industry-sponsored trials and investigator-initiated trials. Specifically, industry-sponsored trials tended to have more formal processes in place to protect PHI and defined mechanisms for sharing data among those directly involved in the trial.

The three primary modes for sharing electronic information in the context of trials were as follows.

By Email

Data sent by email included mostly queries and responses to queries (eg, questions to sites about inconsistent or incomplete data for a particular patient). According to our informants, patient information was rarely encrypted when sent this way.

If PHI data files were sent by email then they were encrypted. This was used to justify the transmission of such files using an inherently insecure medium. If there was no EDC system in use in the trial or it did not support file sharing, then files were exchanged between any of the individuals and organizations working on the trial. If an EDC system that supported file sharing was deployed, then email was used to send data files to individuals who do not have accounts on the system.

Shared Drives

These drives were used within sites to store all trial information, including keys linking pseudonyms to patient names and Case Report Form (CRF) data. All site staff working on the trial would normally have access to the files on the shared drive. If the files were protected, the same password was often used for all of the files, and all staff who needed to access the documents would know that password. Formal processes for changing individual and shared credentials after the departure of staff were often not defined. Generally, individuals would not be taken off the access list once the trial was complete.

The file formats that we considered encourage the sharing of passwords. For example, it is not possible to assign different passwords to each individual who needs to access each of these documents. A single password is used for a document, and all individuals who need to read the document know that same password. If many documents need to be exchanged, it is not practical to have a different password for each one; therefore, often a single password is used for all documents and this password is shared among all users.

EDC Systems

In trials using EDC systems that support file sharing (through either an internal email system or document management features), individual patient-level data would be shared through the EDC system. The amount of access control would depend on the specific EDC system in question. If the EDC system did not support file sharing then most often email would be used.

It should be noted that, given the sensitivity of the topic, the interviewees may have held back some information. Specifically, they may not have been willing to share information about poor security practices in the trials they were participating in. Consequently, our results should be seen as an optimistic view of current practices.

Discussion

Summary

Previous work had indicated that password-protected files containing the PHI of clinical trial participants were being sent by email. Our initial study objective was to examine the strength of the passwords used to protect those files. Strong passwords were seen as an indicator of following good security practices in the context of clinical research.

We obtained a sample of 15 encrypted files that were sent by insecure email and were able to recover the passwords for 93% (14/15) of the files using commercial password recovery tools. Thirteen of those 14 files (93%) had sensitive health information in them. Therefore, in total 13/15 files were recovered *and* had PHI (87%). Since we were able to recover passwords using off-the-shelf tools, then it would be quite easy for an unsophisticated adversary to also do so. This result is consistent with previous research showing that health care professionals choose weak passwords to access patient data when there are no restrictions on password strength [44].

Perhaps more alarming, all of the Office and ZIP files in our sample used the default weak encryption methods. Therefore, an adversary had two different ways to extract the PHI: by attacking the weak algorithm itself or by attacking the weak password. In the current version of the WinZip tool (version 14.5), the default encryption is *still* based on the weak ZIP 2.0 standard.

At the time of this study the default applications for these file formats (ie, Microsoft Office and WinZip) did not enforce any password strengths, which means users could create any password they wished. For example, in earlier versions of WinZip that did provide password protection it was not possible to enforce a particular password strength (older versions of WinZip are still available [45]). Similarly, only recent versions of Microsoft Word have provided password strength enforcement [46]. Therefore, the passwords chosen were those that the stakeholders believed were sufficiently strong.

A follow-up interview study to examine the file-sharing practices of clinical trial study coordinators indicated that some PHI was exchanged by email that was not encrypted (eg, queries about specific patient data). Shared drives were another commonly used mechanism for exchanging files containing participant PHI. Shared drives create additional risks because, in practice, all files posted on the drive share a common password, and this common password is also shared among all stakeholders who need to access any one of the files. Sharing passwords is a violation of best-security practices. Furthermore, this goes against another best practice of limiting access to PHI to only the information that an individual needs (ie, a person who needs to access a single file should not get the password to access all files). From a regulatory perspective, it is also not possible to maintain audit trails of modifications made to files on shared drives.

Recommendations

Encrypt PHI Sent by Email

Protocols can be employed to securely exchange information that was sent by email using PGP (Pretty Good Privacy) or S/MIME (Secure/Multipurpose Internet Mail Extensions) [47]. However, these tools remain quite difficult for people to use [48-50]. Furthermore, in an enterprise setting where the key management complexities are handled by a central information technology department, they are still complicated to use when communicating beyond institutional boundaries and therefore may not be suitable for distributed collaborations that cross such boundaries.

Some products bypass the key management complexities by sending a plaintext notification email to the receiver that they have received a message with a link to a secure website where they can pick up their email [51]. The receiver, however, then needs to create an account on the secure website to pick up the message. In the context of clinical trials with staff joining and leaving throughout, such an option may be workable if creating an account is simple.

Another common approach is to use the built-in password protection capabilities available in tools for common file formats (such as WinZip and Microsoft Office) and then transmit the encrypted files. Instructions for encrypting Microsoft Office and ZIP files are available [41,42,52-54]. However, caution should be exercised when using some of these tools. The default encryption standard may be a weak one. A strong encryption algorithm must be selected or set as the default.

If file encryption tools will be the main mechanism used to protect PHI, then all PHI needs to be in files, including queries and their responses.

Users may get confused between encrypting a file and protecting parts of it with a password (which does not encrypt it). Therefore, an alternative that avoids the potential for confusion

is to use an external file encryption tool [55], whereby it would be clear that the whole file is being encrypted.

Enforce Strong Passwords

Where file encryption with passwords will be used, policies need to be put in place to ensure that strong passwords are also used. Ensuring password strength would mitigate the type of attack we describe in this paper. Standards for passwords are available [56], as well as general guidelines on email security [47] and information management security in the health care context [57].

The default applications for creating Office and ZIP files can enforce passwords, but only if the most recent versions are used, as only these have such capabilities, and they need to be set up to enforce password strength.

This needs to be augmented with privacy training for study coordinators so that they have an appreciation of privacy risks when using information technology in the conduct of trials. Training should cover procedures for the handling of electronic data, as well as providing background on the security risks of the specific technologies used in the study.

Minimize Password Sharing

In collaborative workflows that are common in clinical trials, current methods for file sharing are risky because they require password sharing, for example, by sharing files through email or on shared drives. It does not matter how strong a password is; if many individuals know that password then it is not a secure password.

Shared passwords make it difficult to maintain clear audit trails of individuals responsible for particular changes, which is a critical requirement in 21 CFR Part 11. For example, if multiple individuals at a site are able to view and edit an encrypted document on a shared drive because they all have the password, this would likely run afoul of the regulations because audit trails of modifications made to individual files are not maintained with shared drives.

Encryption of documents today assigns the password to the document rather than to the individual. To eliminate password sharing means creating multiple copies of each document with a unique password for every user. Commonly used contemporary tools cannot handle such additional password management complexity.

A more practical solution is to use collaboration environments, such as Microsoft SharePoint or equivalent ones. These allow the creation of repositories with different access controls for different users without the need to encrypt the documents themselves or store them on hosted email servers. Collaboration environments can also maintain detailed audit trails and version control.

Make File-Sharing Systems Inclusive

Modern EDC systems support secure email communications between stakeholders in the trial within the walls of the system, and some provide secure file sharing and document management mechanisms. Despite this capability, some of the stakeholders in clinical trials do not have access to the EDC system. For example, an external statistician would not normally have an EDC account and therefore may be sent a data file by email. The user base for such systems can be quite large, including individuals across multiple organizations, and these individuals change during a trial [58]. In addition, if there are multiple staff working on a trial within a single site, then they ought to all have EDC system accounts, otherwise mechanisms such as shared drives are used. Therefore, the use of an EDC system with good security practices around file sharing is insufficient insurance against inappropriate security practices unless *everyone* who needs to access files has an account on it.

File-sharing capabilities may not be embedded within an EDC system, but may also be complementing an EDC system (eg, a document management system). In such cases the same conditions noted above would need to apply.

In the future, the use of federated authentication systems could allow file sharing that is more appropriate to the workflows in clinical trials.

Strengthen Data Breach Notification Exemptions

It should not be taken for granted that the default file encryption algorithms used to protect PHI are strong. In fact, we found that emailing the ZIP files in our sample would be considered a data breach under the US Health Information Technology for Economic and Clinical Health (HITECH) Act because they all used the weak ZIP 2.0 standard. Furthermore, the emailing of files encrypted using the default encryption in Word 2003 and earlier would also be a breach under the US HITECH Act. Therefore, the simple technical act of encryption does not ensure that this was done effectively [59,60]. A good example illustrating this is the case of TJX Companies, the parent company of some of the largest retailers in the United States, whereby adversaries were able to crack a weak encryption algorithm and access more than 90 million credit card numbers [61,62]. Encryption exemptions should always require that the algorithms used must meet a minimal standard.

Encryption exemptions in breach notification laws should explicitly consider the strength of the passwords that are used. If, for example, a sensitive document on someone's hacked Gmail account is encrypted and the password is "password," then the encryption is somewhat meaningless, however strong the algorithm itself is. Based on the results of our study, it seems prudent to consider password strength in determining whether an exemption applies: it should not be assumed that encryption, even with a strong algorithm, means that it was done adequately and that the adversary would not be able to figure out the password. Some states, such as North Carolina and Oregon, recognize the risk of an adversary acquiring the decryption key or password [59], and therefore would not allow an encryption exemption from notification under those conditions.

Limitations

Given the small number of trials from which we obtained files, broad generalization of the results is difficult. But we did expect that only trials that had good security and privacy practices would be willing to participate. We also expected that only study coordinators who were comfortable with the quality of their security practices would be willing to participate in the

interviews. Therefore, the findings are expected to be biased toward those who were security-aware and were investing in protecting the data. Should this be case, then the more general state of affairs would be worse than depicted by our conservative results.

All of our data were collected from Canadian trials and Canadian coordinators. While the regulated trials from which we collected data had international sponsors and our interviewees participated in and discussed practices in international trials, our findings are specific to practices within a Canadian geography.

Our results indicate a potential privacy risk rather than an actual risk, since we do not know whether anyone has actually inappropriately accessed these files and cracked their passwords. However, this should not dilute the seriousness of the risk, since one purpose of having good password management practices is to act as a deterrent against an attack.

Conclusions

When sharing files containing PHI in the context of clinical trials, it is critical to encrypt all PHI. However, such a practice does not provide much protection if the passwords are weak or if the passwords are widely shared. Our study indicated that the passwords used are not strong and could be compromised using a commercial password recovery tool, and that some file-sharing practices used in clinical trials promote the wide sharing of passwords among study staff.

These results suggest that stronger oversight is needed on the transfer of health information in the context of clinical trials, and better training and enforcement (technical and procedural) of good security practices.

Acknowledgments

We wish to thank Liam Peyton for reviewing an earlier version of this paper.

Conflicts of Interest

None declared

Multimedia Appendix 1

Background on email file sharing in clinical trials

[PDF file (Adobe PDF File), 48 KB-Multimedia Appendix 1]

Multimedia Appendix 2

A manually constructed password list file

[ZIP file (ZIP Archive), 3168 KB-Multimedia Appendix 2]

Multimedia Appendix 3

The “npasswd” password quality-checking tool dictionary.

[GZ file (GZIP Archive), 5518 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Invitation and questions

[PDF file (Adobe PDF File), 20 KB-Multimedia Appendix 4]

References

1. El Emam K, Jonker E, Sampson M, Krleza-Jerić K, Neisa A. The use of electronic data capture tools in clinical trials: Web-survey of 259 Canadian trials. J Med Internet Res 2009;11(1):e8 [FREE Full text] [doi: 10.2196/jmir.1120] [Medline: 19275984]
2. Irving R. 2002 Report on Information Technology in Canadian Hospitals. Thornhill, Ontario: Canadian Healthcare Technology; 2003.
3. Healthcare Information and Management Systems Society Foundation. Healthcare CIO Results. Chicago, IL: HIMSS Foundation; 2004.
4. Andrews JE, Pearce KA, Sydney C, Ireson C, Love M. Current state of information technology use in a US primary care practice-based research network. Inform Prim Care 2004;12(1):11-18. [Medline: 15140348]
5. Bower AG. The Diffusion and Value of Healthcare Information Technology. Santa Monica, CA: Rand Corp.; 2005.
6. Fonkych K, Taylor R. The State and Pattern of Health Information Technology Adoption. Santa Monica, CA: Rand Corp.; 2005.

7. El Emam K, Jonker E, Sams S, Neri E, Neisa A, Gao T, et al. Pan-Canadian De-Identification Guidelines for Personal Health Information. Ottawa, Ontario: Office of the Privacy Commissioner of Canada; 2007.
8. El Emam K, Dankar FK, Vaillancourt R, Roffey T, Lysyk M. Evaluating the risk of re-identification of patients from hospital prescription records. Can J Hosp Pharm 2009;62(4):307-319 [FREE Full text]
9. El Emam K, Kosseim P. Privacy interests in prescription records, part 2: patient privacy. IEEE Security & Privacy Magazine 2009;7(2):75-78. [doi: 10.1109/MSP.2009.47]
10. El Emam K, Jabbouri S, Sams S, Drouet Y, Power M. Evaluating common de-identification heuristics for personal health information. J Med Internet Res 2006;8(4):e28 [FREE Full text] [doi: 10.2196/jmir.8.4.e28] [Medline: 17213047]
11. Upshur RE, Morin B, Goel V. The privacy paradox: laying Orwell's ghost to rest. CMAJ 2001 Aug 7;165(3):307-309. [Medline: 11517649]
12. Gershon AS, Tu JV. The effect of privacy legislation on observational research. CMAJ 2008 Mar 25;178(7):871-873. [doi: 10.1503/cmaj.061353] [Medline: 18362384]
13. Cavoukian A. Office of the Privacy Commissioner of Ontario. 2007. Order HO-004 URL: http://www.ipc.on.ca/images/Findings/up-3ho_004.pdf [accessed 2007-04-20] [WebCite Cache ID 5OFOzaj1O]
14. El Emam K. Data Anonymization Practices in Clinical Research: A Descriptive Study. Ottawa, Ontario: Health Canada, Access to Information and Privacy Division; 2006.
15. Johnson ML, Bellovin SM, Reeder RW, Schechter SE. Laissez-faire file sharing: access control designed for individuals at the endpoints. In: NSPW '09 Proceedings of the 2009 Workshop on New Security Paradigms. New York NY: ACM Press; 2009.
16. Dalal B, Nelson L, Smetters D, Good N, Elliot A. Ad-hoc guesting: when exceptions are the rule. 2008 Presented at: Usability, Psychology, and Security 2008; Apr 14, 2008; San Francisco, CA URL: http://www.usenix.org/events/upsec08/tech/full_papers/dalal/dalal_html/dalal.html
17. Voida S, Edwards WK, Newman MW, Grinter RE, Ducheneaut N. Share and share alike: exploring the user interface affordances of file sharing. In: Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems. New York, NY: ACM Press; 2006:221-230.
18. Whalen T, Smetters D, Churchill EF. User experiences with sharing and access control. In: Proceedings of the CHI '06 Extended Abstracts on Human Factors in Computing Systems. New York, NY: ACM Press; 2006:1517-1522.
19. Osterman Research. Proofpoint. 2010. Outbound Email and Data Loss Prevention in Today's Enterprise URL: http://www.proofpoint.com/downloads/Proofpoint-Outbound-Email-and-Data-Loss-Prevention-2010.pdf [accessed 2010-12-13] [WebCite Cache ID 5ux9tNwBx]
20. Wheeler D. Applied Clinical Trials Online. 2010. Rethinking Document Sharing: The Benefits of Peer-to-Peer Networking Over Email, Fax, and Hosted Solutions URL: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/CRO%2FSponsor/Rethinking-Document-Sharing/ArticleStandard/Article/detail/660941 [accessed 2010-10-02] [WebCite Cache ID 5tBqD8vNb]
21. Shapiro M. Applied Clinical Trials Online. 2009. Poll Finds Ironic Inefficiency: Uncovering the Risky Communication Methods of Clinical Trials Professionals, While Discovering a Potential Web-Based Replacement URL: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/article/articleDetail.jsp?id=586857 [accessed 2010-10-02] [WebCite Cache ID 5tBqPCPWN]
22. Applied Clinical Trials Online. 2009. Document Management Inefficiencies Cost Sites Time and Money Says Intralinks' Survey URL: http://appliedclinicaltrialsonline.findpharma.com/appliedclinicaltrials/News/Document-Management-Inefficiencies-Cost-Sites-Time/ArticleStandard/Article/detail/601186?contextCategoryId=44911&ref=25[WebCite Cache ID 5tBqd7lO7]
23. Schmidt DA. SANS Institute InfoSec Reading Room. 2003. E-mail Communication With Patients in the Wake of the HIPAA Final Security Rule URL: http://www.sans.org/reading_room/whitepapers/legal/e-mail-communication-patients-wake-hipaa-final-security-rule_1057 [accessed 2010-12-13] [WebCite Cache ID 5uxAd8Ign]
24. HIMSS Foundation. Healthcare Information and Management Systems Society. 2009. 2009 HIMSS Security Survey URL: http://www.himss.org/content/files/HIMSS2009SecuritySurveyReport.pdf [accessed 2010-12-13] [WebCite Cache ID 5uxApn1kW]
25. Commonwealth of Massachusetts. Mass.gov. 201 CMR 17. 00: Standards For the Protection of Personal Information of Residents of the Commonwealth URL: http://www.mass.gov/Eoca/docs/idtheft/201CMR1700reg.pdf [accessed 2011-01-14] [WebCite Cache ID 5vjwvZkj3]
26. Worthen B. Wall Street Journal. 2008. New data privacy laws set for firms URL: http://online.wsj.com/article/SB122411532152538495.html [accessed 2010-10-03] [WebCite Cache ID 5tCoEs06a]
27. Cottis P. Redbridge Primary care Trust (NHS). 2008. Transferring Personal Information Policy and Procedures URL: http://www.redbridge.nhs.uk/files/documents/1030_transferring%20personal%20information%20policy%20080215.pdf [accessed 2010-12-13] [WebCite Cache ID 5uxBTsEuk]
28. Research Computing. Partners Healthcare. 2010. Email Encryption URL: http://rc.partners.org/emailencryption/ [accessed 2010-10-03] [WebCite Cache ID 5tCnUoOn3]

29. Chalmers L. University of Edinburgh, Community Health Sciences. 2007. Guidance to Research and the Data Protection Act URL: http://www.chs.med.ed.ac.uk/cphs/researchTraining/DPResearch.pdf [accessed 2010-12-22] [WebCite Cache ID 5vAdm66WW]
30. US Department of Health and Human Services, Food and Drug Administration, Office of the Commissioner. FDA. 2007. Guidance for Industry: Computerized Systems Used in Clinical Investigations URL: http://www.fda.gov/Cder/Guidance/7359fnl.pdf [accessed 2009-01-11] [WebCite Cache ID 5dkOZy0uG]
31. Good Automated Manufacturing Practice Forum. The Good Automated Manufacturing Practice (GAMP) Guide for Validation of Automated Systems in Pharmaceutical Manufacture. Tampa, FL: International Society for Pharmaceutical Engineering; 2002.
32. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologic Evaluation and Research, Center for Devices and Radiological Health, Center for Food Safety and Applied Nutrition, Center for Veterinary Medicine, Office of Regulatory Affairs. FDA. 2003. Guidance for Industry: Part 11, Electronic Records; Electronic Signatures - Scope and Application URL: http://www.fda.gov/cder/guidance/5667fnl.pdf [accessed 2009-01-11] [WebCite Cache ID 5dkOj5uYQ]
33. US Department of Health and Human Services, Food and Drug Administration. 21 CFR Part 11: electronic records; electronic signatures; final rule. Federal Register 1997;62(54):13430-13466 [FREE Full text]
34. US Department of Health and Human Services, Food and Drug Administration, Center for Biologic Evaluation and Research, Center for Drug Evaluation and Research, Center for Devices and Radiological Health, Center for Food Safety and Applied Nutrition, Center for Veterinary Medicine, Office of Regulatory Affairs. FDA. 1999. Guidance For Industry: Computerized Systems Used in Clinical Trials URL: http://www.fda.gov/ora/compliance_ref/bimo/ffinalcct.pdf [accessed 2011-01-14] [WebCite Cache ID 5vjx3KTzB]
35. International Pharmaceutical Privacy Consortium. IPPC. 2006. Transmission Security Practices of Pharma Sponsors of Clinical Research URL: http://www.pharmaprivacy.org/download/Clinical_Research_Transmission_Security.pdf [accessed 2010-12-23] [WebCite Cache ID 5vCIzGZd7]
36. US Department of Health and Human Services, Food and Drug Administration. FDA. 2010. FDA to Conduct Inspections Focusing on 21 CFR 11 (Part 11) Requirements Relating to Human Drugs URL: http://www.fda.gov/AboutFDA/CentersOffices/CDER/ucm204012.htm [accessed 2010-10-01] [WebCite Cache ID 5t9ze5Jnr]
37. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ 2003 Mar;22(2):151-185. [doi: 10.1016/S0167-6296(02)00126-1] [Medline: 12606142]
38. Schonlau M, Fricker RD, Elliott MN. Conducting Research Surveys Via E-mail and the Web. Santa Monica, CA: RAND; 2002.
39. Miles MB, Huberman AM. In: Huberman AM, editor. Qualitative Data Analysis: An Expanded Sourcebook. 2nd edition. Thousand Oaks, CA: Sage Publications; 1994.
40. Cazier JA, Medlin BD. Password security: an empirical investigation into e-commerce passwords and their crack times. Information Security Journal 2006;15(6):45-55. [doi: 10.1080/10658980601051318]
41. Microsoft. 2003. Microsoft Office 2003 Editions Security Whitepaper URL: http://office.microsoft.com/download/afile.aspx?AssetID=AM102424861033 [accessed 2010-12-17] [WebCite Cache ID 5v3FVk3pw]
42. Microsoft. 2007. 2007 Office System Document: 2007 Microsoft Office System Document Encryption URL: http://www.microsoft.com/downloads/en/details.aspx?FamilyID=0444ea0e-3f62-4da0-8551-52349b70272e&displaylang=en
43. Wu H. International Association for Cryptologic Research. 2005. The Misuse of RC4 in Microsoft Word and Excel URL: http://eprint.iacr.org/2005/007.pdf [accessed 2011-01-14] [WebCite Cache ID 5vjxZJxof]
44. Cazier JA, Medlin BD. How secure is your information system? An investigation into actual healthcare worker password practices. Perspect Health Inf Manag 2006;3:8. [Medline: 18066366]
45. CBCnews.ca. 2010 Aug 04. Hundreds of Ont. Patient Health Files Stolen URL: http://www.cbc.ca/canada/toronto/story/2010/08/04/usb-medical-files-stolen684.html?ref=rss&loomia_si=t0:a16:g2:r5:c0.0440057:b36264812[WebCite Cache ID 5rkr1UjXO]
46. Microsoft Corporation. Microsoft. 2010. Password Policy URL: http://office.microsoft.com/en-ca/excel-help/password-policy-HA010355926.aspx [accessed 2010-09-29] [WebCite Cache ID 5t76jss0L]
47. Bisker S, Tracy M, Jansen W. National Institute of Standards and Technology. 2002. Guidelines on Electronic Mail Security URL: http://www.21cfrpart11.com/files/library/security/guidelines_on_email_sec.pdf [accessed 2010-12-16] [WebCite Cache ID 5v1jynapa]
48. Whitten A, Tygar J. Why Johnny can't encrypt: a usability case study of PGP 5. In: Proceedings. 1999 Presented at: 8th USENIX Security Symposium; Aug 23-26, 1999; Washington, DC.
49. Sheng S, Broderick L, Koranda CA, Hyland JJ. Why Johnny still can't encrypt: evaluating the usability of email encryption software. In: CyLab Usable Privacy and Security Laboratory, Carnegie Mellon University. 2006 Presented at: Symposium On Usable Privacy and Security; Jul 12-14, 2006; Pittsburgh, PA URL: http://cups.cs.cmu.edu/soups/2006/posters/sheng-poster_abstract.pdf
50. Garfinkel SL, Miller RC. Johnny 2: a user test of key continuity management with S/MIME and Outlook Express. In: CyLab Usable Privacy and Security Laboratory, Carnegie Mellon University. 2005 Presented at: Symposium On Usable

Privacy and Security; Jul 6-8, 2005; Pittsburgh, PA URL: http://cups.cs.cmu.edu/soups/2005/2005proceedings/p13-garfinkel.pdf
51. Firstbrook P, Ouellet E. Gartner. 2010. Magic Quadrant for Secure E-mail Gateways URL: http://synctech.com.vn/Casestudy/Gartner%20Magic%20Quadrant%20for%20Secure%20Email%20Gateways-2010.pdf [accessed 2011-01-14] [WebCite Cache ID 5vjxoiPus]
52. WinZip Computing. WinZip.com. How Do You Encrypt Files in a Zip File with WinZip? 2010 URL: http://kb.winzip.com/kb/entry/78/ [accessed 2010-09-28] [WebCite Cache ID 5t5VpbDwr]
53. WinZip Computing. WinZip.com. 2010. Password Policy For Encryption URL: http://kb.winzip.com/kb/entry/260/ [accessed 2010-09-28] [WebCite Cache ID 5t5VlTV7O]
54. Shinder D. TechRepublic. 2007. Safeguard Your Office 2007 Files With Encryption, Document Protection, and Digital Signatures URL: http://articles.techrepublic.com.com/5100-10878_11-6176764.html [accessed 2010-09-29] [WebCite Cache ID 5t779mb5Y]
55. Axantum Software. Axantum.com. 2011. AxCrypt Software Download URL: http://www.axantum.com/AxCrypt/Downloads.html [accessed 2011-01-14] [WebCite Cache ID 5vjxrk6Cb]
56. Federal Information Processing Standards. National Institute of Standards and Technology, Information Technology Laboratory. 1985. Standard for Password Usage URL: http://www.itl.nist.gov/fipspubs/fip112.htm [accessed 2011-01-14] [WebCite Cache ID 5vjxt0qFA]
57. HITRUST Alliance. HITRUSTAlliance.net. 2010. Common Security Framework URL: http://www.hitrustalliance.net/commonsecurityframework/ [accessed 2011-01-14] [WebCite Cache ID 5vjxtZiZ8]
58. Olson L. Electronic record challenges for clinical systems. Drug Inf J 2001;35:721-730.
59. Burdon M, Low R, Reid J. If it's encrypted its secure! The viability of US state-based encryption exemptions. In: Proceedings of the 2010 IEEE International Symposium on Technology and Society. Los Alamitos, CA: IEEE; 2010.
60. Burdon M, Reid J, Low R. Encryption safe harbours and data breach notification laws. Computer Law & Security Review 2010;26(5):520-534. [doi: 10.1016/j.clsr.2010.07.002]
61. Berg GG, Freeman MS, Schneider KN. Analyzing the TJ Maxx data security fiasco: lessons for auditors. CPA Journal 2008;78(8):34-37 [FREE Full text]
62. Pereira J. Breaking the code: how credit card data went out wireless door. Wall Street Journal 2007 May 04.

Abbreviations

CFR: Code of Federal Regulations
CRF: Case Report Form
EDC: electronic data capture
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
HITECH Act: Health Information Technology for Economic and Clinical Health Act
PHI: personal health information

Edited by G Eysenbach; submitted 13.08.09; peer-reviewed by K Shuaib, D Chen, F Manion; comments to author 03.09.09; revised version received 23.12.10; accepted 12.01.11; published 11.02.11

Please cite as:
El Emam K, Moreau K, Jonker E
How Strong are Passwords Used to Protect Personal Health Information in Clinical Trials?
J Med Internet Res 2011;13(1):e18
URL: http://www.jmir.org/2011/1/e18/
doi: 10.2196/jmir.1335
PMID: