

Original Paper

An Evaluation of Personal Health Information Remnants in Second-Hand Personal Computer Disk Drives

Khaled El Emam¹, BEng, PhD; Emilio Neri², BEng; Elizabeth Jonker³, BA

¹University of Ottawa and Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

²TrialStat Corporation, Ottawa, Canada

³Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada

Corresponding Author:

Khaled El Emam, BEng, PhD

CHEO Research Institute

401 Smyth Road

Ottawa, Ontario K1H 8L1

Canada

Phone: ++1 613 797 5412

Fax: ++1 613 741 0397

Email: kelemam@uottawa.ca

Abstract

Background: The public is concerned about the privacy of their health information, especially as more of it is collected, stored, and exchanged electronically. But we do not know the extent of leakage of personal health information (PHI) from data custodians. One form of data leakage is through computer equipment that is sold, donated, lost, or stolen from health care facilities or individuals who work at these facilities. Previous studies have shown that it is possible to get sensitive personal information (PI) from second-hand disk drives. However, there have been no studies investigating the leakage of PHI in this way.

Objectives: The aim of the study was to determine the extent to which PHI can be obtained from second-hand computer disk drives.

Methods: A list of Canadian vendors selling second-hand computer equipment was constructed, and we systematically went through the shuffled list and attempted to purchase used disk drives from the vendors. Sixty functional disk drives were purchased and analyzed for data remnants containing PHI using computer forensic tools.

Results: It was possible to recover PI from 65% (95% CI: 52%-76%) of the drives. In total, 10% (95% CI: 5%-20%) had PHI on people other than the owner(s) of the drive, and 8% (95% CI: 7%-24%) had PHI on the owner(s) of the drive. Some of the PHI included very sensitive mental health information on a large number of people.

Conclusions: There is a strong need for health care data custodians to either encrypt all computers that can hold PHI on their clients or patients, including those used by employees and subcontractors in their homes, or to ensure that their computers are destroyed rather than finding a second life in the used computer market.

(*J Med Internet Res* 2007;9(3):e24) doi: [10.2196/jmir.9.3.e24](https://doi.org/10.2196/jmir.9.3.e24)

KEYWORDS

Privacy; confidentiality; security; data disclosure

Introduction

The adoption of electronic medical records is growing [1-5]. Concurrently, a majority of patients, and the public in general, are concerned about unauthorized disclosure and use of their personal health information (PHI) in an era of the electronic medical record [6-10].

Concern about privacy has caused some members of the public to not be totally honest with their health care provider [10]. A

survey in the United States found that as many as 15% of adults have changed their behavior to protect their privacy [6]. Those behavior changes include going to another doctor, paying out-of-pocket when insured to avoid disclosure, not seeking care to avoid disclosure to an employer, giving inaccurate or incomplete information on medical history, and asking a doctor not to write down the health problem or to record a less serious or embarrassing condition. More than a quarter of teens indicated that they would not seek out health care if they had concerns about the confidentiality of their information [11]. In

a survey of US physicians, nearly 87% reported that a patient had asked that information be kept out of their record, and nearly 78% of physicians said that they had withheld information from a patient's record due to privacy concerns [12]. Similar behaviors have been reported in Canada. A survey estimated that 12% of Canadians have withheld information from a health care provider because of concerns over whom the information might be shared with or how it might be used [13], and an estimated 735000 Canadians decided not to see a health care provider for the same reasons [14]. Such behavior changes can reduce the accuracy of health data [15-18].

Due to inaccurate data, patient safety may be jeopardized: clinicians may make treatment errors [19] or make errors in ordering medications [20]. Furthermore, researchers may underestimate disease prevalence [21], and health system managers may underestimate compliance with standards of care such as vaccination guidelines [22]. Health care organizations may be fined if they report inaccurate data to government agencies [23].

While federal and provincial health care privacy legislation, such as the Health Insurance Portability and Accountability Act in the United States and the Personal Health Information Privacy Act in Ontario, do motivate organizations to properly protect PHI, we do not know the extent to which that has actually been effective in eliminating inadvertent disclosures of PHI.

One relatively easy way to get personal information (PI) about other people is through the disk drives available on the second-hand computer market [24-26]. These computers may have been deliberately resold by their owners (individuals or organizations), donated to good causes (eg, charities or schools) who subsequently sold them, or may have ended up on the second-hand market after they were lost or stolen. In this study we examine the data remnants in second-hand disk drives to determine the extent to which PHI is inadvertently leaking from data custodians. To our knowledge, there have been no studies that have attempted to assess the extent to which PHI can be inappropriately disclosed in this way.

Methods

One approach to evaluate the extent to which PHI is leaking from data custodians is to count the number of security breaches that are publicized in the media. This, however, has a number of disadvantages: (1) not all security breaches involve PHI (eg, many are of financial data) and media reports may not make the distinction, (2) not all security breaches result in PHI being disclosed (eg, the data was encrypted) and such details usually do not appear in media reports, and (3) only some US states and only one Canadian province have breach notification laws [27]. Therefore, it is plausible that many breaches never get reported in the media.

Consequently, in this study we estimate the prevalence of PI and PHI leaks through second-hand disk drives. Our measure is the proportion (percentage) of second-hand disk drives available on the reseller market with PI and PHI on them. We purchased functional computer disk drives from the second-hand computer equipment reseller market across Canada and

examined their contents using digital forensic tools. All nonfunctional drives were returned and replaced.

This study was conducted in the winter of 2006/07. Ethics approval was obtained from the Research Ethics Board at the Children's Hospital of Eastern Ontario (CHEO) Research Institute.

Number and Type of Drives

Our focus was on drives that would be used by individual end users (ie, in their desktop machines and laptops). This means that we excluded drives that were used in servers. Hence, we focused on smaller disk drives with a capacity range of 10 GB to 40 GB.

There is no previous research on PHI leaks in second-hand disk drives; therefore, we relied on data remnant studies of PI to determine the number of disk drives needed to estimate the prevalence of PHI [24-26]. We expected the proportion of drives that leak PHI to be smaller than the proportion leaking PI since there is less health information collected and stored electronically compared to other types of PI (eg, financial and legal information). We therefore expected the proportion of drives with PHI to be closer to the lower end of PI, which is 18% [26]. The size of the 95% confidence interval in previous studies that analyzed more than 12 functional drives ranged from $\pm 9\%$ to $\pm 11\%$ [26]. We then selected an interval value in the middle: $\pm 10\%$, which ensures that the precision of our estimates is within the expected range for this type of study. The minimal number of drives to obtain such a confidence interval is 57. Consequently, we aimed to get data from 60 functional disk drives.

Identifying Vendors

A comprehensive list of 125 Canadian second-hand computer equipment vendors was identified from telephone directories, contacts and experts in the computer industry, Canadian vendors selling on eBay during the study period, local business directory searches, and a Google search to find "used computer equipment in Canada." The results were reviewed to form a list of 40 credible potential vendors distributed geographically across the country. We went down the shuffled list and systematically contacted these vendors via telephone and/or email for more information on their inventory. Used disk drives were purchased either in person, over the phone, or the Web, and were picked up personally or shipped to our lab. We limited the maximum number of drives per vendor to 10 so as to ensure a wider distribution of sources.

After contact, some vendors were excluded for a number of reasons:

- They would not sell individual disk drives. Due to cost constraints, we could only purchase stand-alone drives rather than fully configured second-hand computers from which the drives could be physically extracted.
- Some vendors in rural areas did not want to ship the equipment across the country or did not accept payments by phone or over the Internet.
- Some vendors did not have disk drives within the stipulated size range in stock at the time of the study.

We were able to purchase equipment from 12 different vendors (multiple sites for retail chains were counted as multiple vendors).

Data Recovery

Functioning drives can be classified as blank, recoverable, or securely wiped. Blank drives were readable but there was no data on them at all, current or deleted. Data on drives that have been formatted or repartitioned can be quite easily recovered [24]. Files that are deleted are also recoverable since a delete does not actually remove the data from the drive but only removes the entry from the file directory. We used a commercial software package called Recover My Files (GetData Pty Ltd, Hurstville, NSW, Australia, version 3.98, build 5282) to retrieve the data from the drives and recover the files that have been deleted [28]. The same tool was used to recover data from formatted and repartitioned disk drives. Further information about data recovery is included in the Multimedia Appendix.

It is possible to use tools that implement a specific secure delete algorithm that ensures that the data cannot be recovered. The DOD 5220.22-M standard is a US Department of Defense standard providing specifications for clearing and sanitizing electronic data storage devices [29-31]. There are some commercial and freely available tools that implement that standard (eg, see [32]). It is not possible to extract the data from such drives. For drives that were not blank and that were not recoverable, we used a hexadecimal editor to read the patterns of data on the disk. Disks that have been wiped using this approach either have a single character (usually a zero) written to the disk or have a characteristic pattern of alternate ones and zeros followed by a random character written to the disk.

Data Coding and Analysis

All data from the recovered drives were stored on DVDs. A search of the files on each DVD was performed in order to isolate files that may have contained PI and PHI. The DVDs were searched for Microsoft Word documents, Excel documents, PowerPoint documents, Outlook files, Access database files, Adobe (PDF) documents, and text files. All discovered files were manually reviewed and a summary of the discovered PI and PHI was completed for each disk. An attempt was also made to identify the owner(s) of each disk drive from the information it contained.

PI was defined as identifying information (eg, name, address, social insurance number) about an individual or individuals *plus* other sensitive but nonmedical information (eg, financial information, personal correspondence, divorce documents, legal records). PHI was defined as identifying information about an individual or individuals *plus* any sensitive physical or mental health information. By this definition, if a drive only had a list of names and addresses but no sensitive information associated with them, then such data would not be considered PI or PHI. This definition is somewhat conservative because one can argue that a list of names and addresses suggests that all of these individuals were associated in some way; hence, association information would be revealed. Therefore, our results would be considered a lower bound on the prevalence of PI and PHI in the second-hand drives we analyzed.

Four ratings were made for each drive depending on the type of information it contained: (1) PI about the owner, (2) PI about other individuals apart from the owner, (3) PHI about the owner, and (4) PHI about other individuals apart from the owner. If a drive was clearly owned by multiple people (eg, members of a family), then they were all considered the owner in our coding. This means that, for example, if PI existed on any one of them, then the drive was considered to contain PI on the owner.

When considering whether information about the owner was really PI or not, we needed to decide what to do about *work products*. A work product is the output of an individual's professional or employment responsibilities. For example, a physician's prescription record would be considered a work product of the physician, irrespective of whether the patient is identifiable or not. The Federal Privacy Commissioner of Canada does not consider work products to be PI [33]. However, the European Commission had a different interpretation and considers information on and *relating to* an individual regardless of the position or capacity of the individual, such as a prescription record written by a physician, a communication of PI [34]. Given the uncertainty across jurisdictions, we treated information deemed to be work products as PI in one analysis and not PI in another analysis and present both sets of results.

Two independent individuals rated the drives. The first rater analyzed all of the 60 drives. The second rater analyzed a subset of the drives to ensure that the coding was reliable. Where there was disagreement, the two raters met to discuss their rationale and reach a consensus rating.

To determine how many drives the second rater needed to analyze, we performed a power analysis for using the Kappa statistic [35]. Given that there are no precedents for the interrater reliability of data extraction from second-hand drives, we relied on generally accepted benchmarks for Kappa values. Hartmann notes that Kappa values should exceed 0.6 [36]. Landis and Koch provide a more general benchmark, where values between 0.4 and 0.6 are considered moderate agreement [37]. A similar benchmark is provided by Altman [38]. Fleiss suggests that values between 0.4 and 0.75 represent intermediate to good agreement [39]. To err on the conservative side, we assumed that our value of Kappa would be at least 0.5, which would be considered a moderate level of agreement according to the above benchmarks. At that level of agreement, and 80% power to reject a null hypothesis comparing Kappa to agreement by chance, the second rater needed to code only 32 drives [40,41].

The final results are presented in terms of the percentage of disk drives containing PI and PHI on owners and other individuals, with the 95% confidence interval [42]. Interrater agreement is presented in terms of the Kappa statistic and its 95% confidence interval.

Special Protocols

Three special protocols were put in place for this study:

1. Some disk drives were expected to contain inappropriate/obscene material (eg, pornography). We therefore did not explicitly look through image files (file extensions .gif, .jpg, .psd, .tif, and .bmp). Also, one member of our research team initially screened the drives for files

and directories with suggestive names and flagged these particular drives as potentially containing such materials. No searching of files with suggestive names was done on flagged drives.

2. If any illegal materials were discovered (eg, child pornography), that information was to be passed on to the police.
3. If there were cases of disclosure of particularly sensitive PI or PHI for a large number of individuals, then they were to be reported to the appropriate (federal or provincial) privacy commissioner for follow-up.

Results

All of the 60 drives were from personal computers and ran the Windows operating system. Repartitioning and formatting are two common approaches for manipulating the drives. However,

as noted earlier, much data can be recovered despite these efforts. There were 35 drives that were repartitioned or formatted, and 5 that had had nothing done to them (all data were readily available). Therefore, data was potentially recoverable from 67% (95% CI: 54%-77%) of the drives.

A significant percentage of drives (28%; 95% CI: 19%-41%) were wiped using the DOD 5220.22-M standard. Three of the drives (5%) were completely blank and there had not been any data written to them. Five of the drives had pornographic files on them. Two of the drives were referred to the provincial privacy commissioner's office due to the sensitivity of the data that was found. No cases were referred to the police.

We were able to retrieve data from 39 drives (one of the repartitioned drives had no data on it). This represents 65% (95% CI: 52%-76%) of the total. Overall, we extracted 57 DVDs of data from the 39 drives.

Table 1. Contingency table with marginal totals and percentages showing the status of purchased drives distributed by province of purchase*

Province	Repartitioned	Formatted	DoD 5220.22-M Standard	Data Readily Available	Blank	Total
Ontario	19	11	5	4	3	42 (70%)
Quebec		5				5 (8%)
Alberta			12			12 (20%)
British Columbia				1		1 (2%)
Total	19 (32%)	16 (27%)	17 (28%)	5 (8%)	3 (5%)	60 (100%)

*For store chains, we considered the location of the specific store that we purchased from. The actual owners of the disk drives may be located in a different province or country. Four of the drives bought from Ontario belonged to US-based entities: 2 of them were state government departments, 1 was a municipal department, and 1 belonged to an individual.

There were 7 vendors in Ontario, 1 in Quebec, 3 in Alberta, and 1 in British Columbia. The distribution of drives by province is shown in Table 1. There are relatively more drives purchased from Ontario. Data were extracted from the drives from 6 of the Ontario vendors. All of the drives purchased from Alberta had been securely cleared such that no data were recoverable.

After our data analysis, we contacted the 3 Alberta vendors to understand why they all had used secure methods for deleting data on the drives they resell. They all stated that they had internal standard operating procedures for doing secure deletes

on the second-hand disk drives that they sell because they do not want to get involved in any data breaches.

We were able to determine the address of the drive owner for 26 of the 39 drives with recoverable data. All of the 26 disk drives came from urban areas (using the Canada Post and the Canadian Medical Association definitions of a rural postal code, which is only one of multiple possible definitions [43]). The owners for 22 of these 26 drives were in the same province as the vendor. The other 4 were US-owned but were sold by Ontario vendors.

Table 2. Claims made by the vendors of the drives from which we were able to extract data

Vendor Statement About Wiping Drives	Count
"Like new condition"	1
Verbally stated that the drives were formatted	1
"Recertified to factory settings"	1
None	5

The 8 vendors from whom we bought drives that had data on them did not actually make any claims that the data would be

removed in a secure way (see Table 2). Therefore, they were not in breach of any agreements that they had made.

Table 3. Percentage of drives with recoverable files and percentage of total drives with available personal data

	Owner PI (A)*	Owner PI (B)*	Other PI	Owner PHI	Other PHI
Percentage of Recovered (95% CI)	72% (28/39) (56%-83%)	62% (24/39) (46%-75%)	56% (22/39) (41%-71%)	13% (5/39) (6%-27%)	15% (6/39) (7%-28%)
Percentage of Total (95% CI)	47% (28/60) (35%-59%)	40% (24/60) (29%-53%)	37% (22/60) (26%-49%)	8% (5/60) (7%-24%)	10% (6/60) (5%-20%)
Kappa† (95% CI)	0.8 (0.6-1)	0.6 (0.33-0.88)	0.78 (0.54-1)	0.76 (0.45-1)	0.795 (0.52-1)

* (A) indicates that work products were considered as PI, and (B) indicates that work products were not considered as PI.

† Interrater agreement Kappa scores and their 95% confidence intervals.

A summary of the type of data that was uncovered in these drives is shown in [Table 3](#). All Kappa scores were above our threshold of moderate agreement. The vast majority of drives with data had PI about their owners (close to half of all drives) and about others. Examples of PI found on the disk drives included:

- personal budgets, salary information, tax returns and completed tax filing forms
- letters regarding personal relationships
- information on life insurance policies and inheritances
- payroll records of employees, including addresses, dates of birth, and social insurance numbers
- email correspondence regarding employees and their actions
- police record checks
- divorce documents

A considerable percentage of the total drives had PHI about the owners (8%) and other people apart from the owner (10%). The vast majority of that information was in correspondence (eg, Word documents, PDF documents, and email). Examples of PHI found on the disk drives included:

- psychological assessments of adults and children, correspondence related to custody cases involving children, affidavits, and social history of abuse victims
- medical certificates
- letters regarding alcohol addictions of other individuals (not the owners of the drive)
- reports from a registered nurse about other individuals' health problems, cases of abuse, children's health, and medication lists
- correspondence regarding the placement of adults and children in long-term care facilities

The 6 disk drives with PHI about other people came from 3 different vendors; half of them were for personal use and the other half belonged to organizations.

Discussion

We found that PI, including PHI, was recoverable from 65% of the drives we purchased; 8% of the disk drives had PHI of the owners, and 10% of all drives had PHI on other people. Half of the latter set came from organizations that were directly entrusted with that information and the other half from people working for such organizations (eg, a nurse who took some information home to work on it).

Approximately 8% of personal computers in use worldwide are second-hand machines [44], out of a total installed base of 980 million in 2007 [45]. To the extent that our findings are generalizable, an 8%-10% prevalence of PHI in second-hand disk drives in use paints a disturbing picture about the inappropriate disclosure of PHI. The second-hand computer market is expected to grow significantly in the next few years [46,47] and, with it, the opportunity for further such disclosure of PHI.

In a previous data remnants study done in the United States [24], 158 drives were bought. Of these, 129 were successfully imaged. Approximately 9% were wiped. It was possible to extract data from many of the remaining drives (38%). A similar international study provided the results presented here in [Table 4](#). PI was recoverable from 60% of North American drives, although the sample size was quite small. Based on data from the United Kingdom and Australia, the range of drives with PI is 18% to 49%. While our findings on PI are consistent with previous studies, giving them some face validity, previous studies did not consider PHI.

Table 4. Summary of findings from an international data remnants study [26]

	UK and Australia (2005)	UK (2006)	Australia (2006)	Germany (2006)	North America (2006)
Total Drives	116	200	53	40	24
Faulty Drives	13 (11%)	87 (43%)	3 (6%)	30 (75%)	12 (50%)
Wiped*	17 (16%)	55 (49%)	18 (36%)	4 (40%)	1 (8%)
Had PI*	51 (49%)	35 (31%)	9 (18%)	3 (30%)	7 (60%)

*The percentage of these disk drives that were not faulty.

Prevalence of PHI

Our results indicate that not as much health information is leaking as other types of information, such as financial and legal information. Why is relatively less PHI available electronically on these drives?

In Canada, the use of computers and the Internet is quite common. The majority of the population has access to a home computer [48], and most citizens have access to the Internet [49,50]. However, this does not mean that they have easy electronic access to their own PHI.

There are a number of ways that individuals can get electronic access to their own health data. For example, individuals may request their medical records from the institution that provided them with care. In practice, very few hospitals provide medical records electronically or make them accessible [51]. Another study found that a very small percentage of members of an integrated delivery system used eHealth services when provided to them [52].

It is more likely that PHI will exist in correspondence, such as email. The proportion of US Internet users who reported communicating over the Internet with their health care provider in 2005 was 10% [53]; a European survey found that 4% have approached their family doctor over the Internet, and about 7% of email users in the United States exchange emails with physicians or health professionals [54]. The proportion of physicians who report communicating by email with their patients varies from 3.6% to 24% [55-57]. About a quarter of patients correspond via email with family members [50]. PHI may also be exchanged electronically with peers [58]. This is consistent with our findings in that most of the PHI we found was in correspondence rather than in electronic medical records or database files.

One would expect that as electronic medical records become more widely deployed, more PHI will be available to patients electronically and hence the risk of inappropriate disclosure of PHI will increase over time. The disclosure risk is highest with care and service providers who would have extensive electronic correspondence with and documentation on many patients and clients on their work and personal computers.

Practices for Securing Data on Disk Drives

There is clearly a need for organizations and individuals, and certainly in Ontario and to some extent Quebec, to take actions to reduce the risk of personal data leaking from second-hand disk drives. A disk can leave its custodian in three ways: it is destroyed, it is given away, or it is lost or stolen.

The safest way to dispose of a disk drive is to properly destroy it. Approximately 38% of all used personal computers, including their drives, are destroyed [59]. While there are a large number of techniques that an individual or organization can potentially employ to destroy equipment [60], many of them require specialized equipment or resources and it is therefore not practical for most users to do it themselves. However, destruction of equipment can be outsourced to specialized vendors.

If equipment will be donated or resold, the risks of PHI leaks remain high. Donated equipment may end up in foreign second-hand markets, as demonstrated by a recent case of British computers ending up in Africa [61]. Approximately 6% of second-hand computers are exported [59]. If not exported, local resellers will not necessarily wipe data from the drives they acquire [62].

There are three general approaches that can be pursued to protect data on equipment before it is given away: de-identification of data, encryption of data, and the use of secure delete technology. Such approaches should be applied on all computers that will hold PHI, including the personal computers of staff and contractors who may take data home to work off-site.

Any PHI on a disk drive ought to be de-identified at the earliest opportunity. De-identification ensures that the risk of finding out the identity of the individuals about whom the data pertains is low [63]. This means removing or masking directly identifying data and applying other anonymization techniques to protect indirectly identifying data [64]. However, there will be many cases when data need to be identifiable to be useful. Hence, additional techniques would also need to be used.

Another way to protect data is by using encryption technology. Encryption can be used to create specific virtual drives, and all sensitive information can be stored on the virtual drives. Unless the password used is weak or the encryption algorithm is compromised, it would be extremely difficult to extract the information. However, this is generally not enough. Many programs will store their data, temporary files, cached files, backup files, and registry values outside the encrypted virtual drive. Quite a significant amount of information can be left in these files. Most users would not know to change the settings of their applications to only use the encrypted disk drive, and sometimes that option is not available. Therefore, if one really wants to protect data, this would probably not be the best approach unless one possesses a great deal of technical expertise (to change the setting of the applications to force them to use the encrypted drive).

The best encryption technology to use is whole disk encryption [65] that is invoked before the operating system, during system boot, starts to operate. This ensures that all data on the drive (temporary, backup, and data) are encrypted. Fortunately, this type of technology is becoming more generally available in common operating systems and hardware. Therefore, one would expect that, in the next few years it will be much more widely deployed and will significantly reduce the risks we have identified. Specific stand-alone products are listed by the Privacy Commissioner of Ontario in a fact sheet [65].

The second technology one can use is secure delete. This allows one to delete all of the data on the drive so that they are not recoverable (such as when using the DoD 5220.22-M standard). Secure delete by itself, however, is not enough. One needs to perform a more general disk wipe. Software for wiping disks usually performs a secure delete as well as removing all of the temporary, backup, and cached files from the system.

A recent study noted that commercial software for wiping disks tends to be quite unreliable [66]. In one case, the software did

not even attempt a secure delete because of a software bug. The difficulty with wiping software is that the program needs to determine where each application keeps its information. This is difficult to do for a very large number of applications that change often. It has been argued that because the market for privacy tools is small (and hence the vendors have limited resources), such vendors will not be able to keep up-to-date with the application and operating system changes [66]. Therefore, while the use of wiping software is reassuring, it may not actually be sufficient to protect personal data on disk drives.

Even if an organization does not resell or donate its equipment, theft and loss are real risks. For instance, a recent survey noted that 47% of organizations reported theft of a laptop or mobile computing device [67]. Some recent health care examples: (1) the theft of 2 laptop computers containing the names, birth dates, addresses, PHI, and insurance information for 3000-4000 patients was reported after a break-in at a rehabilitation clinic [68], (2) a laptop computer that contained 51 assessment reports was stolen from the car of a psycho-educational consultant working for the school board [69], and (3) a laptop computer containing data on 2900 patients participating in clinical trials was stolen from a researcher's car [70]. All of the techniques described above that are used to protect data when equipment is donated or sold should therefore be considered even if there is no intention to part with the equipment. One cannot control loss or theft events.

In summary then, it is best to properly destroy equipment when it is no longer in use. Even if that is not possible or desirable, it is still advisable to have full drive encryption to be activated as soon as the computers are purchased. With full drive encryption, there is minimal risk (unless the passwords used are weak) if the disk drive is given away, lost, or stolen at a later date.

Acknowledgments

This work was partially funded by the Office of the Privacy Commissioner of Canada through its contributions program and by Bell Information and Communications Technology Inc through an ORNEC (Ontario Research Network for e-Commerce) grant. This work was approved by the Research Ethics Board of the Children's Hospital of Eastern Ontario Research Institute. We thank Mary Ocampo for her help with coding the data on the disk drives.

Conflicts of Interest

None declared.

Multimedia Appendix

Details of the data recovery process [[PDF file \(Adobe Acrobat\), 49 KB-](#)]

References

1. Irving R. 2002 Report on Information Technology in Canadian Hospitals. Thornhill, ON: Canadian Healthcare Technology; 2003.
2. Healthcare Information and Management Systems Society (HIMSS). 15th Annual HIMSS Leadership Survey Sponsored by Superior Consultant Company. Healthcare CIO Results 2004.
3. 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](#)]
4. Bower A. The Diffusion and Value of Healthcare Information Technology. Pittsburgh, PA: RAND Corporation; 2004.

Limitations

In this study we only examined one source of leakage of PHI from data custodians. However, our results indicate that this is an important source of very sensitive PHI. As more information is stored and exchanged electronically, the risks from such leakage are bound to only increase unless current practices change.

We explicitly limited the study to Canadian vendors since this geographic location has not been studied before and because Canada has strong federal privacy laws. Hence, one would have expected that the ability to find PHI would be quite low—which was not the case.

The representativeness of the 60 drives of all drives in the second-hand market is a concern. We would argue that this sample underestimates the problem for a number of reasons. First, we excluded large disk drives, which eliminated data from servers. Servers would potentially contain large databases of PI and/or PHI. Second, some vendors (eg, those in rural locations) became suspicious of our motives for purchasing disk drives (“Why is someone from Ottawa buying a single drive from rural Alberta?”) and therefore refused to sell. We suspect that vendors with drives containing un-wiped PI and/or PHI were less likely to sell us the equipment. Therefore, our results should be considered a lower bound on the extent to which PI and PHI leaks through second-hand drives.

We did not specifically seek disk drives with health information on them. Had we done so, the PHI proportions would likely have been higher. However, that would not have provided a realistic assessment of the risk. Second-hand equipment vendors do not specialize by domain (ie, there were no vendors that specialize in selling only used equipment from health care facilities). Had we specifically requested equipment from health care institutions, it may have sounded like a suspicious request and dissuaded the vendors from completing the transaction.

5. Fonkych K, Taylor R. The State and Pattern of Health Information Technology Adoption. Pittsburgh, PA: RAND Corporation; 2005.
6. ; California HealthCare Foundation. Medical Privacy and Confidentiality Survey Summary and Overview. Oakland, CA: California HealthCare Foundation; 1999. URL: <http://www.chcf.org/documents/ihealth/survey.pdf> [accessed 2007 Sep 11] [WebCite Cache ID 5Q2YtO8yH]
7. ; Harris Interactive. Health information privacy (HIPAA) notices have improved public's confidence that their medical information is being handled properly. 2005. URL: <http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=894> [accessed 2007 Sep 11] [WebCite Cache ID 5Q2YQkeWM]
8. Grimes-Gruczka T, Gratzner C; The Institute for the Future. Ethics Survey of Consumer Attitudes about Health Web Sites. Oakland, CA: California HealthCare Foundation; 2000.
9. Willison DJ, Keshavjee K, Nair K, Goldsmith C, Holbrook AM; Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness investigators. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. *BMJ* 2003 Feb 15;326(7385):373 [FREE Full text] [Medline: 12586673] [doi: 10.1136/bmj.326.7385.373]
10. Mitchell E, Sullivan F. A descriptive feast but an evaluative famine: systematic review of published articles on primary care computing during 1980-97. *BMJ* 2001 Feb 3;322(7281):279-282 [FREE Full text] [Medline: 11157532] [doi: 10.1136/bmj.322.7281.279]
11. Cheng TL, Savageau JA, Sattler AL, Dewitt TG. Confidentiality in health care. A survey of knowledge, perceptions, and attitudes among high school students. *JAMA* 1993 Mar 17;269(11):1404-1407. [Medline: 8441216] [doi: 10.1001/jama.269.11.1404]
12. ; Association of American Physicians and Surgeons. New poll: doctors lie to protect patient privacy. 2001 Jul 31. AAPS Online. URL: <http://www.aapsonline.org/press/nrnewpoll.htm> [accessed 2007 Sep 11] [WebCite Cache ID 5NzvROMX2]
13. ; EKOS Research Associates. Rethinking the Information Highway. . 2003.
14. Saravamuttoo M. Privacy: changing attitudes in a tumultuous time. Sixth Annual Privacy and Security Workshop; November 3-4, 2005; Toronto, ON. .
15. Gostin LO, Turek-Brezina J, Powers M, Kozloff R. Privacy and security of health information in the emerging health care system. *Health Matrix Clevel* 1995;5(1):1-36. [Medline: 10141742]
16. Hodge JG, Gostin LO, Jacobson PD. Legal issues concerning electronic health information: privacy, quality, and liability. *JAMA* 1999 Oct 20;282(15):1466-1471. [Medline: 10535438] [doi: 10.1001/jama.282.15.1466]
17. Woodward B. The computer-based patient record and confidentiality. *N Engl J Med* 1995 Nov 23;333(21):1419-1422. [Medline: 7477127] [doi: 10.1056/NEJM199511233332112]
18. Goldman J. Testimony before the subcommittee on health of the committee on ways and means on "Patient Confidentiality." 1998 Mar 24. .
19. Leape LL, Bates DW, Cullen DJ, Cooper J, Demonaco HJ, Gallivan T, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA* 1995 Jul 5;274(1):35-43. [Medline: 7791256] [doi: 10.1001/jama.274.1.35]
20. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc* 2004;11(2):104-112 [FREE Full text] [Medline: 14633936] [doi: 10.1197/jamia.M1471]
21. Johnson N, Mant D, Jones L, Randall T. Use of computerised general practice data for population surveillance: comparative study of influenza data. *BMJ* 1991 Mar 30;302(6779):763-765 [FREE Full text] [Medline: 2021767]
22. Wilton R, Pennisi AJ. Evaluating the accuracy of transcribed computer-stored immunization data. *Pediatrics* 1994 Dec;94(6 Pt 1):902-906. [Medline: 7971009]
23. Davidson B, Lee Y, Wang R. Developing data production maps: meeting patient discharge data submission requirements. *Int J Healthcare Technol Manage* 2004;6(2):223-240. [doi: 10.1504/IJHTM.2004.004978]
24. Garfinkel S, Shilat A. Remembrance of data passed: a study of disk sanitization practices. *IEEE Secur Priv* 2003:17-27. [doi: 10.1109/MSECP.2003.1176992]
25. Jones A, Mee V, Meyler C, Gooch J. Analysis of data recovered from computer disks released for resale by organizations. *J Inform Warfare* 2005;4(2):45-53.
26. Jones A, Valli C, Sutherland I, Thomas P. The 2006 analysis of information remaining on disks offered for sale on the second hand market. *J Digit Forensics Secur Law* 2006;1(3):23-36.
27. ; Canadian Internet Policy and Public Interest Clinic. Approaches to Security Breach Notification: A White Paper. Ottawa, ON: Canadian Internet Policy and Public Interest Clinic; 2007. URL: http://www.cippic.ca/documents/bulletins/BreachNotification_9jan07-print.pdf [accessed 2007 Sep 11] [WebCite Cache ID 5QwUai54N]
28. GetData. Recover My Files. URL: <http://www.recovermyfiles.com/> [accessed 2007 Sep 11] [WebCite Cache ID 5QCzWP9C4]
29. ; Department of Defense - Department of Energy - Nuclear Regulatory Commission - Central Intelligence Agency. DoD 5220.22-M National Industrial Security Program Operating Manual (NISPOM). Washington, DC: US Government Printing Office; 1995. URL: <http://download.ehealthinformation.ca/cite/nispo-522022m-1995.pdf> [accessed 2007 Sep 11] [WebCite Cache ID 5QJEDu55S]

30. ; Department of Defense. DoD 5220.22-M: National Industrial Security Program Operating Manual (NISPOM). Washington, DC: Department of Defense; 2006. URL: <http://www.dss.mil/isp/odaa/documents/nispom2006-5220.pdf> [accessed 2007 Sep 11] [WebCite Cache ID 5QExtKegU]
31. ; Updated DSS Clearing and Sanitization Matrix. 2007. DSS. URL: http://www.dss.mil/isp/odaa/documents/clearing_and_sanitization_matrix.pdf[WebCite Cache ID 5QJEQOMOT]
32. Russinovich M. SDelete v.1.51. 2006 Nov 1. Microsoft Technet. URL: <http://www.microsoft.com/technet/sysinternals/Security/SDelete.msp> [accessed 2007 Sep 11] [WebCite Cache ID 5QeDFx2bn]
33. Radwanski G. PIPED Act Case Summary #15. 2001 Oct 2. Office of the Privacy Commissioner of Canada. URL: http://www.privcom.gc.ca/media/an/wn_011002_e.asp [accessed 2007 Sep 11] [WebCite Cache ID 5QCtVSGEC]
34. ; The Working Party on the Protection of Individuals With Regard to the Processing of Personal Data. Article 29 Data Protection Working Party. Opinion 4/2007 on the concept of personal data. Adopted on 20th June. 2007. European Commission. URL: http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf [accessed 2007 Sep 11] [WebCite Cache ID 5Q2YBu0CR]
35. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;XX(1):37-46. [doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)]
36. Hartmann DP. Considerations in the choice of interobserver reliability estimates. *J Appl Behav Anal* 1977;10(1):103-116 [FREE Full text]
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 Mar;33(1):159-174. [Medline: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)] [doi: [10.2307/2529310](https://doi.org/10.2307/2529310)]
38. Altman D. *Practical Statistics for Medical Research*. London, UK: Chapman and Hall; 1991.
39. Fleiss J. *Statistical Methods for Rates and Proportions*. Hoboken, NJ: Wiley; 1981.
40. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005 Mar;85(3):257-268. [Medline: [15733050](https://pubmed.ncbi.nlm.nih.gov/15733050/)]
41. Flack V, Afifi A, Lachenbruch P. Sample size determinations for the two rater kappa statistic. *Psychometrika* 1988;53(3):321-325. [doi: [10.1007/BF02294215](https://doi.org/10.1007/BF02294215)]
42. Agresti A, Coull B. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* 1998;52(2):119-126. [doi: [10.2307/2685469](https://doi.org/10.2307/2685469)]
43. Pong R, Pitblado J. Don't take geography for granted! Some methodological issues in measuring geographic distribution of physicians. *Can J Rural Med* 2001;6(2):103-112.
44. Escherich M. Thriving secondary PC market puts old PCs to good use. Gartner Inc. 2005.
45. Shiffler GIII. Forecast: PC installed base worldwide, 2003-2011. Gartner Inc. 2007.
46. Escherich M, Shao J, Maita K. Mature regions fuel supply of used PCs.. Gartner Inc - Dataquest. 2005.
47. Escherich M, Teng A. Thirst for technology drives used PC demand in emerging markets.. Gartner Inc. 2005.
48. ; Organization for Economic Development and Cooperation. OECD Key ICT indicators: households with access to a home computer. 2006. OECD. URL: <http://www.oecd.org/dataoecd/19/46/34083096.xls> [accessed 2007 Sep 11] [WebCite Cache ID 5QLh1C9bv]
49. ; Organization for Economic Development and Cooperation. OECD Key ICT indicators: households with access to the Internet in selected OECD countries. 2006. OECD. URL: <http://www.oecd.org/dataoecd/19/45/34083073.xls> [accessed 2007 Sep 11] [WebCite Cache ID 5QLgvVT4D]
50. Fox S, Fallows D. Internet health resources: health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access. Washington, DC: Pew Internet & American Life Project; 2003. URL: http://www.pewinternet.org/pdfs/PIP_Health_Report_July_2003.pdf[WebCite Cache ID Nba]
51. Fioriglio G, Szolovits P. Copy fees and patients' rights to obtain a copy of their medical records: from law to reality. *AMIA Annu Symp Proc* 2005:251-255 [FREE Full text] [Medline: [16779040](https://pubmed.ncbi.nlm.nih.gov/16779040/)]
52. Hsu J, Huang J, Kinsman J, Fireman B, Miller R, Selby J, et al. Use of e-Health services between 1999 and 2002: a growing digital divide. *J Am Med Inform Assoc* 2005;12(2):164-171 [FREE Full text] [Medline: [15561786](https://pubmed.ncbi.nlm.nih.gov/15561786/)] [doi: [10.1197/jamia.M1672](https://doi.org/10.1197/jamia.M1672)]
53. Beckjord EB, Finney Rutten LJ, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). *J Med Internet Res* 2007;9(3):e20 [FREE Full text] [Medline: [17627929](https://pubmed.ncbi.nlm.nih.gov/17627929/)] [doi: [10.2196/jmir.9.3.e20](https://doi.org/10.2196/jmir.9.3.e20)]
54. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. *BMC Public Health* 2007;7(1):53 [FREE Full text] [Medline: [17425798](https://pubmed.ncbi.nlm.nih.gov/17425798/)] [doi: [10.1186/1471-2458-7-53](https://doi.org/10.1186/1471-2458-7-53)]
55. Liebhaber AB, Grossman JM. Physicians slow to adopt patient e-mail. *Data Bull (Cent Stud Health Syst Change)* 2006 Sep(32):1-2. [Medline: [16995287](https://pubmed.ncbi.nlm.nih.gov/16995287/)]
56. Grant RW, Campbell EG, Gruen RL, Ferris TG, Blumenthal D. Prevalence of basic information technology use by U.S. physicians. *J Gen Intern Med* 2006 Nov;21(11):1150-1155. [Medline: [16879417](https://pubmed.ncbi.nlm.nih.gov/16879417/)] [doi: [10.1111/j.1525-1497.2006.00571.x](https://doi.org/10.1111/j.1525-1497.2006.00571.x)]
57. Brooks RG, Menachemi N. Physicians' use of email with patients: factors influencing electronic communication and adherence to best practices. *J Med Internet Res* 2006;8(1):e2 [FREE Full text] [Medline: [16585026](https://pubmed.ncbi.nlm.nih.gov/16585026/)] [doi: [10.2196/jmir.8.1.e2](https://doi.org/10.2196/jmir.8.1.e2)]

58. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ* 2004 May 15;328(7449):1166 [FREE Full text] [Medline: [15142921](#)] [doi: [10.1136/bmj.328.7449.1166](#)]
59. Escherich M, Smulders C. Market focus: worldwide secondary PC market sizing methodology. Gartner Inc. 2005.
60. ; Communications Security Establishment. Clearing and Declassifying Electronic Data Storage Devices. Ottawa, ON: Government of Canada; 2006. URL: <http://www.cse-cst.gc.ca/documents/publications/gov-pubs/itsg/itsg06.pdf> [accessed 2007 Sep 11] [WebCite Cache ID 50IbcPcnJ]
61. UK bank details sold in Nigeria. BBC News. 2006. URL: <http://news.bbc.co.uk/2/hi/business/4790293.stm> [accessed 2007 Sep 11] [WebCite Cache ID 5QczZlZLg]
62. Sullivan B. I just bought your hard drive. MSNBC News. 2006. URL: http://redtape.msnbc.com/2006/06/one_year_ago_ha.html [accessed 2007 Sep 11] [WebCite Cache ID 5QwUjM9pR]
63. 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] [Medline: [17213047](#)] [doi: [10.2196/jmir.8.4.e28](#)]
64. El Emam K, Jonker E, Sams S, Neri E, Neisa A, Gao T, et al. Pan-Canadian De-Identification Guidelines for Personal Health Information. Report Prepared for the Office of the Privacy Commissioner of Canada. 2007. URL: <http://www.ehealthinformation.ca/documents/OPCReportv11.pdf> [accessed 2007 Sep 11] [WebCite Cache ID 5Ow1Nko5C]
65. Cavoukian A. Office of the Privacy Commissioner of Ontario. Encrypting Personal Health Information on Mobile Devices. Fact Sheet. 2007 May. URL: http://www.ipc.on.ca/images/Resources/up-fact_12e.pdf [accessed 2007 Sep 11] [WebCite Cache ID 5Qd1NU27A]
66. Geiger M, Cranor L. Scrubbing stubborn data: an evaluation of counter-forensic privacy tools. *IEEE Secur Priv* 2006;4(5):16-25. [doi: [10.1109/MSP.2006.132](#)]
67. Gordon LA, Loeb MP, Lucyshyn W, Richardson R. Eleventh Annual CSI/FBI Computer Crime and Security Survey. Computer Security Institute. 2006.
68. Personal Health Information Protection Act Report HI-050047-1: A Physiotherapy and Rehabilitation Centre. Toronto, ON: Office of the Privacy Commissioner of Ontario. 2006. URL: http://www.ipc.on.ca/images/Findings/up-I_050047_1.pdf [accessed 2007 Sep 11] [WebCite Cache ID 5OF67opQF]
69. Personal Health Information Protection Act Report HI-050044-1: A Psychologist Working for a School Board. Toronto, ON: Office of the Privacy Commissioner of Ontario. 2006. URL: http://www.ipc.on.ca/images/Findings/up-HI_050044_1_Report.pdf [accessed 2007 Sep 11] [WebCite Cache ID 5OF7qHLuW]
70. Order HO-004. Toronto, ON: Office of the Privacy Commissioner of Ontario. 2007. URL: http://www.ipc.on.ca/images/Findings/up-3ho_004.pdf [accessed 2007 Sep 11] [WebCite Cache ID 5OFOzaj1O]

Abbreviations

PHI: personal health information

PI: personal information

submitted 25.05.07; peer-reviewed by P Abdel Malik, P Chhanabhai; comments to author 08.06.06; revised version received 12.08.07; accepted 06.09.07; published 30.09.07

Please cite as:

El Emam K, Neri E, Jonker E

An Evaluation of Personal Health Information Remnants in Second-Hand Personal Computer Disk Drives

J Med Internet Res 2007;9(3):e24

URL: <http://www.jmir.org/2007/3/e24/>

doi: [10.2196/jmir.9.3.e24](#)

PMID:

© Khaled El Emam, Emilio Neri, Elizabeth Jonker. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>, 30.09.2007). Except where otherwise noted, articles published in the Journal of Medical Internet Research are distributed under the terms of the Creative Commons Attribution License (<http://www.creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.