Working Safely with Electronic Protected Health Information in Research

Introduction and Scope

The scope of this document is primarily electronic protected health information (“ePHI”) in the context of research. Nevertheless, you may need to be concerned about other sensitive data types as well. Additionally, your specific contract/grant may require more protective measures than are described here.

What is PHI?

Protected Health Information (PHI) is individually identifiable health information, defined as information that:

- Is created or received by a health care provider, health plan, or health care clearinghouse;
- Relates to
  - The past, present or future physical or mental health or condition of an individual
  - The provision of health care to an individual
  - The past, present or future payment for the provision of health care to an individual; and
- Identifies the individual or information that the disclosing entity might reasonably believe could be used to identify the individual

PHI in all forms-verbal, written on paper, and electronic – is governed by HIPAA. It is both a HIPAA requirement and the policy of the University of Michigan Health System that the patient’s written authorization must be obtained before using or disclosing that patient’s PHI for research purpose unless a waiver has been obtained from the IRB.

More information on authorization and waivers for research is available at:

UMHS Policy 01-04-360, Use of Protected Health Information (PHI) in Research

What is ePHI?

To clear up any confusion, the terms “ePHI” and “PHI” need to be explained. As mentioned earlier, PHI is Protected Health Information. The “e” prefix in ePHI stands for “electronic”, so ePHI is “electronic Protected Health Information”, or PHI in electronic form.

Why Do I Care about PHI/ePHI?

First, you have the obvious ethical duty of respecting patient confidentiality. Failing to do so can also have a number of real-world consequences for researchers.

HIPAA/HITECH mandates reporting of ePHI breaches and data exposure, and can assess penalties and fines up to $1.5 million for each violation cited in connection with a single breach. Violations are associated with failure to properly secure data; however, failure to report the breach itself is also a violation. One breach can thus result in several violations, with aggregate penalties reaching several million dollars.

In addition, funding agencies typically require that breaches be reported to them as well. This could put future funding in jeopardy since it may bring into question a researcher’s ability to work with sensitive data. Furthermore, if you cannot show that your research data was not tampered with, it may be rendered worthless for the purposes of your research.

Examples of PHI

HIPAA defines PHI in a technical manner to include 18 specified “elements” or types of identifiers. For ease in understanding, below are some examples of what may be considered to be PHI. Note that this list is not all-inclusive, bit is shown here to give you an idea of what to look for.

- Patient names
- Street address, city, country, zip code
- Dates related to individuals (e.g. birthdate)
- Phone numbers
- Social Security Number
- Account numbers
- Patient admission date
- Patient discharge date
- Medical record number

- Patient number (facility assigned)
- Unique patient number
- ORs assigned
- Procedure date, date of service
- Full-face photographic images and any comparable images
- Biometric identifiers such as finger and voice prints
- Any other unique identifying number, characteristic, or code
- Carrier codes (insurance/HMO name)
- Health care professional ID
- Health care facility ID
- Fax number
- Health plan beneficiary numbers
- Email addresses
- Internet protocol address numbers (IP addresses)
- Web Universal Resource Locators (URLs)
- Device identifiers and serial numbers
- Certificate/license numbers
- Vehicle identification numbers and serial numbers

How Can PHI/ePHI Be Used for Research?

Uses and disclosures of PHI/ePHI for research purposes are subject to HIPAA's Minimum Necessary Use standard. Researchers should use and disclose only those data elements that are absolutely essential to the research being conducted. This means that, where possible, researchers should use/disclose either:

- Data that has been de-identified by HIPAA standards (preferred) or
- A limited data set (LDS) as defined by HIPAA in which a limited set of identifiers are used but for which it remains difficult to identify a particular patient.

Patient authorization is not required to use or disclose PHI to create either de-identified or limited data sets. HIPAA allows UMHS faculty and staff to create the data sets themselves or disclose the PHI to a Business Associate to create the data set. When PHI is given to a Business Associate to create the data set, an appropriate agreement approved by the Health System Legal Office ("HSLO") must be in place. See "What Agreements Are Necessary to Share Research Data" below.

For additional information on what data elements comprise de-identified and limited data sets, please see:

UMHS Policy 01-04-340, De-identification and Re-identification of Protected Health Information (PHI)

UMHS Policy 01-04-342, Limited Data Sets

For additional information on Business Associate Agreements, please see:

UMHS Policy 01-04-400, Business Associate Agreements

De-identification

De-identification transforms PHI so that it is no longer classified as protected. This requires either 1) removing 18 HIPAA-designated elements or 2) statistically de-identifying the data if a qualified statistician documents and attests that the information cannot be used to identify an individual.

If you can use de-identified data in your research, you will have many fewer concerns. This is the lowest risk form of data. Use it if you can. A security breach in which only de-identified data is involved does not have to be reported to HHS/OCR. However, you may still need to report it to sponsors, depending on the terms of your contract/grant. There may still be data integrity issues as well.

Limited Data Sets

A limited data set is similar to de-identified data; however, it may include the following identifiers:

- Address information
- Dates
- Unique identifying numbers

HIPAA allows limited data sets to be used/disclosed only for purposes of research, public health, or health care operations. Limited data sets may be shared/disclosed without subject/patient authorization, but this requires a data use agreement to be in place.

Don’t Guess if It’s PHI/ePHI

While different forms of data can lower your risk, it is unwise to attempt to make your own determination of whether data is PHI, a limited data set, or de-identified data. It is easy to get this wrong. For example, whether or not it is technically PHI will depend on how the data was collected.

Another problem often seen in practice is that researchers have declared data to be de-identified when in fact it is not. Very often this is found to be at least a limited data set, which is PHI, and is therefore subject to breach notification.

The best course of action is to consult with the IRB and UMHS Compliance to determine if the data that you are using for your research is PHI. If you make this determination yourself and guess wrong, you may be held personally responsible.

Recommendations

There are some best practices that you can follow in your research to limit your liability for data. You should always limit the data elements that you have to only those absolutely needed for your research. There is no reason to accept liability for data that you do not need, and most recently, the government has been penalizing entities for violations of the “minimum necessary” rule under HIPAA.

Within your study team, not all team members may need access to all the data elements. Limiting unnecessary access to additional data elements will reduce your risk.

Again, work with the IRB or UMHS Compliance to decide if what you have is full PHI, a limited data set, or de-identified data, and protect it accordingly. After this is determined, do not add data elements without consulting with the IRB or Compliance.

What Agreements are Necessary to Share Research Data?

If you are sharing or disclosing PHI outside of UMHS, you will need to have appropriate agreements in place.

- “Full” PHI (i.e. PHI that is not either de-identified or in the form of a limited data set) requires a data sharing agreement approved by the HSLO. This may be a Business Associate Agreement (“BAA”) or another type of data sharing agreement, depending on the purpose for which the full PHI is being shared.

- A limited data set requires a specific Data Use Agreement (DUA) that specifies how the limited data set may be used and that the limited data set is subject to HIPAA's breach notification and reporting requirements.
- A de-identified data set may require a UMHS agreement or attestation; this differs from the DUA used for limited data sets because de-identified data is not subject to HIPAA breach notification and reporting requirements.

Using Personal Systems or Storage

While personal systems are used in research, this is not a good idea for a number of reasons:

- Systems are lost or stolen.
- Systems become infected with malware.
- Sensitive data gets everywhere on your system.
- Your backups get filled with sensitive data.
- You have to clean all data from the system and backups when you are done. Some forms of storage can never be 100% cleaned.
- Cloud services you personally use can copy data-this is usually not permitted.
- In the event of an investigation, these personal systems may be forensically examined or come under the umbrella of a litigation hold.

If you are already doing this, there are steps that you can take:

- Utilities such as CCleaner are available for Mac and Windows for free (personal use) or at very low cost. This can clean out the nooks and crannies of your system.
  - It does mean that backups will have to be carefully pruned or re-created, depending on your system.
  - Some contracts/grants have specific requirements on how the data is destroyed.

Using Provided Systems

Using a provided rather than a personal system is preferred. However, since this system may come from a number of sources, here are some basic things to look for in such a system:

- Disk encryption-talk to MSIS or MCIT
- Current Antivirus-not 100% effective, but considered a “baseline”
- Systems should have a firewall
- Automated system updates, as applicable, or manual where necessary

- Deactivation of unnecessary services and disabling of unneeded ports
- Periodic vulnerability scanning with system patching to keep the system secure

Some Alternatives

There are some alternatives that can be used to reduce risk when doing research from a personal system.

VDI (Virtual Desktop Infrastructure) can be used to remotely connect to a computer with research data and tools. Very little data leaks to the connecting system.

Other options include using web applications to conduct research that can have a very small footprint on the connecting system. This can be good, but you must be extremely careful about viewing or exporting spreadsheets, PDFs, etc. Copies of these documents will remain where you view them.

Avoid DIY (Do It Yourself) Servers

Although it is often done, setting up a server in your lab is not recommended. In doing so, you will be taking on a number of responsibilities including maintenance, security, backups, as well as conducting security risk assessments mandated by HIPAA.

MSIS, MCIT, and ITS have options for systems with ePHI. Third-party vendors are a possibility, but you must get a Business Associate Agreement between them and the University. You must also exercise due diligence that they can keep your data secure via audit reports.

If you are already running your own server, here are some things to consider:

- If you are on a public IP address, it's particularly critical that you harden your system against attack.
- You are at lower risk if you are on the "UMHS Private" network.
- MSIS can advise you on locking your system down.
- You will be responsible for doing a self-assessment of risk and will be responsible for the reporting and consequences of a breach.

For more information contact MSIS at 734-763-7770 or msis_help@umich.edu.

Other Services

Many internal/external cloud services are NOT approved for ePHI. It's safer to assume that a service or vendor is not approved for use until you find out otherwise. There is a good summary at: Sensitive Data Guide to IT Services | Sensitive Data Guide

Note: not all MSIS/MCIT services are mentioned at this site, but they are still available for ePHI.

External Media

Using external media can be very dangerous if proper precautions are not taken. Lost media/devices are the cause of many HHS "wall of shame" entries. Here are some rules to follow:

- Don't use unencrypted thumb drives, external USB drives, CDs, DVS, flash card memory.
- Look for "FIPS 140-2" certified products.
- Properly encrypted (AES, 3DES) means "Safe Harbor"-we don't have to report it if it is lost or stolen.
- Use non-trivial passwords:
  - The longer the better, preferably above 9 characters
  - Mixed letters, numbers, and punctuation
  - Don't put a sticky note with the password on the device.
- Even encrypted, external media can be a vector for malware, so current anti-malware software is a must.
- Where possible, use department file shares or UMHS storage area networks.
- To share files, use the UMHS secure file transfer service, MiShare. See the information provided under "Transferring Files-MiShare".

Email

Be careful using email for ePHI. You could strongly encrypt attachments, but there are many ways that this can go wrong:

- You could forget to encrypt an attachment.
- You could attach the wrong file.
- The password you used to encrypt could be compromised.
- Once an email is out, you can't recall it.

- If a password is compromised, there is no way of securing the data when it is out of your control.
- Email stays in people's inboxes for years.

Email is only secure for email correspondence within the UMHS Outlook System to another UMHS user within the same system. These users can be identified within UMHS Outlook through the address book function in UMHS Outlook. UMHS Outlook users all have an "@med.umich.edu" account. If a user is outside of Outlook but part of the University, the user will have an "@umich.edu" account. Email from Outlook to an @umich.edu" account is not considered secure and should not be used to transmit ePHI.

If you need to send attachments to individuals outside of the UMHS Outlook system, use MiShare. See information on MiShare below.

Mobile Devices

Mobile devices are everywhere, but there are hazards in their use with research data:

- iPhone, iPad and Android devices are convenient, but watch out for syncing with cloud services. Example: If you view research data with ePHI on your iPad and have document syncing enabled, you may be copying data to Apple's servers. We don't currently have a BAA with Apple, so this is not compliant with HIPAA.
- The same is true if you are backing up your device to the cloud.
- Some apps will automatically sync with Dropbox or similar services. This is also a problem.

Encrypted Files

Encryption, when properly done, greatly reduces risk. But just because something is "encrypted" does not necessarily mean that it is sufficiently secured. Here are some things to be aware of:

- Many forms of file encryption are not secure.
- Zip native encryption isn't safe unless you get a zip utility that supports AES encryption.
- Some MS Word files can be strongly encrypted, but it depends on the version. Many "crackers" for these files are available on the Internet.
- Once a file is in an attacker's hands, they can run cracking software for weeks. They won't be "locked out" after too many bad tries. At this point it will be too late to pick a very strong password that can withstand this sort of an attack.

Meta Information

When exchanging files between team members or external collaborators, it can be straightforward to remove ePHI, once it is clear what that ePHI is. On the other hand, some types of files carry hidden ePHI information that is not immediately obvious, even if tools that are used to work with special file formats. This is found in picture files, MS Office documents, PDFs, videos, and special format files like DICOM files.

“Redacting” or removing this information can be difficult. Be very wary of editing files to remove ePHI since some can still remain, even if you do not see it. The common practice of drawing “boxes” over documents doesn’t always work. Sometimes it’s best to export the data into a flat text file and then import elsewhere. Seek technical help if you have any questions.

About P2P

Peer-to-Peer (P2P) software such as BitTorrent is common. It does have some legitimate uses. Of course, copyright issues can land you in trouble; enough said. But another way that you can get into trouble with P2P applications is that you may accidentally “share” research data without realizing it. Many clients will “auto start” so that you’re not aware of what /when you’re sharing.

Transferring Files-MiShare

Assuming that you’ve got all the necessary permissions to transfer data internally or externally, the preferred way of doing this is to use MiShare. This is the institutionally approved service for transferring sensitive data. This service has a number of good features:

- Transferring multi-gigabyte files internally/externally works just fine.
- Web interface and sftp supported
- Can be automated with scripts
- Is self-cleaning: files automatically disappear after a few days

This service may be found at https://mishare.med.umich.edu/

Safe Behaviors

- No matter what system you’re using, there are a number of safe behaviors that you can adopt to reduce the risk to your data:

- Don’t leave your system unlocked-use an auto lock in case you forget.
- Use a computer cable to physically secure your devices.
- Know about phishing-don’t click on links in emails.
- Beware of certificate warnings. This is a warning that you will get in your browser when there is a problem establishing an encrypted connection. Many times this is harmless, but you need to verify that this is the case with the service to which you are trying to connect.
- Avoid public wireless access points. If you must utilize them, use a Virtual Private Network (VPN) to access UM/UMHS resources. See https://wiki.umms.med.umich.edu/display/UMHSHELPDESK/VPN for information on installing the appropriate VPN client for your situation.
- Beware of “social engineering” attacks. These are essentially attempts to “con” you into revealing sensitive information, such as passwords, in a variety of ways including emails, phone calls, or even in person.
- Be careful with URL shorteners, such as TinyURL.com. These can be convenient, but they can be used to hide an attack since you can’t see the full URL. They can also be used with legitimate URLs to attack vulnerabilities in web applications.
- Be careful about using publicly accessible computers, such as kiosk computers at conferences and hotels. While using UMHS kiosks is safe, the use of other kiosks may put passwords at risk of being compromised. Furthermore, copies of any sensitive data that you access may remain on these systems.
- Don’t install any software unless you have a high level of trust for the people providing it. Some versions of well-known software have been infected and redistributed by a malicious third party.
- Consider using a browser plugin such as NoScript. Plugins such as this prevent the execution of JavaScript and potentially hazardous objects in web pages. This unfortunately “breaks” many web applications or sites until you mark them as “trusted”. This approach is typically for the dedicated and somewhat paranoid, but it can be an effective countermeasure.

Where to Get Technical Help

For sensitive regulated data, please see: M Safe Computing Sensitive Guide to IT Services: http://safecomputing.umich.edu/dataguide/.

Within the Medical School or NCRC

- Call the MSIS Solutions Center Service Desk (734) 763-7770
- Submit request or email msishelp@umich.edu