Unit 2

Ethical Management of Human Subjects Data

Source: These notes were developed in consultation with

Stephen W. Bernstein, Esq.
McDermott, Will & Emery
28 State Street
Boston, Massachusetts 02109

"It is not the fault of our doctors that the medical service of the community, as at present provided for, is a murderous absurdity ... To give a surgeon a pecuniary interest in cutting off your leg, is enough to make one despair of political humanity. And the more appalling the mutilation, the more the mutilator is paid. He who corrects the in-growing toe-nail receives a few shillings; he who cuts your insides out receives hundreds of guineas, except when he does it to a poor person for practice."

- George Bernard Shaw

A key aspect of data management is the protection of human subjects through the ***guarantee of confidentiality of the data***. The maintenance of confidentiality is a requirement of Human Subjects Review Boards, and it is often also a requirement of funding agencies. The enactment of the Health Insurance Portability and Accountability Act (HIPAA) in April of 2003 imposes explicit requirements for protection of privacy.

This unit is a discussion of HIPAA, strategies for the maintenance of data security, and an introduction to methods for defining unique and confidential study identification numbers (or strings).

Table of Contents

Learning Objectives

When you have finished this unit, you should be able to:

- Explain the main ideas of the April 2003 **Health Insurance Portability and Accountability Act (HIPAA);**
- Implement appropriate procedures for protecting the confidentiality of human subjects data that is recorded on paper forms;
- Set security features to your computer files (eg – password protection) that will protect the confidentiality of human subjects data that are stored in electronic files; and
- Explain the main ideas of the 1979 **Belmont Report, Ethical Principles and Guidelines for the Protection of Human Subjects of Research"**;

and have completed the University of Massachusetts/Amherst Office of Research and Engagement, Research Administration and Compliance CITI Training for Human Subjects Research (html)

- Specifically, you must completed (with a passing grade of at least 80%) either of two courses (CHOOSE ONE)
  - **Group 1: Biomedical Research Investigators and Key Personnel – Basic Course;** ***or***
  - **Group 2: Social Behavioral and Education Research Investigators and Key Personnel – Basic Course**

Glossary

Common Rule - A set of guidelines developed by the federal government for the review and conduct of research on human subjects. Its statement provides a framework for ethical research, generally, and the development of informed consent, specifically.

Covered Entity - A health care provider, health plan, payor, clearing house or any other entity that processes health data electronically.

HIPAA - Health Insurance Portability and Accountability Act

IRB - Institutional Review Board.

Limited Data Set - A data set comprised of facially de-identified information for research, public health and health care operations purposes.

PHI - Protected health information.

Research - Systematic investigation including research development, testing and evaluation, designed to develop or contribute to general reliable knowledge.

1. Introduction to HIPAA*

Dear class – Much of the material presented in this section was provided by Stephen W. Bernstein, Esq, of McDermitt, Will & Emery, 28 State Street, Boston, MA 02109. I confess to finding the "legal-ese" hard to follow in some spots. So I've done my best to clarify the wording where I can – cb.

The acronym **HIPAA** stands for the ***H***ealth ***I***nsurance ***P***ortability and ***A***ccountability ***A***ct, a law that was first enacted in 1996. Interestingly, ***portability*** was the original purpose. The legislation was to ensure smoother continuity of healthcare by providing standards for storing and reporting Health Information, so that as your health insurance changes, protected health information travels to new insurers.

The goals of HIPAA are now broader, in appreciation of the need for strong guarantees of protection of human subjects privacy. New standards went into effect on April 14, 2003; these have had a large impact on accessibility of data for research purposes. There are stricter regulations on access to, and handling and transport of ***P*rotected *H*ealth *I*nformation (PHI).** These, in turn, can have a large impact on design of a study (especially subject recruitment), and management of the data once it is collected.

HIPAA privacy standards directly affect:

- Ability of health care providers to obtain or use protected health information ("PHI") for research purposes (e.g., treatment-related research); and
- Access by researchers to PHI records and databases maintained by health care providers, health plans and health care clearinghouses and their business associates (e.g., access to academic medical center patient records or Medicare records)

Violation of HIPAA standards can be construed as evidence of negligence in health care. HIPAA provides for liability in case of violations – including fines and imprisonment.

---

*Source of much of material on HIPAA:

Stephen W. Bernstein, Esq.
McDermott, Will & Emery
28 State Street
Boston, Massachusetts 02109

HIPPA regulations directly impact allowed access to data, the ways in which data are stored and the ways in which data are shared.

Final Health Information Privacy Rule:

"A **Covered Entity** may not use or disclose protected health information, except as otherwise permitted or required."

Definition: covered entity. A covered entity is a healthcare provider, health plan, payor, clearing house or ***any other entity that processes health data electronically***. "Covered entities" include

- Health plans (e.g., Kaiser, Medicare, Health New England, …)
- Health care clearinghouses
- Health care providers who transmit health information in electronic form
- Indirectly - the business associates of Covered Entities that receive protected information
- Academic medical centers, teaching hospitals, clinics, physicians and other providers are covered entities when they:
  - Conduct research themselves
  - Disclose information to biotech/pharm companies that conduct research

UMass is considered a **"hybrid"** entity. Within the system, the University Health Services (UHS) is a covered entity; other parts of the University are not.

What is covered by HIPAA:

- Individually identifiable health information, known as **"Protected Health Information" (PHI)**
  - Relates to the past, present or future physical or mental health condition
  - Relates to the provision of health care
  - Past, present or future payment for the provision of health care

  And that

  - Identifies individual or could reasonably be used to identify individual

  And that

  - Has been transmitted or maintained in any form or medium (electronic, paper, oral)

Sometimes, individuals who are not members of a covered entity will still need to obtain certain authorizations. If a researcher is ***not*** part of a covered component, no additional approval is necessary to conduct research, but he or she does. . .

- need Institutional Review Board (IRB) approval for human subjects research which is federally funded, federally conducted, or subject to FDA jurisdiction
- may need **patient authorization or waiver of patient authorization** by **IRB** or **Privacy Board** to access PHI from Covered Entities (e.g., medical center) for research purposes
- ***need assurance of IRB and privacy compliance in order to publish research results***

Note! - HIPAA does not protect PHI once in the hands of researchers who are not covered providers

De-Identified information is not PHI.

Once all subject identifiers are removed from a database, it is considered to be "de-identified". Once approved as "de-identified", de-identified databases are not subject to HIPAA regulation, beyond adequate demonstration that the data are indeed fully de-identified.

1.1 Gaining Access to PHI for Research

Definition – research. In the ***Federal Policy for the Protection of Human Subjects*** (here) the "final rule" defines **"research "**as "a systematic investigation including research development, testing and evaluation, designed to develop or contribute to general reliable knowledge."

There are five pathways for permission to use PHI for ***research*** related purposes.

Access is solely for healthcare operations – this means for a specific entity or region, where the result of the "research" will be applied locally and will not result in a research publication, presentation and will not be shared beyond the institution or locality providing data.

- "Health care operations" include
  - protocol development
  - quality assurance
  - clinical guidelines and outcomes studies
  - population-based activities relating to improving health or reducing health care costs

If you are working for a hospital, state health department, or other covered entity and your project fits the above, then access to PHI **"for healthcare operations"** is allowed by your agency.

Design ..... **Data Collection** ..... **Data Management** ..... **Data Summarization** ..... **Statistical Analysis** ..... **Reporting**

1. **Access to PHI is Authorized by the subject**

   - Access to PHI is obtained after signed statement of permission by subject.

     - It is required ***in addition*** to IRB/Common Rule informed consent
     - Authorization exceptions are allowed for protocol development and decedents
     - Authorization includes full disclosure to subject of use and access to data.

2. **Access to PHI is Allowed by Waiver of Authorization Provided by IRB/Privacy Board**

   - In some instances, access to PHI is permitted without authorization. This is known as **Waiver of Authorization**. Determination of waiver status is by the institution's IRB/Privacy board. Waivers may be given for studies using only retrospective medical records or identifiable database research where Authorization is impracticable (no contact with study subjects, no impact on care of subjects is intended).

3. **Access is Allowed because De-Identification has been performed (recall: de-identified information is no longer PHI)**

   - Database research involving de-identified PHI is permitted if the information does not identify an individual ***and there is no reasonable basis to believe the information can be used to identify an individual.***

   - Adequate ***de-identification*** of PHI occurs in one of two ways:

     - #1 – A statistical expert determines and documents that the risk is very small that the information could be used to identify individual. Some of the issues considered are:

       - What are appropriate qualifications of statistical expert?
       - Feasibility of determining risk?
       - Risk of identification is rarely very small
       - Inherent identifiability of genetic materials?

- #2 – De-identification is accomplished by the removal of **all** of the following **18 specified identifiers**:

1 -	Names
2 -	All geographic subdivisions smaller than a State, including: street address, city, county, precinct, zip code, and their equivalent geocode, except for the initial three digits of a zip code, if according to the current publicly available data from the Bureau of the Census: (1) the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and (2) the initial three digits of a zip code for all such geographic units containing 20,000 or few people is changed to 000.
3 -	All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4 -	Phone numbers
5 -	Fax numbers
6 -	Electronic mail addresses
7 -	Social security numbers
8 -	Medical record numbers
9 -	Health plan beneficiary numbers
10 -	Account numbers
11 -	Certificate/License numbers
12 -	Vehicle identifiers and serial numbers, including license plat numbers
13 -	Device identifiers and serial numbers
14 -	Web Universal Resource Locators (URLs)
15 -	Internet Protocol (IP) address numbers
16 -	Biometric identifiers, including finger and voice prints
17 -	Full face photographic images and any comparable images
18 -	Any other unique identifying number, characteristic, or code (this does not mean the unique code assigned by the investigator to code the data).

Under de-identification, there can be no actual knowledge that the data can be combined with other information in such a way as to identify any individual.

4. **Access is Allowed because the Information Qualifies as a *Limited Data Set* and there is a *Data Use Agreement***
   PHI can be used for research purposes if it is a ***"limited data set"*** provided the rules of an accompanying ***"data use agreement"*** are followed.

   - **Definition: Limited data set.** A limited data set is comprised of **facially de-identified** information for research, public health and health care operations purposes

     - It can include zipcodes, geocodes, date of birth, date of admission/discharge/service, non-excluded identifiers
     - It excludes direct identifiers: name, postal address (other than state, city, precinct, zipcode, geocode), telephone #, fax #, email address, social security number, certificate #, license #, vehicle ID/serial number, URLs, IP address, full face or comparable images, medical record #, prescription #, health plan beneficiary #, account #, medical device identifiers and serial numbers, biometric identifiers, fingerprints, voiceprints

   - **Definition: Data Use Agreement.**
     A data use agreement

     - Defines who can use or receive data
     - Defines for what purpose the data may be used
     - The user agrees not to re-identify data or contact data subject
     - It provides adequate assurances that data will be safeguarded (similar to business associate agreement) and not used for unauthorized purposes
     - The user agrees to report improper uses and disclosures
     - The user agrees to "push-down" privacy protection obligations to subcontractors.

Issues in de-identification:

- It is problematic in relational databases (e.g., research using genetic database with clinical database) where identification is required for linkage of database components
- It may not be feasible in certain longitudinal studies (e.g., addition of new data on individuals requires identification)
- It is problematic in certain outcomes studies (e.g., inability to use date of event information other than year may undermine study)
- It may not be useful for epidemiological studies (e.g., dates may be needed to track disease), studies involving infants (e.g., might need date of birth), studies of environmental factors of disease (e.g., might need zip codes)
- ***Paradoxically***, de-identification may cause researchers to seek more PHI through waiver, authorization than if de-identification standards were less stringent

Standards for access to potential subjects and PHI of subjects are now conditional on approval of both IRB and Privacy Board. The Privacy Board of an institution may choose to give their authority to the IRB for research-related access to PHI. There are thus **2 key forms** that patients must sign to show they are fully informed participants in a research study:

- **#1. The Common Rule** (Informed Consent) applied by the Institutional or Human Subjects Review Board. It must contain full disclosure of purpose and procedures, risks and benefits of study, and rights as a participant.
- **#2. HIPAA Authorization**, applied by the Privacy Board. HIPAA authorization requires signed permission from participant to share PHI with researchers. Defines information to be shared; who will have access; purpose of use; limitations on use of PHI.

1.2 The Common Rule

Tip!!!! - The Common Rule below is your road map for developing an informed consent. The **Common Rule** is a set of guidelines developed by the federal government for the review and conduct of research on human subjects. Its statement provides a framework for ethical research, generally, and the development of informed consent, specifically.

The Common Rule (Informed Consent) – applied by IRB

- Understandable language
- No exculpatory language or release of investigator, sponsor or institution
- Statement that study involves research, explanation of purposes of research, expected duration of subject's participation, description of procedures, identification of experimental procedures
- Description of reasonably foreseeable risks and discomforts to subject
- Description of benefits to subject or others reasonably expected from research
- Disclosure of appropriate alternative treatment that might be advantageous
- Statement of extent to which confidentiality will be maintained
- Explanation of whether compensation will be paid and if injury occurs, whether treatment is available and where further information may be obtained
- Explanation of whom to contact about the research, the subject's rights, and any research related injury
- Statement that participation is voluntary; refusal to participate or discontinuance carries no penalty or loss of benefits
- Statement that the treatment or procedure may involve risks to the subject which are currently unforeseeable
- Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent
- Any additional costs to the subject that may result from participation in the study
- The consequences of the subject's decision to withdraw from the research and procedures for orderly termination
- A statement that significant new findings developed during the course of the research which may relate to the subject's willingness to continue participation will be provided to the subject
- The approximate number of subjects involved in the study

1.3 HIPAA Authorization

A **HIPAA Authorization** is a signed permission, by the human subject **(often the patient sitting in the waiting room avoiding the outdated Field and Stream magazine!!)**, that permits the disclosure of that person's PHI for the purposes that are spelled out in the authorization document itself. Like the Common Rule, the National Institutes of Health has established guidelines for the core elements of the authorization. Again, these are to ensure ethical conduct and the maintenance of human subjects protection.

HIPAA Authorization – applied by Privacy Board

- Description of information requested in specific and meaningful fashion
- Identification of person(s) authorized to make the requested use/disclosure
- Identification of person(s) authorized to receive request
- Description of each purpose of the requested use or disclosure; or "at the request of the individual" if initiated by the individual
- Expiration date or event that relates to the purpose of the use or disclosure or the individual
- Signature (or authorized representative's signature), date
- Individual's right to revoke and exceptions
- Information may be subject to re-disclosure and not protected by the federal privacy regulations
- Statement that will not condition treatment on providing authorization, except as permitted, and consequences of refusal to sign; condition permitted for research
- Plain language
- Right to inspect or copy PHI disclosed
- Right to refuse to authorize
- Right to signed copy of authorization

1.4 Issues in HIPAA Compliance

Compliance with HIPAA is an issue at several levels of health care delivery and public health research.

- How will clinical databases and tissue banks for research purposes be developed?
  - Users must obtain Authorization and/or Waiver of Authorization in order to access PHI
  - Research meets the definition of research; it is not treatment, payment or healthcare operations
  - Users can sometimes compile limited data sets with use agreement – ***but these are specific to each use***
  - Users can sometimes compile de-identified information, but the result limits the user's ability to link clinical data and tissue bank information.

- Will researchers be able to access and use clinical data compilations?
  - Users are required to obtain Waiver of Authorization from IRB/Privacy Board if the compiled database includes PHI
  - Sometimes, the user will be allowed to use "limited data set with data use agreement"
  - Users can use a de-identified database

- Can databases created for one purpose be used for future unspecified research projects?
  - Waivers of authorization are ***protocol specific***
  - Reliance on authorization can be problematic in practice for a variety of reasons:
    - It may be impractical to obtain from all subjects
    - There may be pre-existing consent exceptions
    - Authorization may have an expiration date or expiration event
      - Date: If expiration date or event is listed as "none", Authorization permits creation and maintenance of database in perpetuity
      - Event: An expiration event can be "end of research study" (or similar language) – allows researcher to meet record retention requirements
      - ***But does not authorize use for other research purposes***
    - The maintenance of integrity and scientific validity of the database may require updates of authorization that are not practical

HIPAA has significant implications.

HIPAA puts a large data management burden on organizations that maintain databases **to be able appropriately screen and de-identify data that is made available to researchers**.

Design ····· **Data Collection** ····· **Data Management** ····· **Data Summarization** ····· **Statistical Analysis** ····· **Reporting**

Sometimes, a waiver of authorization is given for the development of a research protocol. This raises the question: Where do the Authorization exception for development of a research protocol end and the need for authorization for research begin?

- Disclosure of PHI is permitted, without an Authorization or Waiver of Authorization, when its use pertains to developing a research protocol or for similar purposes preparatory to research.
  - An IRB or Privacy Board will still require a researcher to file a notification of use of PHI for activities preparatory to research. In particular, the
  - Researcher must represent in writing that:
    - Use or disclosure of PHI is sought solely for protocol/research purposes
    - The PHI for which access is sought is necessary for research purposes
    - PHI cannot be removed from premises during review
    - ***He or she will only record de-identified information*** (that is, can read and review PHI, but can only abstract and record de-identified data)
    - Use of data is restricted to developing hypotheses, protocol, characteristics of research cohort
- Under the research protocol exception, <u>can a researcher comb through medical records to identify potential research subjects</u>? Probably …. For example,
  - Clinicians may discuss a study with patients without authorization, discuss option of enrolling in study – without disclosing information on PHI to the researcher, until authorization (permission from patient) is given.
  - **However, clinicians' disclosure to a third party for purposes of recruitment requires authorization or waiver of authorization** (i.e., agreement of patient to pass their PHI to researcher for further contact)

- **What happens when a study participant withdraws his/her consent?** If an Authorization is revoked or expires after data has been included in an identifiable database, can the database continue to be used for research purposes? **Short answer: some of that person's data can be used; this is known as the "reliance" exception. The rest cannot.**
  - What uses (Reliance Exceptions) <u>are allowed</u> after withdrawal?
    - The researcher is allowed continued appropriate use to preserve the integrity of the research study, e.g., "to account for the individual's withdrawal from the study"
    - The researcher is not required to remove PHI from a completed database.
    - The researcher is allowed to continue to analyze already collected data.
    - The researcher is NOT allowed to disclose additional PHI after revocation

- The researcher is allowed to use previously collected PHI for purposes of: an FDA application; (2) investigation of scientific misconduct; or (3) reporting adverse events
- **Note! Upon expiration or revocation of authorization, the researcher is not allowed to use PHI for additional research purposes without a new Waiver of Authorization**

- **Is the participant who withdrew his/her consent allowed access to his/her records to determine whether he or she is in a placebo or treatment group?**

  - It depends.

    - Generally, he/she has right of access to his/her PHI upon (written) request.
    - Generally, withdrawal of consent (revocation) during course of research yields a re-instatement of right of access to PHI, because research is "complete" with respect to revoking patient.
    - A special exception to this right is temporary suspension of right of access until research is complete.
    - **Example of exception: Reinstatement of access to PHI may invalidate the blinding in a blinded study if patient is matched to PHI for access purposes.**

- What is the "**practicability**" standard and how will it be applied?
  note – Oh dear, I'm not sure I understand what Stephen Bernstein is saying here! cb.

  - This is an untested standard/uncertain, non-uniform results
  - Is "impracticality" a scaleable concept?

- **Under what circumstances is the researcher/user justified in NOT destroying identifiers?**

  - In general, the researcher/user is required to ***destroy*** identifiers at end of research
  - The requirement to destroy may be waived if ***justified***

    - Need for continued analysis/reanalysis
    - Need for future research – development of new protocol
    - For purposes of research misconduct investigations
    - Governmental or sponsor audits

Does an IRB/Privacy Board Waiver at one site suffice for purposes of conducting multi-site research?

- Probably not.
  - Sometimes, the researcher/user can obtain Waiver from any IRB or Privacy Board (no sponsorship or location requirement)
  - However, the researcher/user cannot disclose PHI to an unaffiliated IRB/Privacy Board, unless he/she has
    - (1) Authorization/ Waiver of Authorization, or
    - (2) limited data set/data use agreement
  - Typically, the other sites' can accept and then rely on the decision of another site's IRB/ Privacy Board. But it is also free to reject another site's decision.
  - Thus, waiver at one institution does not assure access to PHI of other institutions
  - Thus, there is a need for joint research cooperation agreements
  - Joint research cooperation agreements typically include specific acceptance of other site's waiver, or waiver specific to site

Can 'Covered Entities' provide reports containing PHI to privately sponsored patient registries? Can Covered Entities maintain patient registries?

Typically, no. However, there are exceptions where authorization may be given, as in the following:

- The production of state and federally mandated reports (e.g., state birth record; STD reporting)
- FDA-Type Reporting; eg -
  - The reporting is under the jurisdiction of the FDA, or the disclosure relates to FDA-regulated products or activities, and information relates to safety, effectiveness, or quality of the FDA-regulated product or activity, and
  - The reporting has "public health" purpose
  - Under these exceptional circumstances, however, minimum necessary standards apply
- The patient registry qualifies as a ***"business associate"*** to which PHI may be disclosed without patient Authorization for quality improvement purposes of the disclosing entity
- Registry can re-disclose on de-identified or limited data set basis
- Patient registries can receive limited data sets pursuant to a data use agreement
- Patient registries can receive de-identified data

- **What are the accounting requirements for research disclosures?**

  - No accounting is required for research conducted pursuant to an Authorization
  - No accounting is required for limited data set disclosure for research purposes
  - No accounting is required for disclosure of de-identified information for research purposes
  - Accounting is required if research conducted pursuant to a Waiver, UNLESS:

    - >50 subjects
    - research involves PHI of decedents
    - use of PHI is preparatory to research
    - "simplified" accounting for large database research

- ***TIP!* – *Be mindful of conflicts of reporting needs, authorization, and access to data***

  - Be aware of reporting requirements of all concerned parties
  - Do not make guarantees of access to data to a funding agency or research partner that

    - conflicts with guarantees given to clinical institution that is source of data
    - conflicts with allowable disclosure authorized by patient

    If funding agency requires access to database, this must be disclosed to patients in the authorization.

1.5 How should you proceed as a researcher?

- Identify and evaluate your sources of data

  - Who has the data that you need (internal & external), and how do you get it?
  - ***How can you make the source of data comfortable in sharing it with you*** (legally)?

- Assess what it will take for you to obtain the data

  - Learn where to access Research Authorization form for the institution

    - Obtain Authorization for research from patients when possible

  - Learn how to apply for a Waiver of Authorization for the institution
  - Use joint research cooperation agreements for multi-site projects
  - Develop limited data set policy and use agreement
  - Identify and contract with business associates

- Assess what obligations are imposed on you and your use of data:
    - From the source
    - From your own organization
    - From funding agencies
    - From oversight agencies
    - By law
    - By practice
- Adopt a de-identification policy
    - Consider feasibility of de-identification
        - We will address some of the issues in de-identification of data within a relational database in a later unit
    - Consider obtaining opinion of statistical expert

2. Security Procedures

2.1 General

Appropriate security of data, both on paper and on disk, means designing rules for managing and storing forms and databases to maintain confidentiality and comply with assurances given within the authorization and/or waiver of authorization to use data. This is particularly true of PHI.

For cybersecurity this includes:
Source: https://www.healthit.gov/sites/default/files/Top_10_Tips_for_Cybersecurity.pdf

This resource discusses the following 10 tips in detail

1. Establish a Security Culture
2. Protect mobile devices
3. Maintain good computer habits
4. Use a firewall
5. Install and maintain anti-virus software
6. Plan for the unexpected
7. Control access to PHI
8. Use strong passwords and change them regularly
9. Limit network access
10. Control physical access

For paper forms this includes:

- Rules for copying / storing / filing forms, such as:
    - Locked filing cabinets and/or locked rooms with limited access
    - No forms left on desk, table, in computer room where access is readily available to non-study personnel
    - Locking door or filing forms when leaving – even for a bathroom break
- Rules for mailing / faxing (faxing seems out of date now) info, such as:
    - **Do not fax forms with PHI** – cannot guarantee privacy
    - **Do not email PHI** – unless using HIPAA-standard secure email system, and permission within consent and authorization to do so.

- Mailed data forms must be addressed to clearly designated staff to ensure that the data are not opened and read by any non-authorized persons.

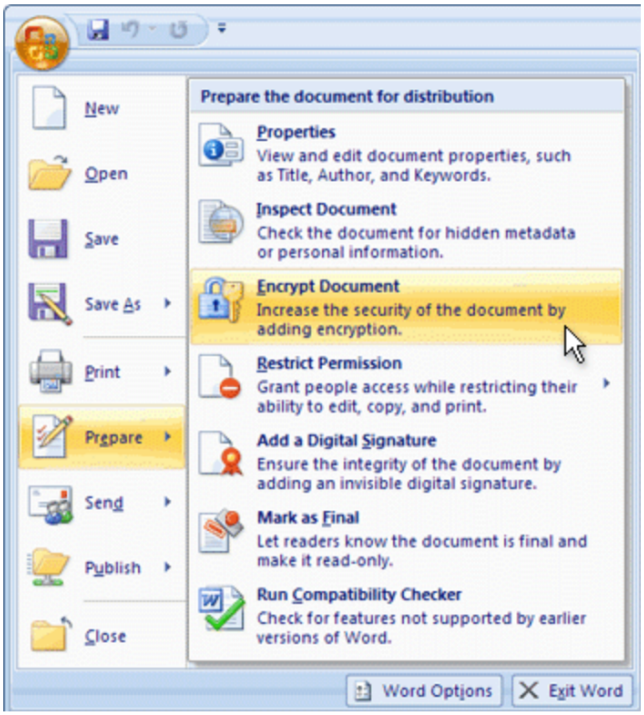
2.2 PC Users: How to Set up a User Account

Windows 10 Users
https://www.howtogeek.com/226540/how-to-create-a-new-local-user-account-in-windows-10/

2.3 PC Users: How to Password Protect a MS Word File

Step 1:

Click the Microsoft Office Button **, point to** Prepare**, and then click** Encrypt Document**.**



2.4 MAC Users: How to Set up a User Account

For mac OS Mojave 10.14 and **mac OS High Sierra**
https://support.apple.com/guide/mac-help/set-up-other-users-mtusr001/mac
Tip – Be sure you are logged in as administrator

Video (4:53) Link: https://www.youtube.com/watch?v=aj3G4xrzNm8

2.5 MAC Users: How to Password Protect a MS Word File

Word for MAC 2016
Source:
https://support.office.com/en-us/article/Password-protect-a-document-in-Word-for-Mac-5dc20870-62ea-43b1-ab0b-39426a57cff1

Word for MAC 2011
Source:
https://support.office.com/en-us/article/Password-protect-a-document-in-Word-for-Mac-5dc20870-62ea-43b1-ab0b-39426a57cff1#ID0EAABAAA=2011

2.6 Electronic File Protection –WinZip

In addition to the goal of file protection (discussed below), it's handy to be able to zip and unzip files anyway. Here is the link to WinZip for Windows, Mac and Mobile.
http://www.winzip.com/index.htm?sc_cid=go_us_b_search_wz_brand

WINZIP is an additional option for password protecting files that is available for download. It is software that can be used to create compressed files – or sets of files and folders – into a single **.zip** format file. It is extremely handy for saving space on a disk, or fitting large files or folders onto a disk for back-up storage. When "zipping" files it is possible to password protect the zip file, so that the files it contains cannot be accessed without the password.

Design ..... **Data Collection** ..... **Data Management** ..... **Data Summarization** ..... **Statistical Analysis** ..... **Reporting**

Unfortunately, many computer worms and viruses have been spread as **.zip** file email attachments – so that many servers do not accept emails with zip files attached, and it is no longer a valuable way to transmit data by email.

Within most programs there are options on level of security applied – the type of encryption used. If your database contains PHI, then go with the strongest level of encryption available.

2.7 Guidelines for Setting Passwords

* Source: Carnegie Mellon University Information Security Office – Guidelines for Password Management
https://www.cmu.edu/iso/governance/guidelines/password-management.html

Guidelines

The following are general recommendations for creating a Strong Password:

A Strong Password **should** -

- Be at least 8 characters in length
- Contain both upper and lowercase alphabetic characters (e.g. A-Z, a-z)
- Have at least one numerical character (e.g. 0-9)
- Have at least one special character (e.g. ~!@#$%^&*()_-+=)

A Strong Password **should not** -

- Spell a word or series of words that can be found in a standard dictionary
- Spell a word with a number added to the beginning and the end
- Be based on any personal information such as user id, family name, pet, birthday, etc.

These recommendations make "brute force" hacking efforts (that is, running through all possible combinations of letters, numbers and other keyboard characters) too time-consuming to be practical.

However it becomes important that you keep a record of your passwords (and a back-up!) – strange combinations are harder to memorize – just keep it separate from the files!

Silly Tip – Come up with a sentence that you can remember easily and build your password from this using the Carnegie Mellon guidelines on the previous page.
Example - "Did you walk to school or bring your lunch" →
The password might then be ***dYwtS#obYl***.

3. Identification (ID) Numbering Systems

Unique identification numbers for each subject are crucial for linking data forms to computerized data records and for linking data from multiple sources on the same subject. **HIPAA regulations have made it required that a unique, study-specific ID number be created for a research study** and that this number ***cannot be linked to any PHI***. This study-specific ID becomes the key linking variable for information on study subjects within a relational database.

ID variables, also known as KEY variables or key fields, are of two main types:

- **sequential numbering**
- **informative**

or, very often, a **combination of the two types**.

Informative ID numbers have information coded within the ID. An example is the subject CodeID used in the CPB Study. The first 4 letters of the subject's last name followed by the 2-digit year of birth defined the ID variable. For example, Ellen Smith born in 1927 would have CodeID = SMIT-27. Note that this type of ID is ***not acceptable as a de-identified*** ID variable, since it is derived directly from PHI.

This type of ID functions well for staff who are familiar with patients by name, and helps the staff in gathering data and filing forms appropriately. It is effective in decreasing the incidence of staff writing patient names on forms where they didn't belong – a potential violation of privacy and a common occurrence when clinical research staff wish to ensure that the information from the correct patient goes on the form.

The major drawback to this type of ID is that is cannot be guaranteed to be unique. In the CPB study, we had 3 sets of duplicates in close to 300 subjects. Joseph Smithfield born in 1927 will have the same ID as Ellen Smith, born in the same year. While he may not be in the hospital at the same time as Ellen Smith, this ID would not be useful for updating or merging computer files.

A sequential ID number is a simple counting number, starting from some designated value, and adding one for each new subject. Starting from 1 is reasonable and acceptable. I have found that leading zeros confuse some staff, so if a 3-digit number is needed, then one option is to start with 101 rather than 001. An example of a simple sequential ID is the LOGID in the CPB study, which began with 1001. This type of ID number is appropriate to serve as a KEY variable, and can remain in a de-identified database, as long as it is study specific and not entered into the patient medical record.

A good system for keeping track of assignment of sequential numbers is required so duplicates are not assigned. This can be a simple log, where sequential numbers are pre-listed, and subject information (name, medical record number, ...) is recorded as the subject in enrolled. A pre-printed list of numbers prevents duplicate assignment, as does pre-printing numbers on forms, or use of an automatic numbering system.

If an error is made, for example a subject listed twice (assigned to 2 numbers), simply cross out one of the two numbers, and ***do not re-use it***. Re-using tends to lead to confusion, and inclusion of forms combining data from two subjects. By crossing out some numbers, you will end up with gaps in the sequence, but that is not important -- ***the goal of an ID system is not to maintain a subject count, but to supply a unique ID for each subject***.

If survey forms are mailed or given out without assigned ID numbers attached, a sequential ID can be assigned as a form is returned and assigned to a batch for processing. Another option is to use an automatic numbering field available in some software. This assigns a unique, sequential ID number as a form is entered. The number assigned should then be written onto the form once the data is entered, to provide a direct link with the paper copy and the database record. This facilitates data cleaning.

Combination ID numbers are also common. In many studies with multiple data collection centers or sites, a unique site ID is assigned. This is used as the first digit(s) of a subject ID number, followed by sequential numbering within site.

For example in the study on cystic fibrosis related diabetes (CFRD) the pilot study was conducted at 5 sites. The University of Minnesota (UMinn) was site 1, Baystate Medical Center was site 2, and so on. Patients enrolled in the study at UMinn were assigned numbers 101, 102, 103, ... Patients enrolled at Baystate were assigned numbers 201, 202, 203, ... The first digit indicates site in this study. In the operations manual instructions might be written as:

Subject ID numbers (SID) are assigned using the format S*nn* where S is the site number, and *nn* a 2-digit sequential number, starting with 01, 02, ... at each site.

Site Name	Site Code (S)
U Minn	1
Baystate	2
Ohio State	3
...	...

I have often found it convenient to use a double ID system -- a combination of an informative ID such as the CODEID described above, with a sequential ID. This satisfies the needs of both the data collection staff (who no longer feel the need to write a patient's name on the edge of a confidential form) and in addition, provides a unique ID. When either ID fails to match when merging data files -- an alert to a possible error is made. Additionally, you can choose to enter only the sequential ID into the database, creating a de-identified database.

Use informative IDs with caution. Problems occur when using informative IDs if an error occurs in the information leading to the creation of the ID. Later correction could lead to a change in ID that would undermine the linking system between data files. For example, in the ANC soil ingestion study we needed a system for labeling food and fecal samples turned in each day by each child's family.

> Children were assigned a sequential ID -- in the range 01 to 99. The prefix 'A' was used to distinguish this study from other ongoing soil ingestion studies with samples also undergoing analysis in the same lab; the code 'N' for nutrition and 'F' for feces defined the sample type, and then the study day, from 01 to 07 indicated the day the sample was collected.
>
> In this manner, a child was identified as A*nn*, for example **A64**. The food sample from the first day of the study for that child was labeled **A64N01** (format A*nn*N*dd*). A fecal sample from this child on the third day was labeled **A64F03** (format A*nn*F*dd*).
>
> Problems arose when we checked into missing or mislabeled samples, corrected these and had to try to follow the corrections through the system -- chemical processing, lab books, lab reports, ... We found it easier to forget the meaning within the ID, and simply use it as a label. But this meant that I could no longer read sub-string information from the label to decide which sample I was dealing with.

Self-Validating ID Numbers. Loss of data due to failure to match information collected from different sources can be a real frustration in large longitudinal studies or studies using multiple data sources. One system used to minimize errors in ID numbers is to design a self-validating ID. This type of ID number includes a digit (or digits) that serve as a check. For example, in a setting where an ID is a combination of a 2-digit site code (SS), and a 3-digit sequential number (nnn), a validation digit can be defined as 1 if the sum of the digits is odd, 0 if the sum of the digits is even. That is, the ID takes the form SSnnnV, where V is 1 or 0. For example, the 24th subject at site 11 would have the ID: 110240. The final digit is 0 since 1+1+0+2+4=8 is even. The 5th subject at site 15 would have ID 150051.

This validation or check digit can help guard against some common errors, but doesn't help, for example with transposition of numbers (150501 in place of 150051).

Other more complex algorithms for defining check digits are possible, but can lead to frustration of staff and result in more errors, if the algorithms are too complicated and difficult to implement.

3.1 Maintaining Subject Confidentiality

Always create an ID number unique to the research study.

It is inappropriate, and in fact illegal to use a subject's social security number, phone number or hospital or clinic medical record number as a research study ID. If these numbers appear on any study reports or forms that are faxed or otherwise open to view by anyone outside of study personnel, this is a breach of confidentiality. A study ID should identify the subject only for purposes of the study. While other types of ID may be collected in order to link data from different sources (e.g., hospital medical record, clinic record, birth certificate registry) – once the files are entered and linked to the unique study ID, ***these other identifiers should not be kept in analysis data files***, to remain in compliance with HIPAA.

If subject PHI is kept in study files (e.g., for studies where subjects are re-contacted at follow-up periods), ***the link between study IDs and personal information should be kept in a separate file, accessible only to study personnel who require that information to contact subjects***. In all other communications regarding information on a subject, only the unique study numbers should be used (For example faxing a list to a study center requesting clarification of data values on subjects with LOGID numbers 1531 and 1479 is acceptable since the numbers have meaning only to those with access to the study materials. Faxing the same request using a medical record number would be a violation of patient confidentiality and HIPAA.)

Protecting confidentiality will also mean that you will often need to create a separate page or form for collecting patient contact information, and enter and store this data in a separate data file that is only accessible to limited staff. New laws protecting patient confidentiality are tightening up the rules on data access – attention to such details is required.