Marshfield Clinic

Research Data Protection

1. SCOPE

1.1. System-Wide

2. DEFINITIONS & EXPLANATIONS OF TERMS

2.1. **Data Encryption:**

- The conversion of data into a form, through use of an algorithm which renders electronic data, unusable, unreadable, or indecipherable by unauthorized persons. Decryption is the process of converting encrypted data back into its original form so that it can be usable and understood.

2.2. **De-identified Data:**

- Means data that does not identify an individual and there is no reasonable basis to believe that the information can be used to identify the individual. De-identified data will not contain any HIPAA identifiers.

2.3. Direct Patient Identifiers:

- Information (data) that can be used to readily identify the patient/subject. For subject protection they are considered by the IRB and HIPAA to be protected health information and are listed below:
  - ◊ Names;
  - ◊ Postal address that is more specific than the level of city or town (county, state, zip code etc. are NOT specific enough to be considered direct identifiers);
  - ◊ Telephone numbers;
  - ◊ Fax numbers;
  - ◊ Electronic mail address;
  - ◊ Social security numbers;
  - ◊ Medical record numbers;
  - ◊ Health plan beneficiary numbers;
  - ◊ Account numbers;
  - ◊ Certificate/license numbers;
  - ◊ Vehicle identifiers and serial numbers (including license plate numbers);
  - ◊ Device identifiers and serial numbers;
  - ◊ Web Universal Locators (URLs);
  - ◊ Internet Protocol (IP) address numbers;
  - ◊ Biometric identifiers (including finger and voice prints)

◊ Full face photographic images and any comparable images;

2.4. Limited Data Set:

- A set of data containing Protected Health Information (PHI) that excludes direct identifiers of the individual(s) for whom the data set represents, or any relatives, employer or household member of the individual.

2.5. Research Data Warehouse (RDW):

- A data repository containing clinical, claims and operational data obtained from the Marshfield Clinic System. This data is currently available for research purposes through Bioinformatics Research Center (BIRC) Research Analytics. This data repository is administratively segregated into four areas: General (see the following definition), PHI (containing direct patient identifiers), Admin (containing administrative, human resources and other employment-type data), and Work (containing temporary datasets loaded by RDW users for programmatic purposes.). For researchers without direct access to the RDW this data is made available through data queries to the BIRC who will abstract the desired data for the researcher.

3. RESOURCE GUIDE BODY

The protection of personal and confidential information of research participants is one of Marshfield Clinic's highest priorities. Marshfield Clinic conducts research compliant with all applicable state and federal laws and regulations. 45 CFR 46.111 requires that research studies involving human subjects include adequate provisions to protect the privacy interests of participants and to maintain the confidentiality of data. In its review of research, the MCRF IRB considers whether adequate provisions exist for the security of research data, in all its forms, before, during and after the research. Researchers must ensure that research data are protected in a manner consistent with human subject protection regulations and the HIPAA Privacy Rule. Researchers are responsible for ensuring that adequate controls are described in the materials submitted to the IRB, and are followed. A related policy, "Use and Disclosure of Protected Health Information in Research," focuses on data protection to meet the HIPAA Privacy Rule. This document focuses on acceptable strategies for protecting identifiable data during its capture, use, storage and transfer.

3.1. Access to and Collection of Data

a. Data for research is primarily collected through the course of health care delivery. Patient data is captured using the electronic health record (EHR) for the Marshfield Clinic's patient population.

   - Structured (coded) data collected in the EHR is transferred and integrated into Marshfield's enterprise data warehouse and then distributed to the research data warehouse (RDW) where it is made available for research approved by the IRB. The RDW is administratively segregated into four areas: PHI (containing direct patient identifiers), Admin (containing administrative, human resources and other employment-type data), General (stores Marshfield Clinic System operational and clinical data that has been stripped of any direct patient identifiers for use in generating limited data sets for research purposes), and Work (containing temporary datasets loaded by RDW users for programmatic purposes).

b. Additional data are available through review and abstraction of the electronic health record (EHR). CattailsMD, Marshfield Clinic's EHR, may be reviewed, after IRB approval, to validate electronic algorithms for disease states or to identify specific elements needed for adjudication or research. All access to the EHR is monitored via regular random and targeted audits.

c. Data unavailable by other means may be collected, after IRB approval, via patient survey or questionnaire.

3.2. Use and Storage of Research Data

a. Irrespective of the source or data collection method, specific mechanisms must be implemented to protect and limit access to datasets that may contain sensitive data.
b. Research data should be coded and identifiers removed by the PI or a research team member as soon as feasible, with a master list containing the identifiers secured and kept in separate file cabinets (paper records) or on a separate physical device (electronic data).
c. Marshfield Clinic's *Network Security Policy* includes a description of general methods in place to secure and maintain security, privacy and integrity of the Marshfield Clinic data network and all of the devices and data contained within.
d. Marshfield Clinic's policy *Portable or Mobile Devices* includes guidelines on the use and storage of data on portable devices. Because electronic portable devices are particularly susceptible to loss or theft, the storage of identifiable subject data, even encrypted files, should be limited and data transferred to a secure system as soon as possible. When not in use, electronic portable devices must be securely stored.
e. Access to the Research Data Warehouse (RDW) is via a password-protected logon script, and viewing of individual tables containing direct identifiers is limited to a small number of predefined users. This data is currently available for research purposes through BIRC Research Analytics.
f. Data that contains direct identifiers should be hosted on servers secured and housed in a HIPAA compliant environment. Currently an example of such a server at Marshfield Clinic is the DWSERV01 server, which is a server designated to store sensitive data extracted from the RDW for data requests. This security measure is distinct from the access-control measures related to the Research Data Warehouse.
g. Study specific electronic questionnaires and survey data should be stored as coded access or redcap files and held on user group protected network directories inside password secured database management systems.
h. Paper records must be kept in a secure location and only be accessible to appropriate study staff.

3.3. Secure Transfer of Research Data

a. Transfer of research data outside Marshfield Clinic is to follow the procedure, "Sharing and Transferring Research Data and Materials."

b. PHI should only be transmitted over secure networks, regardless of location, or as encrypted data files over public networks. PHI may not be transmitted via e-mail unless encrypted.

c. The intent to share data containing direct identifiers and the methods to be employed must be disclosed in the IRB application and approved by the IRB. This includes investigators leaving the employ of Marshfield Clinic who desire to take a copy of the data generated from their research.

3.4. IRB review of data protection measures

a. The material submitted to the IRB must include a description of how research data will be kept secure. The following should be addressed:

- How data will be collected and recorded
- Type of identifiers linked to the research data
- How data will be stored and secured
- At what point identifiers will be removed
- Description of the disposition of study data upon study completion
- Planned future uses of research data

b. The IRB will evaluate the protections in place based upon the level of identifiers, sensitivity of the data and risk of breach.

4. ADDITIONAL RESOURCES

4.1. References:
- None.

4.2. Supporting documents available:
- Policy #1611 – Network Security Policy
- Policy #4502 – Portable or Mobile Devices

5. DOCUMENT HISTORY

Version No.	Revision Description
1.0	New Document in Document Control system transferred from Policy & Handbook Library - #5289.0 (no changes made)
2.0	
3.0	

6. DOCUMENT PROPERTIES

Primary Author: Scheller, Lori A
Co-Author(s):
Approver(s): This document has been electronically signed and approved by: Ziemba, Steven J PHD on: 10/5/2015 2:02:22 PM