SPECTRUM HEALTH

Office of Research Integrity (ORI)
100 Michigan NE, MC 090
Grand Rapids, MI 49503
616.391.5800

System Privacy
100 Michigan NE, MC 843
Grand Rapids, MI 49503
616.486.4113
privacy@spectrumhealth.org

Protecting Health Information Privacy in Research

A Spectrum Health ORI and Privacy Guidance Document

Purpose

This document provides guidance on important steps to take to protect privacy when collecting, storing, analyzing, and sharing health information for research purposes.

Regulatory Guidance

45 CFR 46.111(a)(7) and 21 CFR 56.111(a)(7): In order to approve research …adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.

45 CFR 164.512(i): Standard: Uses and disclosures for research purposes

Discussion

The Spectrum Health orientation, online research ethics course (CITI), and privacy training all informed you of the importance of protecting the privacy of health information as a researcher. All researchers at Spectrum Health are subject to the organization's privacy policies and procedures, either as an employee, resident, student, medical staff member, faculty advisor, or other "workforce member." This document provides best practices on collecting, storing, analyzing, and sharing health information for research purposes.

Appendix A of this document lists the different types of research data.

Appendix B of this document lists the data points that are considered *identifiable* under HIPAA and therefore constitute Protected Health Information (PHI).

Collecting Research Data Using Microsoft Excel

- If you copy study data from the electronic medical record (EMR), immediately remove the patient name, medical record number (MRN) and all other direct identifiers (see **Appendix B**). Identify the individual data with a unique study subject number.
- The unique study number cannot be linked to or derived from the patient's medical record number (MRN) or other identifying number. For instance, the number cannot include the last four digits of the patient's Social Security Number, the patient's initials, or all or a portion of the MRN.
- Create an Answer Key/Correlation Tool (see **Appendix C**) to identify which unique study subject number. For example, Subject #1 = John Doe, DOB 2/10/75, MRN#2223890. Your research data set is then limited to subject number as the main identifiable item and the study data points. For example, Subject #1, Surgery date 2/1/15, comorbidities – diabetes, 2 lesions removed, etc.
- Store the Answer Key/Correlation Tool (i.e., identifying which study number corresponds to which patient) in a separate electronic document that is password protected and stored in a secure/limited access folder on the Spectrum Health network.

Collecting Research Data Using REDCap

- As you build each field in your project make sure to click the check box signifying that it is a PHI identifier.
- At the end of the Project Setup and before moving the project to Production, click on "Check For Identifier" link to perform a verification on all fields (located under the Online Designer button).
- After you are completely finished with extracting all statistical data from your REDCap project, but prior to marking the project as an "Archive" status, locate the direct PHI identifier fields and delete those fields. (You will likely need to go in to DRAFT mode to make this change)
- When you have finished deleting those fields, click on "Submit Changes for Review". The REDCap Admin will need to click on their email link to approve the changes (which will send you an email). This will signify that you can now Archive the project.

Avoid the following:

- Storing any research data on paper, a collection/abstraction sheet, or in a Microsoft Excel spreadsheet that *also* contains the patient name, address, phone number, and/or other identifying information. See the above guidance regarding creation of unique study numbers and a correlation tool.
- Storing the correlation tool in the exact same location as the research data, unless each electronic document has a unique password.
- Storing the correlation tool in the same Microsoft Excel workbook on different tabs/sheets.

Storing Research Data

- If possible, you must use REDCap as the database for storing study information. **Please visit the REDCap User Group InSite page here to learn more about using REDCap.**
- If you are unable to use REDCap (and the IRB approves your study without REDCap as your data collection & storage method), you must store study data on a Spectrum Health network folder to which your Spectrum Health IRB approved research team members have access.
- Secure your research data set and correlation tool each with a unique password. Store the research data on drives/folders assigned to your department.

Avoid the following:

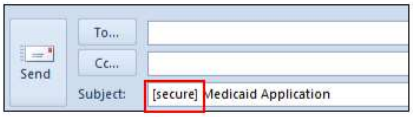
- Storing any research data in hard copies. If you need to print copies of data for analysis, the hard copies should be appropriately destroyed as soon as no longer needed.
- Storing any research data on a personal laptop, iPad, or mobile phone unless you have explicit approval to do so from the Spectrum Health IRB.
- Storing any research data on a computer network or email outside of Spectrum Health, unless approved by the Spectrum Health IRB. Outside networks are non-Spectrum Health owned and/or controlled computers/servers.
- Storing any data on a portable disc, key fob, or cloud storage.
- Using drives publicly available to all Spectrum Health employees (i.e., H drive).
- Keeping identifiable data after you have completed your study analysis. At that time, the correlation tool should be destroyed via deletion by the REDCap administrator(s), or deletion of the Excel or other document (if use was approved by the IRB).

Sharing Research Data for Analysis

- To send research data via email to a study team member or collaborator, use your Spectrum Health email address.
- Per Spectrum Health policy, emails containing any study data (even if de-identified in accordance with HIPAA) to be sent *outside the Spectrum Health network* must be sent securely. Visit the HIPAA & Privacy Matters InSite Page here for more information about how to send external email securely. The below screenshots also demonstrate three ways to send external email securely.

[secure]

- Type **[secure]** in brackets.
- Press the **Spacebar**.
- Type the subject in the subject line.

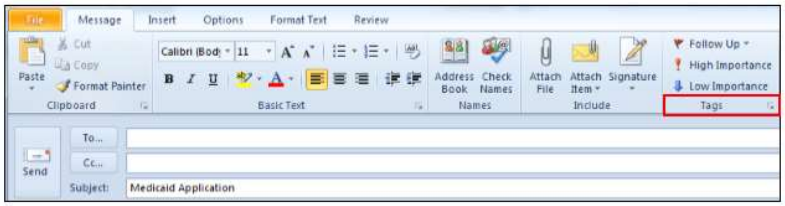


[marked for secure delivery]

- Type the subject in the subject line.
- Press the **Spacebar**.
- Type **[marked for secure delivery]** in brackets.

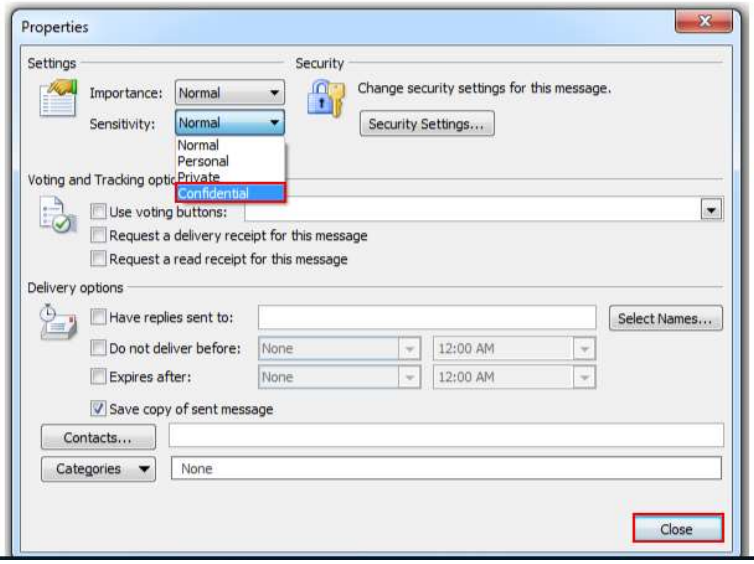
Confidential Setting

- Click the arrow in the bottom right-hand corner of the **Tags** section.



The **Properties** window appears.

- Click the dropdown arrow in the **Sensitivity** field.
- Select **Confidential**.
- Click the **Close** button.



- Please be aware that if the recipient of your research data is not a Spectrum Health employee, you may be required to have an agreement in place prior sending the data. **For assistance regarding necessary study agreement(s) contact researchassist@spectrumhealth.org.**

Avoid the following:

- Emailing any research data or research datasets/spreadsheets using your personal or university's email address (e.g., student@gmail.com, student@msu.edu).
- Emailing any research data to anyone outside of Spectrum Health unless you have been granted approval to do so by Spectrum Health. Failure to do so is a violation of Spectrum Health policy and may lead to corrective action. **Contact researchassist@spectrumhealth.org for more information about review and approval of research studies involving the use and transfer of health information to an outside organization or collaborator.**

Conclusions

If you have any further questions, please contact Spectrum Health Office of Research Integrity at 616-486-2060 or the Spectrum Health Privacy Team at 616-486-4113 or privacy@spectrumhealth.org. When emailing the Spectrum Health Privacy Team please indicate in the subject line if the question is research related.

Appendix A

Types of Research Data

It is important to distinguish the difference between different types of research data sets. Below are the definitions and corresponding examples.

- **Anonymous:** The research data cannot be traced back to the identity of the individual. To be considered anonymous research data it cannot have any number/code used to replace the patient name or have anything else unique included in the dataset that someone could use to determine the patient's identity (e.g. exact date of birth).

  Example: Exact age in years and for those >90 stating just >90, gender, disease

- **De-identified:** The research data can only be traced back to the identity of an individual with access to the correlation tool. De-identified also refers to the HIPAA standard for de-identification—that is, information for which all identifiers have been removed, or for which there is no reasonable basis to believe the information could be used to identify an individual. De-identified research data sets do not include dates or other indirect identifiers linked to the identity of the individual. See Appendix A.

  Example: Patient Code, Gender, Pathology Findings

- **Limited Data Set:** The research data includes indirect identifiers, as defined by HIPAA. These indirect identifiers typically include any dates tied to an individual's medical treatment (e.g., Date of Admission/Surgery) and could also include date of birth, city, state, and/or zip code. Sharing a limited data set with anyone not at Spectrum Health requires a legal document called a Data Use Agreement (DUA), which requires the recipient to keep the data private and secure and only use it for its intended purpose, i.e., for the purpose of conducting the approved research study.

  Example: Patient code, Date of Birth, Gender, Date of Surgery, Pathology Findings

- **Identifiable:** The research data contains data points in the dataset that would allow the identity of the patient to be discovered or known. A patient name (or even just patient initials), street address, phone number, or anything else personally tied to the individual and only that individual is considered an identifier (see Appendix A). An identifier enables someone to ascertain identity. A correlation tool must be used to keep readily identifiable information separate from the main data set.

  Example: Patient name or initials, MRN, DOB, Gender, Home Address, Surgery Date

Appendix B

List of 18 Identifiers

1. Names;
2. All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, 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 fewer 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 plate 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; and
18. Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data)

There are also additional standards and criteria to protect an individual's health information from re-identification. Any code used to replace the identifiers in datasets cannot be derived from any information related to the individual and the master codes, nor can the method to derive the codes be disclosed. For example, a subject's initials cannot be used to code their data because the initials are derived from their name. Additionally, the researcher must not have actual knowledge that the research subject could be re-identified from the remaining identifiers in the PHI used in the research study. In other words, the information would still be considered identifiable is there was a way to identify the individual even though all of the 18 identifiers were removed.

Appendix C

Sample Correlation Tool (Answer Key)

Subject#	Patient Name	Date of Birth	Patient MRN
001	John Doe	1/5/1970	1234567
002	Jane Doe	2/25/1962	2345678
003	Mike Smith	3/15/1980	3456789
004	Sue Smith	2/10/1942	4567890

Sample Research Data Set

Subject#	Age	Gender	BMI	Admit Date	Admit Reason	Blood Collection Date	HGB (GM/DL)	HCT %
001	45	M	23	1/3/15	MVA	1/3/15	14.0	38.0
002	53	F	28	1/4/15	CVA	1/4/15	12.0	46.0
003	35	M	18	1/4/15	GI	1/5/15	15.0	40.0
004	72	F	21	1/5/15	CHF	1/6/15	11.0	32.0