Protecting Patient Privacy in Research

A Spectrum Health IRB and Privacy & Security Guidance Document

Purpose

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

Regulatory Guidance

45 CFR 46.111(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

As a medical researcher, you have been informed during your Spectrum Health orientation, online research ethics course (CITI), and in your Health Information Portability and Accountability Act (HIPAA) training the importance of protecting patient privacy. All researchers at Spectrum Health are subject to the institution’s privacy policies and procedures, either as an employee, student researcher, faculty advisor, or other “workforce member.” This document provides best practices on collecting, storing and sharing patient health information used for research purposes.

Appendix A of this document lists the data points that are specifically considered identifiable under HIPAA and constitute Protected Health Information (PHI) relevant to this guidance.
Collecting the Research Data

Best Practice:

- If you have an employee at Spectrum Health run a report for you pulling the data from the electronic medical record (EMR) (e.g., Data Warehouse, Business Objects, Billing, Technology & Information Systems, Quality Department) immediately remove the patient name and medical record number (MRN) and replace it with a study subject number.
- Create an answer key (aka Correlation Tool – see Appendix B) to identify which unique study subject number goes with which patient. For example, Subject #1 = John Doe, DOB 2/10/75, MRN#2223890. Your research dataset then is limited to subject number as the main identifiable item and the study data points – see Appendix B. For example, Subject #1, Surgery date 2/1/15, comorbidities – diabetes, 2 lesions removed, etc.
- Store the answer key for which number goes with which patient either paper-based only in a file cabinet, or in a separate file folder ideally on a different drive location password protected.

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. Replace the patient name with a number that is not 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.
- Storing the answer key of which number goes with which patient in the exact same location as the research data unless each has a unique password.
- Storing the answer key in the same Microsoft Excel workbook on different tabs/sheets.

Electronically Storing the Research Data

Best Practice:

- Store the research data on a Spectrum Health computer or in a locked file cabinet to which your Spectrum Health IRB approved research team members and/or faculty advisor has access.
- Secure your research data set and answer key with a unique password. Search online how to password protect Microsoft Excel or Word to learn how.
- Store the research data on drives assigned to your department or your Spectrum Health faculty advisor’s department.

Avoid the following:

- 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 data on a portable disc, key fob, or in cloud storage unless you have explicit approval to do so from the Spectrum Health IRB.
- Using drives publicly available to all Spectrum Health employees (i.e., H drive).
- Storing any research data on your home or university’s computers, personal computers, or on your university’s campus servers.
• Keeping identifiable data after you have completed your analysis and been accepted for publication. At that time, the answer key to the identity of the patient should be destroyed via placing in one of the Spectrum Health approved Rapid Shred locked containers or deleting it from the drive.

Sharing the Research Data for Analysis
Best Practice:
• Use your Spectrum Health email address assigned to you as part of your orientation to send research data.
• Email research data to your statistician, faculty advisor and IRB approved research team members Spectrum Health assigned email address only.
• Contact Spectrum Health’s Privacy Security Team at privacy@spectrumhealth.org prior sending any research data information to anyone without a Spectrum Health email address. They will tell you how to send the email and research data per Spectrum Health policy, i.e., encrypted and “secure.” You can also search on the Spectrum Health’s internal website (InSite) “sending secure email,” for instructions. Please be aware if the recipient is not a Spectrum Health employee/provider or GRMEP employee then a data use agreement must be in place prior sending the data to be compliant with HIPAA regulations. Email researchassist@spectrumhealth.org for assistance with a data use agreement.

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 or Grand Rapids Medical Education Partners (GRMEP is a supported entity of Spectrum Health) unless you have been granted approval to do so by the Spectrum Health IRB. Failure to do so is a violation of Spectrum Health policy and may lead to dismissal from the Spectrum Health research program, if you are a student, or other consequences, if you are an employee or member of the workforce.

Types of Research Data
When communicating with the IRB, your research team, and collaborators on your research study, it is important to distinguish the difference between different types of research data sets. Below are the Spectrum Health IRB 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 answer key of which number belongs to which patient identity. De-identified also refers to information for which all patient identifiers have been removed, in accordance with HIPAA regulations. De-identified research data sets do not include dates linked to the identity of the individual.

  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 has to be in place to ensure the receiver agrees to keep the data private and secure and only use it for its intended purpose.

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

• **Identified:** 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. This type of “all in one” research data set cannot be used for retrospective chart reviews conducted under a waiver of HIPAA authorization. An answer key (aka correlation tool) which keeps the subject readily identifiable information separate from the main data set must be used as a condition of granting the waiver.

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

**Conclusions**

Please contact Spectrum Health Office of the IRB at 616-486-2031 or at irbassist@spectrumhealth.org or the Spectrum Health Privacy & Security Team at 616-486-4113 or at privacy@spectrumhealth.org if you have any further questions.
Appendix A

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 individual's privacy 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.
### Sample Correlation Tool (Answer Key)

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

### Sample Research Data Set

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