

# HIPAA and Protected Health Information (PHI)

---

## Definition

**Protected Health Information (PHI)** refers to any information within a medical record or a designated record set that can identify an individual **and** that is created, used, or disclosed during the provision of healthcare services such as diagnosis or treatment.

Under the **Health Insurance Portability and Accountability Act (HIPAA)**, researchers may access and use PHI when it is necessary for research purposes. However, HIPAA applies **only** when the research **creates, uses, or discloses PHI** that is part of a medical record or healthcare service, such as treatment, payment, or operations.

---

## Examples of PHI Usage in Research

- Retrospective chart reviews using existing medical records.
  - Clinical trials that diagnose health conditions or evaluate new drugs or devices, where data is added to the patient's medical record.
  - Sponsored studies that submit health data to the U.S. Food and Drug Administration (FDA).
- 

## FERPA and Student Health Records

Researchers should be aware that **student health records** at postsecondary institutions that receive funding from the U.S. Department of Education (DoED) are considered “**education records**” under the **Family Educational Rights and Privacy Act (FERPA)**.

- **Subject to FERPA:** Student health records from campus services such as University Health Services (UHS) and the Optometry Clinic.

- **Subject to HIPAA:** Non-student health records and other clinical data associated with treatment.
- 

## What is Not Considered PHI

Some research may involve **personally identifiable health-related information** that is *not* considered PHI because:

- It is not related to a healthcare service (treatment, payment, operations).
- It is not recorded in or derived from a medical record.
- It remains within the researcher's personal research files.

Such information is referred to as **Research Health Information (RHI)** and is **not subject to HIPAA**, although **human subjects protections still apply**.

### Examples of Research Using Only RHI

- Aggregated (non-individual) data analyses.
- Diagnostic tests where results are not entered into medical records or disclosed to participants.
- De-identified genetic studies for exploring markers or promoter elements (not linked to care).

In contrast, **genetic testing done for diagnosis or treatment purposes** *is* considered PHI and falls under HIPAA regulations.

---

## Health Information Without Identifiers

Health data without any of HIPAA's 18 identifiers is **not considered PHI**.

**Example:**

- A dataset of anonymous vital signs = *not PHI*.
  - The same dataset + medical record numbers = *PHI*.
- 

## The 18 HIPAA Identifiers

Any one of the following makes a dataset PHI under HIPAA:

1. Names
2. All geographical subdivisions smaller than a State (e.g., street address, city, ZIP code with fewer than 20,000 people)
3. All elements of dates (except year) directly related to an individual, including birthdate, admission/discharge dates, death date; and all ages over 89
4. Phone numbers
5. Fax numbers
6. Email addresses
7. Social Security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate or license numbers
12. Vehicle identifiers and serial numbers, including license plates
13. Device identifiers and serial numbers
14. Web URLs
15. IP address numbers

16. Biometric identifiers (e.g., finger/voice prints)
17. Full-face photographs and comparable images
18. Any other unique identifying number, characteristic, or code

*(Note: This does not include study-assigned random subject IDs as long as the linking key is securely stored.)*

---

## Protections Against Re-identification

- Codes used to replace identifiers must **not be derived from personal information**, such as names or initials.
- The **method to derive codes must be confidential**.
- Researchers **must not have actual knowledge** that the individual could still be identified based on the remaining information.

**Key Point:** Even if all 18 identifiers are removed, the dataset is still considered identifiable if **any method exists** to link the data back to the individual.

-