Extracting value from HIPAA Data

James Yaple
Jackson-Hannah LLC

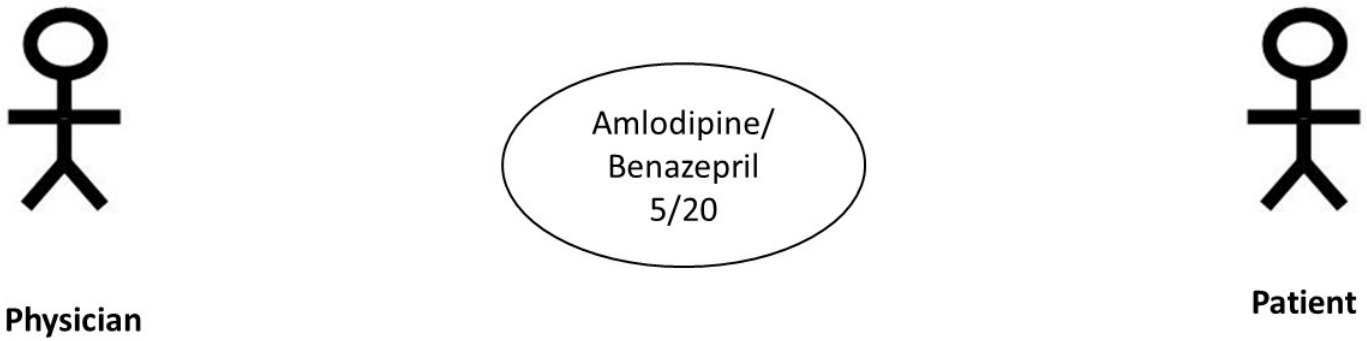
Session Objectives

- Examine the value of realistic information in research and software testing
- Explore the challenges of de-identifying health data in accordance with HIPAA
- Identify the 18 data elements that must be de-identified and the value they represent
- Compare and contrast the two standard methods for de-identifying health data in accordance with HIPAA

HIPAA/PHI

- Health Insurance Portability and Accountability Act of 1996 (HIPAA)
- PHI - Individually identifiable health information
- HIPAA Rules apply to **covered entities and business associates**. Covered entities must comply with the Rules' requirements to protect the privacy and security of health information and must provide individuals with certain rights with respect to their health information.

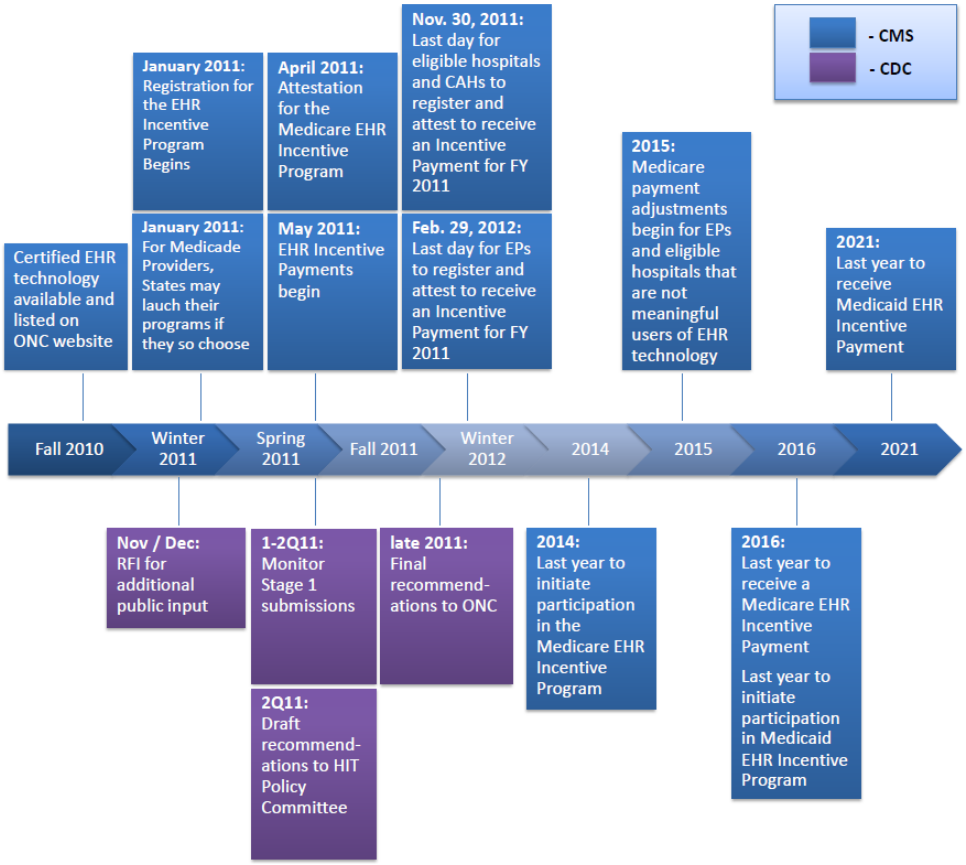
When does information become PHI?



EHR Meaningful Use (the challenge)

- Use EHR to:
  - Improve quality, safety, efficiency, and reduce health disparities.
  - Engage patients and family.
  - Improve care coordination, and population and public health.
- Stage 1 – Data capture and sharing
- Stage 2 – Advanced clinical processes
- Stage 3 – Improved outcomes

Meaningful Use Timeline [cms.gov]



Veterans Health Information Systems & Technology Architecture - VistA

- Department of Veterans Affairs introduced in 1996
- Open source through Open Source Electronic Health Records Association (OSEHRA) (2011).
- VistA is the largest EHR in the world. It was developed by the U.S. Department of Veterans Affairs. VistA is in production today at hundreds of healthcare facilities across the country from small outpatient clinics to large medical centers. The software is currently used by the Indian Health Service and a number of other healthcare organizations around the world.
- Most VistA instances represent a single location (limited sharing).

Million Veteran Program - MVP

- The Veterans Affairs (VA) Office of Research and Development is launching the Million Veteran Program (MVP to better understand how genes affect health and wellness. MVP will establish one of the largest databases of genetic, military exposure, lifestyle, and health information. Research findings based on MVP may lead to new ways of preventing and treating illnesses in Veterans.

- MVP has extensive safeguards in-place to ensure information security and patient confidentiality are top priorities.

Realistic data for software testing

- When you're testing a web application, you're as much testing the data as testing the application behavior.
- Using production data will ensure that what you are testing will be as close as possible to the actual behavior once the feature is released to production users.
- Data needs to be regularly refreshed to accommodate new use cases.
- If your test data is old, the validity of testing is reduced.
- Net result is bad test data yields poor functional results and may affect your largest business asset without realizing.

Black Market Value of EHR/PHI and other data

- "The value of personal financial and health records is two or three times [the value of financial information alone], because there's so many more opportunities for fraud," said David Dimond, chief technology officer of EMC Healthcare.
- Cyber criminals are selling the information on the black market at a rate of $50 for each partial EHR, compared to $1 for a stolen social security number or credit card number. EHR can then be used to file fraudulent insurance claims, obtain prescription medication, and advance identity theft. [Medcape]

Risks to keeping data de-identified

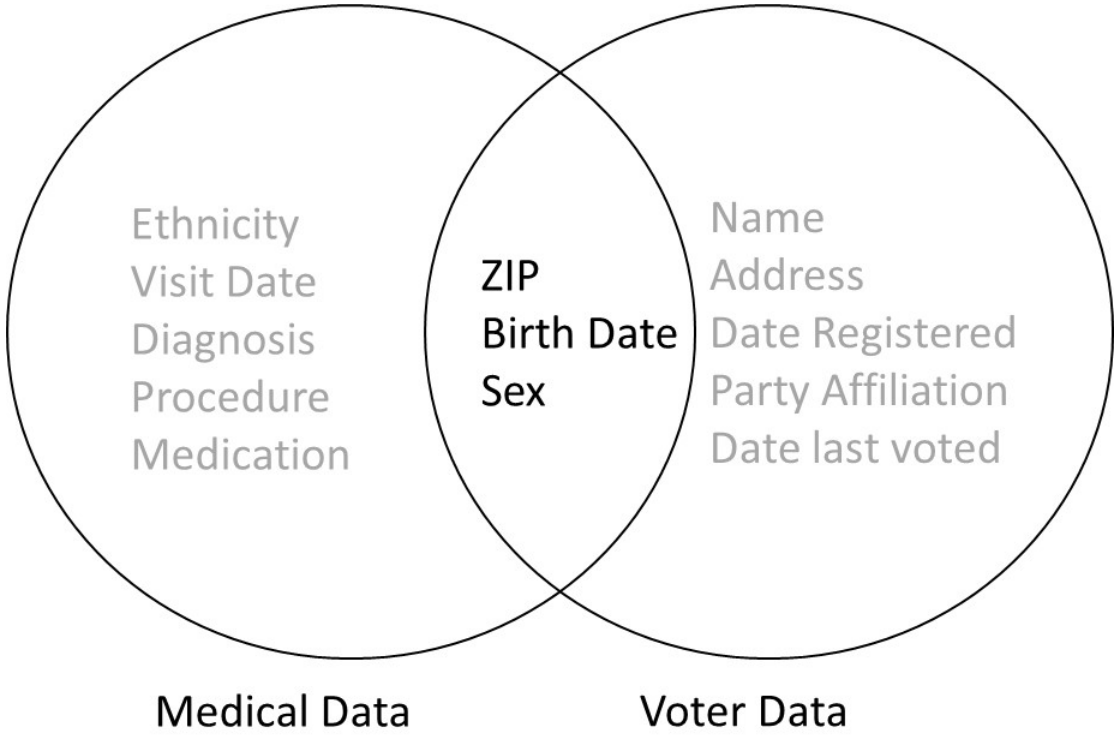
- Release of records with very unique characteristics (e.g., unusual occupation or very high salary or age)
- External availability of records with matching data elements which can be used to link with the de-identified information and identify individuals (e.g., voter registration records or driver's license records).

Risk Assessment for re-identification

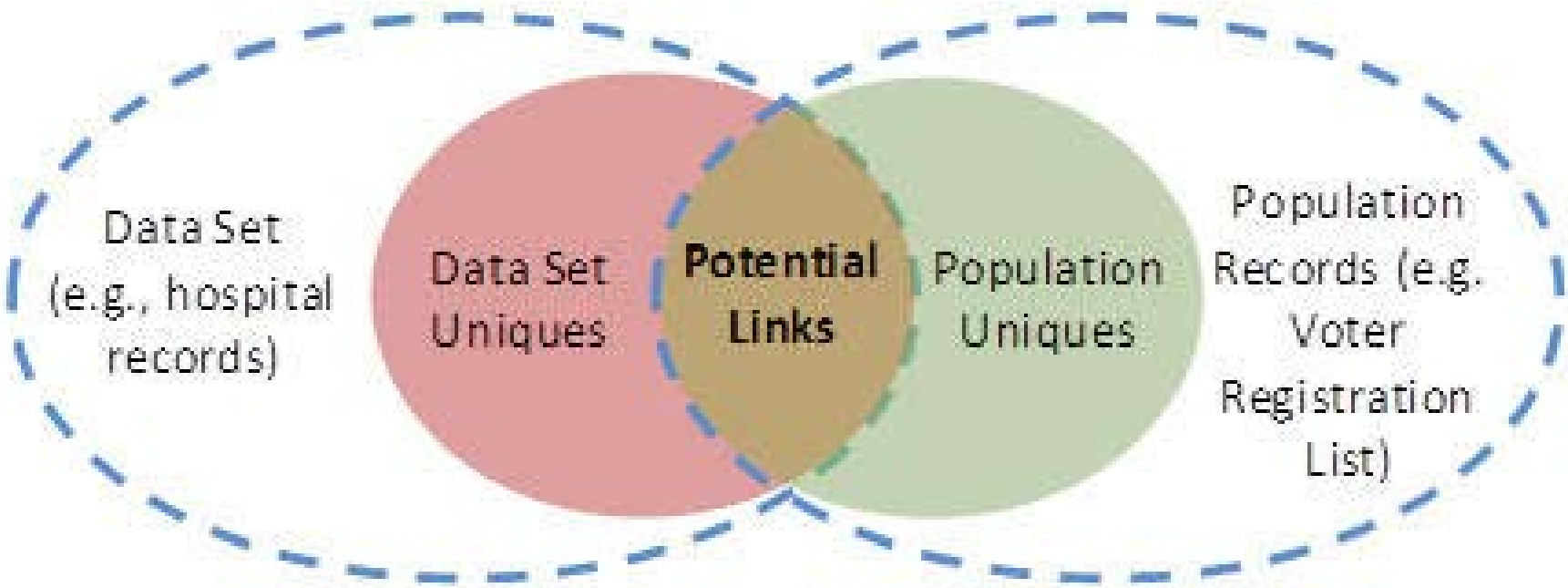
- Assess risk according to probability it will consistently occur in relation to the individual.
- Determine external data sources containing the patient identifiers and the replicable features in the health information.
- Determine the extent to which the subject's data can be distinguished in the health information.
- The greater the replicability, availability, and distinguishability of the health information, the greater the risk for identification.
- Example comparison: Birth date=high, blood glucose result=low.

The William Weld example (c. 2002)

William Weld was governor of Massachusetts. Insurance company provided medical data to researchers. This was matched against voter registration roles (linking data). The de-identified data (pre-HIPAA) was re-identified



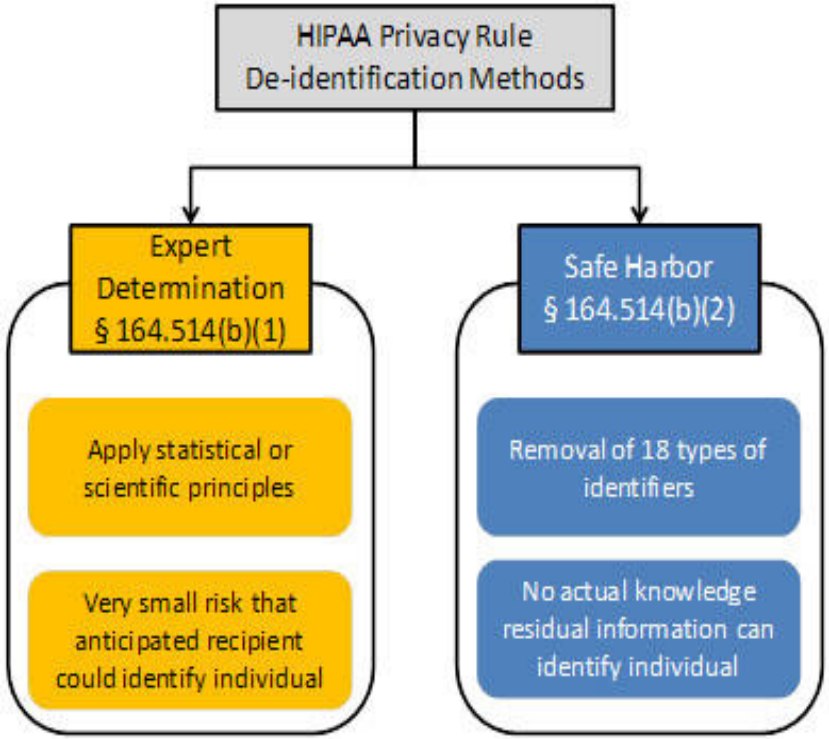
HHS depiction of matching (linking) (2012)



HHS definitions

§ 164.514

- **Other requirements relating to uses and disclosures of protected health information.**
  (a) *Standard: de-identification of protected health information.* Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information.



Expert Determination

164.514(b)(1)

- A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
- (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
- (ii) Documents the methods and results of the analysis that justify such determination;

Experts – 2012 Guidance

164.514(b)(1)

- Who is an expert? no specific professional degree or program
- Expiration – certification can be time-limited
- Multiple solutions – de-identification from same source data
- Methods: Suppression, Generalize/Ranges, Perturbation
- Re-Identification code; unique (hash) identifier to keep records linked

De-identification – Safe Harbor

164.514(b)(2)

- Specific identifiers of the individual or of relatives, employers, or household members of the individual, are removed:

- Examples:
  - Year of birth, gender, 3-digit zip code will match approximately 0.04% of residents in the United States
  - Full DOB, Gender, and 5-Digit ZIP Code will match over 50% of residents in the United States.

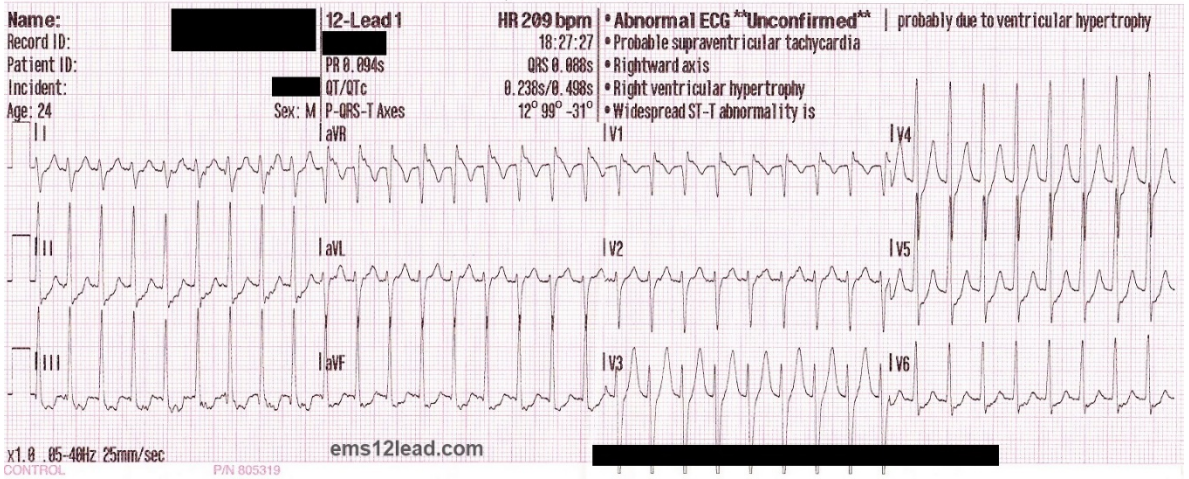
18 HIPAA Identifiers

- Names
- Address (Zip if > 20K)
- Dates (except year) related to an individual, including birth, admission, discharge, death
- Telephone numbers
- Fax numbers
- Email addresses
- 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 Resource Locators (URLs)
- Internet Protocol (IP) addresses
- Biometric identifiers, including finger and voice prints
- Full-face photographs and comparable images
- Any other unique identifying number, or code

Data Elements and Analytic Value

- Of all the HIPAA identifiers that provide clinical value, dates and address information are the most valuable.
- Admission and discharge dates can indicate effectiveness of treatments.
- Address information can show environmental factors. Short ZIP (3-digit) sufficient
- Analog data problematic to de-identify and maintain value.

Complications – Images and Recordings



Complications – Caretaker notes

```
       TITLE: PC ACUTE CARE VISIT
DATE OF NOTE: FEB 04, 2000@11:18     ENTRY DATE: FEB 04, 2000@11:20
      AUTHOR:                      EXP COSIGNER:
     URGENCY:                           STATUS: COMPLETED

Chief Complaint: Patient notes 1 month history of blurred vision and
frequent urination

HISTORY OF PRESENT ILLNESS:

DEMO,FATHER is a 44 year-old MALE  who presents complaining of blurred
vision for the past 1 month. He finds it is difficult for him to read
clearly and is even effecting his driving.  He also notes that he has be
getting up to the bathroom frequently, esp. at night.  He now routinely
to urinate 3-4 times a night.  He is not aware of any particular weight
loss, but does feel thirsty much of the time.
--------------------------------------------------------------------------
PAST MEDICAL HISTORY:
Illnesses: Hypertension

Surgeries: None

Allergies: PENICILLINS

Medications:
1)    HYDROCHLOROTHIAZIDE 25MG TAB**  Qty: 45    ACTIVE
        for 90 days  Sig: TAKE ONE-HALF TABLET  Refills: 0
        MOUTH EVERY MORNING THC BLOOD PRESSURE
2)    METOPROLOL 25MG XL TAB  Qty: 90 for 90     ACTIVE
        days  Sig: TAKE ONE TABLET MOUTH QDAY    Refills: 0
        FOR THE HEART
--------------------------------------------------------------------------
FAMILY HISTORY:
Diabetes  Father, Sibling, Grandparent
```

NAME: SPRIGGS,KENNETH S
UNIT #: 1302620
ACCT #: 2020126
DOB: 12/05/75
DATE: 05/20/98

MODE OF ARRIVAL: Private vehicle.

CHIEF COMPLAINT: Abdominal pain.

HISTORY OF PRESENT ILLNESS: The patient is a 22-year-old male who had an acute onset of abdominal pain about 4 p.m. and began vomiting about 1-2 hours prior to admission. He estimates he has vomited 6 times. He states the vomiting decreases the abdominal pain but it is still present. He describes the pain as feeling as though somebody with a knife tried to cut their way out of his abdomen. The patient's pain waxes and wanes. Sometimes it is very severe and other times it is less so. When I asked him if he would describe this as crampy, he agreed that that was an appropriate description of the quality of pain he was having. The pain is increased by movement. He has not noted increasing or decreasing factors.

PAST MEDICAL HISTORY: The patient has asthma.

CURRENT MEDICATIONS: Proventil, Intal and Azmacort. He has never taken oral steroids for his asthma. He has no known drug allergies.

SOCIAL HISTORY: He is a CSU student. He is a nonsmoker. He says he drinks alcohol about twice per year. He has not had alcohol in the past few days.

REVIEW OF SYSTEMS: GENERAL: The patient felt well prior to the onset of the pain. He ate lunch today. He has not eaten any dinner and is not anorexic at this time. He has had some child, but doesn't know whether he has had a fever or not. GI: He has had no diarrhea. He denies hematemesis. He denies a history of peptic ulcer disease or abdominal surgery.

Key points in de-identification/value

- Continuous process for constantly changing data. (old data not good)
- Match address data to provider (use hospital 3-digit zip)
- Use year of birth.
- Create unique (hashed) identifier to keep records linked.
  - Protect identifier hash to prevent re-identification (destroy?)
- Remove data (analog) with embedded metadata that can't be de-identified.
- Assess changes in data from analog to digital.
- Provide an expert determination along with safe harbor.

Summary

- There is significant market value in EHR/HIPAA data to legitimate (MU, research, software testing) and illegal actors.
- Some EHR/HIPAA data is difficult to de-identify.
- HHS considers removal of 18 data elements as a “safe harbor” method of de-identification.
- Assess risk based on uniqueness of data.
- Characteristics of data (analog/digital) changing.
- Unpredictable release of linking data.
- Know the rules. Penalties are severe.

Questions?

Reference URLs on next slide

James.Yaple@Jackson-Hannah.com

Reference URLs

- http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html
- https://iapp.org/media/pdf/knowledge_center/Re-Identification_of_Welds_Medical_Information.pdf
- https://epic.org/privacy/reidentification/Sweeney_Article.pdf
- http://www.ehealth.va.gov/VistA_Monograph.asp
- http://www.research.va.gov/mvp/
- http://www.medscape.com/viewarticle/824192
- https://fcsm.sites.usa.gov/files/2014/04/spwp22.pdf