HIPAA and Big Data
Twenty Third National HIPAA Summit

March 17, 2015
Mitchell W. Granberg, Optum Chief Privacy Officer

Overview
HIPAA and Big Data

- Big Data Definitions
- Big Data and Health Care
- Benefits and Risks

Big Data – Definitions

- **Big data** is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and information privacy. Source: Wikipedia (OED is similar).

- **The amount of data is exploding:**
  - For $600, can buy a disk drive that can store all of the world's music.[1]
  - As of 2003, an estimated 5 exabytes of digital data had been generated.
  - As of 2013, we now generate 5 exabytes every *two days*, and have digitized 2.72 zettabytes. [2]

---

[1] *Big data: the next frontier for innovation, competition, and productivity,* McKinsey Global Inst. (May 2011).
[2] *A Policy Forum on the Use of Big Data in Health* Care, Bipartisan Policy Center (June 25, 2013).

OPTUM™

Big Data – Definitions

- **Health care data is similarly exploding.** Data is coming available from a variety of new sources, including:
    - Clinical (EHR and natural language processing) and claims data
    - Biometric data
    - Consumer-generated health care data
        - Device
        - Applications
    - Personal data from data brokers
    - Social media and internet data
        - Twitter, Facebook, web browsing

OPTUM™

Big Data – Definitions

- • Since 2013, nearly six in ten (59%) hospitals have adopted at least a basic EHR system. This presents an increase of 34% from 2012 to 2013 and a five-fold increase since 2008. Over nine in ten (93%) hospitals possessed a certified EHR technology in 2013, increasing by 29% since 2011.[3]
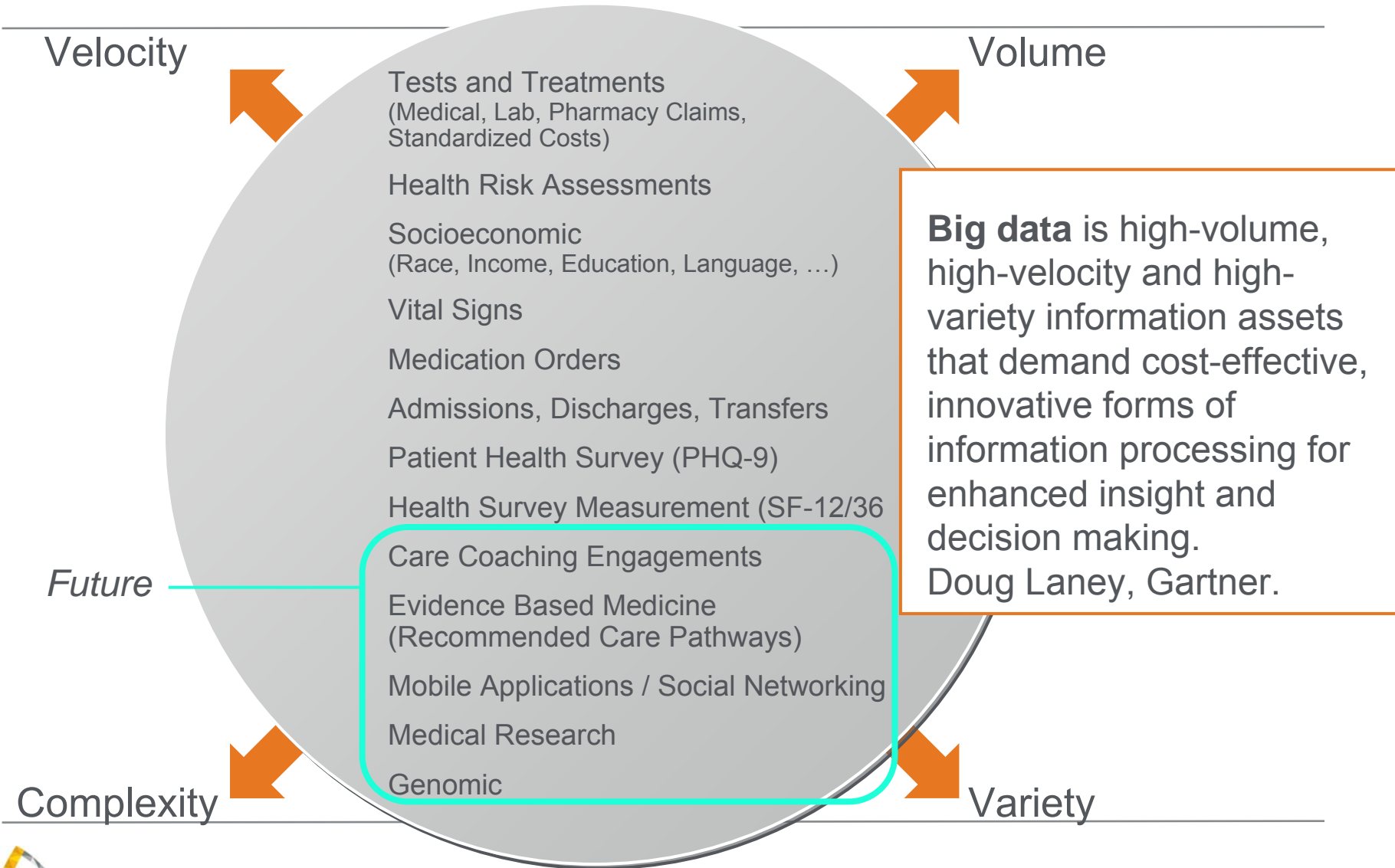
Hospital adoption of EHR systems has increased more than five-fold since 2008.

Figure 1: Percent of non-federal acute care hospitals with adoption of at least a Basic EHR system and possession of a certified EHR: 2008-2013
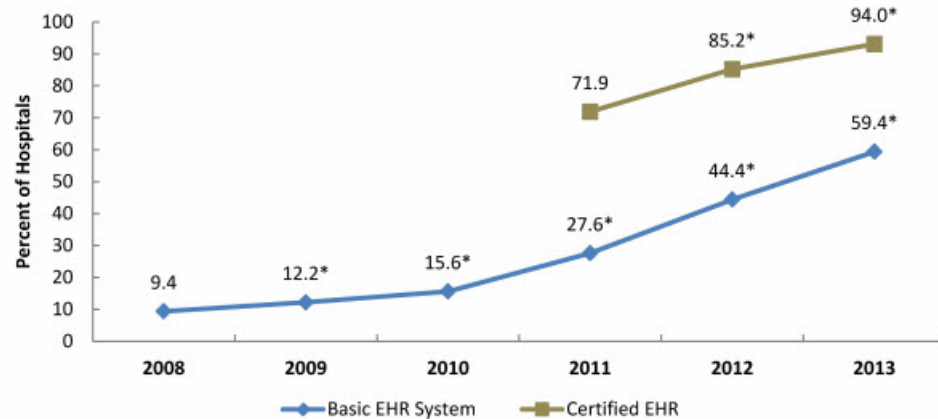
[3] *Adoption of Electronic Health Record Systems Among U.S. Non-federal Acute Care Hospitals: 2008-2013*, ONC Data Brief No. 16, (May 2014).

OPTUM™

Big Data – Definitions

Velocity

Volume

Tests and Treatments
(Medical, Lab, Pharmacy Claims,
Standardized Costs)

Health Risk Assessments

Socioeconomic
(Race, Income, Education, Language, …)

Vital Signs

Medication Orders

Admissions, Discharges, Transfers

Patient Health Survey (PHQ-9)

Health Survey Measurement (SF-12/36

Future

Care Coaching Engagements

Evidence Based Medicine
(Recommended Care Pathways)

Mobile Applications / Social Networking

Medical Research

Genomic

Big data is high-volume,
high-velocity and high-
variety information assets
that demand cost-effective,
innovative forms of
information processing for
enhanced insight and
decision making.
Doug Laney, Gartner.

Complexity

Variety

OPTUM™

6

Big Data – Definitions

- **Big data** is "[t]he ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" and "…things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value."[4]

- In health care, big data is a combination of all 3 definitions:
  - Big
  - Volume, velocity, variety
  - New insights and value

"Data is the new oil – we need to find it, extract it, refine it, and monetize it."

David Buckingham

[4] *Big Data: A Revolution that Will Transform How We Live, Work, and Think,* Viktor Mayer-Schönberger and Kenneth Cukier (Mar. 2013)

OPTUM™

What's Missing in Health Care Big Data and Innovation?

Collaboration across health care ecosystem

- Stakeholders address same problems with different approaches
- Broad collaboration is rare
- Advances in care require coordinated efforts

Availability of high-quality, high-volume data sources

- Commercial claims data sources are not readily available
- High-quality, national EHR data is nonexistent or expensive
- Linked claims/EHR data is unavailable

Opportunity to prototype and test new findings

- Testing new tools, care pathways, etc., is limited
- Rigorous prototyping and testing new ideas with feedback loop is difficult

Vehicles for widespread adoption of new knowledge

- New knowledge doesn't spread easily
- Translation requires significant investments of time and resources

OPTUM™

Big Data in Healthcare

Data Partnerships

Life Sciences/Data and Analytics

Multi-Stakeholder

PCORI CDRN
PCORI PCORnet
FDA Sentinel

Government

Delivery System/Partners

Payer/IT

Life Sciences/Payer

Life Sciences/PBM

Benefits of Big Data

Care and Treatment	• Full view of patient and member using claims, clinical, and consumer data • Can provide insights into treatment variations, behavioral, and patient subgroup response. • Predictive modeling to target care; some estimates are that such use of data could reduce health care costs by $300 to $400 billion.[5]
Member Engagement	• Use of big data can help determine who and when a patient will respond and engage to live a healthier life. • Enables personalized medicine. • Integrate data from outside of the medical system.
Fraud, Waste, and Abuse	• This already occurs on the claims level. • EHR data may now be integrated to perform a greater review than ever before. • Incorporation of non-traditional patient data may also occur.
Research	• New wearable and consumer technologies (such as Apple ResearchKit) allows researchers to move beyond claim and chart studies.

[5] *Big data: the next frontier for innovation, competition, and productivity,* McKinsey Global Inst. (May 2011)

OPTUM™

Dangers of Big Data

Overhyped and Oversold	• Not as powerful or as insightful as advertised! Big data ≠ smart data.
Accuracy	• Can lead to false conclusions without "real world" confirmation. • Is data from third parties accurate or useful?
Privacy	• Troubling questions about patient, member, and consumer privacy and choice. • HIPAA does not required consent for treatment, payment, or health care operations.
Security	• The bigger the data, the more tempting the target . . . .

PHARMA & HEALTHCARE 4/14/2014 @ 11:52PM | 3,755 views

Medicare's Big Data Dump Is Just That - A Dump

+ Comment Now   + Follow Comments

THE WALL STREET JOURNAL ☰ | BUSINESS

BUSINESS

Health Insurer Anthem Hit by Hackers

Breach Gets Away With Names, Social Security Numbers of Customers, Employees

The database of health insurer Anthem Inc., containing personal information for about 80 million customers and employees, has been hacked. WSJ's Stefanie Ilgenfritz reports. Photo: AP

OPTUM™

Covered Entities
HIPAA and Big Data

Covered Entities and Big Data

- Under HIPAA, how do you get from what's in your EHR or claims platform and combine it with data from disparate sources to create normalized, actionable Big Data?

- Covered Entities (providers and payers) have their own challenges with respect to big data.

- While CEs have access to their own data, it can difficult to combine data from other sources on an identifiable basis on a scalable basis.

- Sensitive conditions laws may also complicate sharing – HIV, substance abuse, mental health, etc.

Covered Entities and Big Data Solutions

- For sharing data *within* the health care system, Covered Entities should consider:

  - **An Affiliated Covered Entity (ACE):** legally separate entities that are affiliated may designated themselves as a single covered entity and share Protected Health Information.

  - **An Organized Health Care Arrangement (OHCA):** a clinically integrated or organized health care system that can disclose PHI about an individual for *any* health care operations activities of the organized health care arrangement.

  - Using and Disclosing PHI under HIPAA's *health care operations* exception.
    - Population-based activities relating to improving health or reducing health care costs, case management and care coordination, and health plan performance
    - *But* there must typically be a common relationship with the individual, the information must pertain to that relationship, and only for limited health care operations purposes.

  - Participating in information sharing organizations, such as Health Information Exchanges.

  - Limited Data Sets, which can be used and disclosed for research, public health, and health care operations purposes of the Covered Entity.

  - Privacy Board or IRB approved studies.

Covered Entities and Big Data Solutions

- To incorporate data from outside of the health care system, consider:
  - Working with non-HIPAA entities and using patient consent for consumer-generated data.
    - *Possible* examples include Lose It!, Fitbit, Research Kit, wireless scales, and other applications and devices.
    - Many applications or services will not want to become your Business Associate
    - Need to develop flexible and innovative ways of sharing information without disclosing PHI.
  - Explore possibilities of incorporating identifiable data from outside sources such as data brokers, medical Web sites, etc.
    - May need to link such data with your members or patients by acquiring all individuals in a certain geography or with certain attributes.
    - Consider probabilistic linking of data.
    - Consider general use of data – for instance, generalizing based on income level in a certain zip code.
  - Once any of this data is received by a Covered Entity, it becomes PHI and subject to HIPAA.

OPTUM™

Business Associates
HIPAA and Big Data

Business Associates and Big Data

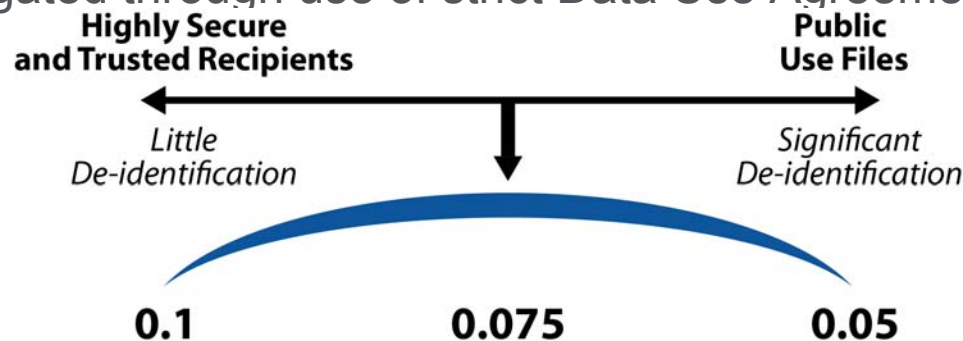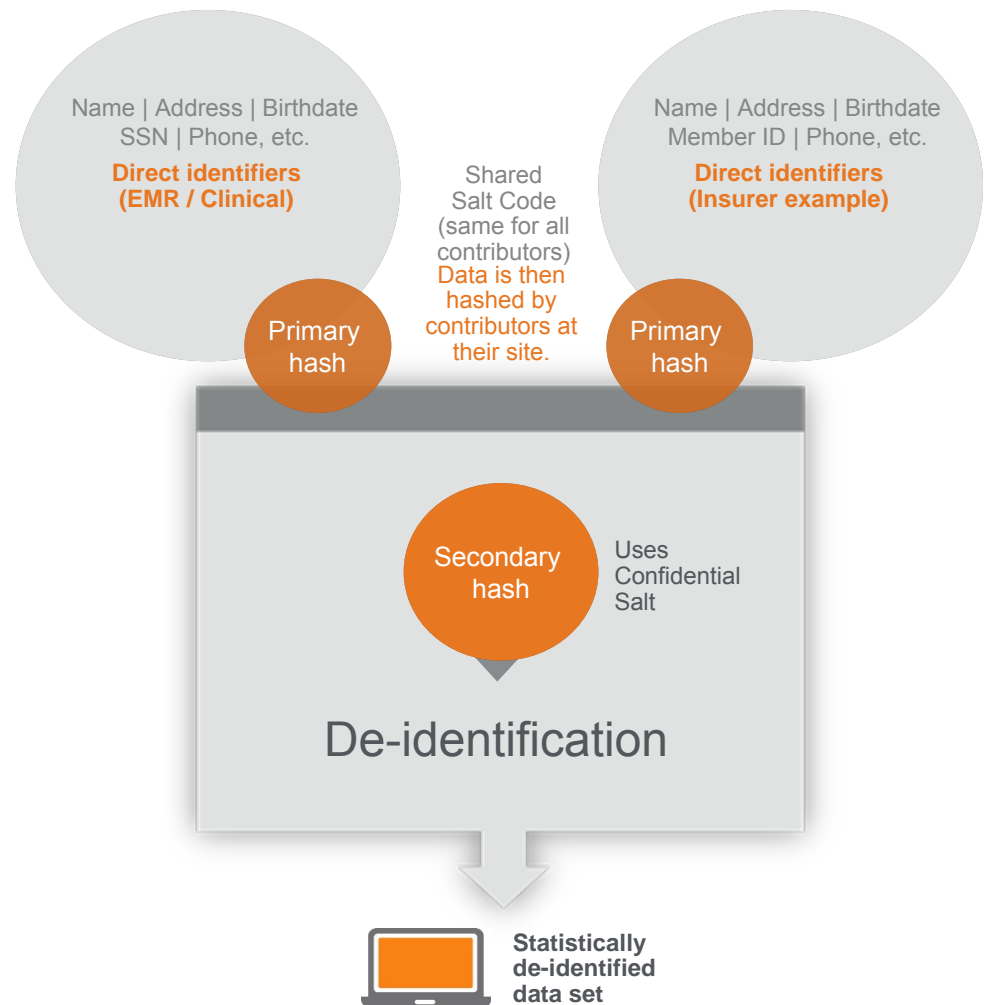|	
Aggregation and Health Care Operations	• HIPAA allows a business associate to perform services for multiple covered entities to permit data analysis for health care operations purposes. • More efficient method of performing data disclosures than individual disclosures between covered entities; in fact, covered entities may not be able to disclose much of this data between themselves. • This does not allow a Business Associate to use the data for a secondary purpose.
De-identification	• HIPAA allows a Covered Entity to disclose PHI to a business associate for the sole purpose of de-identification. • An excellent option for use of Big Data; balances risk and utility. • Allows BA to use data for secondary purposes without restriction. (This may not be attractive to Covered Entities.) • Can be complex and experts trained to de-identify databases are scarce.
Management and Administration of the BA	• Limited usefulness for Big Data • Related to back office, administrative functions such as compliance, cost management, and auditing. • Typically may not be used for secondary purposes – Business Associates may only use data as allowed under their contract.

OPTUM™

De-identified Data

- HIPAA allows a Covered Entity to disclose Protected Health Information to a Business Associate, and the Business Associate can de-identify as a health care operation. The Business Associate can then use it for a secondary purpose because it is now longer considered subject to the HIPAA Privacy Rule.

- Use of the *Expert Method* allows highly flexibly de-identified data sets to be created if the appropriate knowledge, technology, and expert is available.

- Data sets can be tailored to the type of data that is needed the most by trading off other data – for instance, geography can be sacrificed for cell size.

- Risk can be mitigated through use of strict Data Use Agreements and security.

Highly Secure and Trusted Recipients   **Public Use Files**

Little De-identification   *Significant De-identification*

0.1   0.075   0.05

OPTUM™

Data from Multiple Sources Can Be Linked and De-identified

- This is an example of how PHI can be taken from multiple sources, de-identified, and then linked.
- Requires an expert statistician to oversee the process and significant technology investment.
- Balances research utility with privacy safeguards.

Name | Address | Birthdate
SSN | Phone, etc.
**Direct identifiers
(EMR / Clinical)**

Shared Salt Code (same for all contributors) Data is then hashed by contributors at their site.

Name | Address | Birthdate
Member ID | Phone, etc.
**Direct identifiers
(Insurer example)**

Primary hash

Primary hash

Secondary hash

Uses Confidential Salt

De-identification

Statistically de-identified data set

OPTUM™

Conclusion

- Big Data and health care is a reality. HIPAA allows for the creation of Big Data through:
    - Health care operations and disclosing data regarding shared patients and members;
    - Incorporating outside data, whether through consent or matching;
    - Privacy Board approved studies;
    - Business Associates aggregating data of multiple covered entities for health care operations purposes; and
    - Using HIPAA's de-identification rules to link multiple data sets to use for secondary purposes.

OPTUM™

Questions?
Thank You

Mitchell W. Granberg, Optum Chief Privacy Officer

mitchell.granberg@optum.com