EMORY UNIVERSITY | Department of Biomedical Informatics

EHR Data Quality for Healthcare Analytics: Not Secondary Anymore

Andrew Post, MD, PhD
Assistant Professor and Clinical Informatics Architect
Department of Biomedical Informatics
Emory Healthcare Information Services
Emory University

EMORY UNIVERSITY | Department of Biomedical Informatics

Financial Disclosures

- None

EMORY UNIVERSITY | Department of Biomedical Informatics

Secondary Use of Clinical Data

Collected for one purpose, used for another

- Clinical and translational research
- Healthcare quality measurement and improvement
- Personal health records (PHRs)
- Health information exchange (HIE)
- Public health surveillance for emerging threats

EMORY UNIVERSITY | Department of Biomedical Informatics

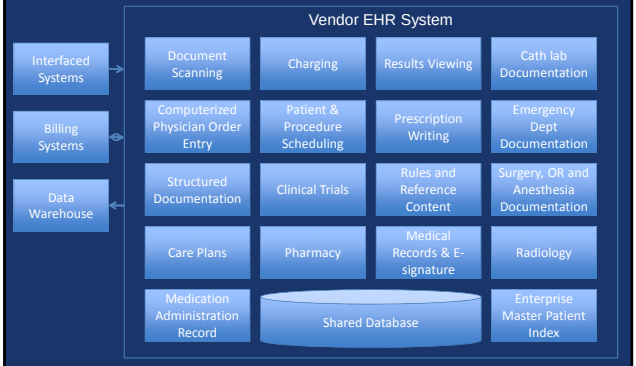
Challenges in Using Existing Data

- "Left censoring": first instance of disease in record may not be when first manifested
- "Right censoring": data source may not cover long enough time interval
- Data from other clinical (other health systems) and non-clinical (pharmacies) settings may not be captured
- Bias in testing and treatment
- Institutional or personal variation in practice or documentation styles
- Inconsistent use of coding or standards

Hersh, W. Secondary Use of Clinical Data from Electronic Health Records. Available from: http://skynet.ohsu.edu/~hersh/secondary-use-pr.pdf (presentation).

EMORY UNIVERSITY | Department of Biomedical Informatics

Where does EHR Data Come From?



EMORY UNIVERSITY | Department of Biomedical Informatics

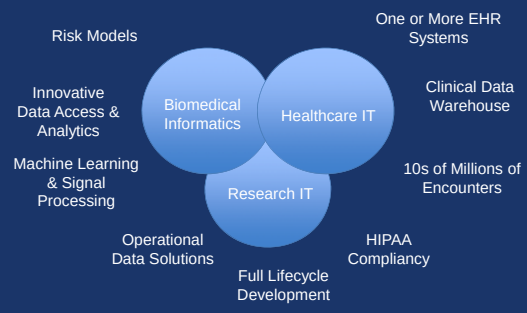
Meaningful Use of an EHR

- Driven by 5 underlying goals for a healthcare system
  - Improving quality, safety and efficiency
  - Engaging patients in their care
  - Increasing coordination of care
  - Improving the health status of the population
  - Ensuring privacy and security
- Three requirements – use of certified EHR tech
  - In a meaningful manner
  - Connected for health information exchange (HIE)
  - To submit information on clinical quality measures

Hersh, W. Meaningful Use and All That: Update in Clinical Informatics. Available from: http://skynet.ohsu.edu/~hersh/clinical-informatics-update.pdf (presentation).

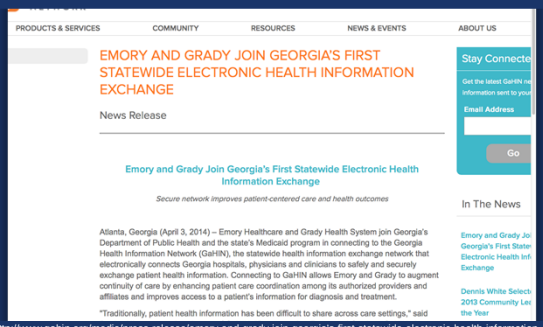
EMORY UNIVERSITY | Department of Biomedical Informatics

A Snapshot of Data Resources of the Modern Academic Health System



EMORY UNIVERSITY | Department of Biomedical Informatics

Health Information Exchanges



http://www.gahin.org/media/press-release/emory-and-grady-join-georgia's-first-statewide-electronic-health-information

EMORY UNIVERSITY | Department of Biomedical Informatics

Big Data

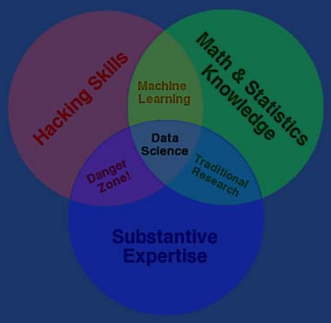
- Storage capacity has outstripped advances in the speed of reading data from disk and the amount of RAM in typical computers for processing it
- Data management challenges around volume, velocity and variety – Doug Laney, META Group, 2001
- "Big Data" requires specialized tools because it doesn't fit into RAM
- Big Data tasks usually involve data reuse for a new purpose

EMORY UNIVERSITY | Department of Biomedical Informatics

Data Variety (Complexity) is the Challenge

EMORY UNIVERSITY | Department of Biomedical Informatics

Data Science/eScience



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

EMORY UNIVERSITY | Department of Biomedical Informatics

Data Science Characteristics

80% of the work
–Aaron Kimball

90% of their work
–Scientists at Univ. of Washington
(the other 10% is science)

- Get and clean lots of data, often from more than one source
- Define a data model describing the "typical" regions in the data
- Look for "outliers" – data that doesn't fit the model
- Look for individual outliers – credit card transaction involving an usually large amount of money
- Look for outliers in sequence data
  - sequence of credit card transactions at businesses near highways in neighboring states
  - spike in tweets about cold symptoms
- Typical behavior may change with time, so update the model using new data frequently

EMORY UNIVERSITY | Department of Biomedical Informatics

Hospital Readmissions within 30 Days

5 years of data from local data warehouse – 230,000 hospital encounters
5 years of data from the UHC Clinical Database – ~20 million hospital encounters

Relevant Medical Record Data

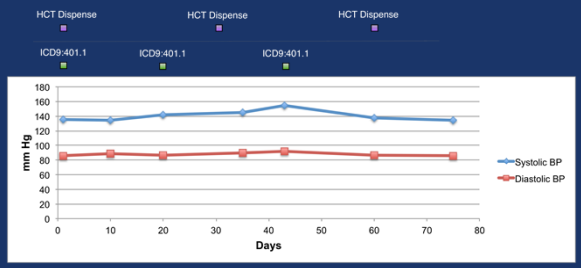
- Admissions/discharges
- Diagnosis codes (18,012)
- Procedure codes (4,665)
- APR DRGs (318)
- MS DRGs (750)
- Medication orders (32,936)
- Laboratory test results (3,117)
- Vital signs
- Geographic information

Variables of Interest

- Chronic co-morbidities
- Specialty providing care
- Variables derived from clinical results
  - Obesity
  - Diabetes/uncontrolled diabetes
  - End-stage renal disease (ESRD)
  - Pressure ulcer
  - Sickle cell disease/sickle cell crisis
- Temporal variables
  - Multiple previous MI
  - Multiple 30-day readmissions
  - Chemotherapy prior to surgery
  - Recent previous encounter

EMORY UNIVERSITY | Department of Biomedical Informatics

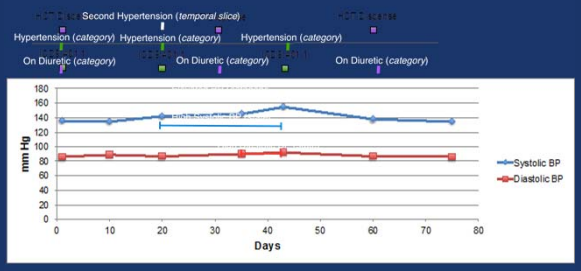
Automated Chart Abstraction



Post AR & Harrison JH. J Am Med Inform Assoc 2007;14(5):674-83

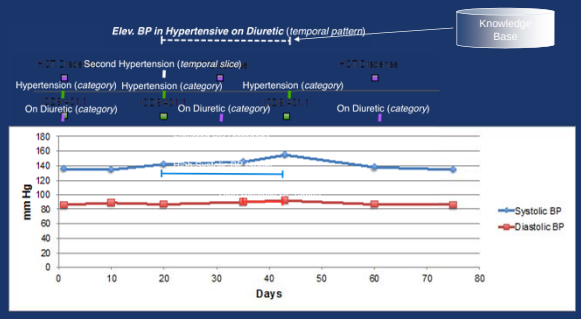
EMORY UNIVERSITY | Department of Biomedical Informatics

Abstraction of Data into Intervals



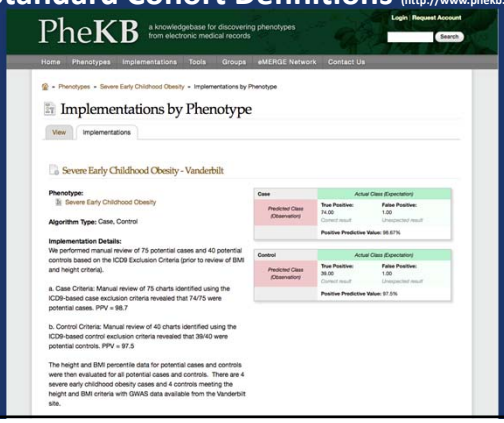
EMORY UNIVERSITY | Department of Biomedical Informatics

Abstraction of Intervals into Information



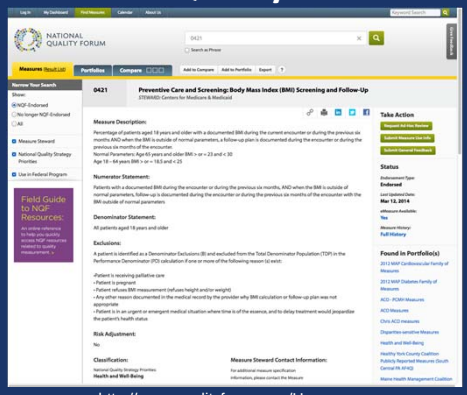
EMORY UNIVERSITY | Department of Biomedical Informatics

Standard Cohort Definitions (http://www.phekb.org)



EMORY UNIVERSITY | Department of Biomedical Informatics

Standard Quality Measures

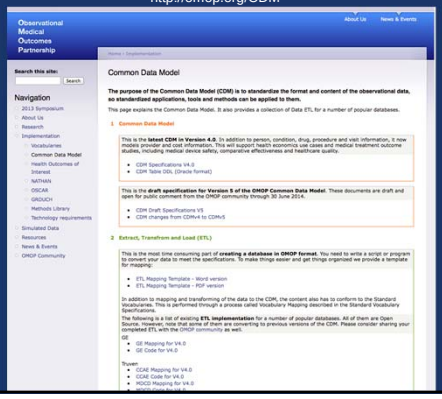


http://www.qualityforum.org/Home.aspx

EMORY UNIVERSITY | Department of Biomedical Informatics

Shared Data Models

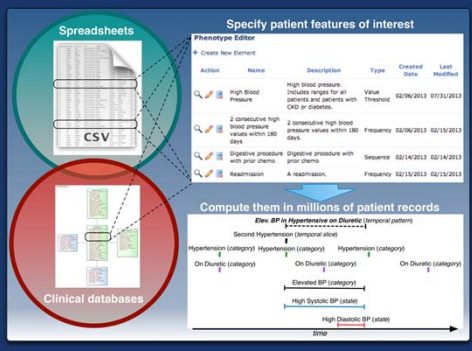
http://omop.org/CDM



EMORY UNIVERSITY | Department of Biomedical Informatics

Computing Derived Measures with Shared Tools

http://eurekaclinical.org

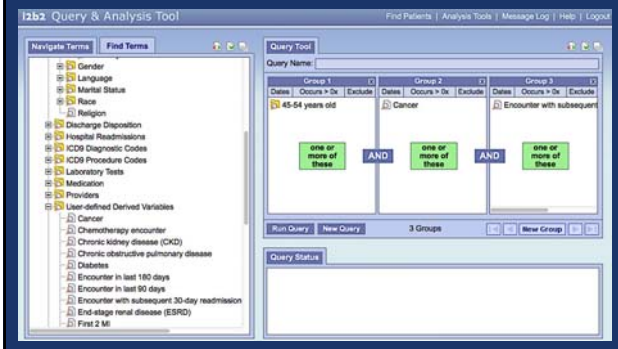


Post AR, Kurc T et al. AMIA Annu Symp Proc 2013;1160-69

EMORY UNIVERSITY | Department of Biomedical Informatics

Cohort Discovery and Data Extraction

http://www.i2b2.org



EMORY UNIVERSITY | Department of Biomedical Informatics

HIEs for Population Studies

Application of Information Technology ■

The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories

GRIFFIN M. WEBER, MD, PHD, SHAWN N. MURPHY, MD, PHD, ANDREW J. MCMURRY, MS, DOUGLAS MACFADDEN, MS, DANIEL J. NIGRIN, MD, MS, SUSANNE CHURCHILL, PHD, ISAAC S. KOHANE, MD, PHD

Abstract The authors developed a prototype Shared Health Research Information Network (SHRINE) to identify the technical, regulatory, and political challenges of creating a federated query tool for clinical data repositories. Separate Institutional Review Boards (IRBs) at Harvard's three largest affiliated health centers approved use of their data, and the Harvard Medical School IRB approved building a Query Aggregator Interface that can simultaneously send queries to each hospital and display aggregate counts of the number of matching patients. Our experience creating three local repositories using the open source Informatics for Integrating Biology and the Bedside (i2b2) platform can be used as a road map for other institutions. The authors are actively working with the IRBs and regulatory groups to develop procedures that will ultimately allow investigators to obtain identified patient data and biomaterials through SHRINE. This will guide us in creating a future technical architecture that is scalable to a national level, compliant with ethical guidelines, and protective of the interests of the participating hospitals.

■ **J Am Med Inform Assoc.** 2009;16:624–630. DOI 10.1197/jamia.M3191.

EMORY UNIVERSITY | Department of Biomedical Informatics

Secondary Use – Where Are We?

1. Comprehensive electronic health record
2. Comprehensive clinical data warehouse ⟵
3. Shared data representations for research and quality
4. Shared definitions of derived measures of interest in research and quality
5. Tools for cohort discovery and export into analytics and clinical research tools
6. Health information networks for research and quality

EMORY UNIVERSITY | Department of Biomedical Informatics

QUESTIONS AND DISCUSSION