Workshop Introduction
Putting the Pieces Together: Precision Medicine Discovery from Electronic Health Records

Sarah A. Pendergrass PhD, MS
Assistant Professor/Investigator I
Biomedical and Translational Informatics
Geisinger Health System

September 22, 2015

Precision Medicine

The right medicine

And the right intervention

The right patient

The right dose

The right time

Also remember we are moving to drug development based on genetic variation

A Key to Precision Medicine

- We collect a tremendous amount of information about health and disease through **electronic health records**
  - Diagnoses
  - Clinical Lab Measurements
  - Medications
  - In patient and out patient

- What if we use these information to help inform better patient treatment?

http://www.brimg.net/images/doctor-using-mobile-chart-checking-patient-corbis_573x300.jpg
http://ihealthtran.com/wordpress/wp-content/uploads/2012/11/EHR-Health-IT.jpg

Integrating EMRs into Genomic Research

Use of EMRs in medicine

Early 1990's

Sequenced the human genome

2001

Improved technologies for biomedical analysis

~ 2005

NHGRI eMERGE Initiative

2007

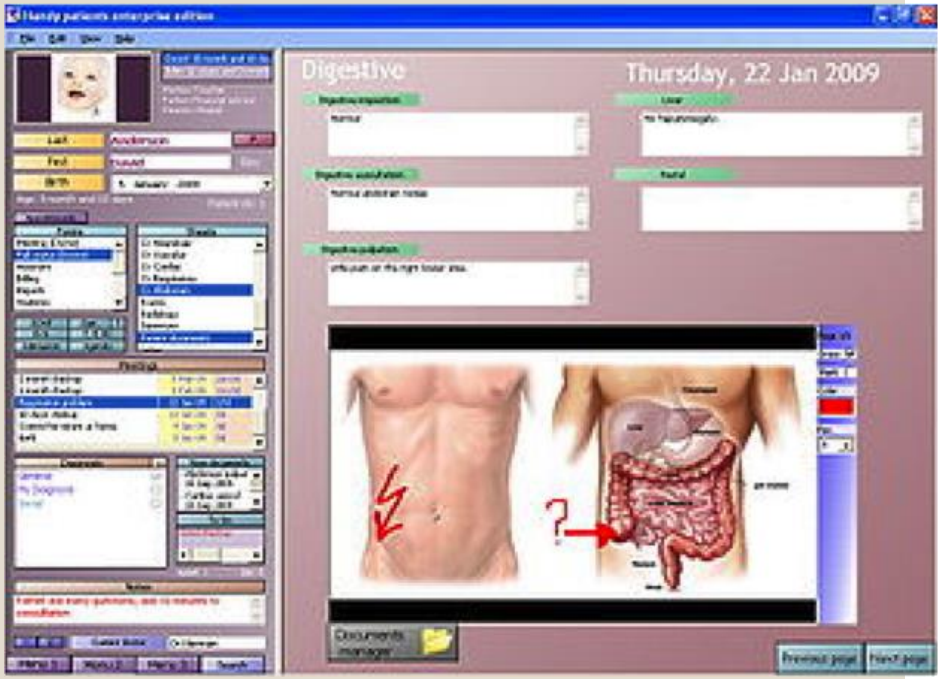
Tools for analyzing large datasets

Now

Janina M. Jeff, PhD, MS

ELECTRONIC HEALTH RECORDS (EHRS)

- Electronic Health Records (EHRs)
- Electronic version of a patients medical history
  - maintained by the provider over time
  - demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, biomarkers, and radiology reports



"EHRs will enable providers to make better decisions and provide better care. "

Electronic Health Record (EHR) Linked to Genetic Data

- Success in research discovery with de-identified patient electronic health records (EHR) linked to de-identified genetic data
  - Identification of novel genetic associations

- Genetic Data
  - Single nucleotide polymorphisms (SNPs)
  - Common frequency variants
  - Genome-wide association studies
  - Pharmacogenomics

- EHR
  - Wide range of patient derived data

Illustration: iStockphoto

Electronic Health Record (EPIC/Clarity)

- Ambulatory (Outpatient)
- Inpatient (Hospital Admissions)
- Emergency Department
- Medication Orders
- Lab (Orders and Results)
- Imaging Orders
- Procedures
- Diagnosis information
- Demographics
- Patient History (Social, Surgical, Medical, etc.)
- Problem List

Disparate Data Sources

- Cardiology Databases (Xcelera [1991 forward], MUSE [1980 forward], Echo [1991 forward], Apollo [1999 forward])
  - Cardiovascular Imaging (MR/CT)
  - Electrocardiogram
  - Echocardiogram
  - Surgical and Catheterization
- Radiology (RIS) [1997 forward]
  - Pre-procedure questions
  - Radiology Reports
- DEXA [1998 forward]

Phenotypic Algorithm Development

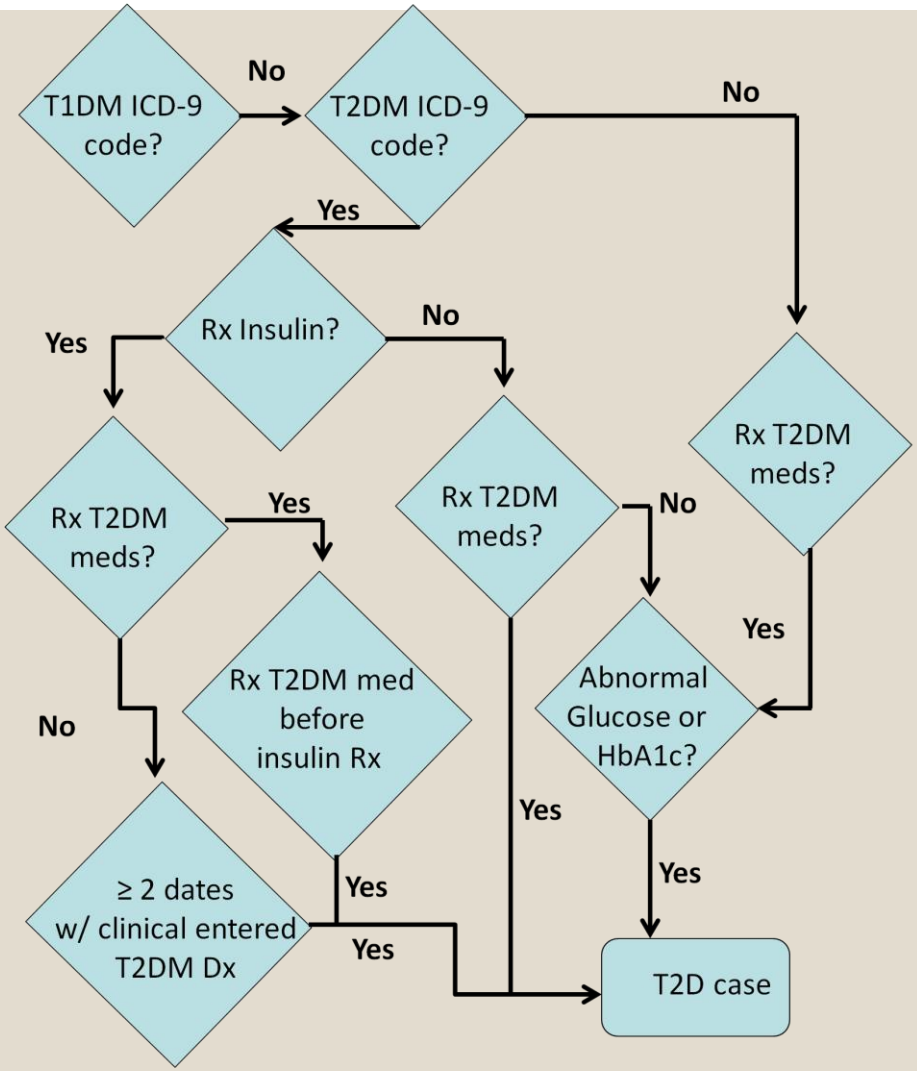
- Terminology
  - Using International Classification of Disease codes (ICD-9, ICD-10)
  - Procedural Codes (CPT)
  - Problem lists (Historical Codes, V,T,E codes)
  - Clinical Laboratory Values
  - Pharmacologic Data
  - Vital Signs
  - Structured Billing Text and Notes

- Note: Variability in these data exists across multiple EHR platforms

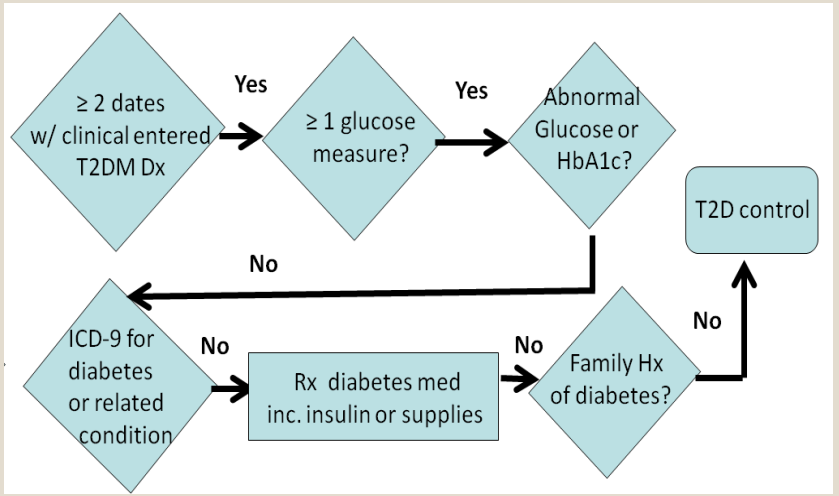
Developing a Phenotypic Algorithm

- Using multiple pieces of information to define case/control status, or a quantitative phenotypic measure
  - What is your phenotype of interest?
  - How can this phenotype be optimally used for increased power in a large-scale 'omic analysis?
  - Who are the individuals that will be affected by this disease or will have this phenotype?
  - Is it a case/control phenotype, quantitative measure?
  - What are medications, comorbidities, other diseases, surgeries that might affect your phenotype or measure?
    - Example: lipid levels after drug treatment suggesting an individual has "normal" lipid levels, impacting who you define as "high lipid levels"
    - Example: patient incorrectly defined as a control due to treatment radically changing their patient medical record, such as gastric bypass surgery resulting in type-2 diabetes reversal
    - Example: patient incorrectly defined as a "case" when their condition is a a result of surgery

Phenotypic Algorithm: Flow Chart

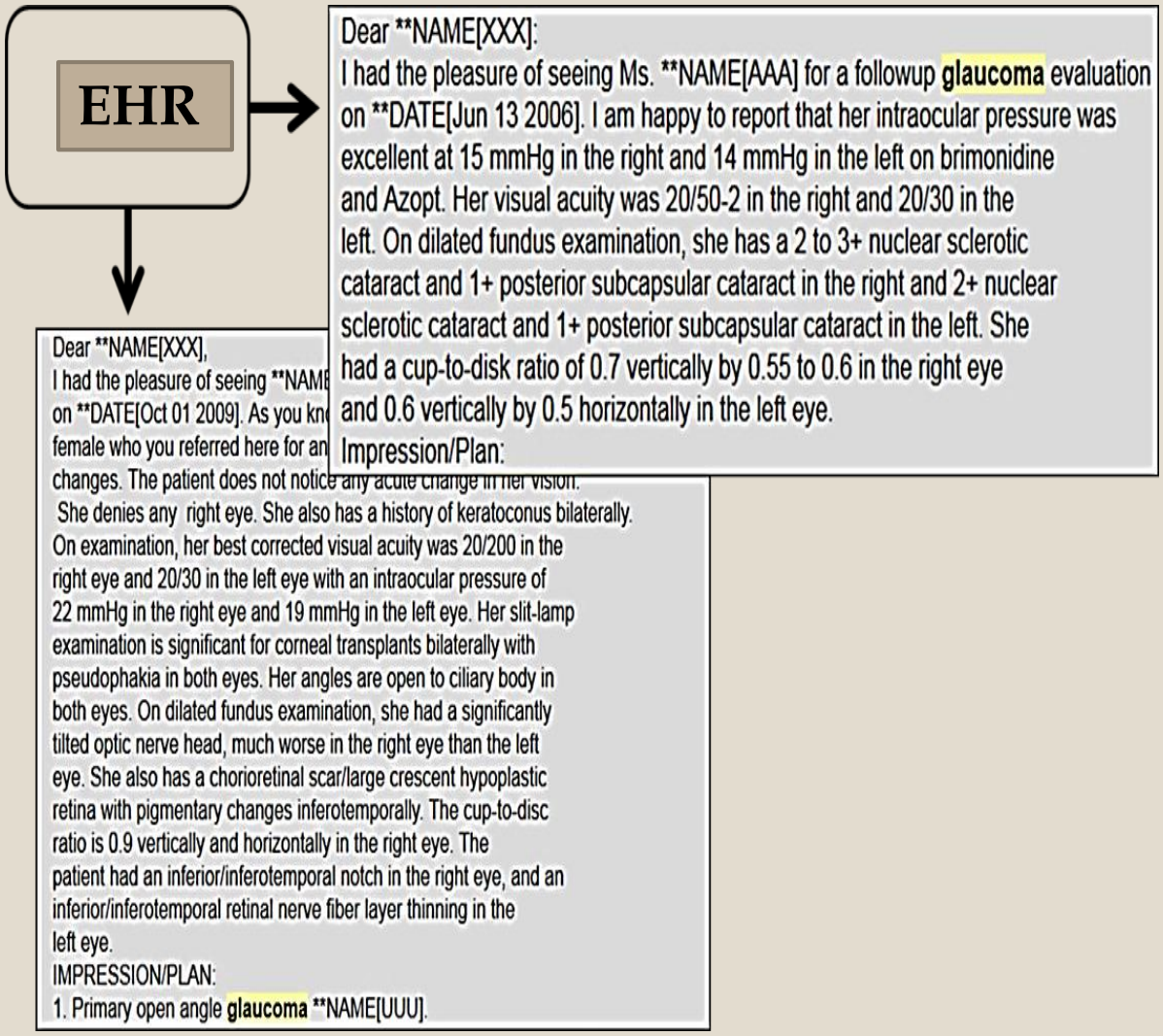


Case definition



Control definition

Clinical notes/communications allow for Indirect Verification



Developing a Phenotypic Algorithm

- How do I know it works?
  - Use algorithm to define group of cases and controls and evaluate PPV and NPV
    - Chart review
  - **PPV: Positive Predictive Value** is the probability that subjects with a positive screening test truly have the disease
    - Iteratively refine case definition through partial manual review until case definition yields PPV ≥ 95%
  - **NPV: Negative Predictive Value** is the probability that subjects with a negative screening test truly don't have the disease
    - For controls, exclude all potentially overlapping syndromes and possible matches, iteratively refine such that NPV ≥ 98%

More Resources?

- PheKB (https://phekb.org/)
  - Documentation and versioning of validated phenotype algorithms
    - Implementation details
  - Can validate existing phenotype algorithms in a different EHR
  - Collaborate on phenotypic algorithm develpment