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

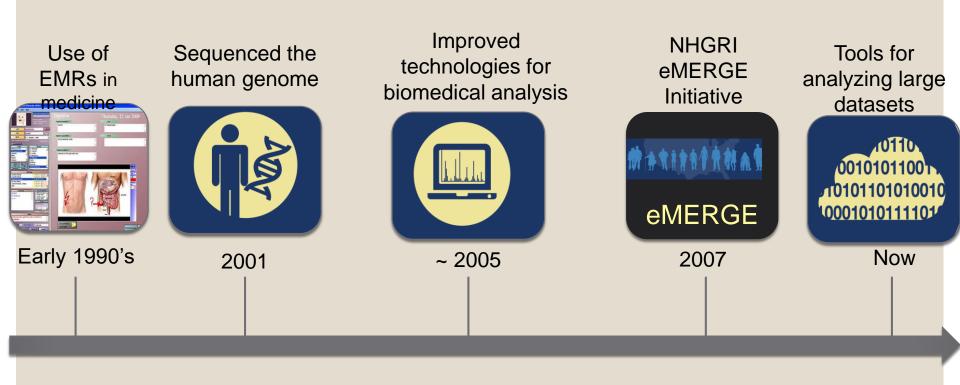


What if we use these information to help inform better patient treatment?



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

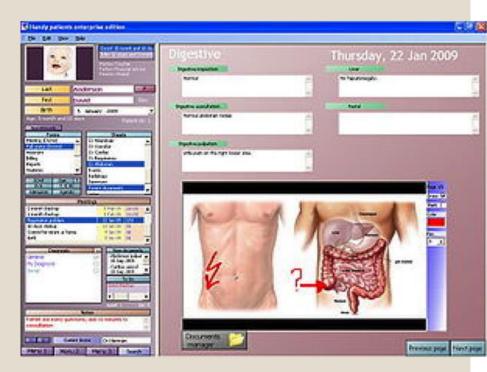
Integrating EMRs into Genomic Research





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



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

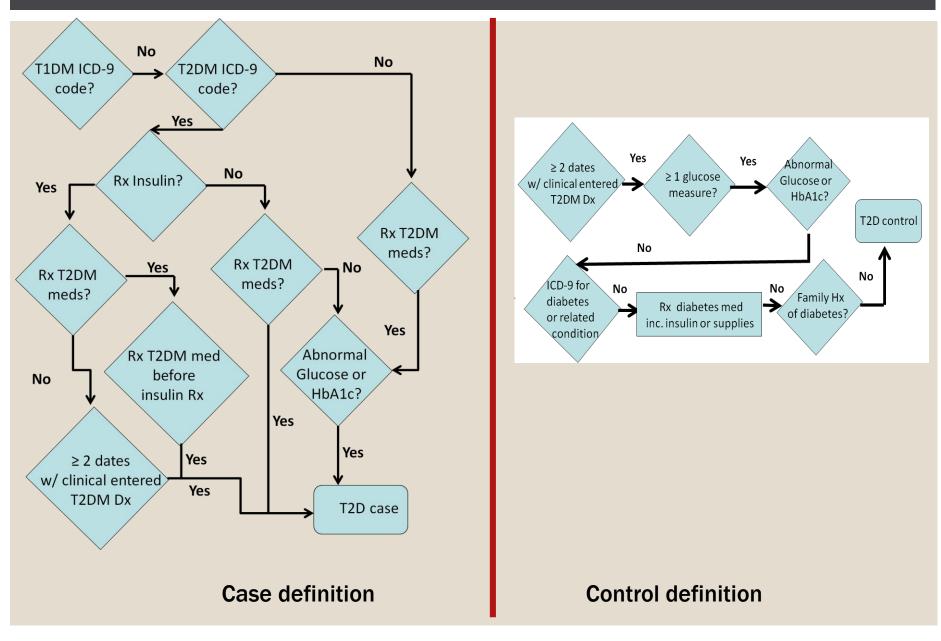
- Terminology
 - Using International Classification of Disease codes (ICD-9, ICD-10)
 - Procedural Codes (CPT)
 - Problem lists (Historical Codes, V,T,E codes)
 - Clinical Laboratory Values
 - Pharmacologic Data
 - Vital Signs
 - Structured Billing Text and Notes

 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



Clinical notes/communications allow for Indirect Verification

Dear **NAME[XXX]:

I had the pleasure of seeing Ms. **NAME[AAA] for a followup glaucoma evaluation on **DATE[Jun 13 2006]. I am happy to report that her intraocular pressure was excellent at 15 mmHg in the right and 14 mmHg in the left on brimonidine and Azopt. Her visual acuity was 20/50-2 in the right and 20/30 in the left. On dilated fundus examination, she has a 2 to 3+ nuclear sclerotic cataract and 1+ posterior subcapsular cataract in the right and 2+ nuclear sclerotic cataract and 1+ posterior subcapsular cataract in the left. She had a cup-to-disk ratio of 0.7 vertically by 0.55 to 0.6 in the right eye I had the pleasure of seeing **NAME on **DATE[Oct 01 2009] As you kn and 0.6 vertically by 0.5 horizontally in the left eye.

female who you referred here for an Impression/Plan: changes. The patient does not notice any acute change in her vision. She denies any right eye. She also has a history of keratoconus bilaterally. On examination, her best corrected visual acuity was 20/200 in the right eye and 20/30 in the left eye with an intraocular pressure of 22 mmHg in the right eye and 19 mmHg in the left eye. Her slit-lamp examination is significant for corneal transplants bilaterally with pseudophakia in both eyes. Her angles are open to ciliary body in both eyes. On dilated fundus examination, she had a significantly tilted optic nerve head, much worse in the right eye than the left eye. She also has a chorioretinal scar/large crescent hypoplastic retina with pigmentary changes inferotemporally. The cup-to-disc ratio is 0.9 vertically and horizontally in the right eye. The patient had an inferior/inferotemporal notch in the right eye, and an inferior/inferotemporal retinal nerve fiber layer thinning in the left eye. IMPRESSION/PLAN: Primary open angle glaucoma **NAME[UUU].

EHR

Dear **NAME[XXX].

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 (<u>https://phekb.org/</u>)

- Documentation and versioning of validated phenotype algorithms
 Implementation details
- Can validate existing phenotype algorithms in a different EHR
- Collaborate on phenotypic algorithm development



a knowledgebase for discovering phenotypes from electronic medical records