Opportunities and Challenges for Clinical Research with Electronic Health Records

Benjamin A. Goldstein, PhD, MPH

ben.goldstein@duke.edu

Department of Biostatistics & Bioinformatics
School of Medicine
Duke University

July 30th, 2017

WHY WE WANT TO USE EHRS FOR CLINICAL RESEARCH

- Data readily available
- Often 100,000's of Patients
- Information collected over a variety of fields
- Can study just about any clinical outcome
- Representative Population

What We Can Do With Electronic Health Records

1. Risk Prediction
    - Near term prediction - *Risk of inhospital sepsis*
    - Long(er) term risk - *30 Day Revisit*
2. Population Health
    - Health Service Utilization - *Assessment of high utilizers*
    - Disease Epidemiology - *Experience of incident diabetes in Durham County*
3. Comparative Effectiveness Research (CER)
    - Retrospective Studies - *Assessment of community intervention for diabetics (SEDI)*
    - Prospective Studies - *Point of care randomization*
4. Association Analyses
    - Risk factors for disease - *Phenome Wide Association Studies*
    - Data mining - *Drug-Drug interactions*

Why We May *Not* Want to use EHRs for Clinical Research

Data are not collected for research

- Data exist in disparate places
- All patients have different pieces of information
- Observational Data

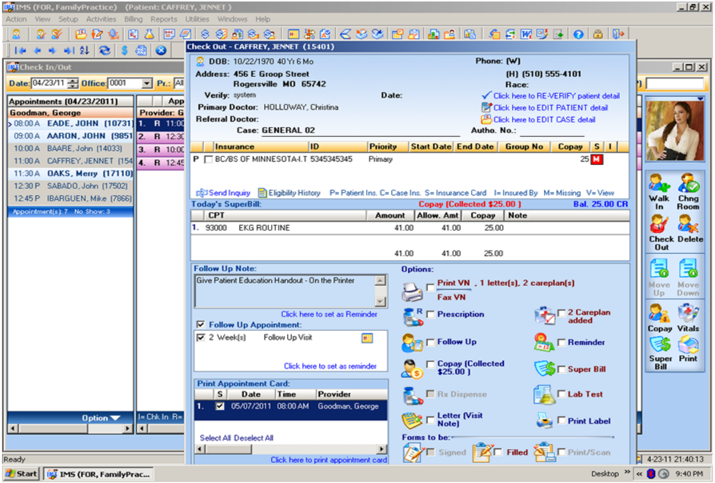
1. Structure of Electronic Health Records

2. Research with EHR Data

3. Concluding Thoughts

1. Structure of Electronic Health Records
2. Research with EHR Data
3. Concluding Thoughts

THE EHR FRONT END: GETTING DATA IN



Date: 04/23/11 Office: 0001 Pr.: All

Appointments (04/23/2011)

Goodman, George

08:00 A	EADE, JOHN (10731)
09:00 A	AARON, JOHN (9851
10:00 A	BAARE, John (14033)
11:00 A	CAFFREY, JENNET (154
11:30 A	OAKS, Merry (17110)
12:30 P	SABADO, John (17502)
12:45 P	IBARGUEN, Mike (7866)

Appointment(s): 7 No Show: 3

App

Provider: G

1.	R	11:00
2.	R	12:30
3.	R	10:00
4.	R	12:45

Check Out - CAFFREY, JENNET (15401)

DOB: 10/22/1970 40 Yr 6 Mo
Phone: (W)
(H) (510) 555-4101
Address: 456 E Groop Street
Rogersville MO 65742
Race:
Verify: system
Date:
✓ Click here to RE-VERIFY patient detail
Primary Doctor: HOLLOWAY, Christina
Click here to EDIT PATIENT detail
Referral Doctor:
Click here to EDIT CASE detail
Case: GENERAL 02
Autho. No.:

		Insurance	ID	Priority	Start Date	End Date	Group No	Copay	S	I
P	□	BC/BS OF MINNESOTA-I.T	5345345345	Primary				25	M	

Send Inquiry Eligibility History P= Patient Ins. C= Case Ins. S= Insurance Card I= Insured By M= Missing V= View

Today's SuperBill: Copay (Collected $25.00) Bal. 25.00 CR

	CPT		Amount	Allow. Amt	Copay	Note
1.	93000	EKG ROUTINE	41.00	41.00	25.00	
			41.00	41.00	25.00	

Follow Up Note:

Give Patient Education Handout - On the Printer

Click here to set as Reminder

☑ Follow Up Appointment:

☑ 2 Week(s) Follow Up Visit

Click here to set as reminder

Print Appointment Card:

	S	Date	Time	Provider
1.	☑	05/07/2011	08:00 AM	Goodman, George

Select All Deselect All

Click here to print appointment card

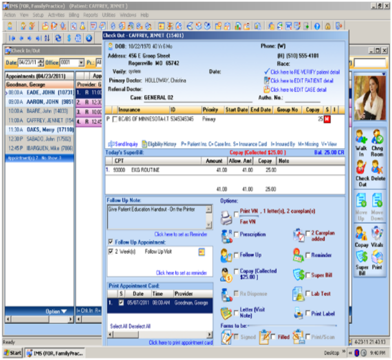
Options:

- □ Print VN , 1 letter(s), 2 careplan(s)
  Fax VN
- □ Prescription
- □ 2 Careplan added
- □ Follow Up
- □ Reminder
- □ Copay (Collected $25.00)
- □ Super Bill
- □ Rx Dispense
- □ Lab Test
- □ Letter (Visit Note)
- □ Print Label

Forms to be:

□ Signed □ Filled □ Print/Scan

Data Move Front End to Data Wharehouse



Date: 04/23/11 Office: 0001 Pr.: All

Appointments (04/23/2011)
Goodman, George
08:00A EADE, JOHN (10731)
09:00A AARON, JOHN (9851
10:00A BAARE, John (14033)
11:00A CAFFREY, JENNET (154
11:30A OAKS, Merry (17110)
12:30P SABADO, John (17502)
12:45P IBARGUEN, Mike (7866)
Appointment(s): 7 No Show: 3

	App
Provider: G	
1.	R 11:00
2.	R 12:30
3.	R 10:00
4.	R 12:45

Check Out - CAFFREY, JENNET (15401)

DOB: 10/22/1970 40 Yr 6 Mo Phone: (W)
Address: 456 E Groop Street (H) (510) 555-4101
Rogersville MO 65742 Race:
Verify: system Date:
Primary Doctor: HOLLOWAY, Christina
Referral Doctor:
Case: GENERAL 02 Autho. No.:

✓ Click here to RE-VERIFY patient detail
Click here to EDIT PATIENT detail
Click here to EDIT CASE detail

		Insurance	ID	Priority	Start Date	End Date	Group No	Copay	S	I
P	□	BC/BS OF MINNESOTA-H.T	5345345345	Primary				25	M	

Send Inquiry Eligibility History P= Patient Ins. C= Case Ins. S= Insurance Card I= Insured By M= Missing V= View

Today's SuperBill: Copay (Collected $25.00) Bal. 25.00 CR

	CPT	Amount	Allow. Amt	Copay	Note
1.	93000 EXG ROUTINE	41.00	41.00	25.00	
		41.00	41.00	25.00	

Follow Up Note:
Give Patient Education Handout - On the Printer
Click here to set as Reminder

☑ Follow Up Appointment:
☑ 2 Week(s) Follow Up Visit
Click here to set as reminder

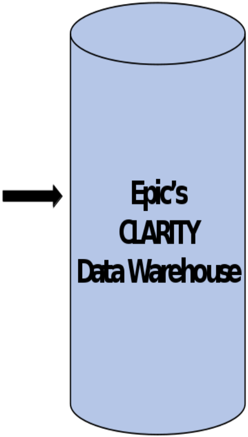
Print Appointment Card:

	S	Date	Time	Provider
1.	☑	05/07/2011	08:00 AM	Goodman, George

Select All Deselect All
Click here to print appointment card

Options:
□ Print VN , 1 letter(s), 2 careplan(s)
Fax VN
□ Prescription
□ 2 Careplan added
□ Follow Up
□ Reminder
□ Copay (Collected $25.00)
□ Super Bill
□ Rx Dispense
□ Lab Test
□ Letter (Visit Note)
□ Print Label
Forms to be:
□ Signed □ Filled □ Print/Scan

4-23-11 21:40:13
9:40 PM



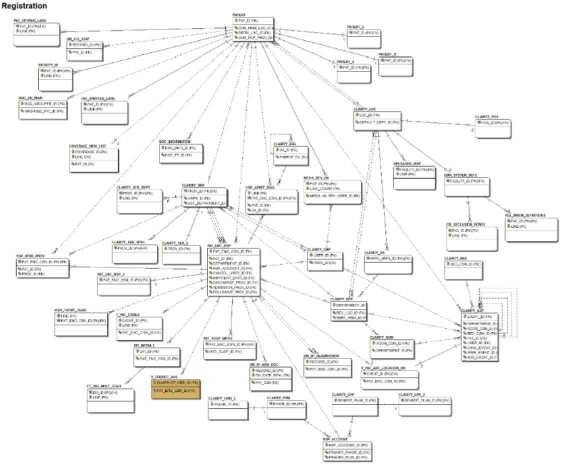
Epic's
CLARITY
Data Warehouse

- •Patient Demographics
- •Encounters (Outpatient/Inpatient)
- •Diagnoses
- •Procedures
- •Lab Results
- •Medications
- •Vital Signs
- •Social History
- •Radiological Results
- •Clinician Notes

Etc.

The Data Structure

It's
Complicated...

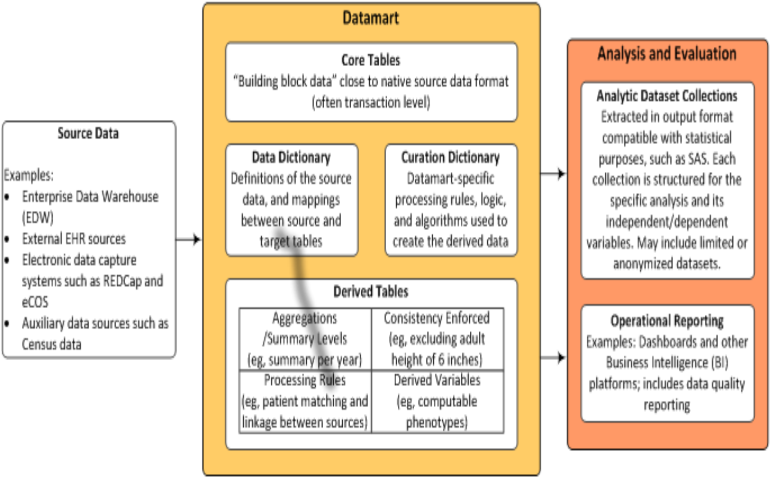


Check the Blind Spots

- Data movement and curation requires decision-making.
- Decisions may not be easily accessible.
- Decisions may not be documented or documentation may not be made available.

Turning EHRs into Data

The analysis pipeline and data platform



Source Data

Examples:

- Enterprise Data Warehouse (EDW)
- External EHR sources
- Electronic data capture systems such as REDCap and eCOS
- Auxiliary data sources such as Census data

Datamart

Core Tables
"Building block data" close to native source data format (often transaction level)

Data Dictionary
Definitions of the source data, and mappings between source and target tables

Curation Dictionary
Datamart-specific processing rules, logic, and algorithms used to create the derived data

Derived Tables

Aggregations /Summary Levels (eg, summary per year)	Consistency Enforced (eg, excluding adult height of 6 inches)
Processing Rules (eg, patient matching and linkage between sources)	Derived Variables (eg, computable phenotypes)

Analysis and Evaluation

Analytic Dataset Collections
Extracted in output format compatible with statistical purposes, such as SAS. Each collection is structured for the specific analysis and its independent/dependent variables. May include limited or anonymized datasets.

Operational Reporting
Examples: Dashboards and other Business Intelligence (BI) platforms; includes data quality reporting

Figure 2. Datamart components and relationship to external systems and processes.

DATA MARTS: STRENGTHS AND WEAKNESSES

- Strengths
  - Registry like
  - Multiple clinical subject areas for cohort
  - Regularly scheduled data refresh

  Ideal For: Posing variety of questions across subject area

- Soft Spots
  - More time and effort to create than data extract
  - Structure not easily adaptable
  - Data are fixed between refreshes

  Not Ideal For: Small, targeted analyses

Adding Information Back into EHR

- Dashboards
- Best Practice Alerts
- Predictive Analytics
- Clinical trial recruitment (Snifters)

DIFFERENT TYPES OF CLINIC ENVIRONMENTS

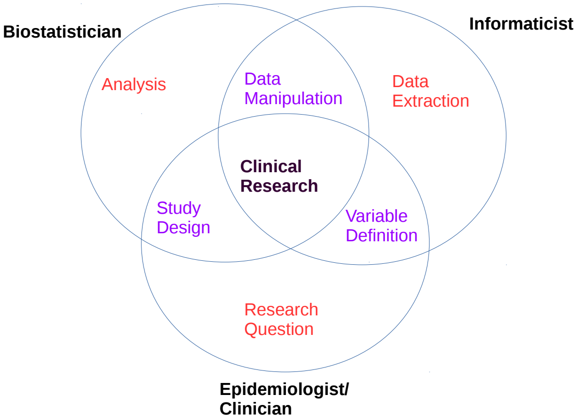
- Clinic Based System (e.g. Practice Fusion, Flatiron)
  - Capture Routine Care
  - Local Population
  - Misses inpatient activity
- Hospital Based System
  - Observe inpatient procedures and events
  - Only observe when sick
  - Referral hospitals may not represent local or stable population
- Comprehensive Medical System (e.g. VA, Kaiser)
  - Observe all types of patient encounters
  - May represent artificial population

1 Structure of Electronic Health Records

2 Research with EHR Data

3 Concluding Thoughts

Collaborative Clinical Research



Biostatistician

Informaticist

Epidemiologist/
Clinician

Analysis

Data
Extraction

Research
Question

Data
Manipulation

Study
Design

Variable
Definition

**Clinical
Research**

Four Ways EHR Data Differ from Traditional Clinical Data

1. We don't have everything we want
2. Outcomes are not defined - need to phenotype data
3. Data irregularly and potentially densely observed
4. Data not observed randomly - Informed Presence

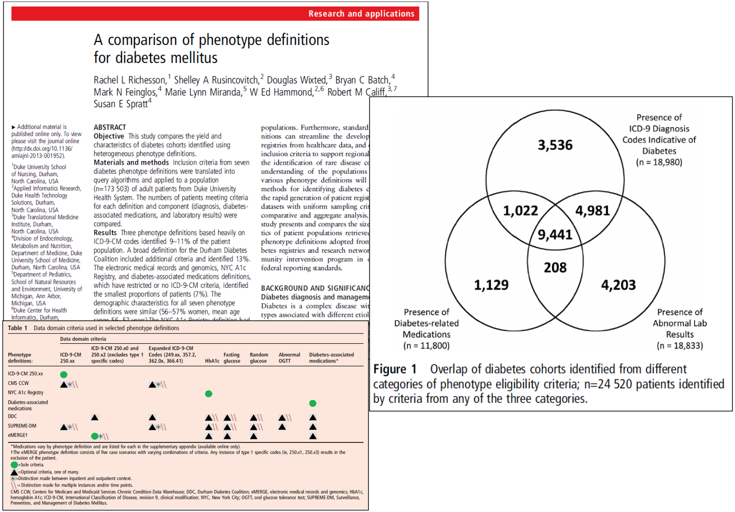
MOST EHRS ARE INCOMPLETE

- Patients seek care at multiple facilities
- Missing information on when individuals are healthy
- EHRs don’t always contain all the data you want

LINKING EHR DATA

- Data from other facilities (PCORNet)
- Claims: Center for Medicaire & Medicaid Services (CMS)
- Mortality: National Death Index (NDI) & Social Security Death Index (SSDI)
- Genetic Data
- GeoCode Information: American Community Survey (ACS)
- Personal Tracking Data: FitBit, sensors

Issues of Data Definition: What is a Diabetic?



Research and applications

A comparison of phenotype definitions for diabetes mellitus

Rachel L Richesson,[1] Shelley A Rusincovitch,[2] Douglas Wixted,[3] Bryan C Batch,[4] Mark N Feinglos,[4] Marie Lynn Miranda,[5] W Ed Hammond,[2,6] Robert M Califf,[3,7] Susan E Spratt[4]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/amiajnl-2013-001952).

[1]Duke University School of Nursing, Durham, North Carolina, USA
[2]Applied Informatics Research, Duke Health Technology Solutions, Durham, North Carolina, USA
[3]Duke Translational Medicine Institute, Durham, North Carolina, USA
[4]Division of Endocrinology, Metabolism and Nutrition, Department of Medicine, Duke University School of Medicine, Durham, North Carolina, USA
[5]Department of Pediatrics, School of Natural Resources and Environment, University of Michigan, Ann Arbor, Michigan, USA
[6]Duke Center for Health Informatics, Durham,

ABSTRACT

Objective This study compares the yield and characteristics of diabetes cohorts identified using heterogeneous phenotype definitions.

Materials and methods Inclusion criteria from seven diabetes phenotype definitions were translated into query algorithms and applied to a population (n=173 503) of adult patients from Duke University Health System. The numbers of patients meeting criteria for each definition and component (diagnosis, diabetes-associated medications, and laboratory results) were compared.

Results Three phenotype definitions based heavily on ICD-9-CM codes identified 9–11% of the patient population. A broad definition for the Durham Diabetes Coalition included additional criteria and identified 13%. The electronic medical records and genomics, NYC A1c Registry, and diabetes-associated medications definitions, which have restricted or no ICD-9-CM criteria, identified the smallest proportions of patients (7%). The demographic characteristics for all seven phenotype definitions were similar (56–57% women, mean age range 56–57 years) The NYC A1c Registry definition had

populations. Furthermore, standard nitions can streamline the develop registries from healthcare data, and inclusion criteria to support regional the identification of rare disease c understanding of the populations various phenotype definitions will methods for identifying diabetes c the rapid generation of patient regist datasets with uniform sampling cri comparative and aggregate analysis. study presents and compares the siz tics of patient populations retrieve phenotype definitions adopted fron betes registries and research netwo munity intervention program in federal reporting standards.

BACKGROUND AND SIGNIFICANC

Diabetes diagnosis and manageme

Diabetes is a complex disease wi types associated with different etiol

Table 1 Data domain criteria used in selected phenotype definitions

	Data domain criteria							
Phenotype definitions:	ICD-9-CM 250.xx	ICD-9-CM 250.x0 and 250.x2 (excludes type 1 specific codes)	Expanded ICD-9-CM Codes (249.xx, 357.2, 362.0x, 366.41)	HbA1c	Fasting glucose	Random glucose	Abnormal OGTT	Diabetes-associated medications*
ICD-9-CM 250.xx	●							
CMS CCW	▲✳\\		▲✳\\					
NYC A1c Registry				●				
Diabetes-associated medications								●
DDC		▲	▲	▲\\	▲\\	▲\\	▲\\	▲
SUPREME-DM	▲✳\\		▲✳\\	▲\\	▲\\	▲\\	▲	▲
eMERGE†		●✳\\		▲	▲	▲		▲

*Medications vary by phenotype definition and are listed for each in the supplementary appendix (available online only).

†The eMERGE phenotype definition consists of five case scenarios with varying combinations of criteria. Any instance of type 1 specific codes (ie, 250.x1, 250.x3) results in the exclusion of the patient.

●=Sole criteria.

▲=Optional criteria, one of many.

✳=Distinction made between inpatient and outpatient context.

\\ = Distinction made for multiple instances and/or time points.

CMS CCW, Centers for Medicare and Medicaid Services Chronic Condition Data Warehouse; DDC, Durham Diabetes Coalition; eMERGE, electronic medical records and genomics; HbA1c, hemoglobin A1c; ICD-9-CM, International Classification of Disease, revision 9, clinical modification; NYC, New York City; OGTT, oral glucose tolerance test; SUPREME-DM, Surveillance, Prevention, and Management of Diabetes Mellitus.

Presence of ICD-9 Diagnosis Codes Indicative of Diabetes (n = 18,980)
3,536
1,022
4,981
9,441
208
1,129
4,203
Presence of Diabetes-related Medications (n = 11,800)
Presence of Abnormal Lab Results (n = 18,833)

Figure 1 Overlap of diabetes cohorts identified from different categories of phenotype eligibility criteria; n=24 520 patients identified by criteria from any of the three categories.

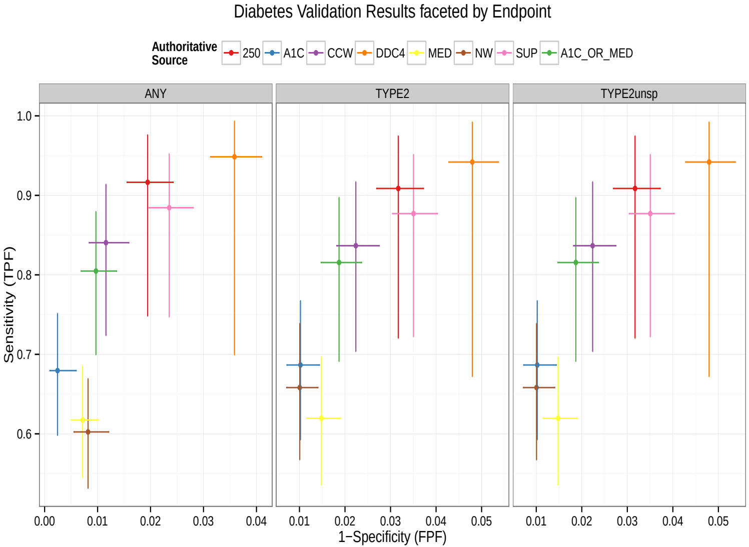
Richesson RL, Rusincovitch SA, Wixted D, Batch BC, Feinglos MN, Miranda ML, Hammond WE, Califf RM, Spratt SE. A Comparison of Phenotyp Definitions for Diabetes Mellitus. J Am Med Inf Assoc 2013 (epub ahead of print). http://www.ncbi.nlm.nih.gov/pubmed/24026307

Issues of Data Definition: What is a Diabetic?

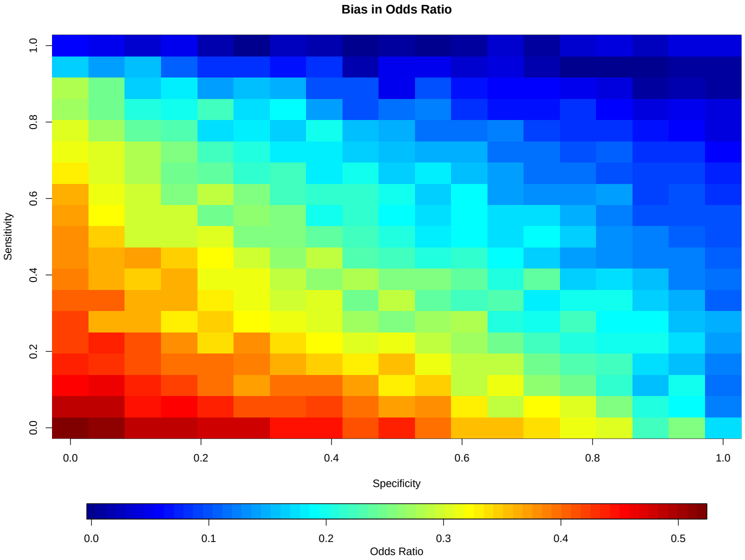
	ICD-9 250.xx	ICD-9 250.x0 & 250.x2 (exclude type I)	Expand. ICD-9 (249.xx, 357.2, 362.0x, 366.41)	HbA1c	Glucose	Abnormal OGTT	Diabetes Meds
ICD-9 250.xx	X						
CMS CCW	X*		X*				
NYC A1c Registry				X			
Meds							X
DDC		X	X	X	X	X	X
SUPREME-DM	X*		X*	X	X	X	X
eMERGE		X*		X	X		X

* Distinction between Inpatient and Outpatient Visits

Definition Differences



Impact of Poorer Definitions



ADDITIONAL PHENOTYPING CHALLENGES

- **Death:** Internal work estimates 20% capture of deaths
- **Disease Incidence:** Need to apply 'burn-in' periods
- **Censoring:** Need to apply 'burn-out' periods

Multiple Measurements per Person

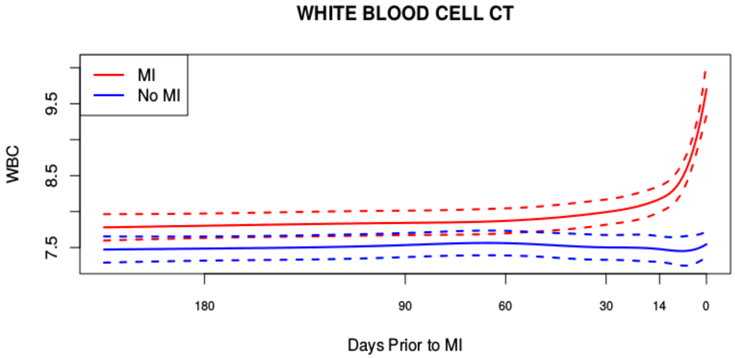
Opportunities

- Get to observe patient’s evolving health status
- More frequent visits than a typical longitudinal study
- Denser visit information

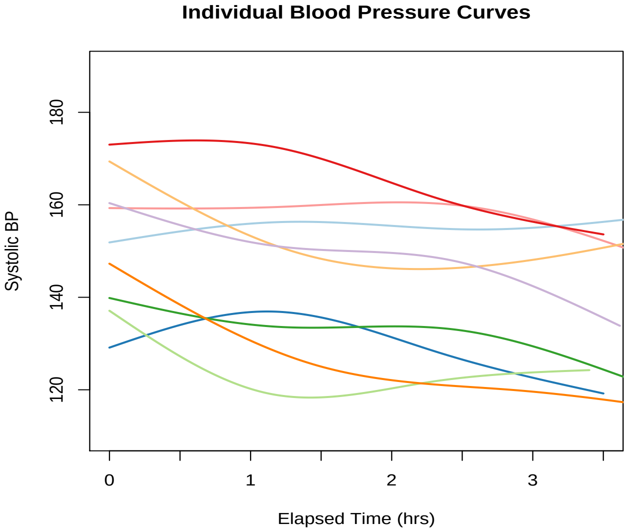
Challenges

- Visits are irregularly spaced
- Different ways to aggregate
- Don’t know what you are not seeing

Look at Changes over Long Periods of Time...



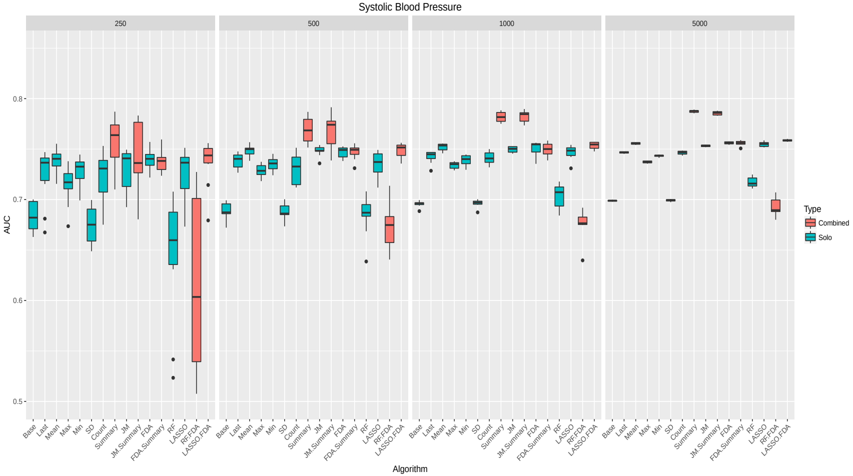
…Or Short Periods of Time



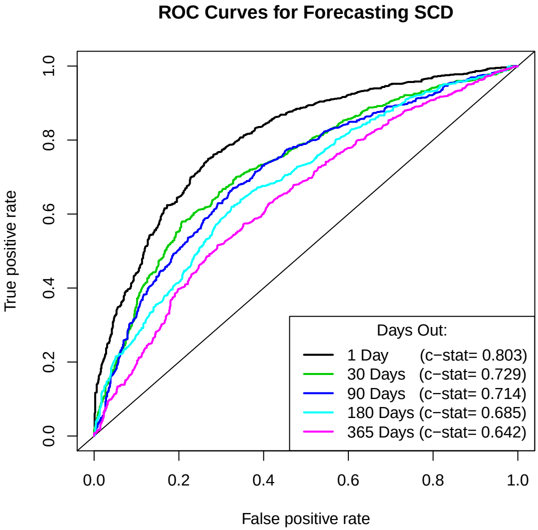
ANALYZING REPEATED MEASURES

Summarizing Data	**Modelling Progression**
Mean/Median Values	Regression Splines
Extreme Values	Functional Data Analysis
Variability	Joint Models
Number of Measurements	

Simpler Methods Often Work Best



EHR Data Optimized for Nearer Term Prediction



Top Predictors

	1 Day	7 Days	30 Days
1	LabValue: Albumin	LabValue: Albumin	LabValue: Albumin
2	Pre Systolic BP	Pre Systolic BP	Pre Systolic BP
3	Pre MAP	Pre MAP	Lowest Systolic BP
4	Pre Pulse Pressure	LabValue: WBC	LabValue: Creatinine
5	LabValue: Hemoglobin	Medication Dose: Epogen	Pre MAP
6	Lowest Systolic BP	LabValue: Creatinine	Post MAP

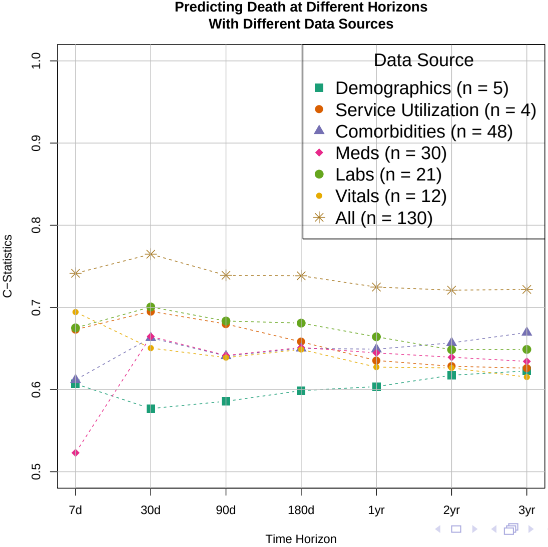
	90 Days	180 Days	365 Days
1	LabValue: Albumin	LabValue: Albumin	LabValue: Albumin
2	Pre Weight	Pre Weight	Medication Dose: Epogen
3	Pre Systolic BP	Pre Map	Age
4	Pre Pulse Pressure	Post Weight	LabValue: Creatinine
5	Medication Dose: Epogen	Medication Dose: Epogen	Pre Systolic BP
6	Post Weight	Pre Systolic BP	Pre Pulse Pressure

Top Predictors

	1 Day	7 Days	30 Days
1	LabValue: Albumin	LabValue: Albumin	LabValue: Albumin
2	Pre Systolic BP	Pre Systolic BP	Pre Systolic BP
3	Pre MAP	Pre MAP	Lowest Systolic BP
4	Pre Pulse Pressure	LabValue: WBC	LabValue: Creatinine
5	LabValue: Hemoglobin	Medication Dose: Epogen	Pre MAP
6	Lowest Systolic BP	LabValue: Creatinine	Post MAP

	90 Days	180 Days	365 Days
1	LabValue: Albumin	LabValue: Albumin	LabValue: Albumin
2	Pre Weight	Pre Weight	Medication Dose: Epogen
3	Pre Systolic BP	Pre Map	Age
4	Pre Pulse Pressure	Post Weight	LabValue: Creatinine
5	Medication Dose: Epogen	Medication Dose: Epogen	Pre Systolic BP
6	Post Weight	Pre Systolic BP	Pre Pulse Pressure

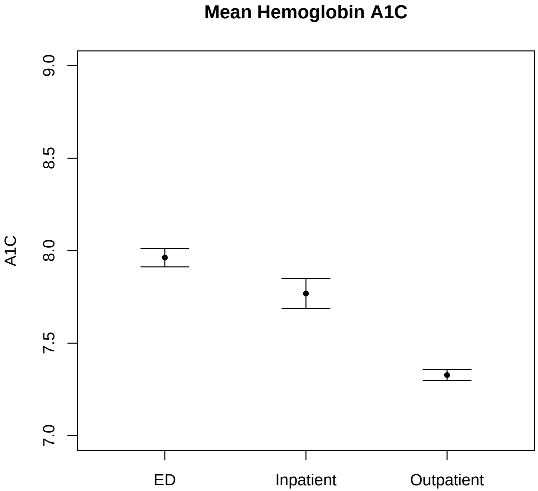
Different Data Elements have Different Predictability



Biases in EHRs: Informed Presence

- We only see patients when they are sick
- We only see information that is deemed important
- Different environments have different policies

Informed Presence I: Where a Person Seeks Care is Informative



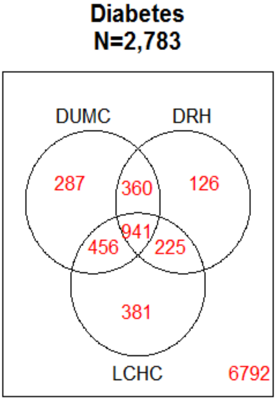
Location Impacts Inference

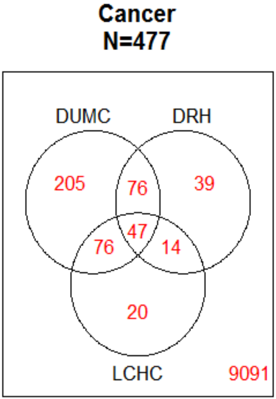
- Hazard Ratio for HgB A1C for time to Myocardial Infarction

Type	Hazard Ratio	P-value
Unadjusted	1.06 (1.01, 1.11)	0.026
Adjusted for Location	0.97 (0.92, 1.02)	0.178
OP Only	1.07 (1.00, 1.14)	0.044
ED Only	0.94 (0.89, 0.99)	0.022

- Interaction between A1C and location

Informed Presence II: Which Hospital a Patient Uses is Informative



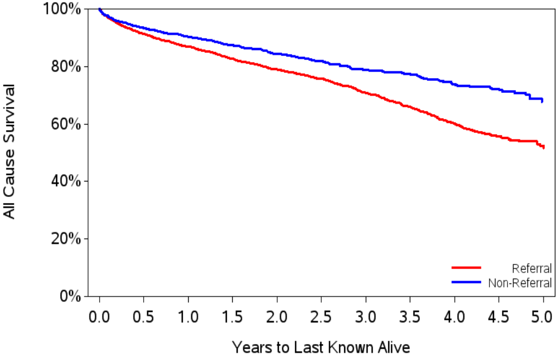


Facility Impacts Inference

- Odds Ratio for Cancer Status on Diabetes

Location	**Odds Ratio**	**95% CI**
All Facilities	1.69	(1.36, 2.10)
DUMC Only	1.46	(1.15, 1.87)
DRH Only	0.89	(0.63, 1.26)
LCHC Only	1.08	(0.74, 1.56)

INFORMED PRESENCE III: REFERAL HOSPITALS ARE AN *Admixed* POPULATION



Number at risk											
Referral	5522	3307	2690	2159	1748	1360	995	697	474	282	65
Non-Referral	2114	1532	1318	1110	882	697	519	387	262	171	64

ADMIXTURE BIAS

- Comparison of Local and Referal Patients at Cardiac Catheterization Lab

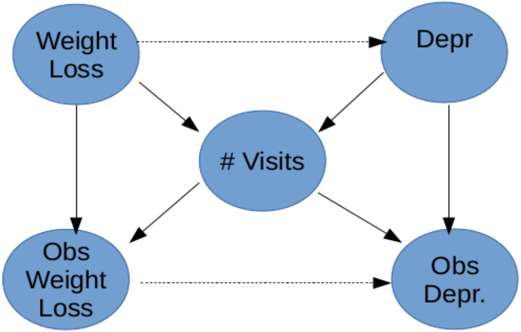
Local Patients	**Referal Patients**
Older	Younger
More Comorbidities	More severe valve disease
Disease due to ageing	Disease due systematic factors
Better outcomes	More follow-up procedures

Informed Presence IV: Need to Account for Number of Encounters

Regression of Depression on Weight Loss

	Odds Ratio	Δ log(OR)	Δ OR
Minimally Adjusted	3.98 (3.81, 4.17)	—	—
+ No. Encounters	2.37 (2.26, 2.50)	-0.52	-1.61
+ Comorbidities	2.82 (2.69, 2.96)	-0.35	-1.16
+ No. Encounters & Comorb	2.30 (2.18, 2.42)	-0.55	-1.68

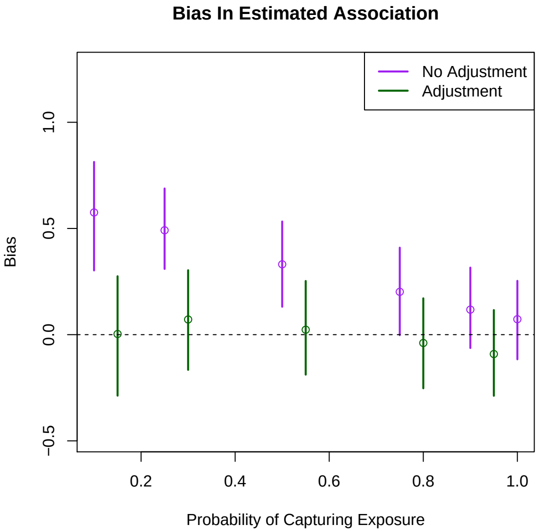
Number of Encounters Potential Confounder



Need to Account for Number of Encounters

		Median Number of Encounters	
	Sensitivity	Without Condition	With Condition
Depression	56.3%	6	38
Weight Loss	9.3%	7	45

Number of Encounters Potential Confounder



1. Structure of Electronic Health Records

2. Research with EHR Data

3. Concluding Thoughts

EXTRA CARE NEEDED

- Need to be mindful from where the data come
- There is not always one way to turn raw data into analytic data
- Which data to *cut* is more important than how you analyze it
- New analytic techniques may be useful/necessary

Questions to ask when Designing EHR Based Studies

- Where in the health system are the data collected?
- What is the coverage/catchment area of your health system?
- Is the patient population receiving care across multiple institutions/centers?
- Do the data constitute different catchments? (Admixture)
- How are you defining exposures and outcomes? (Phenotyping)
- How are you defining person-time?
  - What is an appropriate burn-in period to define a cohort?
  - Is a burn-out period necessary to define censoring?
- Do different populations produce more information (i.e. sicker patients have more encounters)?

ADDITIONAL FRONTIERS

- Micro-randomized trials
- Integration of external data
- Real time risk assessment

IS IT ALL BAD?

A LOT OF OPPORTUNITIES WITH EHRS

- More studies
- Cheaper studies
- Faster studies
- (Perhaps) More representative studies

References

- Phelan, M., Bhavsar, N.A., Goldstein, B.A. Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference. *eGEMS*, In Press.
- Goldstein B.A., Pomann, GM, Winkelmayer, WC., and Pencina, MJ. Comparison of risk prediction methods using repeated observations with application to Electronic Health Record. *Statistics in Medicine*, 2017, 36(17) 2750 2763.
- Goldstein B.A. Pencina M. J., Montez-Rath, M.E. and Winkelmayer W.C. Value of different categories of information available in electronic health records for prediction of mortality in patients on dialysis. *Journal of the American Medical Informatics Association*, 2017, 24(1): 176-181.
- Goldstein B.A., Navar, A.M., Pencina, M.J., and Ioannidis, J.P. Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. *Journal of the American Medical Informatics Association*, 2017, 24(1): 198-208.
- Spratt, SE. Pereira, K., Granger, BB.,Batch, BC., Phelan, M., Pencina, M., Miranda, ML., Boulware, E., Lucas, JE., Nelson, CL., Neely, B., Goldstein, BA., Barth, P., Richesson, RL., Riley, IL., Corsino, L., McPeek Hinz, ER., Rusincovitch, S., Green, J., Barton, AB., and the DDC Phenotype Group., Assessing Electronic Health Record Phenotypes against Gold-Standard Diagnostic Criteria for Diabetes Mellitus. *Journal of the Medical Informatics Association*, 2017, 24, e121-e128.
- Goldstein B.A., Bhavasr, N.A., Phelan, M., and Pencina, M.J. Controlling for informed presence bias due to the number of health encounters in an Electronic Health Record. *American Journal of Epidemiology*, 2016, 184(11): 847-855.
- Goldstein BA, Cheng TI, and Winkelmayer, WC. Classifying Individuals Based on a Densely Captured Sequence of Vital Signs: An Example using Repeated Blood Pressure Measurements during Hemodialysis Treatment. *Journal of Biomedical Informatics*, 2015, 57: 219-224.
- Goldstein, B.A., Assimes TL., Winkelmayer WC, and Hastie T. Detecting clinically meaningful biomarkers with repeated measurements in an Electronic Health Record. *Biometrics*, 2015, 71, 478-486.
- Goldstein, B.A., Chang, TI, Mitani, A.A., Assimes TL., and Winkelmayer WC. Near-term prediction of sudden cardiac death in older patients using the electronic health records. *Clinical Journal of the American Society of Nephrology*, 9, 2014, 82-91.
- Richesson RL, Rusincovitch SA, Wixted D, Batch BC, Feinglos MN, Miranda ML, Hammond WE, Califf RM, and Spratt SE.A comparison of phenotype definitions for diabetes mellitus. *Journal of the American Medical Informatics Association*, 2013, 20, e319-326.