Machine Learning for Healthcare

HST.956, 6.S897

Lecture 4: Risk stratification

David Sontag

CSAIL

INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

HEALTH SCIENCES & TECHNOLOGY

Course announcements

- Recitation Friday at 2pm (4-153) – optional
- No class this Tuesday
- Problem set 1 due next Thursday, Feb 21
- Sign up for lecture scribing or MLHC community consulting
- Readings will be posted several days ahead
- All course communication through Piazza

Outline for today's class

1. Risk stratification
2. Case study: Early detection of Type 2 diabetes
   - Framing as supervised learning problem
   - Deriving labels
   - Evaluating risk stratification algorithms
3. Subtleties with ML-based risk stratification
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

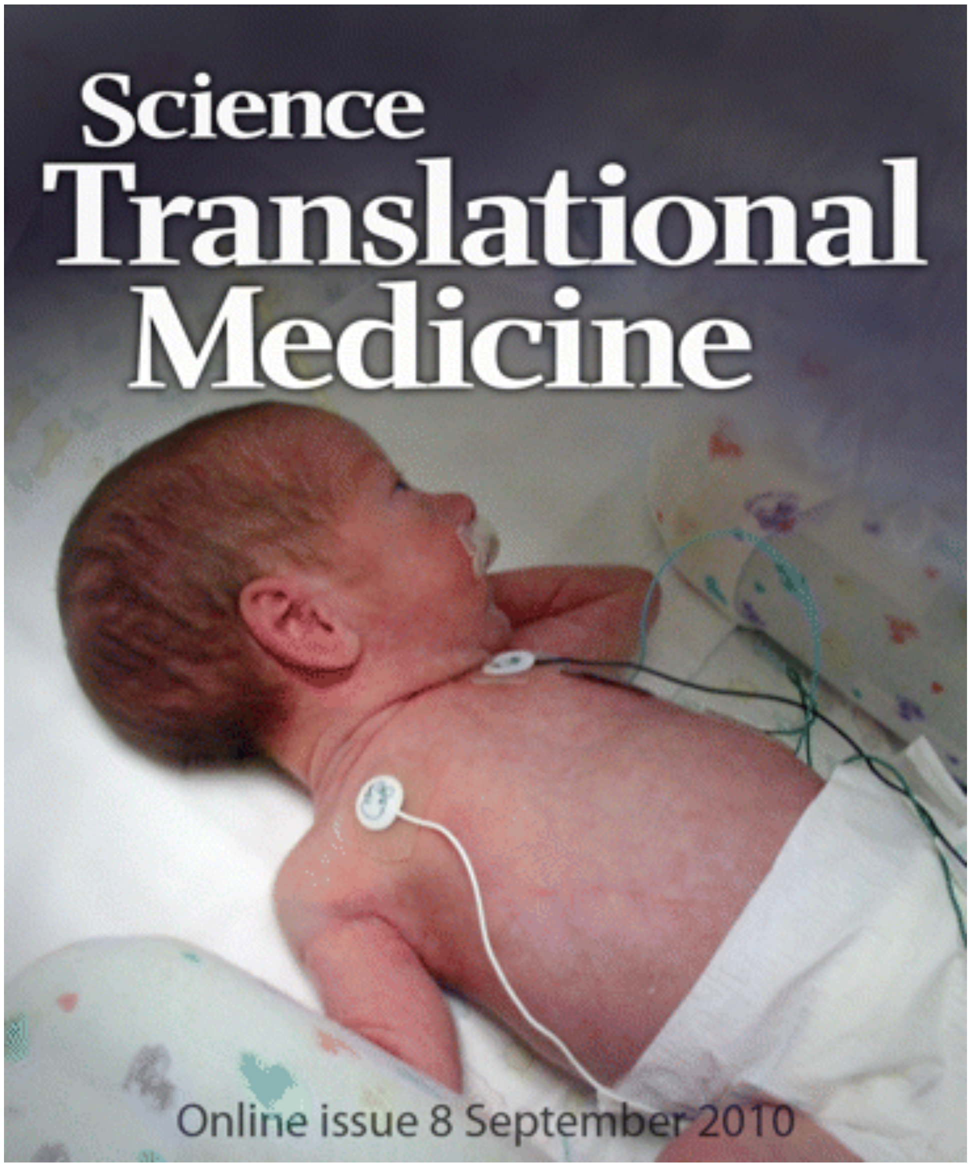
Outline for today's class

1. **Risk stratification**
2. Case study: Early detection of Type 2 diabetes
   - Framing as supervised learning problem
   - Deriving labels
   - Evaluating risk stratification algorithms
3. Subtleties with ML-based risk stratification
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

What *is* risk stratification?

- Separate a patient population into **high-risk** and **low-risk** of having an outcome
  - Predicting something in the future
  - Goal is different from diagnosis, with distinct performance metrics
- Coupled with **interventions** that target high-risk patients
- Goal is typically to reduce cost and improve patient outcomes

Examples of risk stratification



Preterm infant's risk of severe morbidity?

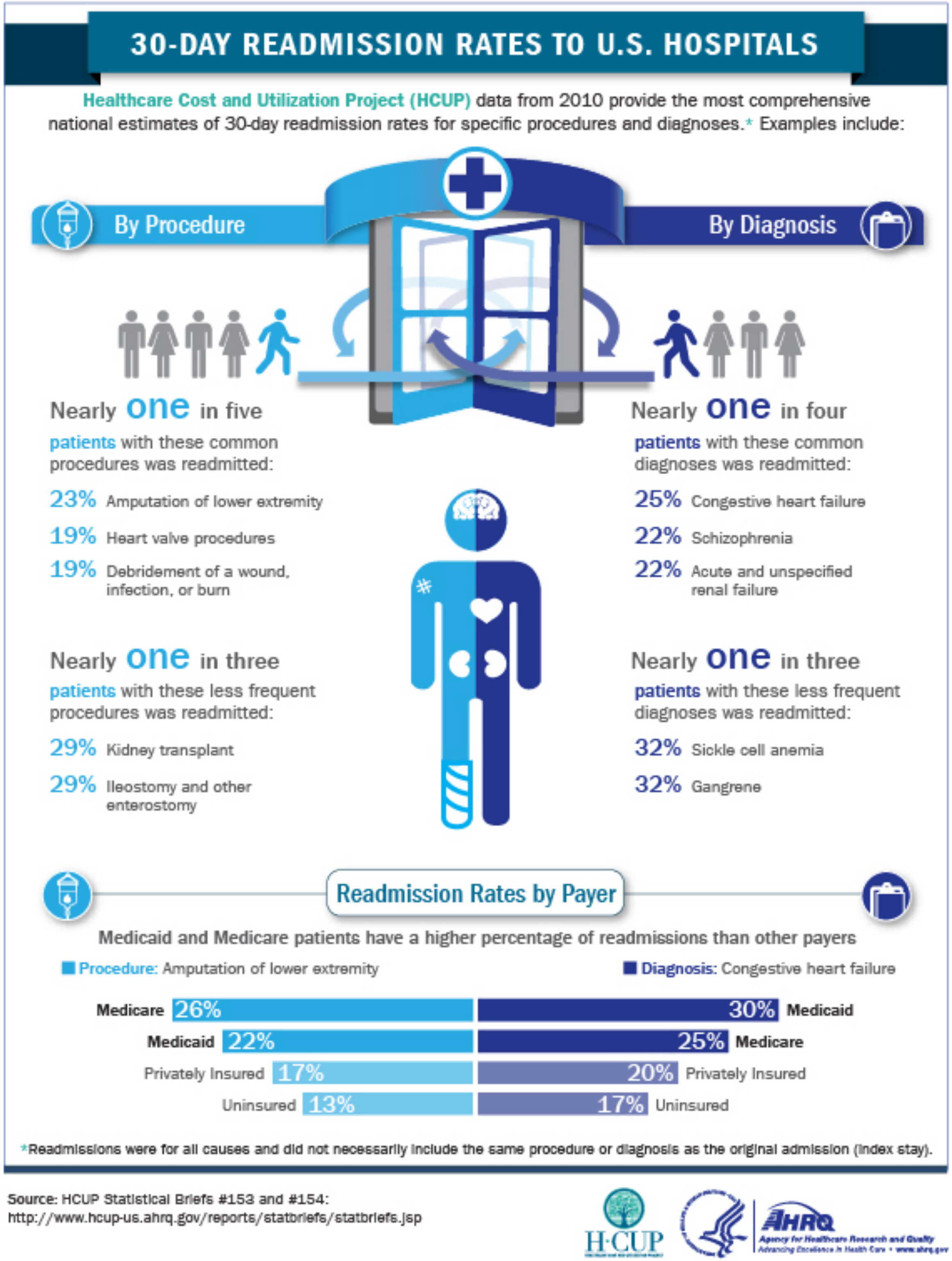
(Saria et al., Science Translational Medicine 2010)

Examples of risk stratification

Figure source: https://www.drmani.com/heart-attack/

Does this patient need to be admitted to the coronary-care unit?

(Pozen et al., NEJM 1984)



Likelihood of hospital readmission?

Figure source: https://www.air.org/project/revolving-door-u-s-hospital-readmissions-diagnosis-and-procedure

Old vs. New

- Traditionally, risk stratification was based on simple scores using human-entered data

APGAR SCORING SYSTEM

	0 Points	1 Point	2 Points	Points totaled
Activity (muscle tone)	Absent	Arms and legs flexed	Active movement	
Pulse	Absent	Below 100 bpm	Over 100 bpm	
Grimace (reflex irritability)	Flaccid	Some flexion of Extremities	Active motion (sneeze, cough, pull away)	
Appearance (skin color)	Blue, pale	Body pink, Extremities blue	Completely pink	
Respiration	Absent	Slow, irregular	Vigorous cry	

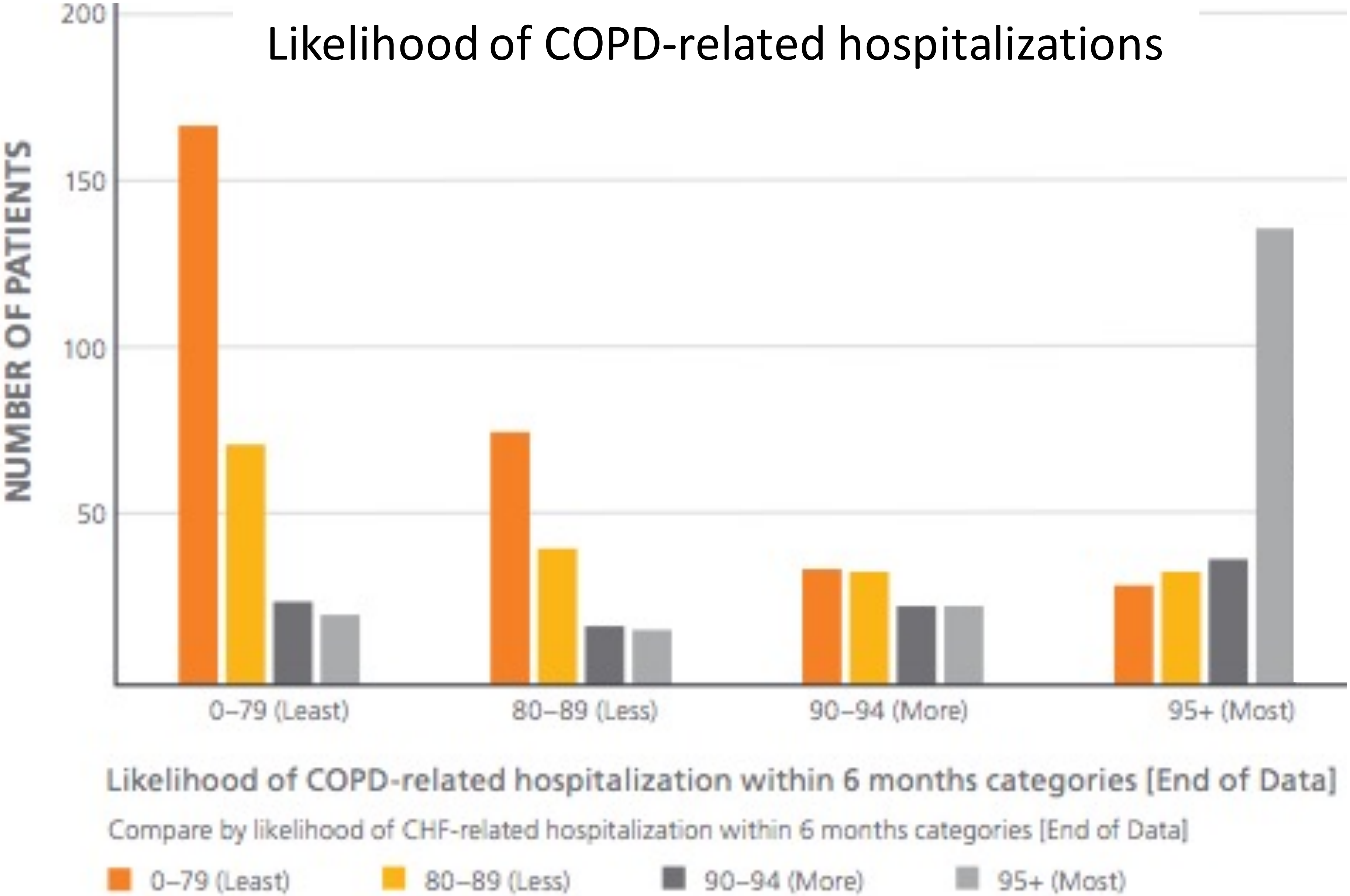
Severely depressed	0-3
Moderately depressed	4-6
Excellent condition	7-10

Old vs. New

- Traditionally, risk stratification was based on simple scores using human-entered data
- Now, based on machine learning on high-dimensional data
  - Fits more easily into workflow
  - Higher accuracy
  - Quicker to derive (can special case)
- **But, new dangers introduced with ML approach – to be discussed**

Example commercial product

Likelihood of COPD-related hospitalizations



Optum Whitepaper, “Predictive analytics: Poised to drive population health"

Example commercial product

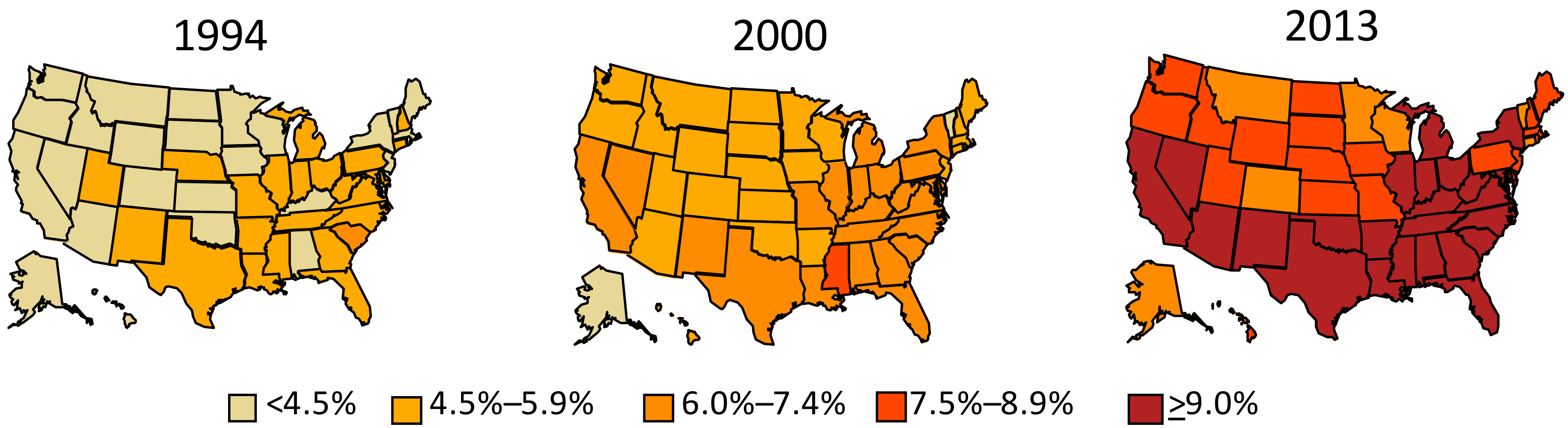
High-risk diabetes patients missing tests	# of A1c tests	# of LDL tests	Last A1c	Date of last A1c	Last LDL	Date of last LDL
Patient 1	2	0	9.2	5/3/13	N/A	N/A
Patient 2	2	0	8	1/30/13	N/A	N/A
Patient 3	0	0	N/A	N/A	N/A	N/A
Patient 4	0	2	N/A	N/A	133	8/9/13
Patient 5	0	0	N/A	N/A	N/A	N/A
Patient 6	0	1	N/A	N/A	115	7/16/13
Patient 7	1	0	10.8	9/18/13	N/A	N/A
Patient 8	0	0	N/A	N/A	N/A	N/A
Patient 9	0	0	N/A	N/A	N/A	N/A
Patient 10	0	0	N/A	N/A	N/A	N/A

Optum Whitepaper, “Predictive analytics: Poised to drive population health"

Outline for today's class

1. Risk stratification
2. **Case study: Early detection of Type 2 diabetes**
   – Framing as supervised learning problem
   – Deriving labels
   – Evaluating risk stratification algorithms
3. Subtleties with ML-based risk stratification
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

Type 2 Diabetes: A Major public health challenge



$245 billion: Total costs of diagnosed diabetes in the United States in 2012
$831 billion: Total fiscal year federal budget for healthcare in the United States in 2014

Type 2 Diabetes Can Be Prevented *

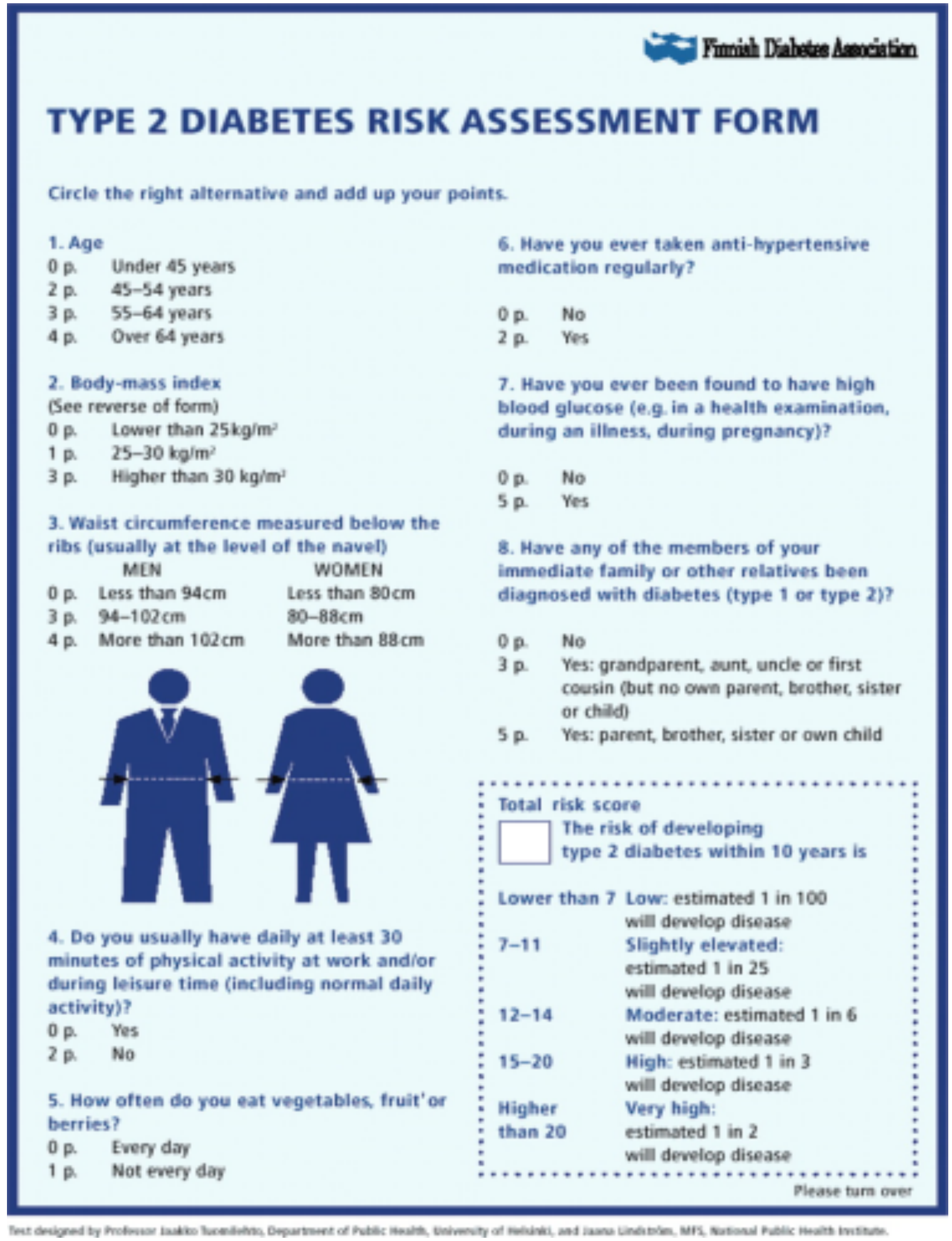
Requirement for successful large scale prevention program

1. Detect/reach truly at risk population
2. Improve the interventions
3. Lower the cost of intervention

* Diabetes Prevention Program Research Group. "Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin." The New England journal of medicine 346.6 (2002): 393.

Traditional Risk Prediction Models

- Successful Examples
  - ARIC
  - KORA
  - FRAMINGHAM
  - AUSDRISC
  - FINDRISC
  - San Antonio Model
- Easy to ask/measure in the office, or for patients to do online
- Simple model:
  can calculate scores by hand



Finnish Diabetes Association

TYPE 2 DIABETES RISK ASSESSMENT FORM

Circle the right alternative and add up your points.

1. Age
0 p. Under 45 years
2 p. 45–54 years
3 p. 55–64 years
4 p. Over 64 years

2. Body-mass index
(See reverse of form)
0 p. Lower than 25 kg/m²
1 p. 25–30 kg/m²
3 p. Higher than 30 kg/m²

3. Waist circumference measured below the ribs (usually at the level of the navel)

	MEN	WOMEN
0 p.	Less than 94 cm	Less than 80 cm
3 p.	94–102 cm	80–88 cm
4 p.	More than 102 cm	More than 88 cm

4. Do you usually have daily at least 30 minutes of physical activity at work and/or during leisure time (including normal daily activity)?
0 p. Yes
2 p. No

5. How often do you eat vegetables, fruit or berries?
0 p. Every day
1 p. Not every day

6. Have you ever taken anti-hypertensive medication regularly?
0 p. No
2 p. Yes

7. Have you ever been found to have high blood glucose (e.g. in a health examination, during an illness, during pregnancy)?
0 p. No
5 p. Yes

8. Have any of the members of your immediate family or other relatives been diagnosed with diabetes (type 1 or type 2)?
0 p. No
3 p. Yes: grandparent, aunt, uncle or first cousin (but no own parent, brother, sister or child)
5 p. Yes: parent, brother, sister or own child

Total risk score
The risk of developing type 2 diabetes within 10 years is

Lower than 7	Low: estimated 1 in 100 will develop disease
7–11	Slightly elevated: estimated 1 in 25 will develop disease
12–14	Moderate: estimated 1 in 6 will develop disease
15–20	High: estimated 1 in 3 will develop disease
Higher than 20	Very high: estimated 1 in 2 will develop disease

Please turn over

Text designed by Professor Jaakko Tuomilehto, Department of Public Health, University of Helsinki, and Jaana Lindström, MFS, National Public Health Institute.

Challenges of Traditional Risk Prediction Models

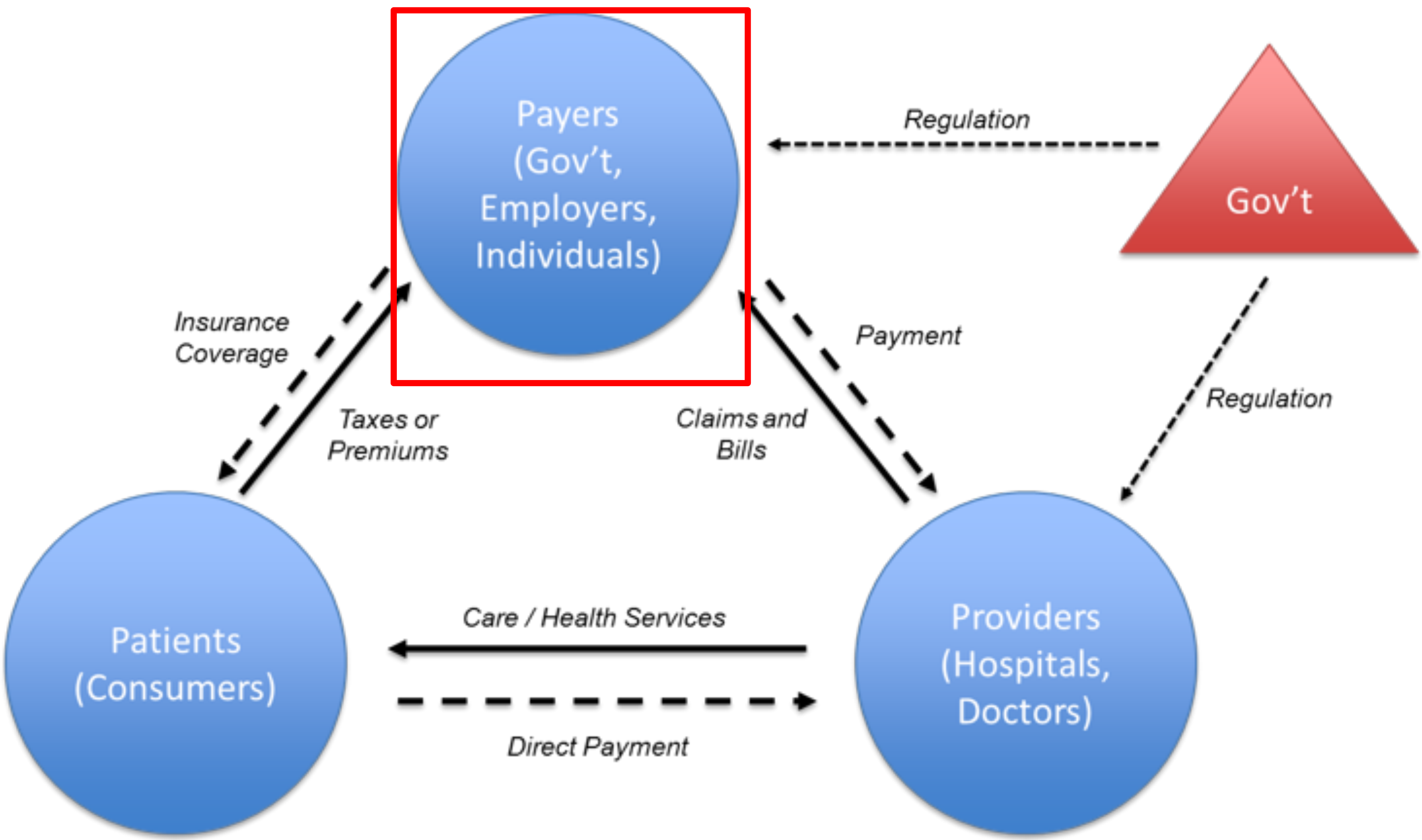
- A screening step needs to be done for every member in the population
  - Either in the physician's office or as surveys
  - Costly and time-consuming
  - Infeasible for regular screening for millions of individuals
- Models not easy to adapt to multiple surrogates, when a variable is missing
  - Discovery of surrogates not straightforward

Population-Level Risk Stratification

- Key idea: Use readily available administrative, utilization, and clinical data
- Machine learning will find surrogates for risk factors that would otherwise be missing
- Perform risk stratification at the population level – millions of patients

[Razavian, Blecker, Schmidt, Smith-McLallen, Nigam, Sontag. *Big Data.* '16]

Health stakeholders



Source for figure:
http://www.mahesh-vc.com/blog/understanding-whos-paying-for-what-in-the-healthcare-industry

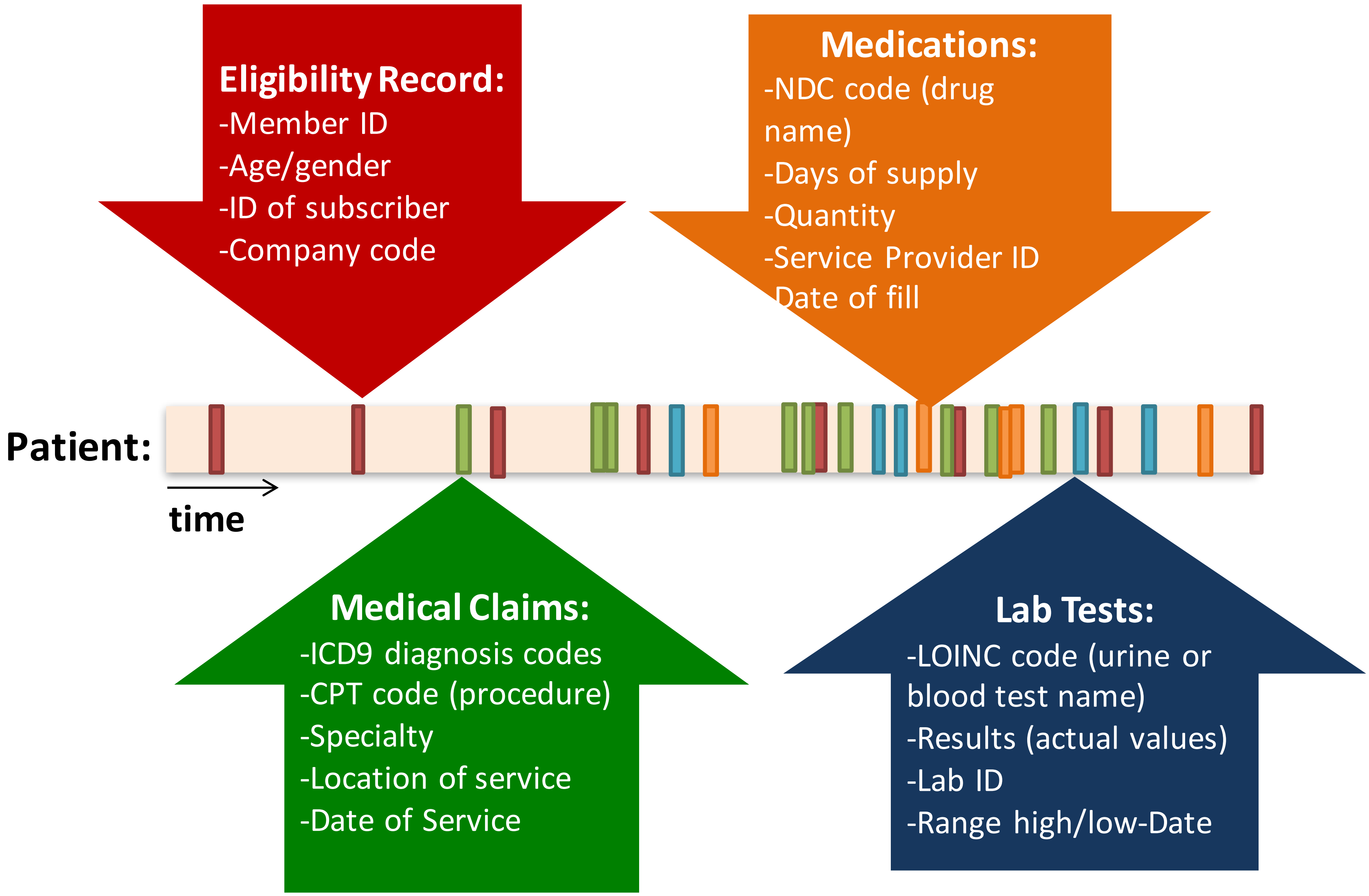
A Data-Driven approach on Longitudinal Data

- Looking at individuals who got diabetes *today,* (compared to those who didn't)
  - Can we infer which variables in their record could have predicted their health outcome?

A Few Years Ago

Today

Administrative & Clinical Data



Top diagnosis codes

Disease	count
4011 Benign hypertension	447017
2724 Hyperlipidemia NEC/NOS	382030
4019 Hypertension NOS	372477
25000 DMII wo cmp nt st uncntr	339522
2720 Pure hypercholesterolem	232671
2722 Mixed hyperlipidemia	180015
V7231 Routine gyn examination	178709
2449 Hypothyroidism NOS	169829
78079 Malaise and fatigue NEC	149797
V0481 Vaccin for influenza	147858
7242 Lumbago	137345
V7612 Screen mammogram NEC	129445
V700 Routine medical exam	127848

Disease	count
53081 Esophageal reflux	121064
42731 Atrial fibrillation	113798
7295 Pain in limb	112449
41401 Crnry athrscl natve vssl	104478
2859 Anemia NOS	103351
78650 Chest pain NOS	91999
5990 Urin tract infection NOS	87982
V5869 Long-term use meds NEC	85544
496 Chr airway obstruct NEC	78585
4779 Allergic rhinitis NOS	77963
41400 Cor ath unsp vsl ntv/gft	75519

Disease	count
71947 Joint pain-ankle	28648
3004 Dysthymic disorder	28530
2689 Vitamin D deficiency NOS	28455
V7281 Preop cardiovsclr exam	27897
7243 Sciatica	27604
78791 Diarrhea	27424
V221 Supervis oth normal preg	27320
36501 Opn angl brderln lo risk	26033
37921 Vitreous degeneration	25592
4241 Aortic valve disorder	25425
61610 Vaginitis NOS	24736
70219 Other sborheic keratosis	24453
3804 Impacted cerumen	24046

Out of 135K patients who had laboratory data

Top lab test results

Lab test	
2160-0 Creatinine	1284737
3094-0 Urea nitrogen	1282344
2823-3 Potassium	1280812
2345-7 Glucose	1299897
1742-6 Alanine aminotransferase	1187809
1920-8 Aspartate aminotransferase	1187965
2885-2 Protein	1277338
1751-7 Albumin	1274166
2093-3 Cholesterol	1268269
2571-8 Triglyceride	1257751
13457-7 Cholesterol.in LDL	1241208
17861-6 Calcium	1165370
2951-2 Sodium	1167675

Lab test	
2085-9 Cholesterol.in HDL	1155666
718-7 Hemoglobin	1152726
4544-3 Hematocrit	1147893
9830-1 Cholesterol.total/Cholesterol.in HDL	1037730
33914-3 Glomerular filtration rate/1.73 sq M.predicted	561309
785-6 Erythrocyte mean corpuscular hemoglobin	1070832
6690-2 Leukocytes	1062980
789-8 Erythrocytes	1062445
787-2 Erythrocyte mean corpuscular volume	1063665

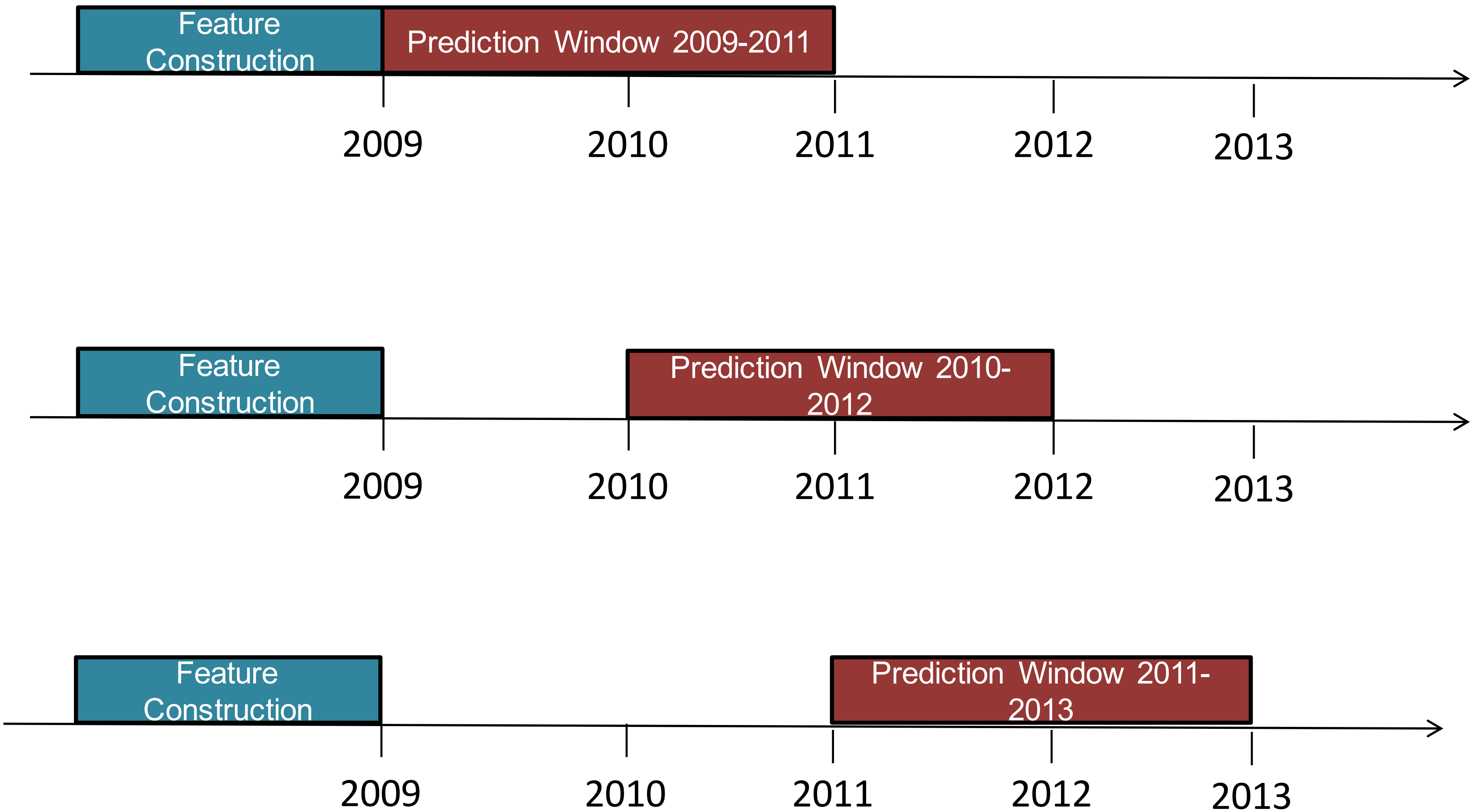
Lab test	
770-8 Neutrophils/100 leukocytes	952089
731-0 Lymphocytes	943918
704-7 Basophils	863448
711-2 Eosinophils	935710
5905-5 Monocytes/100 leukocytes	943764
706-2 Basophils/100 leukocytes	863435
751-8 Neutrophils	943232
742-7 Monocytes	942978
713-8 Eosinophils/100 leukocytes	933929
3016-3 Thyrotropin	891807
4548-4 Hemoglobin A1c/Hemoglobin.total	527062

Count of people who have the test result (ever)

Outline for today's class

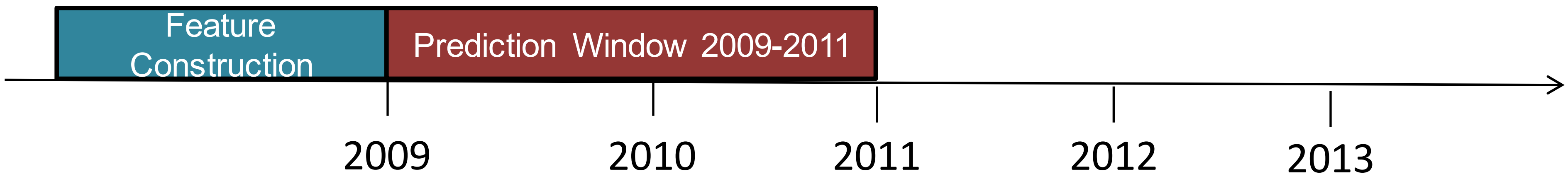
1. Risk stratification
2. Case study: Early detection of Type 2 diabetes
   - **Framing as supervised learning problem**
   - Deriving labels
   - Evaluating risk stratification algorithms
3. Subtleties with ML-based risk stratification
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

Framing for supervised machine learning



Gap is important to prevent label leakage

Framing for supervised machine learning

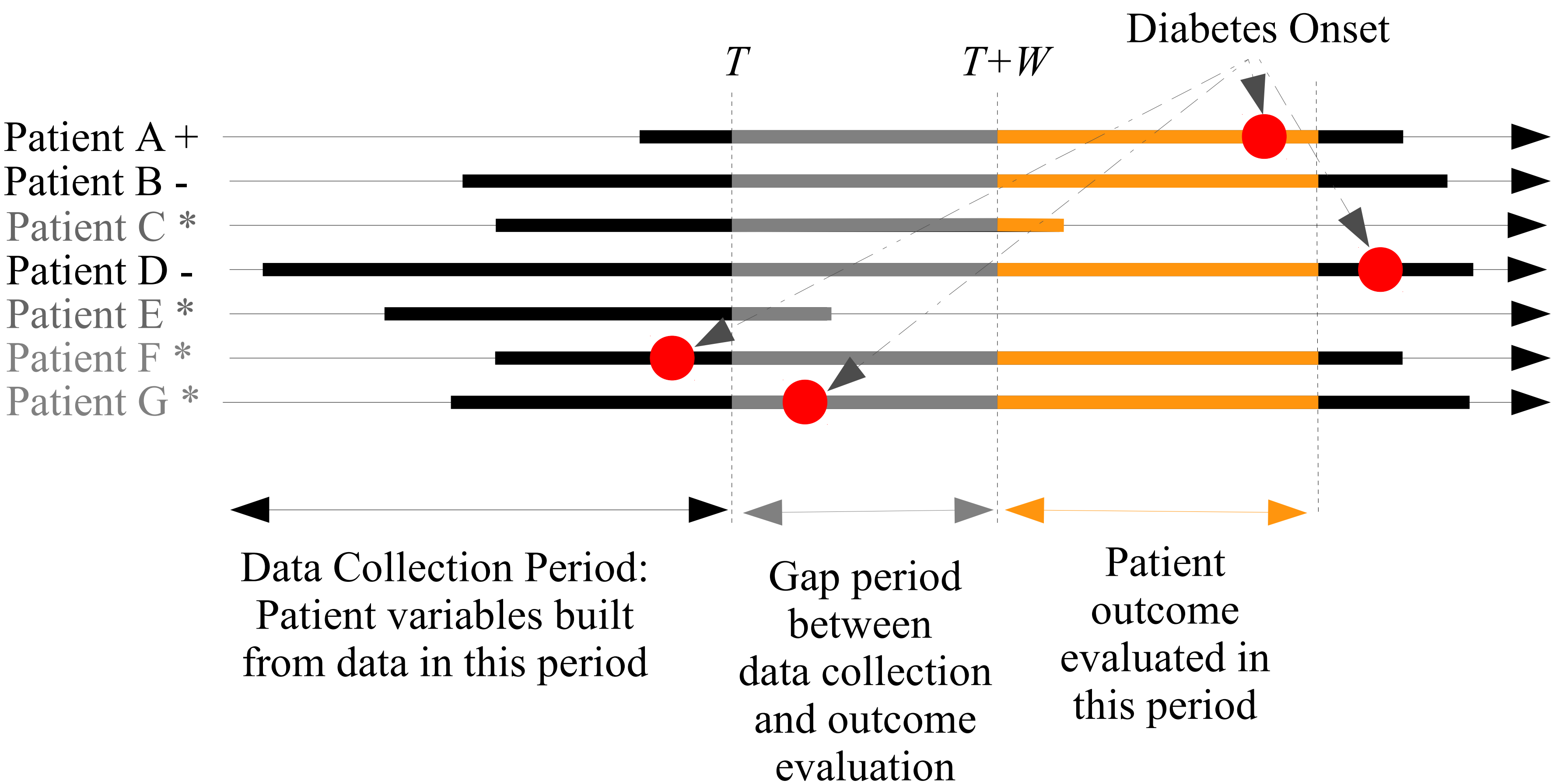


Problem: Data is censored!

- Patients change health insurers frequently, but data doesn't follow them
- *Left censored*: may not have enough data to derive features
- *Right censored*: may not know label

Reduction to binary classification

Exclude patients that are left- and right-censored.



This is an example of alignment by *absolute time*

Alternative framings

- Align by relative time, e.g.
  - 2 hours into patient stay in ER
  - Every time patient sees PCP
  - When individual turns 40 yrs old
- Align by data availability

NOTE:

- If multiple data points per patient, make sure each patient in *only* train, validate, or test

Methods

- L1 Regularized Logistic Regression
  - Simultaneously optimizes predictive performance *and*
  - Performs feature selection, choosing the subset of the features that are most predictive
- This prevents overfitting to the training data

L1 regularization

- Penalizing the L1 norm of the weight vector leads to *sparse* (read: many 0's) solutions for *w.*

$$\min_w \sum_i \ell(x_i, y_i; w) + \lambda ||w||_1 \qquad ||\vec{w}||_1 = \sum_d |w_d|$$

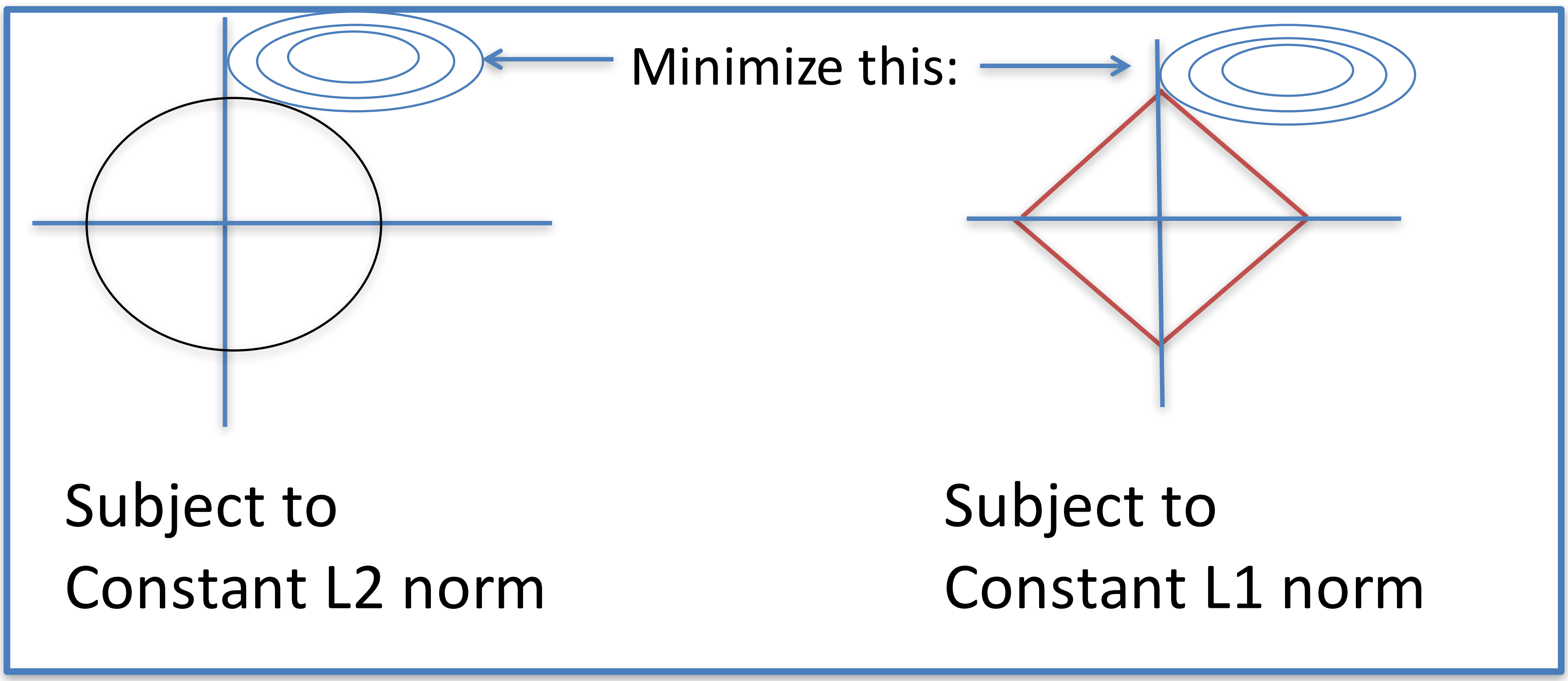
instead of

$$\min_w \sum_i \ell(x_i, y_i; w) + \lambda ||w||_2^2 \qquad ||\vec{w}||_2^2 = \sum_d w_d^2$$

- Why?

L1 regularization

- Penalizing the L1 norm of the weight vector leads to *sparse* (read: many 0's) solutions for *w.*



L1 regularization

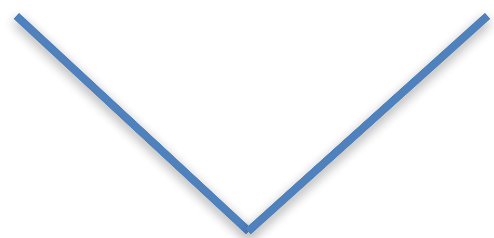
- Penalizing the L1 norm of the weight vector leads to *sparse* (read: many 0's) solutions for *w.*

Intuition #2 – w.w.g.d.d

(What would gradient descent do?)

$$\frac{d}{dw_i}\lambda||w||_2^2 = \pm\lambda w_i 2$$

$$\frac{d}{dw_i}\lambda|w| = \pm\lambda$$



L1 regularization

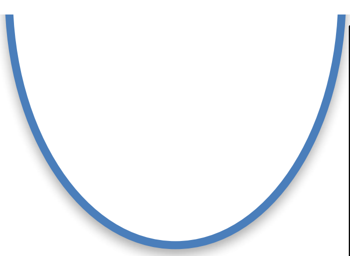
- Penalizing the L1 norm of the weight vector leads to *sparse* (read: many 0's) solutions for *w.*

Intuition #2 – w.w.g.d.d

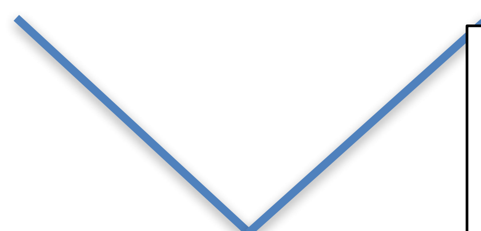
(What would gradient descent do?)

$$\frac{d}{dw_i}\lambda||w||_2^2 = \pm\lambda w_i 2$$

$$\frac{d}{dw_i}\lambda|w| = \pm\lambda$$



The push towards 0 gets weaker as wi gets smaller



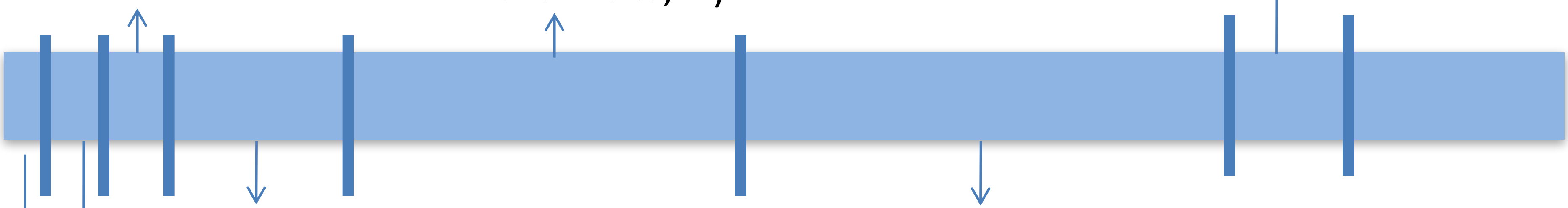
Always pushes elements of wi towards 0

Features used in models

Service place
(urgent care, inpatient, outpatient, …)

Medications taken (999 features)
(laxatives, metformin, anti-arthritics, …)

Procedures performed (457 features)



Specialty of doctors seen
(cardiology, rheumatology, …)

Laboratory indicators (7000 features)

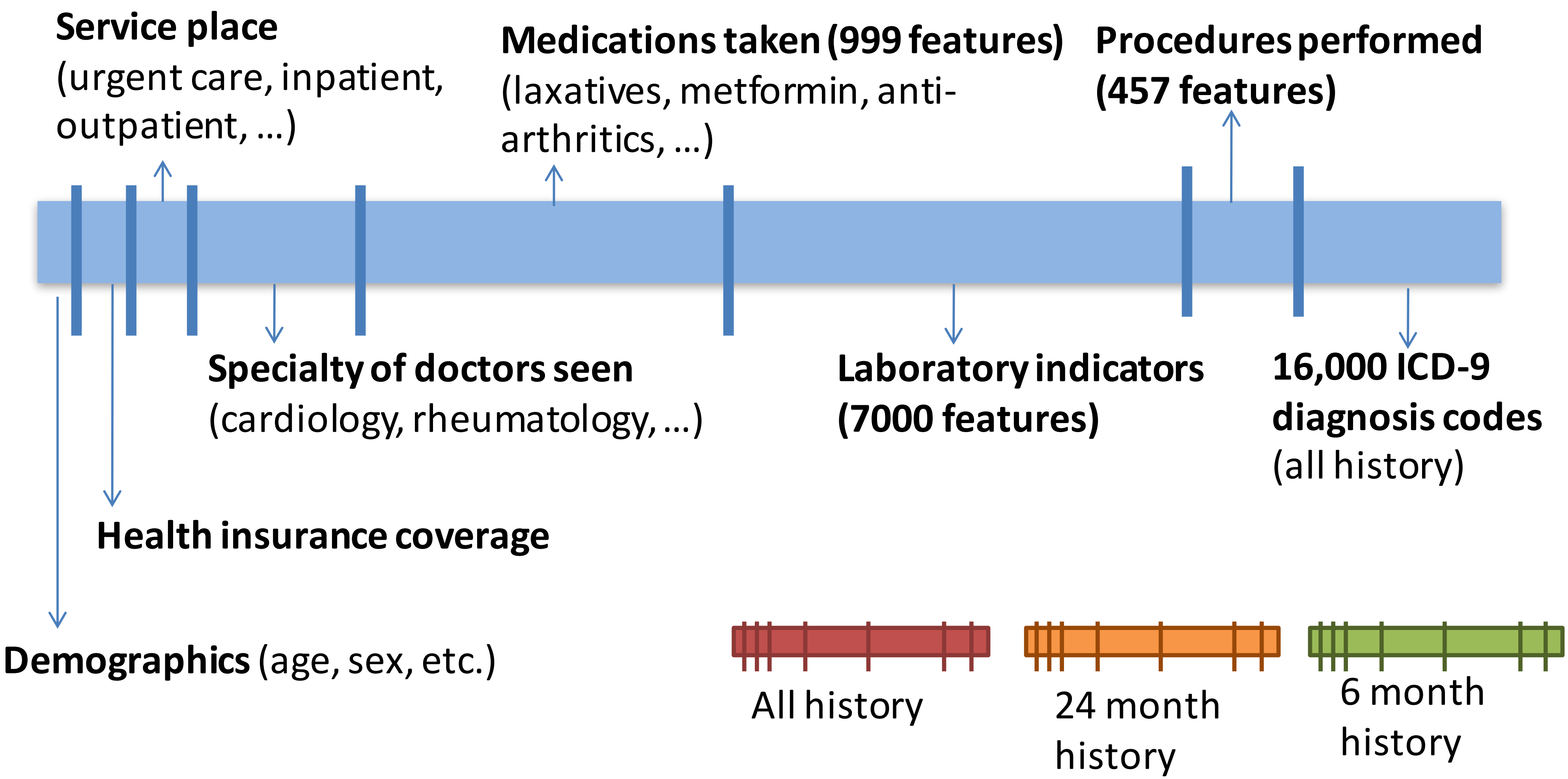
Health insurance coverage

Demographics (age, sex, etc.)

For the 1000 most frequent lab tests:

- Was the test ever administered?
- Was the result ever low?
- Was the result ever high?
- Was the result ever normal?
- Is the value increasing?
- Is the value decreasing?
- Is the value fluctuating?

Features used in models

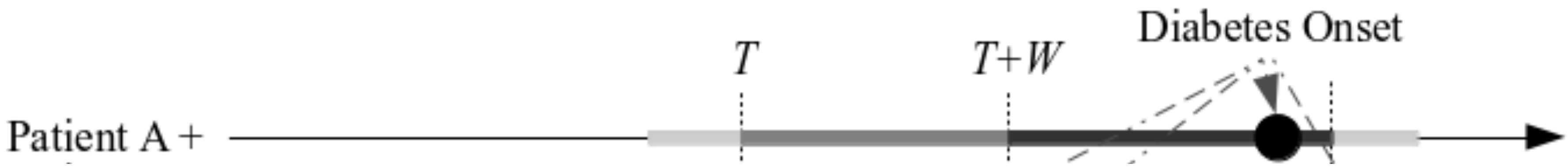


Total features per patient: 42,000

Outline for today's class

1. Risk stratification
2. Case study: Early detection of Type 2 diabetes
   – Framing as supervised learning problem
   – **Deriving labels**
   – Evaluating risk stratification algorithms
3. Subtleties with ML-based risk stratification
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

Where do the labels come from?

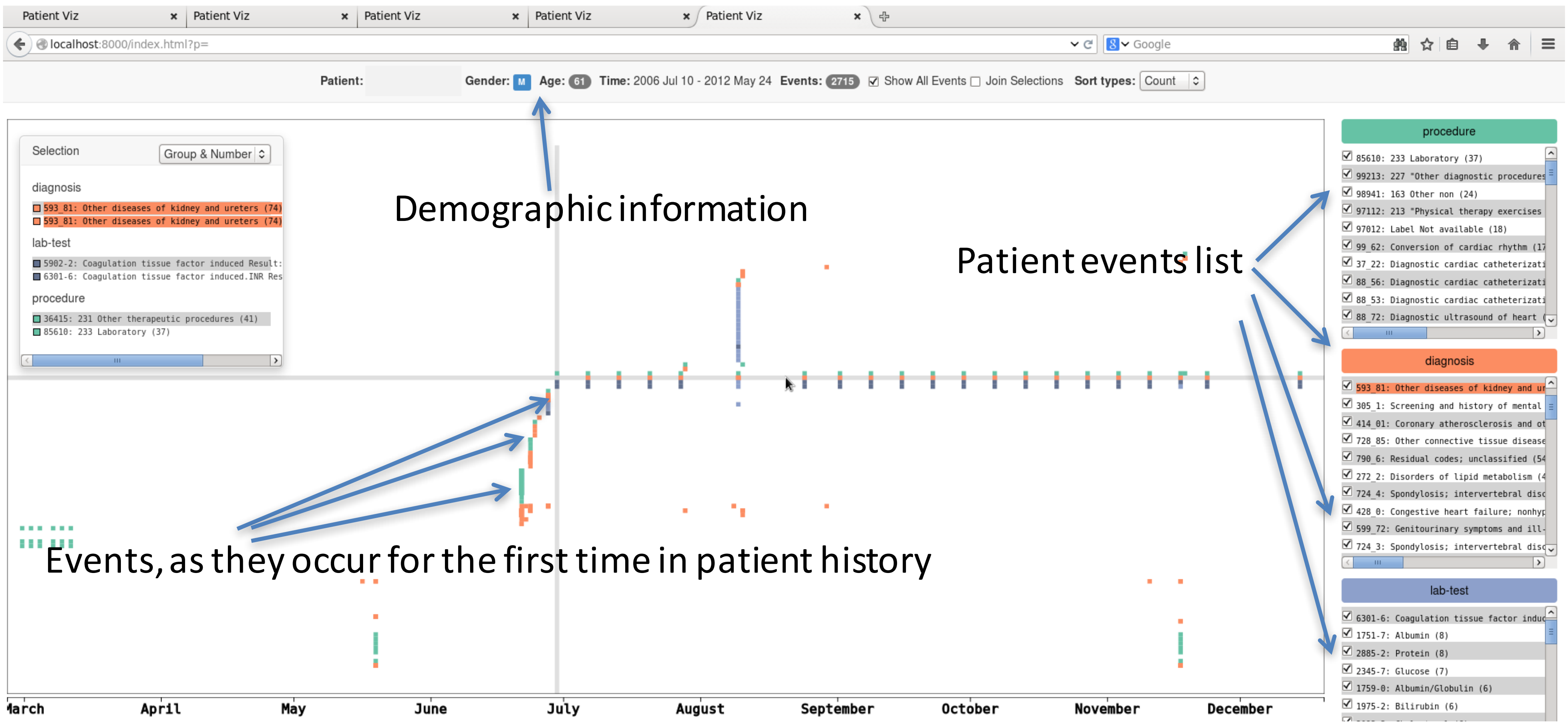


Typical pipeline:

1. Manually label several patients' data by "chart review"
2. A) Come up with a simple rule to automatically derive label for all patients, **or**

   B) Use machine learning to get the labels themselves

Visualization of individual patient data is an important part of chart review



https://github.com/nyuvis/patient-viz

https://www.phekb.org/phenotypes?field_pgx_type_tid_1=398&field_data_model_value=All

PheKB

a knowledgebase for discovering phenotypes from electronic medical records

Login | Request Account

Search

Home | Phenotypes | Resources | Contact Us

Public Phenotypes

Public | Collaboration

Public phenotypes are believed to be complete and final by their authors. When you are logged in you can view and edit phenotypes in your groups that are non public and in various stages of development.

Login To View Private Group Phenotypes

Institution

Type of Phenotype
Disease or Syndrome

Owner Phenotyping Groups

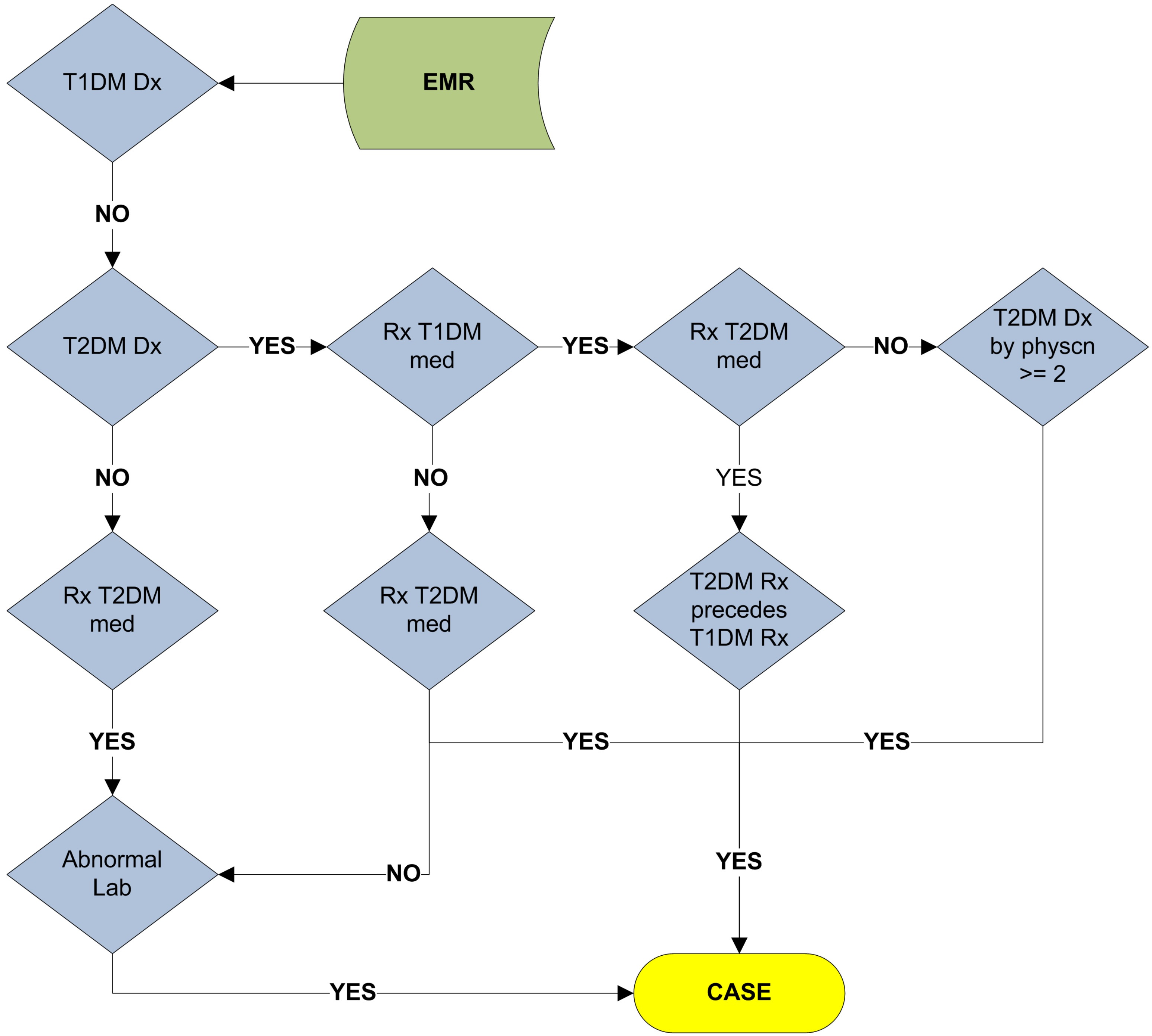
View Phenotyping Groups

Data Model
- Any -

Apply

Title	Institution	Data Modalities and Methods Used	Owner Phenotyping Groups	View Groups	Has new content	Status	Type
Abdominal Aortic Aneurysm (AAA)	Geisinger	CPT Codes, ICD 9 Codes, Vital Signs	eMERGE Geisinger Group	eMERGE Geisinger Group, eMERGE Phenotype WG		Final	Disease or Syndrome
ADHD phenotype algorithm	CHOP	ICD 9 Codes, Medications, Natural Language Processing	eMERGE CHOP Group	eMERGE Phenotype WG		Final	Disease or Syndrome
Appendicitis	Cincinnati Children's Hospital Medical Center	CPT Codes, ICD 9 Codes, Medications, Natural Language Processing	eMERGE CCHMC/BCH Group	eMERGE Phenotype WG		Final	Disease or Syndrome
Atrial Fibrillation - Demonstration Project	Vanderbilt University	CPT Codes, ICD 9 Codes, Natural Language Processing	Vanderbilt - SD/RD Group	Vanderbilt - SD/RD Group		Final	Disease or Syndrome
Autism	Cincinnati Children's Hospital Medical Center	ICD 9 Codes, Medications, Natural Language Processing	eMERGE CCHMC/BCH Group	eMERGE Phenotype WG		Final	Disease or Syndrome
Cataracts	Marshfield Clinic Research Foundation	CPT Codes, ICD 9 Codes, Medications, Natural Language Processing	eMERGE Marshfield Group	eMERGE Phenotype WG		Final	Disease or Syndrome
Crohn's Disease -	[illegible]	ICD 9 Codes, Medications,	Vanderbilt -	Vanderbilt -		[illegible]	Disease

Figure 1: Algorithm for identifying T2DM cases in the EMR.



Source: https://phekb.org/sites/phenotype/files/T2DM-algorithm.pdf

Outline for today’s class

1. Risk stratification
2. Case study: Early detection of Type 2 diabetes
   – Framing as supervised learning problem
   – Deriving labels
   – **Evaluating risk stratification algorithms**
3. Subtleties with ML-based risk stratification
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

What are the Discovered Risk Factors?

- 769 variables have non-zero weight

Top History of Disease	Odds Ratio
Impaired Fasting Glucose (Code 790.21)	4.17 (3.87 4.49)
Abnormal Glucose NEC (790.29)	4.07 (3.76 4.41)
Hypertension (401)	3.28 (3.17 3.39)
Obstructive Sleep Apnea (327.23)	2.98 (2.78 3.20)
Obesity (278)	2.88 (2.75 3.02)
Abnormal Blood Chemistry (790.6)	2.49 (2.36 2.62)
Hyperlipidemia (272.4)	2.45 (2.37 2.53)
Shortness Of Breath (786.05)	2.09 (1.99 2.19)
Esophageal Reflux (530.81)	1.85 (1.78 1.93)

Diabetes
1-year gap

What are the Discovered Risk Factors?

- 769 variables have non-zero weight

Top History of Diseas	
Impaired Fasting Glucose (Code	
Abnormal Glucose NEC (790.29)	
Hypertension (401)	
Obstructive Sleep Apnea (327.23)	
Obesity (278)	
Abnormal Blood Chemistry (790.6	
Hyperlipidemia (272.4)	
Shortness Of Breath (786.05)	(1.99 2.19)
Esophageal Reflux (530.81)	1.85 (1.78 1.93)

Additional Disease Risk Factors Include:
Pituitary dwarfism (253.3), Hepatomegaly(789.1), Chronic Hepatitis C (070.54), Hepatitis (573.3), Calcaneal Spur(726.73), Thyrotoxicosis without mention of goiter(242.90), Sinoatrial Node dysfunction(427.81), Acute frontal sinusitis (461.1), Hypertrophic and atrophic conditions of skin(701.9), Irregular menstruation(626.4), …

Diabetes 1-year gap

What are the Discovered Risk Factors?

- 769 variables have non-zero weight

Top Lab Factors	Odds Ratio
Hemoglobin A1c /Hemoglobin.Total (High - past 2 years)	5.75 (5.42 6.10)
Glucose (High- Past 6 months)	4.05 (3.89 4.21)
Cholesterol.In VLDL (Increasing - Past 2 years)	3.88 (3.53 4.27)
Potassium (Low - Entire History)	2.58 (2.24 2.98)
Cholesterol.Total/Cholesterol.In HDL (High - Entire History)	2.29 (2.19 2.40)
Erythrocyte mean corpuscular hemoglobin concentration -(Low - Entire History)	2.25 (1.92 2.64)
Eosinophils (High - Entire History)	2.11 (1.82 2.44)
Glomerular filtration rate/1.73 sq M.Predicted (Low -Entire History)	2.07 (1.92 2.24)
Alanine aminotransferase (High Entire History)	2.04 (1.89 2.19)

Diabetes
1-year gap

What are the Discovered Risk Factors?

- 769 variables have non-zero weight

Top Lab Factors

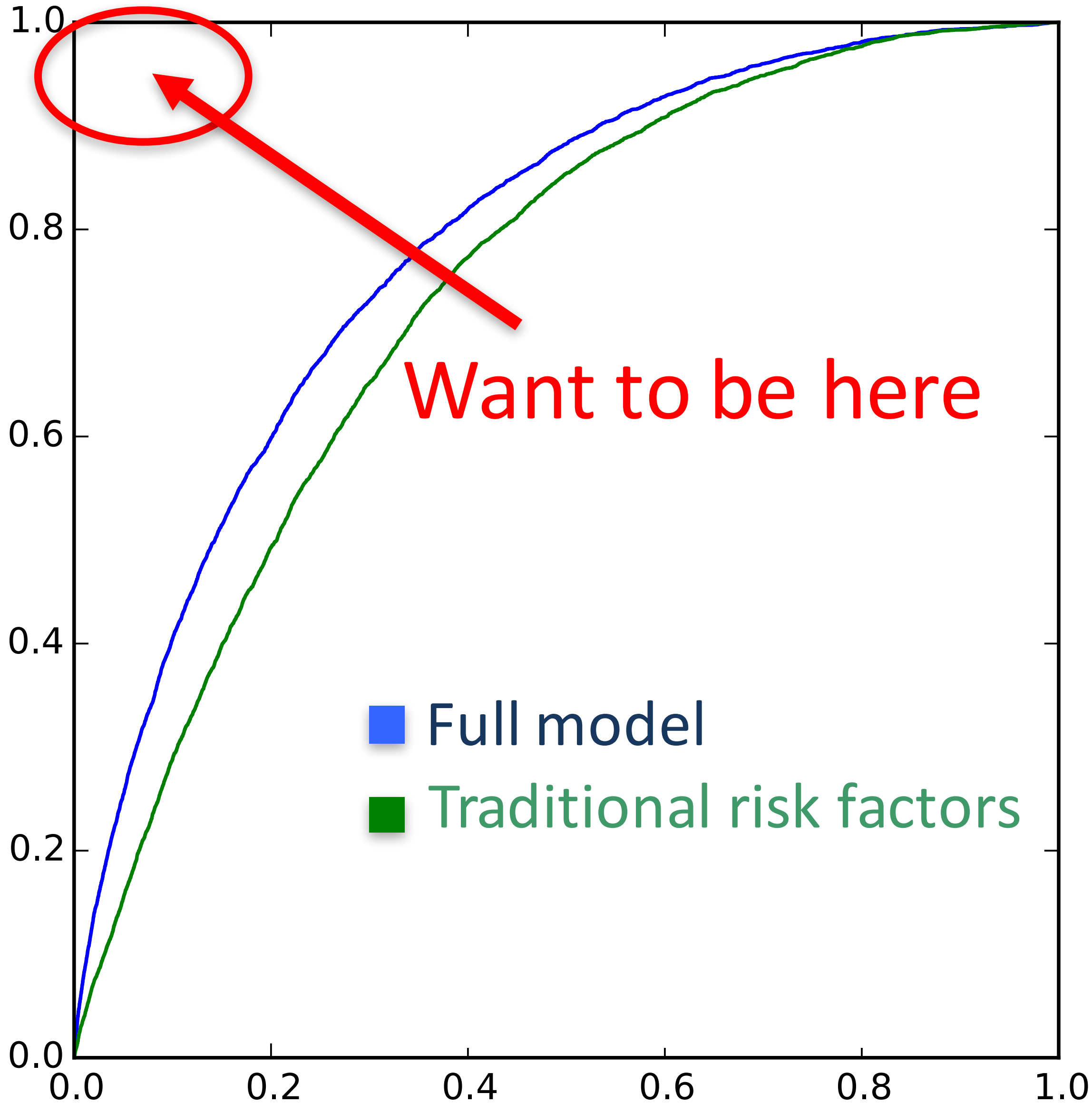
Hemoglobin A1c /Hemoglobin.Total (High	
Glucose (High- Past 6 months)	
Cholesterol.In VLDL (Increasing - Past 2 y	
Potassium (Low - Entire History)	
Cholesterol.Total/Cholesterol.In HDL (Hig	
Erythrocyte mean corpuscular hemoglobin concentration -(Low - Entire History)	2.25 (1.92 2.64)
Eosinophils (High - Entire History)	2.11 (1.82 2.44)
Glomerular filtration rate/1.73 sq M.Predicted (Low -Entire History)	2.07 (1.92 2.24)
Alanine aminotransferase (High Entire History)	2.04 (1.89 2.19)

Additional Lab Test Risk Factors Include:
Albumin/Globulin (Increasing -Entire history), Urea nitrogen/Creatinine -(high - Entire History), Specific gravity (Increasing, Past 2 years), Bilirubin (high -Past 2 years),...

Diabetes
1-year gap

Receiver-operator characteristic curve

True positive rate



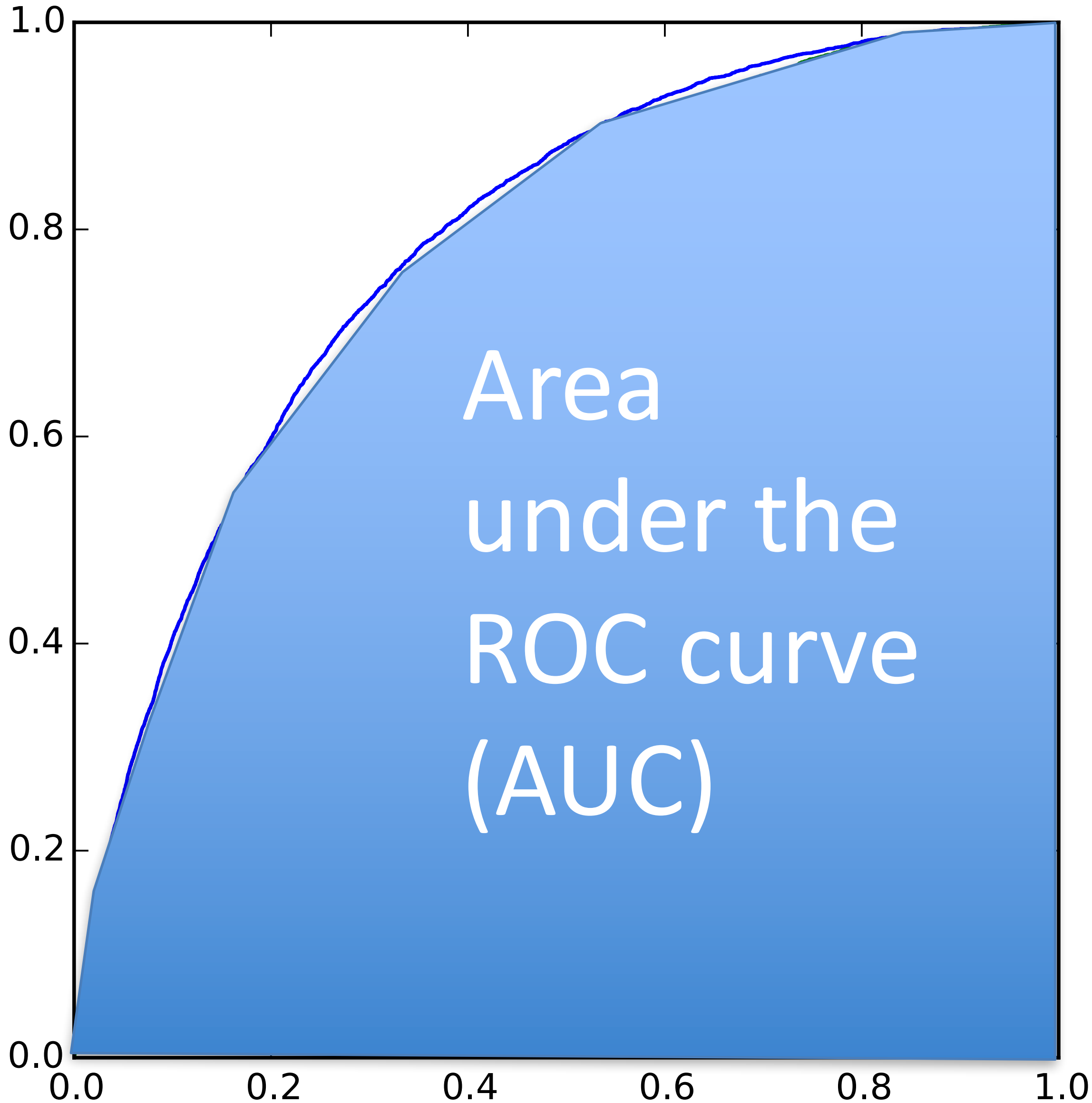
False positive rate

Obtained by varying prediction threshold

Diabetes
1-year gap

Receiver-operator characteristic curve

True positive rate



False positive rate

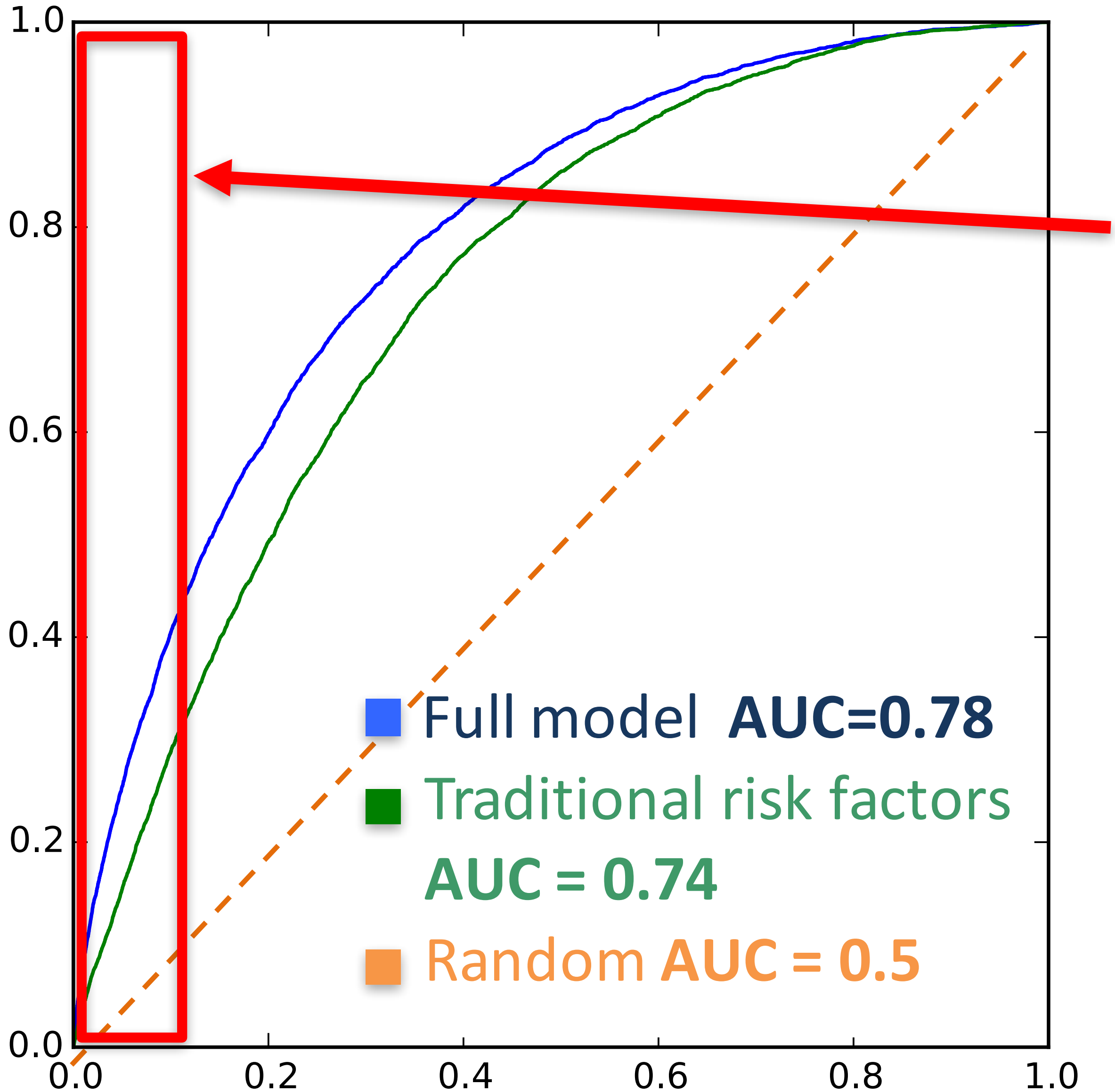
Diabetes 1-year gap

AUC = Probability that algorithm ranks a positive patient over a negative patient

Invariant to amount of class imbalance

Receiver-operator characteristic curve

True positive rate



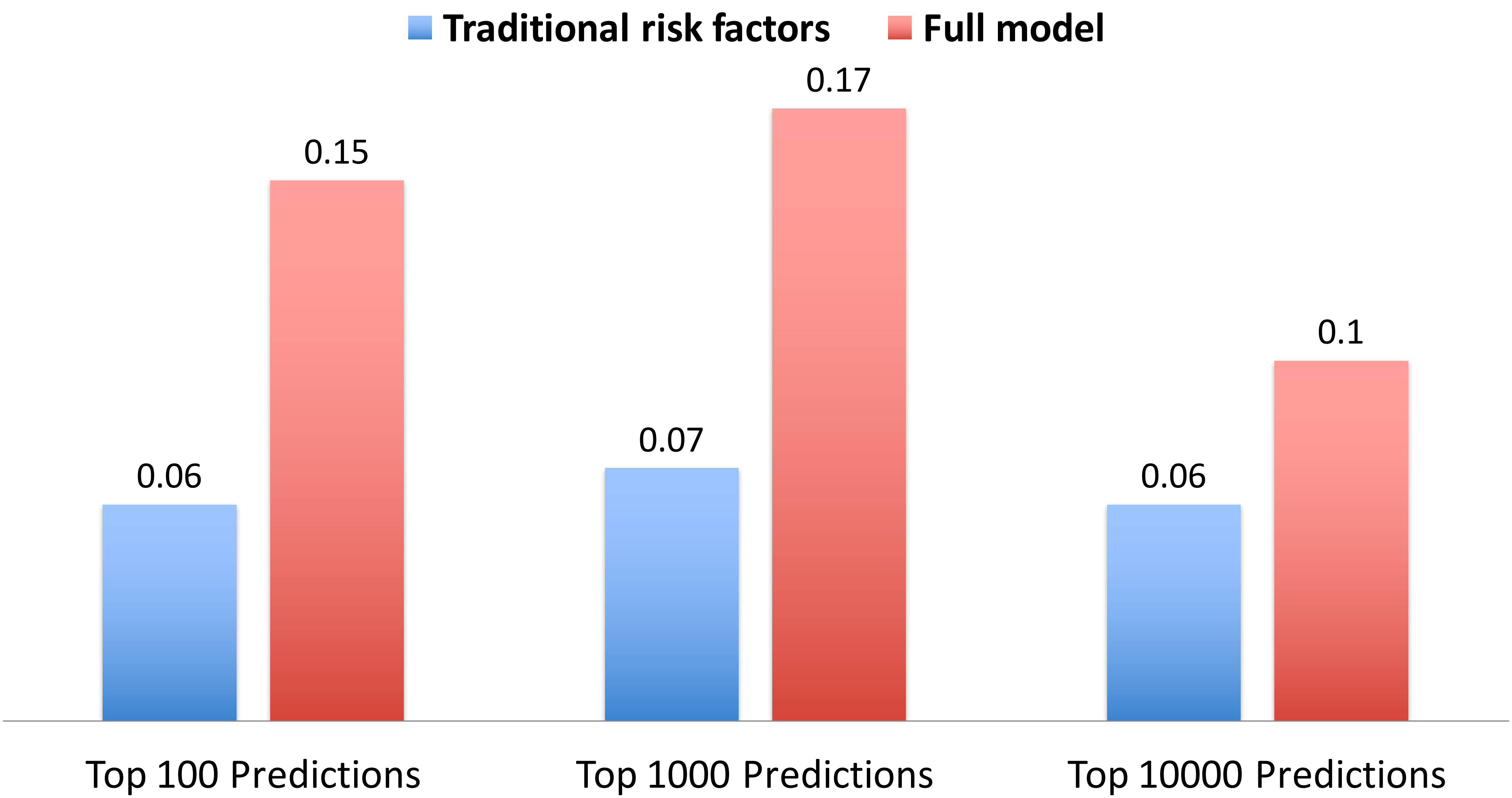
False positive rate

Risk stratification usually focuses on just this region

(because of the cost of interventions)

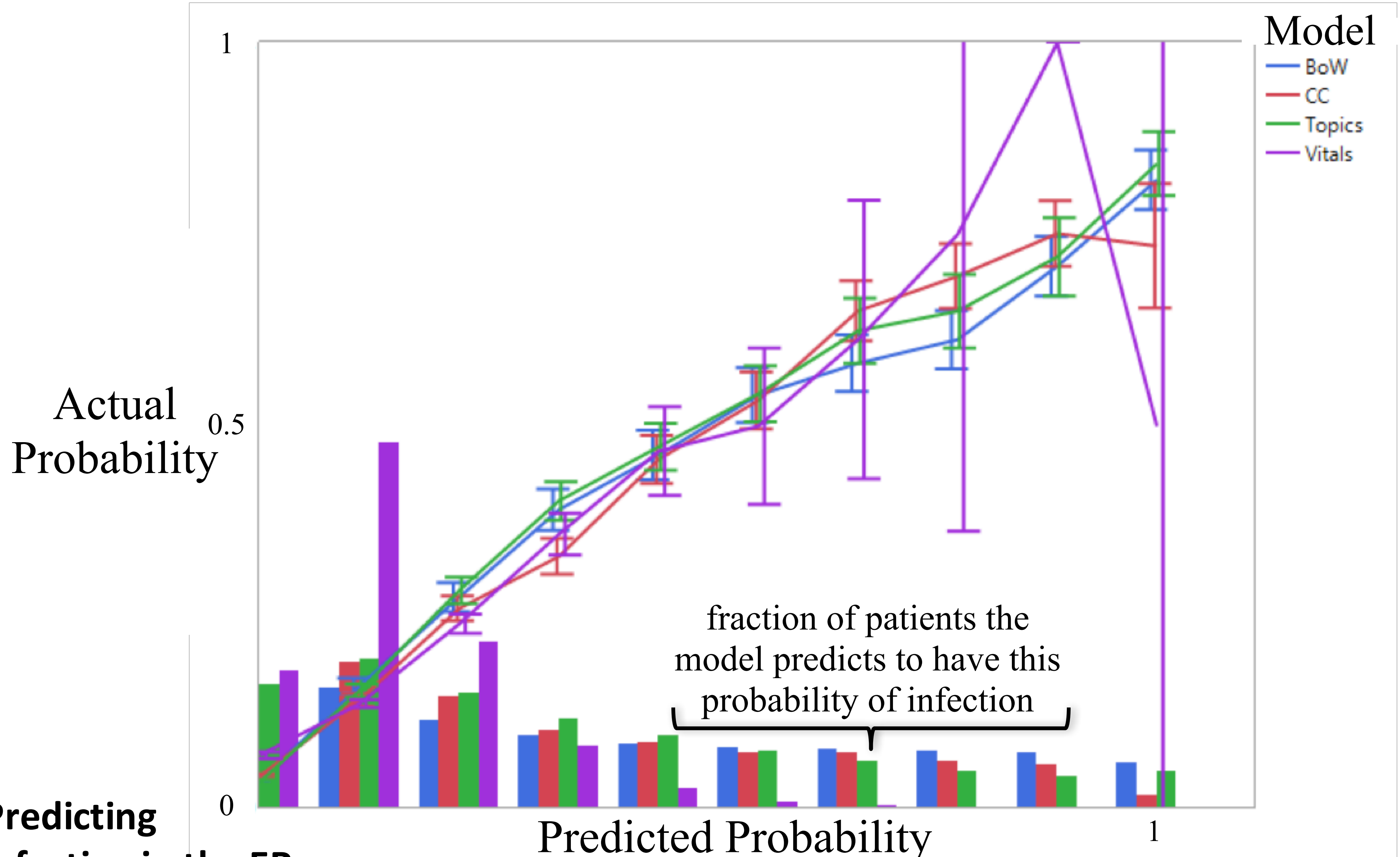
Diabetes 1-year gap

Positive predictive value (PPV)



Diabetes 1-year gap

Calibration (*note: different dataset*)



Predicting infection in the ER

Outline for today’s class

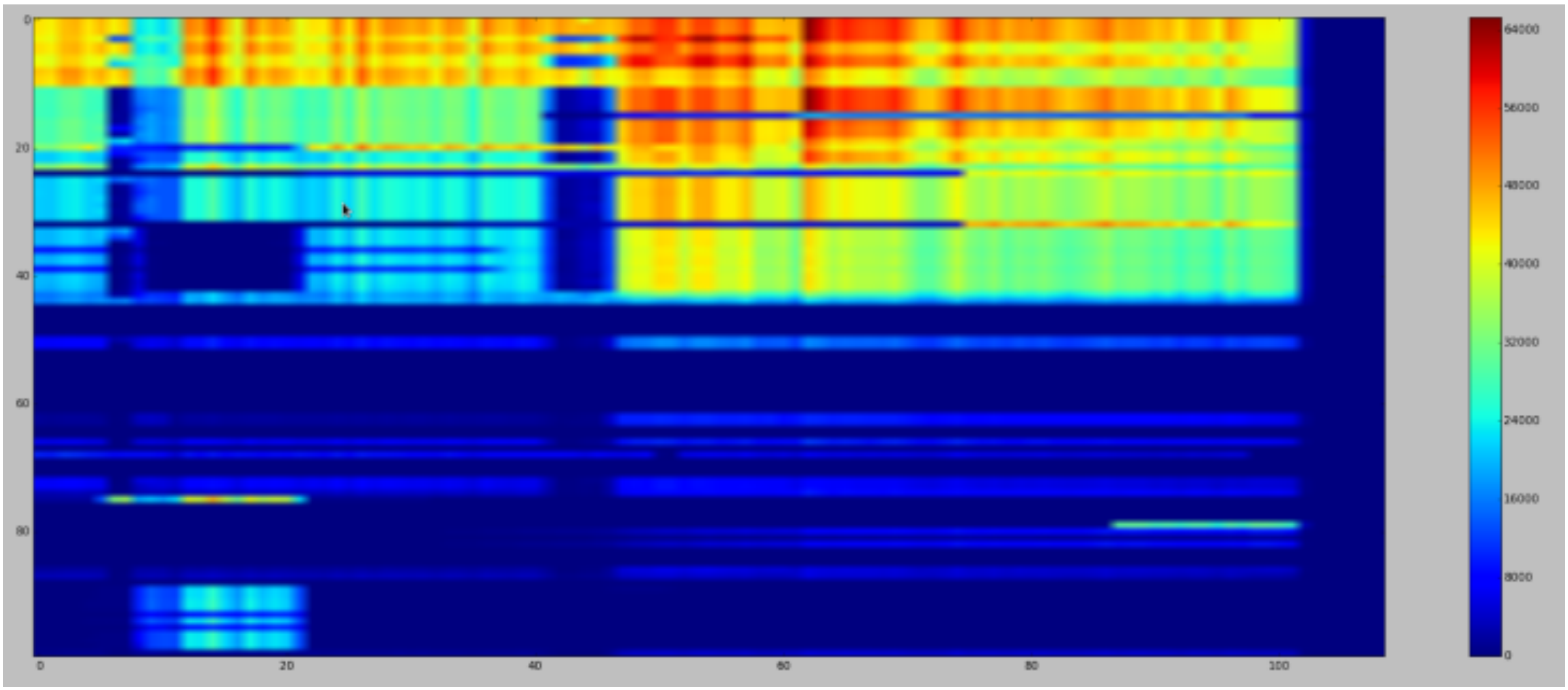
1. Risk stratification
2. Case study: Early detection of Type 2 diabetes
   - Framing as supervised learning problem
   - Deriving labels
   - Evaluating risk stratification algorithms
3. **Subtleties with ML-based risk stratification**
4. Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)

Major challenge: non-stationarity

- ICD10 rolled out in 2015: predictive models learned using ICD9 features are no longer useful!
- Logistical issues => some features may not be available!
- Prevalence and significance of features may change over time
- Automatically derived labels may change meaning

Top 100 lab measurements over time

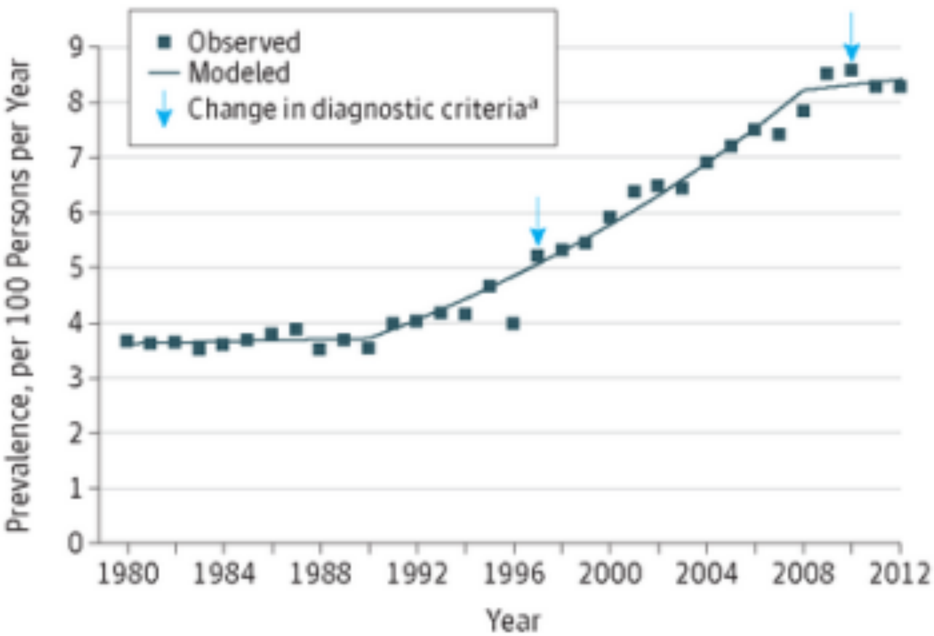
Labs

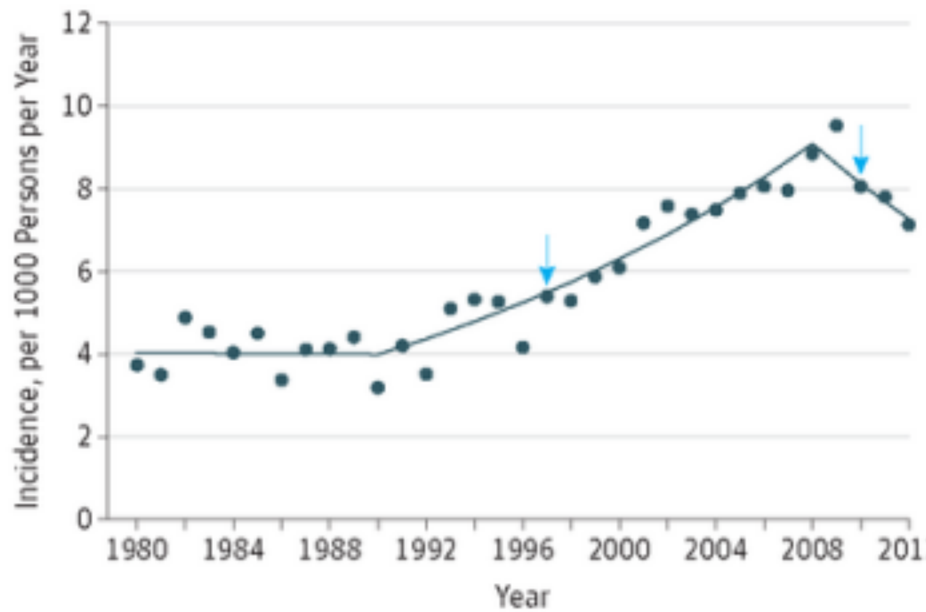


Time (in months, from 1/2005 up to 1/2014)

[Figure credit: Narges Razavian]

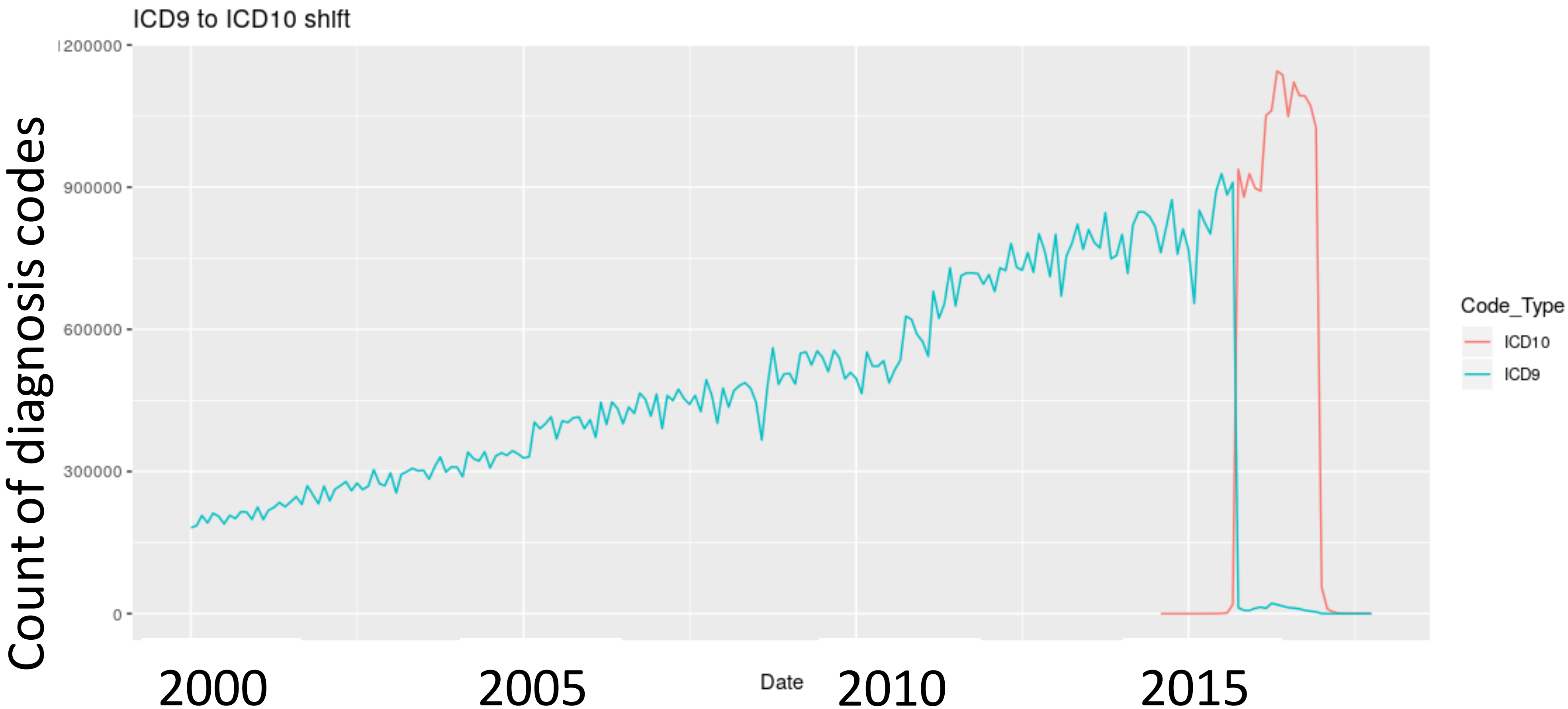
Diabetes Onset after 2009





Geiss LS, Wang J, Cheng YJ, et al. Prevalence and Incidence Trends for Diagnosed Diabetes Among Adults Aged 20 to 79 Years, United States, 1980-2012. *JAMA.* 2014;312(12):1218-1226.

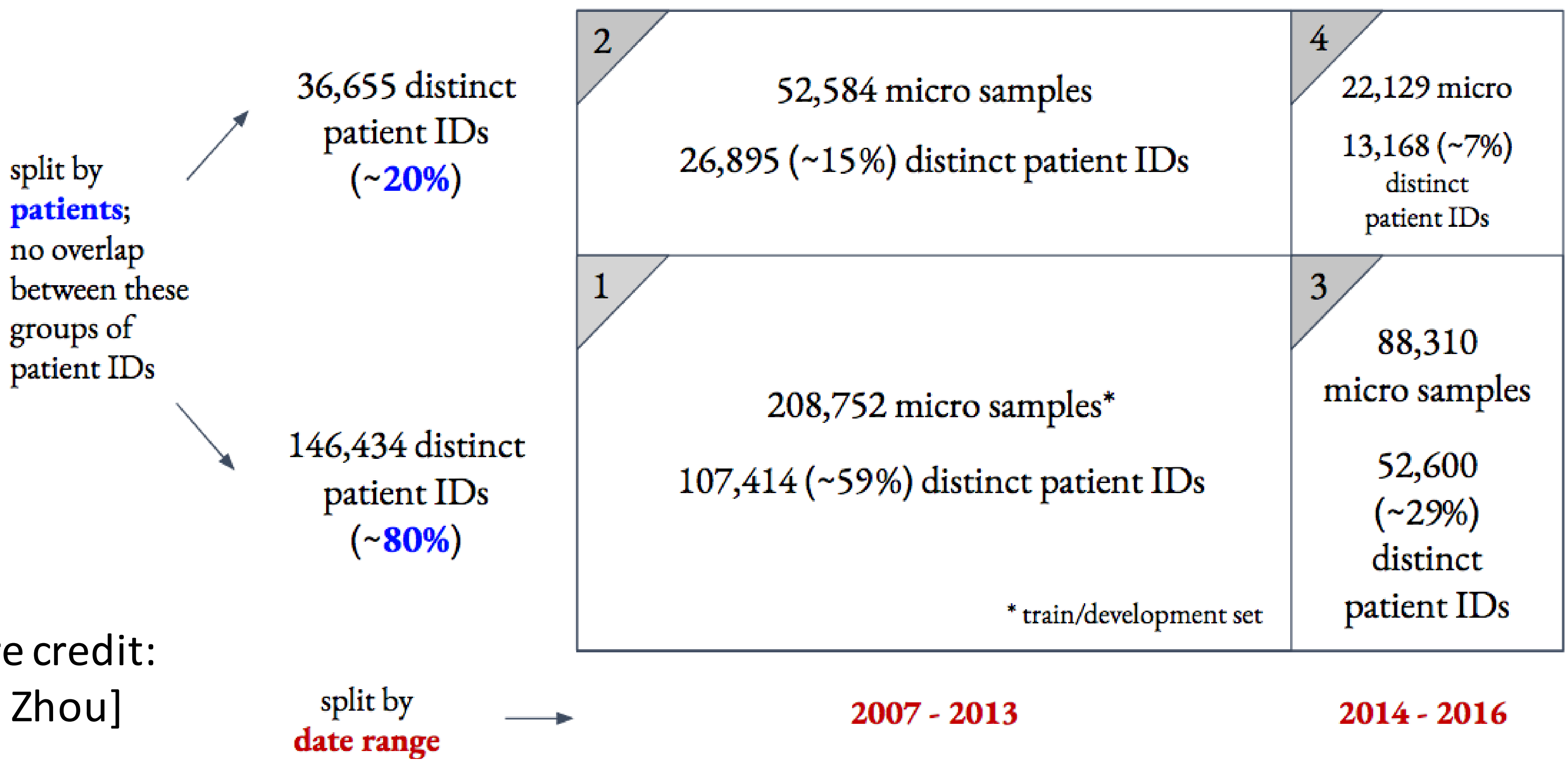
ICD-9 to ICD-10 shift



[Figure credit: Mike Oberst]

Re-thinking evaluation in the face of non-stationarity

- Given all this, do you notice anything wrong with how we evaluated the diabetes model?
- Good practice is to let the test data be from a future year:



[Figure credit: Helen Zhou]

Outline for today’s class

1. Risk stratification
2. Case study: Early detection of Type 2 diabetes
   - Framing as supervised learning problem
   - Deriving labels
   - Evaluating risk stratification algorithms
3. Subtleties with ML-based risk stratification
4. **Discussion with Leonard D'Avolio (Assistant Professor at HMS, CEO @ Cyft)**