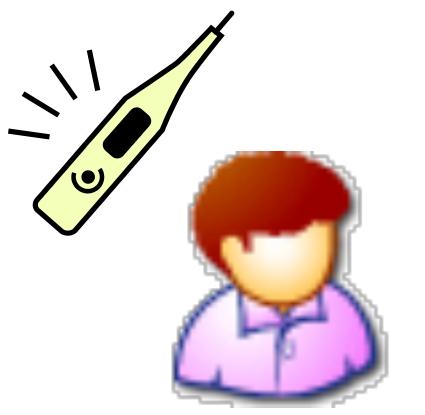


# Data-Driven Healthcare: Visual Analytics for Exploration and Prediction of Clinical Data

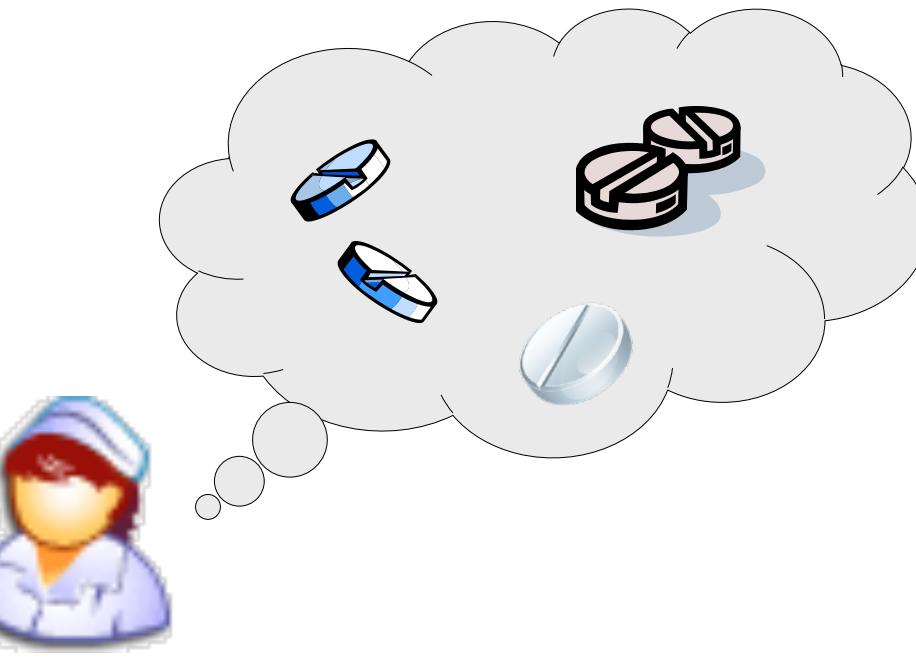
**Adam Perer**

IBM Research

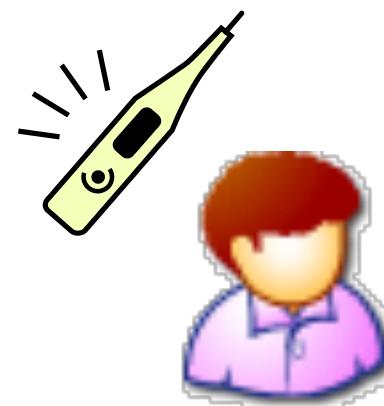




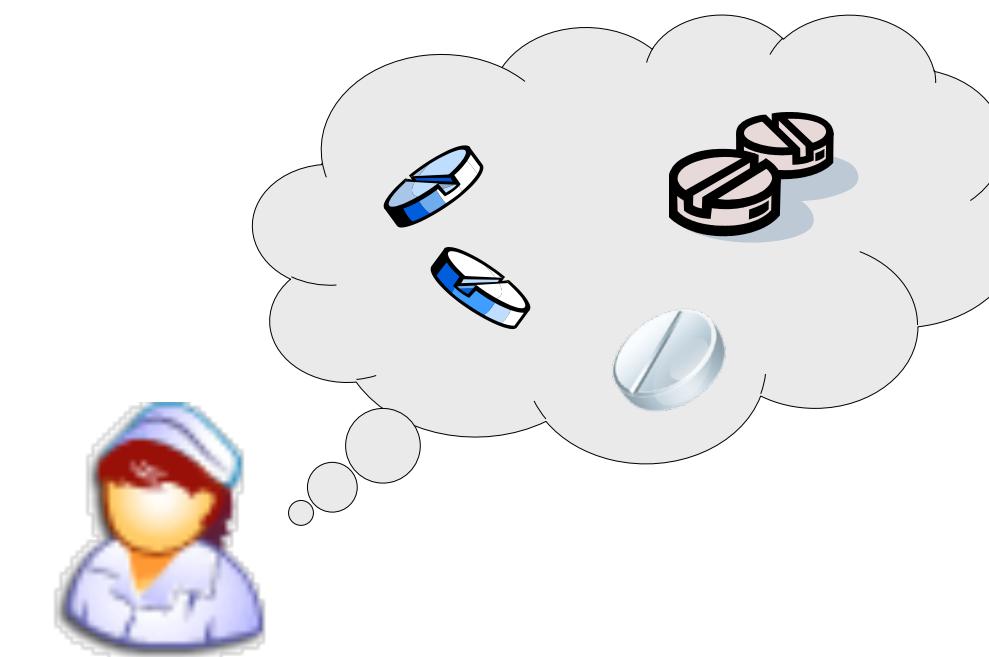
Patient



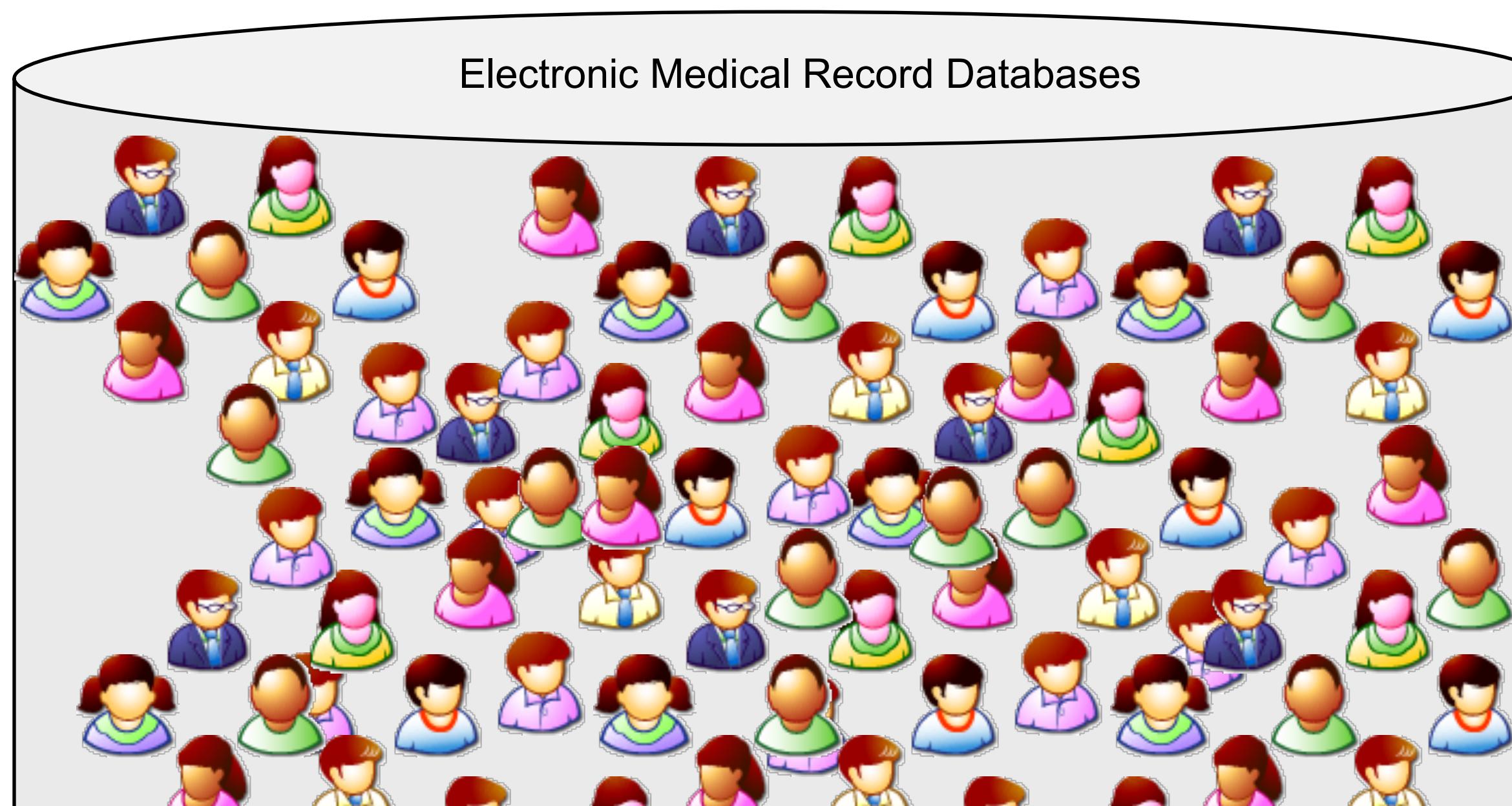
Clinician



Patient

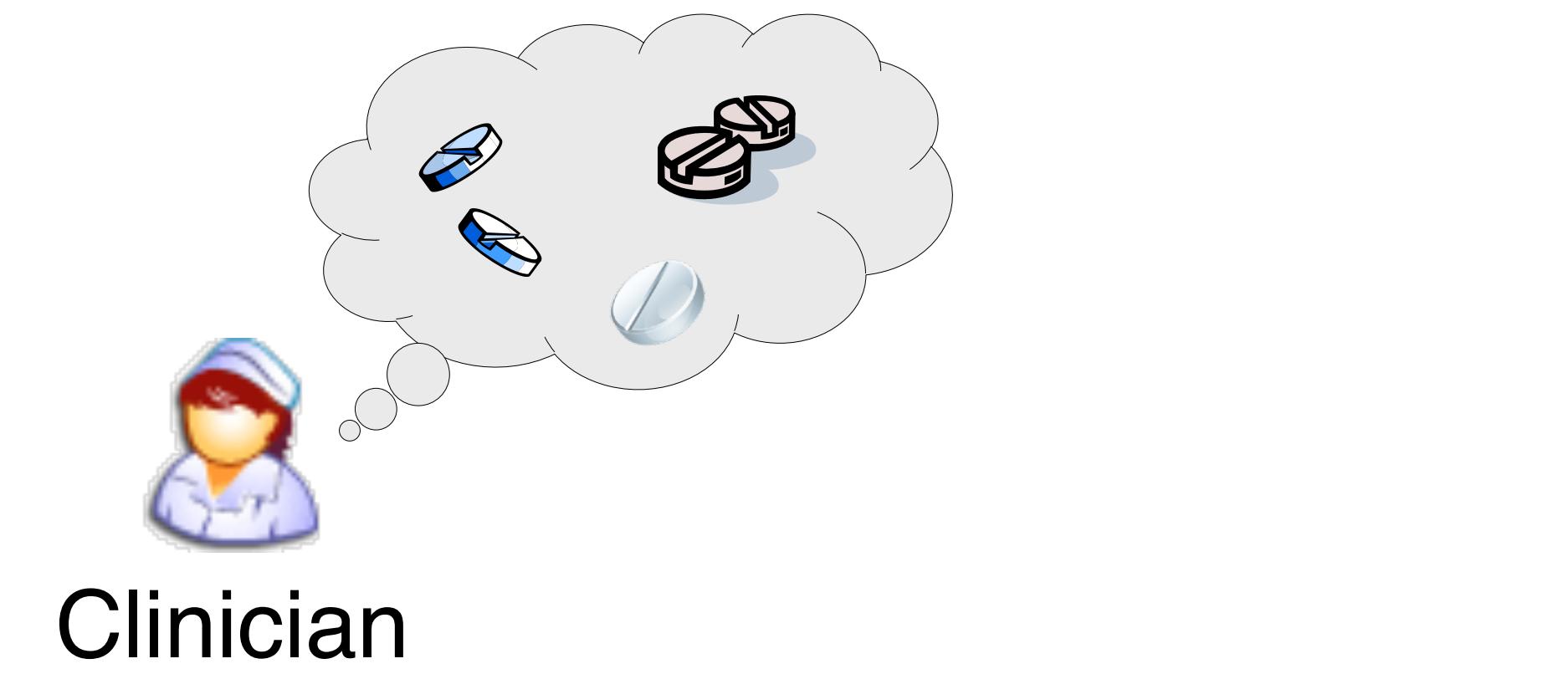
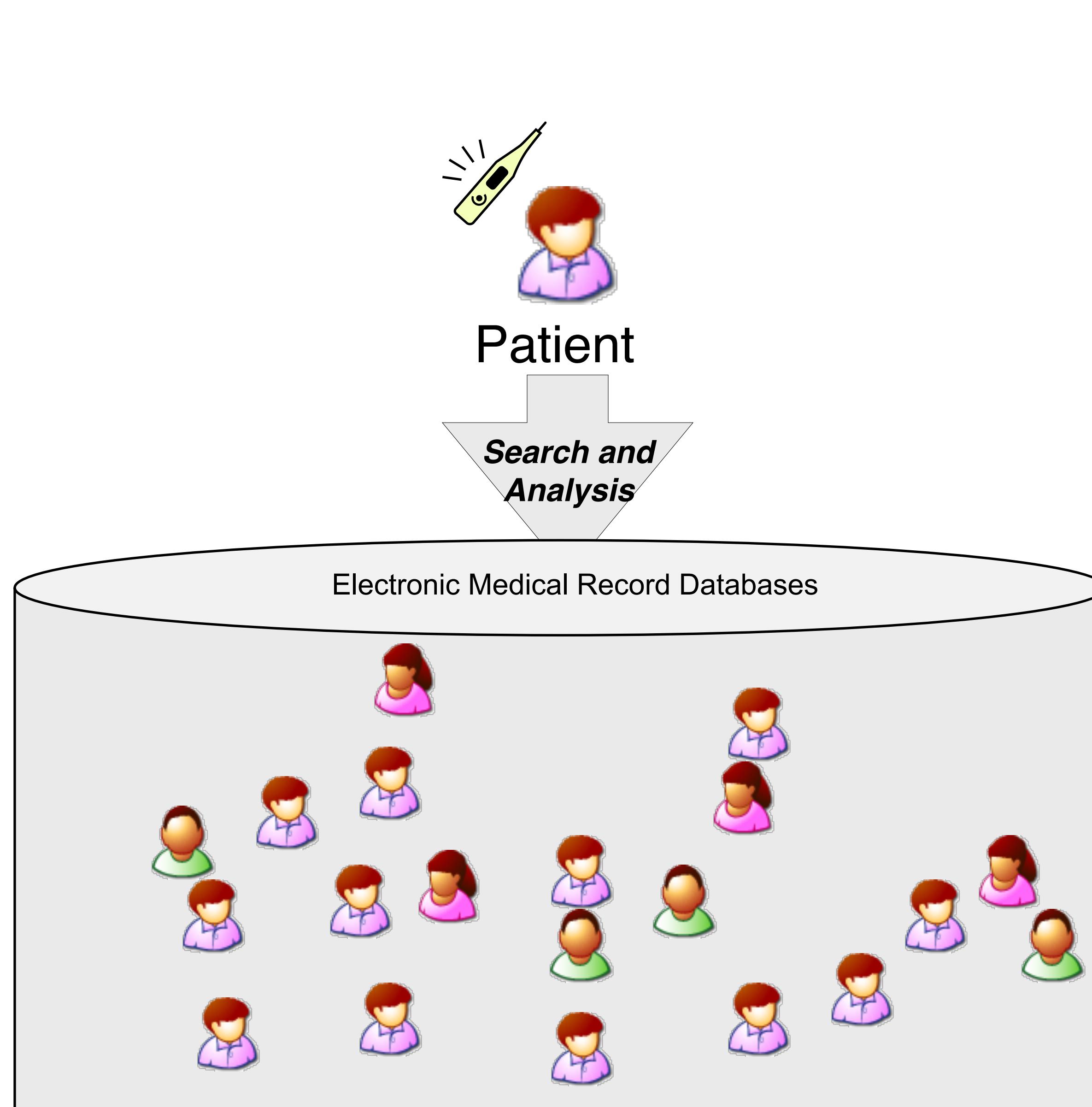


Clinician



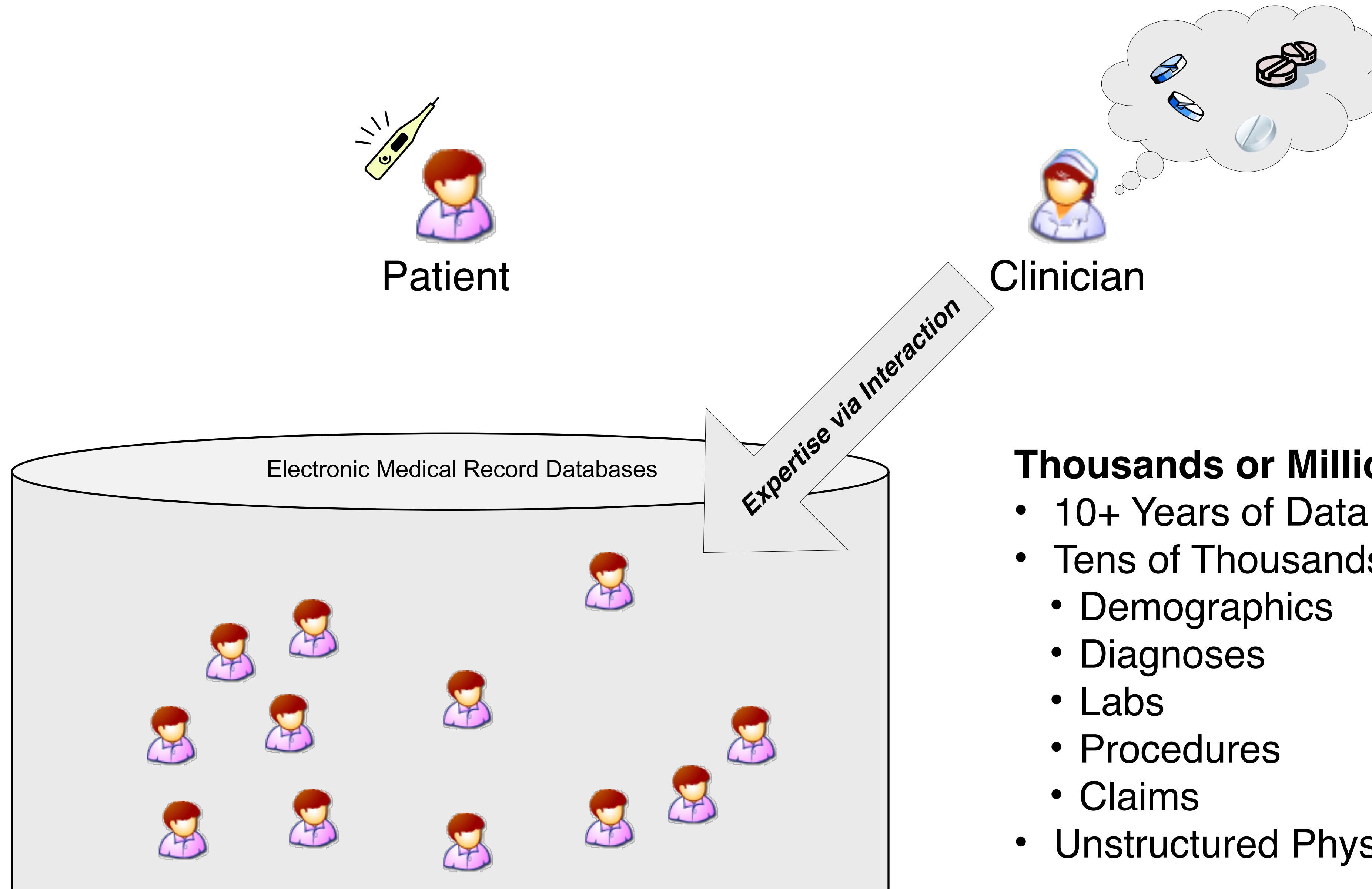
## Thousands or Millions of Patients

- 10+ Years of Data Per Patient
- Tens of Thousands of Features
  - Demographics
  - Diagnoses
  - Labs
  - Procedures
  - Claims
- Unstructured Physician Notes



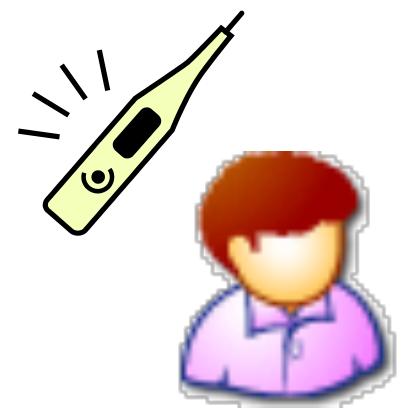
## Thousands or Millions of Patients

- 10+ Years of Data Per Patient
- Tens of Thousands of Features
  - Demographics
  - Diagnoses
  - Labs
  - Procedures
  - Claims
- Unstructured Physician Notes

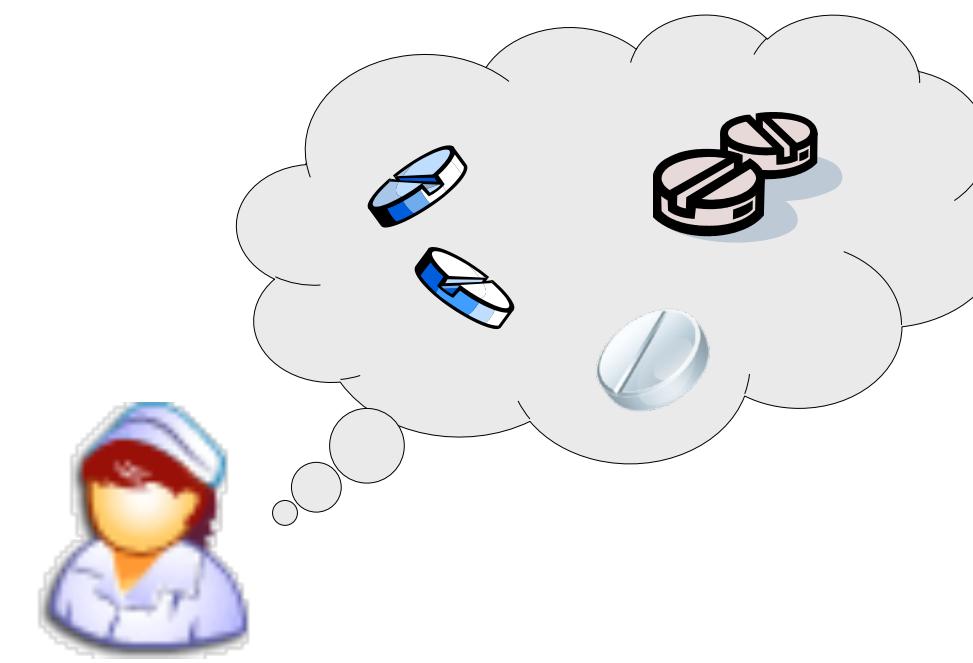


## **Thousands or Millions of Patients**

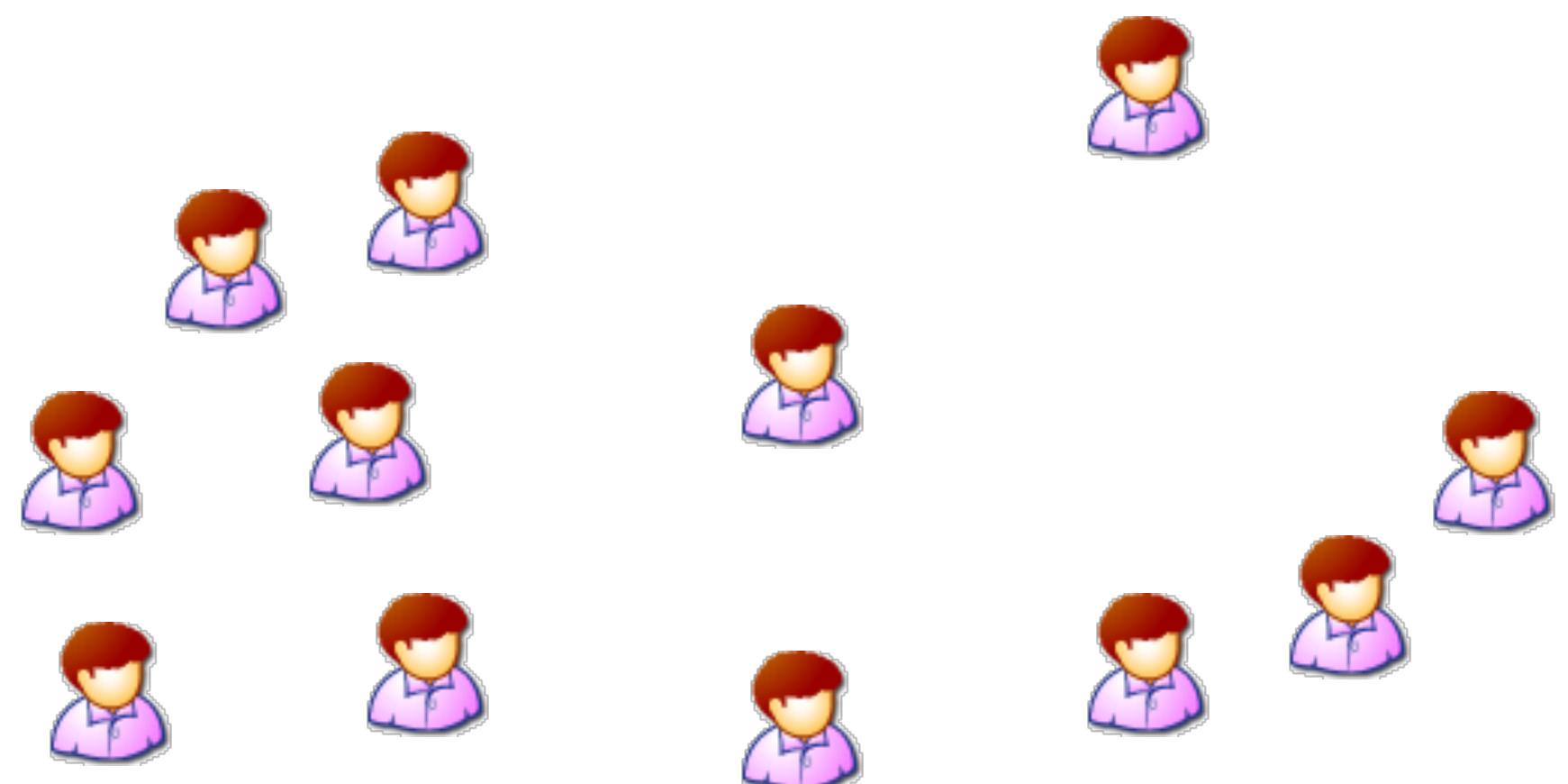
- 10+ Years of Data Per Patient
- Tens of Thousands of Features
  - Demographics
  - Diagnoses
  - Labs
  - Procedures
  - Claims
- Unstructured Physician Notes

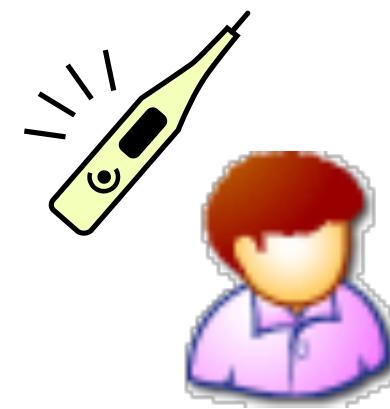


Patient

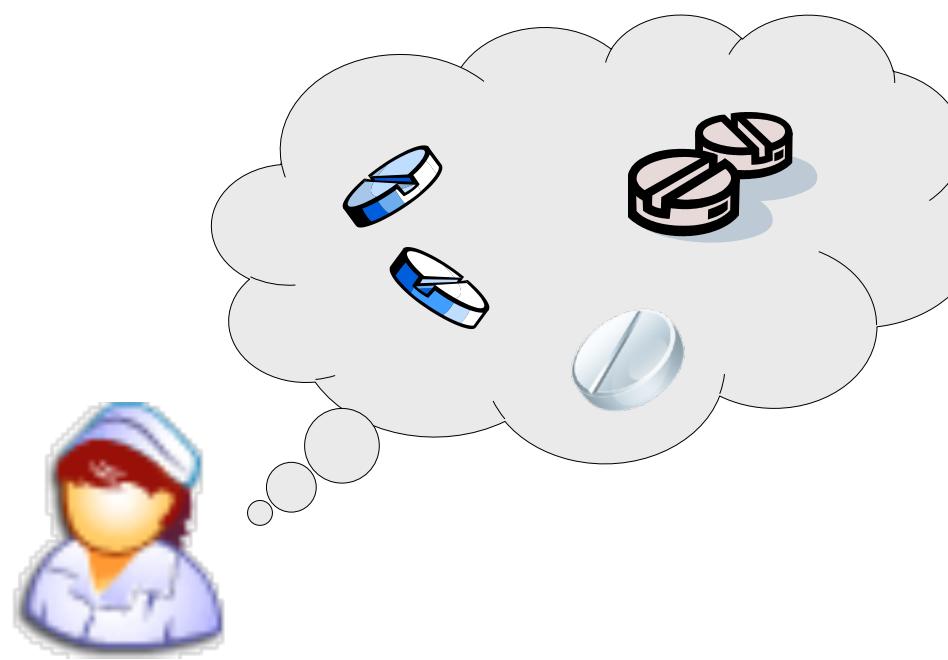


Clinician

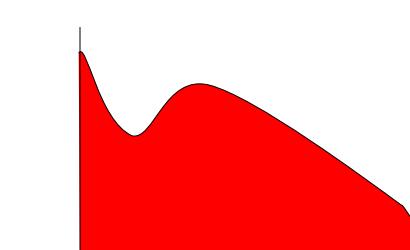
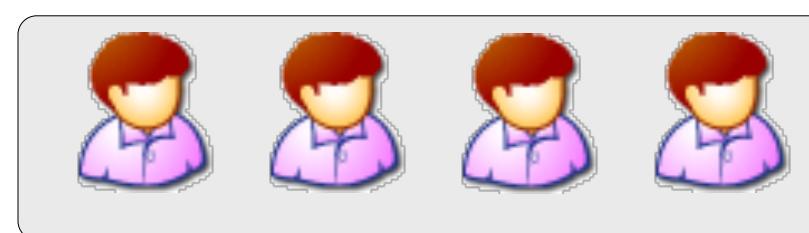
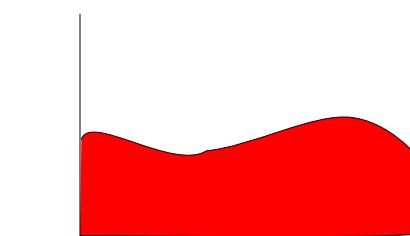
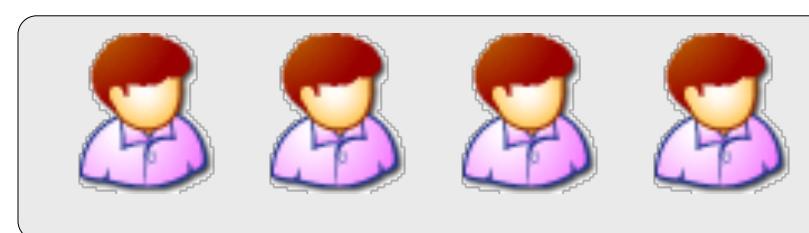
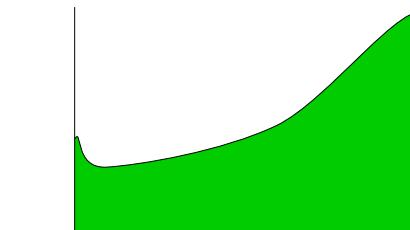
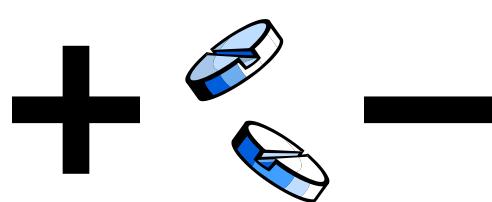
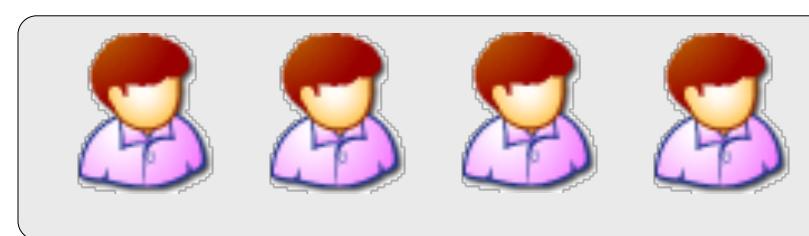


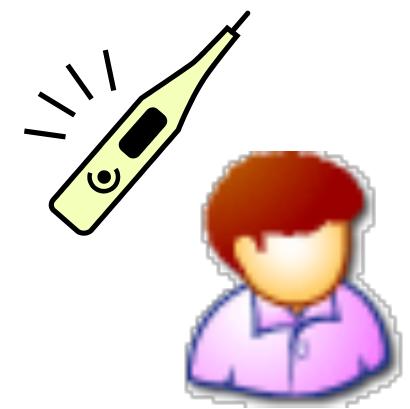


Patient

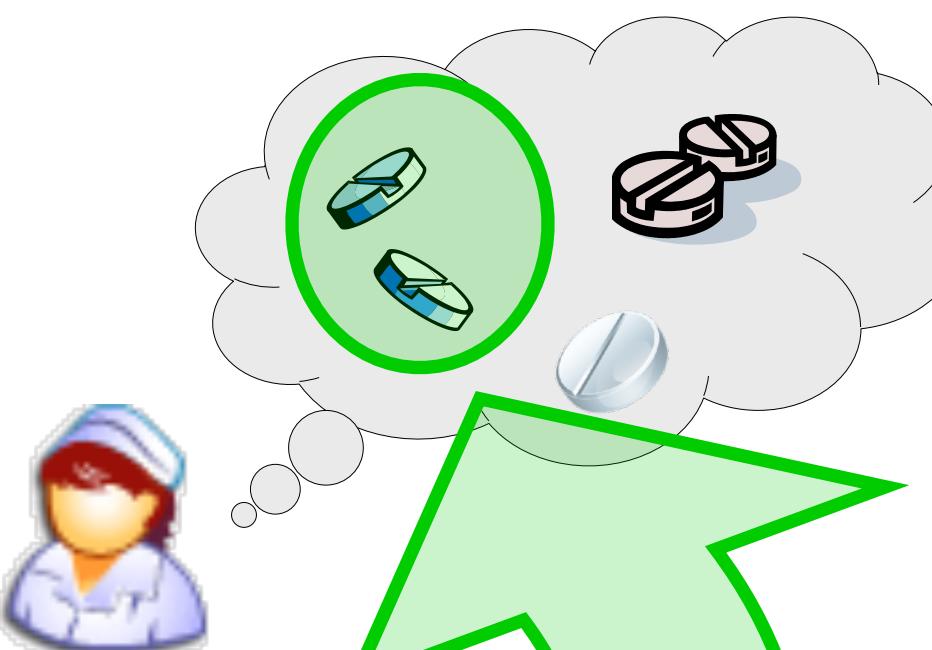


Clinician

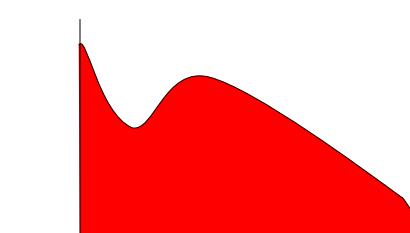
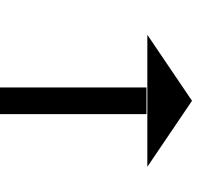
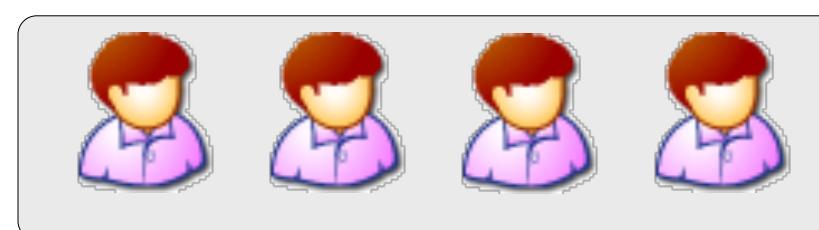
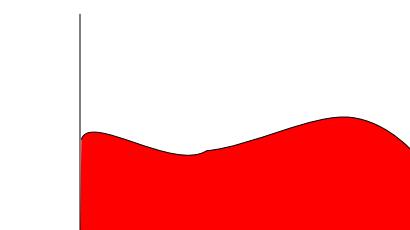
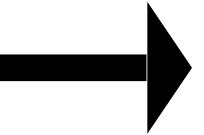
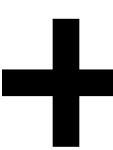
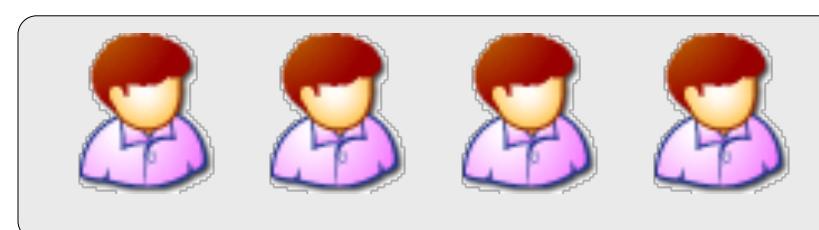
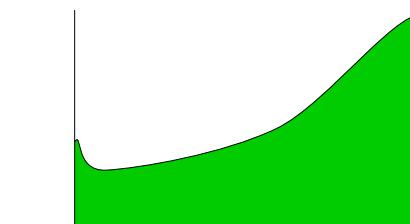
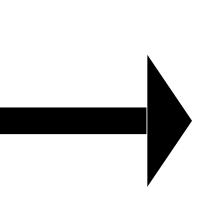
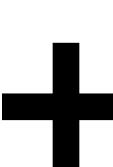
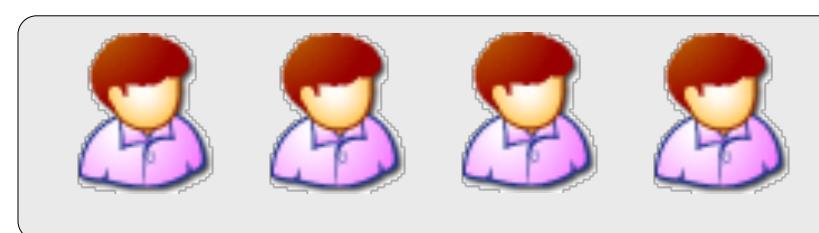




Patient



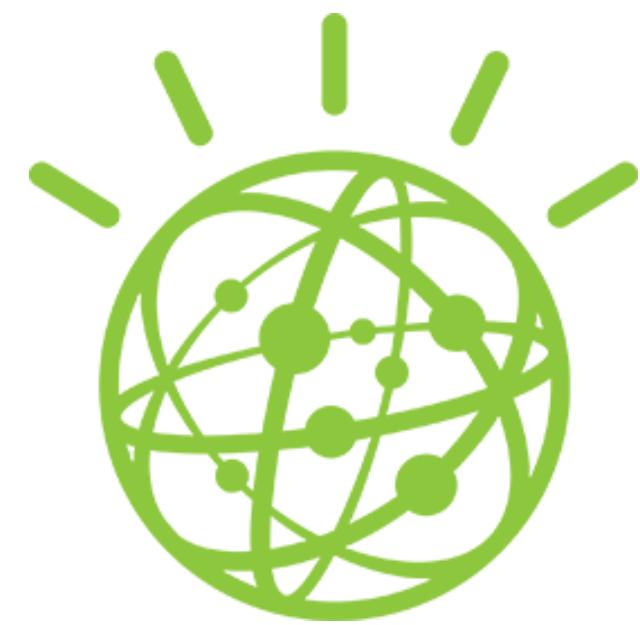
Clinician



# outline

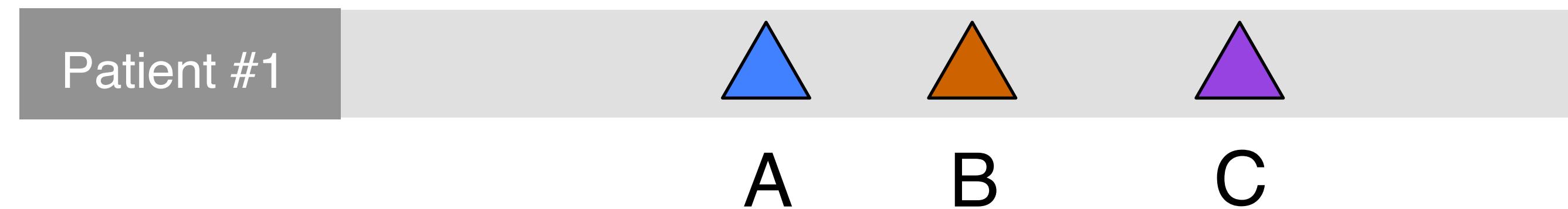
- Visual tools for exploring clinical data support unearthing insights from clinical records
  - **CareFlow**
- Beyond exploration, clinical researchers often want predictions, too.
  - **Coquito, Prospector**

# CareFlow

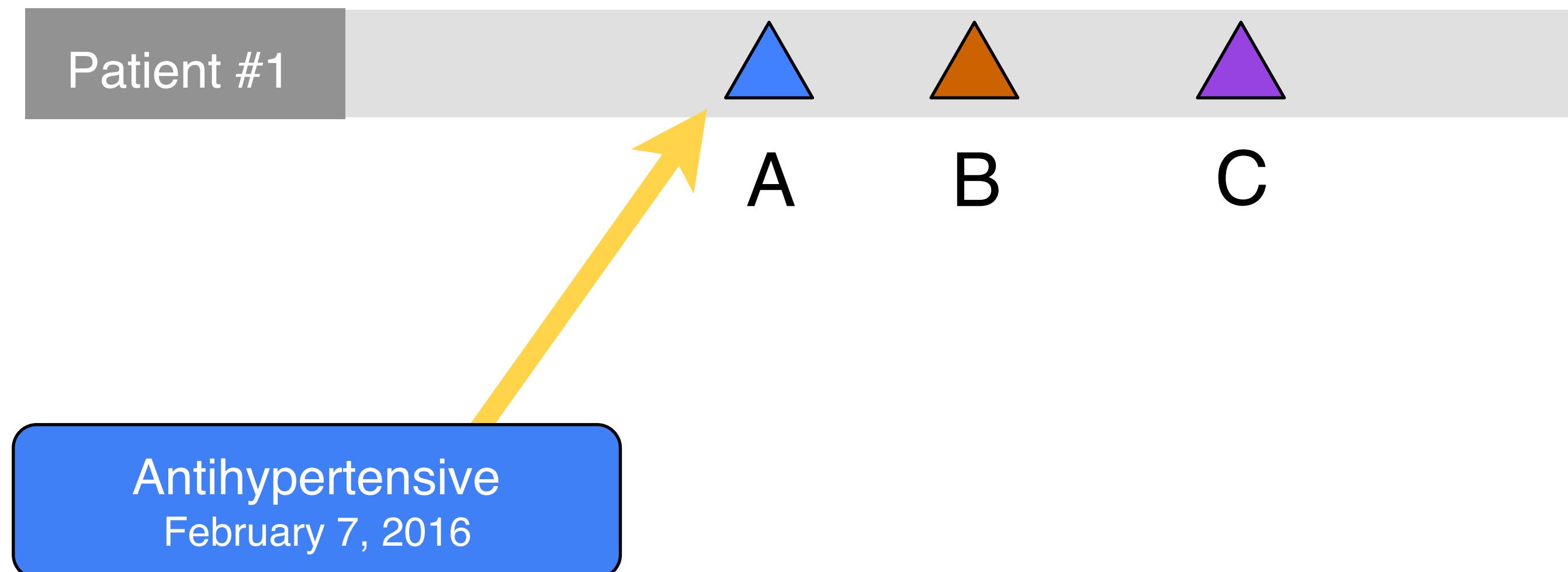


IBM WatsonHealth

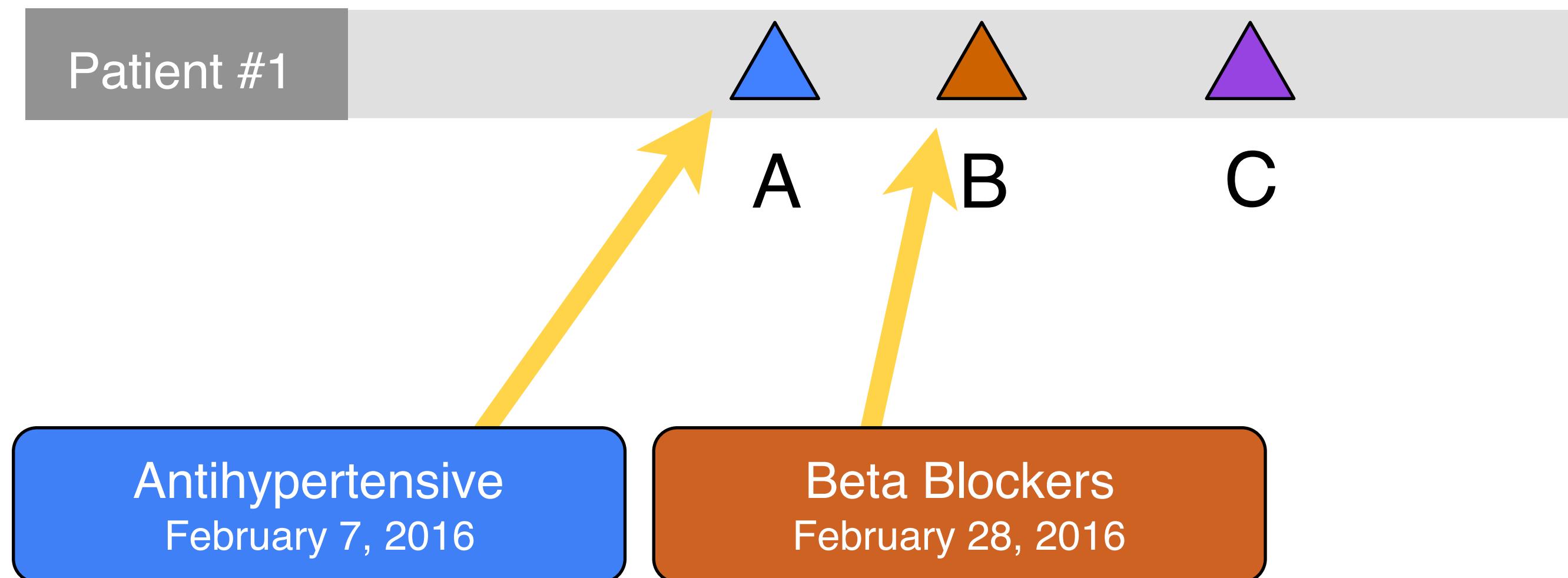
# electronic medical records



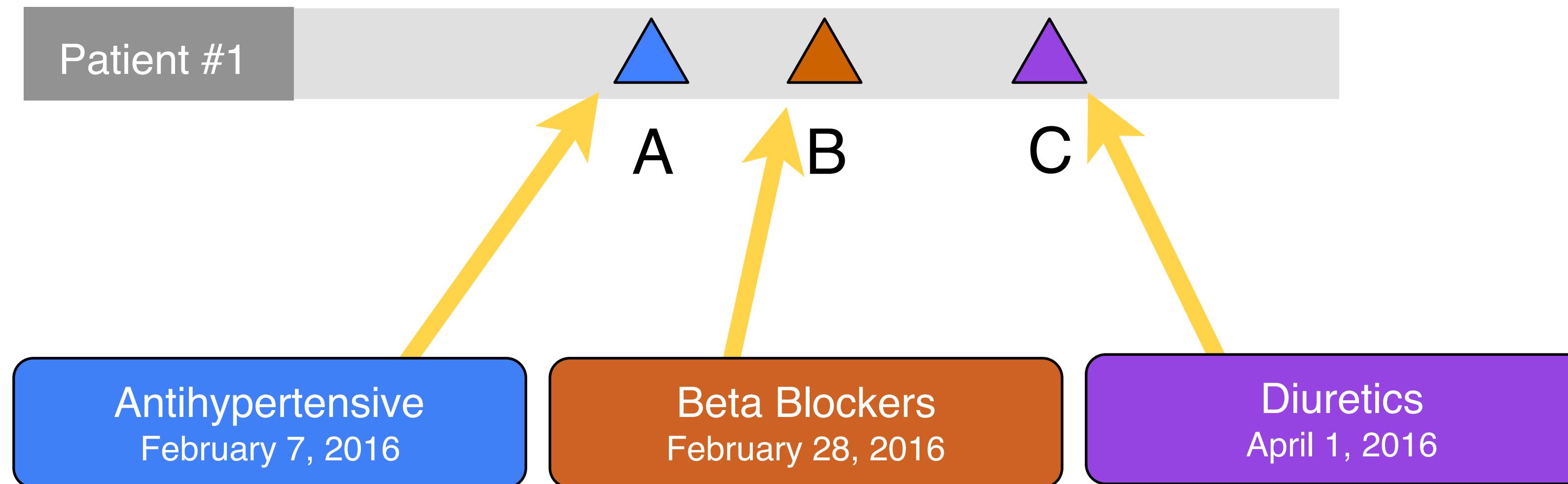
# electronic medical records



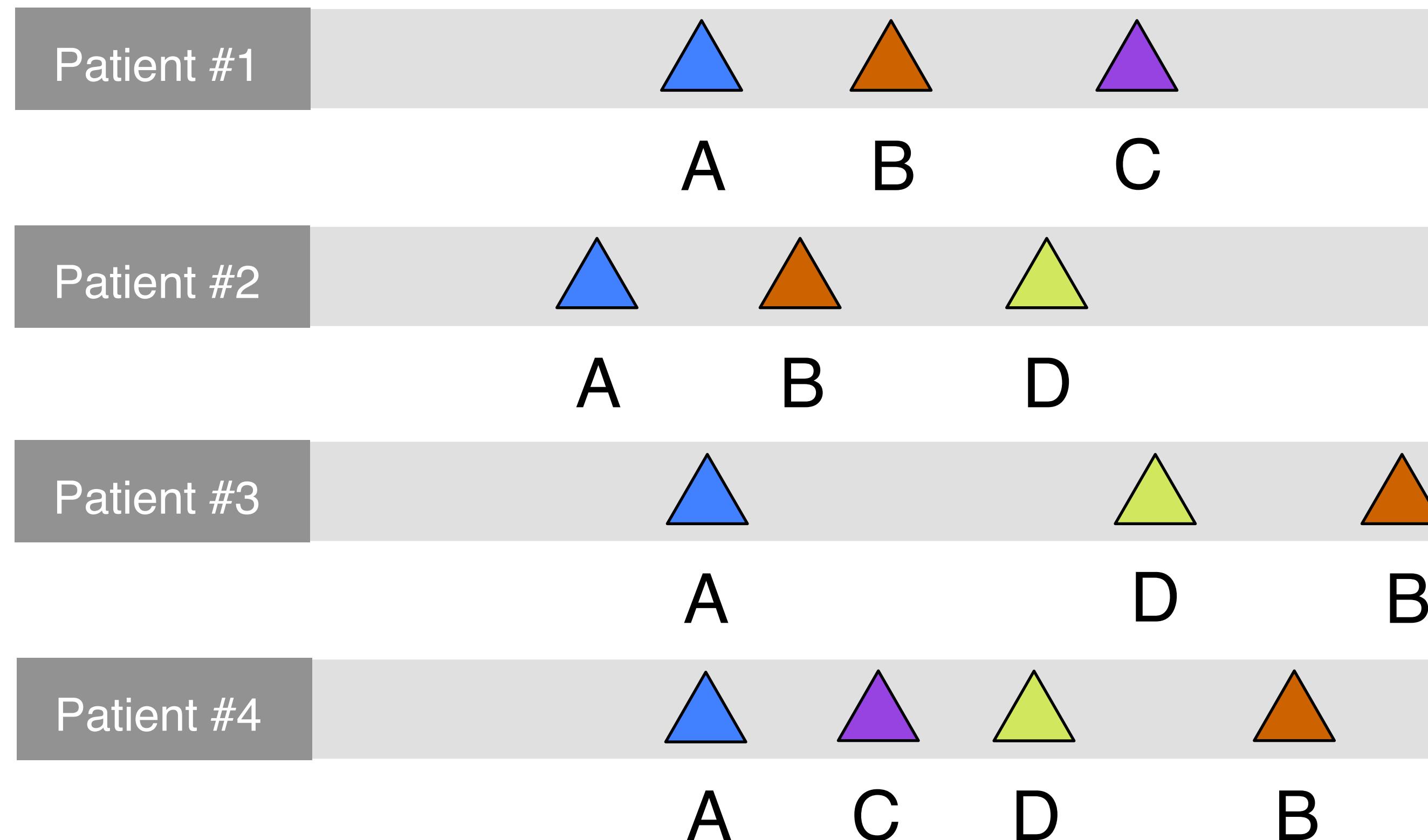
# electronic medical records



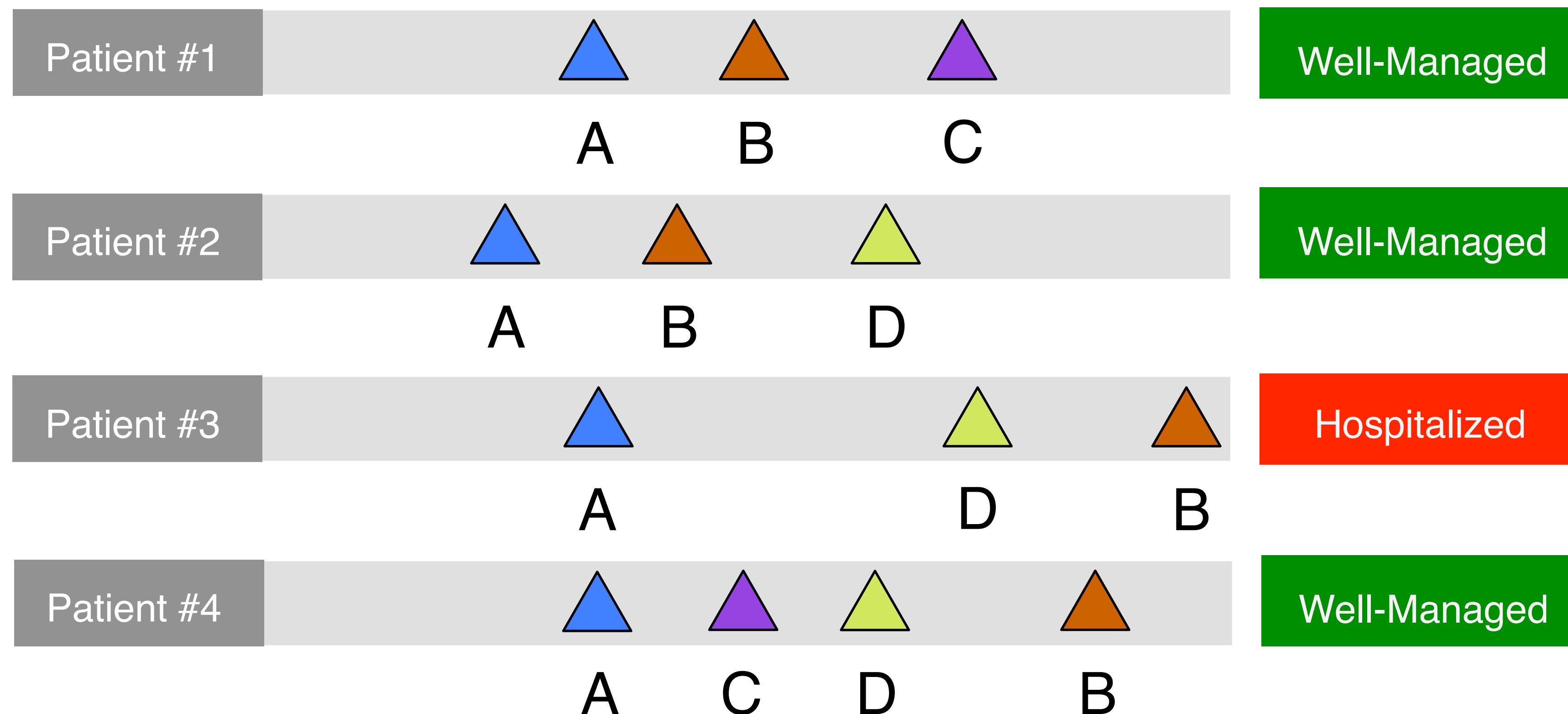
# electronic medical records



# electronic medical records

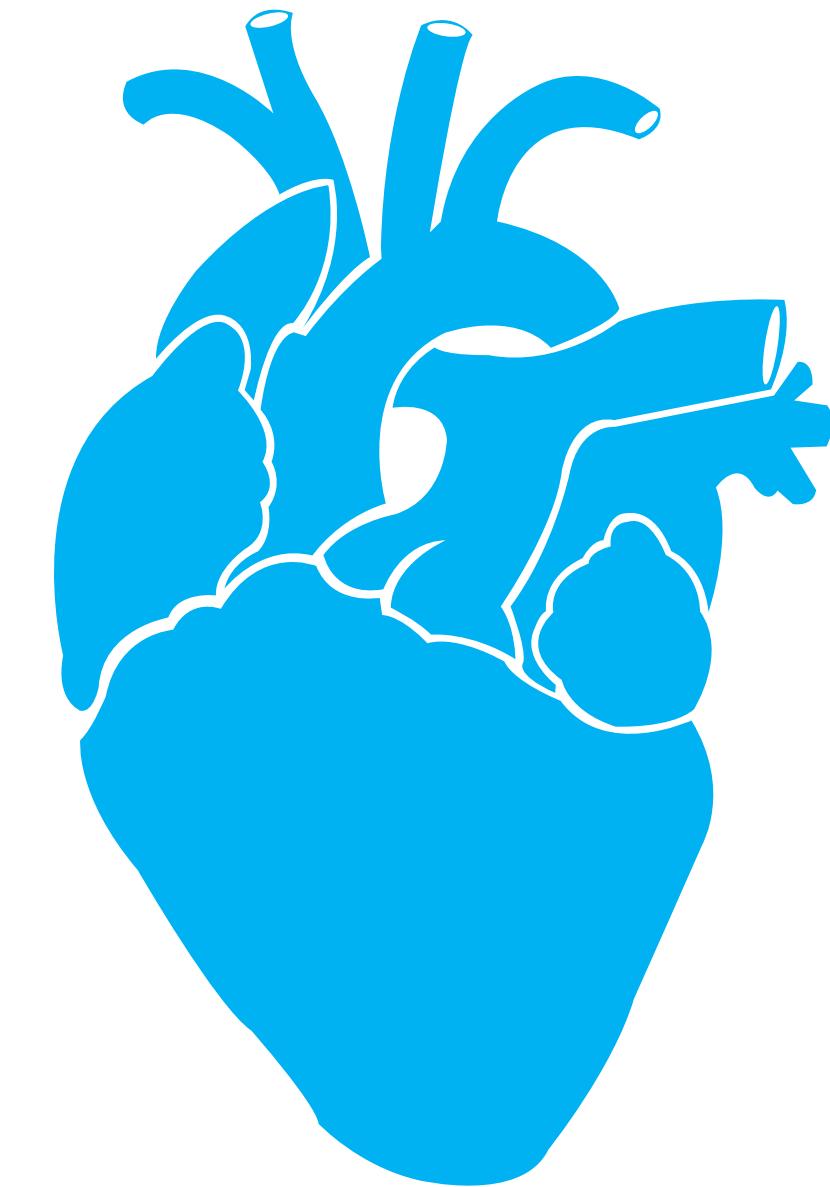


# electronic medical records



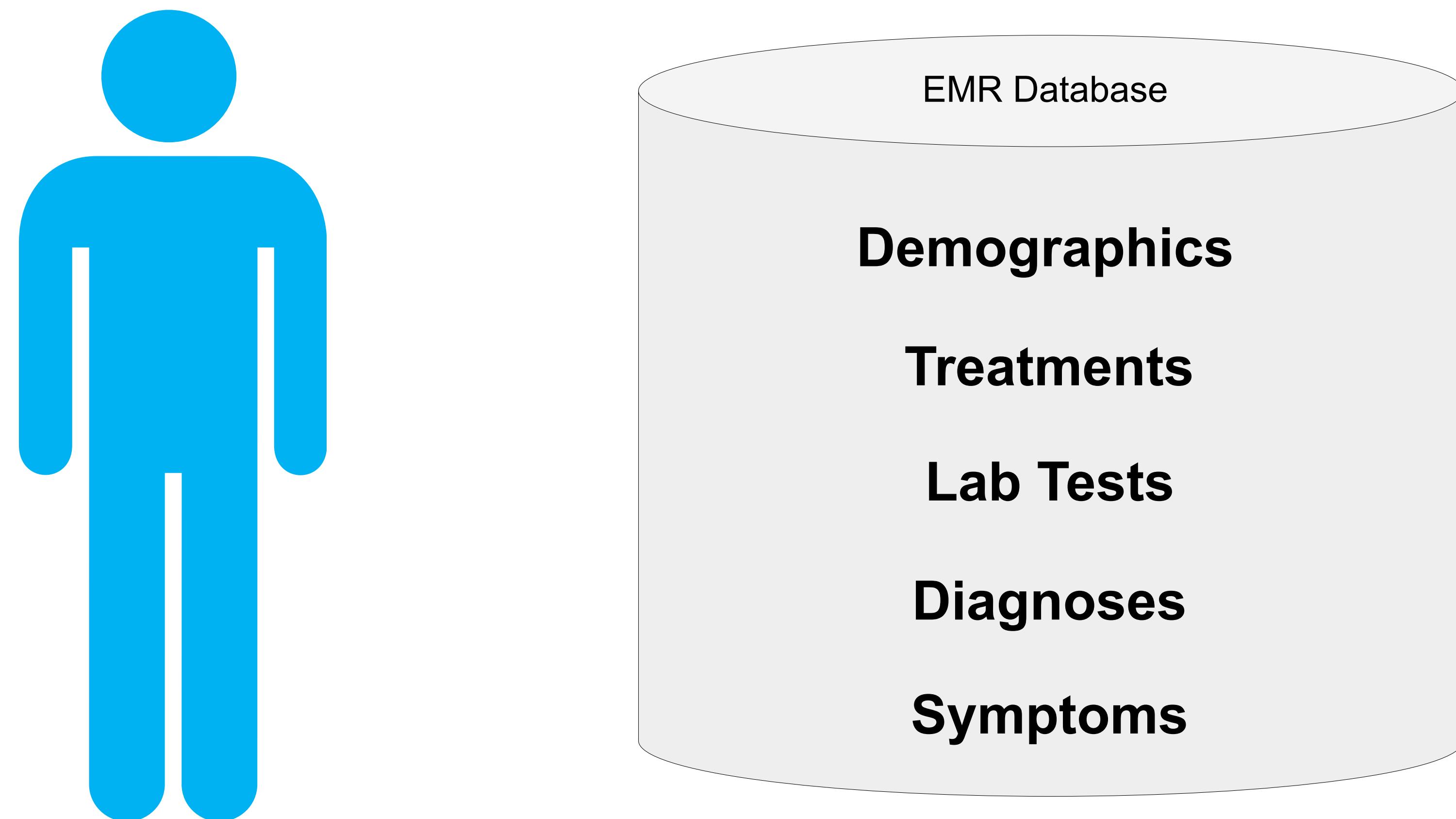
# heart failure

- Potentially fatal disease that affects **2%** of adults in developed countries
- Difficult to manage
- No systematic clinical guidelines for treating Heart Failure
- Presence of co-morbidities affects treatment recommendations.



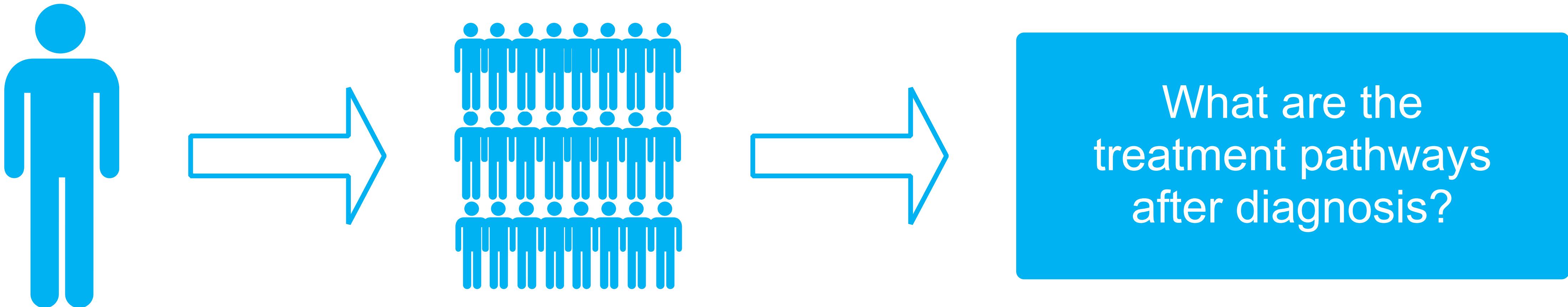
# population

- Hundreds of Thousands of Patients diagnosed with Congestive Heart Failure



# aggregation

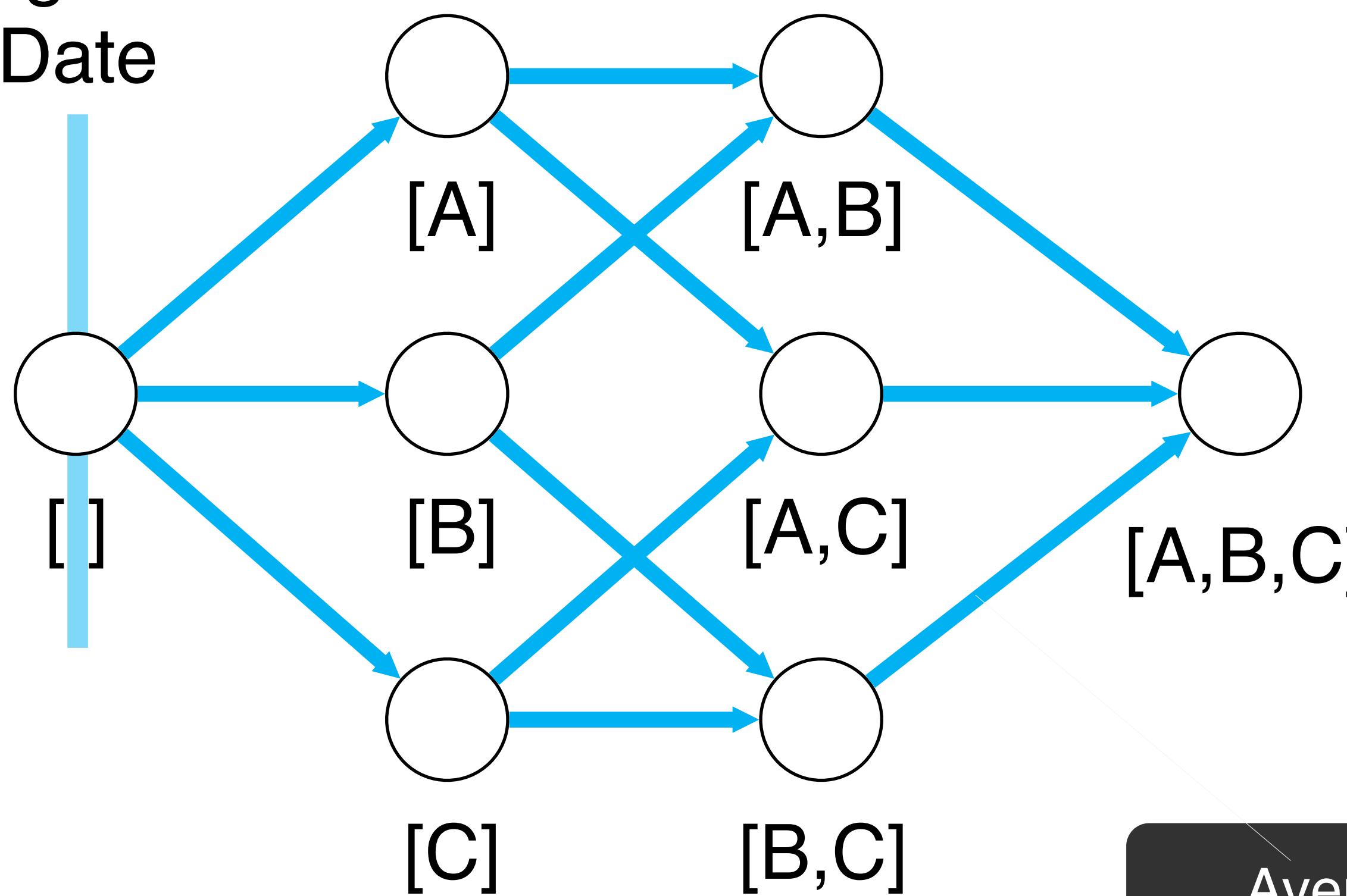
- Start with target patient
- Find similar patients
  - Using our similarity analytics on relevant data
    - Features include medications, symptoms, and diagnoses, and lab tests
- Align all patients by disease diagnosis



# aggregation

Diagnosis

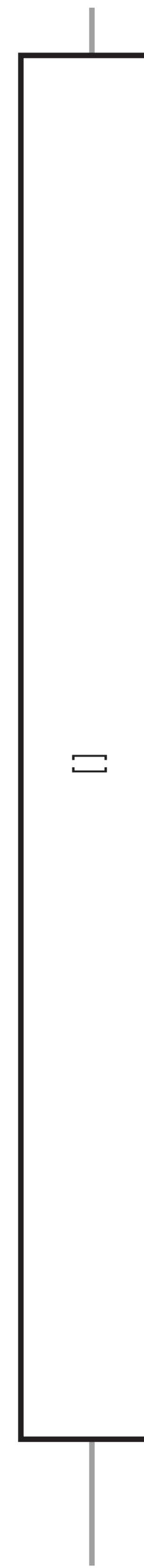
Date



Average outcome = 0.4  
Average time = 10 days  
Number of patients = 10

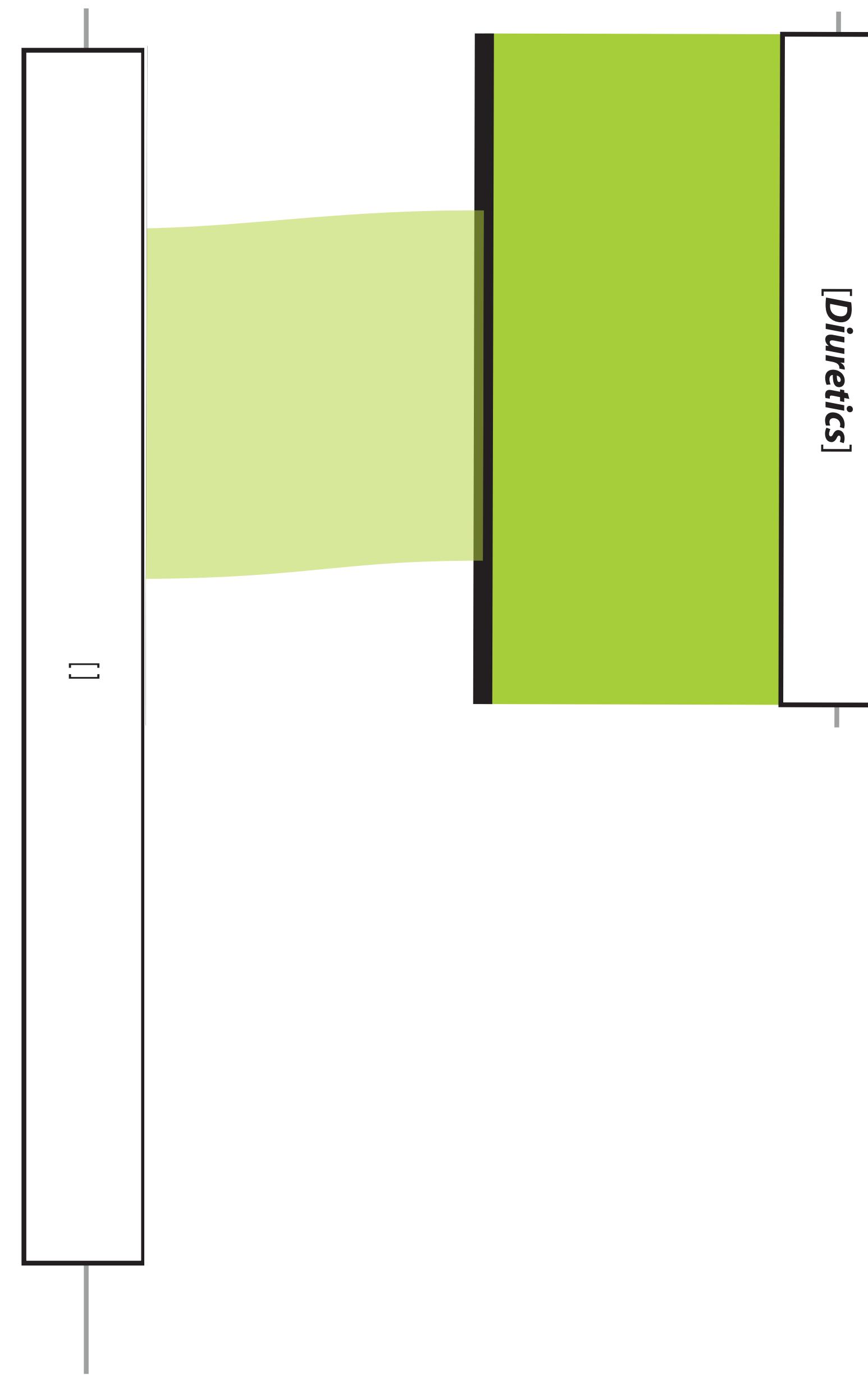
Hospitalized

Managed

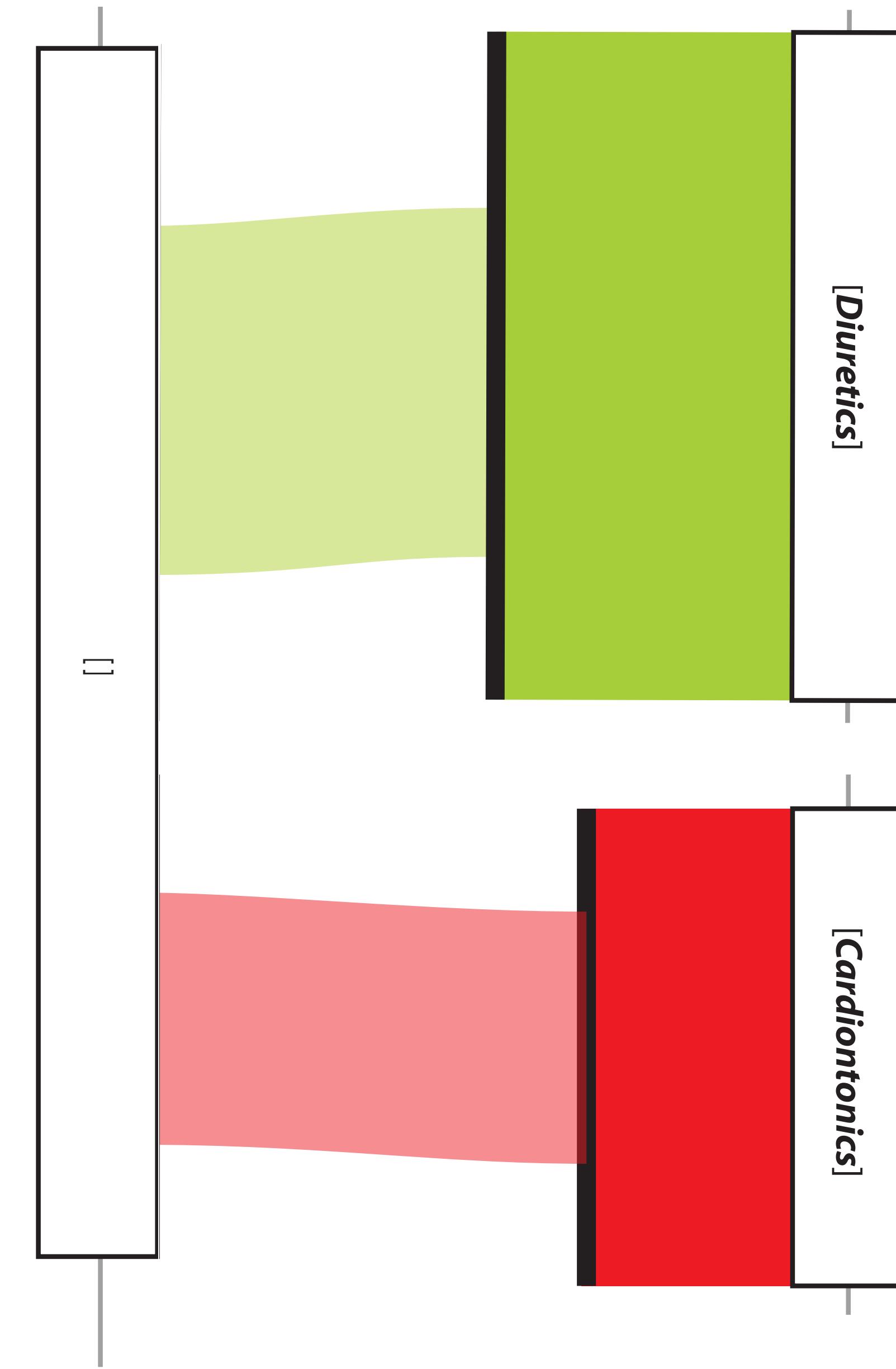


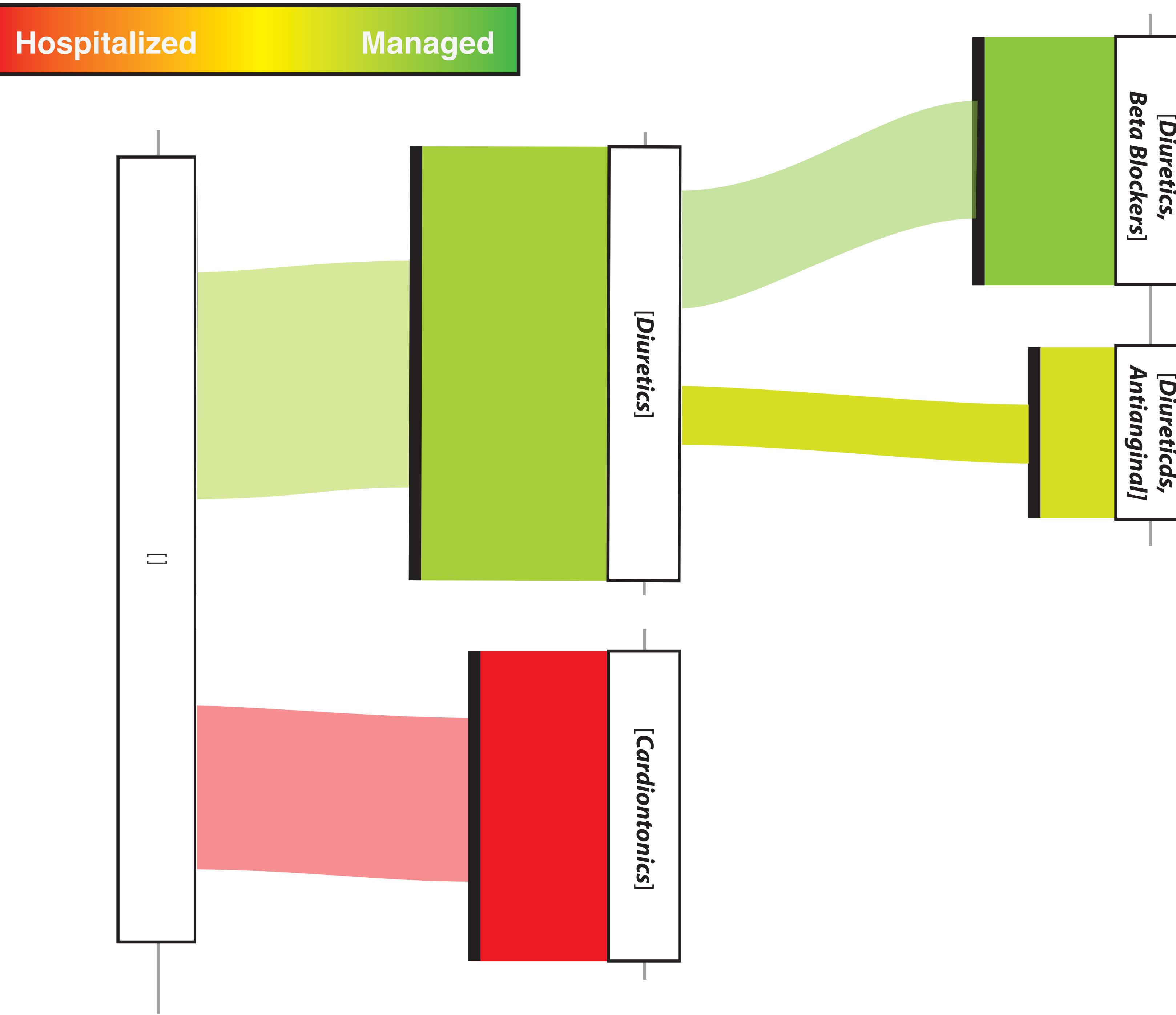
Hospitalized

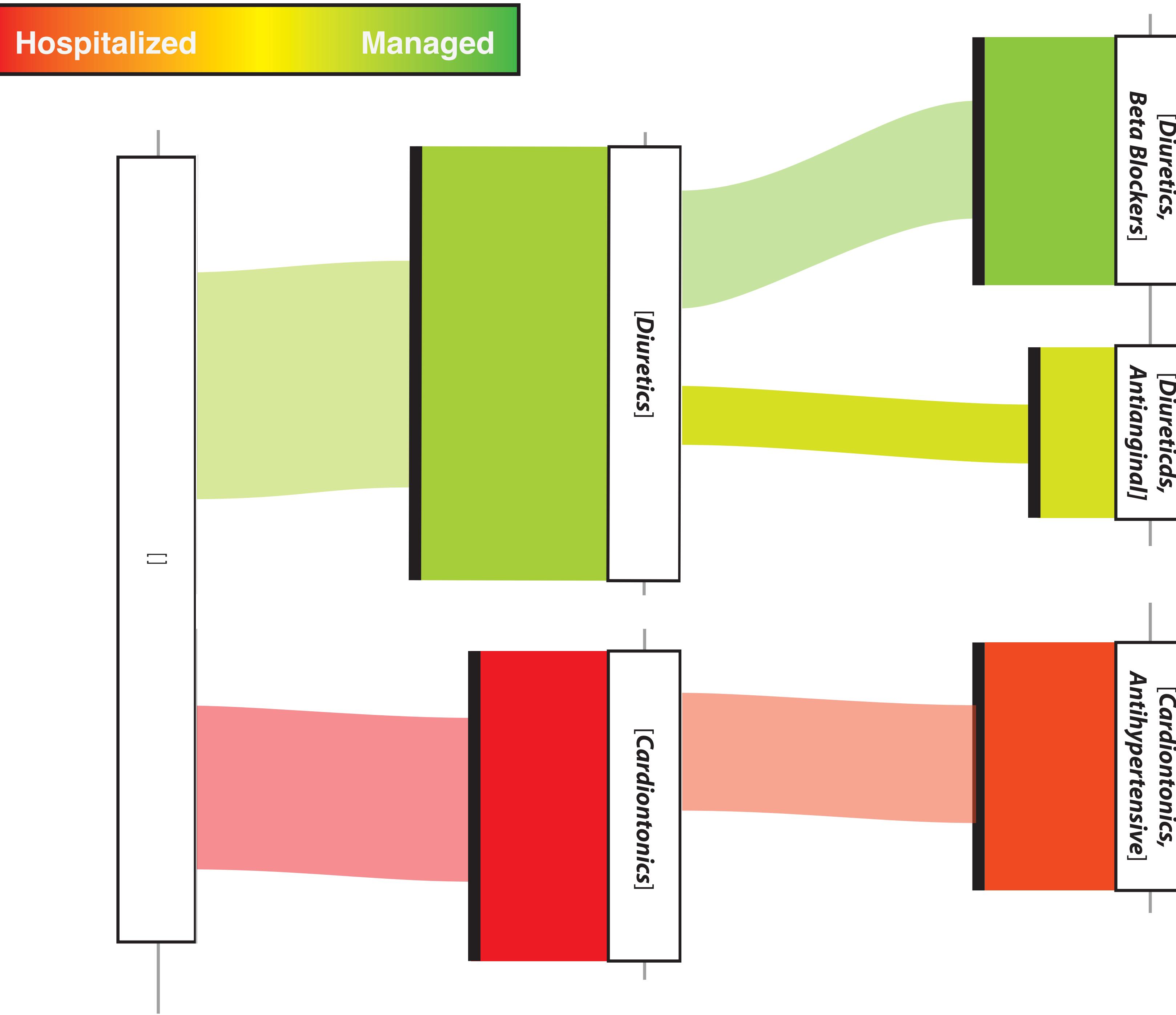
Managed



Hospitalized      Managed

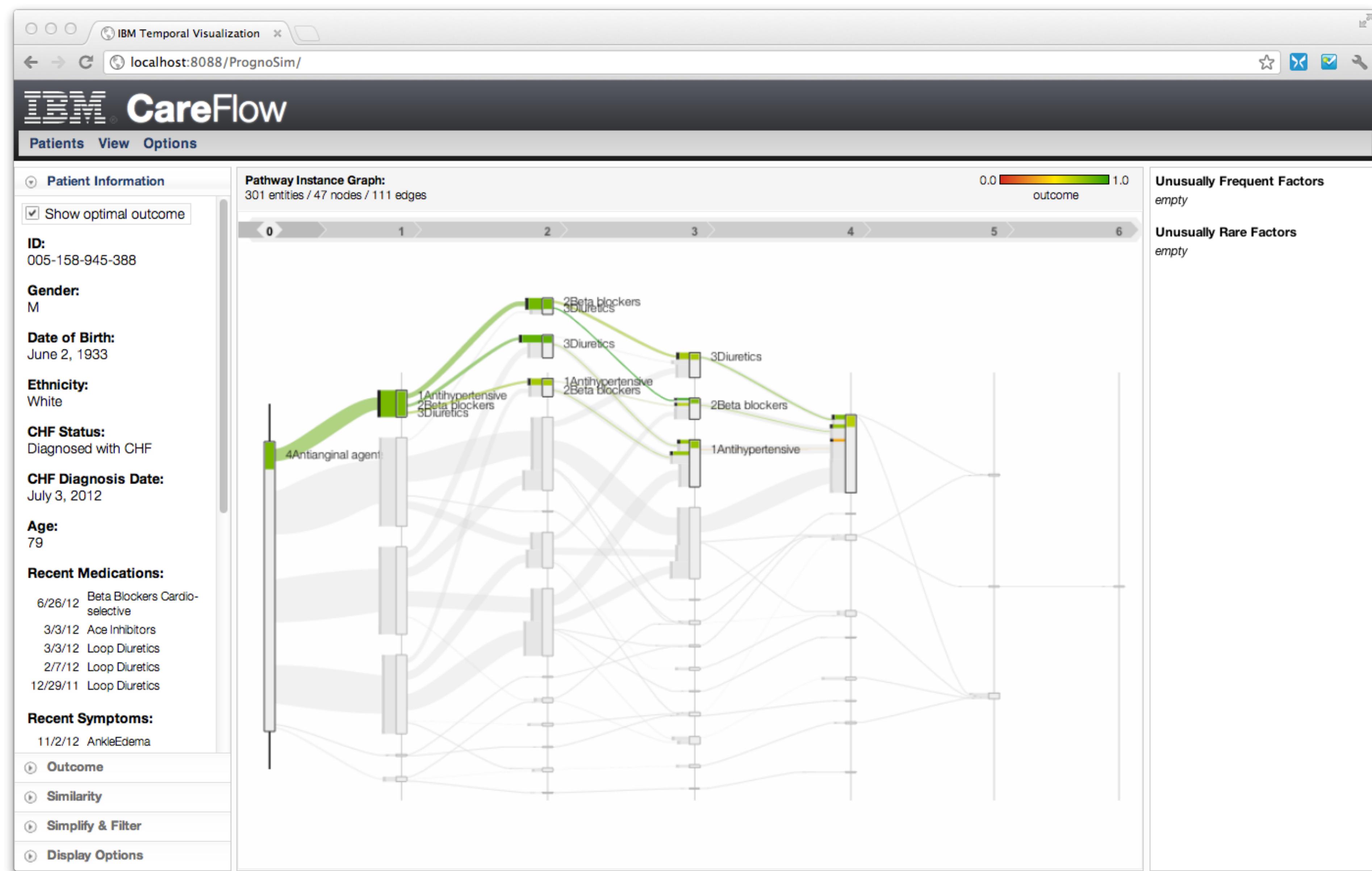








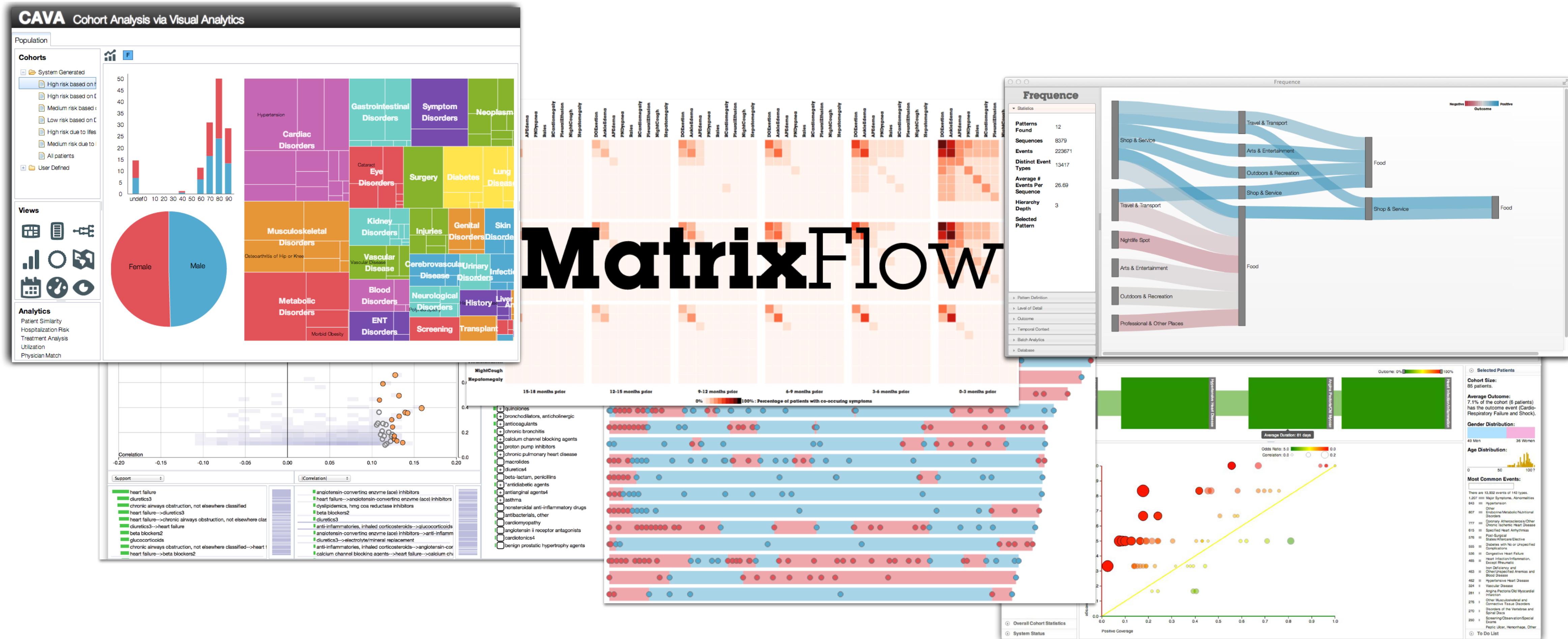
# Care Pathways of 300 similar patients



Optimal Care Pathway  
among 300 similar patients

# other tools for clinical exploration

# MatrixFlow



Videos and Papers at <http://perer.org>

**clinical researchers:**

**clinical researchers:**



# clinical researchers:



# clinical researchers:



# clinical researchers:

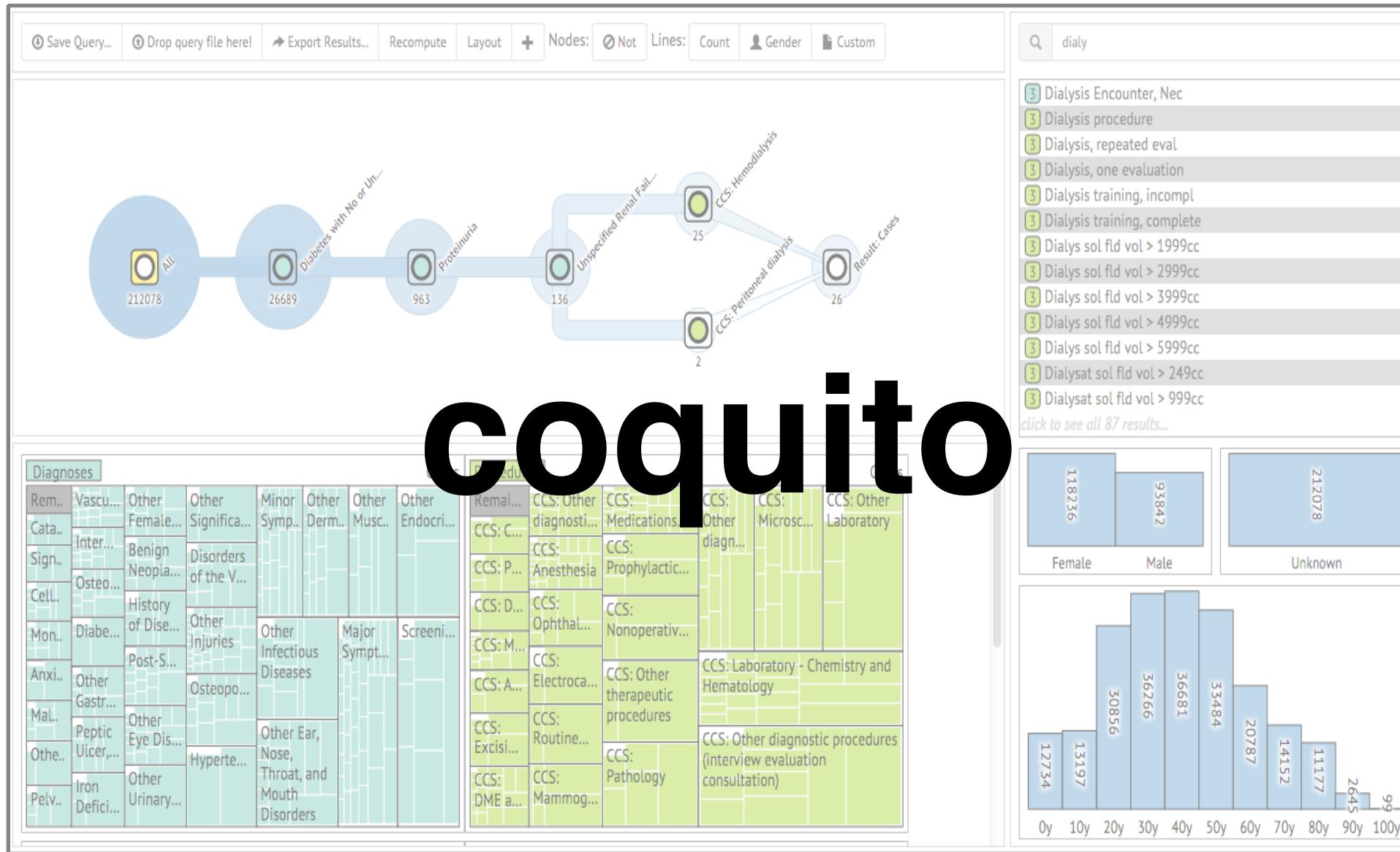


the role  
of visualization  
in prediction

# what can visualization do?

START

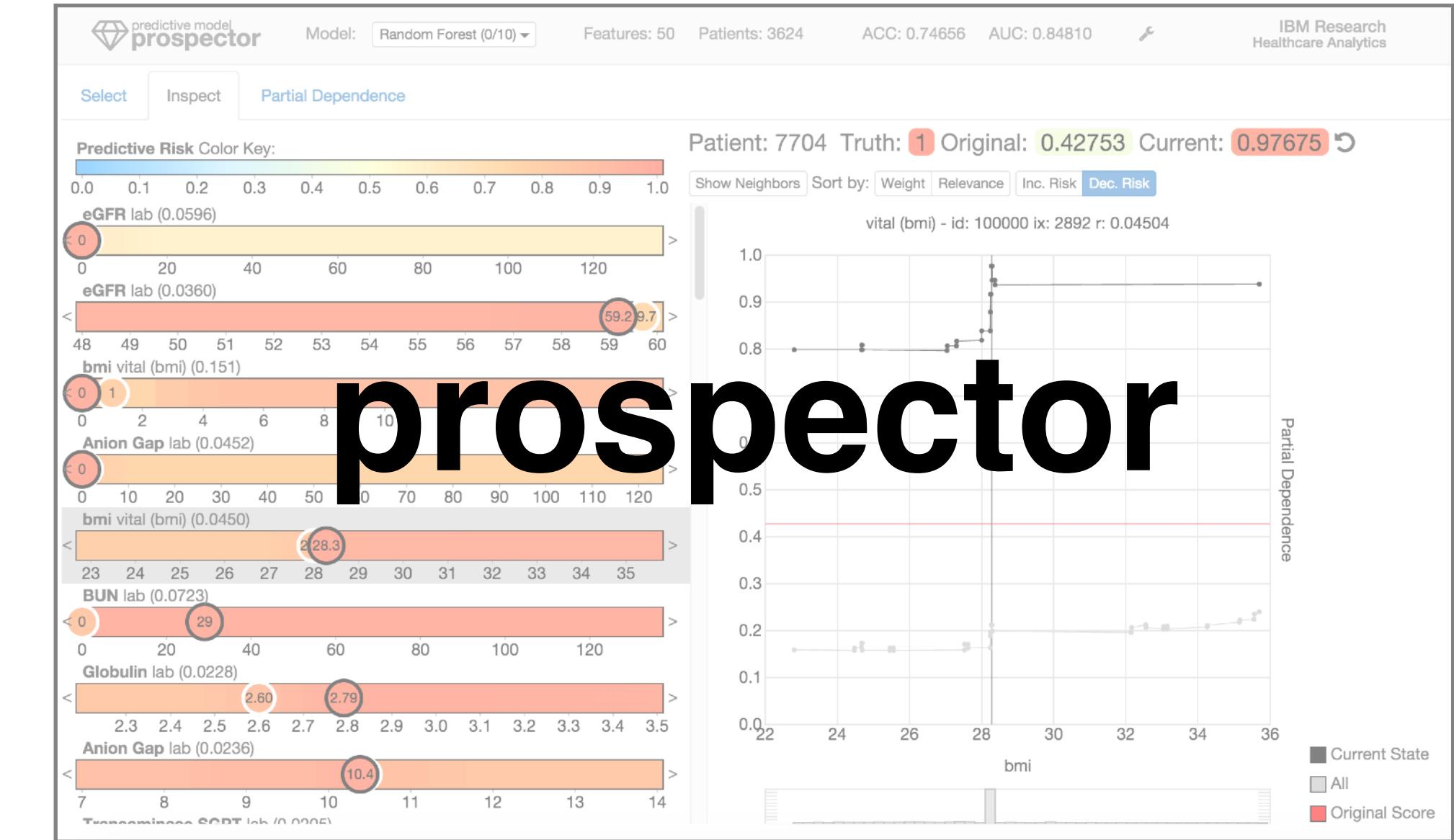
Cohort Definition



coquito

END

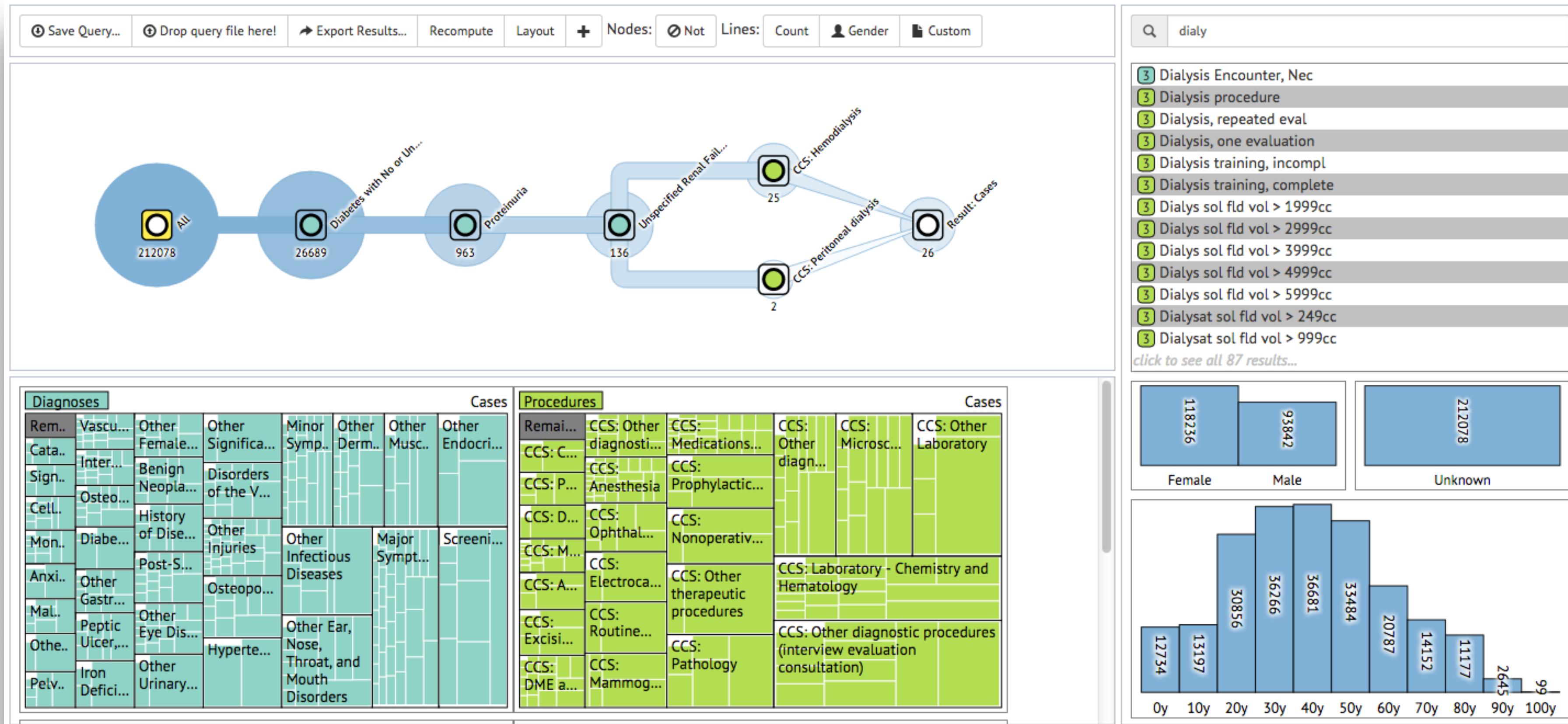
Model Interpretability



prospector

# 1 Cohort Construction

# cohort queries with iterative overviews



Josua Krause, Adam Perer, and Harry Stavropoulos. [Supporting Iterative Cohort Construction with Visual Temporal Queries](#). IEEE Visual Analytics Science and Technology (VAST 2015).

# defining cohorts

- Typically, defining cohorts is a slow process:

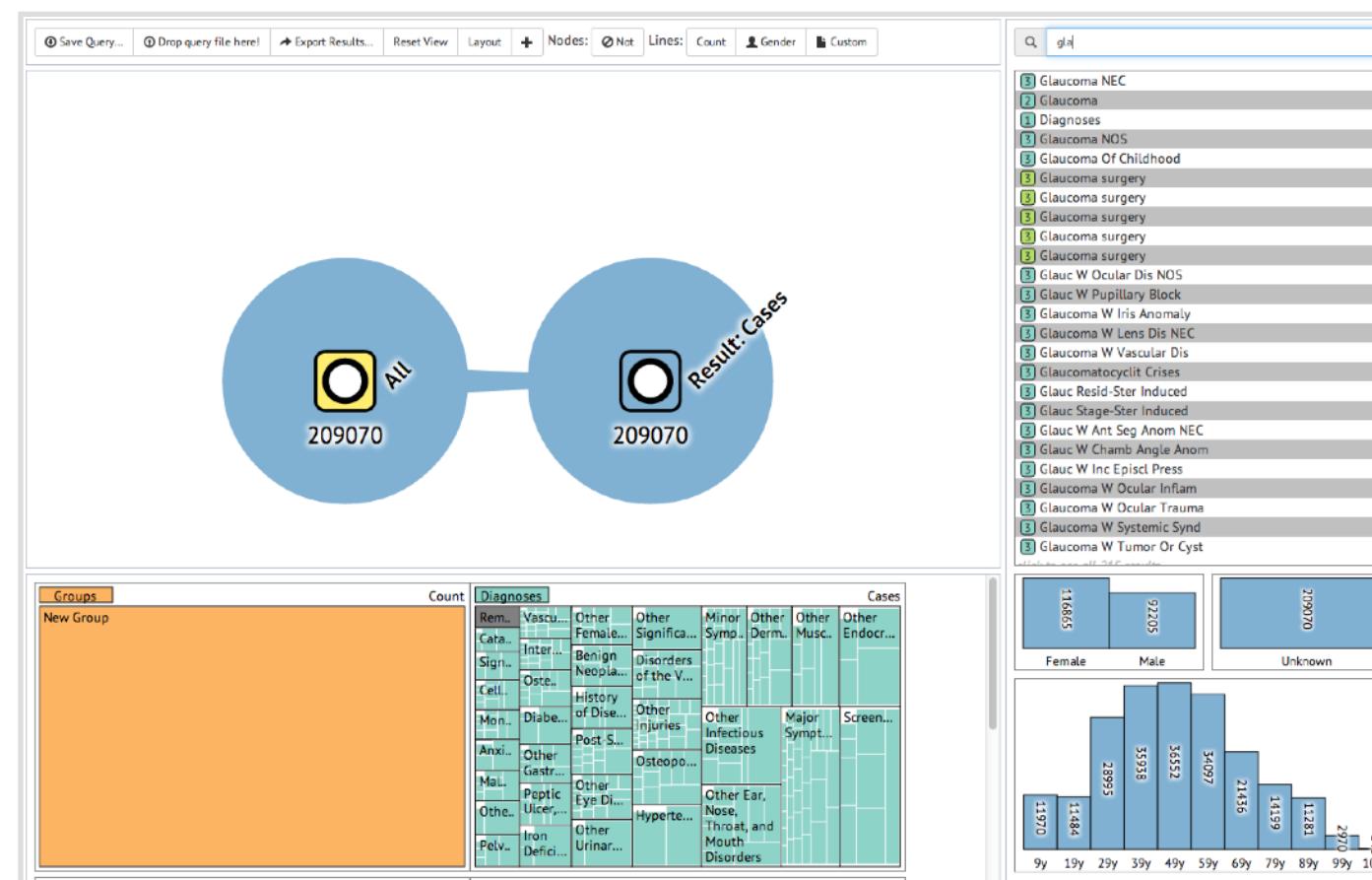
- First, medical researchers define requirements.
- Then, Technologists write SQL queries and deliver them to medical researchers.
- But, often too many patients or too few patients, and the process must restart.



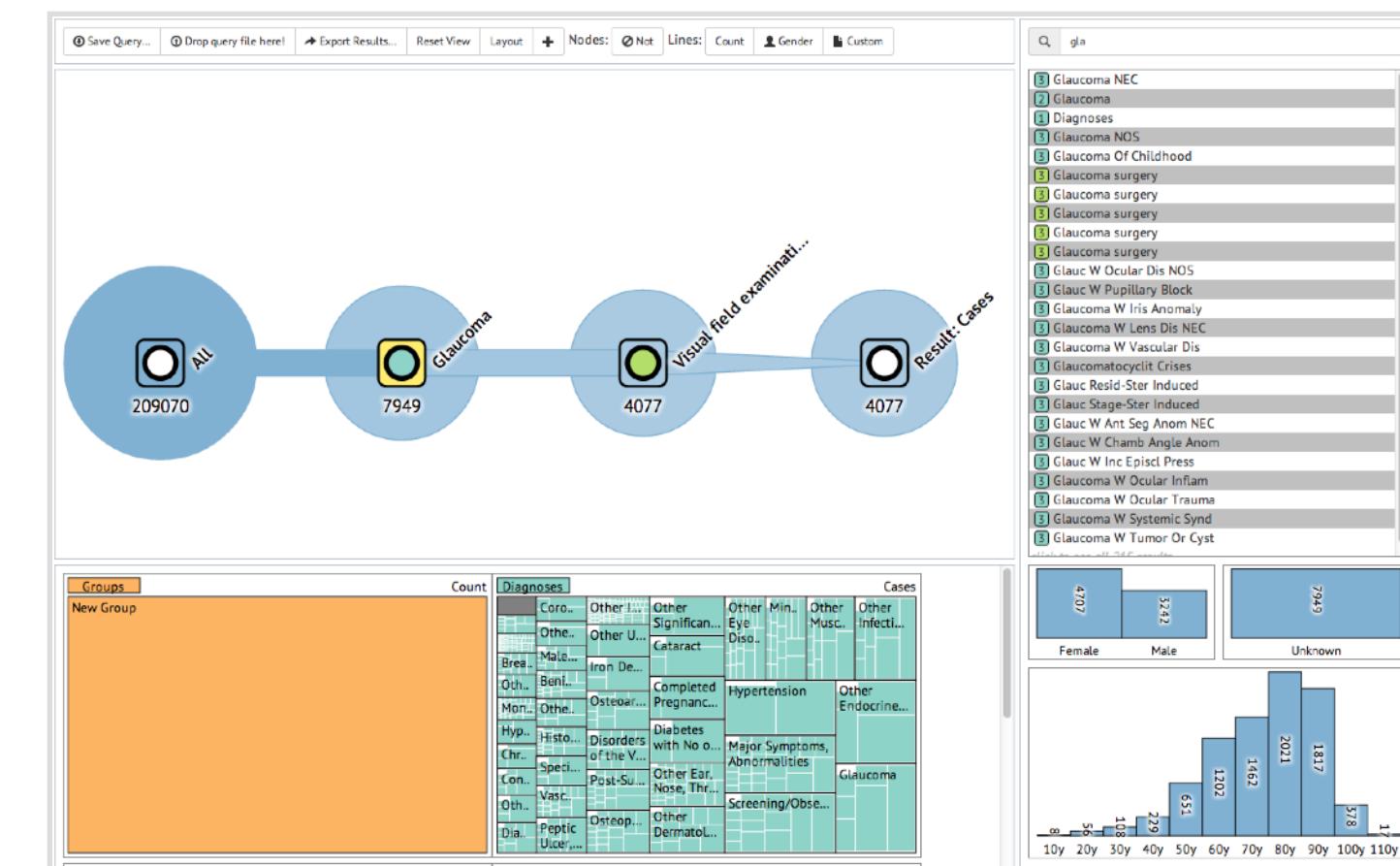
```
215862, 1123300800000), (232190, 1112932800000), (237154, 1113105600000), (248873, 1117252800000), (249428, 1092196800000), (250877, 1112331600000), (251852, 1110269990, 1109048400000), (283194, 1051070400000), (297322, 1123646400000), (304412, 1119067200000) as tmp (pid, from_time) where table.level_1_id is not null and level_2_id is not null and level_3_id is not null and level_4_id is not null and table.gender_description in ('Female','Male') and table.ethnic_group in ('Unknown', 'Asian', 'Black or African American', 'Hispanic', 'American Indian or Alaskan Native', 'White', 'Two or more races') and table.actual_day > 1115956800000 and tmp.pid = table.patient_id and table.actual_day > tmp.from_time) as res group by res.level_1_id, res.level_2_id, res.level_3_id, res.level_4_id
```

```
select count(*), res.level_1_id, res.level_2_id, res.level_3_id, res.level_4_id from (select distinct table.patient_id, table.level_1_id, table.level_2_id, table.level_3_id, table.level_4_id from visual_model_creation_workbench.hierarchies_dates_demographics_sho as table, (values (181, 1030507200000), (381, 1116388800000), (1111, 10562057, 1033704000000), (2197, 1049259600000), (2689, 1039064400000), (3617, 1071550800000), (3962, 1120795200000), (5752, 1070686800000), (6448, 1046840400000), (6661, 1070686800000), (7089432000000), (7785, 1102914000000), (7881, 1080709200000), (7883, 1056427200000), (8245, 1080882000000), (8561, 1075352400000), (8877, 1091592000000), (9081, 1116388800000), (9092, 1073106000000), (9413, 1084680000000), (11480, 1118203200000), (12340, 1112932800000), (13146, 1063339200000), (15220, 1116820800000), (16851, 1124337600000), (17054180800000), (19207, 1074402000000), (19330, 1031284800000), (19880, 1050724800000), (19937, 1124942400000), (20838, 1046754000000), (21927, 1119931200000), (22270, 1045976400000), (23595, 1121140800000), (23651, 1047531600000), (24554, 1048050000000), (25251, 1056600000000), (26443, 1106888400000), (27052, 1052712000000), (29043, 1083643200000), (30803, 1116734400000), (31693, 1095307200000), (32170, 1092888000000), (32777, 1048741200000), (34373, 1116216000000), (35777, 1077771600000), (35177, 1048309200000), (36184, 1042088400000), (37060, 10632239200000), (39414, 1121054100000), (40015, 1052984800000), (41565, 1040000400000), (42000, 1040000400000), (42500, 1040000400000), (43000, 1040000400000), (43500, 1040000400000), (44000, 1040000400000), (44500, 1040000400000), (45000, 1040000400000), (45500, 1040000400000), (46000, 1040000400000), (46500, 1040000400000), (47000, 1040000400000), (47500, 1040000400000), (48000, 1040000400000), (48500, 1040000400000), (49000, 1040000400000), (49500, 1040000400000), (50000, 1040000400000), (50500, 1040000400000), (51000, 1040000400000), (51500, 1040000400000), (52000, 1040000400000), (52500, 1040000400000), (53000, 1040000400000), (53500, 1040000400000), (54000, 1040000400000), (54500, 1040000400000), (55000, 1040000400000), (55500, 1040000400000), (56000, 1040000400000), (56500, 1040000400000), (57000, 1040000400000), (57500, 1040000400000), (58000, 1040000400000), (58500, 1040000400000), (59000, 1040000400000), (59500, 1040000400000), (60000, 1040000400000), (60500, 1040000400000), (61000, 1040000400000), (61500, 1040000400000), (62000, 1040000400000), (62500, 1040000400000), (63000, 1040000400000), (63500, 1040000400000), (64000, 1040000400000), (64500, 1040000400000), (65000, 1040000400000), (65500, 1040000400000), (66000, 1040000400000), (66500, 1040000400000), (67000, 1040000400000), (67500, 1040000400000), (68000, 1040000400000), (68500, 1040000400000), (69000, 1040000400000), (69500, 1040000400000), (70000, 1040000400000), (70500, 1040000400000), (71000, 1040000400000), (71500, 1040000400000), (72000, 1040000400000), (72500, 1040000400000), (73000, 1040000400000), (73500, 1040000400000), (74000, 1040000400000), (74500, 1040000400000), (75000, 1040000400000), (75500, 1040000400000), (76000, 1040000400000), (76500, 1040000400000), (77000, 1040000400000), (77500, 1040000400000), (78000, 1040000400000), (78500, 1040000400000), (79000, 1040000400000), (79500, 1040000400000), (80000, 1040000400000), (80500, 1040000400000), (81000, 1040000400000), (81500, 1040000400000), (82000, 1040000400000), (82500, 1040000400000), (83000, 1040000400000), (83500, 1040000400000), (84000, 1040000400000), (84500, 1040000400000), (85000, 1040000400000), (85500, 1040000400000), (86000, 1040000400000), (86500, 1040000400000), (87000, 1040000400000), (87500, 1040000400000), (88000, 1040000400000), (88500, 1040000400000), (89000, 1040000400000), (89500, 1040000400000), (90000, 1040000400000), (90500, 1040000400000), (91000, 1040000400000), (91500, 1040000400000), (92000, 1040000400000), (92500, 1040000400000), (93000, 1040000400000), (93500, 1040000400000), (94000, 1040000400000), (94500, 1040000400000), (95000, 1040000400000), (95500, 1040000400000), (96000, 1040000400000), (96500, 1040000400000), (97000, 1040000400000), (97500, 1040000400000), (98000, 1040000400000), (98500, 1040000400000), (99000, 1040000400000), (99500, 1040000400000), (100000, 1040000400000), (100500, 1040000400000), (101000, 1040000400000), (101500, 1040000400000), (102000, 1040000400000), (102500, 1040000400000), (103000, 1040000400000), (103500, 1040000400000), (104000, 1040000400000), (104500, 1040000400000), (105000, 1040000400000), (105500, 1040000400000), (106000, 1040000400000), (106500, 1040000400000), (107000, 1040000400000), (107500, 1040000400000), (108000, 1040000400000), (108500, 1040000400000), (109000, 1040000400000), (109500, 1040000400000), (110000, 1040000400000), (110500, 1040000400000), (111000, 1040000400000), (111500, 1040000400000), (112000, 1040000400000), (112500, 1040000400000), (113000, 1040000400000), (113500, 1040000400000), (114000, 1040000400000), (114500, 1040000400000), (115000, 1040000400000), (115500, 1040000400000), (116000, 1040000400000), (116500, 1040000400000), (117000, 1040000400000), (117500, 1040000400000), (118000, 1040000400000), (118500, 1040000400000), (119000, 1040000400000), (119500, 1040000400000), (120000, 1040000400000), (120500, 1040000400000), (121000, 1040000400000), (121500, 1040000400000), (122000, 1040000400000), (122500, 1040000400000), (123000, 1040000400000), (123500, 1040000400000), (124000, 1040000400000), (124500, 1040000400000), (125000, 1040000400000), (125500, 1040000400000), (126000, 1040000400000), (126500, 1040000400000), (127000, 1040000400000), (127500, 1040000400000), (128000, 1040000400000), (128500, 1040000400000), (129000, 1040000400000), (129500, 1040000400000), (130000, 1040000400000), (130500, 1040000400000), (131000, 1040000400000), (131500, 1040000400000), (132000, 1040000400000), (132500, 1040000400000), (133000, 1040000400000), (133500, 1040000400000), (134000, 1040000400000), (134500, 1040000400000), (135000, 1040000400000), (135500, 1040000400000), (136000, 1040000400000), (136500, 1040000400000), (137000, 1040000400000), (137500, 1040000400000), (138000, 1040000400000), (138500, 1040000400000), (139000, 1040000400000), (139500, 1040000400000), (140000, 1040000400000), (140500, 1040000400000), (141000, 1040000400000), (141500, 1040000400000), (142000, 1040000400000), (142500, 1040000400000), (143000, 1040000400000), (143500, 1040000400000), (144000, 1040000400000), (144500, 1040000400000), (145000, 1040000400000), (145500, 1040000400000), (146000, 1040000400000), (146500, 1040000400000), (147000, 1040000400000), (147500, 1040000400000), (148000, 1040000400000), (148500, 1040000400000), (149000, 1040000400000), (149500, 1040000400000), (150000, 1040000400000), (150500, 1040000400000), (151000, 1040000400000), (151500, 1040000400000), (152000, 1040000400000), (152500, 1040000400000), (153000, 1040000400000), (153500, 1040000400000), (154000, 1040000400000), (154500, 1040000400000), (155000, 1040000400000), (155500, 1040000400000), (156000, 1040000400000), (156500, 1040000400000), (157000, 1040000400000), (157500, 1040000400000), (158000, 1040000400000), (158500, 1040000400000), (159000, 1040000400000), (159500, 1040000400000), (160000, 1040000400000), (160500, 1040000400000), (161000, 1040000400000), (161500, 1040000400000), (162000, 1040000400000), (162500, 1040000400000), (163000, 1040000400000), (163500, 1040000400000), (164000, 1040000400000), (164500, 1040000400000), (165000, 1040000400000), (165500, 1040000400000), (166000, 1040000400000), (166500, 1040000400000), (167000, 1040000400000), (167500, 1040000400000), (168000, 1040000400000), (168500, 1040000400000), (169000, 1040000400000), (169500, 1040000400000), (170000, 1040000400000), (170500, 1040000400000), (171000, 1040000400000), (171500, 104000040
```

# defining cohorts with coquito



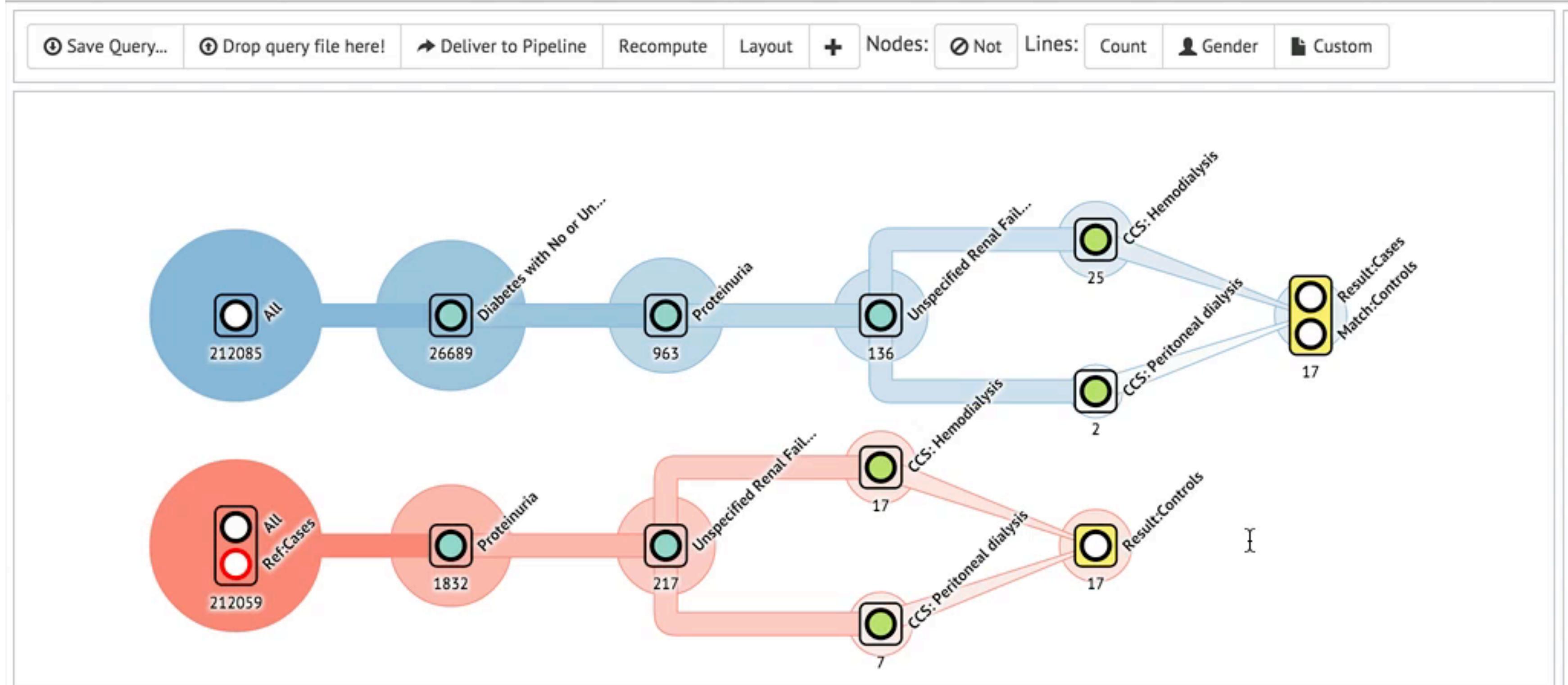
drag and drop constraints  
with immediate feedback  
and hints for query  
refinement



supports using complex  
temporal logic



support for multiple  
queries side-by-side (for  
cases and controls)

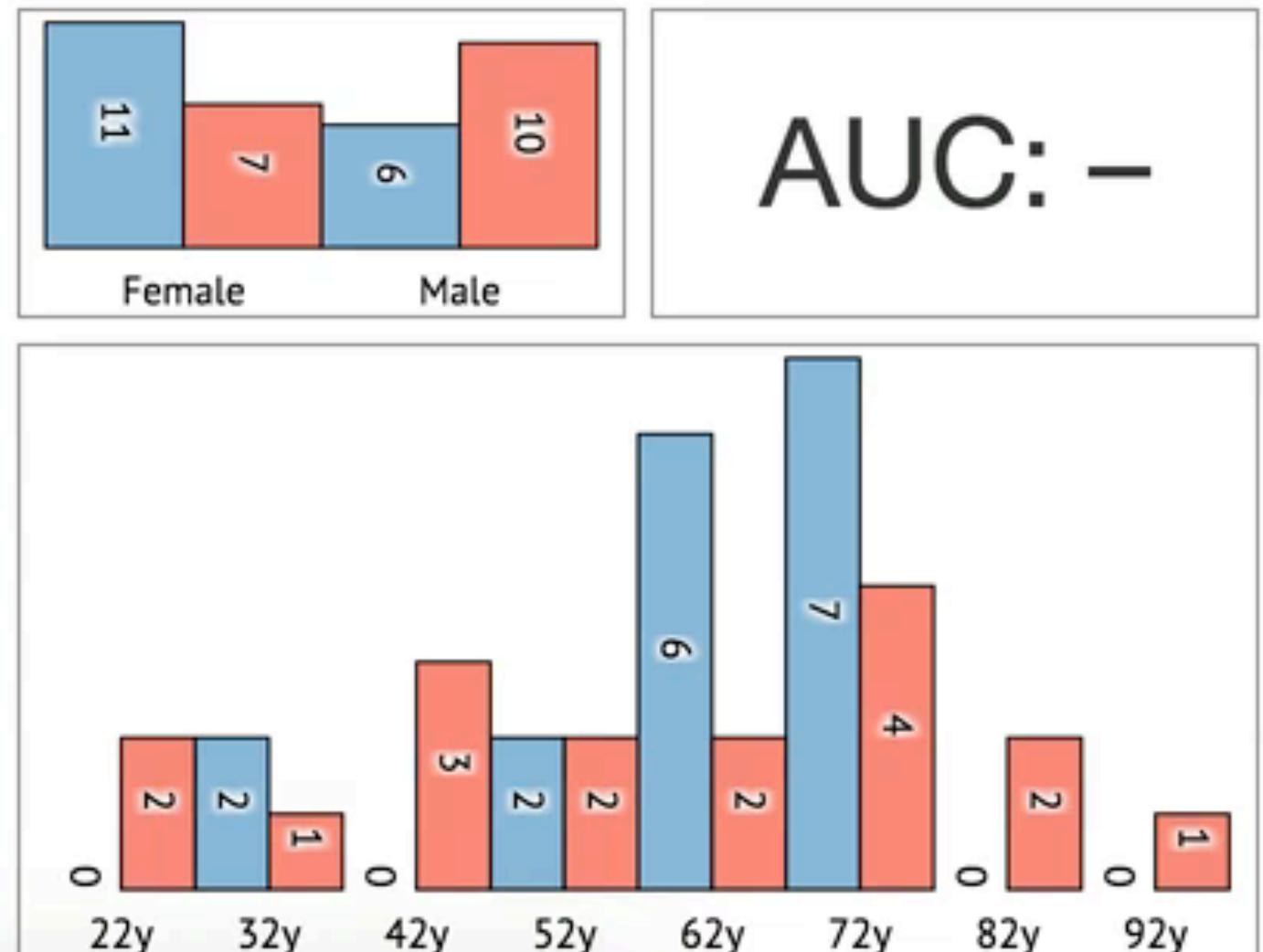


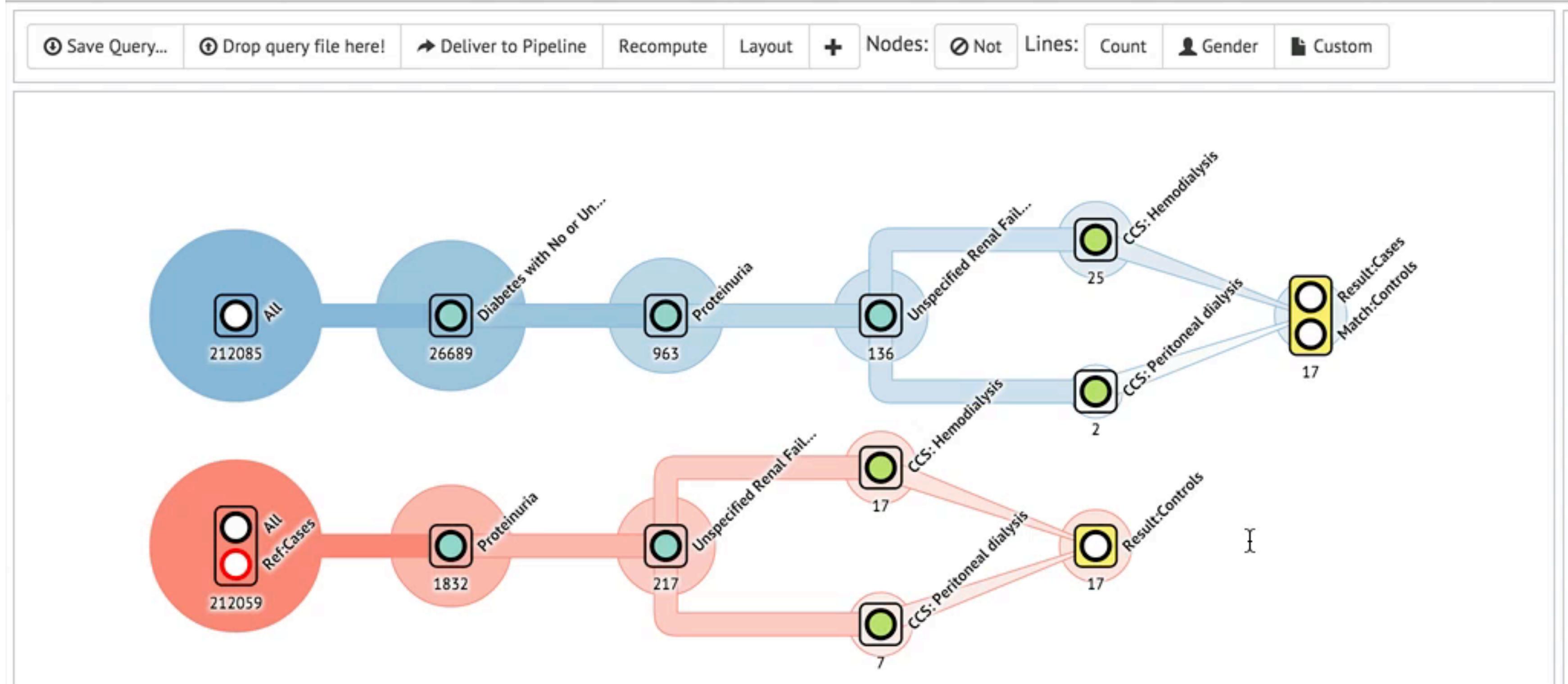
3 Peritoneal Abscess  
3 Peritoneal Adhesions  
3 Peritoneal Dialysis Test  
3 Peritoneal Disorder NEC  
3 Peritoneal Disorder NOS  
3 Peritoneal Effusion  
2 CCS: Peritoneal dialysis  
3 Drain to peritoneal cavity  
3 Puncture, peritoneal cavity  
3 Colpopexy, extraperitoneal  
3 Colpopexy, intraperitoneal  
2 CCS: Excision lysis peritoneal adhesions  
3 Hemoperfusion  
3 Dialysis procedure  
3 Dialysis, repeated eval  
3 Dialysis, one evaluation  
3 Dialysis training, incompl

click to see all 19 results...

Diagnoses				Controls				
Rem..	Othe..	Osteo...	Other M...	Chro..	Dial..	Urin..	Other Urin..	Diso..
Hist..	Othe..	Other...	Viral an...					
Spec..	Chro..	Other	Other Ear, Nose, T...	Other Infectiou...	Iron Defic...	Chro..	Post...	
Othe..	Beni..	Gastr...	Other					
Coro..	Diso..	Peptic Ulcer,...	Dermat...	Chronic Kidney...				
Com..	Cata..	Chronic Obstr...	Nephritis	Screenin...	Hyperten...	Min..	Maj..	
Inte..	Chro..	Other Eye D...	Acute Renal F...					
Diab..	Inte..	Chro..	Vascul...	Other Endocrin...	Unspecifi...			
Coag..			Chronic Kidney...					

Procedures				Controls				
Rema..	CCS: Other v...	CCS: Other n...	CCS: Other...	CCS: DME and sup...	CCS: Diag..	CCS: Elect...	CCS: Medi..	CCS: Other therap...
CCS:...	CCS: Cardiac...	CCS: Nonoper...	CCS: Anesthe...	CCS: Other Labo...	CCS: Other OR proc...	CCS: Other diagn...		
CCS: I...	CCS: Radiois...	CCS: Excisio...	CCS: Other diagnostic ra...	CCS: Prophyl...	CCS: Routine chest X-ray	CCS: Laboratory	CCS: Hemodia...	
CCS:...	CCS: Chronic...	CCS: Periton...	CCS: Ophthal...	CCS: Pathology	CCS: Microscopic e...			
CCS:...	CCS: Other d...	CCS: Other d...	CCS: Other d...	CCS: Other d...	CCS: Other d...			



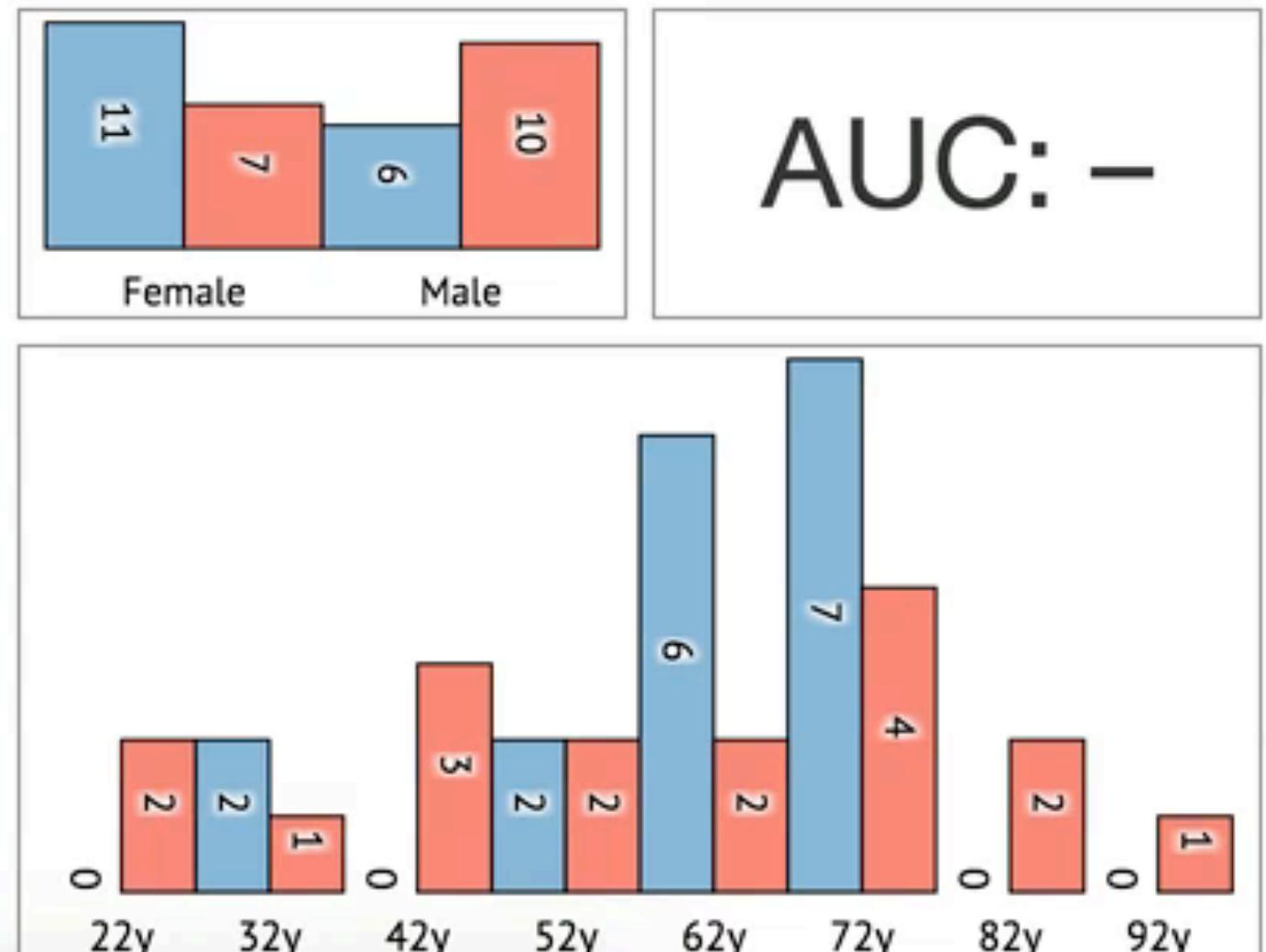


Peritoneal Abscess  
Peritoneal Adhesions  
Peritoneal Dialysis Test  
Peritoneal Disorder NEC  
Peritoneal Disorder NOS  
Peritoneal Effusion  
CCS: Peritoneal dialysis  
Drain to peritoneal cavity  
Puncture, peritoneal cavity  
Colpopexy, extraperitoneal  
Colpopexy, intraperitoneal  
CCS: Excision lysis peritoneal adhesions  
Hemoperfusion  
Dialysis procedure  
Dialysis, repeated eval  
Dialysis, one evaluation  
Dialysis training, incompl

click to see all 19 results...

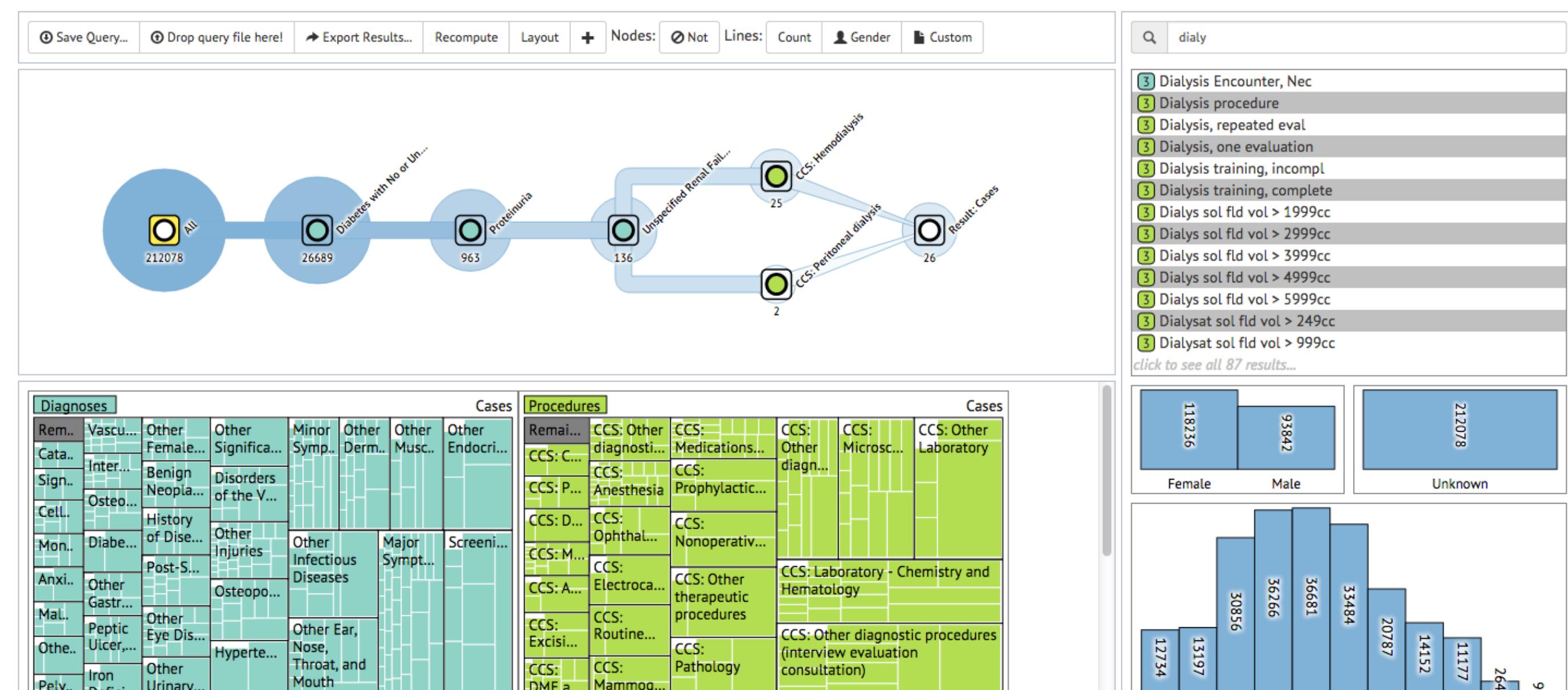
Diagnoses				Controls				
Rem..	Othe..	Osteo...	Other M...	Chro..	Dial..	Urin..	Other Urin..	Diso..
Hist..	Othe..	Other...	Viral an...					
Spec..	Chro..	Other	Other Ear, Nose, T...	Other Infectiou...	Iron Defic...	Chro..	Post...	
Othe..	Beni..	Gastr...	Other					
Coro..	Diso..	Peptic Ulcer,...	Dermat...	Chronic Kidney...				
Com..	Cata..	Chronic Obstr...	Nephritis	Screenin...	Hyperten...	Min..	Maj..	
Inte..	Chro..	Other Eye D...	Acute Renal F...					
Diab..	Inte..	Chro..	Vascul...	Other Endocrin...	Unspecifi...			
Coag..			Chronic Kidney...					

Procedures				Controls				
Rema..	CCS: Other v...	CCS: Other n...	CCS: Other...	CCS: DME and sup...	CCS: Diag..	CCS: Elect...	CCS: Medi..	CCS: Other therap...
CCS:...	CCS: Cardiac...	CCS: Nonoper...	CCS: Anesthe...	CCS: Other Labo...	CCS: Other OR proc...	CCS: Other diagn...		
CCS: I...	CCS: Radiois...	CCS: Excisio...	CCS: Other diagnostic ra...	CCS: Prophyl...	CCS: Routine chest X-ray	CCS: Laboratory	CCS: Hemodia...	
CCS:...	CCS: Periton...	CCS: Ophthal...	CCS: Pathology	CCS: Microscopic e...				
CCS:...	CCS: Creatio...	CCS: Other d...	CCS: Pathology	CCS: Microscopic e...				
CCS:...	CCS: Ancill...	CCS: Other d...	CCS: Pathology	CCS: Microscopic e...				



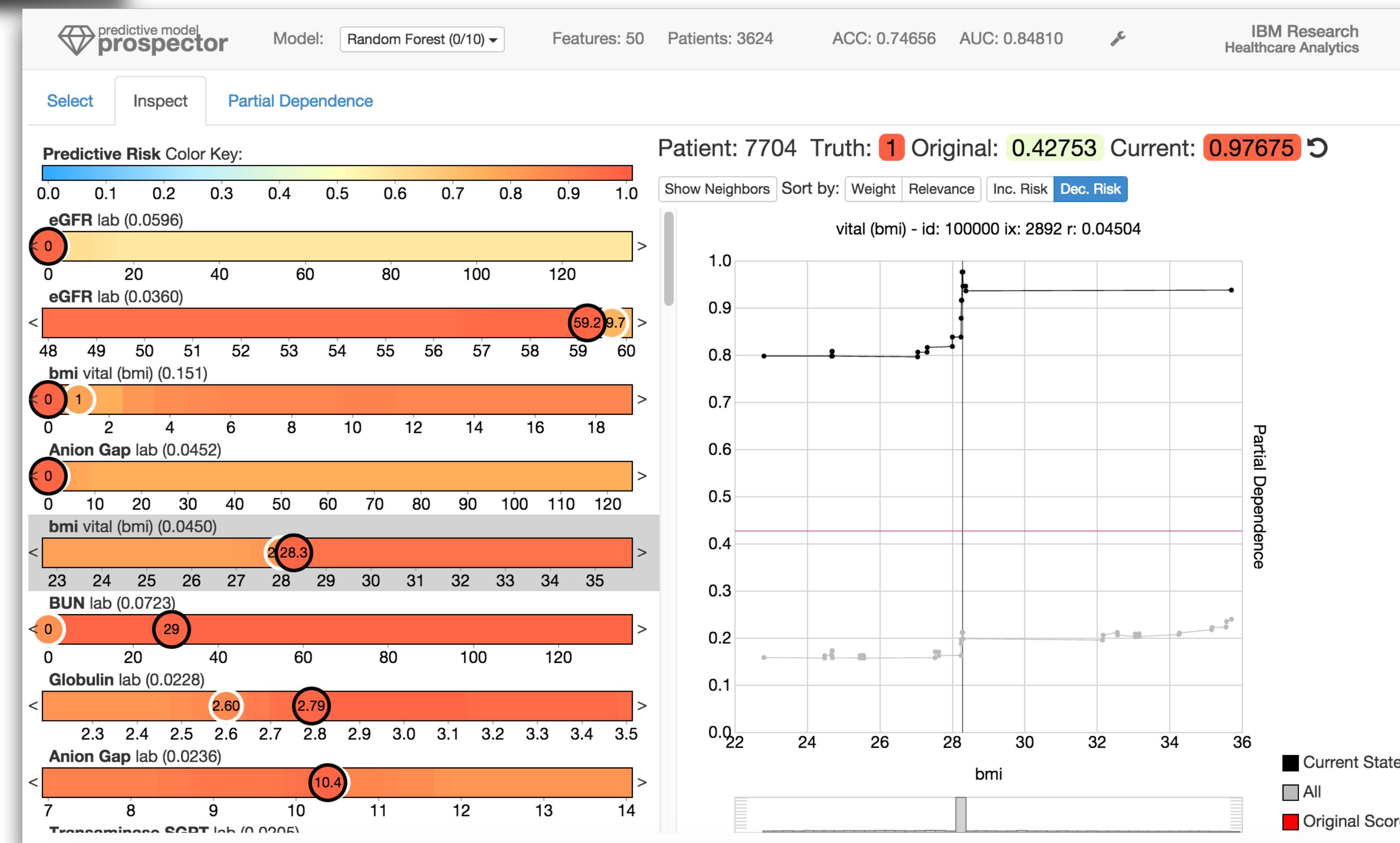
# coquito lessons

- Easy and interactive query formulation lets domain experts explore the data
- Visible intermediate results provide critical feedback
- Hints for query refinements are helpful in improving queries



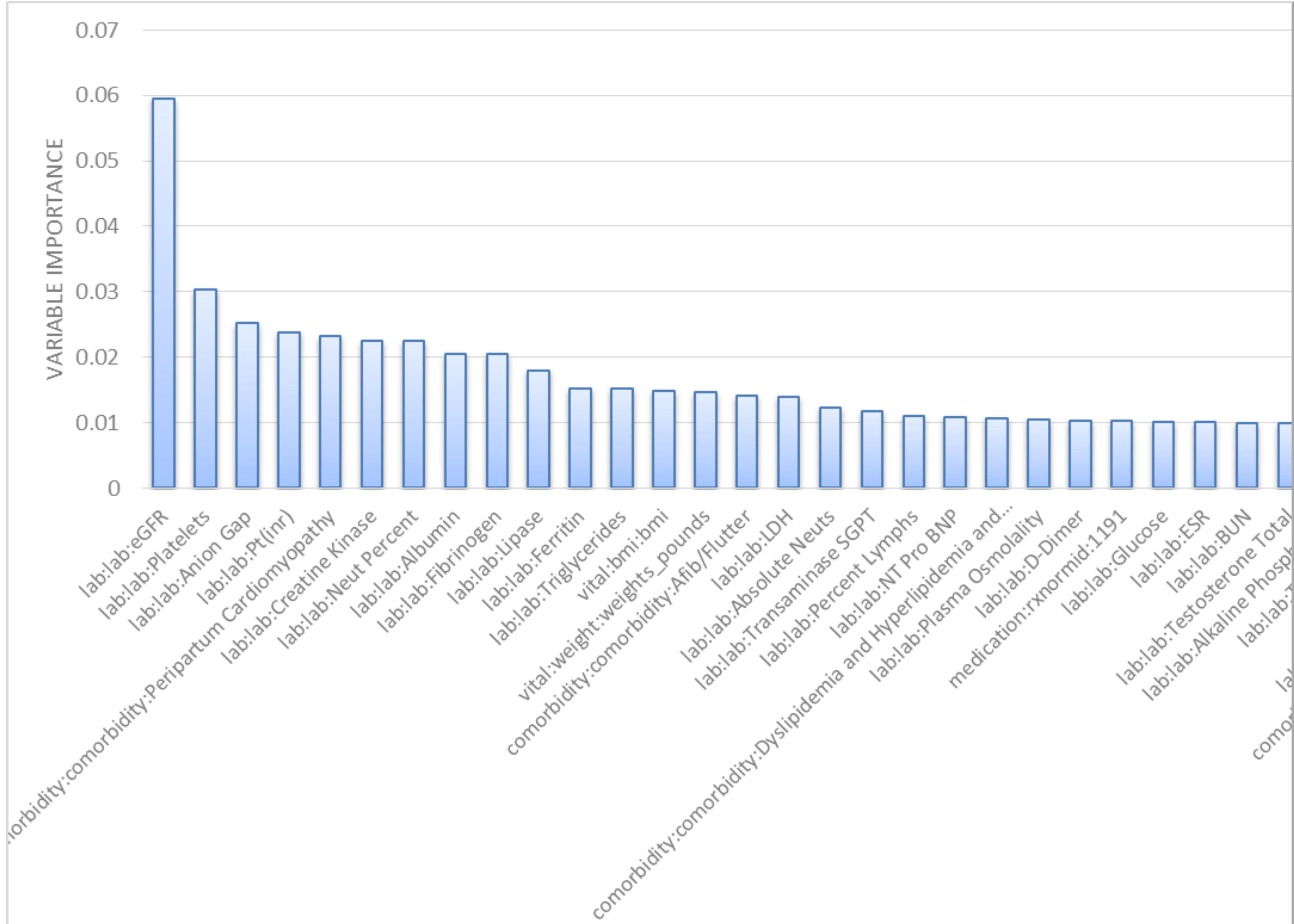
# Model Interpretability

# predictive model prospector



Josua Krause, Adam Perer, and Kenney Ng. [Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models](#). ACM Conference on Human Factors in Computing Systems (CHI 2016). San Jose, California. (2016).

# typical predictive model report



Typically simply a list of top features and their weights

## Why?

Difficult to summarize complex models

## Issues

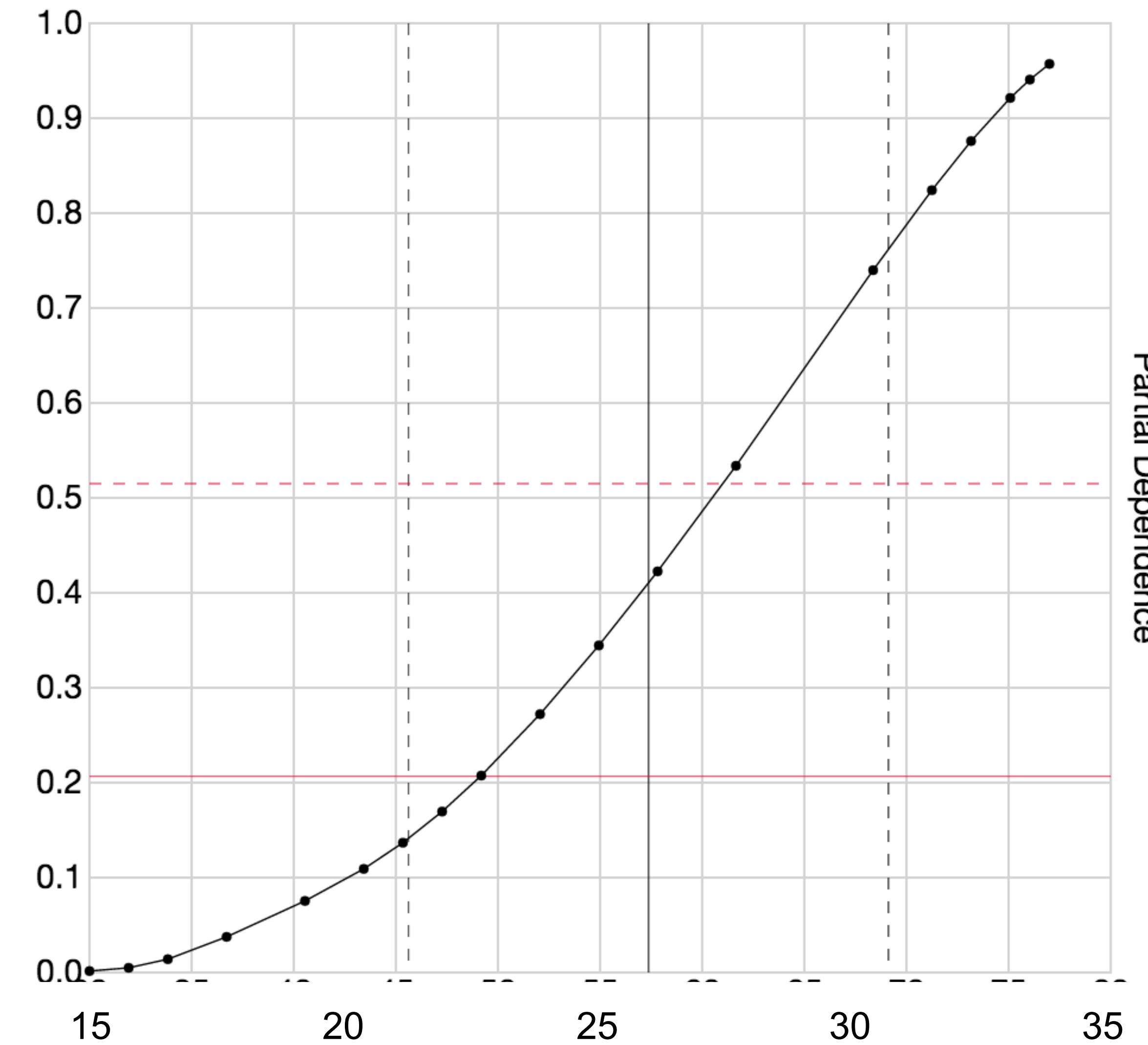
One cannot interpret how the values of each feature impact the prediction

One cannot interact with the model to test hypotheses

# partial dependence

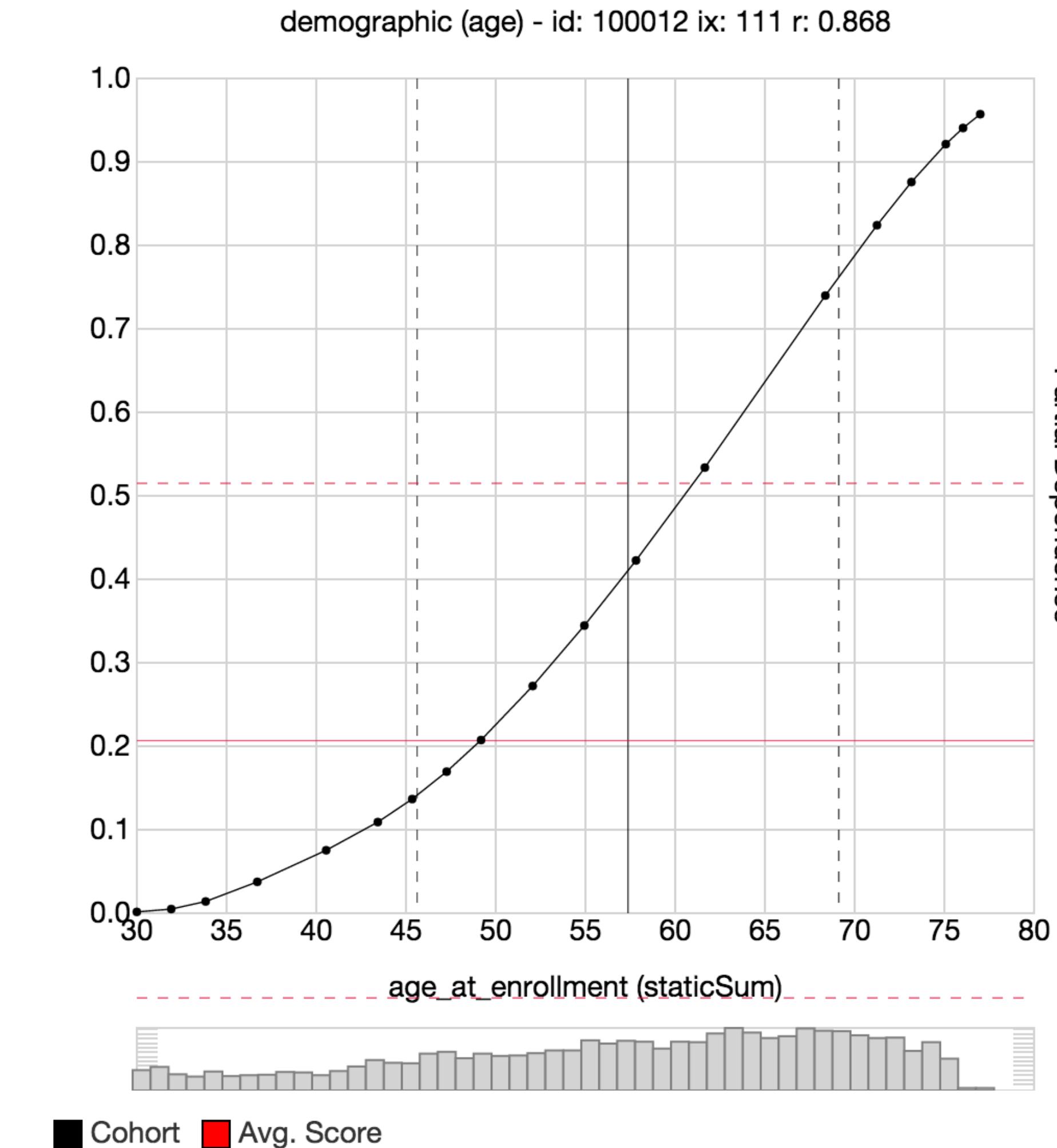


# partial dependence

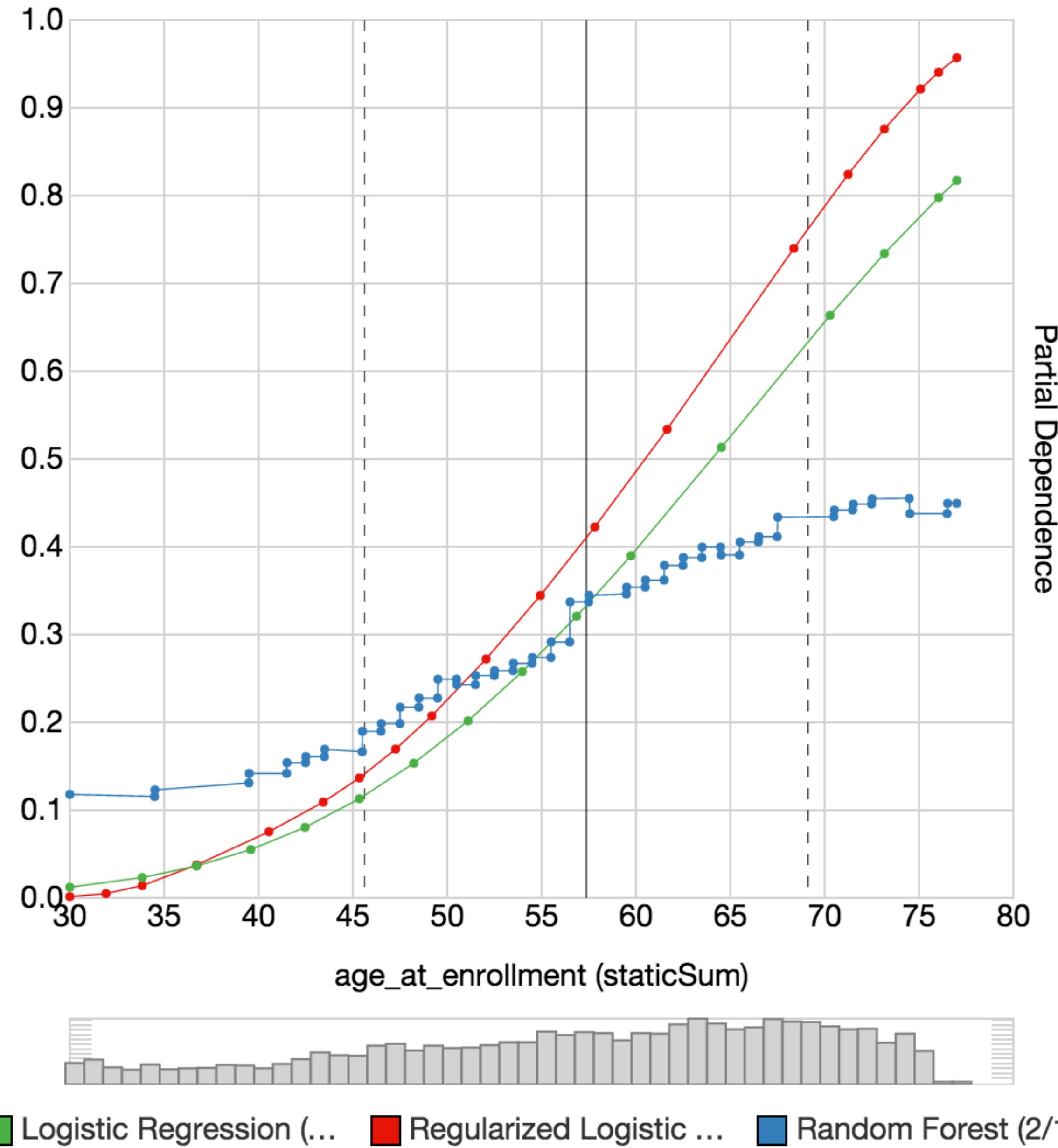


Partial Dependence

# partial dependence



demographic (age) - id: 100012 ix: 111



# localized inspection



diabetes  
diagnoses

7

bmi

22

glucose  
level

160

teeth

N

eyebrows

N

**prediction**

**0.95**

# localized inspection

diabetes  
diagnoses

bmi

glucose  
level

teeth

eyebrows

**prediction**



7

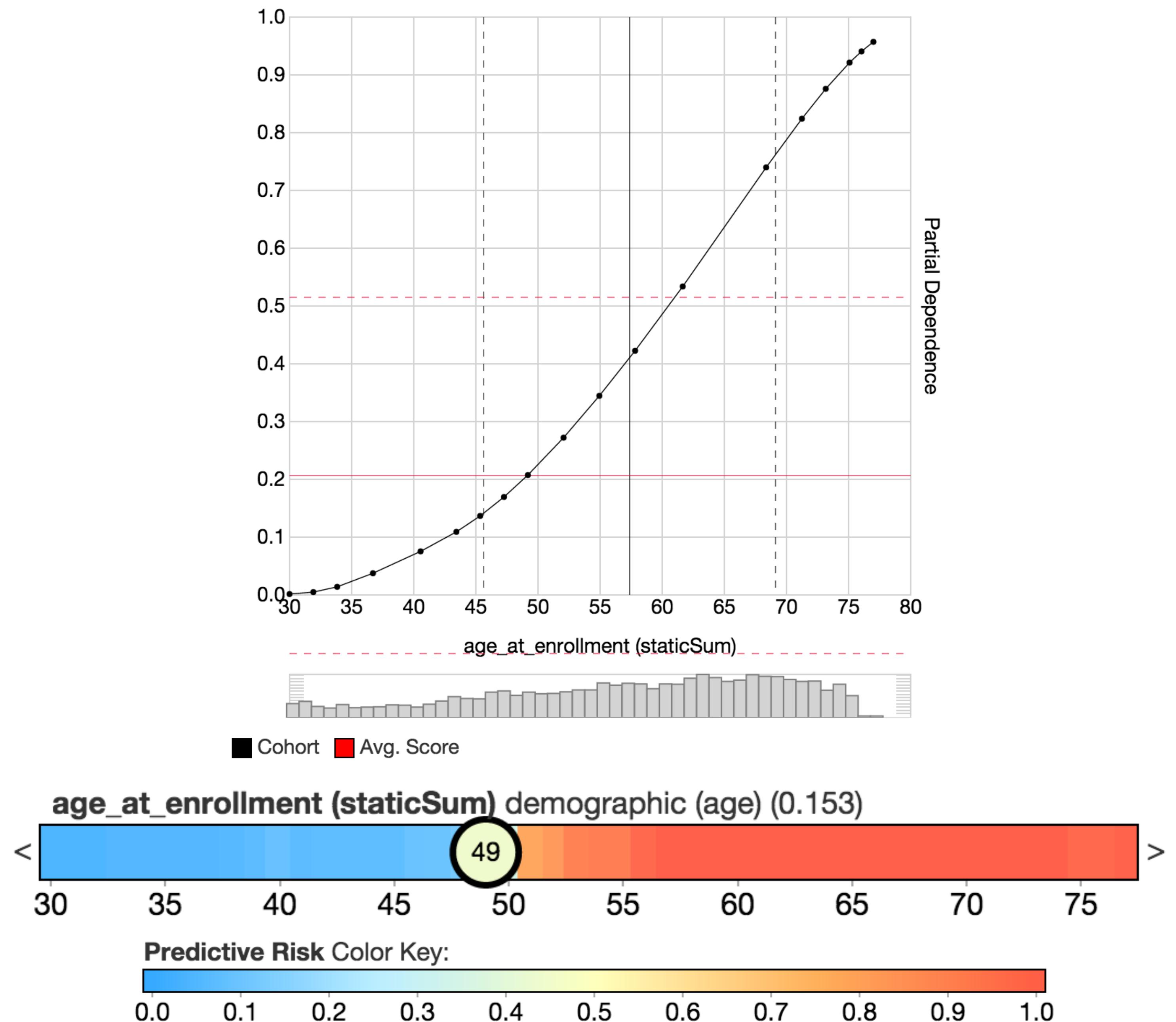
20

140

N

N

0.45



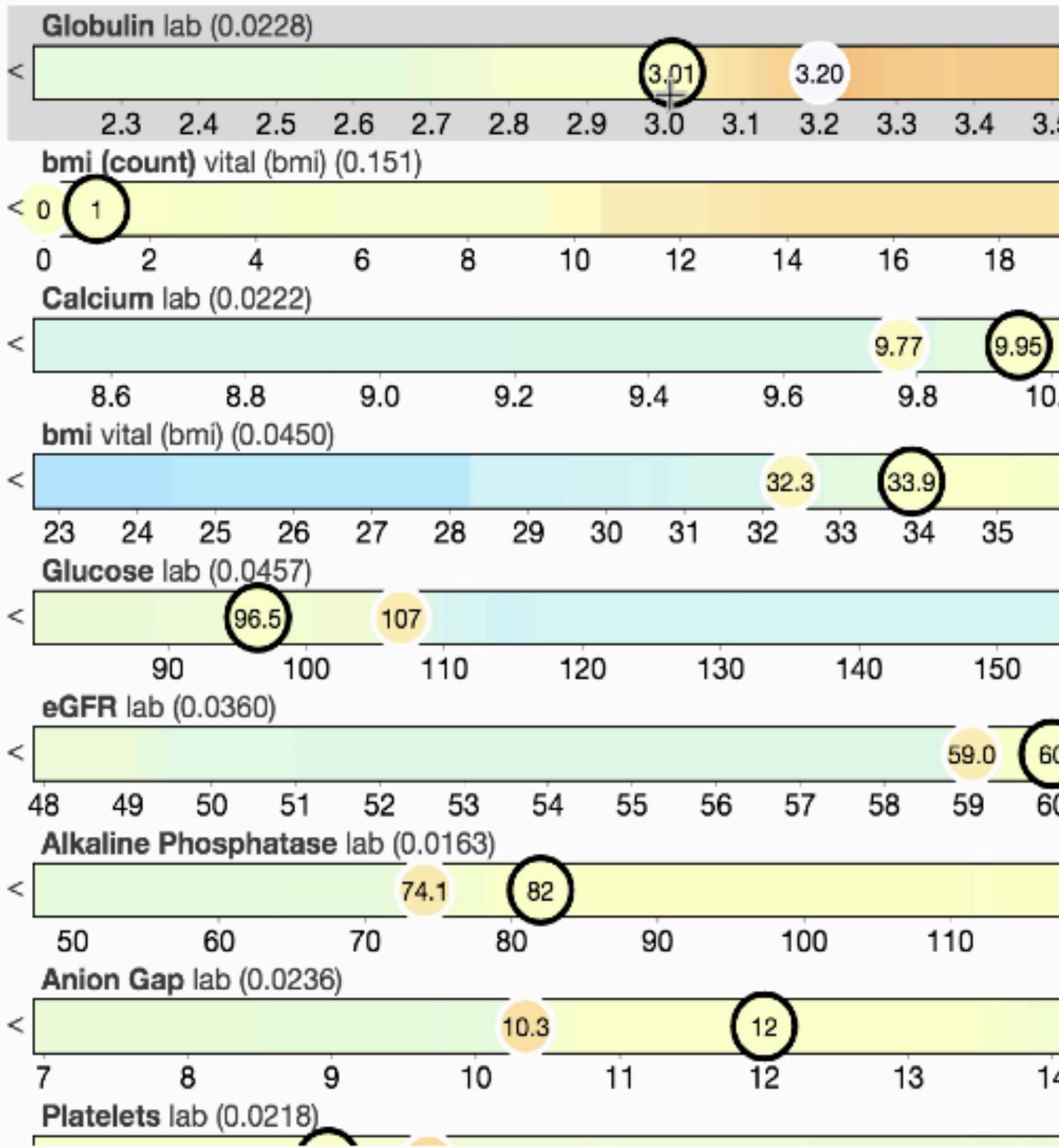
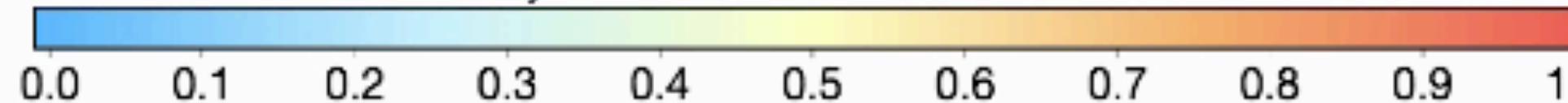
- **predicting onset of diabetes for 4000 patients**
- **4 month long term case study with 5 data scientists**
- **stories of visualization-driven insights in the paper**

Select

Inspect

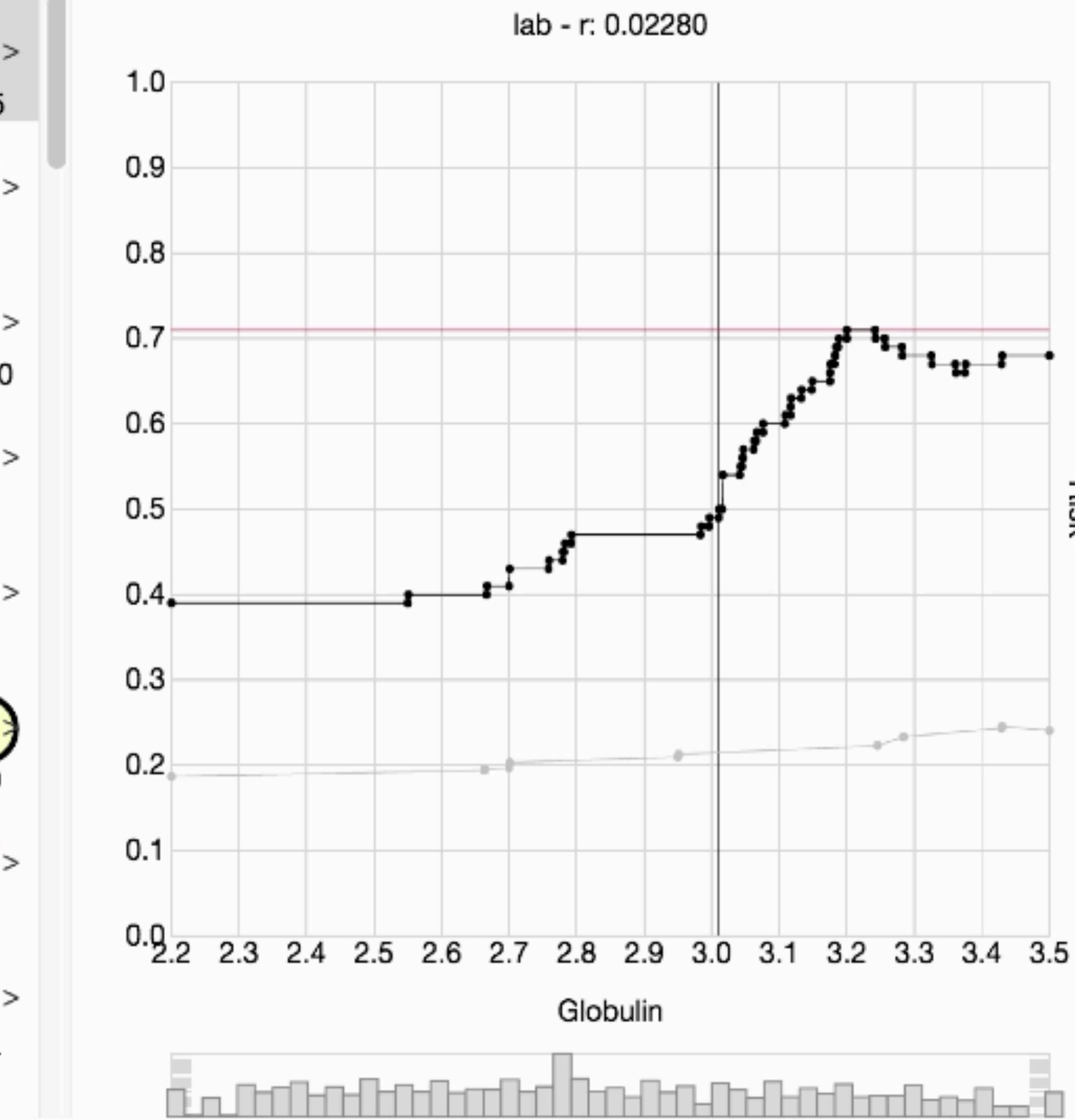
Partial Dependence

Predictive Risk Color Key:



Patient: 5754 Truth: 1 Original: 0.71000 Current: 0.49000 ⚡

Show Neighbors Sort by: Weight Relevance Inc. Risk Dec. Risk

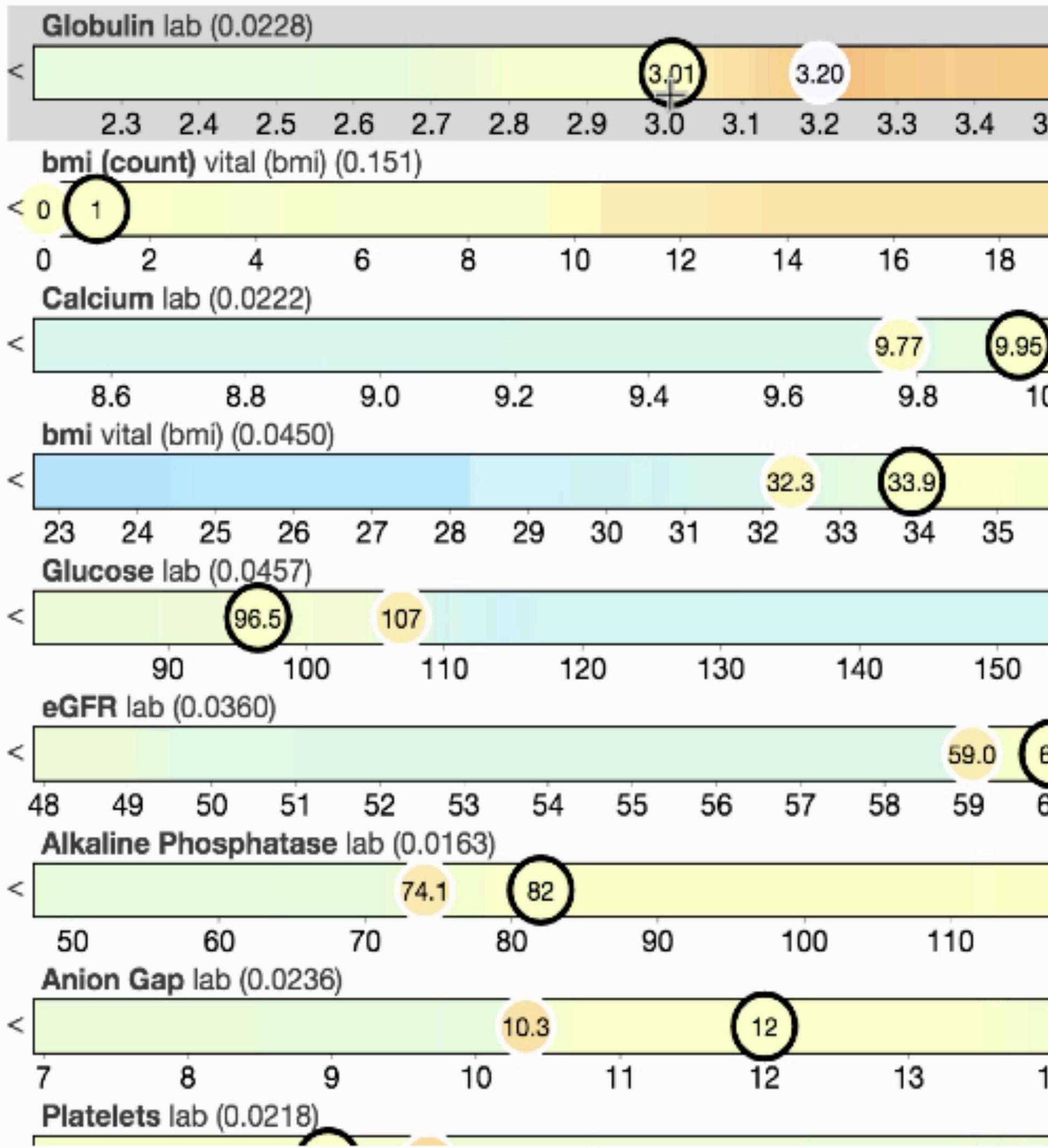
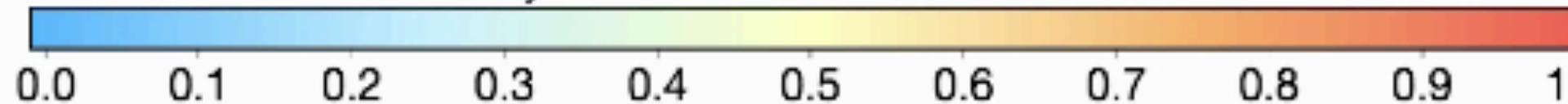


Select

Inspect

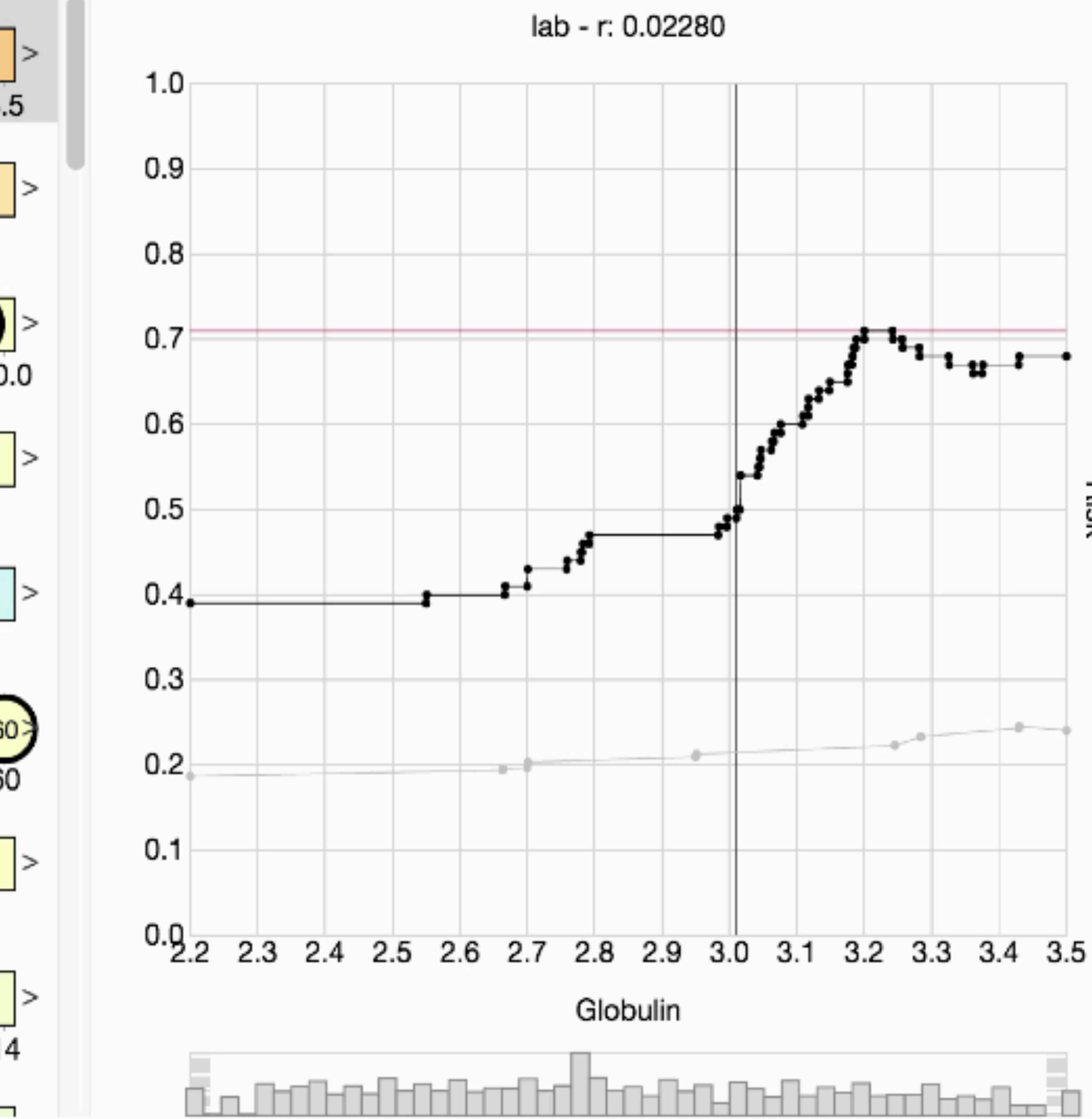
Partial Dependence

Predictive Risk Color Key:



Patient: 5754 Truth: 1 Original: 0.71000 Current: 0.49000 ⚡

Show Neighbors Sort by: Weight Relevance Inc. Risk Dec. Risk



# take-aways

Clinical Data is complex and messy.

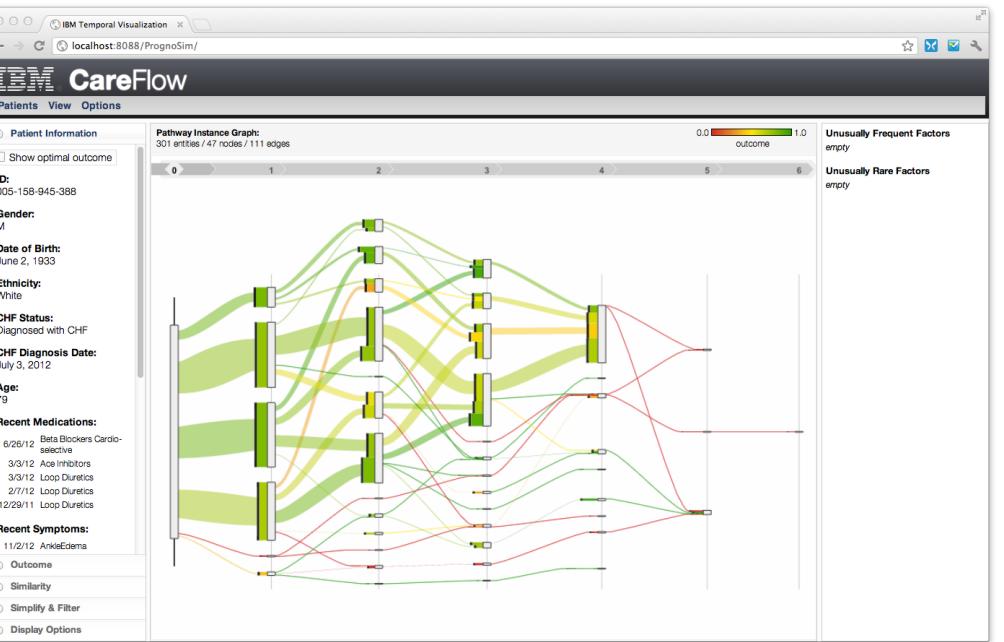
Exploratory visual analytics tools fill a much needed gap.

However, exploratory tools alone do not address their predictive desires.

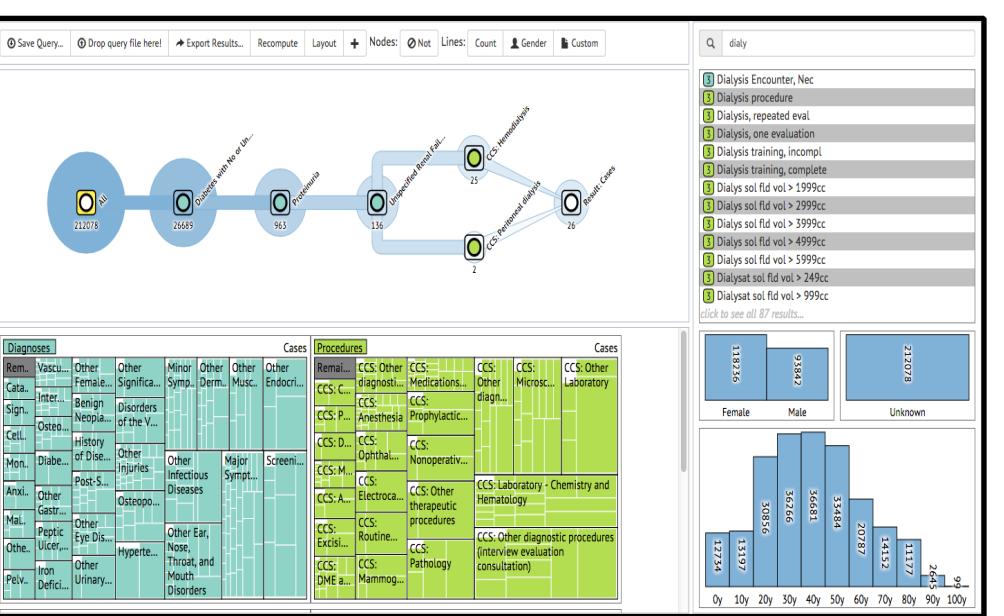
There is a strong role for visualization in predictive tasks.

Adam Perer

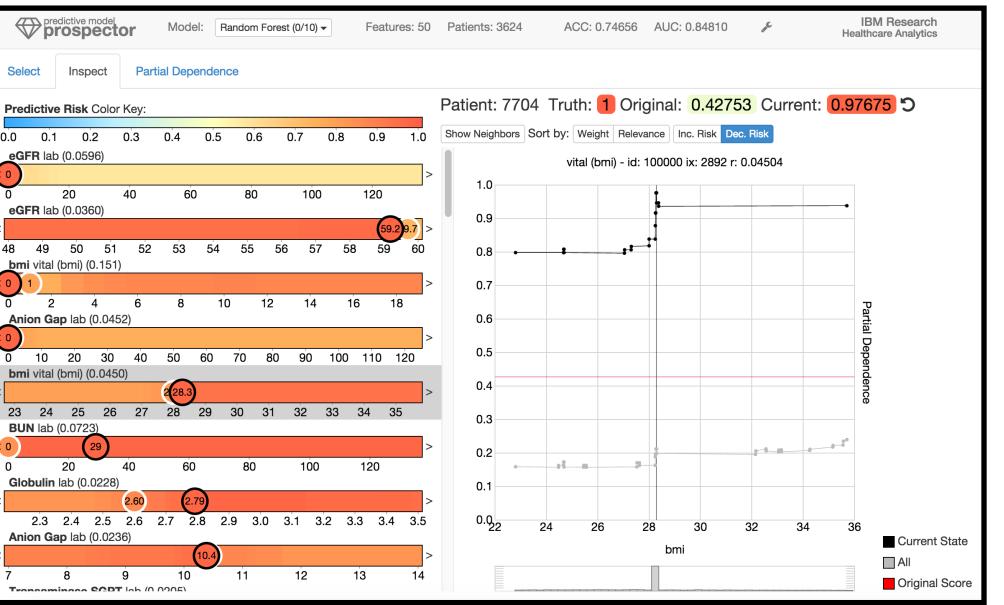
[ papers and videos at <http://perer.org> ]



CareFlow (CHI 2013)



COQUITO (VAST 2015)



Prospector (CHI 2016)