Predicting, and preventing cost-blooms

Nigam Shah, MBBS, PhD

nigam@stanford.edu

STANFORD
SCHOOL OF MEDICINE

Healthcare in the United States

- What is the system for?
- Who are the key players, what are their roles, and what are their interests?
- How does the system function economically?
- What are the trends, failures, and opportunities?
- How, where and why, are data produced?

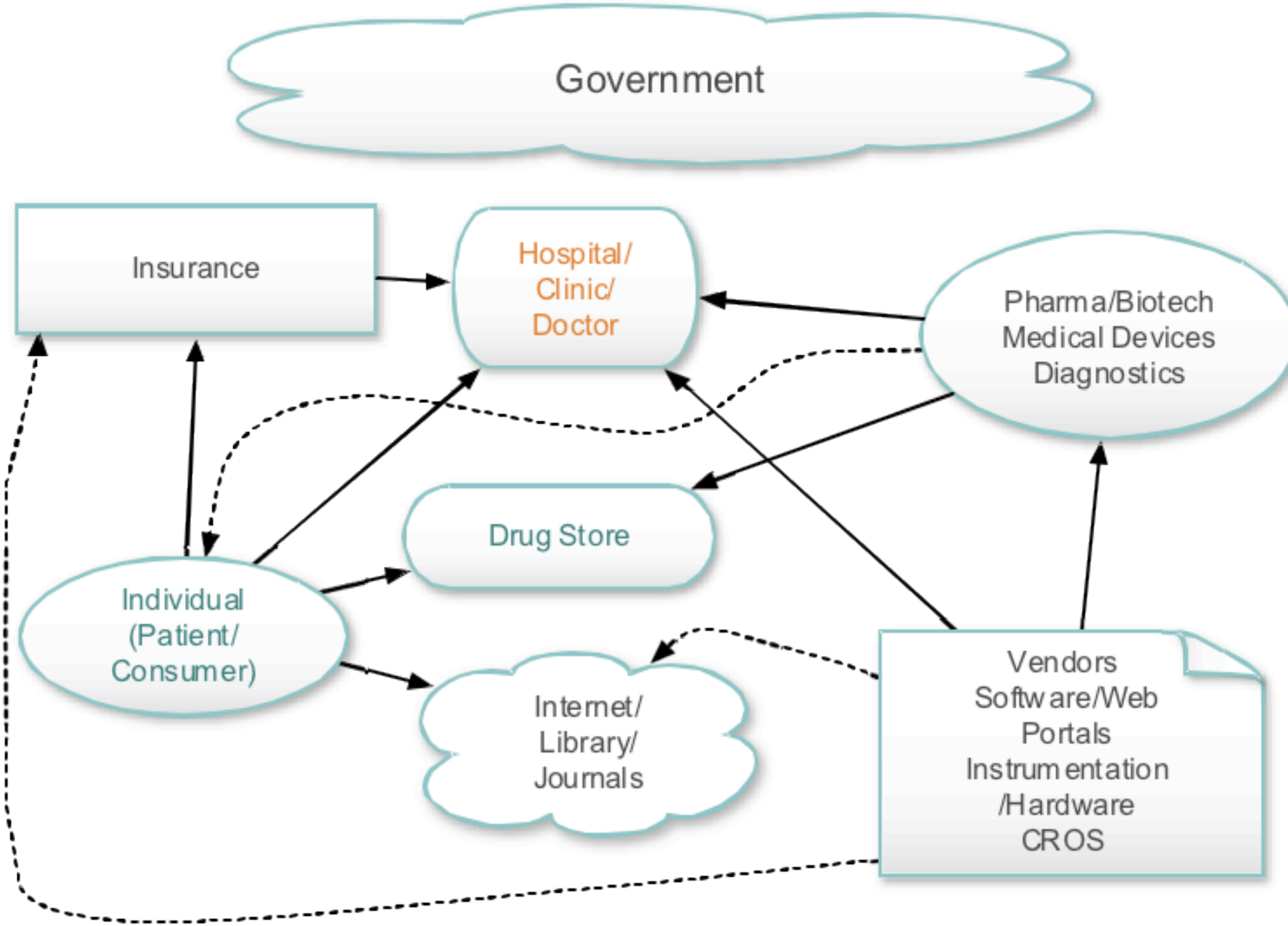

Government
Insurance
Hospital/
Clinic/
Doctor
Pharma/Biotech
Medical Devices
Diagnostics
Drug Store
Individual
(Patient/
Consumer)
Internet/
Library/
Journals
Vendors
Software/Web
Portals
Instrumentation
/Hardware
CROS

Table 2. Number of US Facilities in Health Care Sectors, 2000-2010

Subsector	No. of Facilities in Health Care Sectors		
	2000	2010	Annual Growth Rate, 2000-2010, %[a]
Offices of physicians	195 655	223 797	1.4
Social assistance	129 053	158 764	2.1
Offices of dentists	116 494	129 830	1.1
Nursing and residential care	63 005	79 047	2.3
Pharmacies and drug stores	40 614	41 672	0.3
Home health care services	16 092	27 314	5.4
Outpatient care centers	19 700	27 202	3.3
Medical and diagnostic laboratories	9750	13 220	3.1
General hospitals[b]	6588	5836	−1.2
Urgent care centers	2503	5419	8.0
Retail clinics[c]	3	1200	82.1
Specialty hospitals	499	956	6.7
All others	165 773	221 615	2.9
Total	**765 729**	**935 872**	2.0

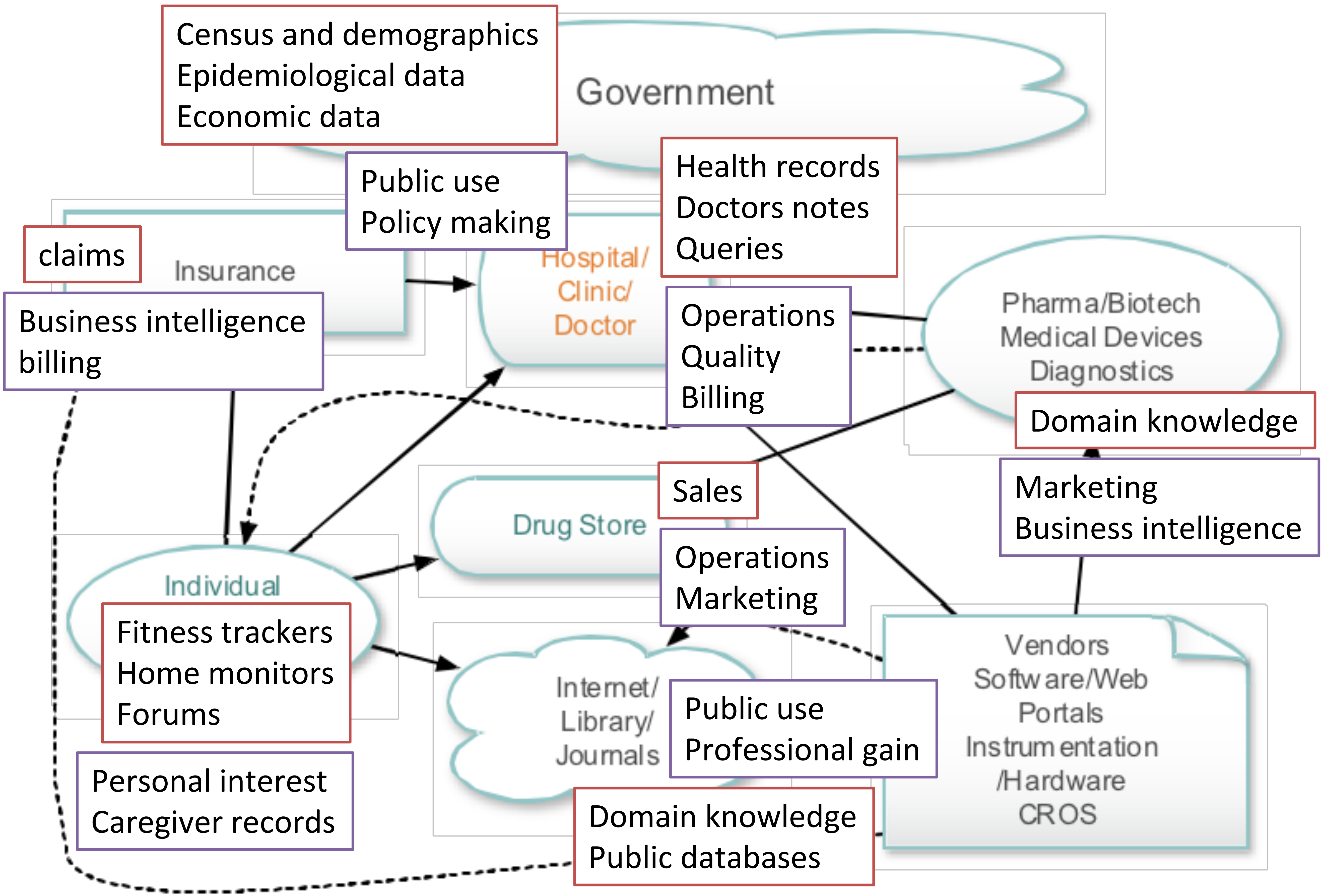
Anatomy of the US Healthcare System

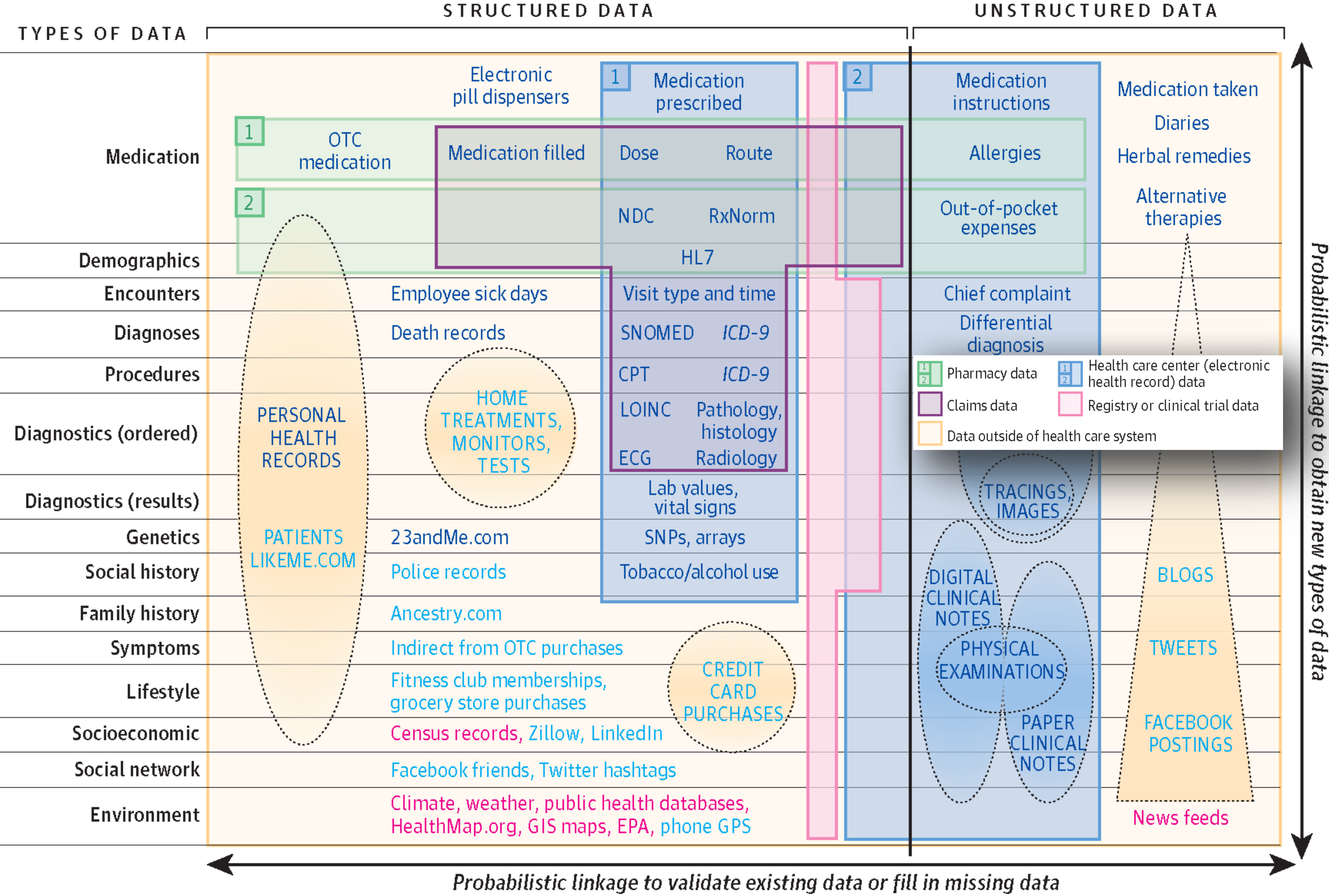
Take a minute to think, then work with your neighbor to answer the following question on your concept map:

What are the kinds of data that each of these entities generate? For what purpose?

Example: individual patients generate fitness tracker data for their own personal interest

Where and why are the data generated?





Weber et al, JAMA 2014

The Anatomy of Health Care in the United States

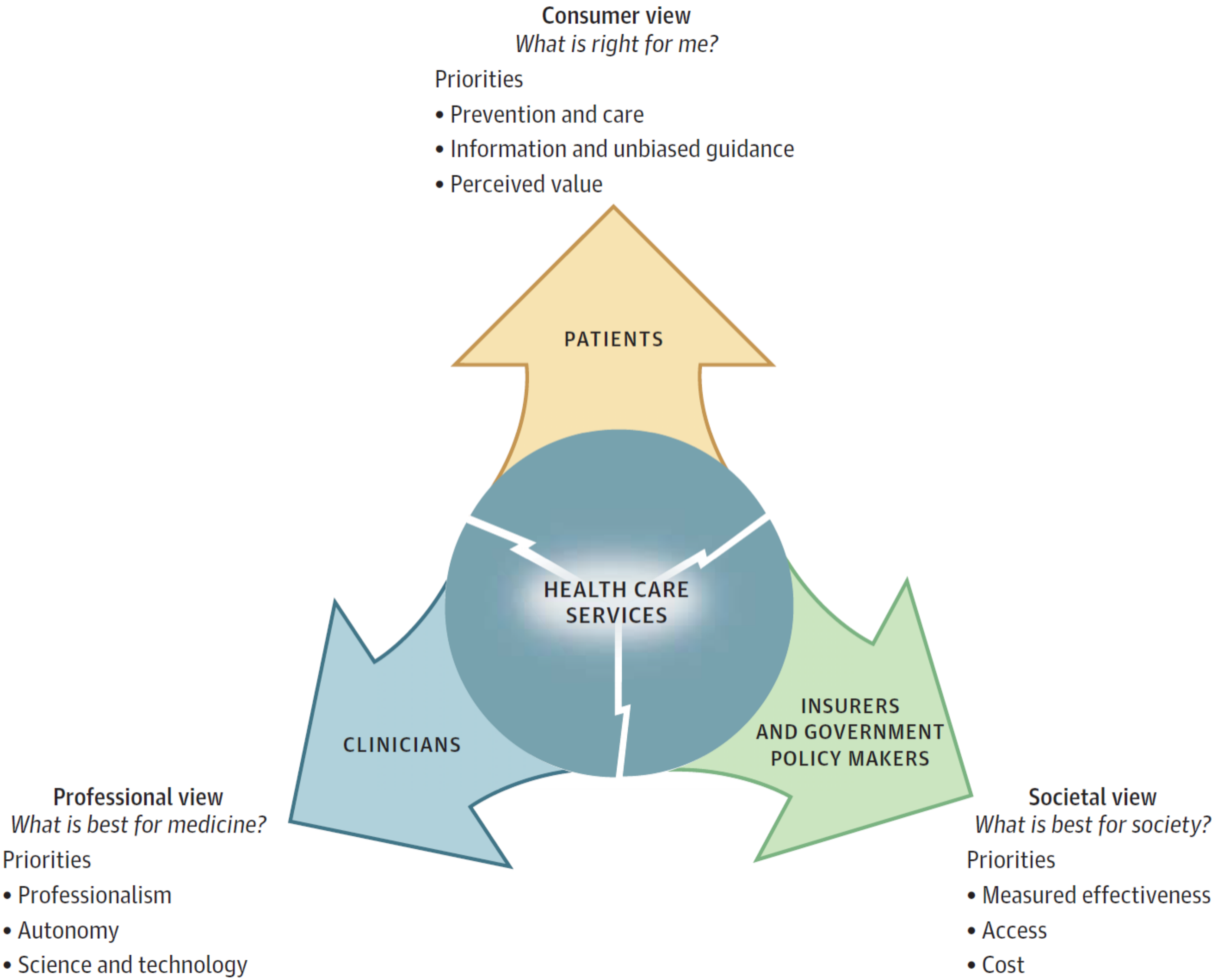
Hamilton Moses III, MD; David H. M. Matheson, MBA, JD; E. Ray Dorsey, MD, MBA; Benjamin P. George, MPH; David Sadoff, BA; Satoshi Yoshimura, PhD

- Publicly available data from 1980 to 2011, on the source and use of funds.
- In 2011, US health care employed 15.7% of the workforce, with expenditures of $2.7 trillion, and being 17.9% of GDP.

- Three factors have produced the most change:
  - consolidation, producing financial concentration
  - information technology, in which investment has occurred but value is elusive;
  - patient empowerment, whereby influence is sought outside traditional channels.

Follow the money … it will lead you to the problems that really need to be solved

Conflicting interests



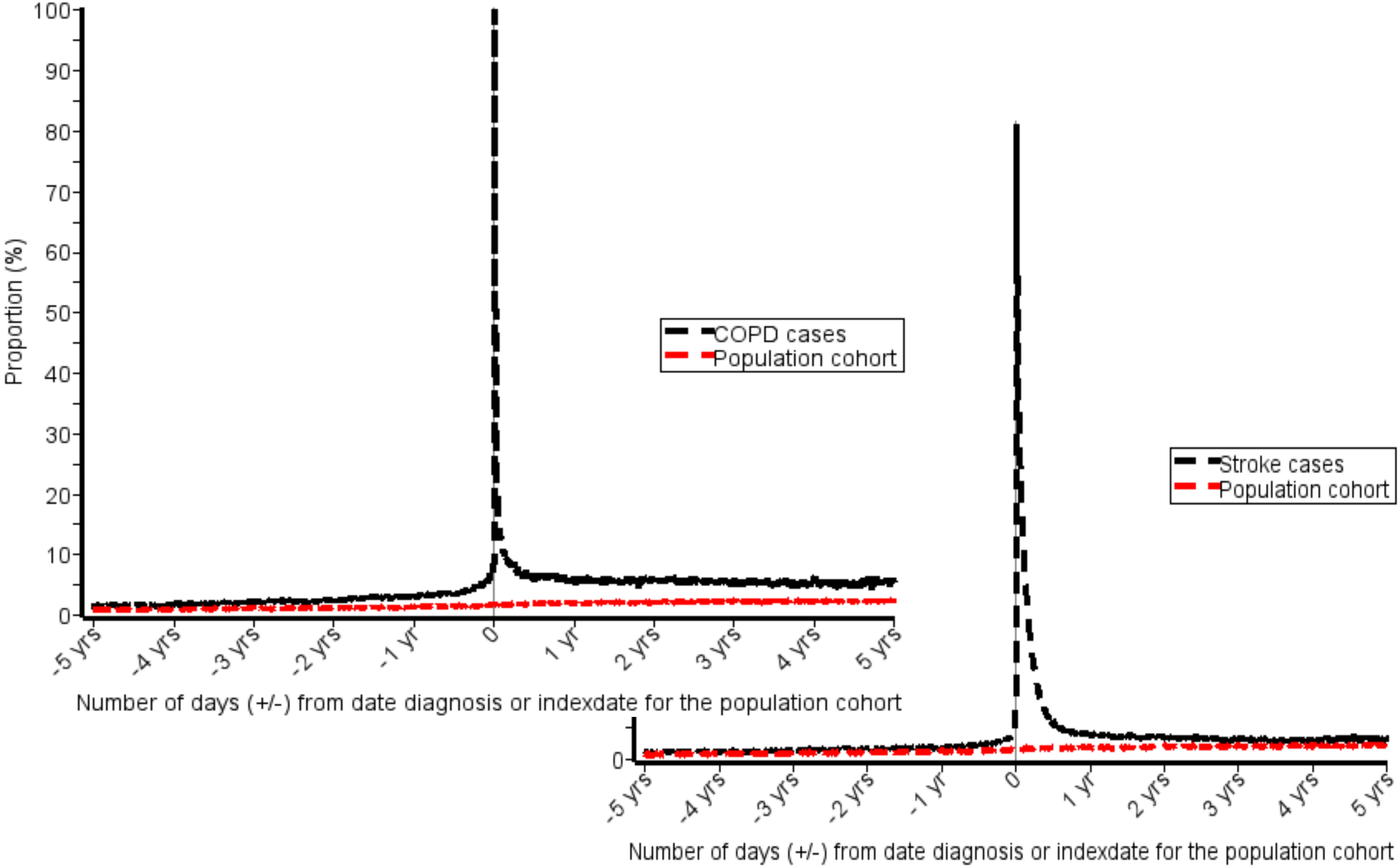
When you use these data:

- Know that priorities are different for each stakeholder, which affects the data that are generated.
- Design studies to leverage strengths and protect from weaknesses of the data. Using multiple sources is beneficial.
- Think about who is interested in the results. Targeting studies to the intersections of two or more interests is impactful.

Why predict cost?

- For “risk-adjustment”
  - Risk assessment → measuring the expected healthcare costs of individuals enrolled in a plan.
  - Risk adjustment → moving funds from plans that have less than their fair-share of high-risk enrollees to plans that have more high-risk enrollees.
- For “risk-contracting”
  - In a fee for performance model, where the provider is assuming total risk for caring for an individual, they need to know their risk exposure.
- For deciding which insurance to buy
  - As an individual, knowing your true risk allows you to buy the appropriate plan with adequate coverage.
    - E.g. should you enroll in a high deductible plan or not?

Cost at the population level

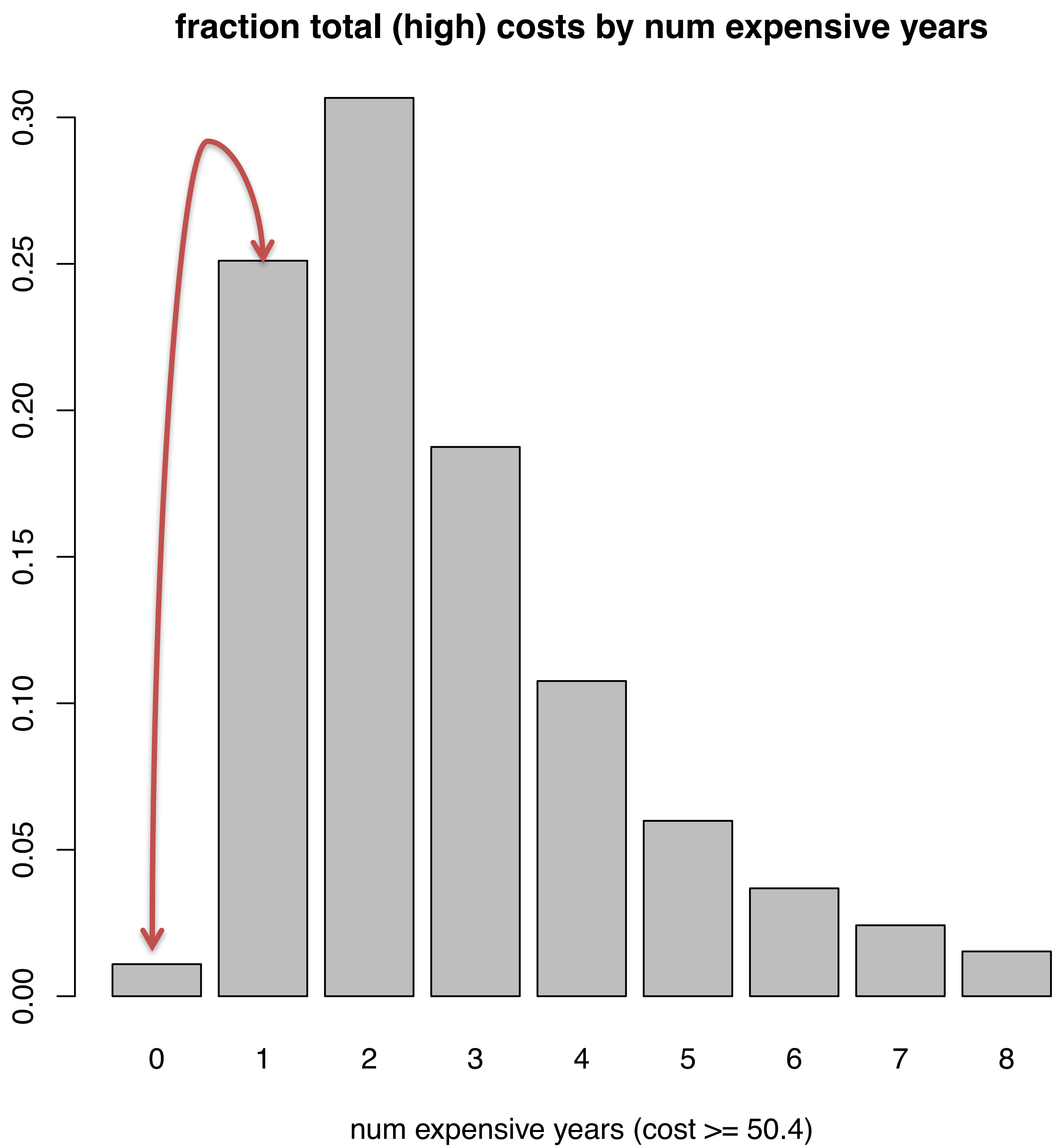


What is worth predicting?

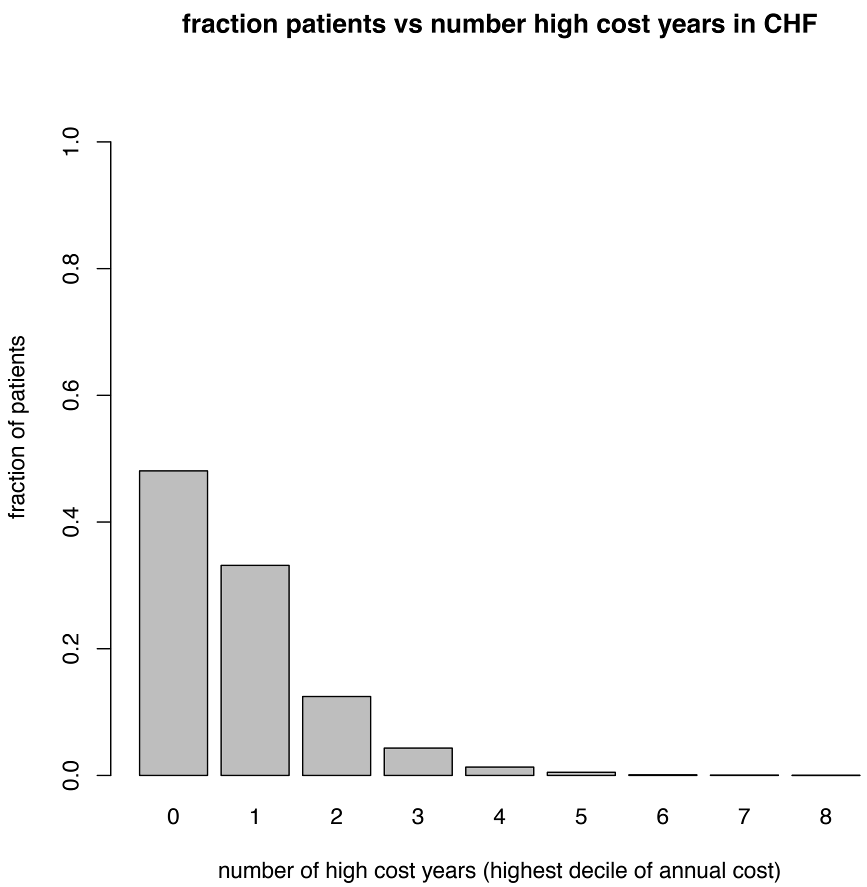
- If you have a high cost year, what is the probability that the next year is high cost?
  - 0.26 overall
  - 0.37 in high cost population
  - 0.03 in low cost population → If they become high-cost, it’s an unexpected event

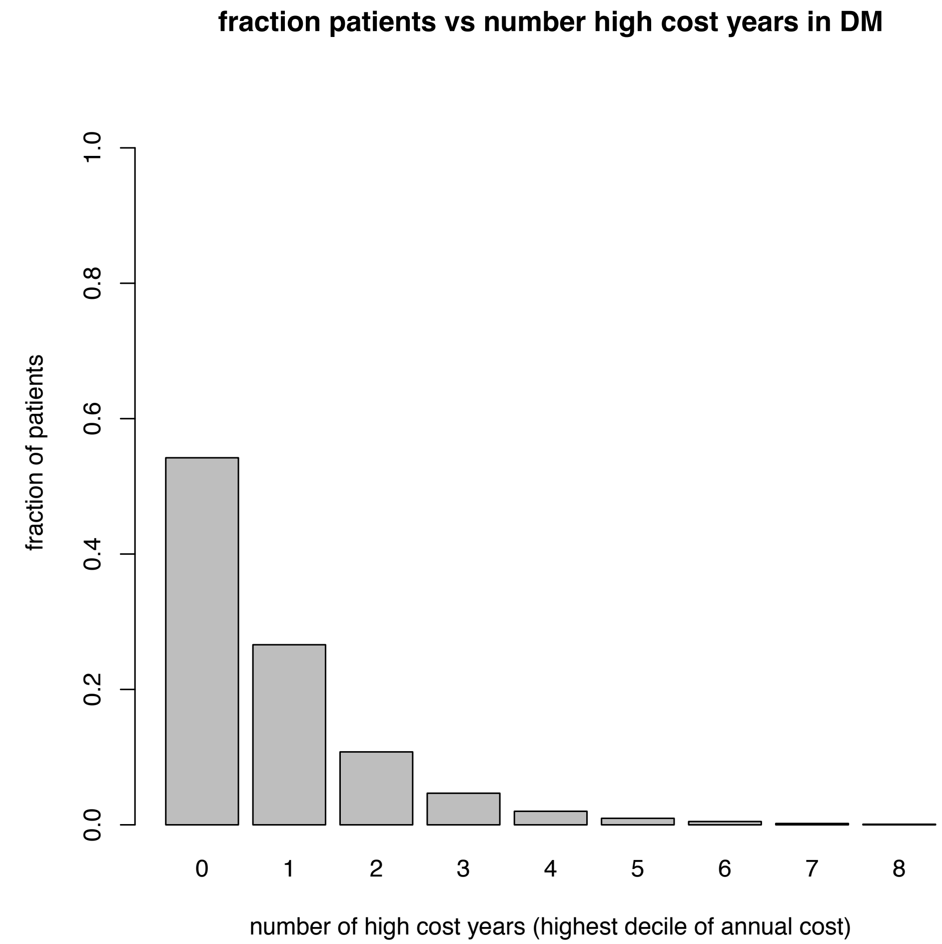
- High Cost vs. a Cost bloom

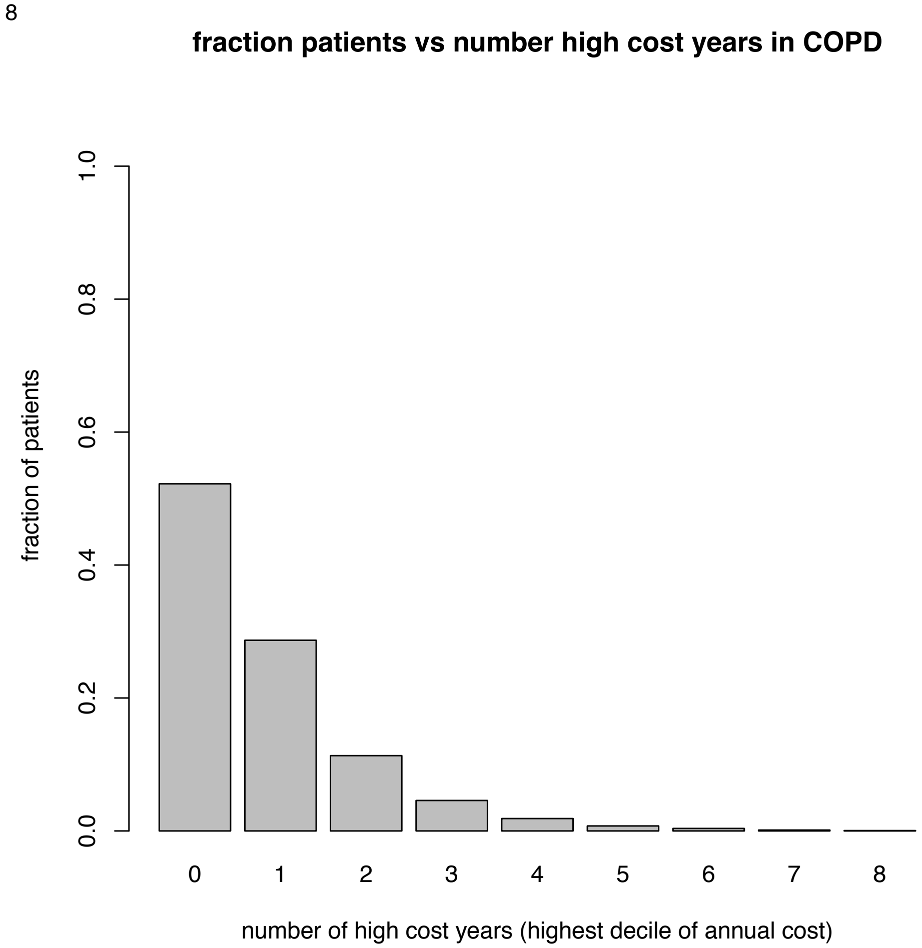
Anatomy of “high cost”



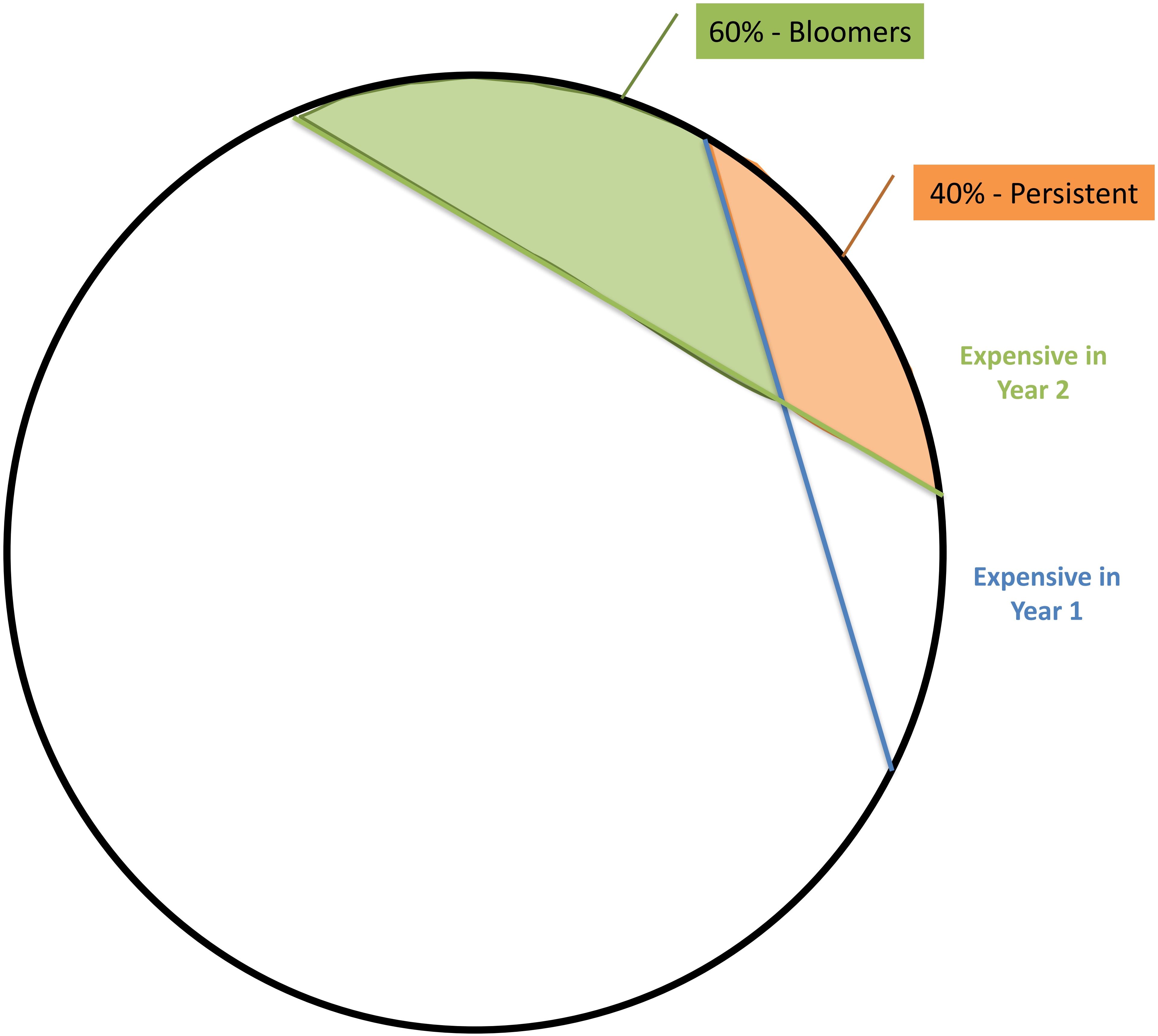
Anatomy of "high cost"



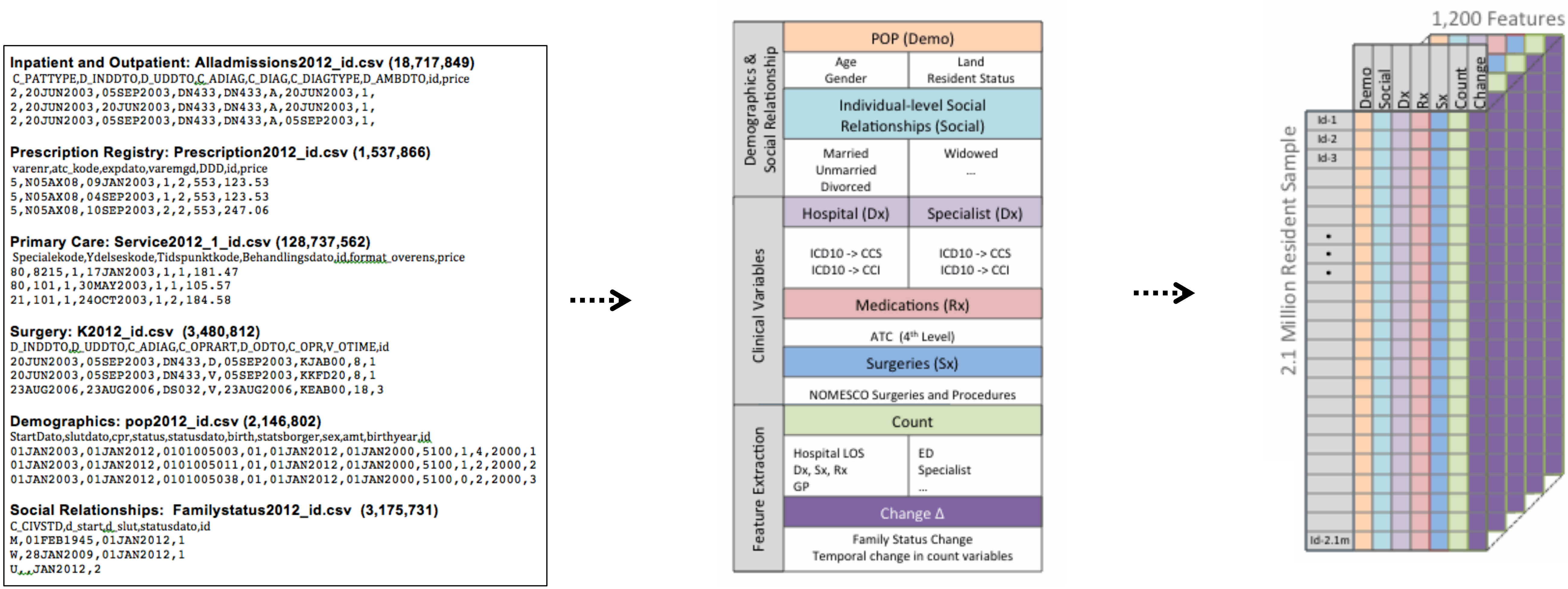


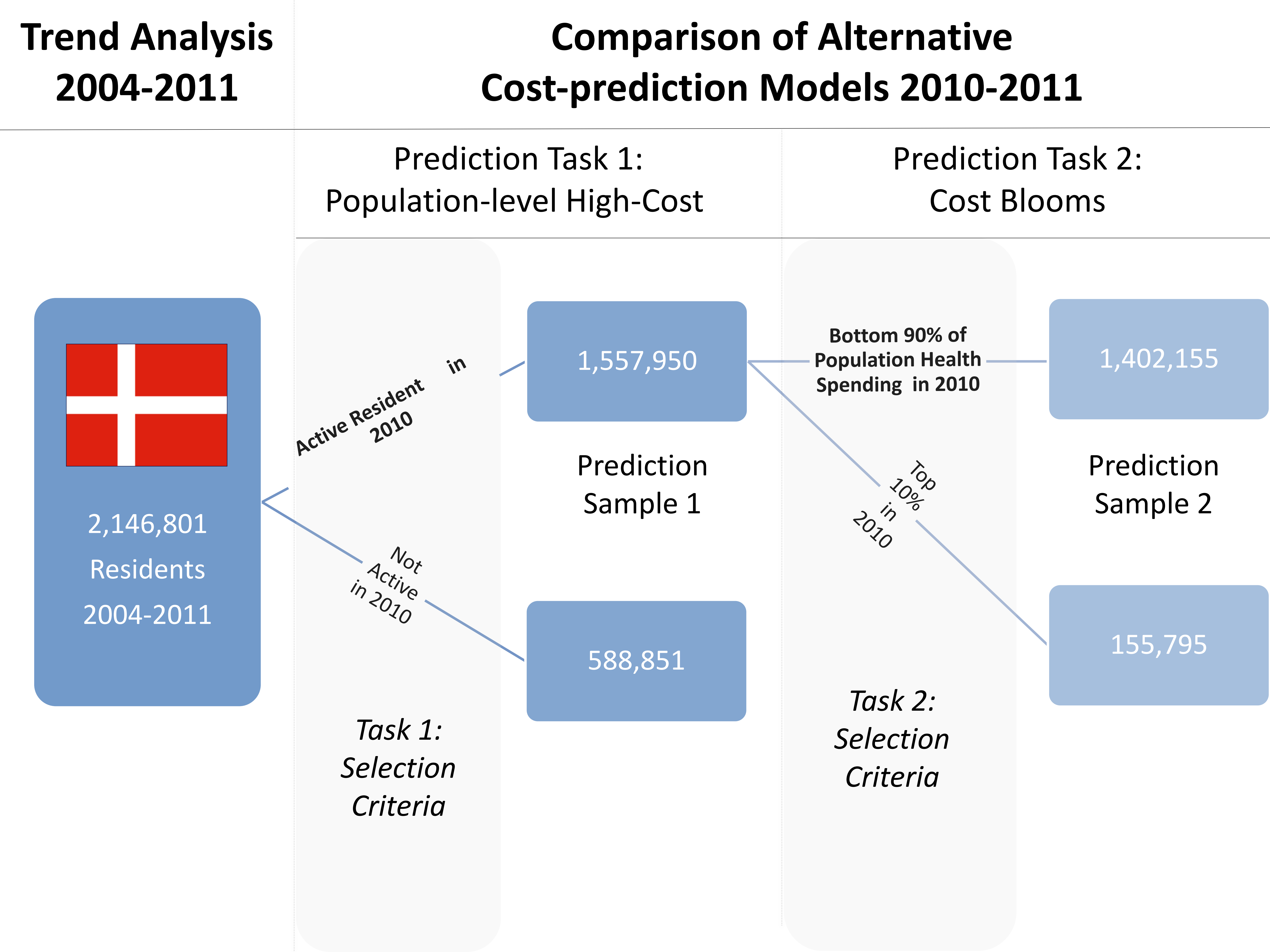


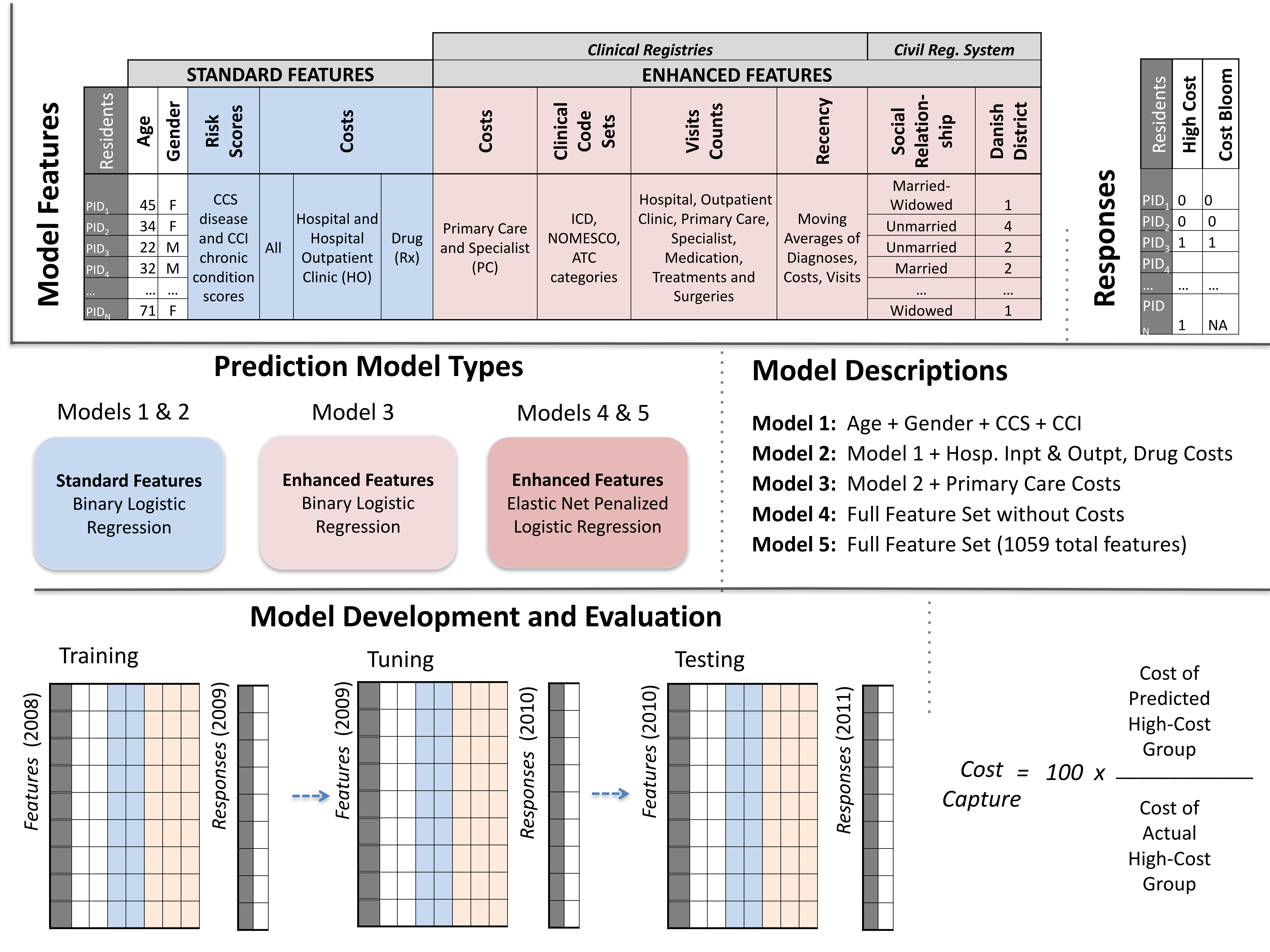
Anatomy of the cost



Predicting cost vs. cost bloom



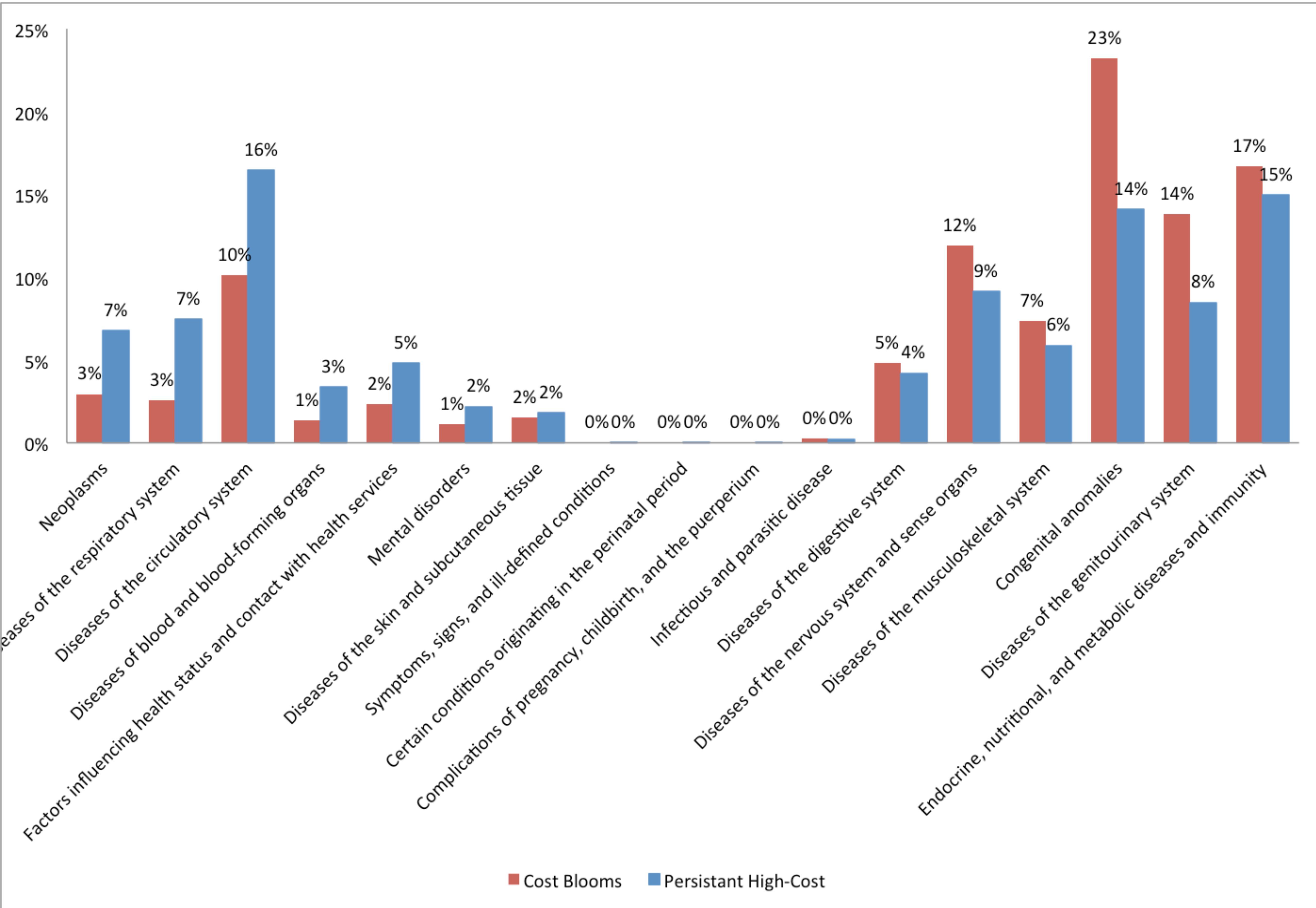

Trend Analysis 2004-2011
Comparison of Alternative Cost-prediction Models 2010-2011
Prediction Task 1: Population-level High-Cost
Prediction Task 2: Cost Blooms
2,146,801 Residents 2004-2011
Active Resident in 2010
Not Active in 2010
1,557,950
588,851
Prediction Sample 1
Task 1: Selection Criteria
Bottom 90% of Population Health Spending in 2010
Top 10% in 2010
1,402,155
155,795
Prediction Sample 2
Task 2: Selection Criteria



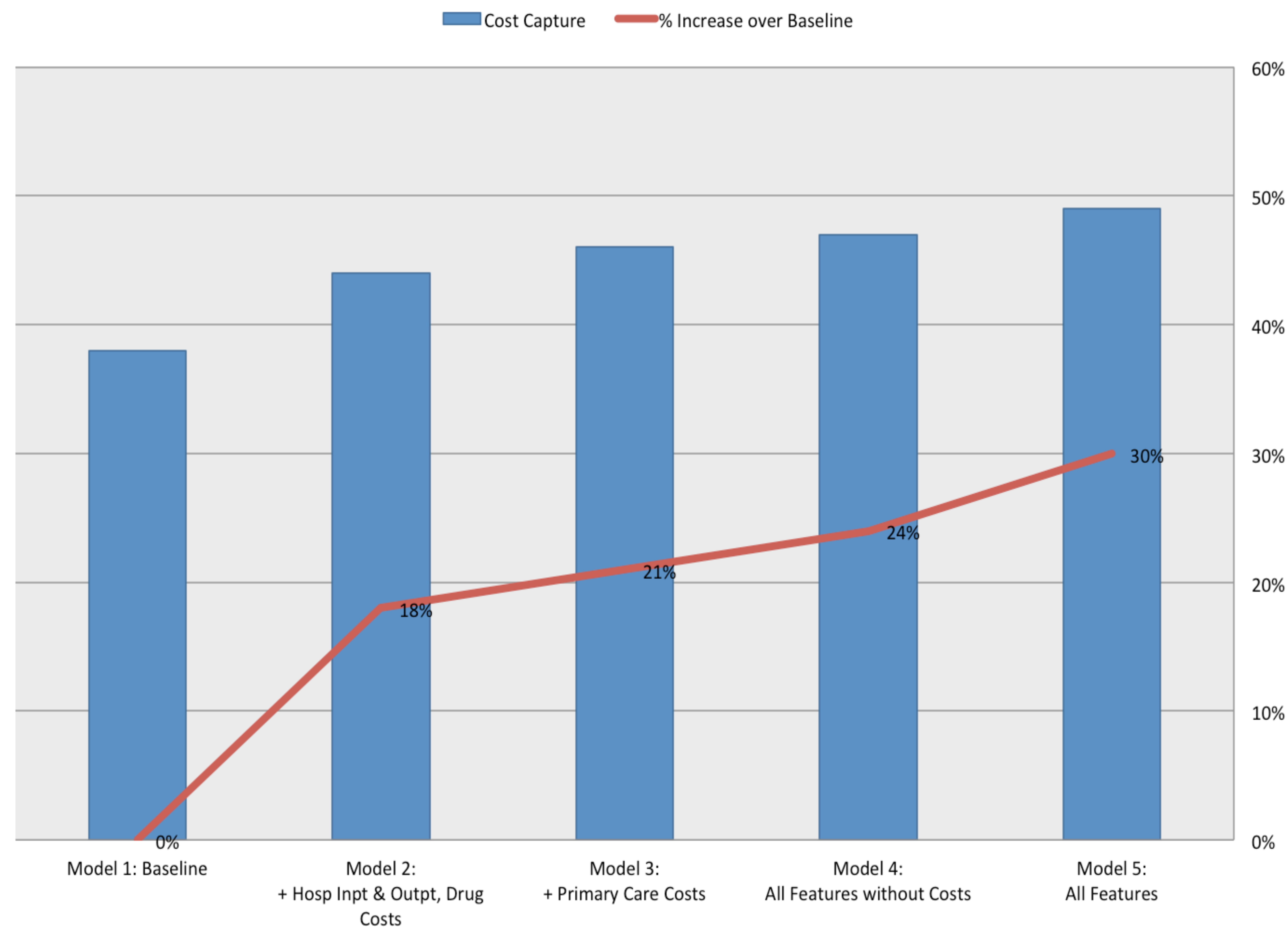
Model Features
Clinical Registries
Civil Reg. System
STANDARD FEATURES
ENHANCED FEATURES
Residents
Age
Gender
Risk Scores
Costs
Costs
Clinical Code Sets
Visits Counts
Recency
Social Relation-ship
Danish District
PID1 45 F
PID2 34 F
PID3 22 M
PID4 32 M
... ...
PIDN 71 F
CCS disease and CCI chronic condition scores
All
Hospital and Hospital Outpatient Clinic (HO)
Drug (Rx)
Primary Care and Specialist (PC)
ICD, NOMESCO, ATC categories
Hospital, Outpatient Clinic, Primary Care, Specialist, Medication, Treatments and Surgeries
Moving Averages of Diagnoses, Costs, Visits
Married-Widowed 1
Unmarried 4
Unmarried 2
Married 2
... ...
Widowed 1
Responses
Residents
High Cost
Cost Bloom
PID1 0 0
PID2 0 0
PID3 1 1
PID4
... ...
PIDN 1 NA
Prediction Model Types
Models 1 & 2
Standard Features
Binary Logistic Regression
Model 3
Enhanced Features
Binary Logistic Regression
Models 4 & 5
Enhanced Features
Elastic Net Penalized Logistic Regression
Model Descriptions
Model 1: Age + Gender + CCS + CCI
Model 2: Model 1 + Hosp. Inpt & Outpt, Drug Costs
Model 3: Model 2 + Primary Care Costs
Model 4: Full Feature Set without Costs
Model 5: Full Feature Set (1059 total features)
Model Development and Evaluation
Training
Features (2008)
Responses (2009)
Tuning
Features (2009)
Responses (2010)
Testing
Features (2010)
Responses (2011)
Cost Capture = 100 x Cost of Predicted High-Cost Group / Cost of Actual High-Cost Group

Results

Prediction Task	Evaluation Metric	Model 1: Baseline
High-cost (N=1,557,950)	AUC	0.775
	Cost Capture	0.495
Cost-bloom (N=1,402,155)	AUC	0.719
	Cost Capture	0.376



25%
20%
15%
10%
5%
0%
3%
7%
3%
7%
10%
16%
1%
3%
2%
5%
1%
2%
2%
2%
0%
0%
0%
0%
0%
0%
0%
0%
5%
4%
12%
9%
7%
6%
23%
14%
14%
8%
17%
15%
Neoplasms
eases of the respiratory system
Diseases of the circulatory system
Diseases of blood and blood-forming organs
Factors influencing health status and contact with health services
Mental disorders
Diseases of the skin and subcutaneous tissue
Symptoms, signs, and ill-defined conditions
Certain conditions originating in the perinatal period
Complications of pregnancy, childbirth, and the puerperium
Infectious and parasitic disease
Diseases of the digestive system
Diseases of the nervous system and sense organs
Diseases of the musculoskeletal system
Congenital anomalies
Diseases of the genitourinary system
Endocrine, nutritional, and metabolic diseases and immunity
Cost Blooms
Persistant High-Cost


Cost Capture
% Increase over Baseline
60%
50%
40%
30%
20%
10%
0%
0%
18%
21%
24%
30%
Model 1: Baseline
Model 2:
+ Hosp Inpt & Outpt, Drug Costs
Model 3:
+ Primary Care Costs
Model 4:
All Features without Costs
Model 5:
All Features

Predictions and Actions

Take on Risk					
Service					
Intervention					
List	✔				
	Cost-bloom	Mortality	Chronic Pain	Pre-diabetes to Diabetes	Risk of Opioid abuse

Possible further work:

- Summarize the bloomers.
- Exploratory analyses to design interventions.

Possible intervention types

- **Relationship-based Interventions:** Suggest high value interventions to attending physicians, healthcare system medical directors, and/or patients.

- **Rules-based Interventions:** Where relationships with providers are insufficiently developed, alteration of plan rules governing coverage, pre-cert, provider network inclusion, provider incentives, patient incentives, formulary tiers, and/or DUR screens.

Summary

1. Important to distinguish cost-bloomers from persistent high-cost patients.

2. 30% improvement in cost capture over a standard diagnosis-based claims model.

3. Including a patient’s social relationship status, and temporal information such as the frequency and recency of healthcare events, improved prediction.

4. Predictions enables precise targeting of the subset of patients who are at the most risk of a cost bloom.

5. Example of machine learning that matters.

Tips for your predictive modeling projects

Data clean up will take about 80% of the time

- If you took a short cut here, stop.

Try simple things first

- "Deep learning" is not the right answer every time!

Ask whether:

- More data will increase performance
- More features will increase performance
- Errors from different models are correlated

Don't get fooled by AUC

- Examine precision recall, calibration, net-reclassification

Don't get attached to one model

Remember that the data are changing under you

Think about model deployment

- Ease of applying the model
- Think about the cost of taking action
- Precision @ K

Open research problems

- Handling data nonstationarity
- Local vs. Global models
- Handling unstructured data
- Outcome ascertainment (and censoring)
- Evaluation: Looking beyond discrimination (calibration, net-reclassification)
- Bridging the “last mile”

Credits

- Suzanne Tamang
- Arnold Milstein
- Alan Glaseroff
- Thomas Wang