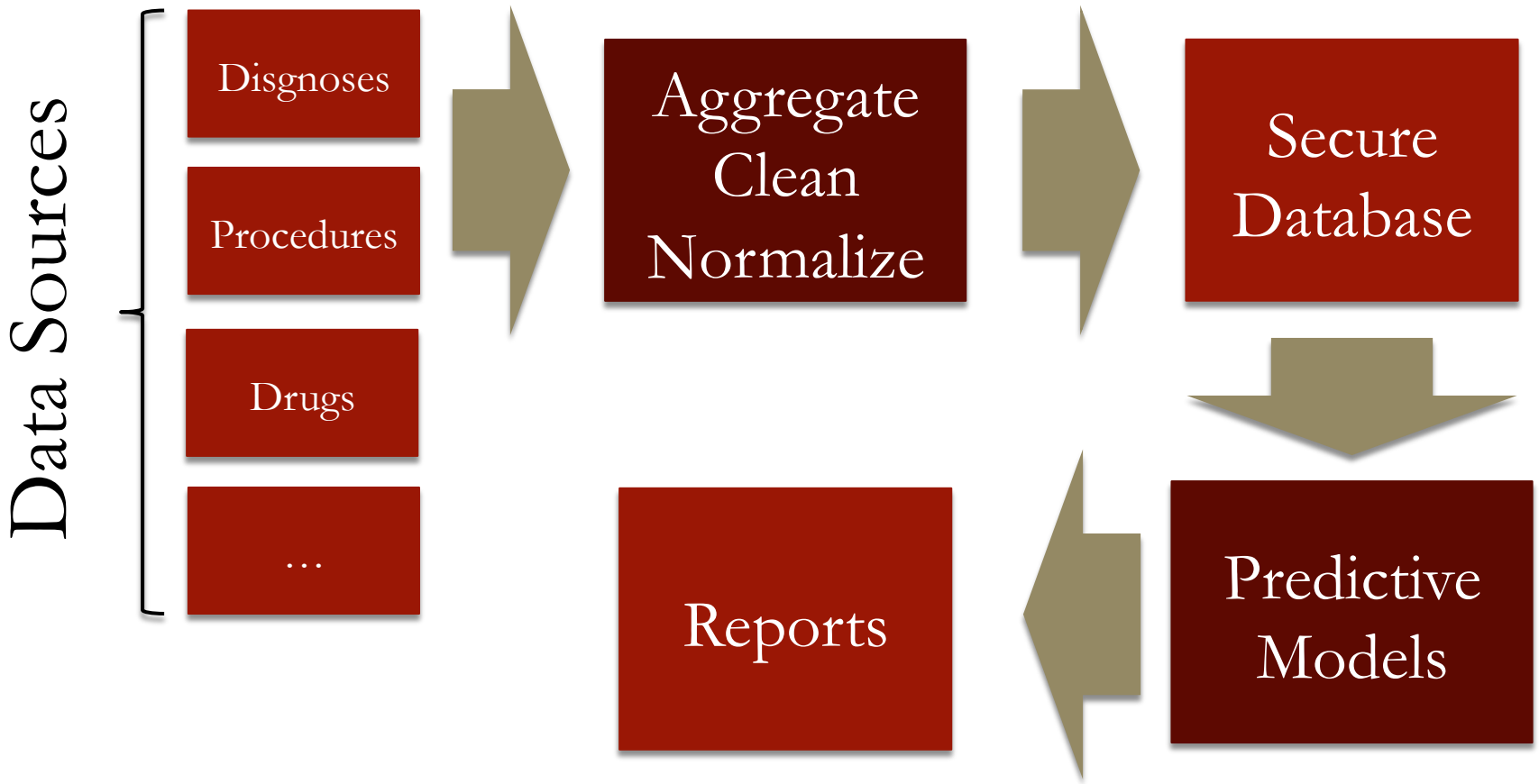
KEEPING AN EYE ON HEALTHCARE COSTS

The D2Hawkeye Story

D2Hawkeye

- Founded by Chris Kryder, MD, MBA in 2001
- Combine expert knowledge and databases with analytics to improve quality and cost management in healthcare
- Located in Massachusetts USA, grew very fast and was sold to Verisk Analytics in 2009

D2Hawkeye


Data Sources
Disgnoses
Procedures
Drugs
…
Aggregate
Clean
Normalize
Secure
Database
Predictive
Models
Reports

Healthcare Case Management

- D2Hawkeye tries to improve healthcare case management
  - Identify high-risk patients
  - Work with patients to manage treatment and associated costs
  - Arrange specialist care
- Medical costs often relate to severity of health problems, and are an issue for both patient and provider
- Goal: improve the quality of cost predictions

Impact

- Many different types of clients
  - Third party administrators of medical claims
  - Case management companies
  - Benefit consultants
  - Health plans
- **Millions of people** analyzed monthly through analytic platform in 2009
- **Thousands of employers** processed monthly

Pre-Analytics Approach

- Human judgment – MDs manually analyzed patient histories and developed
- Limited data sets
- Costly and inefficient
- Can we use analytics instead?

Data Sources

- Healthcare industry is data-rich, but data may be hard to access
  - Unstructured – doctor's notes
  - Unavailable – hard to get due to differences in technology
  - Inaccessible – strong privacy laws around healthcare data sharing
- What is available?

Data Sources

- Claims data
  - Requests for reimbursement submitted to insurance companies or state-provided insurance from doctors, hospitals and pharmacies.
- Eligibility information
- Demographic information

Claims Data

ClaimType	ProviderName	DiagCode	DiagDesc	Source DiagCode	SourceDiagDesc	ProcNDC Code	ProcNDCDesc	ServiceDate	PaidAmount
DEN	SOUTHEASTERN MINNESOTA ORAL & MAX	DD0238	Dental Diseases	5206	Unspecified Anomaly of Tooth Position	DD007	Anesthesia - General	4/22/2005	$ -
DEN	ASSOCIATED ORAL & MAXILLOFACIAL SURGEONS PA	DD0238	Dental Diseases	5206	Disturbances in ToOther Eruption	DD025	Dental	7/8/2005	$ 272.68
DEN	CENTRAL FLORIDA ORAL SURGERY	DD0238	Dental Diseases	5206	Disturbances in ToOther Eruption	DD025	Dental	11/11/2005	$ 568.13
Med	ALPHARETTA INTERNA	DD0004	ENT and Upper Resp Disorders	4610	Acute Maxillary Sinusitis	DD147	Office Visit - Established Patient	5/26/2005	$ 125.85
Med	CUMMING FAMILY MEDICINE	DD0170	Neurotic and Personality Disorders	30000	Neurotic Disorders-30000	DD149	Office Visit - New Patient	6/20/2005	$ -
Med	ATLANTA WOMENS HEALTH GROUP-582483738.20	DD0102	Screening	V776	Special Screening for Cystic Fibrosis	DD077	Lab - Blood Tests	7/29/2005	$ 1.52

Claims Data

- Rich, structured data source
- Very high dimension
- Doesn't capture all aspects of a persons treatment or health – many things must be inferred
- Unlike electronic medical records, we do not know the results of a test, only that a test was administered

D2Hawkeye's Claims Data

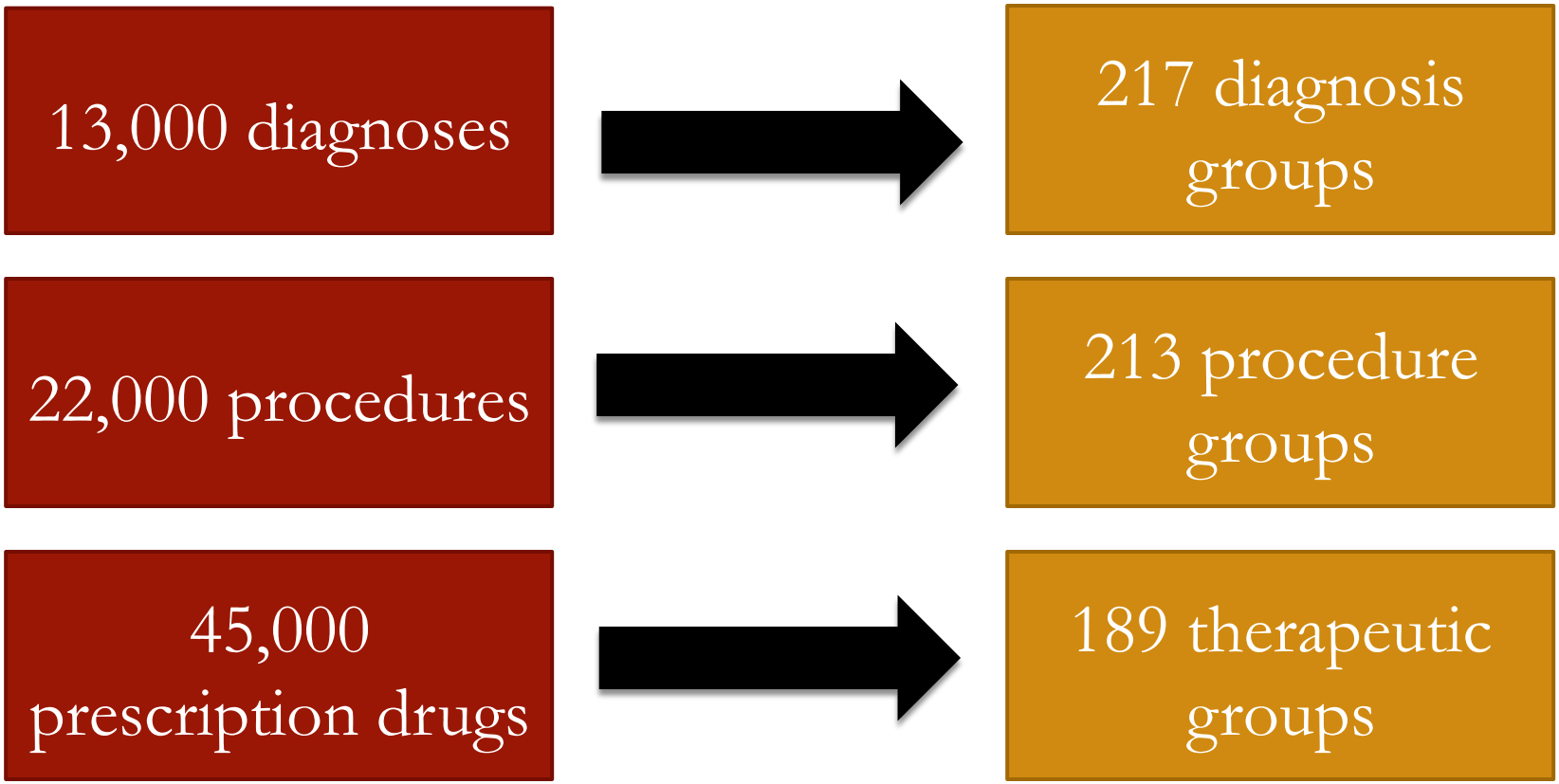
- Available: claims data for **2.4 million people** over a span of **3 years**

"Observation" Period 2001-2003

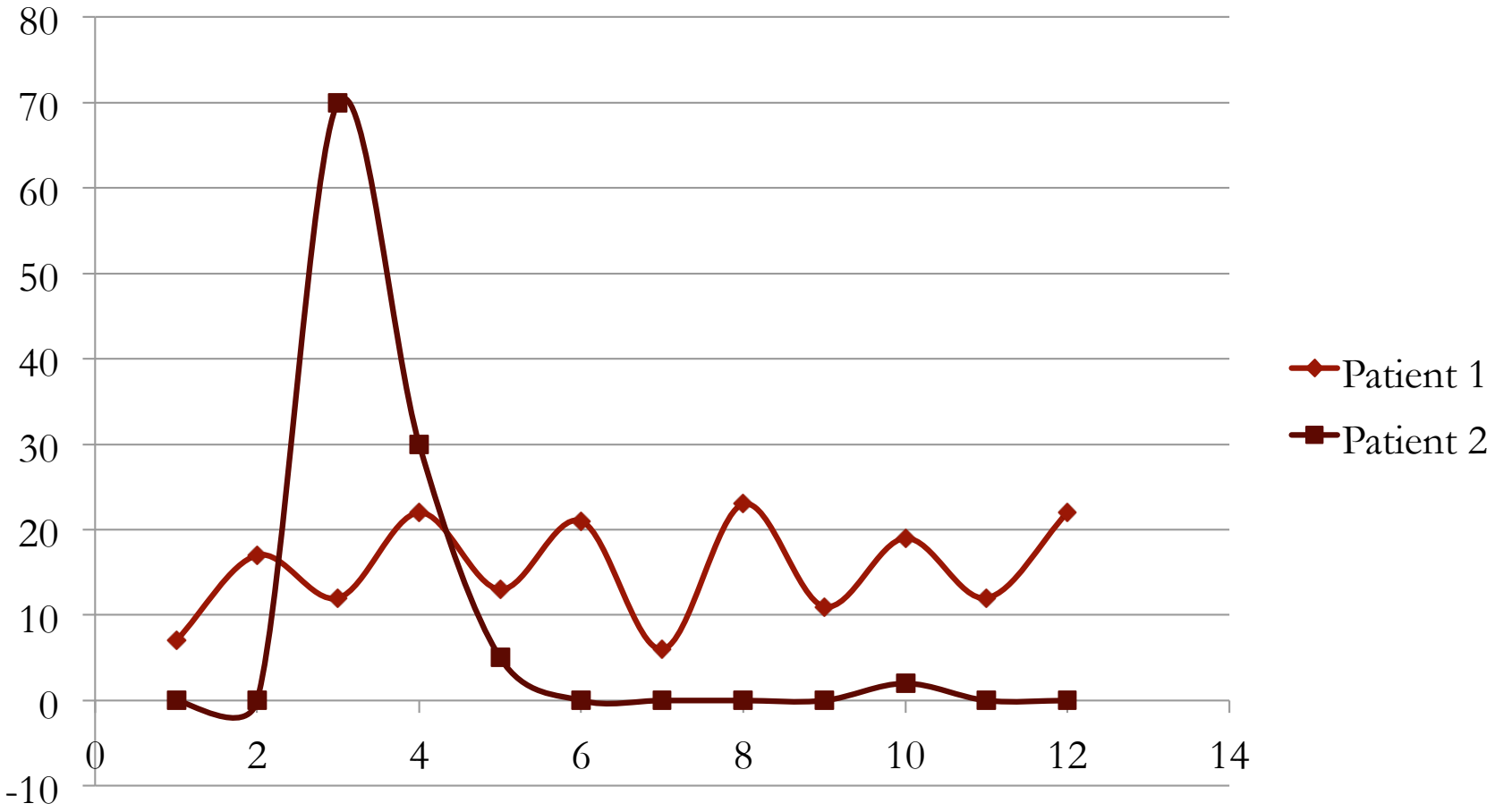
"Results" Period 2003-2004

- Include only people with data for at least 10 months in both periods – **400,000 people**

Variables


13,000 diagnoses
217 diagnosis groups
22,000 procedures
213 procedure groups
45,000 prescription drugs
189 therapeutic groups

Variables – Cost Profiles



Additional Variables

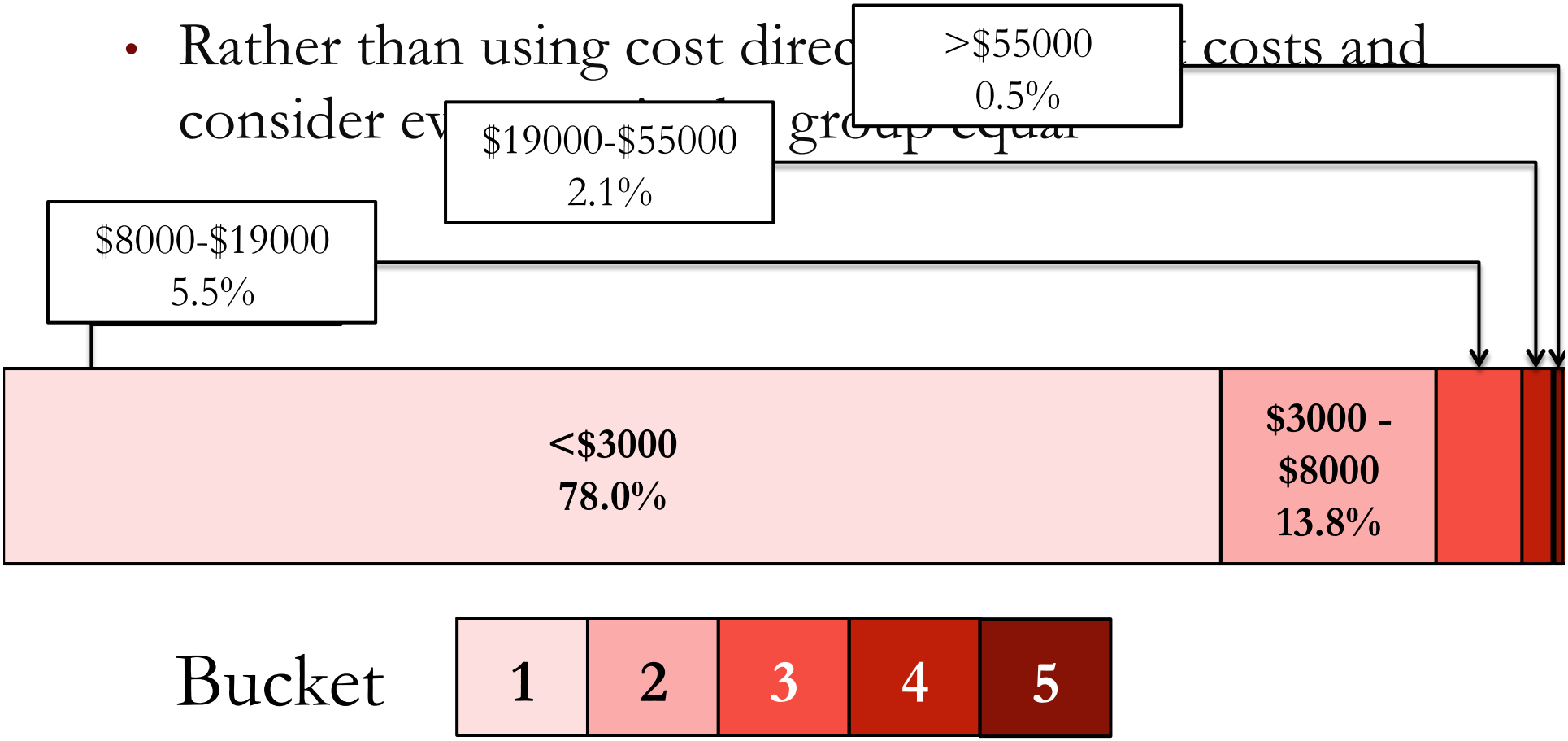
- Chronic condition cost indicators

- 269 medically defined risk rules
  - Interactions between illnesses obesity – depression
  - Interactions between diagnosis and age
  - Noncompliance to treatment
  - Illness severity

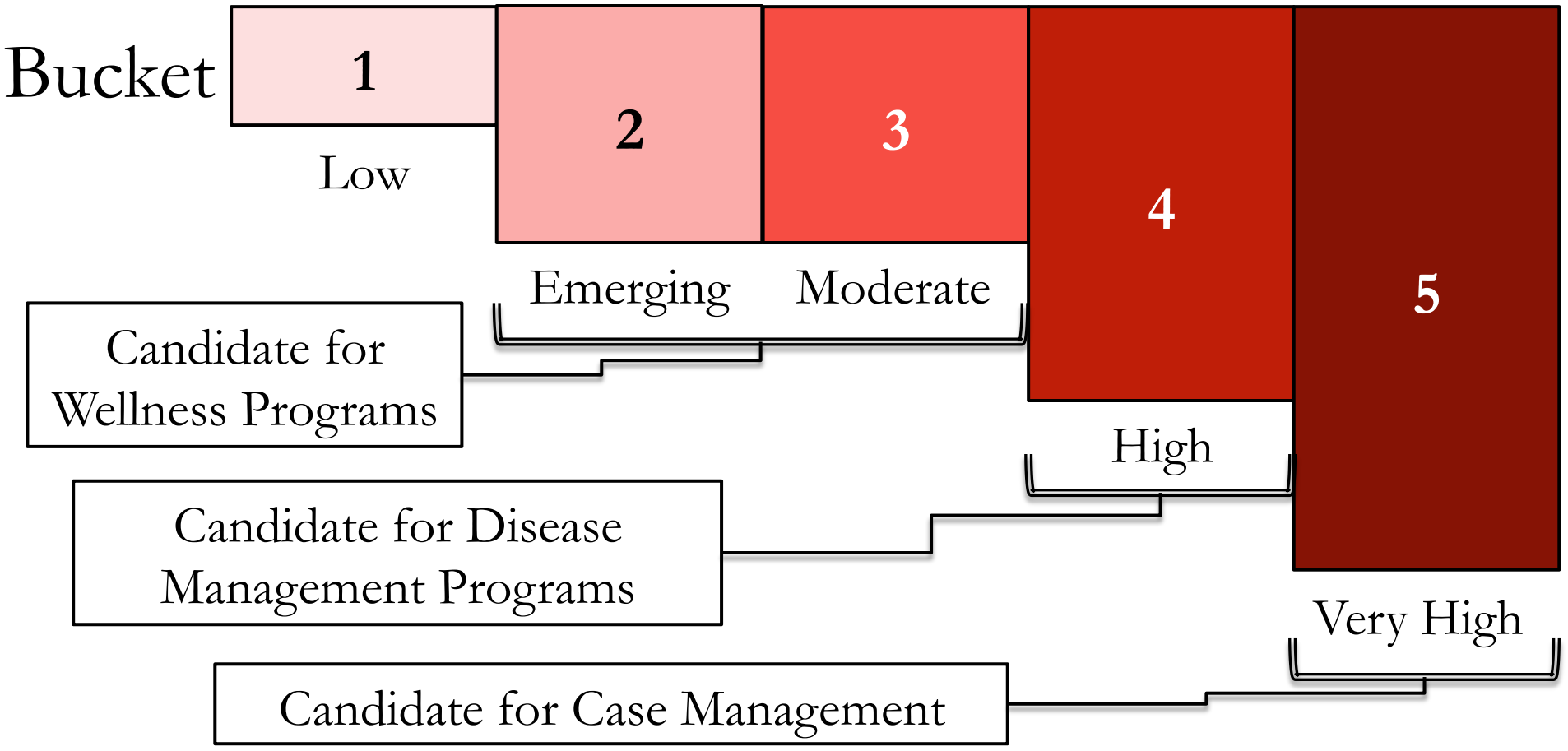
- Gender and age

Cost Variables

- Rather than using cost direc costs and consider ev group equal



Medical Interpretation of Buckets



Error Measures

- Typically we use R^2 or accuracy, but others can be used
- In case of D2Hawkeye, failing to classify a **high-cost patient** correctly is **worse** than failing to classify a **low-cost patient** correctly
- Use a "penalty error" to capture this asymmetry

Penalty Error

- Key idea: use asymmetric penalties
- Define a "penalty matrix" as the cost of being wrong

For example, whenever we classify a low-risk patient as high-risk, we pay a penalty of 2, which is a difference of 3 minus 1, the difference in the error.
But inversely, when you classify a bucket 3 patient as bucket 1 patient, this is double.

		Outcome				
		1	**2**	**3**	**4**	**5**
Forecast	**1**	0	2	4	6	8
	2	1	0	2	4	6
	3	2	1	0	2	4
	4	3	2	1	0	2
	5	4	3	2	1	0

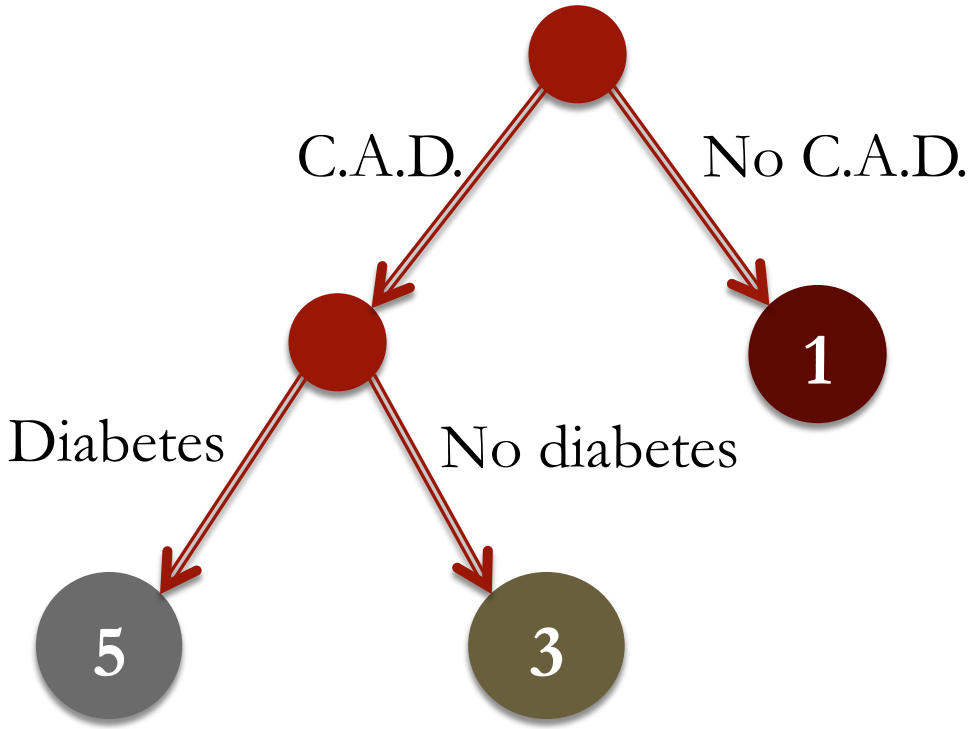
The cost-- the penalty-- is double the amount.
So you observe that the off diagonal penalties are double the corresponding penalties in the lower diagonal.

Baseline

- Baseline is to simply predict that the cost in the next "period" will be the cost in the current period
- Accuracy of 75%
- Penalty Error of 0.56

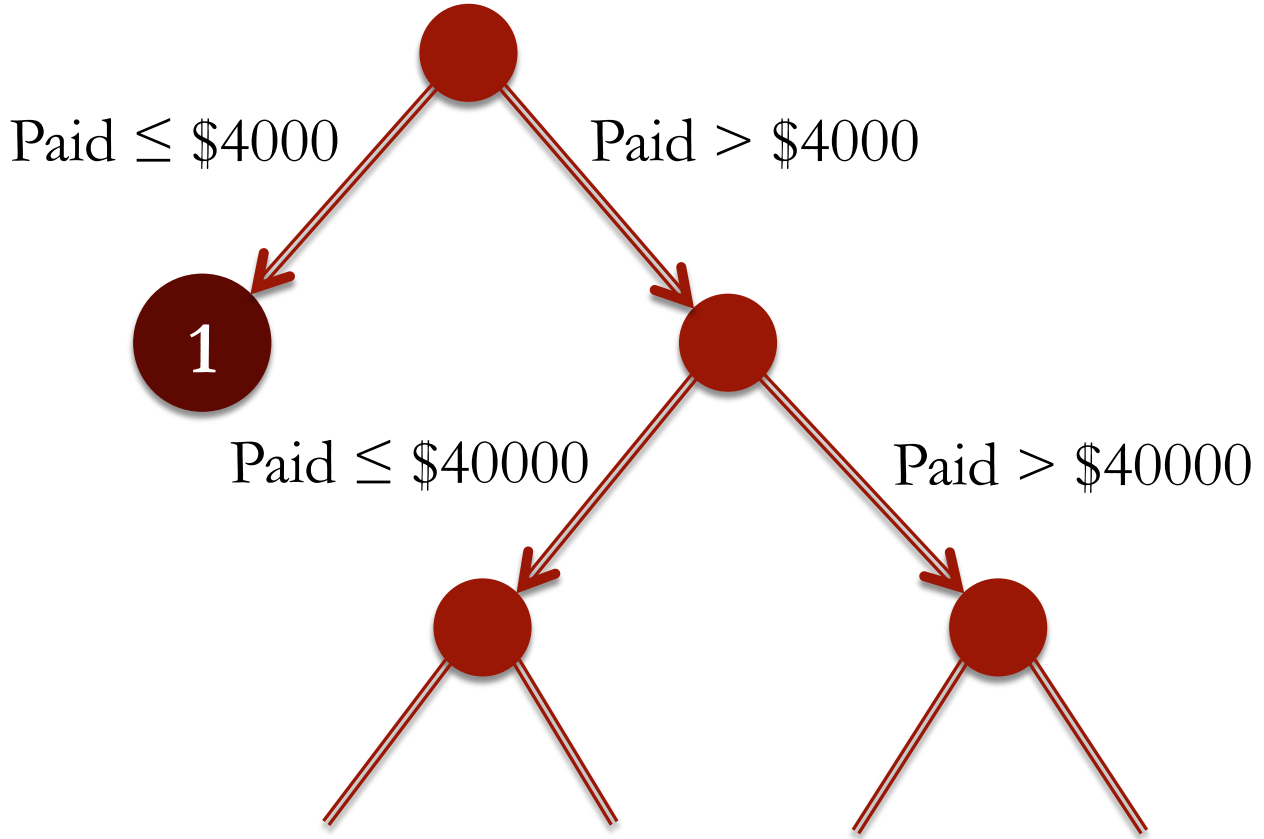
Multi-class Classification

- We are predicting a bucket number
- Example



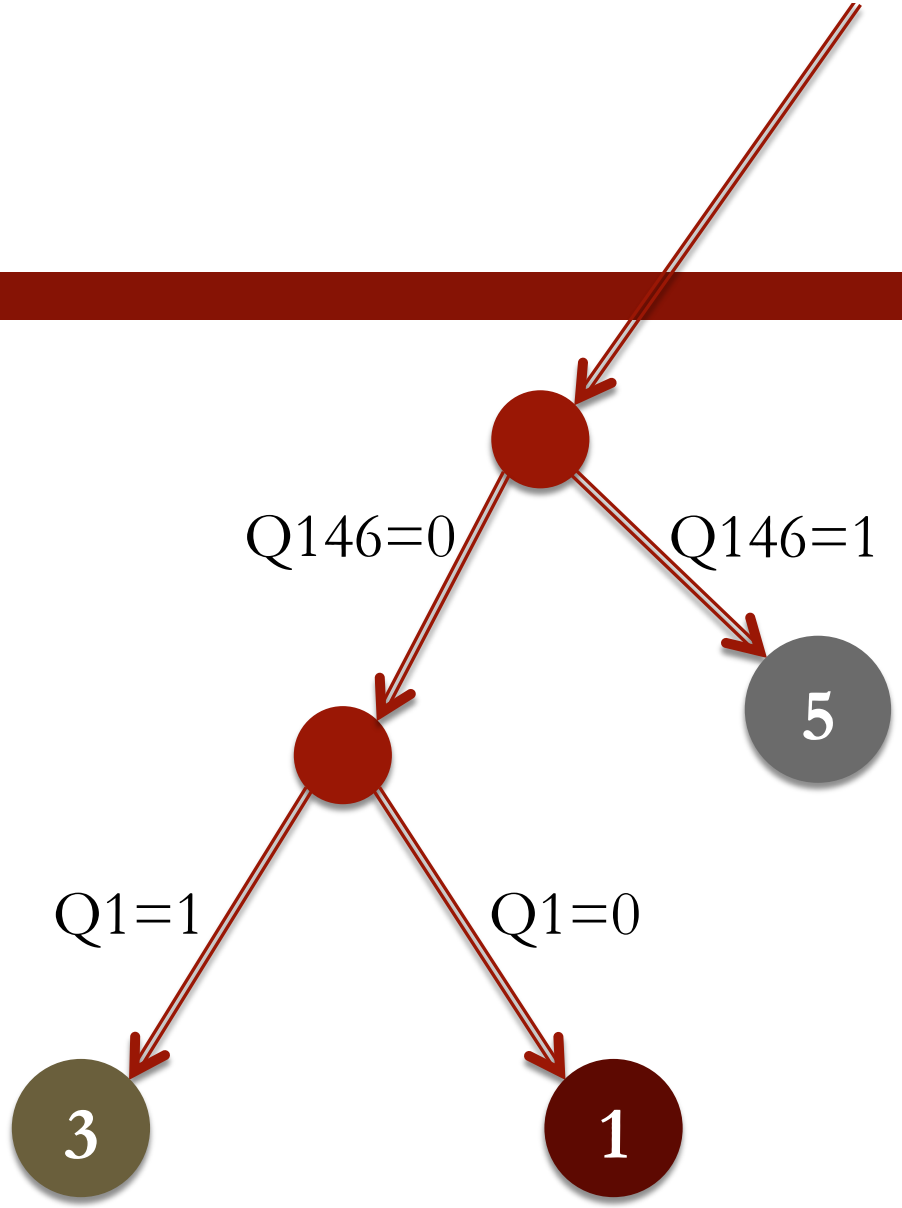
Most Important Factors

- First splits are related to cost



Secondary Factors

- Risk factors
- Chronic Illness
- **"Q146"**
  - Asthma + depression
- **"Q1"**
  - Risk factor indicating hylan injection
  - Possible knee replacement or arthroscopy



Example Groups for Bucket 5

- Under 35 years old, between $3300 and $3900 in claims, C.A.D., but no office visits in last year
- Claims between $3900 and $43000 with at least $8000 paid in last 12 months, $4300 in pharmacy claims, acute cost profile and cancer diagnosis
- More than $58000 in claims, at least $55000 paid in last 12 months, and not an acute profile

Results

	Accuracy		Penalty Error	
Bucket	**Trees**	**Baseline**	**Trees**	**Baseline**
All	80%	75%	0.52	0.56
1	85%	85%	0.42	0.44
2	60%	31%	0.89	0.96
3	53%	21%	1.01	1.37
4	39%	19%	1.01	1.72
5	30%	23%	1.01	1.88

Insights

- **Substantial improvement** over the baseline
- **Doubled accuracy** over baseline in some cases
- Smaller accuracy improvement on **bucket 5**, but **much lower penalty**

Analytics Provide an Edge

- Substantial improvement in D2Hawkeye's ability to identify patients who need more attention
- Because the model was interpretable, physicians were able to improve the model by identifying new variables and refining existing variables
- Analytics gave D2Hawkeye an edge over competition using "last-century" methods