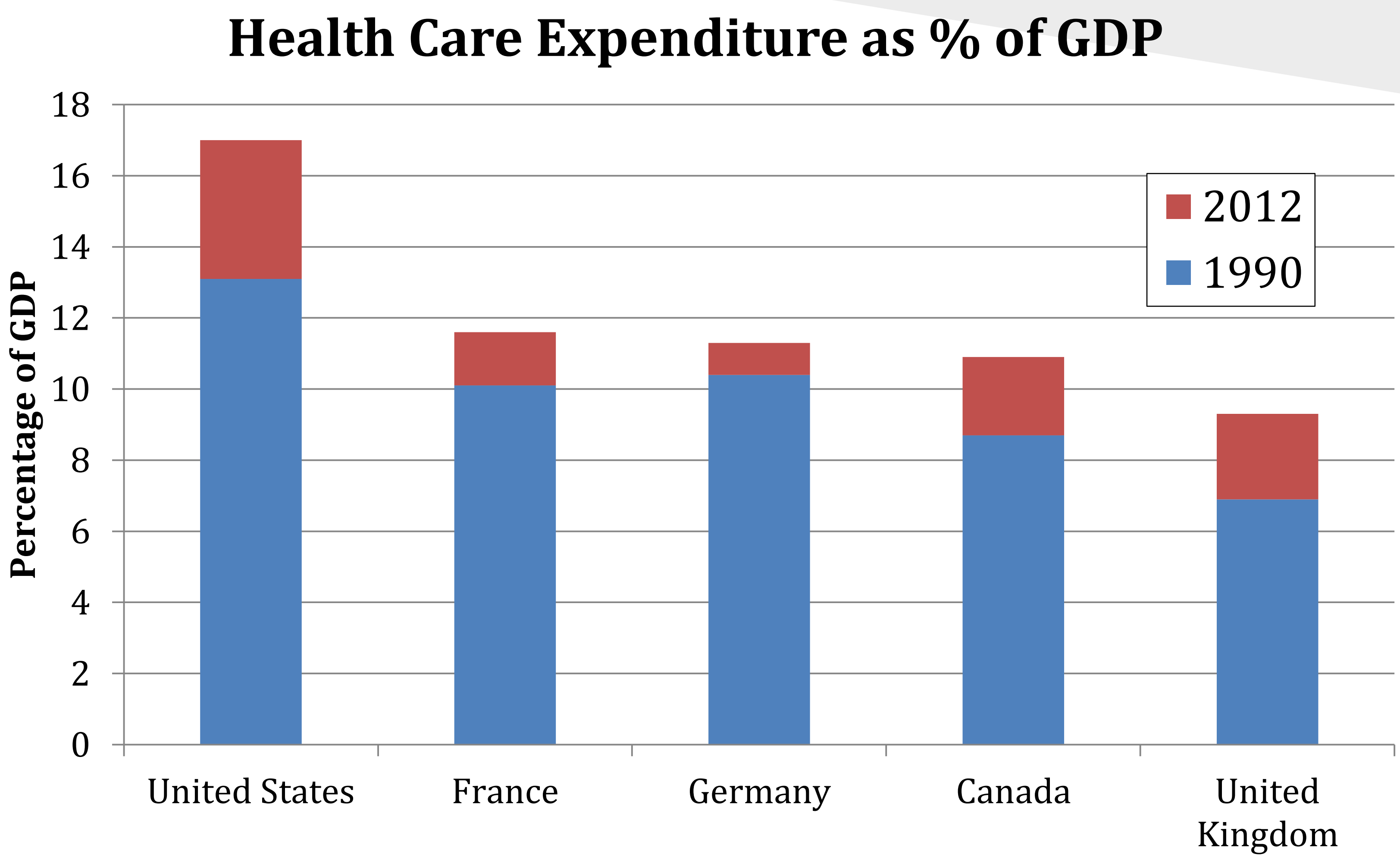
Machine Learning for Medical Decision Support and Individualized Treatment Assignment

Finn Kuusisto
Department of Computer Sciences
Doctoral Defense
August 14, 2015

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Health Care Expenditure



*World Health Statistics 2015, World Health Organization (WHO)

Precision Medicine Initiative

"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

-President Barack Obama, State of the Union Address, January 20, 2015

Precision Medicine

- Tailoring medical treatment to individual characteristics of each patient
- Classify individuals into subpopulations that differ in:
  - Susceptibility to particular diseases
  - Biology and/or prognosis of diseases they develop
  - Response to specific treatments

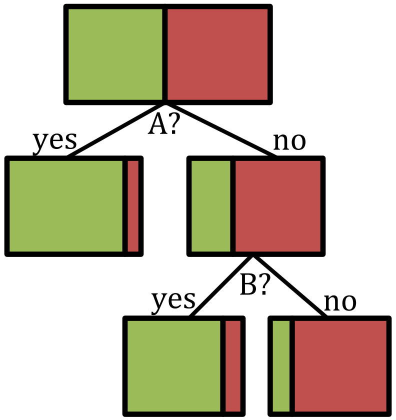
Supervised Learning

Given: Values of the input features and the output feature (response, class) for many patients

Do: Build a model that can accurately predict the unknown value of the output class for new (previously unseen) patients whose values of the input features are known

Classical methods: linear and logistic regression

Other methods: decision trees, random forests, support vector machines, Bayesian networks, artificial neural networks, etc.



Thesis Statement

Machine learning results can be made more clinically-relevant by tailoring current approaches to meet clinical objectives through the development of new algorithms to model individual response to treatment, and by incorporating clinical expertise into model development and refinement.

Publications

Clinical Collaboration

F. Kuusisto, I. Dutra, M. Elezaby, E. Mendonca, J. Shavlik, and E. S. Burnside. "Leveraging Expert Knowledge to Improve Machine-Learned Decision Support Systems". *AMIA Joint Summits on Translational Science*, 2015.

M. Elezaby, **F. Kuusisto**, J. Shavlik, Y. Wu, A. Gegios, H. Neuman, W. B. DeMartini, E. S. Burnside. Core Needle Biopsies: A Predictive Model that Identifies Low Probability (≤2%) Lesions to Safely Avoid Surgical Excision. *Radiological Society of North America (RSNA) 101st Scientific Assembly and Annual Meeting*, 2015.

A. Gegios, M. Elezaby, W. B. DeMartini, J. Cox, C. Montemayor-Garcia, H. Neuman, **F. Kuusisto**, J. M. Hampton, E. S. Burnside. Differential Upgrade Rates for Non-Definitive Image-Guided Core Needle Breast Biopsies Based on BI-RADS Features. *Radiological Society of North America (RSNA) 101st Scientific Assembly and Annual Meeting*, 2015.

F. Kuusisto, I. Dutra, H. Nassif, Y. Wu, M. E. Klein, H. Neuman, J. Shavlik, and E. S. Burnside. "Using Machine Learning to Identify Benign Cases with Non-Definitive Biopsy". *IEEE International Conference on e-Health Networking, Applications & Services*, 2013.

Individualized Treatment Effects

J. Weiss, **F. Kuusisto**, K. Boyd, J. Liu, D. Page. "Machine Learning for Treatment Assignment: Improving Individualized Risk Attribution". *AMIA Annual Symposium*, 2015.

F. Kuusisto, V. Santos Costa, H. Nassif, E. S. Burnside, D. Page, and J. Shavlik. "Support Vector Machines for Differential Prediction". *European Conference on Machine Learning*, 2014.

H. Nassif, **F. Kuusisto**, E. S. Burnside, D. Page, J. Shavlik, and V. Santos Costa. "Score As You Lift (SAYL): A Statistical Relational Learning Approach to Uplift Modeling". *European Conference on Machine Learning*, 2013.

H. Nassif, **F. Kuusisto**, E. S. Burnside, and J. Shavlik. "Uplift Modeling with ROC: An SRL Case Study". *International Conference on Inductive Logic Programming*, 2013.

Publications

Clinical Collaboration

F. Kuusisto, I. Dutra, M. Elezaby, E. Mendonca, J. Shavlik, and E. S. Burnside. "Leveraging Expert Knowledge to Improve Machine-Learned Decision Support Systems". *AMIA Joint Summits on Translational Science*, 2015.

M. Elezaby, **F. Kuusisto**, J. Shavlik, Y. Wu, A. Gegios, H. Neuman, W. B. DeMartini, E. S. Burnside. Core Needle Biopsies: A Predictive Model that Identifies Low Probability (≤2%) Lesions to Safely Avoid Surgical Excision. *Radiological Society of North America (RSNA) 101st Scientific Assembly and Annual Meeting*, 2015.

A. Gegios, M. Elezaby, W. B. DeMartini, J. Cox, C. Montemayor-Garcia, H. Neuman, **F. Kuusisto**, J. M. Hampton, E. S. Burnside. Differential Upgrade Rates for Non-Definitive Image-Guided Core Needle Breast Biopsies Based on BI-RADS Features. *Radiological Society of North America (RSNA) 101st Scientific Assembly and Annual Meeting*, 2015.

F. Kuusisto, I. Dutra, H. Nassif, Y. Wu, M. E. Klein, H. Neuman, J. Shavlik, and E. S. Burnside. "Using Machine Learning to Identify Benign Cases with Non-Definitive Biopsy". *IEEE International Conference on e-Health Networking, Applications & Services*, 2013.

Individualized Treatment Effects

J. Weiss, **F. Kuusisto**, K. Boyd, J. Liu, D. Page. "Machine Learning for Treatment Assignment: Improving Individualized Risk Attribution". *AMIA Annual Symposium*, 2015.

F. Kuusisto, V. Santos Costa, H. Nassif, E. S. Burnside, D. Page, and J. Shavlik. "Support Vector Machines for Differential Prediction". *European Conference on Machine Learning*, 2014.

H. Nassif, **F. Kuusisto**, E. S. Burnside, D. Page, J. Shavlik, and V. Santos Costa. "Score As You Lift (SAYL): A Statistical Relational Learning Approach to Uplift Modeling". *European Conference on Machine Learning*, 2013.

H. Nassif, **F. Kuusisto**, E. S. Burnside, and J. Shavlik. "Uplift Modeling with ROC: An SRL Case Study". *International Conference on Inductive Logic Programming*, 2013.

Outline

- Introduction
- Advice-Based Learning Framework
- Support Vector Machines for Uplift Modeling
- Conclusions

Outline

- Introduction
- Advice-Based Learning Framework
- Support Vector Machines for Uplift Modeling
- Conclusions

Decision Support

Great opportunities for machine-learned decision support systems

But…

Standardized, complete, and sufficient training data is rarely available

ABLe

Comprises two parts

1) Categories of advice sources

2) Iterative process for model refinement

ABLe - Advice Categories

Task

- What is the problem and scope?
- What predictor variables are important?
- How should the problem be modeled?

Relationships Among Variables

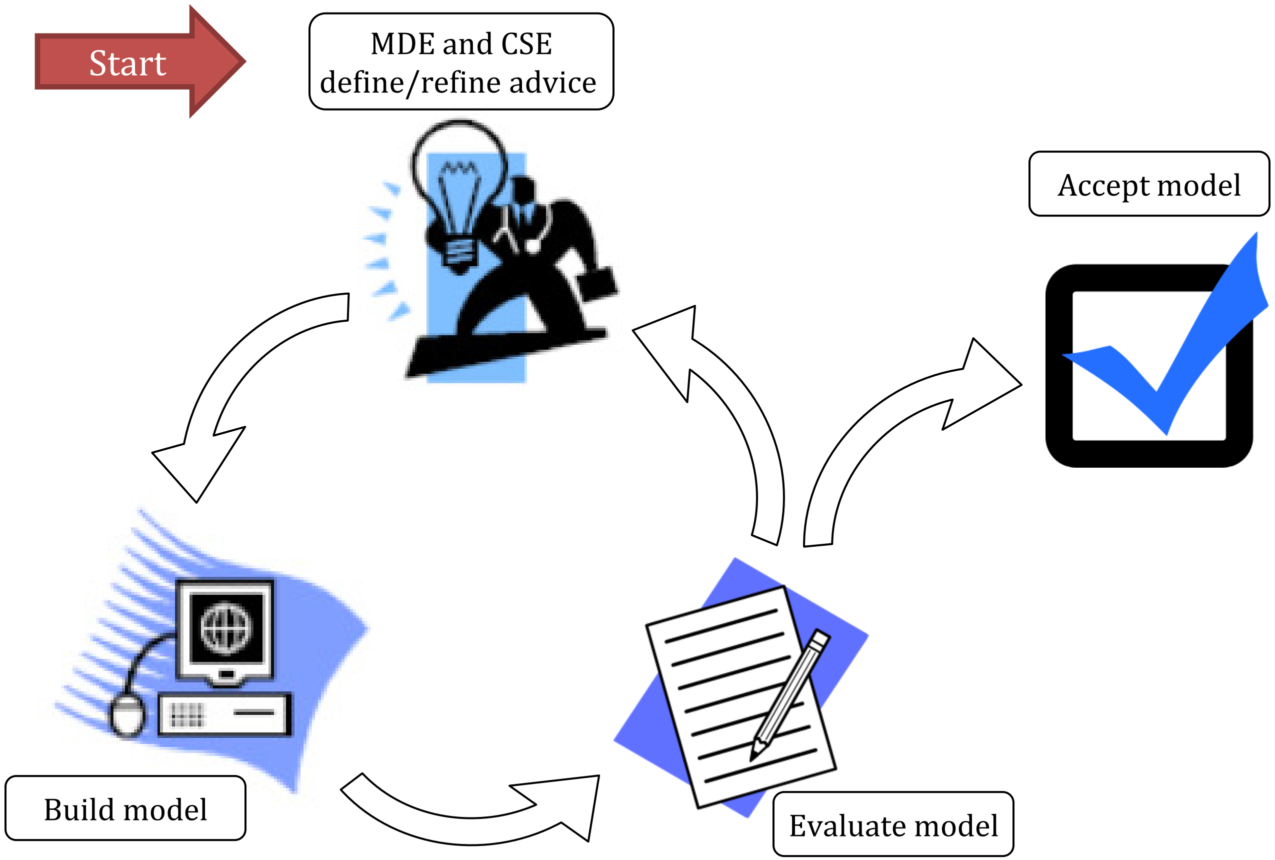
- What combinations of variables are important to the task?

Parameter Values

- What is the clinical objective?
- What model parameters best represent that objective?

ABLe - Iterative Process

Repeated iterations to optimize performance

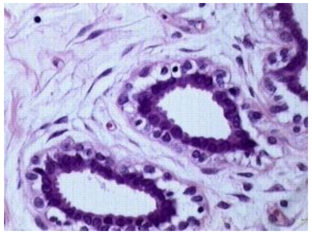


Upgrade Prediction

1
Mammogram

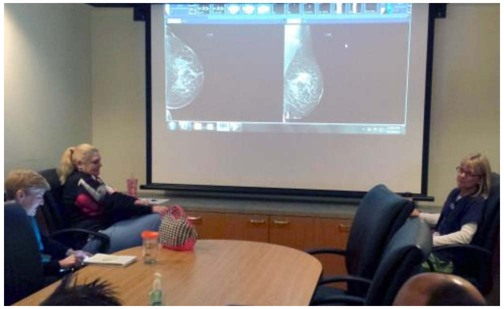
Abnormality

2
Needle Biopsy



Benign Tissue

3
Radiologic-Histologic Correlation



Non-definitive Diagnosis

4
Excision

Final Diagnosis

Malignant
=
"Upgrade"

Image Sources:
1. NIH - wikimedia.org/wiki/File:Woman_receives_mammogram.jpg
2. Itayba - wikimedia.org/wiki/File:Normal.jpg
3. UW Hospital and Clinics
4. NIH - wikimedia.org/wiki/File:Surgical_breast_biopsy.jpg

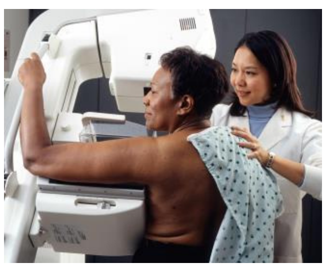
Upgrade Prediction

- 5-15% of core needle biopsies non-definitive
- Approximately 35,000-105,000* per year
- 80-90% of non-definitive biopsies are **benign**

* Based on 2010 annual breast biopsy utilization rate in the United States

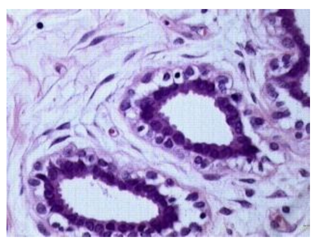
Upgrade Prediction

1
Mammogram



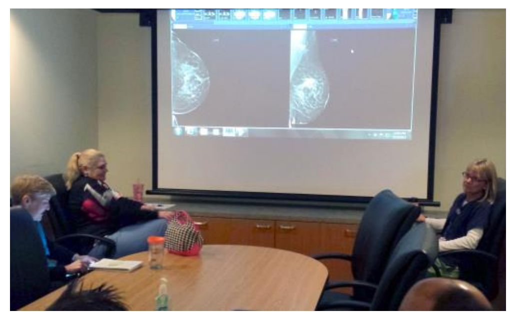
Abnormality

2
Needle Biopsy



Benign Tissue

3
Radiologic-Histologic Correlation



Non-definitive Diagnosis

4
Excision

Final Diagnosis

Malignant
=
"Upgrade"

Image Sources:
1. NIH - wikimedia.org/wiki/File:Woman_receives_mammogram.jpg
2. Itayba - wikimedia.org/wiki/File:Normal.jpg
3. UW Hospital and Clinics
4. NIH - wikimedia.org/wiki/File:Surgical_breast_biopsy.jpg

Phase 1

Task

- Simple probabilistic model (Naïve Bayes)
- Standardized BI-RADS descriptor features
- Some non-standard pathology features and demographics
- Predict probability of **malignancy**
- Assume excision at ≥ 0.02 model score (to balance risk)

Relationships Among Variables

- Rules predicting **increase**/**decrease** risk of **malignancy**

Parameter Values

- None

Relationships Among Variables

If-Then rules from domain expert (Beth) that suggest **increase**/**decrease** risk of **upgrade**.

High-risk mass rule:

IF
Irregular mass shape is present **OR**
Spiculated mass margin is present **OR**
High density mass is present **OR**
Increasing mass size
THEN
Risk of upgrade increases

Biopsies in Practice (2006-11)

Core Needle Biopsies
2,808

Core Needle Biopsies + Dx Mammogram
1,910

Malignant Biopsy
601

Benign Biopsy
1,309

Non-definitive
157

Malignant (upgrade)
29

Benign (non-upgrade)
128

Phase 1 Results

- Naïve Bayes to predict **malignancy**
- Assume excision at ≥ 0.02 model score
- Experiments with and without expert rule features

	Data	Rules	Data + Rules
Malignant Excisions Missed (%)	8 (27.6%)	1 (3.4%)	9 (31.0%)
Benign Excisions Avoided (%)	46 (35.9%)	5 (3.9%)	63 (49.2%)

Observations & Refinements

Observations

- No output threshold with acceptable performance
- Non-definitive biopsies broken into 3 categories at diagnosis
  - Atypical/Radial Scar (ARS)
  - Insufficient (I)
  - Discordant (D)
- ARS and I cases consistently mislabeled

Refinements

- Focus exclusively on discordant cases

Discordant Biopsies (2006-11)

Discordant Biopsy
60

Malignant (upgrade)
10

Benign (non-upgrade)
50

Phase 2 Results

- Naïve Bayes to predict **malignancy** of discordants
- Assume excision at ≥ 0.02 model score
- Experiments with and without expert rule features

	Data	Rules	Data + Rules
Malignant Excisions Missed (%)	**3 (30.0%)**	**1 (10.0%)**	**3 (30.0%)**
Benign Excisions Avoided (%)	**29 (58.0%)**	**17 (34.0%)**	**27 (54.0%)**

Observations & Refinements

Observations

- Good ranking of cases by output model scores
- Most cases assigned less than 0.02 risk

Refinements

- Make model conservative
  - Different costs for false negatives (FN) versus false positives (FP)
  - Take from utility analysis literature in mammography

Phase 3 Results

- Naïve Bayes to predict **malignancy** of discordants
- Cost ratio of 150:1 for FN:FP
- Assume excision at ≥ 0.02 model score
- Experiments with and without expert rule features

	Data	Rules	Data + Rules
Malignant Excisions Missed (%)	**0 (0.0%)**	**0 (0.0%)**	**0 (0.0%)**
Benign Excisions Avoided (%)	**5 (10.0%)**	**5 (10.0%)**	**12 (24.0%)**

Outline

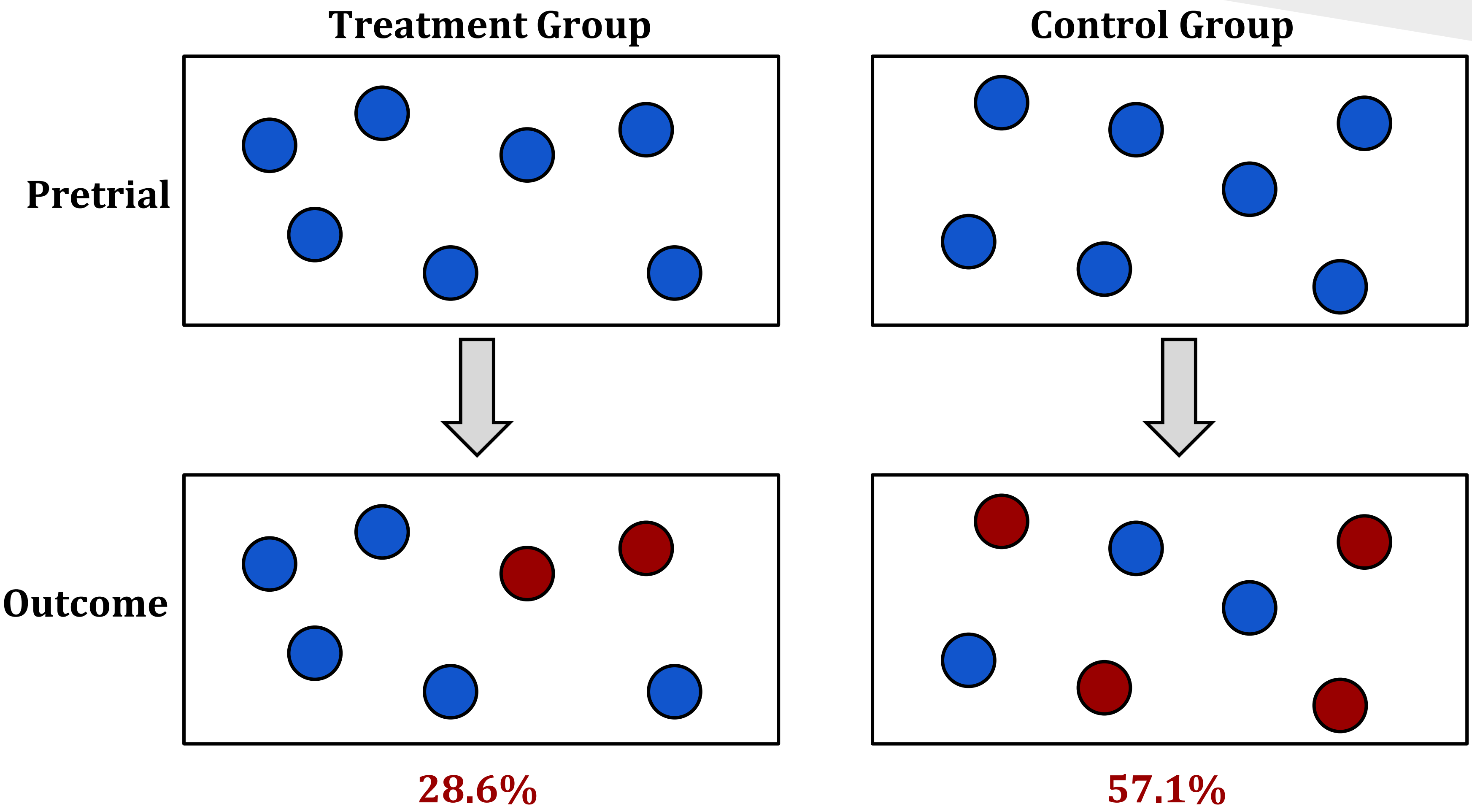
- Introduction
- Advice-Based Learning Framework
- Support Vector Machines for Uplift Modeling
- Conclusions

Clinical Trial

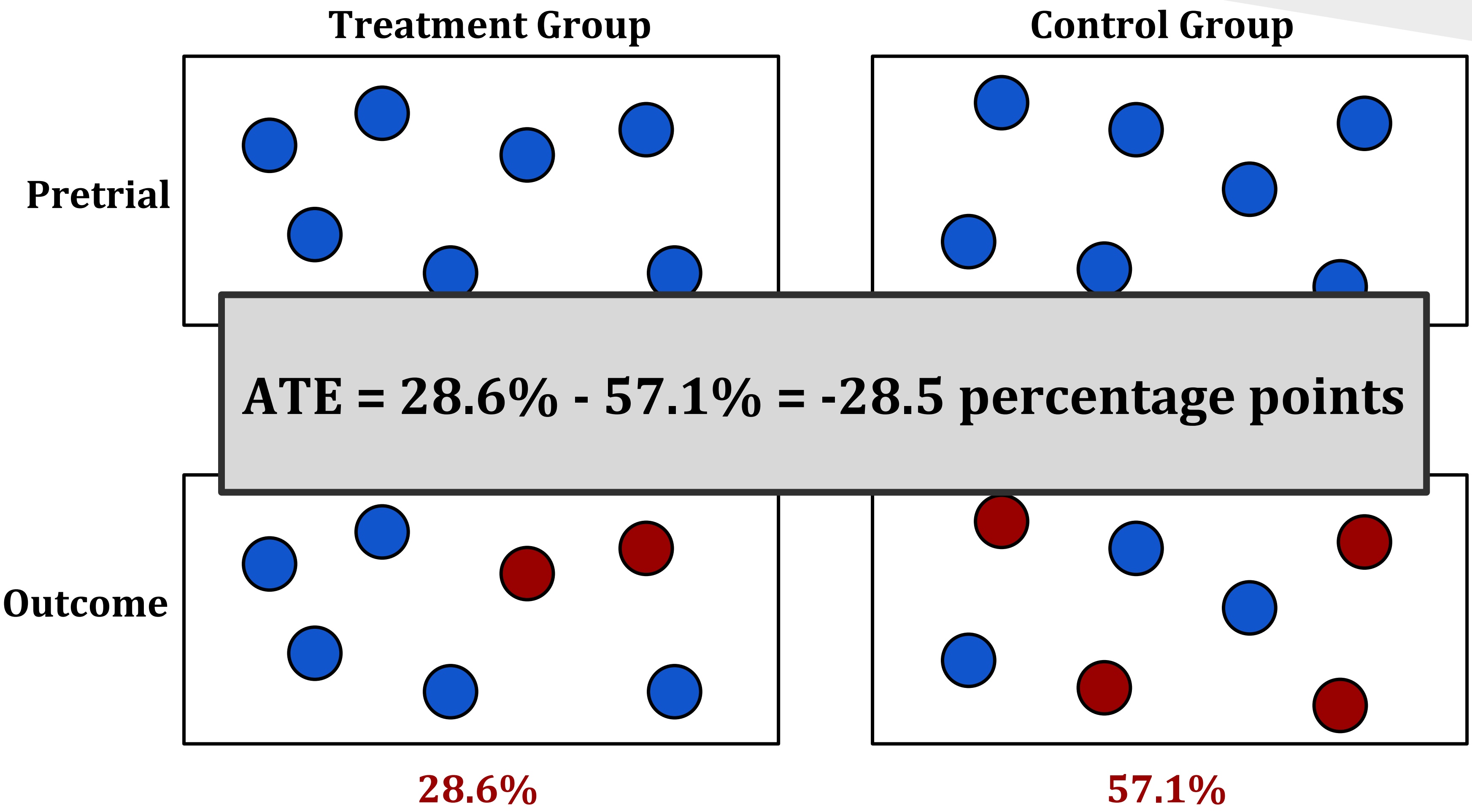
Clinical experiment to determine the average effect of some treatment for:

- Safety
- Efficacy

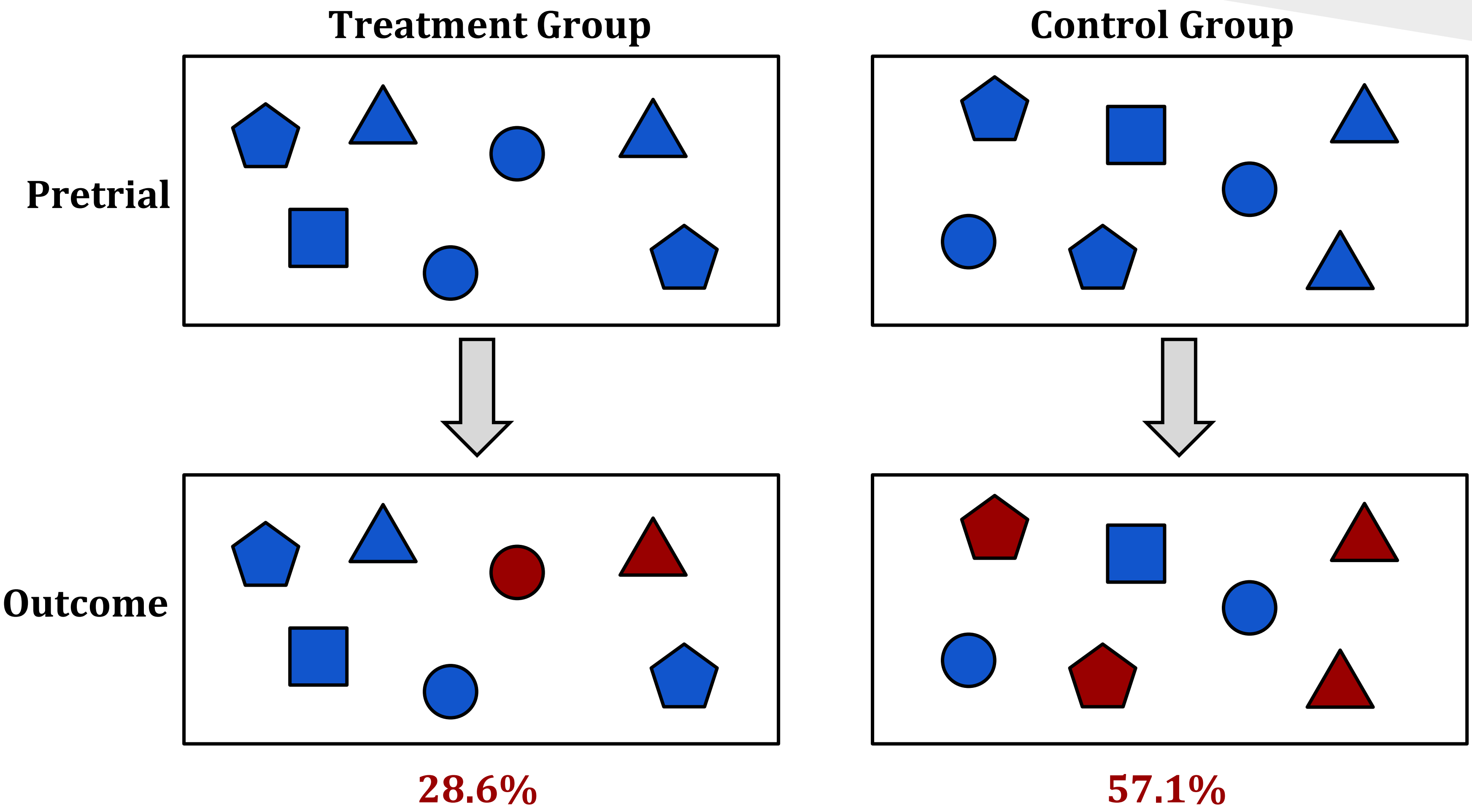
Clinical Trial



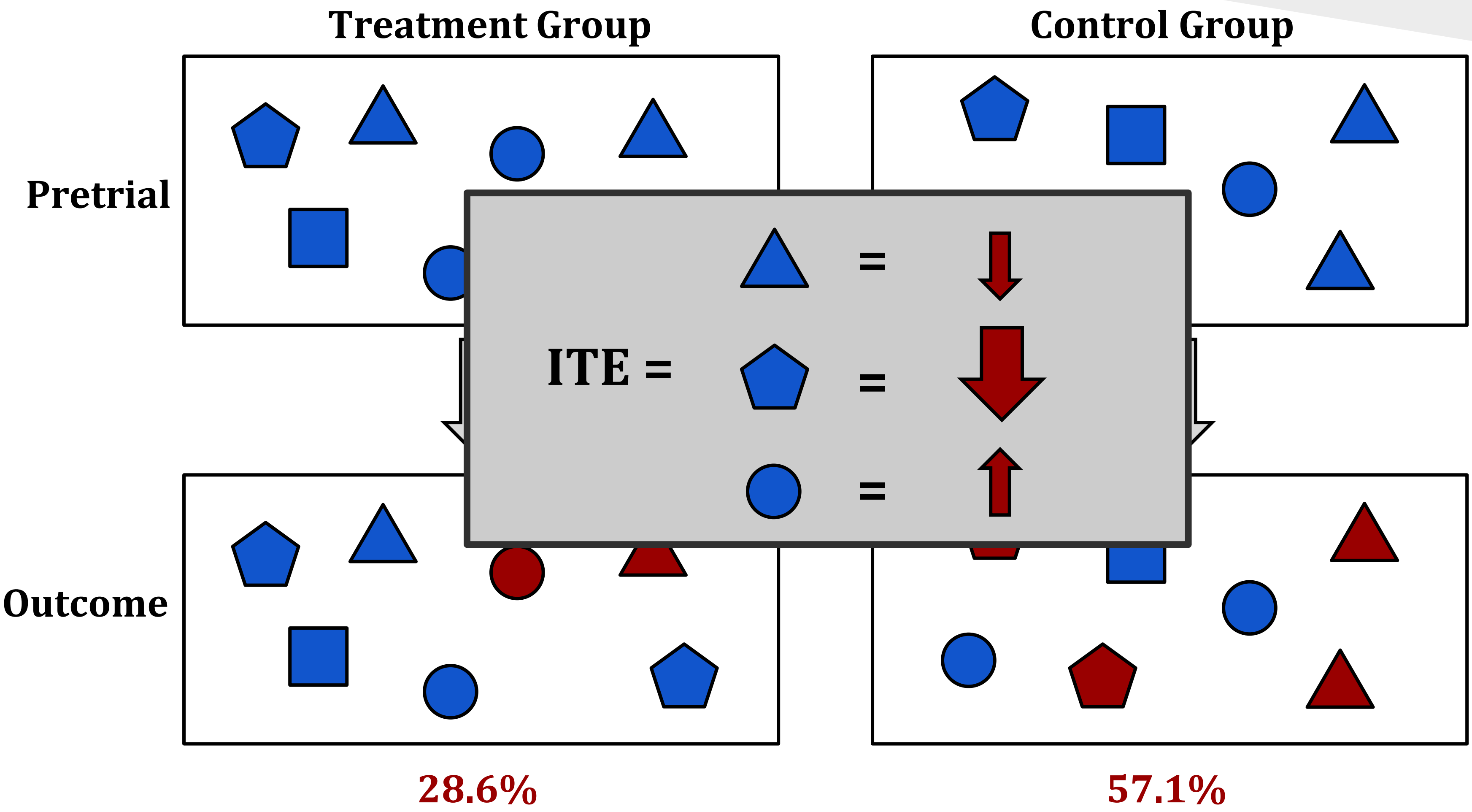
Clinical Trial



Clinical Trial



Clinical Trial



ITE Challenge

- Cannot observe both treatment and control outcomes for any one individual

- Need a lot of data to model ITE for even a moderate number of individual features

Image by Toni Barros - https://www.flickr.com/photos/12793495@N05/3233344867/

Uplift Modeling

(RADCLIFFE & SIMPSON, 2008)

How do we choose which customers to target with some marketing activity?

Persuadables	Customers who respond positively to marketing activity.
Sure Things	Customers who respond positively regardless.
Lost Causes	Customers who respond negatively regardless.
Sleeping Dogs	Customers who respond negatively to marketing activity.

Uplift Modeling

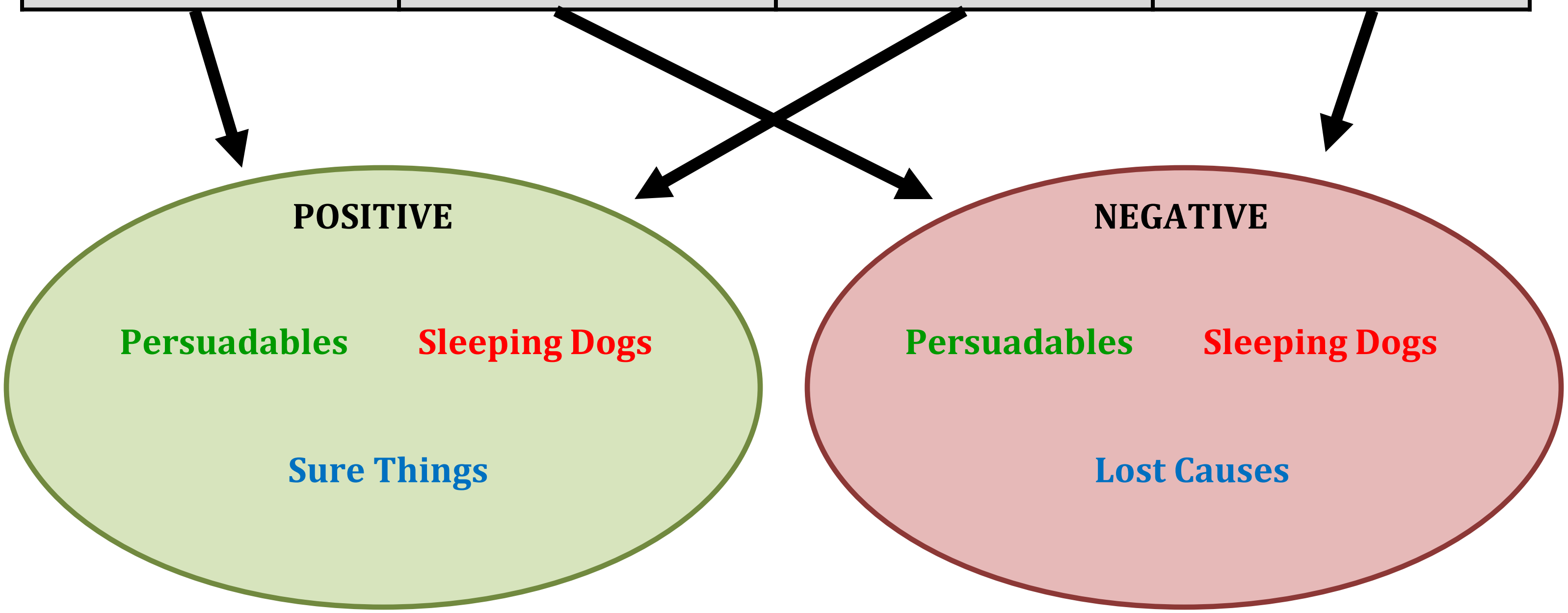
(RADCLIFFE & SIMPSON, 2008)

True customer groups are <u>unknown</u>

Treatment		Control	
Response	No Response	Response	No Response
Persuadables, Sure Things	Sleeping Dogs, Lost Causes	Sleeping Dogs, Sure Things	Persuadables, Lost Causes

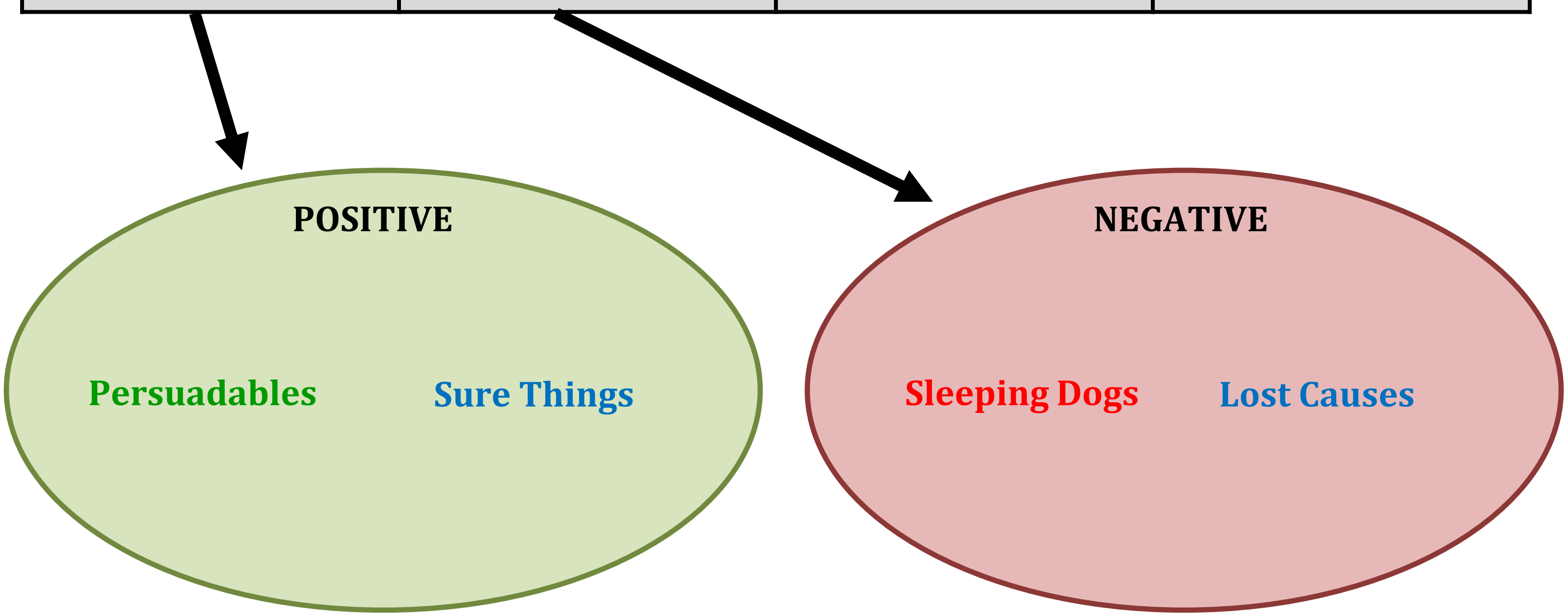
Standard Model

Treatment		Control	
Response	No Response	Response	No Response
Persuadables, Sure Things	Sleeping Dogs, Lost Causes	Sleeping Dogs, Sure Things	Persuadables, Lost Causes



Response Model

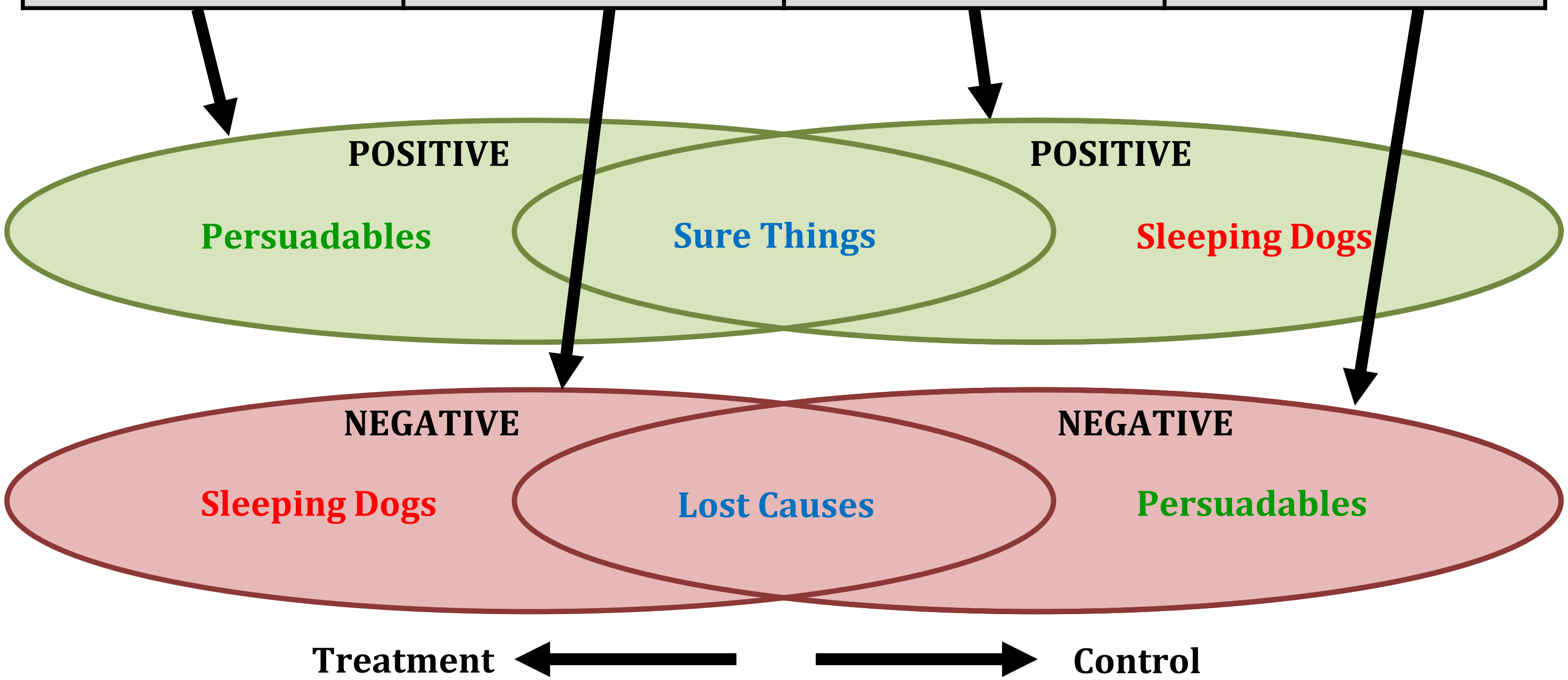
Treatment		Control	
Response	No Response	~~Response~~	~~No Response~~
Persuadables, Sure Things	Sleeping Dogs, Lost Causes	~~Sleeping Dogs, Sure Things~~	~~Persuadables, Lost Causes~~



Uplift Modeling

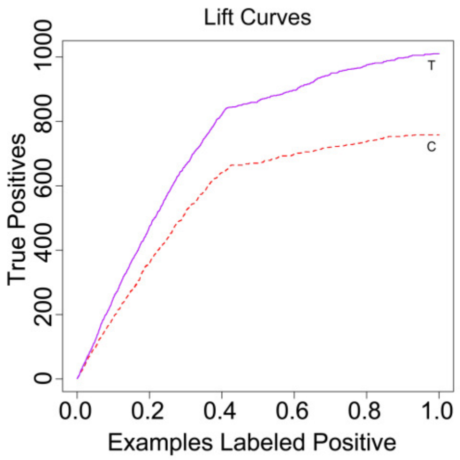
(RADCLIFFE & SIMPSON, 2008)

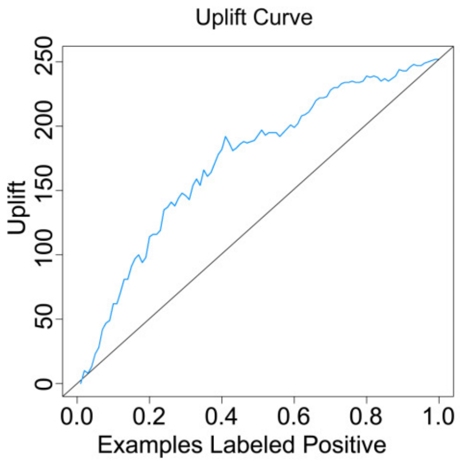
Treatment		Control	
Response	**No Response**	**Response**	**No Response**
Persuadables, Sure Things	Sleeping Dogs, Lost Causes	Sleeping Dogs, Sure Things	Persuadables, Lost Causes



Uplift Modeling

(RADCLIFFE & SIMPSON, 2008)





Lift

The number of **responders** that a classifier identifies **at a given proportion of the population targeted.**

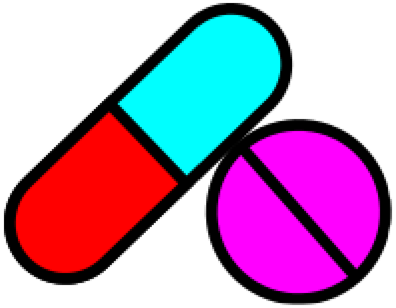
Uplift

The **difference in lift** produced by a classifier between treatment and control subgroups.

$$AUU = AUL_T - AUL_C$$

COX-2 Inhibitors

- Non-steroidal anti-inflammatory drug (NSAID)
- Significantly reduced occurrence of adverse gastrointestinal effects common to other NSAIDs (e.g. ibuprofen)
- Wide use for treatment of ailments such as arthritis
- Later clinical trials showed increased risk of myocardial infarction (MI), or "heart attack"



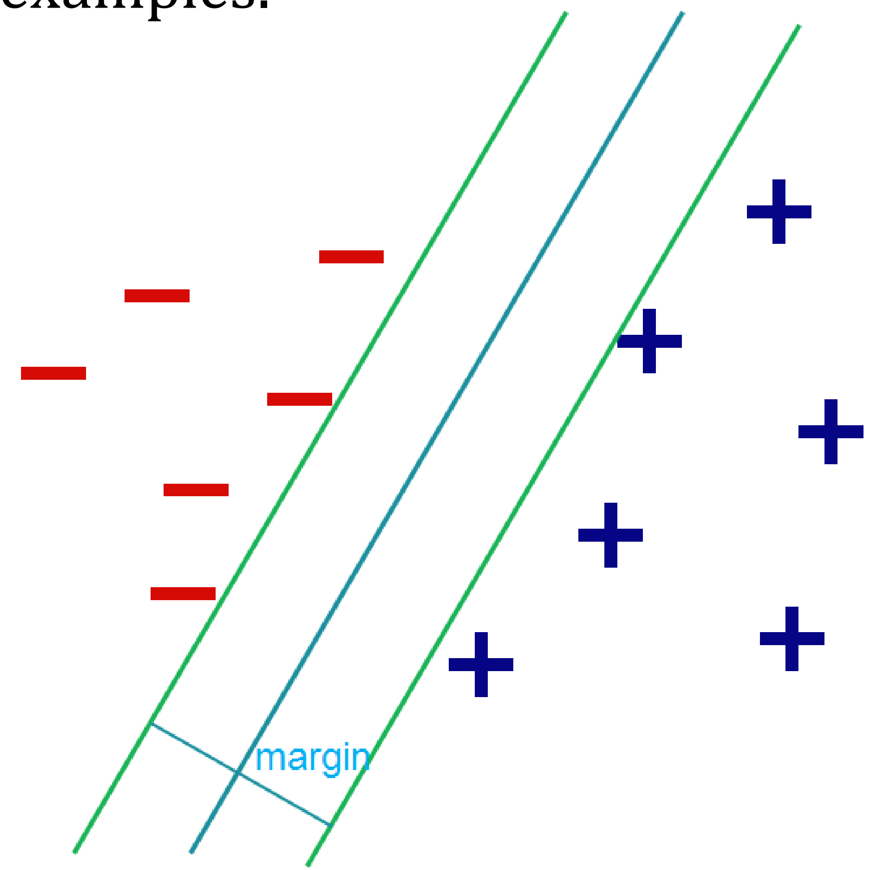
COX-2 Inhibitors

Main Assumption

Patients with an increased risk of MI due to treatment with COX-2 inhibitors are directly analogous to **Persuadables**.

Support Vector Machines

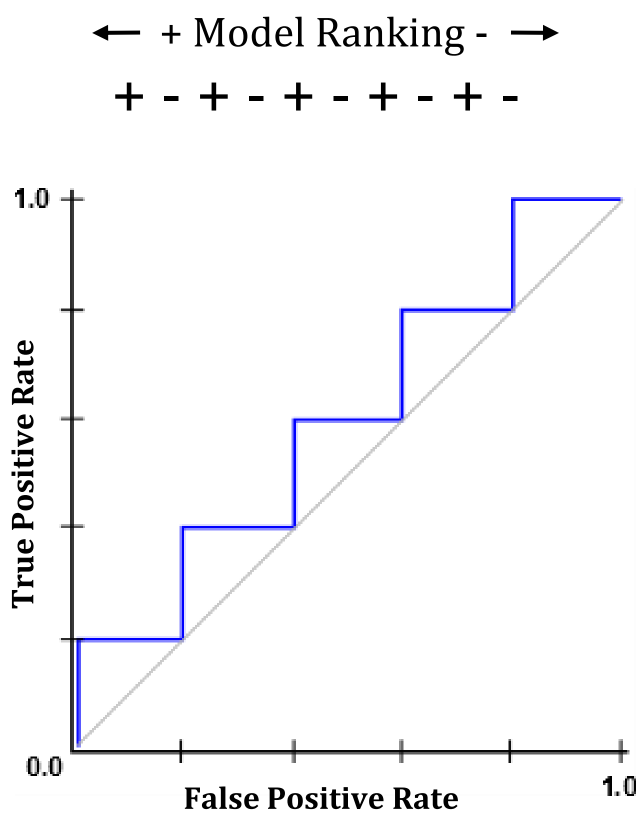
Find maximum-margin separating plane between positive and negative examples.



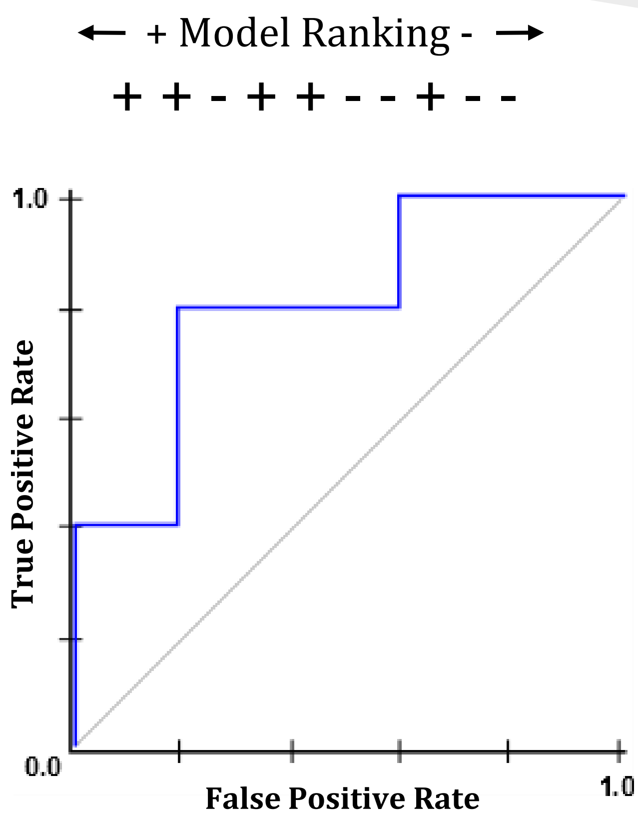
SVM for Uplift

Extend previous SVM work maximizing AUC (Joachims, 2005) to maximize AUU instead.

ROC and AUC



+ Model Ranking -
+ - + - + - + - + -
1.0
True Positive Rate
0.0
False Positive Rate
1.0



+ Model Ranking -
+ + - + + - - + - -
1.0
True Positive Rate
0.0
False Positive Rate
1.0

SVM for Uplift

Let the positive skew of data be:

$$\pi = \frac{P}{P + N}$$

Then (Tuffery, 2011):

$$AUL = P \times \left(\frac{\pi}{2} + (1 - \pi)AUC\right)$$

SVM for Uplift

$$AUU = AUL_T - AUL_C = P_T \times \left(\frac{\pi_T}{2} + (1 - \pi_T)AUC_T\right) - P_C \times \left(\frac{\pi_C}{2} + (1 - \pi_C)AUC_C\right)$$

$$max(AUU) \equiv max(P_T \times (1 - \pi_T)AUC_T - P_C \times (1 - \pi_C)AUC_C)$$

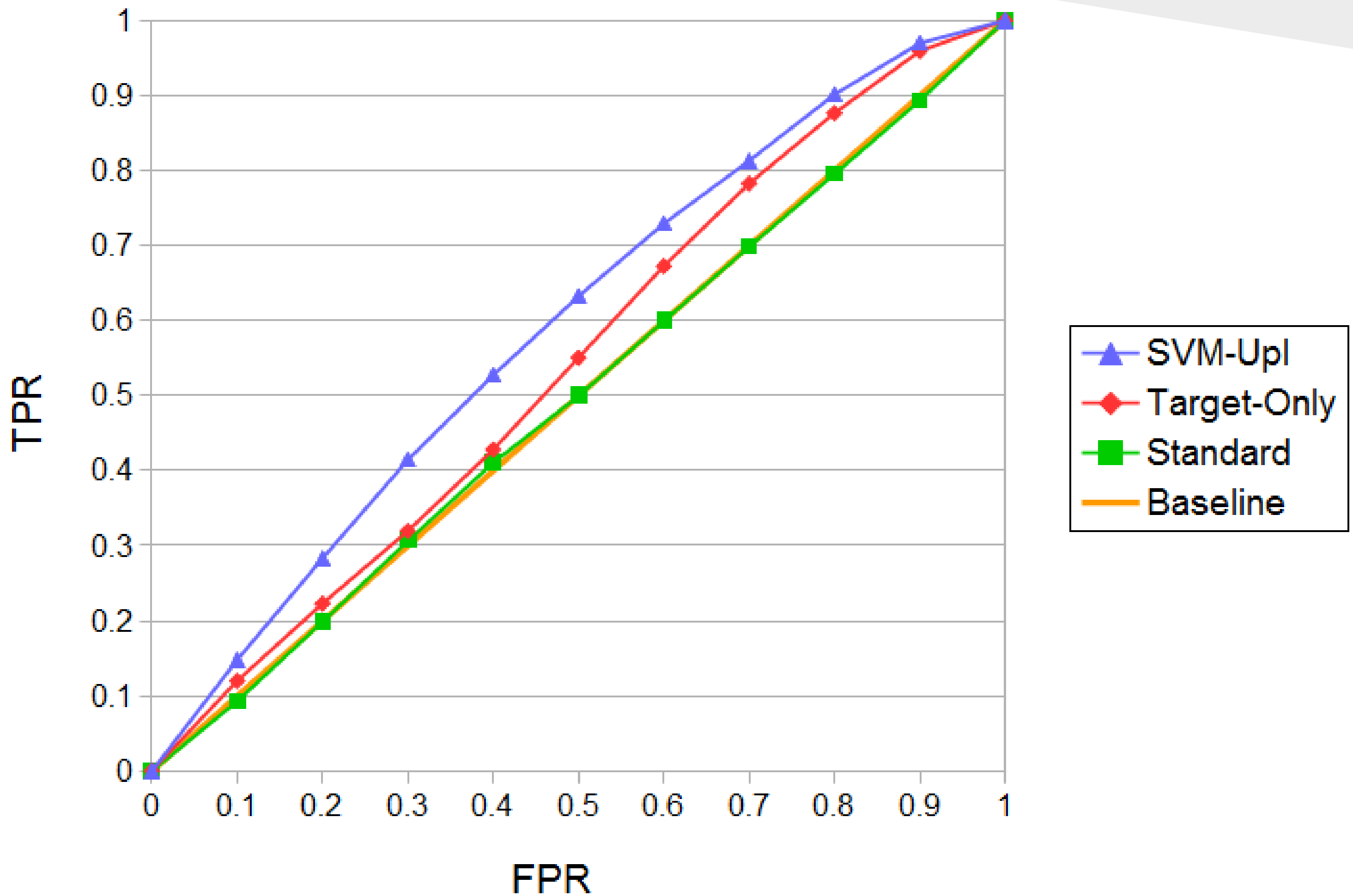
$$\propto max\left(AUC_T - \underbrace{\frac{P_C \times (1 - \pi_C)}{P_T \times (1 - \pi_T)}}_{\lambda}AUC_C\right)$$

$$\boxed{max(AUU) \equiv max(AUC_T - \lambda AUC_C)}$$

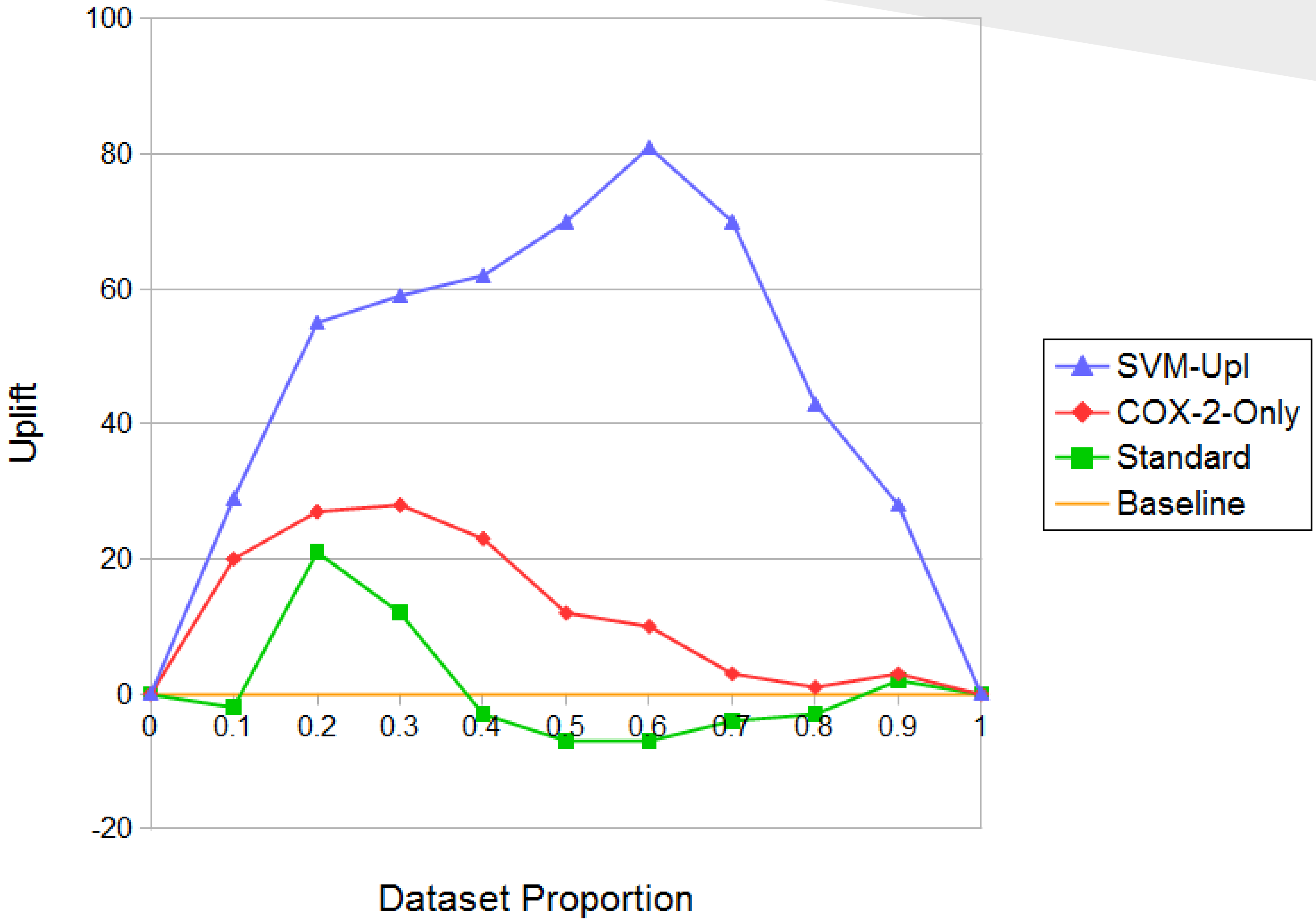
Uplift Modeling Simulation: Persuadable ROC

- Generated synthetic customer population
- Subjected customer population randomly to simulated marketing activity
- Measured ROC with **Persuadables** as the positive class, others as negative

Uplift Modeling Simulation: Persuadable ROC



COX-2 Inhibitor Results



COX-2 Inhibitor Results

Model	AUU	COX-2 AUL	No COX-2 AUL	AUU p-value
SVM^{Upl}	50.7	123.4	72.7	-
COX-2-Only	13.8	151.5	137.7	**0.002***
Standard	1.2	147.7	146.5	**0.002***
Baseline	0.0	0.0	0.0	**0.002***

Outline

- Introduction
- Advice-Based Learning Framework
- Support Vector Machines for Uplift Modeling
- Conclusions

Contributions

In This Presentation

- Developed framework for collaboration between clinicians and machine learning experts to address challenges in decision support (Kuusisto et al., 2015)
- Developed support vector machine for uplift modeling to address COX-2 inhibitor treatment and understand indolent breast cancer in older patients (Kuusisto et al., 2014)

Contributions

In This Presentation

- Developed framework for collaboration between clinicians and machine learning experts to address challenges in decision support (Kuusisto et al., 2015)
- Developed support vector machine for uplift modeling to address COX-2 inhibitor treatment and understand indolent breast cancer in older patients (Kuusisto et al., 2014)

Other Contributions

- Investigated use of machine learning for accurately estimating individualized treatment effects versus traditional approaches with RCT and observational data (Weiss et al., 2015)
- Developed statistical relational uplift modeling algorithm to understand factors contributing to indolent breast cancer in older patients (Nassif et al., 2013)
- Applied inductive logic programming with rule evaluation function tailored to meet clinical objective (Kuusisto et al., 2013)

Overall Conclusions

- Close collaboration with clinicians is essential to develop models to meet clinical objectives
- Leveraging clinical expertise in model-building can alleviate challenges of gathering sufficient data for rare diseases
- Machine learning and uplift modeling have potential applications in treatment assignment and knowledge discovery

Acknowledgements

Advisors:	Jude Shavlik, David Page
Committee:	Elizabeth Burnside, Vitor Santos Costa, Charles Dyer
Collaborators:	Ines Dutra, Houssam Nassif, Yirong Wu
Funding:	NLM R01LM010921, NIH R01CA165229, NIH R01LM011028, NIGMS R01GM097618
Friends:	Eric Lantz, Aubrey Barnard, Alex Cobian, Kendrick Boyd, Jeremy Weiss, Jie Liu, Brandon Smith, Sarah Edlund
Family:	Maggie Kuusisto, Larry Kuusisto, Elina Kuusisto

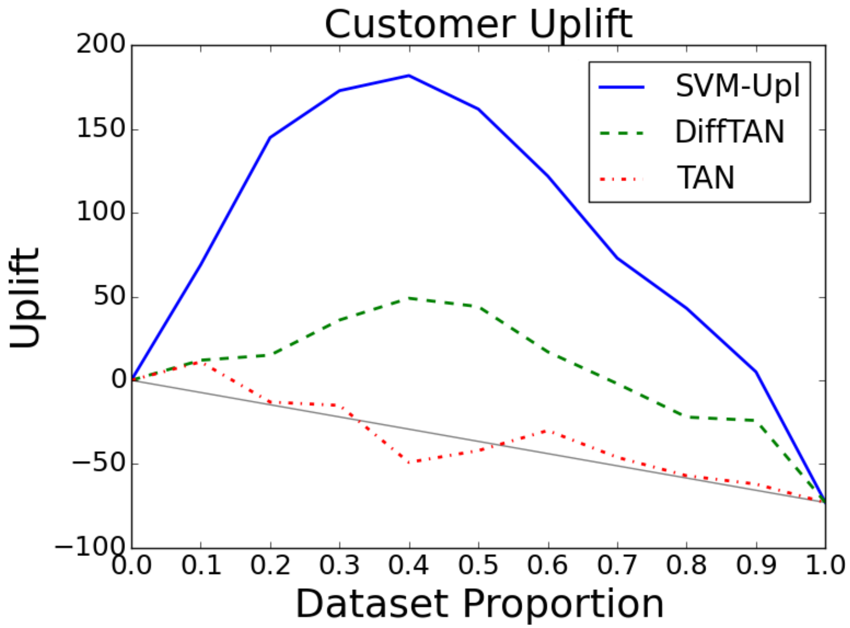
Thank You!

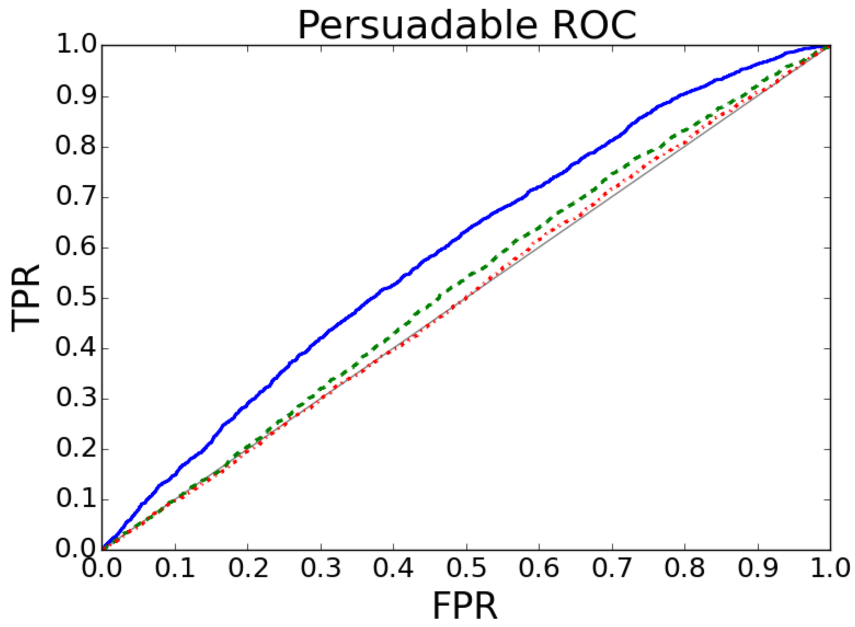
Future Directions

Uplift Bayesian Networks

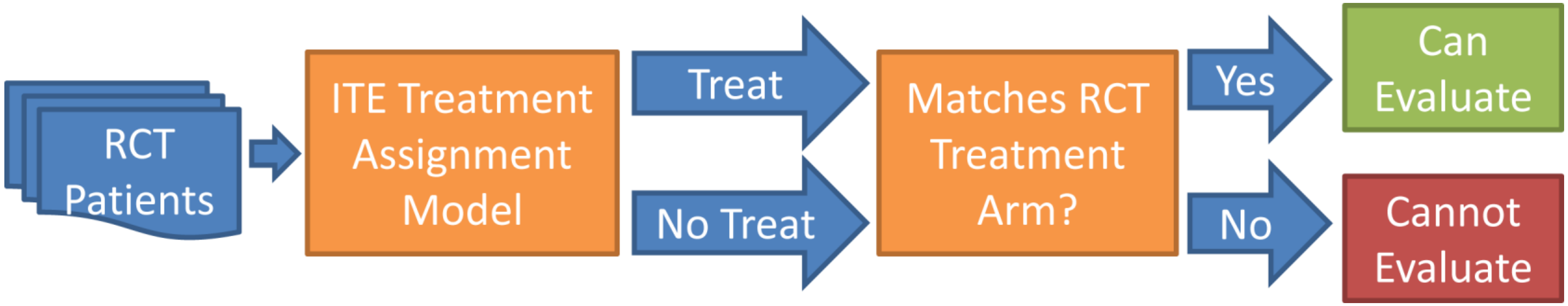
<u>Uplift TAN</u>

$$I_{DIFF}(A;B|Class) = I_{treat}(A;B|Class) - I_{control}(A;B|Class)$$





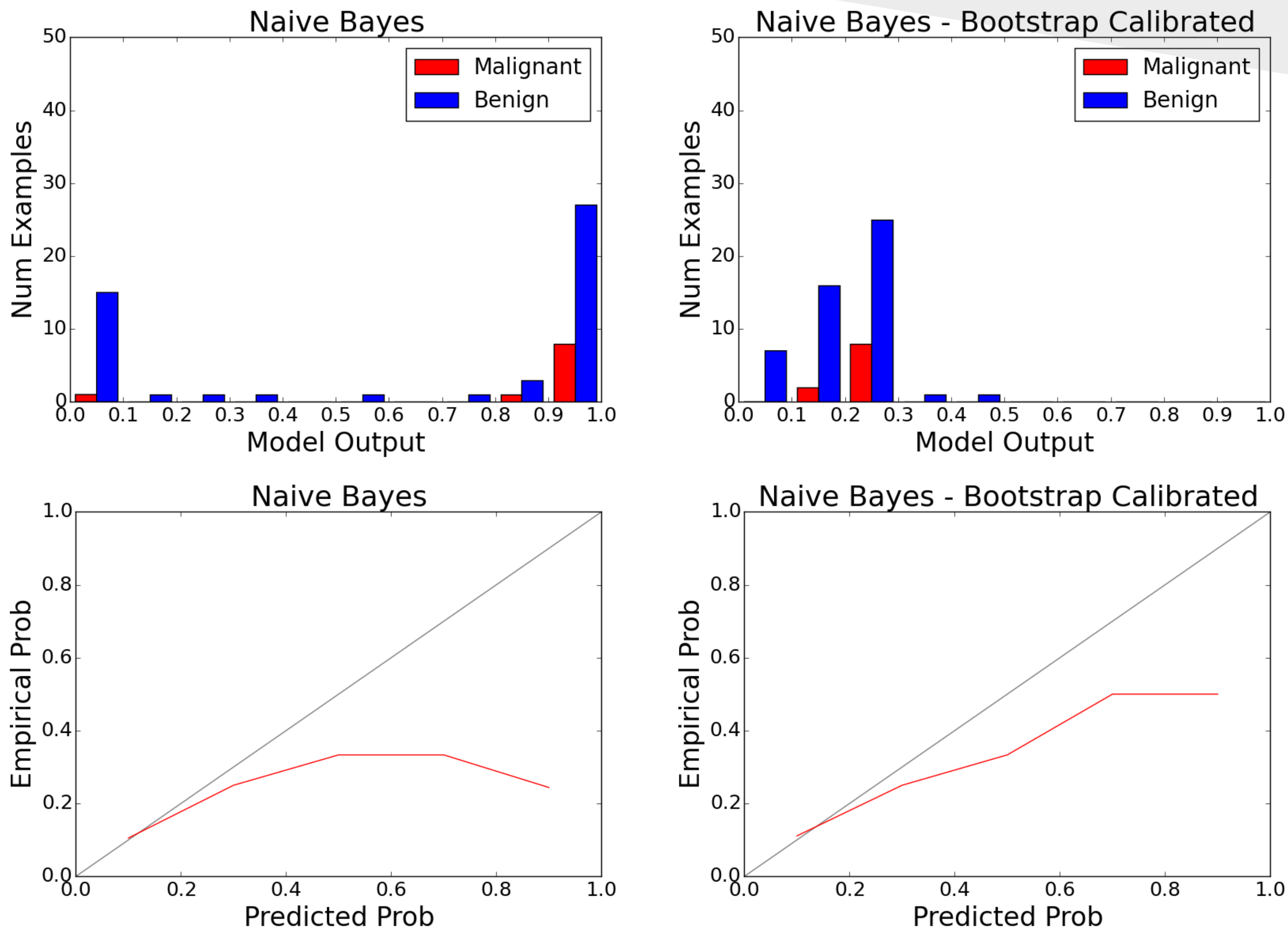
Net Benefit Maximization



- Can evaluate treatment assignment model on RCT data (Vickers et al., 2007)

- Could optimize for treatment assignment directly

Model Calibration



Other Work

Breast Cancer States

In Situ

- Earlier state
- Cancer localized

Invasive

- Later state
- Cancer has invaded surrounding tissue

Breast Cancer Age Differences

Older

- Cancer tends to progress ***less aggressively***
- Patient has ***less*** time for progression

Younger

- Cancer tends to progress ***more aggressively***
- Patient has ***more*** time for progression

Uplift SVM Older In Situ Rules

10 = Clinically Interesting 1 = Clinically Counter-Intuitive

Rank	Feature	Older In Situ Correlation	Radiologist Assessment
1	Linear Calc. Distribution Present	Positive	10
2	Spiculated Mass Margin Present	Negative	10
3	Palpable Lump Present	Positive	3
4	Irregular Mass Shape Present	Negative	9-10
5	No Family History	Negative	8

Upgrade Rules

Use F-score to learn precise rules to predict benign non-definitive biopsies

Algorithm Rule Learning Procedure

for $\mathrm{Train}, \mathrm{Test} \in \mathrm{Folds}$ **do**

$\quad \mathrm{Theory} \leftarrow \mathrm{Aleph}(\mathrm{Train}, \mathrm{minpos} = 2, \mathrm{noise} = 0, \mathrm{evalfn} = F_\beta);$

$\quad \mathrm{Rule}^* \leftarrow \mathrm{argmax}\ F_\beta(\mathrm{Theory}, \mathrm{Train});$

$\quad \mathrm{Evaluate}(\mathrm{Rule}^*, \mathrm{Test});$

end for

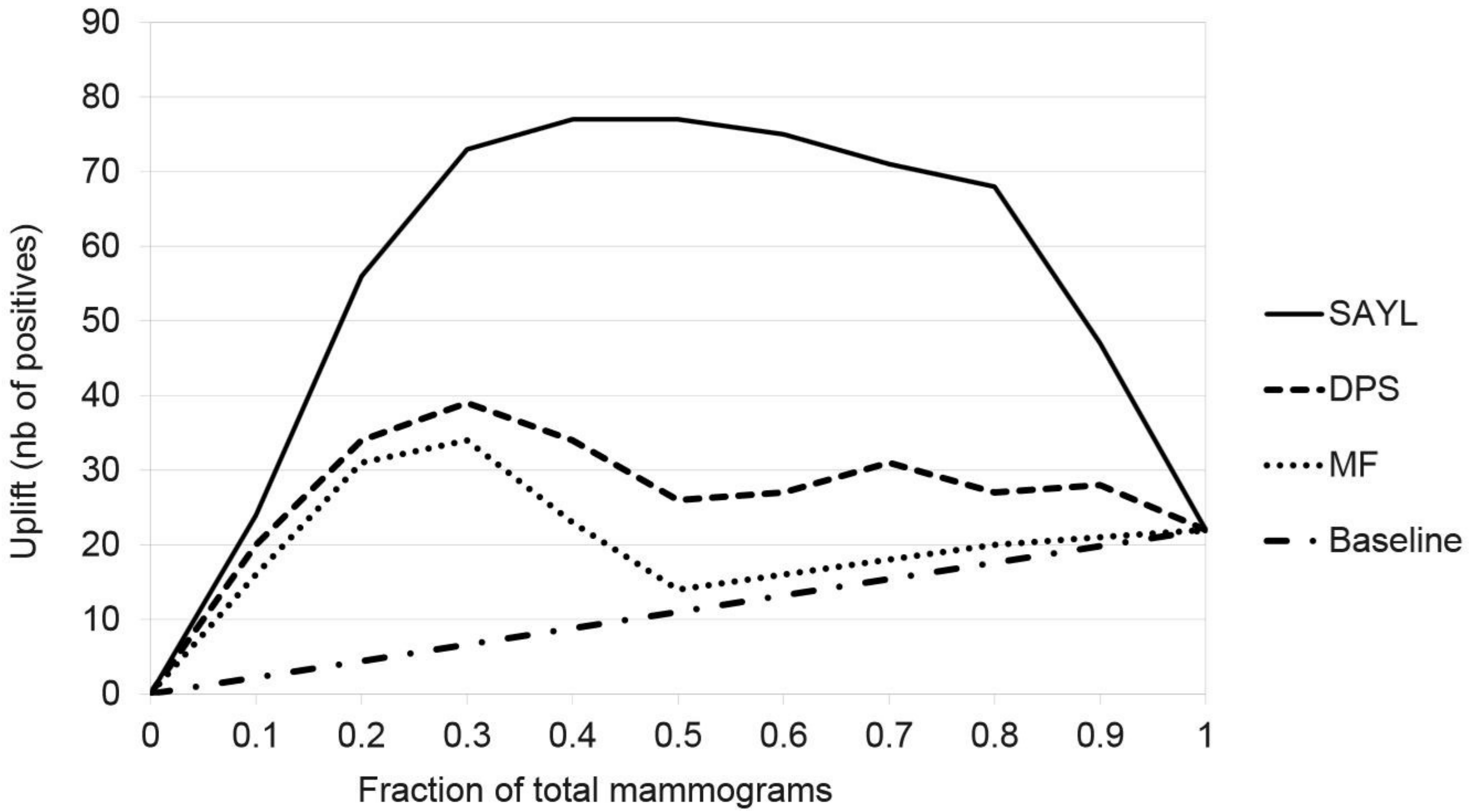
Upgrade Rules

	Rule	Benign Avoided	Malignant Missed
1	The patient did not have a previous surgery, imaging did not present a spiculated mass margin, and the abnormality did not disappear in post-biopsy imaging	30	0
2	Imaging did not present an indistinct mass margin, imaging did not present a spiculated mass margin, and the abnormality did not disappear in post-biopsy imaging	29	0
3	Imaging did not present a spiculated mass margin, and the abnormality did not disappear in post-biopsy imaging	34	1
4	Imaging did not present an indistinct mass margin, and the abnormality did not disappear in post-biopsy imaging	31	1
5	The patient has no personal history of breast cancer, and the abnormality did not disappear in post-biopsy imaging	28	0

SAYL

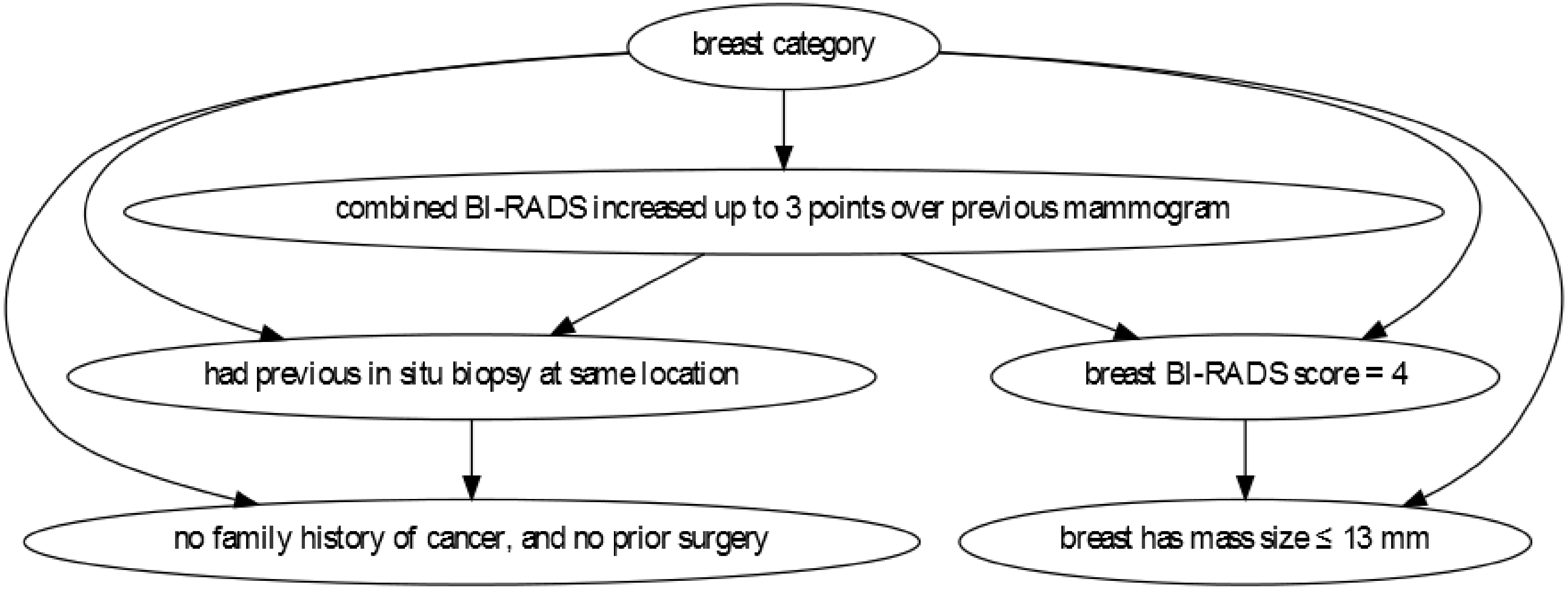
Use ILP to induce feature set used by BN that maximizes uplift.

SAYL

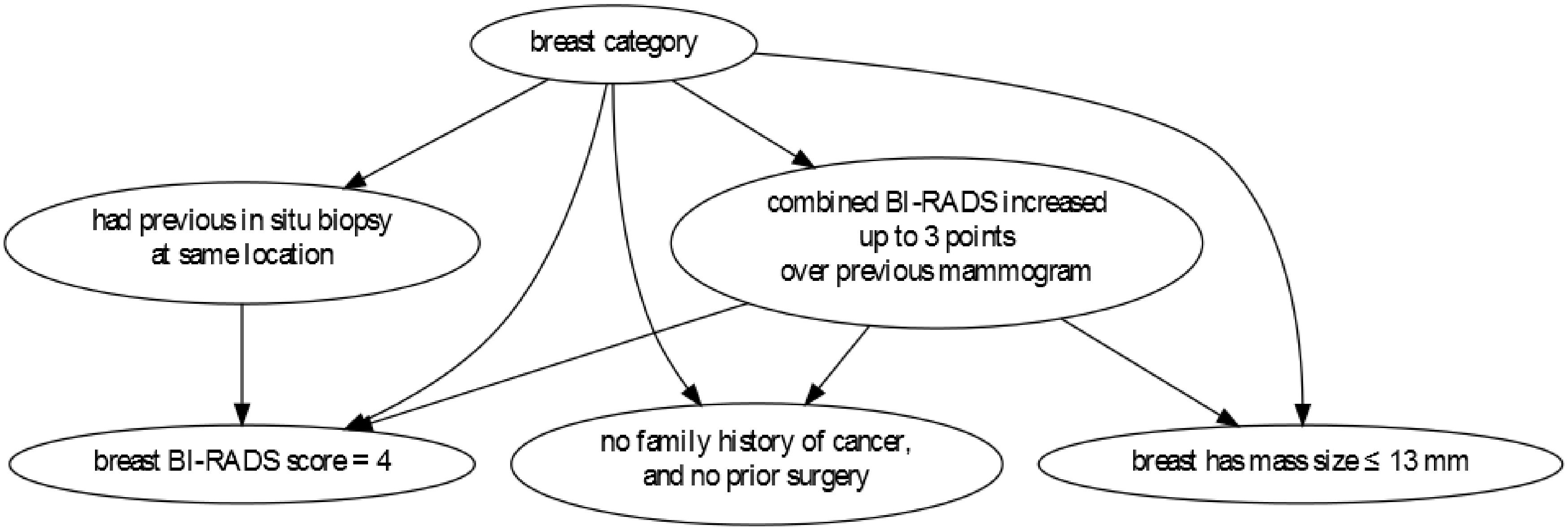


90
80
70
60
50
40
30
20
10
0
Uplift (nb of positives)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Fraction of total mammograms
SAYL
DPS
MF
Baseline

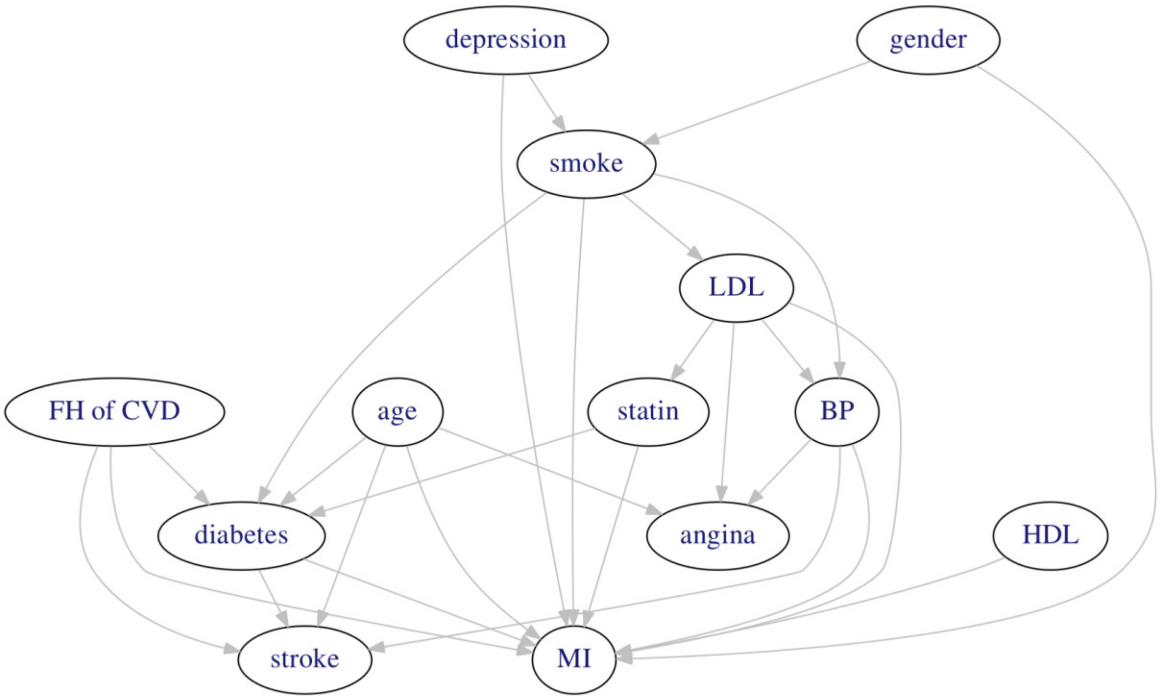
SAYL - Older Model



SAYL - Younger Model



Individualized Treatment



Variable	Obs.	RCT
age (older)	55	55
smoke	28	28
gender (male)	48	48
HDL	27	27
LDL	39	39
diabetes	39	42
family history of CVD	27	27
blood pressure	30	30
history of angina	35	35
history of stroke	5	6
history of depression	27	27
statin use	25	50
MI	8	9

Individualized Treatment