Stochastic dynamic optimization under ambiguity

Lauren N. Steimle

Department of Industrial and Operations Engineering

University of Michigan

Optimal sequential decision-making under uncertainty

Finance

Inventory management

Machine maintenance

Medical decision making

Prevention of heart disease involves balancing benefits and harms of treatment

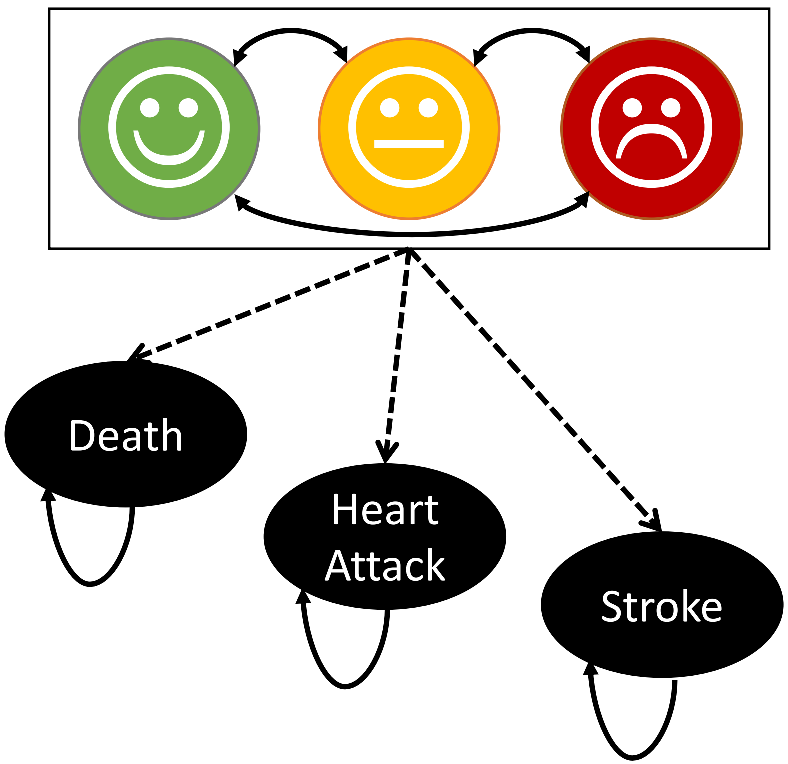
Uncertain Future Benefits

- Delay the onset of potentially deadly and debilitating heart attacks and strokes

Immediate harms

- Side effects (e.g., muscle pain, frequent urination)

Markov decision processes generalize Markov chains to incorporate decisions



Health states
- Blood pressure levels
- Cholesterol levels
- Current medications

Steimle, L. N., & Denton, B. T. (2017). Markov decision processes for screening and treatment of chronic diseases. In *Markov Decision Processes in Practice* (pp. 189-222). Springer, Cham.

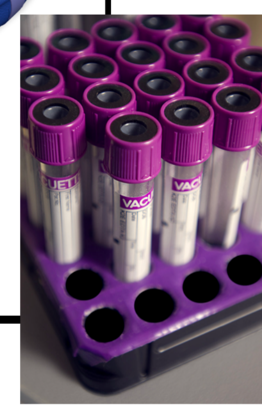
Markov decision processes can improve sequential decision making under uncertainty

Markov Process

$P(a_t)$

State

s_t



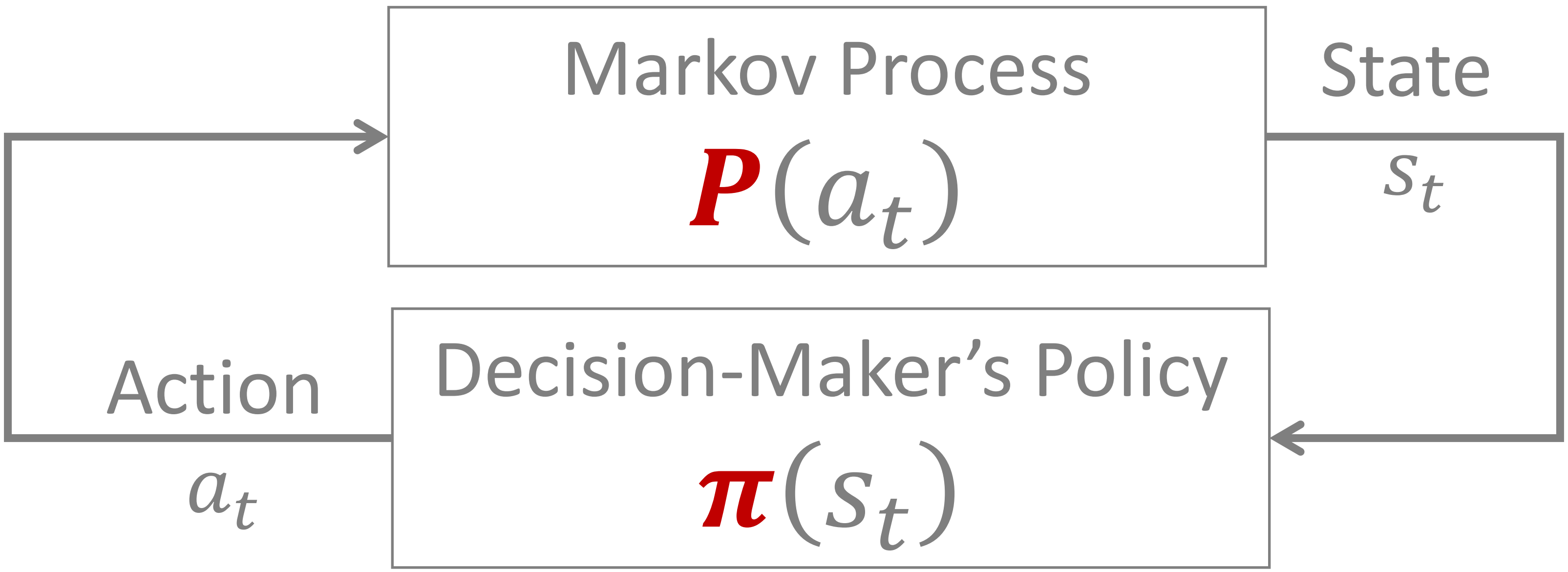
Action

a_t

Decision-Maker's Policy

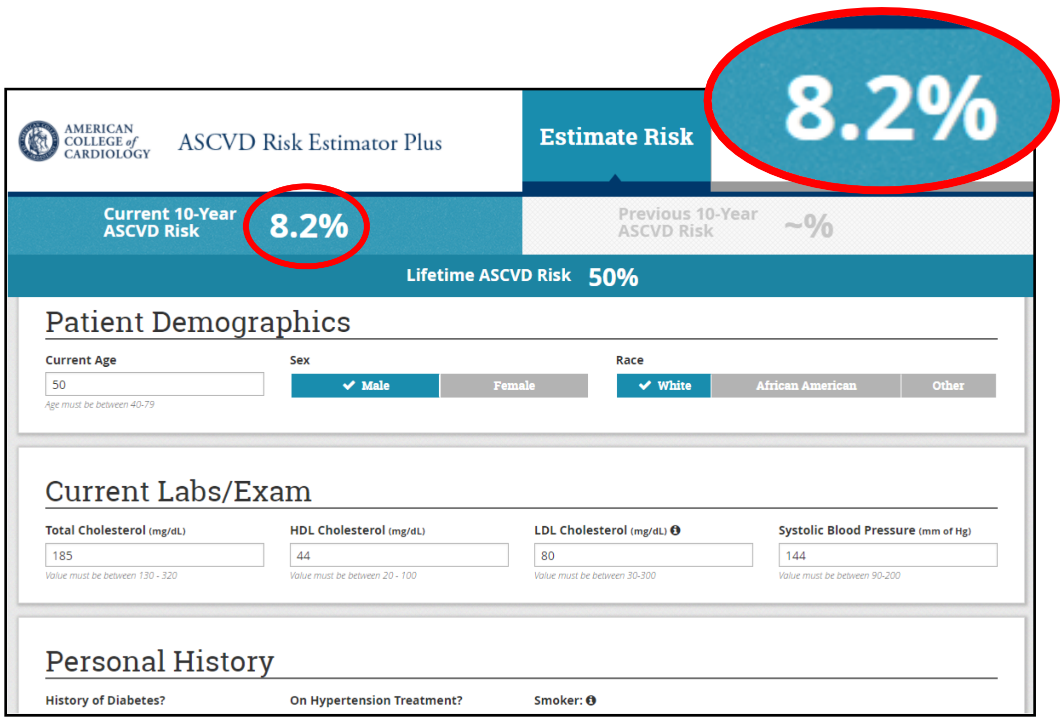
$\pi(s_t)$

Markov decision processes can improve sequential decision making under uncertainty



$$\max_{\pi \in \Pi} \left\{ \mathbb{E}^{\pi, P} \left[\sum_{t=1}^{T} r_t(s_t, a_t) + r_{T+1}(s_{T+1}) \right] \right\}$$

Clinical risk calculators are used to estimate a patient's risk



Inputs:

- Age
- Sex
- Race
- Total Cholesterol
- HDL Cholesterol
- LDL Cholesterol
- Systolic Blood Pressure
- History of Diabetes
- On Hypertensive Treatment
- Smoker

Output:

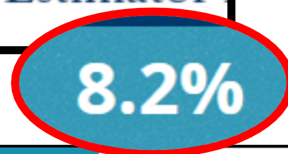
Current 10-Year Risk

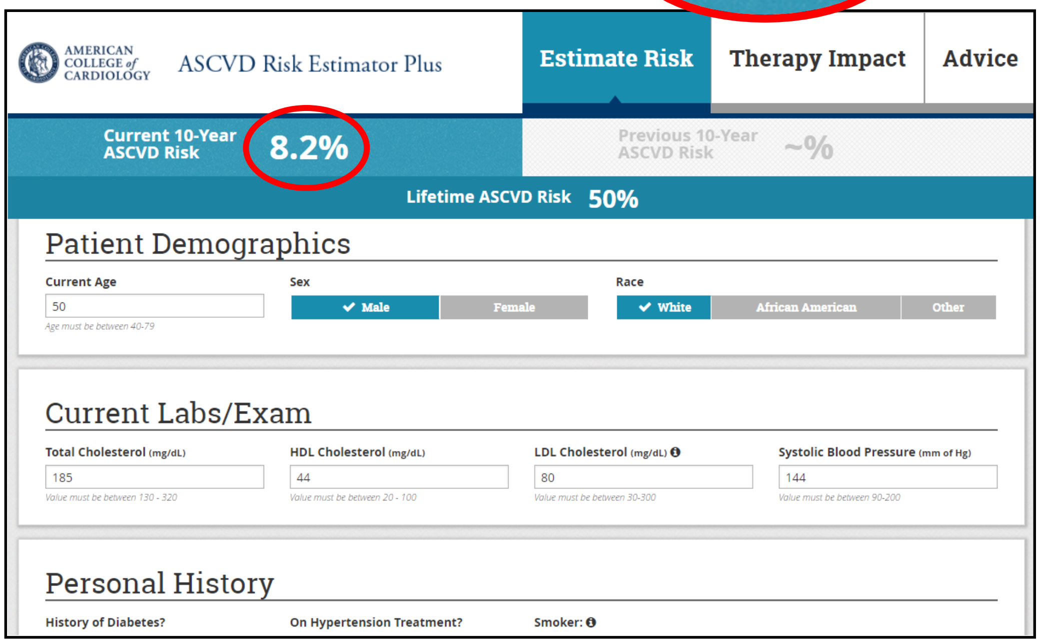
2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2014

Well-established clinical studies give conflicting estimates about CVD risk

AMERICAN COLLEGE of CARDIOLOGY ASCVD Risk Estimator

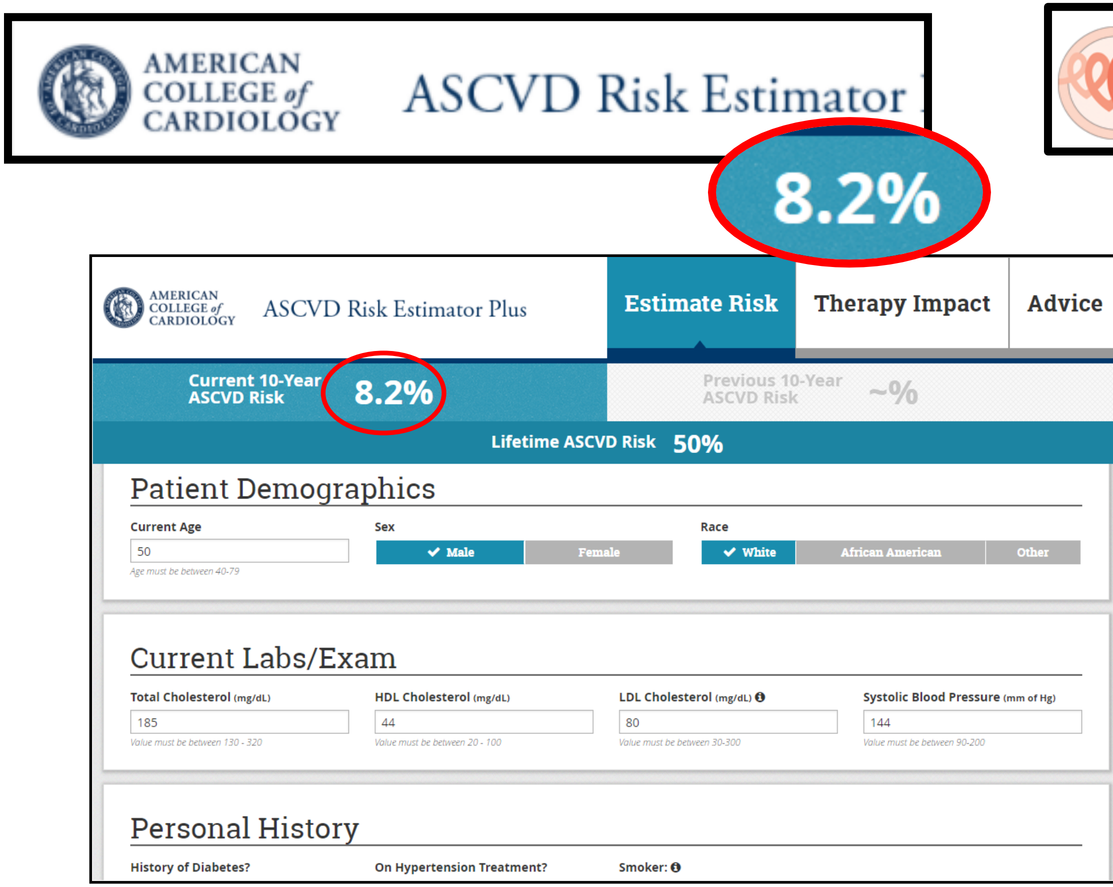
Framingham Heart Study
A Project of the National Heart, Lung, and Blood Institute and Boston University

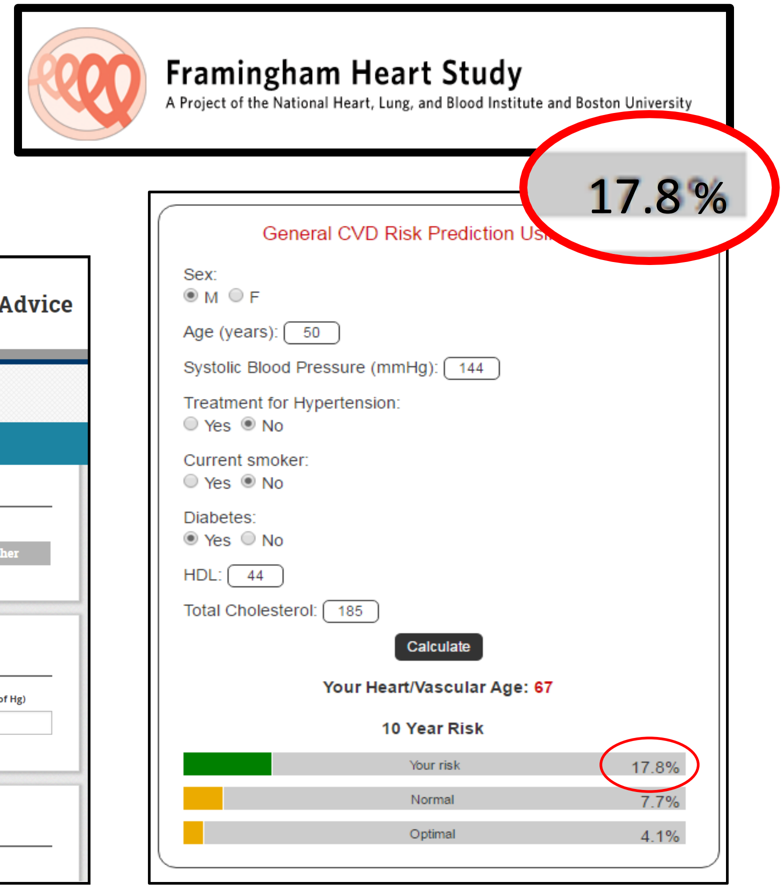




1 Wilson et. al. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation.* 1998
Wolf et. al. Probability of stroke: a risk profile from the Framingham Study. *Stroke.* 1991
2 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2014

Well-established clinical studies give conflicting estimates about CVD risk





1 Wilson et. al. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation.* 1998
Wolf et. al. Probability of stroke: a risk profile from the Framingham Study. *Stroke.* 1991
2 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2014

Research Questions

How can we improve Markov decision processes to account for ambiguity?

How much benefit is there in doing so?

Stochastic dynamic optimization under ambiguity

Multi-model Markov decision processes

Decomposition methods

Other ambiguity-aware formulations

Stochastic dynamic optimization under ambiguity

Multi-model Markov decision processes

Decomposition methods

Other ambiguity-aware formulations

We have two layers of uncertainty in our problem

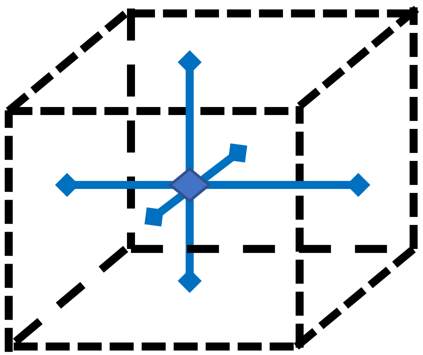
Optimal control of a stochastic system…

- Markov decision processes

…under parameter uncertainty

- Robust optimization
- Stochastic optimization

Robust optimization approach to ambiguity in Markov decision processes



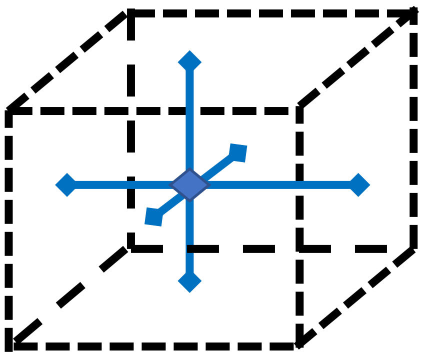
Assume that *P* lies within some *ambiguity set*

e.g., Interval Model

Goal is to maximize worst-case performance

(s,a)-rectangularity property gives a tractable model for MDPs

(s,a)-rectangularity is computationally attractive, but has its drawbacks

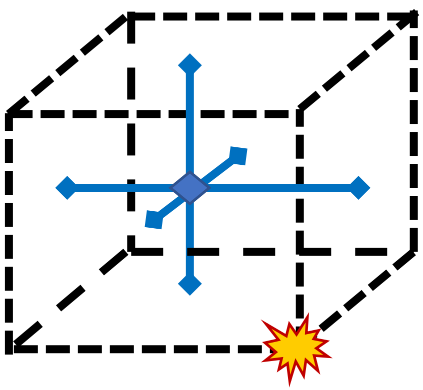


Leads to overly-protective policies

- ➢Optimizing for case where all parameters take on worst-case values simultaneously

Transition matrices might lose known structure

- ➢Ambiguity is realized independently across states, actions, and/or decision epochs



Relaxing (s,a)-*rectangularity* causes max-min problem to be NP-hard*

*Wiesemann, Wolfram, Daniel Kuhn, and Berç Rustem. "Robust Markov decision processes." *Mathematics of Operations Research* 38.1 (2013): 153-183.

The **Multi-model Markov Decision Process** is a new framework for handling ambiguity

Generalizes a Markov decision process

- State space, $\mathcal{S} \equiv \{1, \ldots, S\}$
- Action space, $\mathcal{A} \equiv \{1, \ldots, A\}$
- Decision epochs, $\mathcal{T} \equiv \{1, \ldots, T\}$
- Rewards, $R \in \mathbb{R}^{S \times A \times T}$

Finite set of models, $\mathcal{M} = \{1, \ldots, |\mathcal{M}|\}$

- Model m: An MDP $(\mathcal{S}, \mathcal{A}, \mathcal{T}, R, P^m)$
- Transition probabilities P^m are model-specific

Steimle, L. N., Kaufman, D.L., and Denton B.T. (2018) "Multi-model Markov Decision Processes." *Optimization Online.*

The **weighted value problem** seeks to find a single policy that performs well in each model

Performance of policy π in model m

$$v^m(\pi) = \mathbb{E}^{\pi,P^m}\left[\sum_{t=1}^{T} r_t(s_t, a_t) + r_{T+1}(s_{T+1})\right]$$

Weighted value of policy π

$$W(\pi) = \sum_{m\in\mathcal{M}} \lambda_m v^m(\pi)$$

Weighted value problem

$$W^*(\pi) = \max_{\pi\in\Pi} W(\pi) = \max_{\pi\in\Pi} \sum_{m\in\mathcal{M}} \lambda_m v^m(\pi)$$

We propose exact and approximate solution methods with bounds

Adaptive: Allow for history-dependent policies

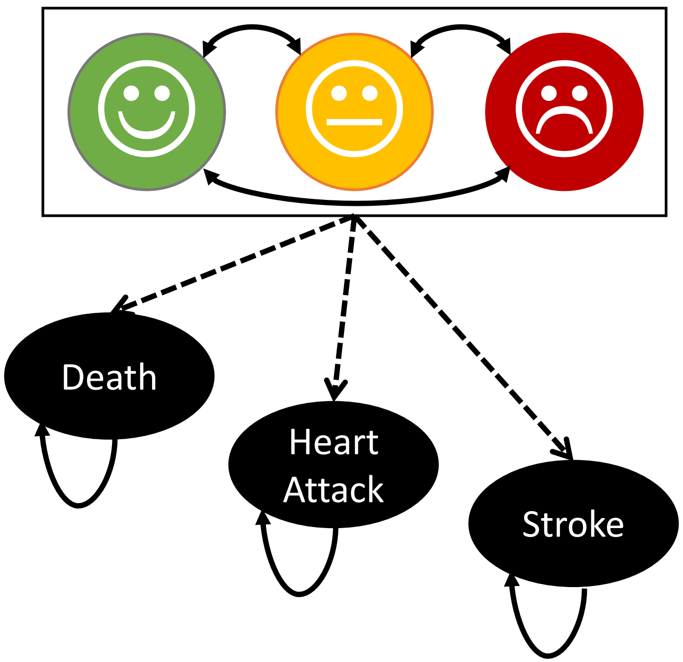
Outer linearization with state-wise pruning

Non-adaptive: Only Markov deterministic policies

Mixed-integer programming (MIP)

Weight-Select-Update (WSU)

We used an approximation algorithm to solve a heart disease management problem



Multi-model Markov decision process

- 4,096 states
- 64 actions
- 20 decision epochs
- 2 models

Case study data

- Longitudinal data from Mayo Clinic
- Framingham, ACC risk calculators
- Disutilities from medical literature

Mason, J. E., Denton, B. T., Shah, N. D., & Smith, S. A. (2014). Optimizing the simultaneous management of blood pressure and cholesterol for type 2 diabetes patients. *European Journal of Operational Research*, *233*(3), 727-738.

We compared our algorithm to policies that ignore ambiguity

Quality-Adjusted Life Years Gained
Over No Treatment, per 1000 Men

Optimal Decisions for FHS Model

MMDP Decisions

Optimal Decisions for ACC Model

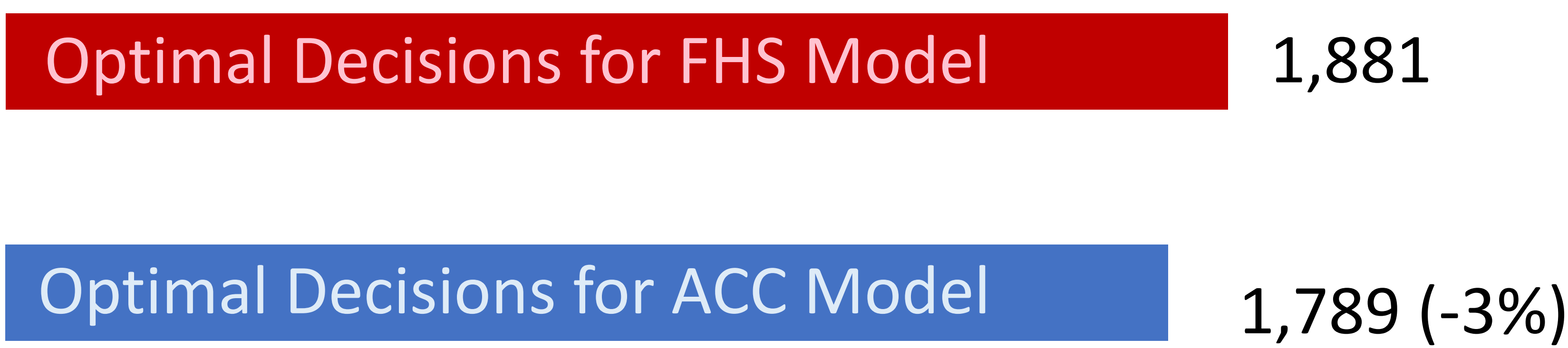
In some cases, ignoring ambiguity has relatively minor implications

Quality-Adjusted Life Years Gained
Over No Treatment, per 1000 Men

Framingham Heart Study Model

In some cases, ignoring ambiguity has relatively minor implications

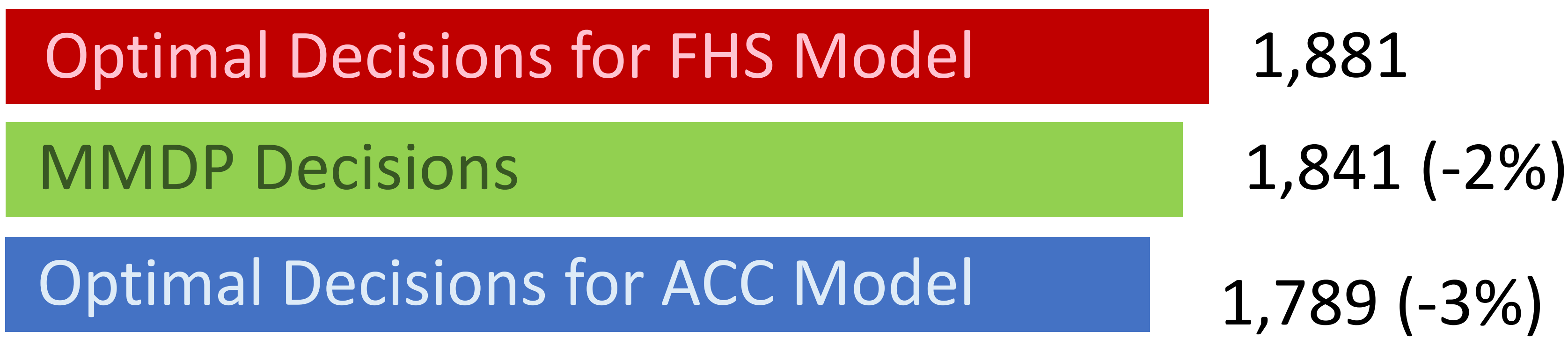
Quality-Adjusted Life Years Gained
Over No Treatment, per 1000 Men



Framingham Heart Study Model

In some cases, ignoring ambiguity has relatively minor implications

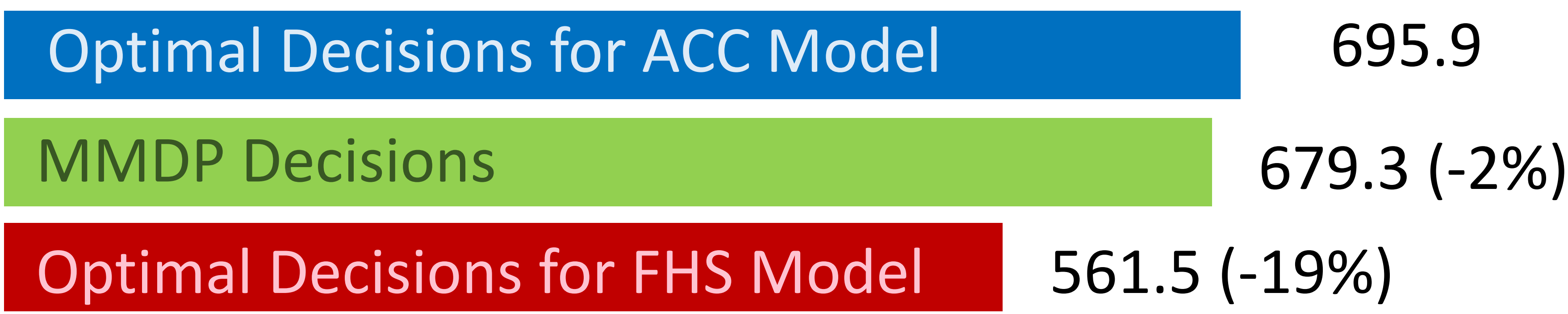
Quality-Adjusted Life Years Gained
Over No Treatment, per 1000 Men



Framingham Heart Study Model

But in other cases, ignoring ambiguity can have major implications

Quality-Adjusted Life Years Gained
Over No Treatment, per 1000 Men



American College of Cardiology Model

Conclusions

The MMDP allows for multiple models of stochastic system in the design of policies

The MMDP is difficult to solve computationally

A polynomial-time approximation algorithm can provide near-optimal solutions in many instances

Using a CVD case study, we showed can be important to address ambiguity arising from multiple models

Stochastic dynamic optimization under ambiguity

Multi-model Markov decision processes

Decomposition methods

Other ambiguity-aware formulations

We have created exact solution methods for solving the weighted value problem

Mixed-integer programming (MIP)

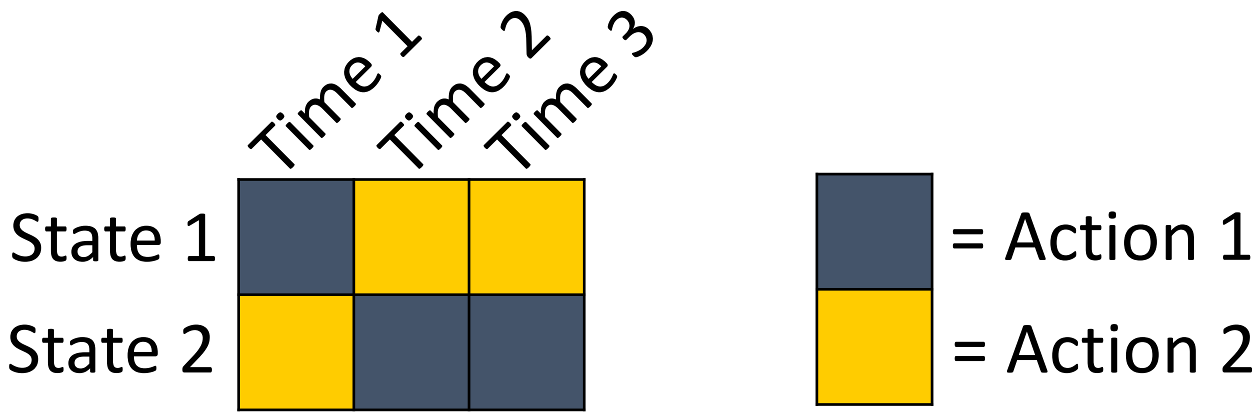
Branch-and-cut

Custom branch-and-bound

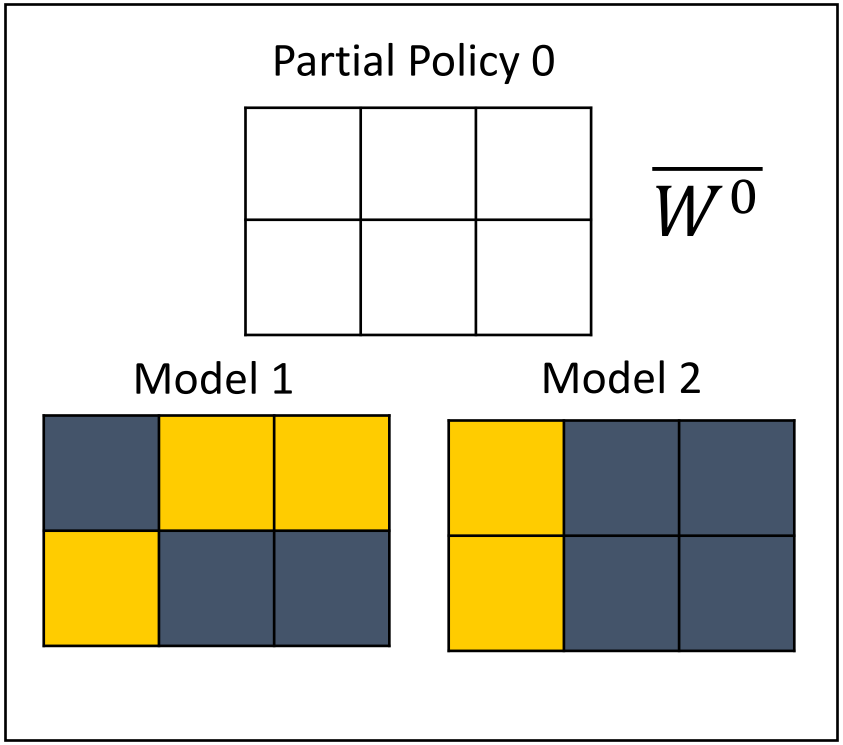
Branch-and-bound works towards finding policies that match across all models

Relax requirement that policy must be same in each model

Goal: Find an *implementable policy* (policy is the same in all models) that maximizes weighted value



B&B begins by solving each model independently

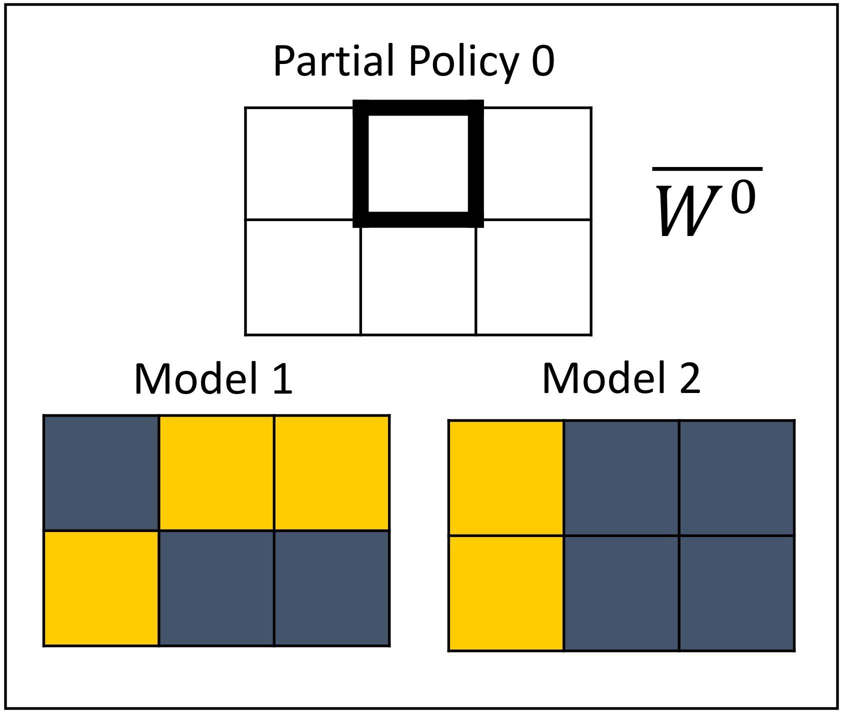


No actions have been fixed in this partial policy

Each model solved independently via backwards induction

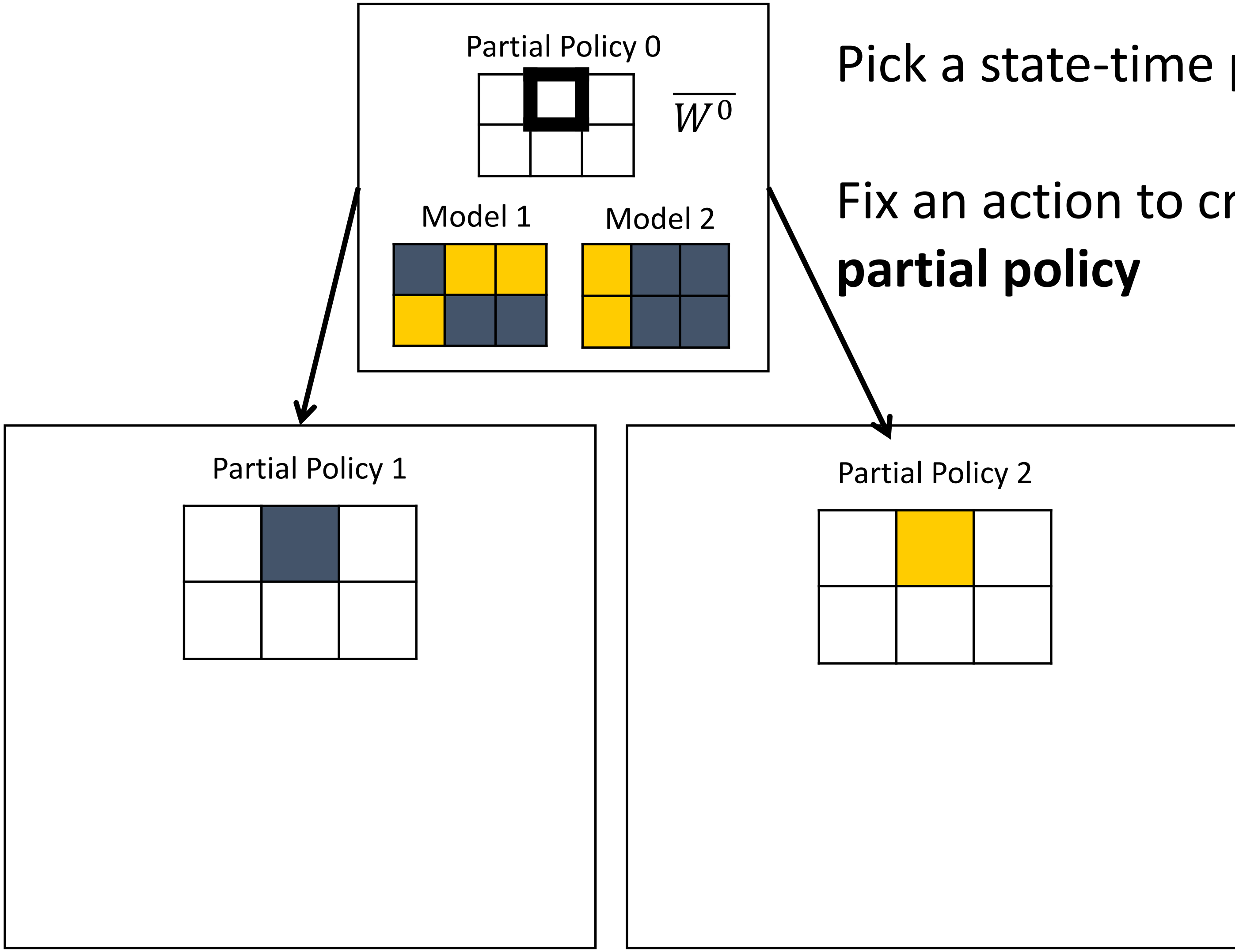
Gives an upper bound $\overline{W^0}$

B&B proceeds by fixing a part of the policy that must match in all models



Pick a state-time pair to branch on

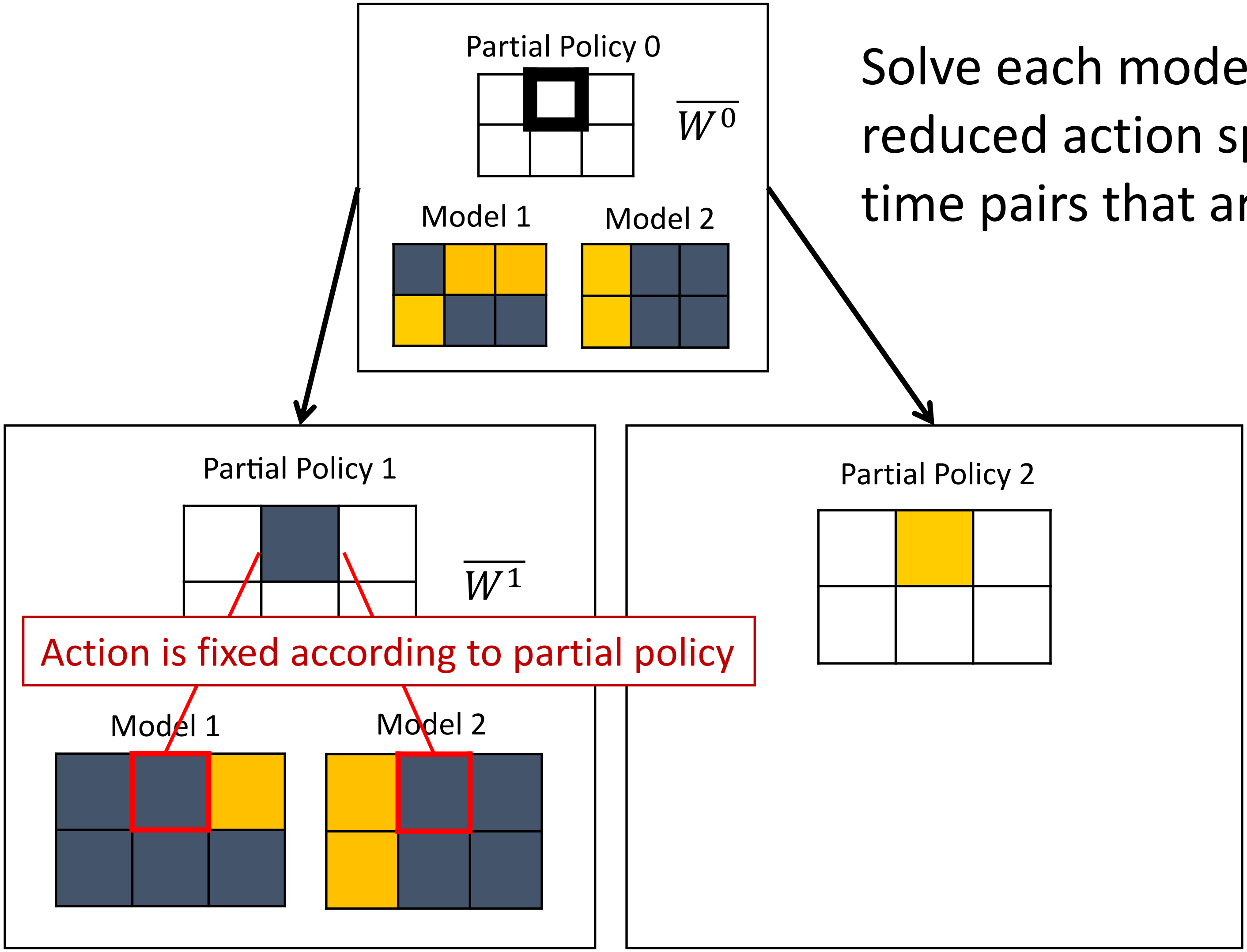
B&B proceeds by fixing a part of the policy that must match in all models



Pick a state-time pair to branch on

Fix an action to create add to the **partial policy**

B&B solves a relaxation using backward induction to obtain upper bound

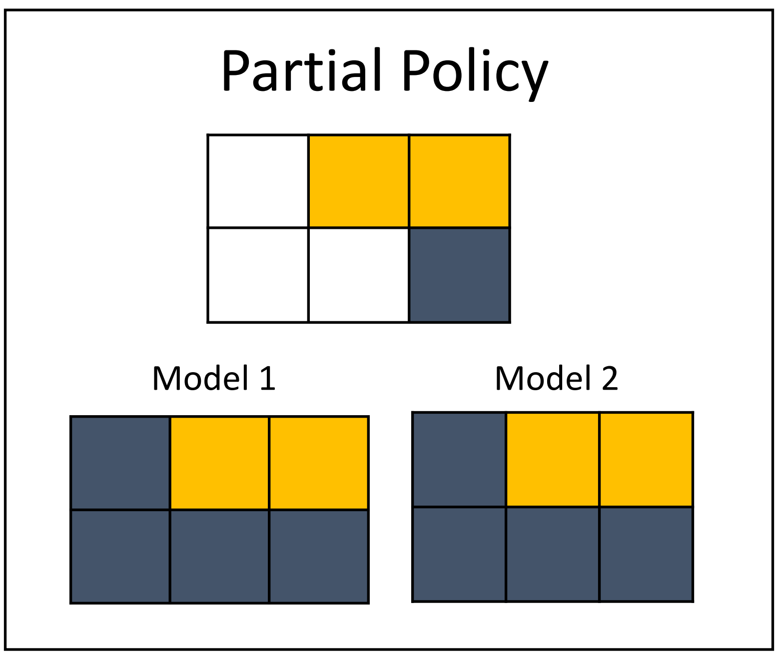


Solve each model's MDP with reduced action space for state-time pairs that are fixed

Pruning eliminates the need to explore all possible policies

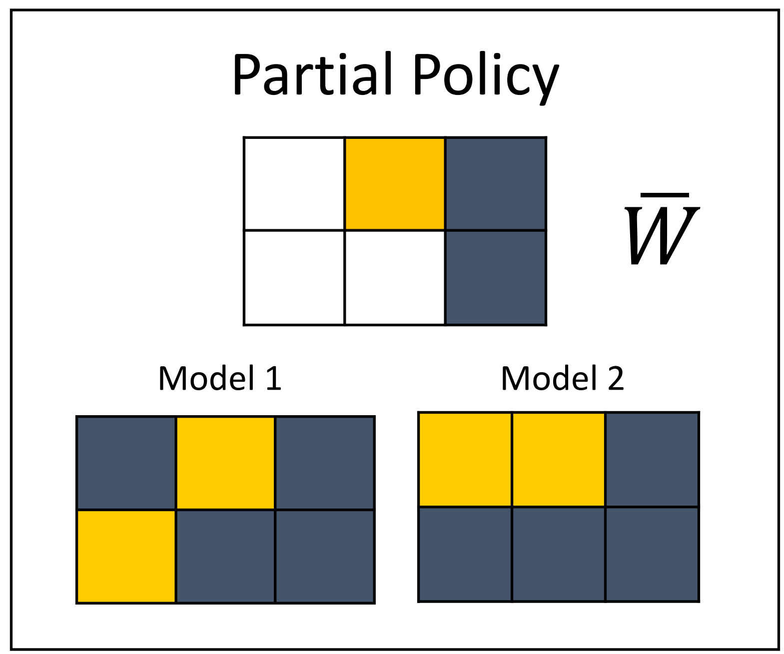
<u>Prune by optimality</u>

Solving the relaxation gives an *implementable policy*



<u>Prune by bound</u>

The incumbent is better than any possible completion of the partial policy



We compared 3 exact methods on 240 instances of MMDPs

Solution Method	Implementation	% solved in 5 minutes?	Optimality Gap (avg.)
MIP Extensive Form	Gurobi		
MIP Branch-and-cut	Gurobi with Callbacks		
Branch-and-Bound	Custom code in C++		

[1] Steimle, L. N., Ahluwalia, V., Kamdar, C., and Denton B.T. (2018) “Decomposition methods for solving Multi-model Markov decision processes.” *Optimization Online.*

[2] Gurobi Optimization, LLC (2018) “Gurobi Optimizer Reference Manual", http://www.gurobi.com

Our custom branch-and-bound approach is the fastest of the solution methods

Solution Method	**Implementation**	**% solved in 5 minutes?**	**Optimality Gap (avg.)**
MIP Extensive Form	Gurobi	0%	12.2%
MIP Branch-and-cut	Gurobi with Callbacks	0%	13.1%
Branch-and-Bound	Custom code in C++	**97.9%**	**1.11%**

Conclusions

A custom branch-and-bound approach outperforms MIP-based solution methods

MMDPs tend to be harder to solve when there is more variance in the models' parameters

In many cases, the mean value problem provides an optimal or near-optimal solution.

Stochastic dynamic optimization under ambiguity

Multi-model Markov decision processes

Decomposition methods

Other ambiguity-aware formulations

So far, we have considered a decision-maker that maximizes expected value

Value of policy π in model m

$$v^m(\pi) = \mathbb{E}^{\pi,P^m}\left[\sum_{t=1}^{T} r_t(s,a) + r_{T+1}(s)\right]$$

Weighted value problem maximizes <u>expectation</u> of model performance

$$W^*(\pi) = \max_{\pi \in \Pi^{MD}} \{\mathbb{E}^{\mathcal{M}}[v^m(\pi)]\}$$

What if the decision-maker wants to protection against undesirable outcomes resulting from ambiguity?

We modified the branch-and-bound algorithm to solve other ambiguity-aware formulations

Max-min

$$\max_{\pi \in \Pi^{MD}} \min_{m \in \mathcal{M}} v^m(\pi)$$

Min-max-regret[1]

$$\min_{\pi \in \Pi^{MD}} \max_{m \in \mathcal{M}} \left\{ \max_{\bar{\pi} \in \Pi} v^m(\bar{\pi}) - v^m(\pi) \right\}$$

Percentile optimization[2]

$$\begin{aligned} \max_{z \in \mathbb{R}, \pi \in \Pi^{MD}} \quad & z \\ \text{s.t.} \quad & \mathbb{P}(v^m(\pi) \geq z) \geq 1 - \epsilon \end{aligned}$$

[1] Ahmed A, Varakantham P, Lowalekar M, Adulyasak Y, Jaillet P (2017) Sampling Based Approaches for Minimizing Regret in Uncertain Markov Decision Processes (MDPs). *Journal of Artificial Intelligence Research* 59:229–264
[2] Merakli, M. and Kucukyavuz, S. (2019) "Risk-Averse Markov Decision Processes under Parameter Uncertainty with an Application to Slow-Onset Disaster Relief." *Optimization Online.*

These problems are still NP-hard. We compared to polynomial-time alternatives

Mean Value Problem

$$\max_{\pi \in \Pi^{MD}} \left\{ \mathbb{E}^{\pi,\bar{P}} \left[\sum_{t=1}^{T} r_t(s,a) + r_{T+1}(s) \right] \right\}$$

(s,a)-rectangular finite scenario MDP*

$$\max_{a \in \mathcal{A}} \min_{p_t(s,a) \in \mathcal{P}_t(s,a)} \left\{ r_t(s,a) + \sum_{s' \in \mathcal{S}} p_t(s'|s,a) v_{t+1}(s) \right\}$$

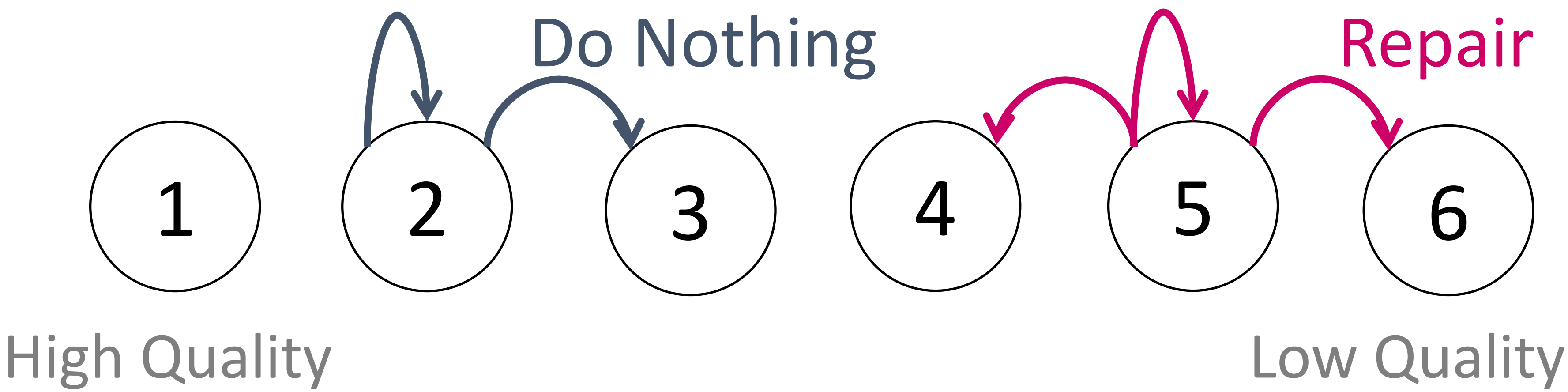
Nilim, Arnab, and Laurent El Ghaoui. "Robust control of Markov decision processes with uncertain transition matrices." *Operations Research* 53.5 (2005): 780-798.

We compared these formulations in two case studies

Machine maintenance

Cardiovascular disease management

Machine maintenance:
Optimal timing of machine repairs



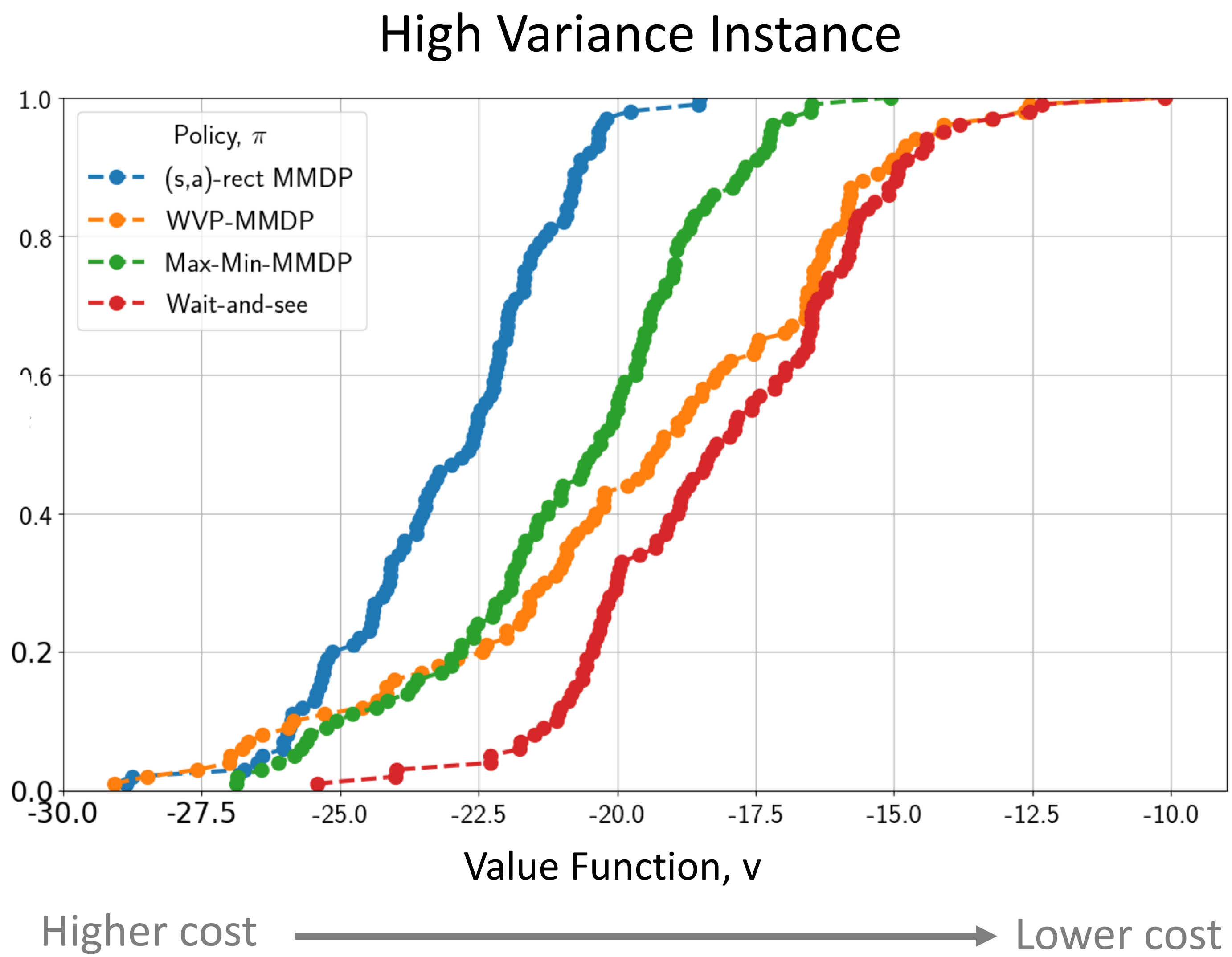
Operating costs depend on quality of machine

Options:

- Do Nothing at no cost
- Minor repair at low cost
- Major repair at high cost

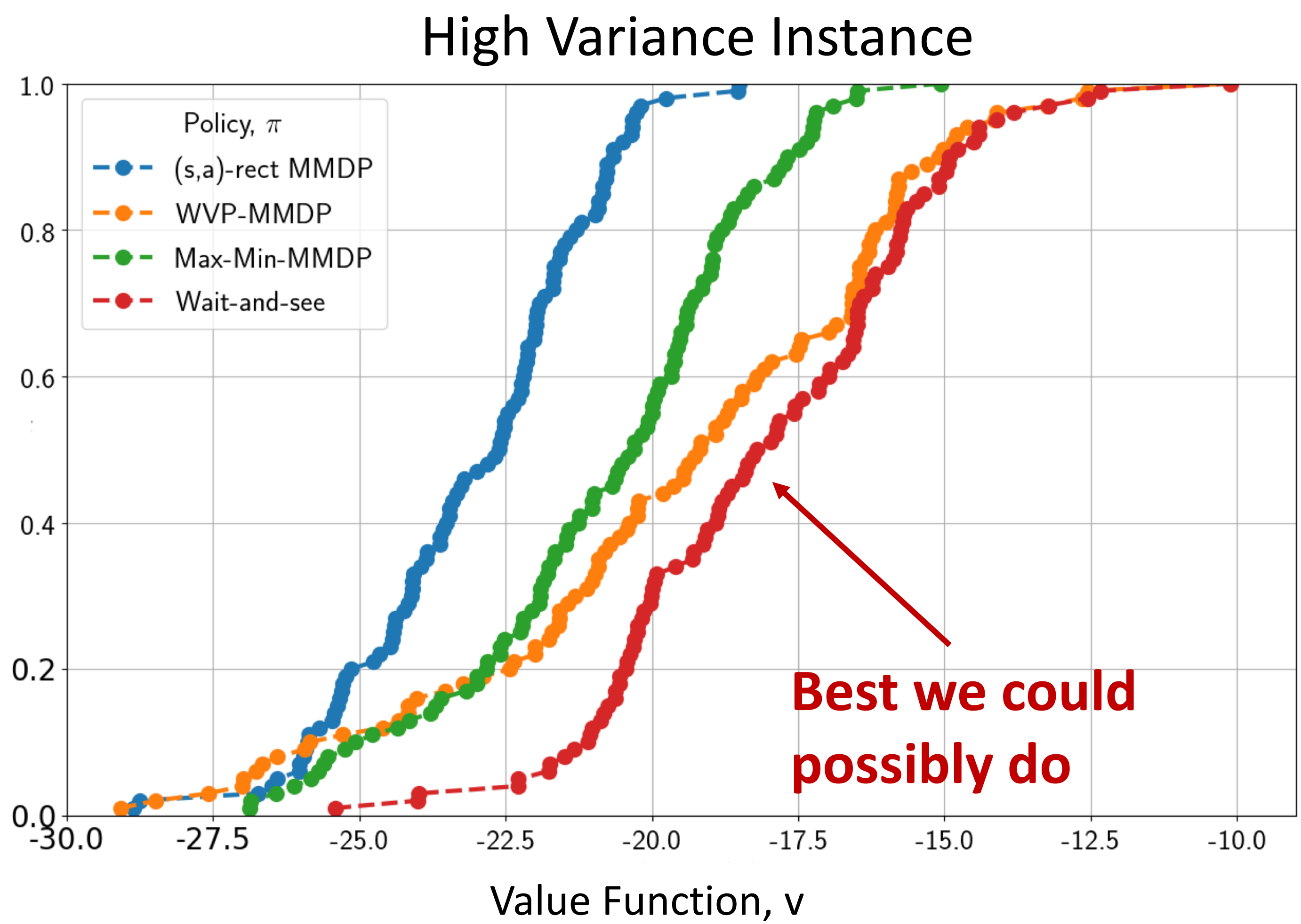
The measure of protection against can distribution of performance among models

$\mathbb{P}(v^m(\pi) \leq v)$



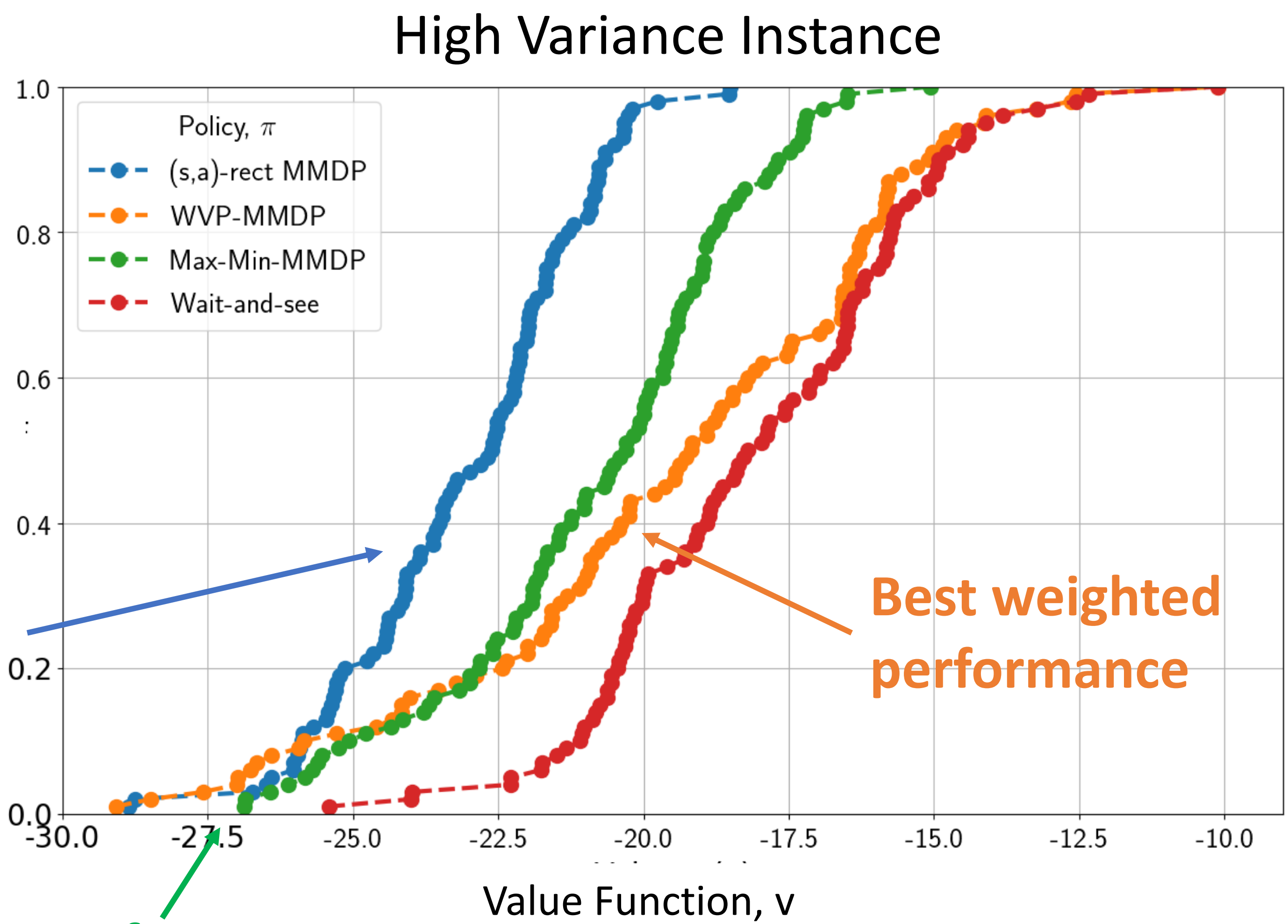
The measure of protection against can distribution of performance among models

$\mathbb{P}(v^m(\pi) \leq v)$



The measure of protection against can distribution of performance among models

$\mathbb{P}(v^m(\pi) \leq v)$

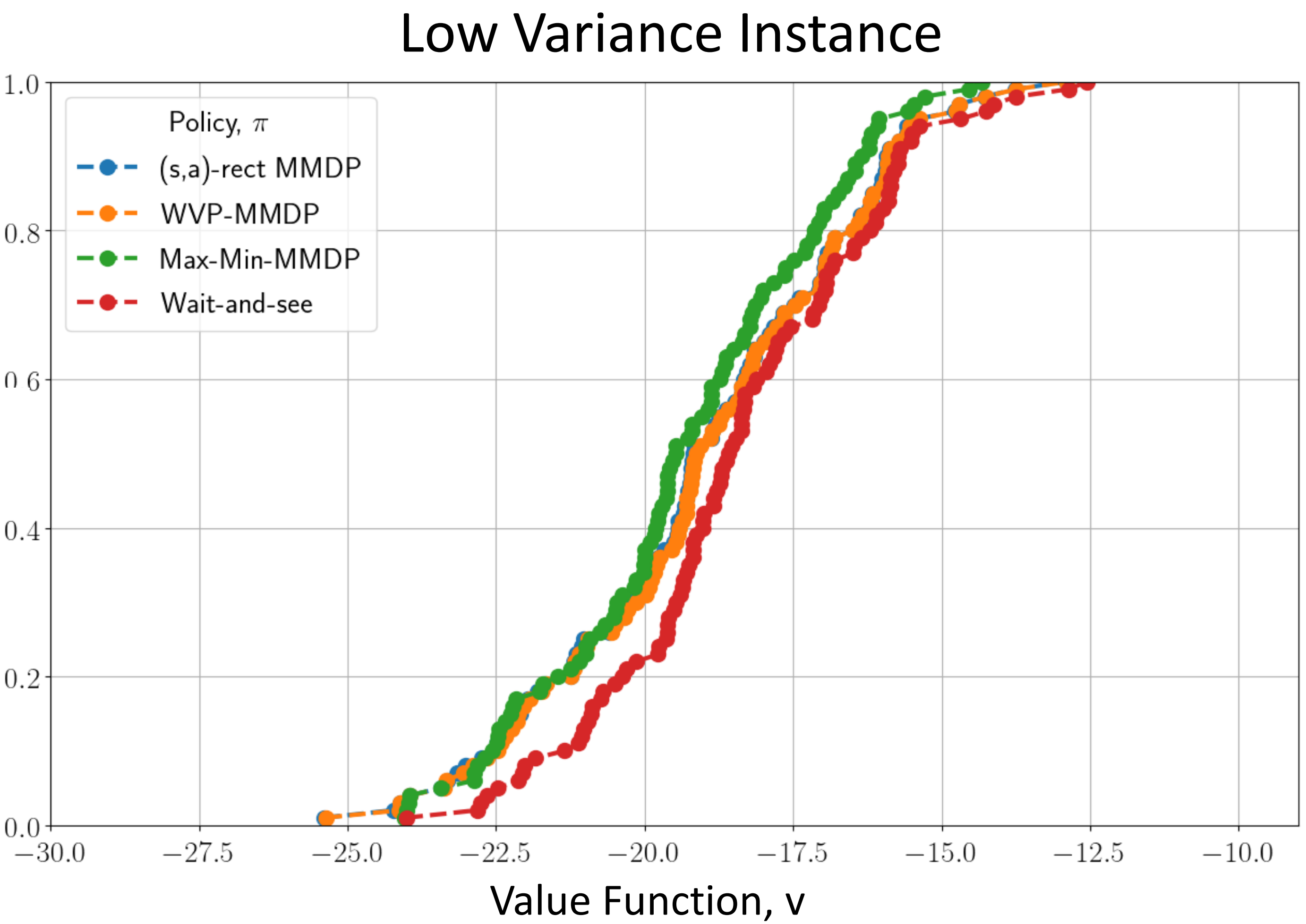


(s,a)-rect-MMDP does not mitigate ambiguity well

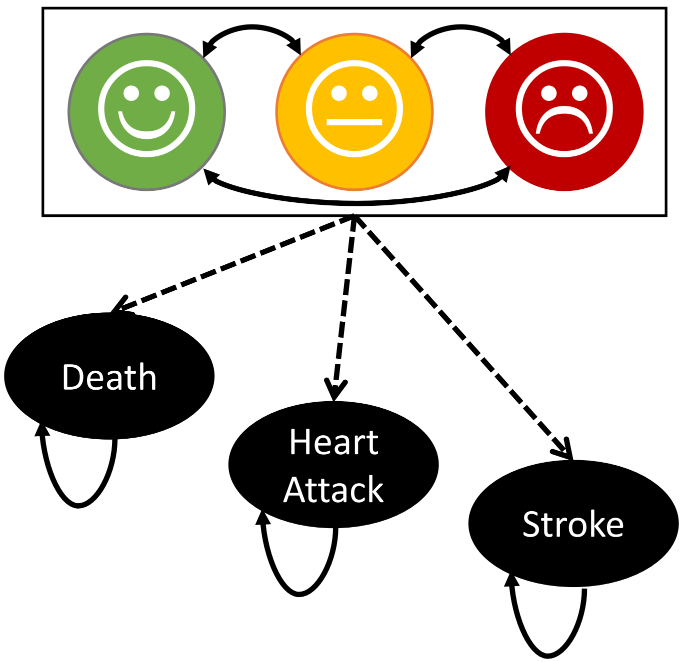
Best worst-case performance

As variance in models decreases, the form of protection against ambiguity matters less

$\mathbb{P}(v^m(\pi) \leq v)$



We considered these formulations to determine the optimal time to start statins



Multi-model Markov decision process

- 64 states (HDL/TC Levels)
- 3 actions (Wait, low-dose, high-dose)
- 34 decision epochs
- 30 models

Case study data

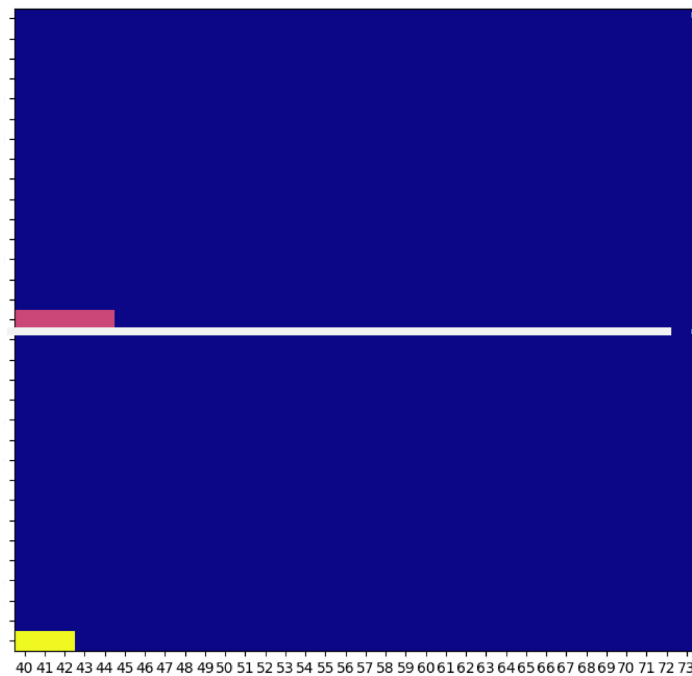
- Longitudinal data from Mayo Clinic
- ACC risk calculator
- Disutilities from medical literature

Mason, J. E., Denton, B. T., Shah, N. D., & Smith, S. A. (2014). Optimizing the simultaneous management of blood pressure and cholesterol for type 2 diabetes patients. *European Journal of Operational Research*, *233*(3), 727-738.

Most formulations of the MMDP recommend similar policies

WVP-MMDP
Perc-Opt-MMDP (20%)
Max-min-MMDP

HDL/TC State



Not on a statin

On a low-dose statin

Better cholesterol

Worse cholesterol

Age

Most MMDP policies are similar; (s,a)-rect-MMDP treats more aggressively

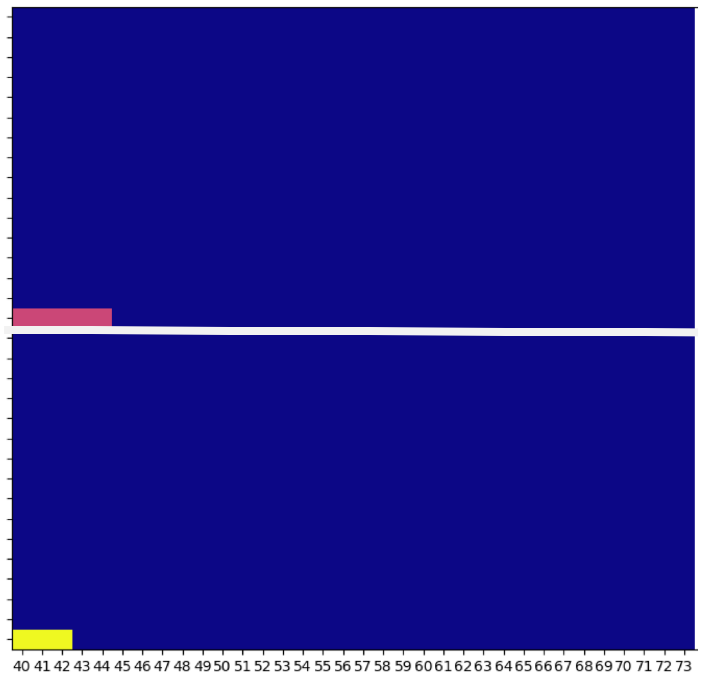
WVP-MMDP
Perc-Opt-MMDP (20%)
Max-min-MMDP

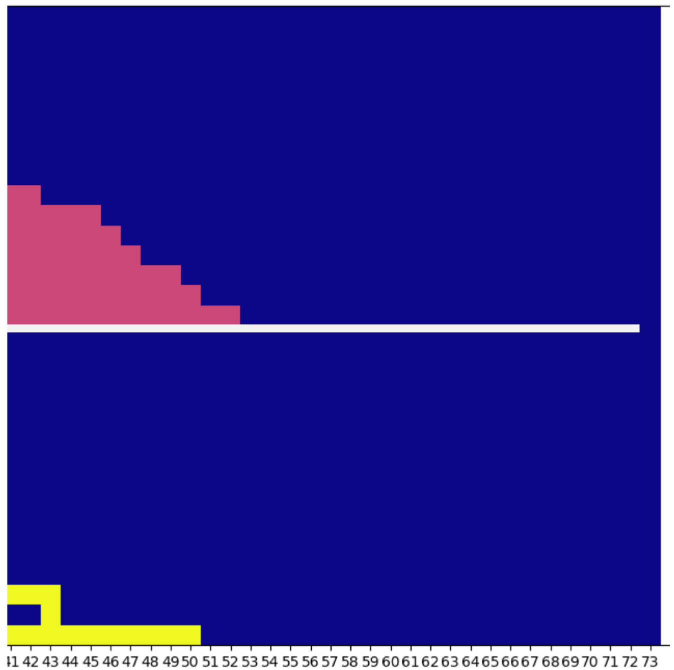
(s,a)-rect-MMDP

HDL/TC State

Lower risk

Higher risk





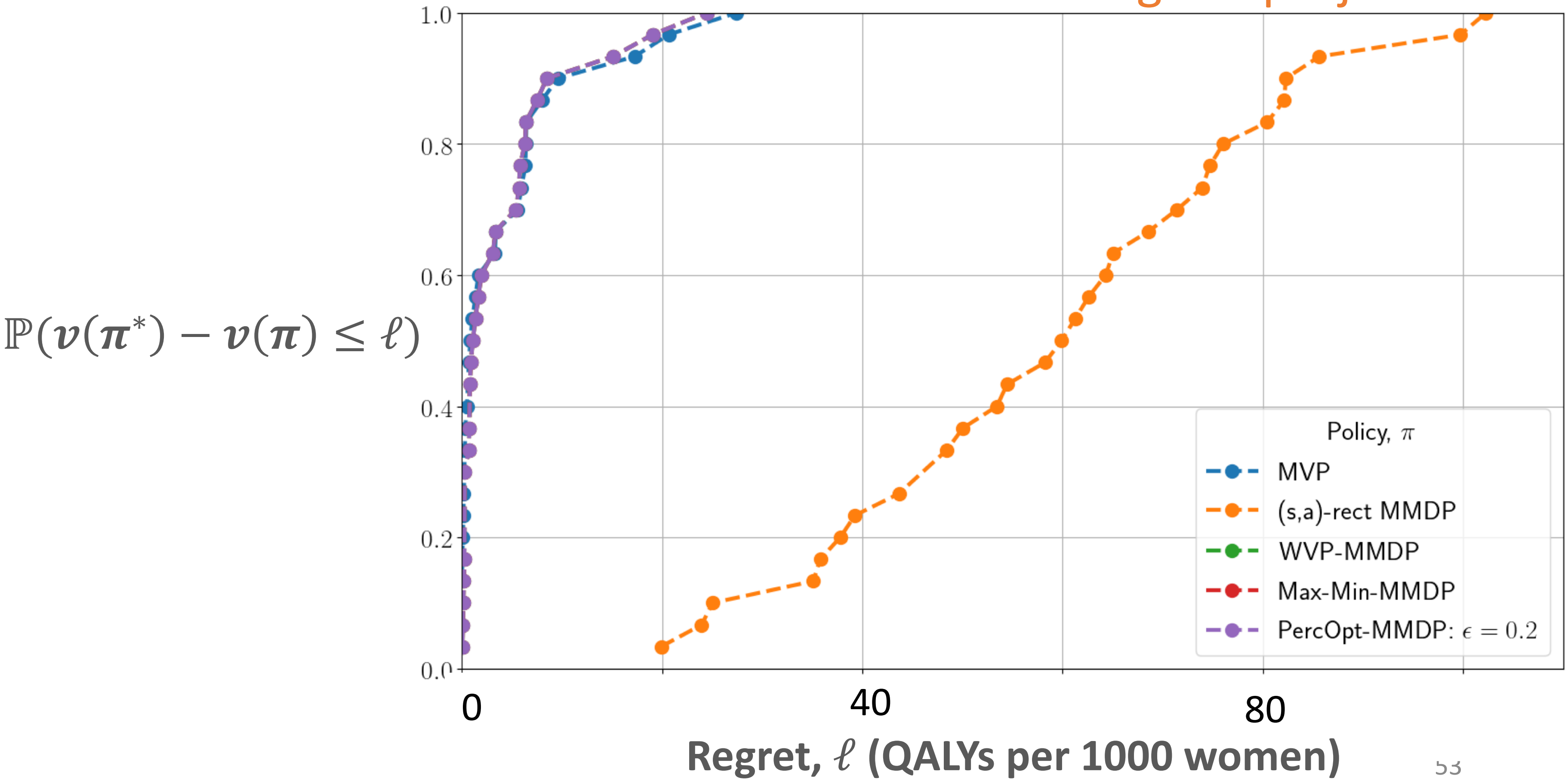
Age

Age

High variance instance

(s,a)-rect-MMDP can perform worse than MVP in all models

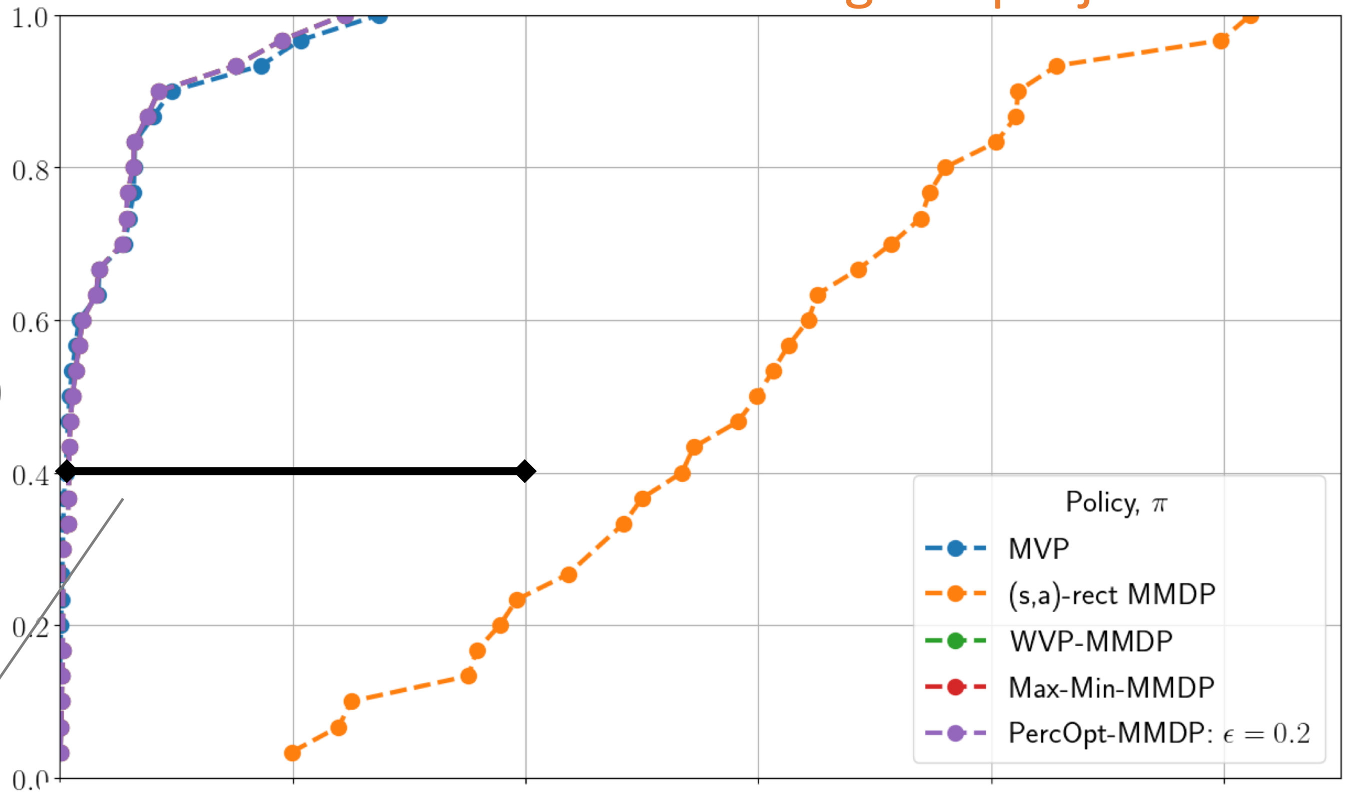
Regret from (s,a)-rectangular projection



(s,a)-rect-MMDP can perform worse than MVP in all models

Regret from (s,a)-rectangular projection

$\mathbb{P}(v(\pi^*) - v(\pi) \leq \ell)$



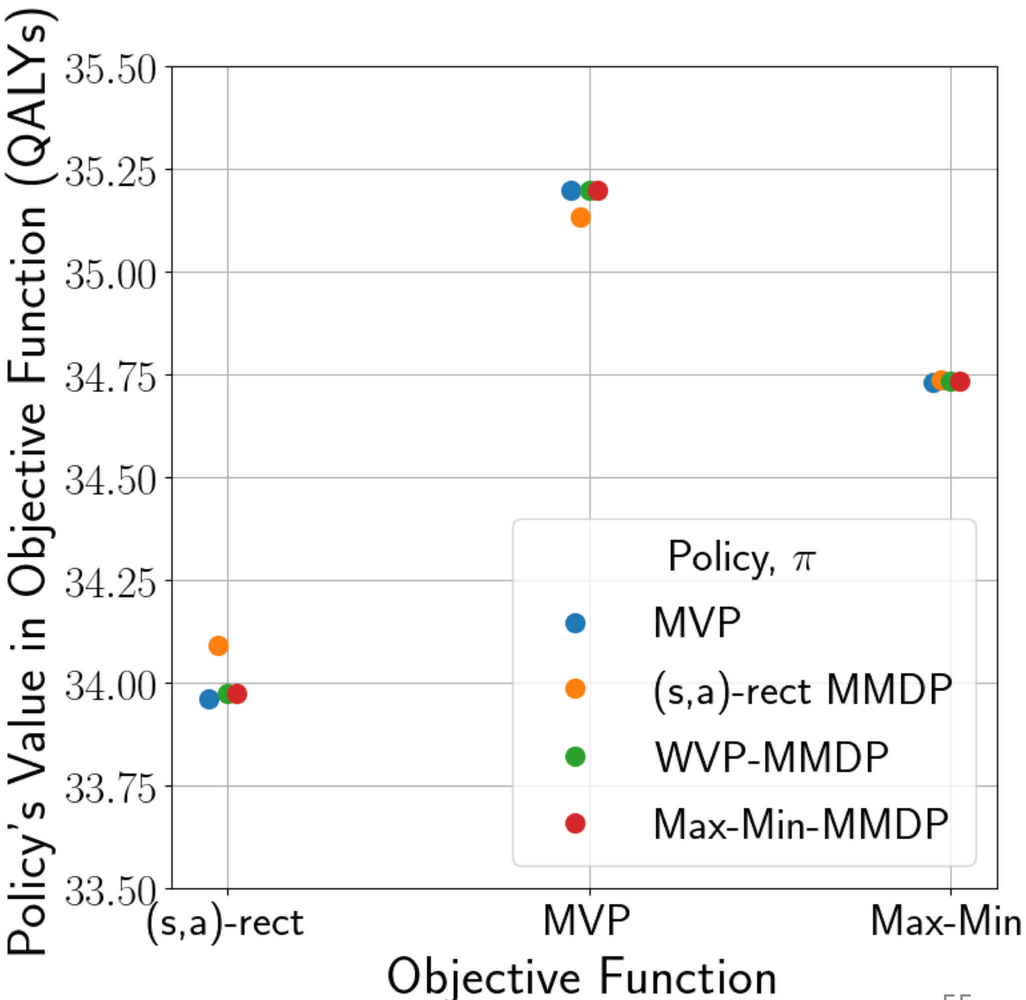
0 40 80

Regret, ℓ (QALYs per 1000 persons)

Benefit from aspirin, an important intervention for CVD prevention

(s,a)-rect-MMDP may not be good indicator of worst-case performance

Difference between worst-case in (s,a)-rect-MMDP and max-min-MMDP



Conclusions

Branch-and-bound can be modified to incorporate other protective measures towards ambiguity

Considering multiple models is most important when the models are quite different; MVP tends to perform well for MDPs with imprecise parameters

Use caution before employing the (s,a)-rectangularity property if not a supported assumption

Summary of contributions

We considered the issue of ambiguity in MDPs arising from multiple plausible models

We created solution methods that allow for DM to consider performance in different models

We characterized when it is most important to consider ambiguity

Laid foundations for future work on incorporating ambiguity in stochastic dynamic optimization

Acknowledgments

Michigan Engineering

Brian T. Denton, Ph.D.

Vinayak Ahluwalia

Charmee Kamdar

UM-Dearborn School of Business

David Kaufman, Ph.D.

Mayo Clinic

Nilay Shah, Ph.D.

U.S. Department of Veterans Affairs

Rodney Hayward, MD

Jeremy Sussman, MD

This material is based upon work supported by the National Science Foundation under Grant Number CMMI- 1462060 (Denton) and Grant Number DGE-1256260 (Steimle). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Lauren N. Steimle

Department of Industrial and Operations Engineering

University of Michigan, Ann Arbor, MI

steimle@umich.edu

@LaurenSteimle

www.umich.edu/~steimle