&

Electronic Health Record Analytics: The Case of Optimal Diabetes Screening

Michael Hahsler[1], Farzad Kamalzadeh[1]
Vishal Ahuja[1], and Michael Bowen[2]

[1] Southern Methodist University

[2] UT Southwestern Medical Center and Parkland Health and Hospital System

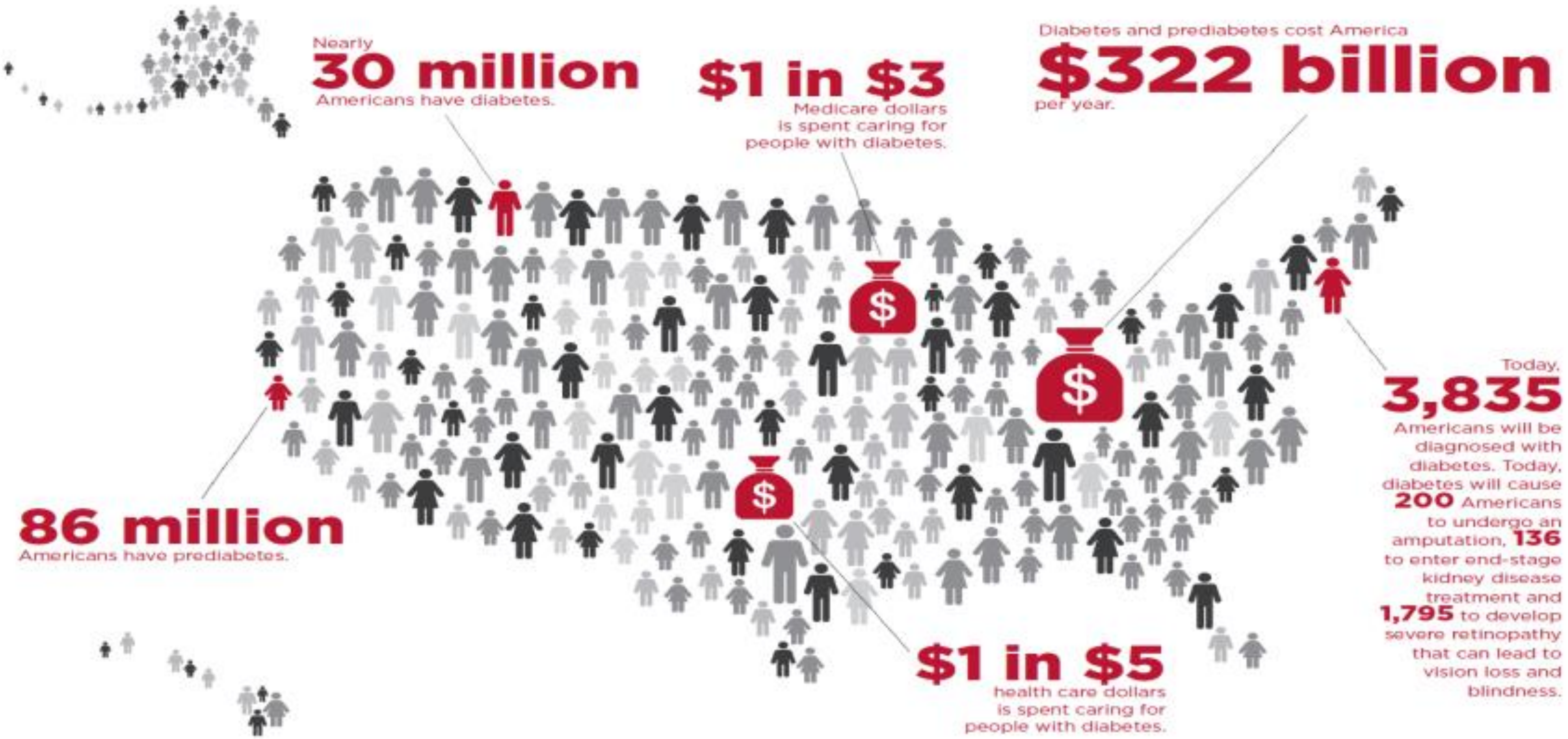
EMIS Industry Advisory Board and Outreach Meeting

December 3, 2018

World Changers Shaped Here

SMU®

THE STAGGERING COSTS OF DIABETES IN AMERICA



American Diabetes Association®

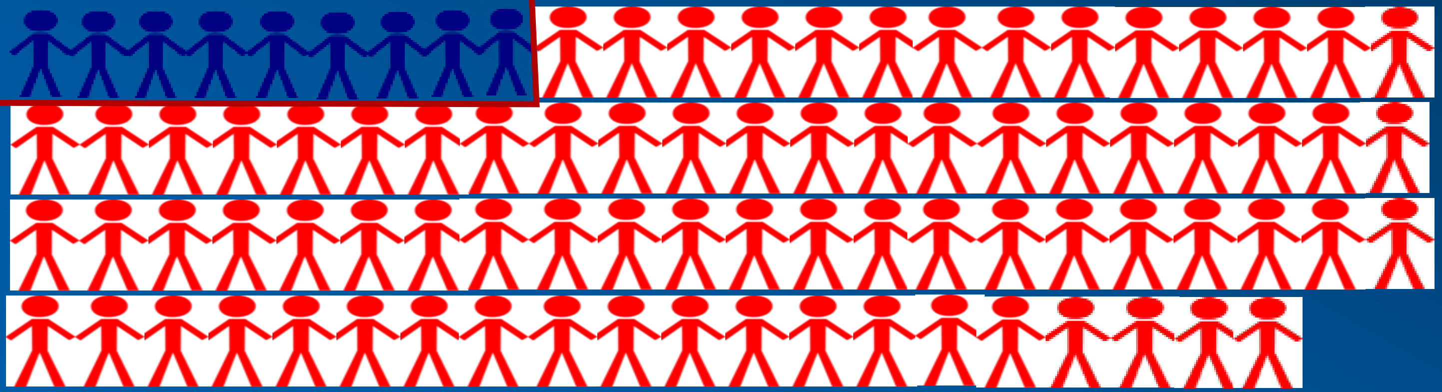
http://main.diabetes.org/dorg/images/infographics/adv-cost-of-diabetes.pdf;
Diabetes Care 2013; 36:1033-1046.

Prevalence of Diagnosed and Undiagnosed Type 2 Diabetes and Prediabetes

29.1 million people in the US have T2DM (9.3% of population)

8.1 Million Undiagnosed

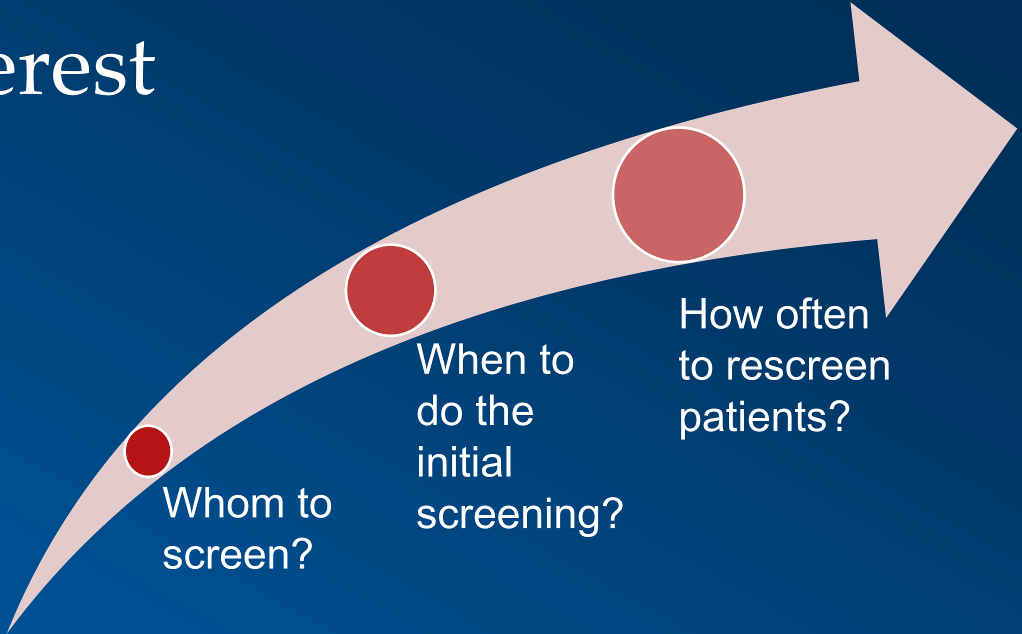
Over 86 million adults in the US with pre-diabetes (37% of population)



77 Million with Undiagnosed Pre-diabetes

CDC National Diabetes Statistics Report, 2014.

Questions of Interest



Optimal screening decision under constraints and uncertainty

- Constraints on resources and patient availability. Population screening is not feasible.
- Individualize the decision based on cohort and patient characteristics.
- Focus on catching the disease (i.e., prevalence) at earlier stages.

Common Screening Strategies

1. Opportunistic Screening

2. Screening Guidelines
   - American Diabetes Association (ADA)

     All adults over age 45 OR any age if BMI ≥ 25 (or ≥ 23 in Asians) AND an additional risk factor
   - U.S. Preventive Services Task Force (USPSTF) 2015

     Adults 40-70 AND BMI≥25

3. Diabetes Risk Score
   - Incidence/prevalence risk score.
   - Not widely used in the US.

Jaana Lindström and Jaakko Tuomilehto, The Diabetes Risk Score: A practical tool to predict type 2 diabetes risk, Diabetes Care 2003 Mar; 26(3): 725-731.

Setting and Data

- **Setting:** Parkland Health and Hospital System, a large integrated, safety-net healthcare system in North Texas.
- **Data Source:** Epic Electronic Medical Record (EHR)
- **Retrospective cohort** (N = 34,297 patients, 2012-2015)
- Eligibility
  - Ages 18-65
  - Established patients (≥1 primary care visit every 18 month)
  - Only unscreened patients with no known diabetes during first 12 month

Available Data Extracted from EHR

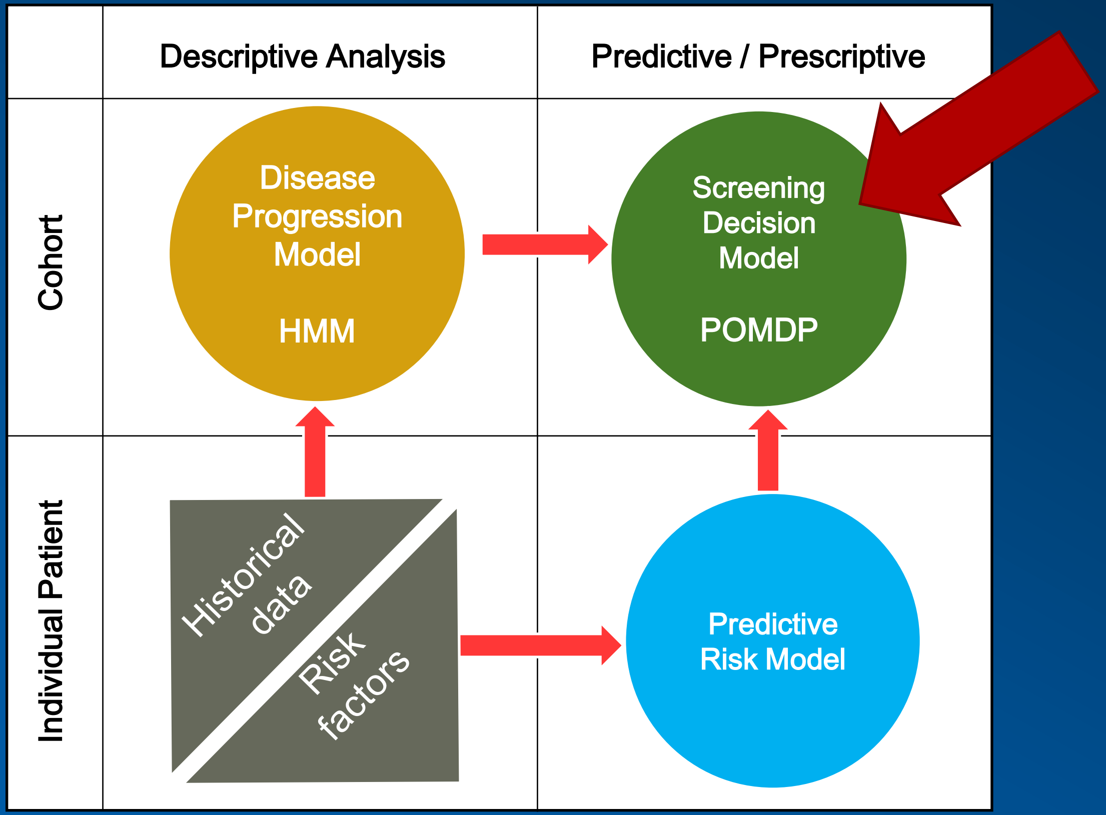
Parkland

105 Features including

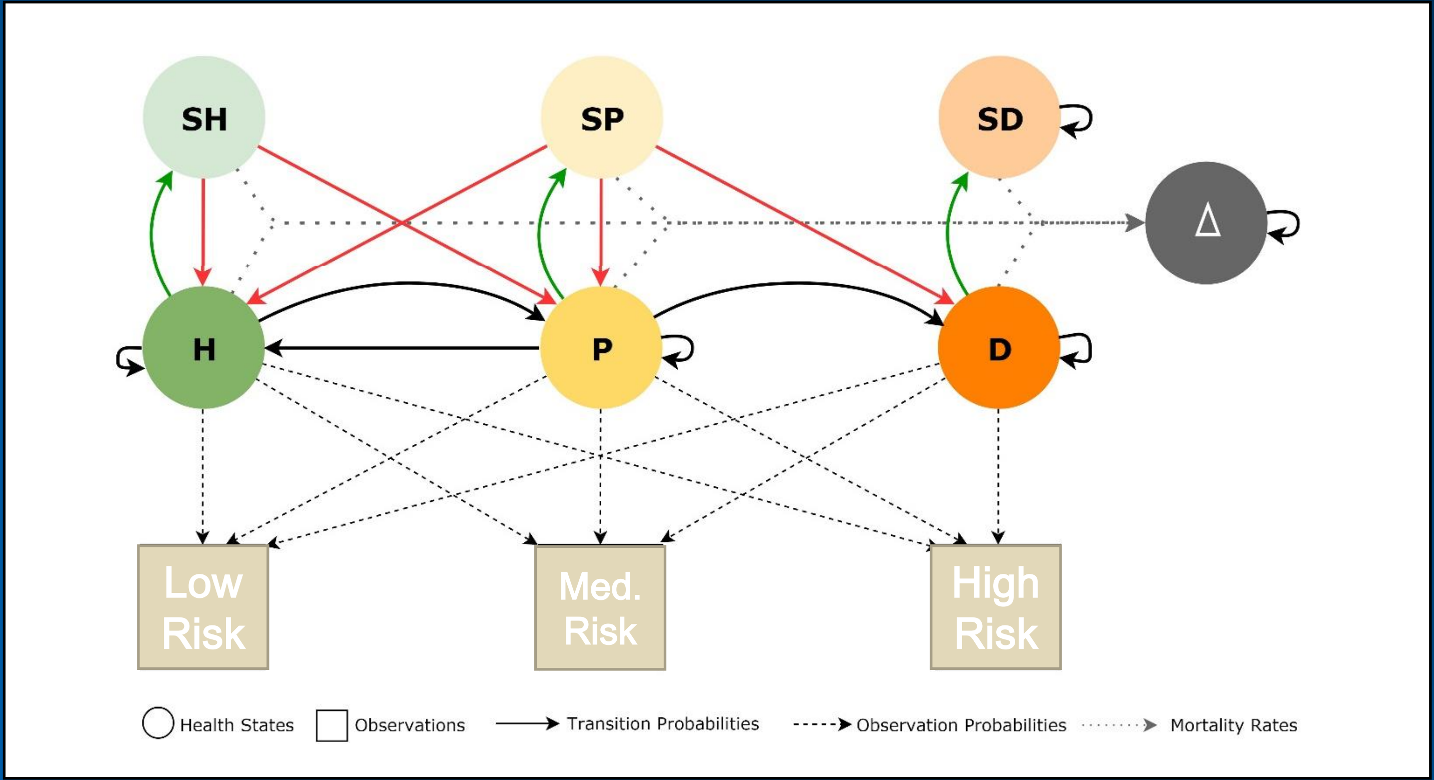
- **Demographic information:** Age, gender, ethnicity, etc.
- **Vitals:** Blood pressure, etc.
- **BMI**
- **Risk factors** (co-morbidities): Hypertension, family history, etc.
- **Lab values:** Cholesterol, random blood glucose, etc.
- **Medications** (prescribed): Blood pressure, cholesterol, etc.
- **Health care utilization:** Office encounters, ER visits, etc.
- **Screening results:** Hemoglobin A1C, fasting plasma glucose, oral glucose tolerance test

Only demographic information, BMI and vitals are widely available.
>20% of the data values are missing overall.
>50% of lab values missing.

Health Analytics Framework

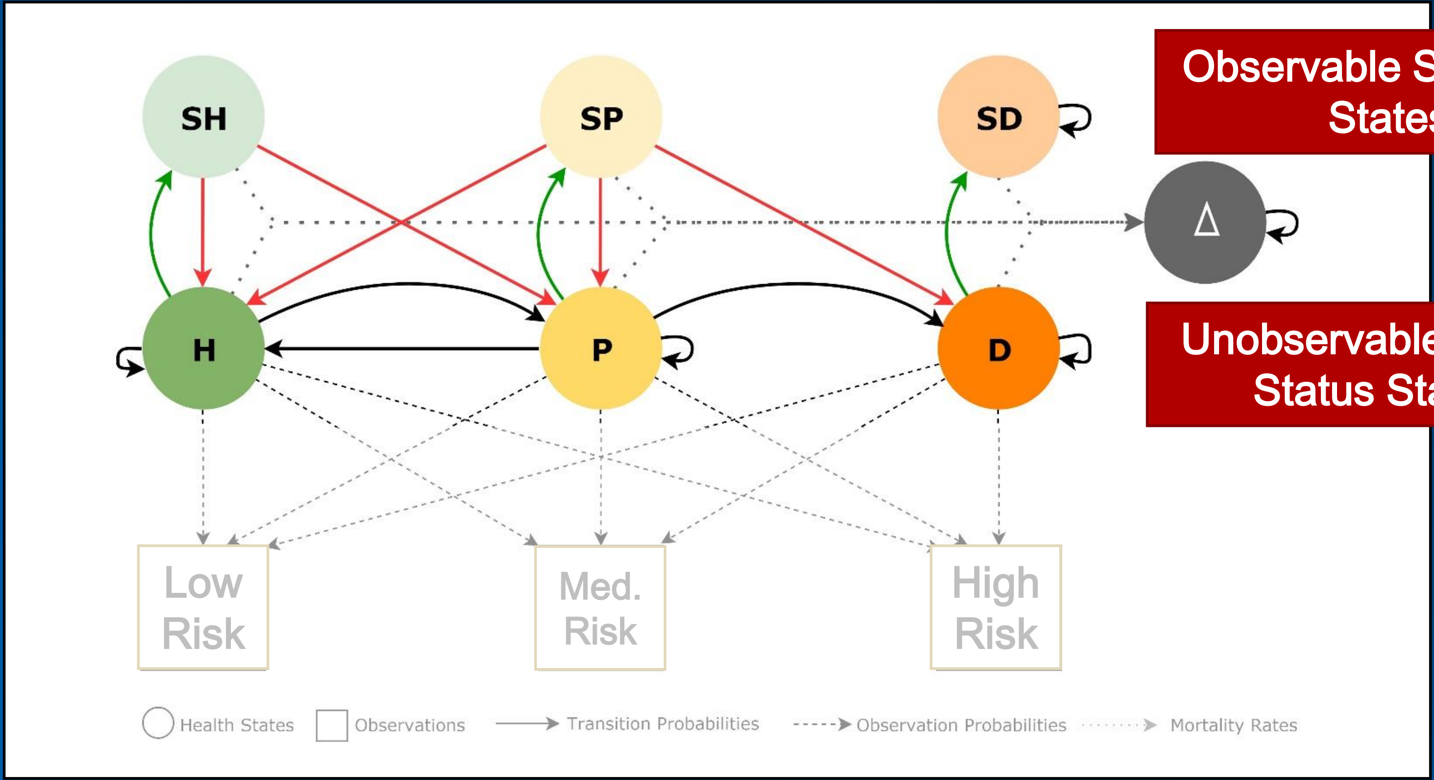


Partially Observable Markov Decision Process



Sondik, E.J. (1978). "The optimal control of partially observable Markov processes over the infinite horizon: discounted cost". Operations Research. 26 (2): 282–304.

POMDP: Discrete Health Status States



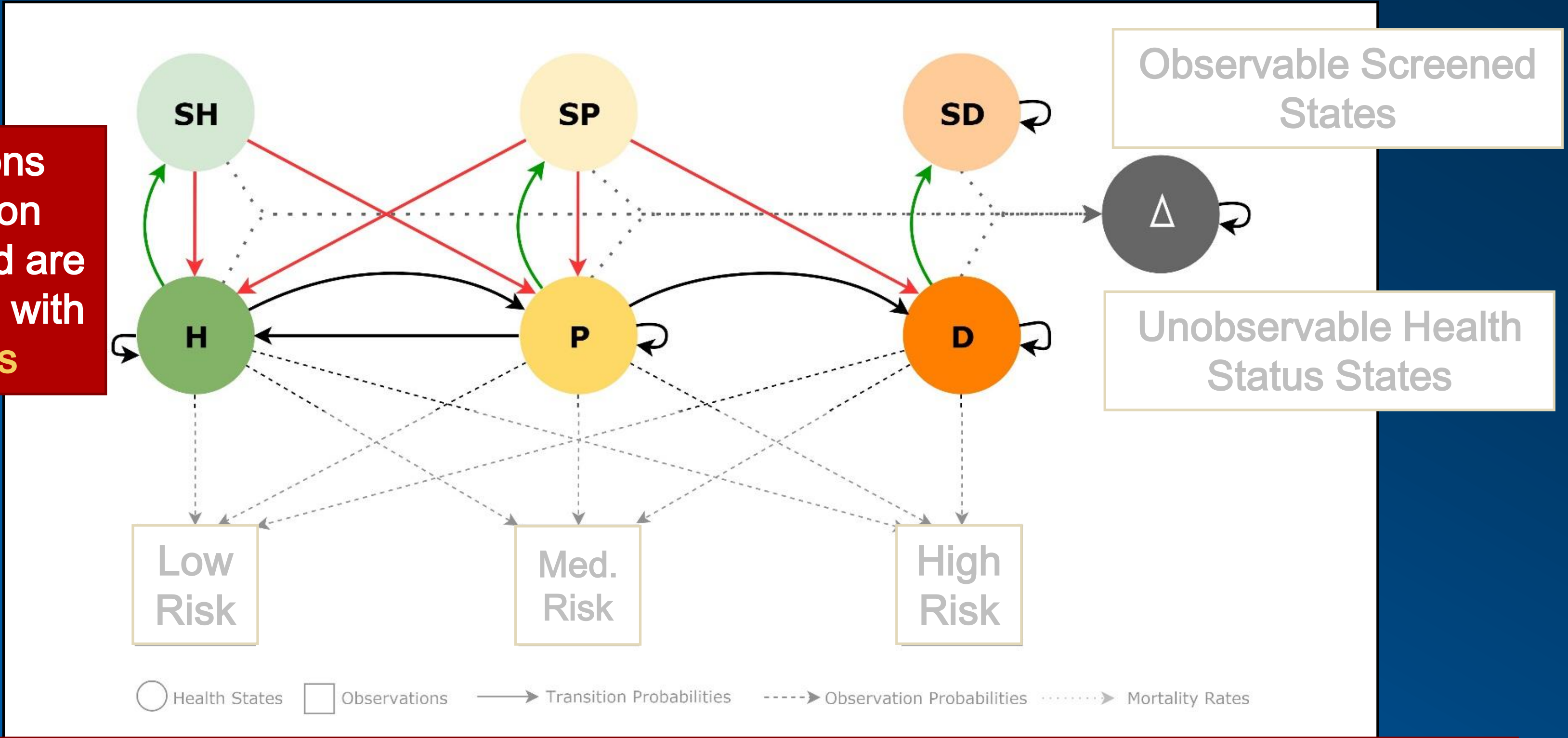
Observable Screened States

Unobservable Health Status States

Note: We only know if a patient has (pre)diabetes if we screen the patient.

POMDP: Actions, Transitions and Rewards

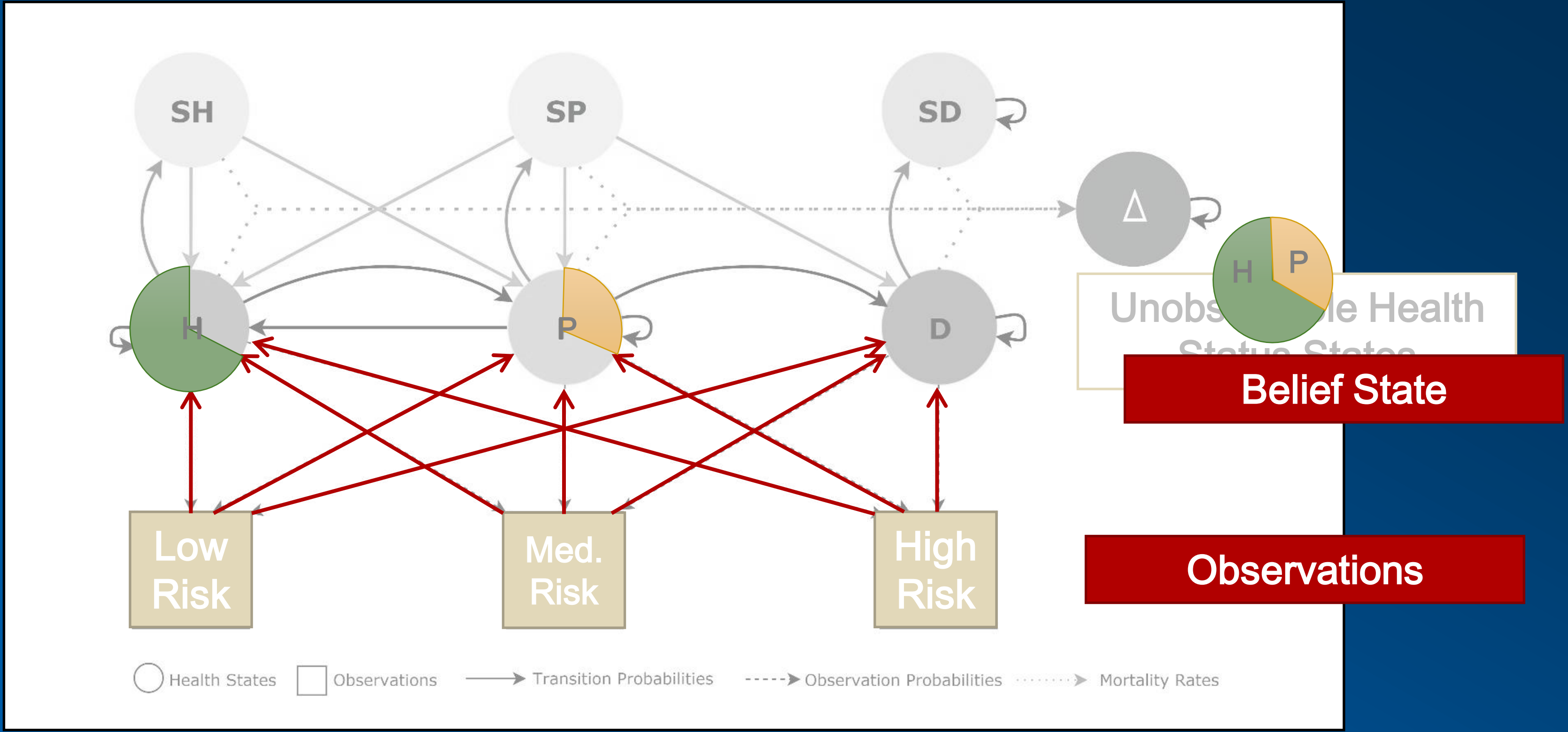
Transitions depend on actions and are associated with rewards



Actions: Screen/don't screen

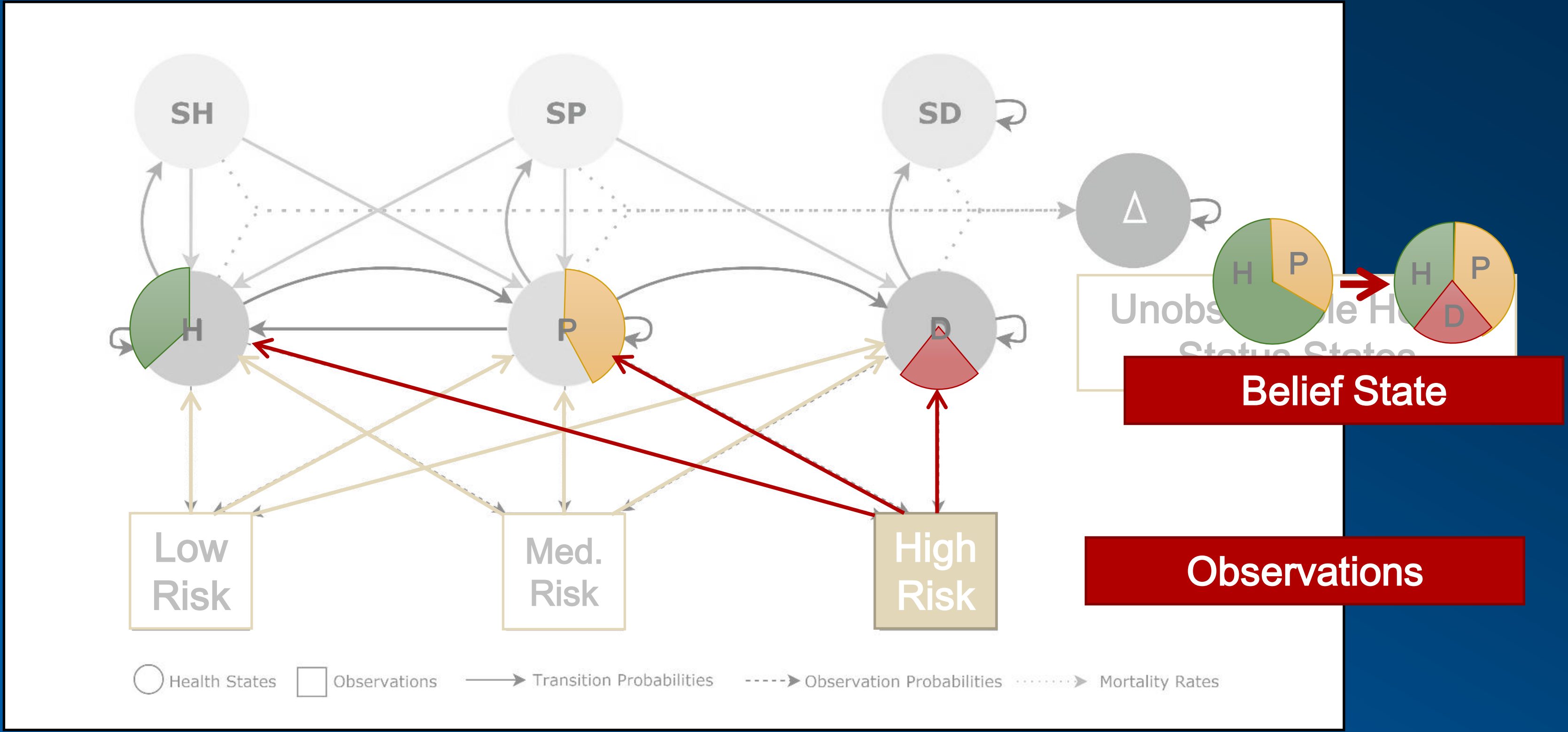
Rewards: Cost of screening, medical expenses, reduced quality of life, lost income

POMDP: Observations and Belief States



Observations give us information about the unobservable health status → **"Belief State"**

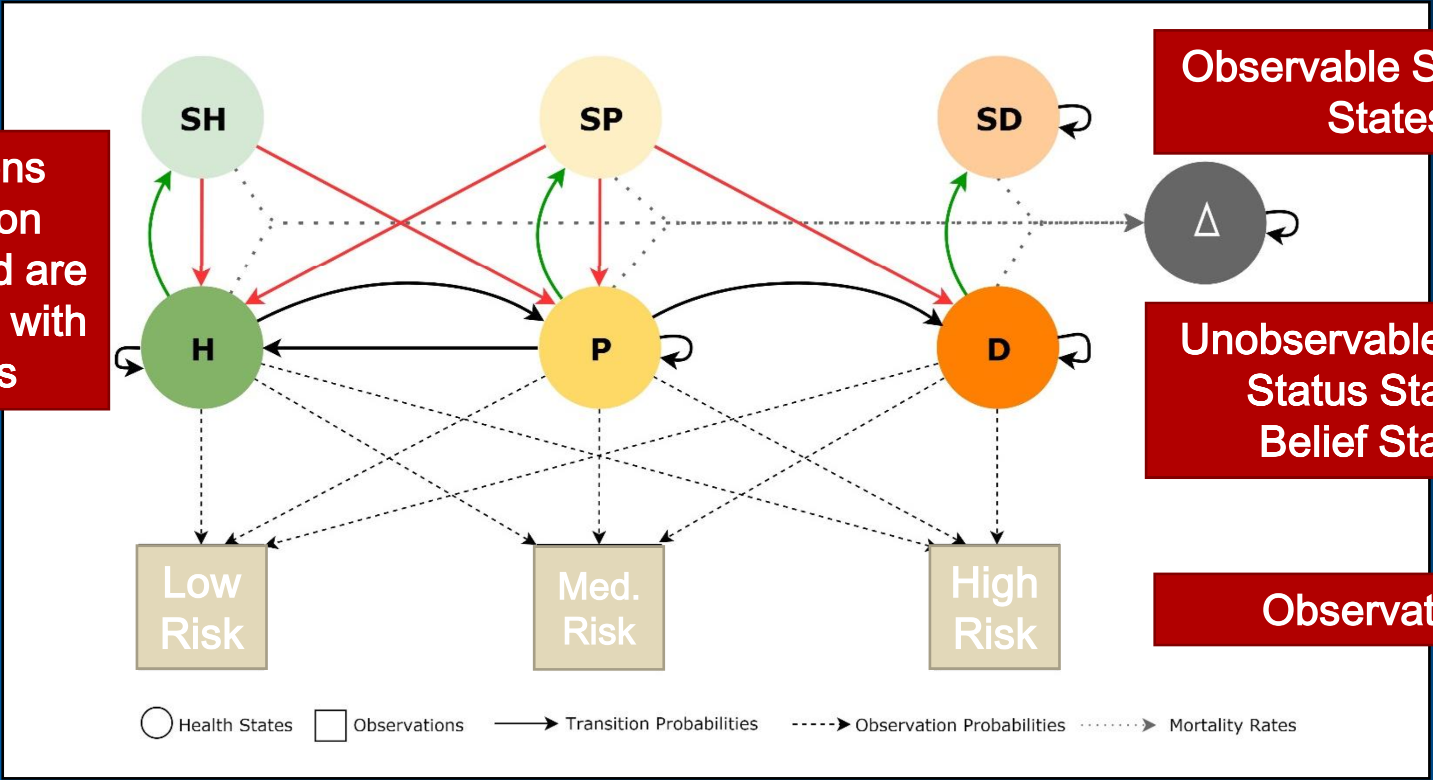
POMDP: Observations and Belief States



A **new observation** results in a change of our "Belief State."

POMDP: Screening Decision Model

Transitions depend on actions and are associated with rewards



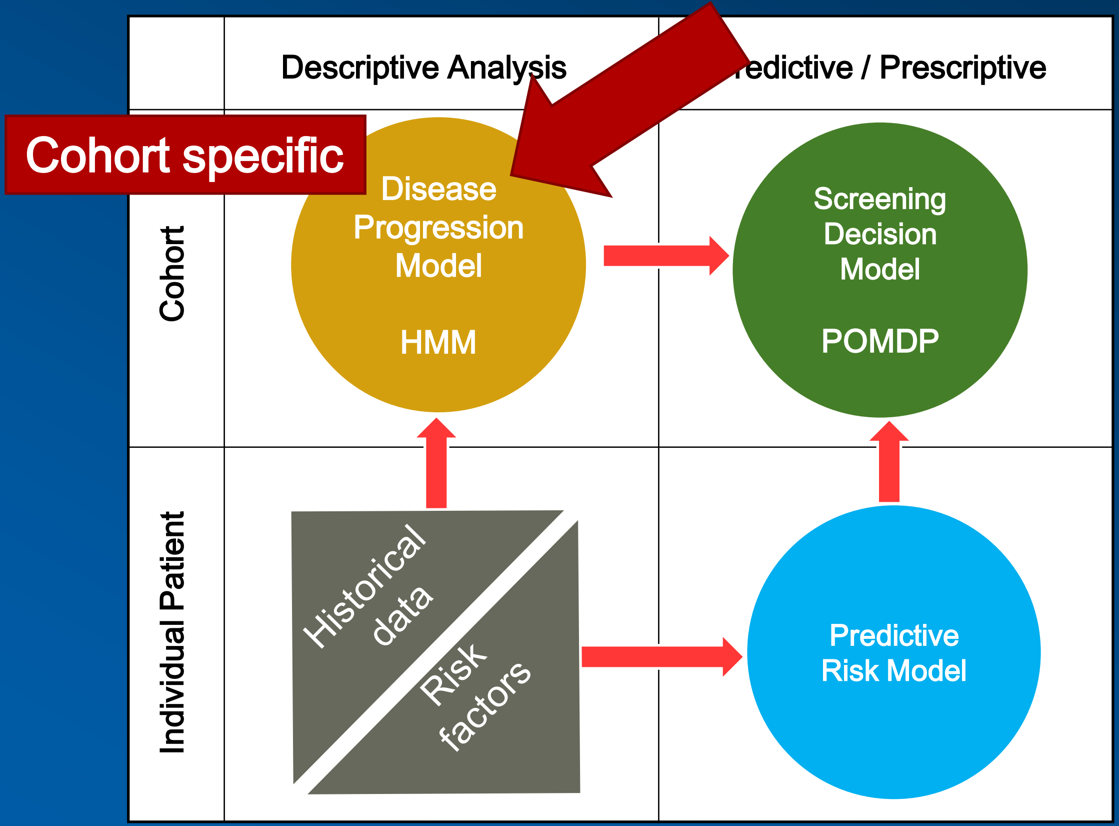
Observable Screened States

Unobservable Health Status States/ Belief States

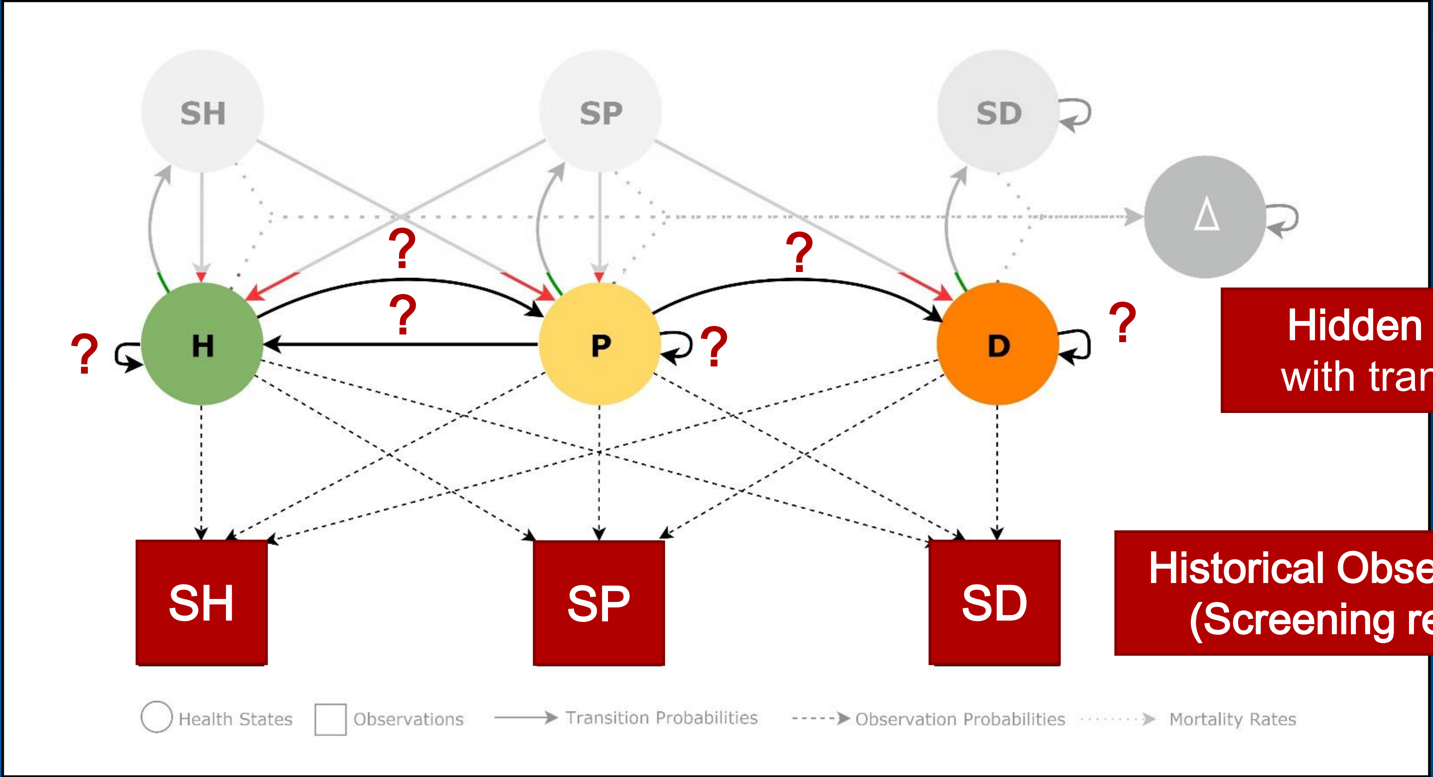
Observations

Goal: optimal policy. I.e., optimal action for each state to maximize the expected future reward.

Health Analytics Framework



HMM: Learn a Cohort-Specific Disease Progression Model

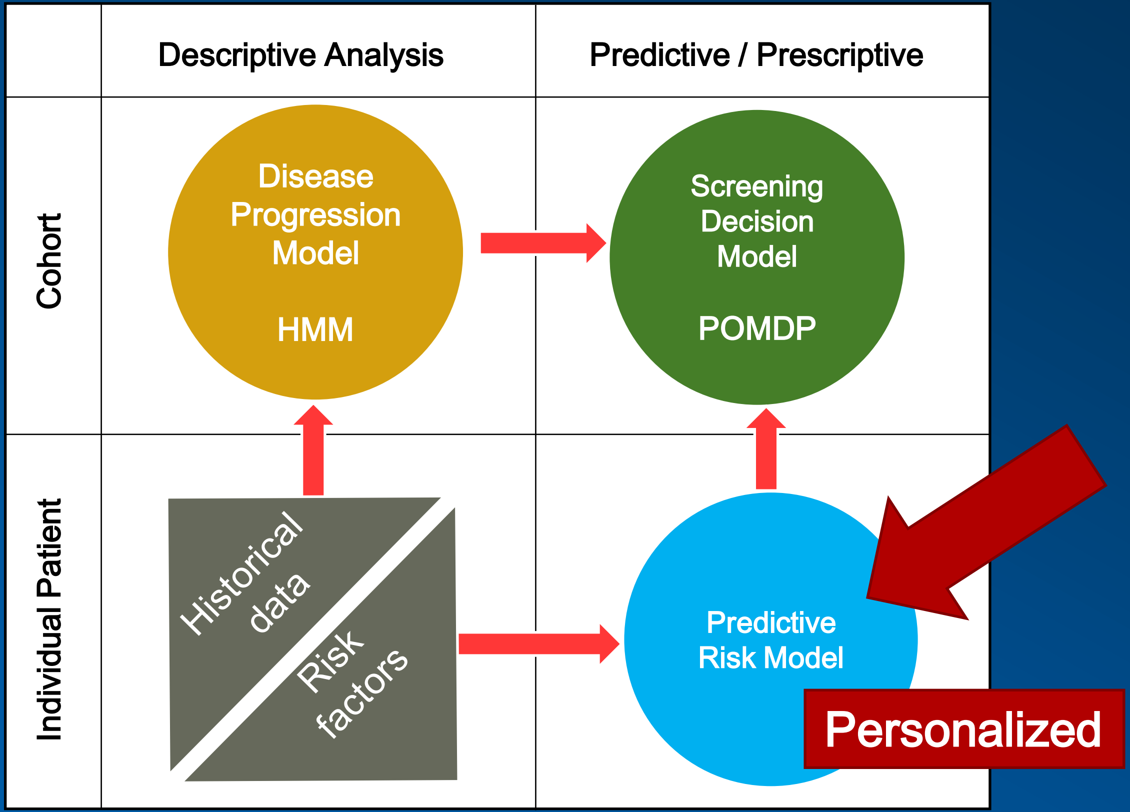


Hidden States with transitions

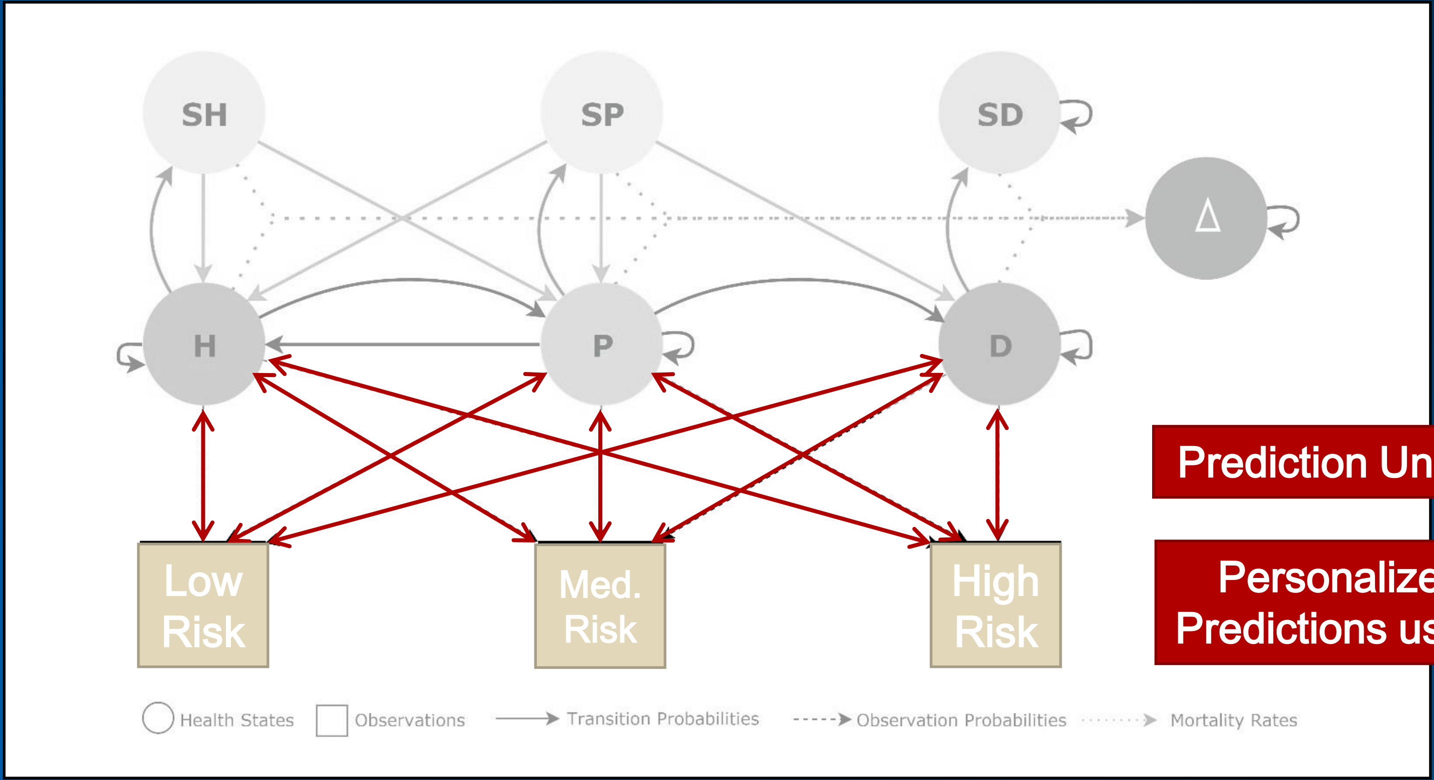
Historical Observations (Screening results)

Sukkar R, Katz E, Zhang Y, Raunig D, Wyman BT, Disease progression modeling using Hidden Markov Models. Conf Proc IEEE Eng Med Biol Soc. 2012;2012:2845-8.

Health Analytics Framework



Predictive Risk Model



Prediction Uncertainty

Personalized Risk Predictions using EHR

Collins et al., Developing risk prediction models for type 2 diabetes: A systematic review of methodology and reporting, BMC Medicine 2011 9:103

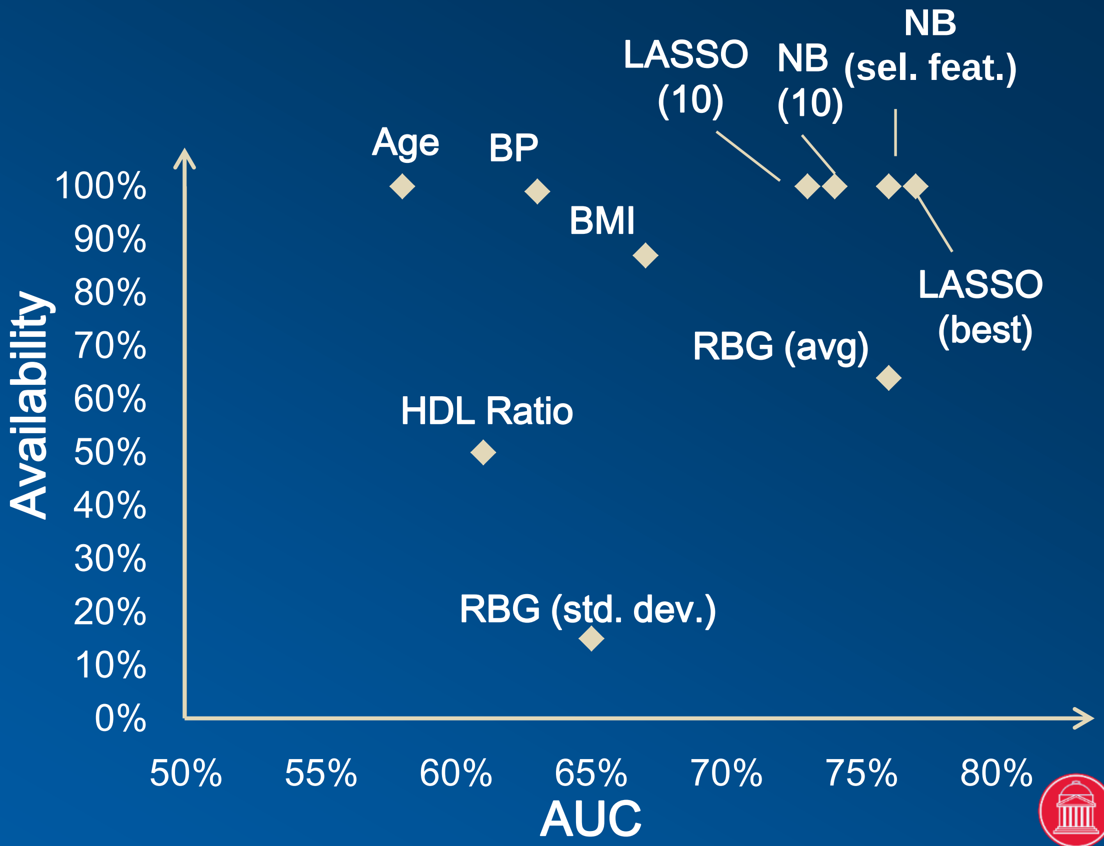
Observations via Predictive Modeling

- Idea: Use predictive modeling (classification) to learn the relationship between clinical observations recorded in EHR and the unobservable health state. Predictions can be used as personalized observations resembling a "Virtual Screening."
- Our key questions are:
  - How to we produce **simple predictive models** to guide screening using only already available data?
  - How do we deal with a large quantity of **missing data** and **data quality issues**?
- Desired properties:
  - Applicable to all patients, no matter how much information we have.
  - Can guide us to what missing patient information would be most valuable.

Comparison of Some Predictive Models

	AUC	Availability
LASSO (best)	77%	100%
NB (select feat.)	76%	100%
NB (10)	74%	100%
LASSO (10)	73%	100%
RBG (avg)	76%	64%
BMI	67%	87%
RBG (std. dev.)	65%	15%
BP	63%	99%
HDL Ratio	61%	50%
Age	58%	100%

LASSO: Logistic Regression with Regularization
NB: Naïve Bayes Classifier
RBG: Random Blood Glucose Test



POMDP: Parameters

Disease Progression (Transitions)

$$\mathcal{P} = \begin{array}{c} H \\ P \\ D \\ \Delta \end{array} \begin{pmatrix} 0.9438 & 0.048 & 0 & 0.0082 \\ 0.0328 & 0.9242 & 0.0348 & 0.0082 \\ 0 & 0 & 0.9916 & 0.0084 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Risk Prediction Performance

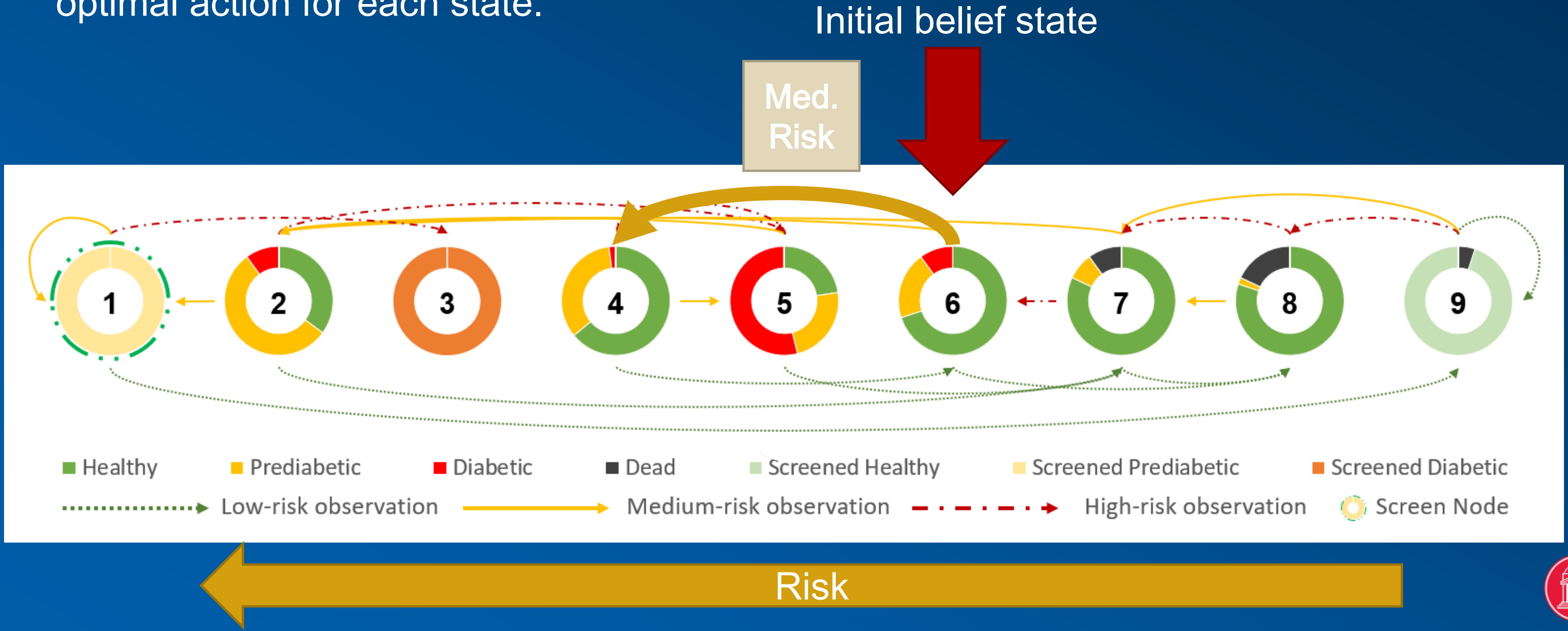
$$\mathcal{O}(o|s) = \begin{array}{c} H \\ P \\ D \end{array} \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.15 & 0.7 & 0.15 \\ 0.05 & 0.25 & 0.7 \end{pmatrix}$$

Rewards (from Literature)

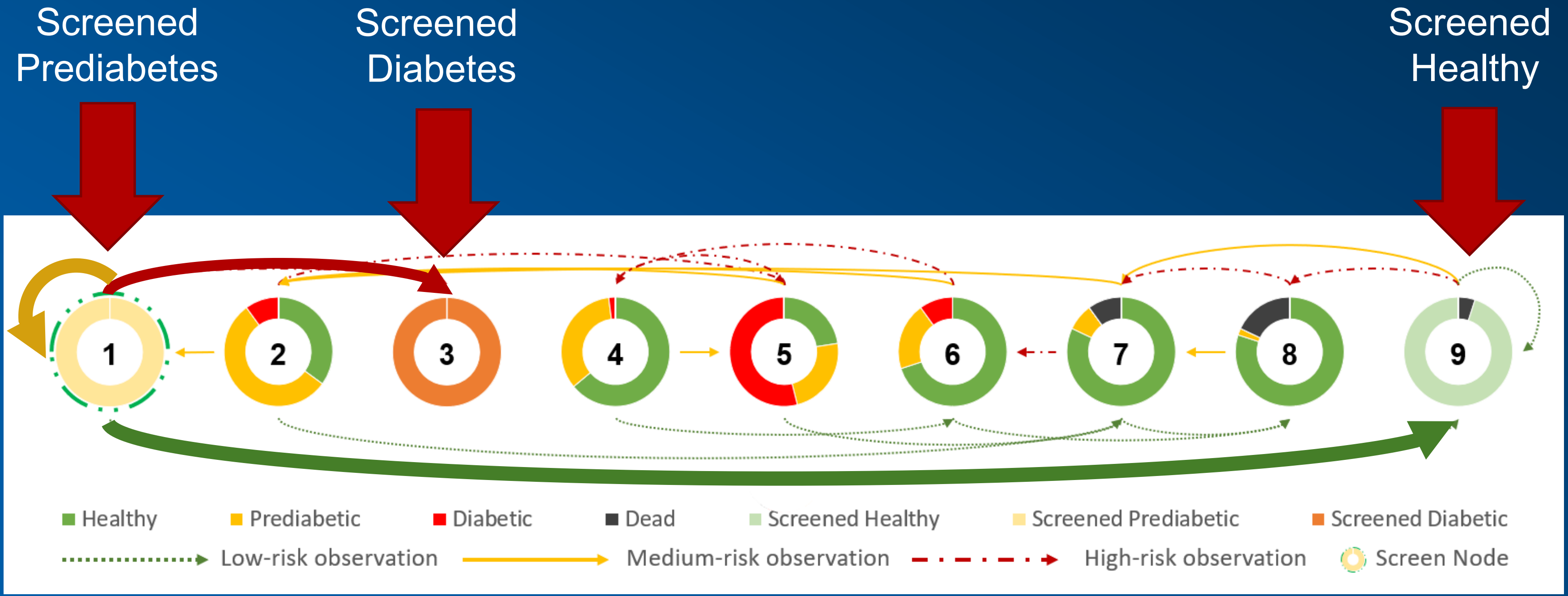
Parameter	Description	Source	Patient	Healthcare system	Society
C_s	Cost of a diabetes screening test	[55][56][57][16]	$134+$192	$8020	$8346
Q	Quality-Adjusted Life Year in U.S. dollars	[58]	$50,000		$50,000
C_D	Direct medical costs per year for new-onset diabetes	[55]		$4,174	$4,174
C_P	Incremental direct medical costs per year for a patient with prediabetes	[55]		$1,316	$1,316
α_P	Annual utility decrease of living with prediabetes	[59][60]	0.16		
α_{UD}	Annual utility decrease of living with undiagnosed diabetes	[59][61][62][63]	0.2		
α_{DD}	Annual utility decrease of living with diagnosed diabetes	[59][61][62][63]	0.18		
m_T	Age-Adjusted mortality rate in U.S. in 2016	[53][64]	0.0084		
m_D	Age-adjusted mortality rate for Diabetes in 2016	[53][64]	0.00021		
l_e	Life expectancy for the U.S. population in 2016	[53]	78.7		
l_d	Lifespan decrement due to Diabetes	[65]	5		
u_r	Uptake rate of Diabetes screening	[66][67][68][69][70]	0.644		

POMDP: Optimal Screening Policy

- We maintain for each patient a belief state.
- The belief state is updated with each new observation.
- The policy is a set of all considered belief states with the optimal action for each state.



POMDP: Optimal Screening Policy



Effectiveness compared to Opportunistic Screening

Screening Policy	ICER (incr. cost per QALY) (SD)	Years Gained (SD)	QALYs gained (SD)	Diagnosis lead time reduction (SD)	Macrovascular events prevented (SD)	Microvascular events prevented (SD)	Deaths prevented (SD)
30+, every 3 years	$27,042 (1268)	0.75 (0.04)	2.04 (0.05)	19 (0.2)	22 (1.6)	207 (4)	48 (2)
45+, every year	$37,366 (1755)	0.62 (0.04)	1.18 (0.03)	14 (0.1)	21 (1.5)	178 (4)	45 (2)
45+, every 3 years	$31,155 (1791)	0.61 (0.04)	0.96 (0.03)	11 (0.1)	20 (1.4)	165 (4)	44 (2)
45+, every 5 years	$29,644 (2175)	0.60 (0.04)	0.86 (0.03)	9 (0.1)	20 (1.5)	157 (4)	44 (2)
60+, every 3 years	$32,201 (2966)	0.59 (0.04)	0.60 (0.03)	6 (0.1)	19 (1.4)	142 (4)	42 (2)
Maximum screening 30+	$36,801 (1233)	0.83 (0.05)	2.63 (0.05)	25 (0.2)	23 (1.5)	229 (4)	50 (2)
Proposed optimal policy	**$20,426** (13[illegible]	**0.81** (0.04)	**2.06** (0.[illegible]	**18 (0.2)**	**23 (1.5)**	**219 (5)**	**49 (2)**

ADA

-30%

x2

Tony Hsiu-Hsi Chen, Ming-Fang Yen, Tao-Hsin Tung. A computer simulation model for cost–effectiveness analysis of mass screening for Type 2 diabetes mellitus, Diabetes Research and Clinical Practice 54 Suppl. 1 (2001) S37– S42

Limitations and Future Steps

- **HMM**: Estimation of transition probabilities may be biased because it is based on actually screened patients.
- **Predictive Model**: Missing data and data quality are a big issues.
- **POMDP**
  - Cost/reward structure in POMDP (e.g., real cost depends on time in a state)
  - Process is most likely not Markovian (more states can represent dependence on past information).
  - Other dimensions for the state space (E.g., age or BMI)? Make the model harder to solve due to an explosion of the number of belief states.
  - Set of possible/available actions (e.g., other interventions including diet and exercise).