# Learning Across Bandits in High Dimension via Robust Statistics

**Hamsa Bastani**

Wharton School, University of Pennsylvania

Joint work with Kan Xu

# Small Data – Why?

- New predictive task
- Imbalanced, rare outcomes
- Nonstationary rewards
- High-dimensional features

# Two Approaches

Improve prediction and downstream decisions despite small data

- **Bandits:** adaptive data collection
- **Transfer learning:** leverage auxiliary data

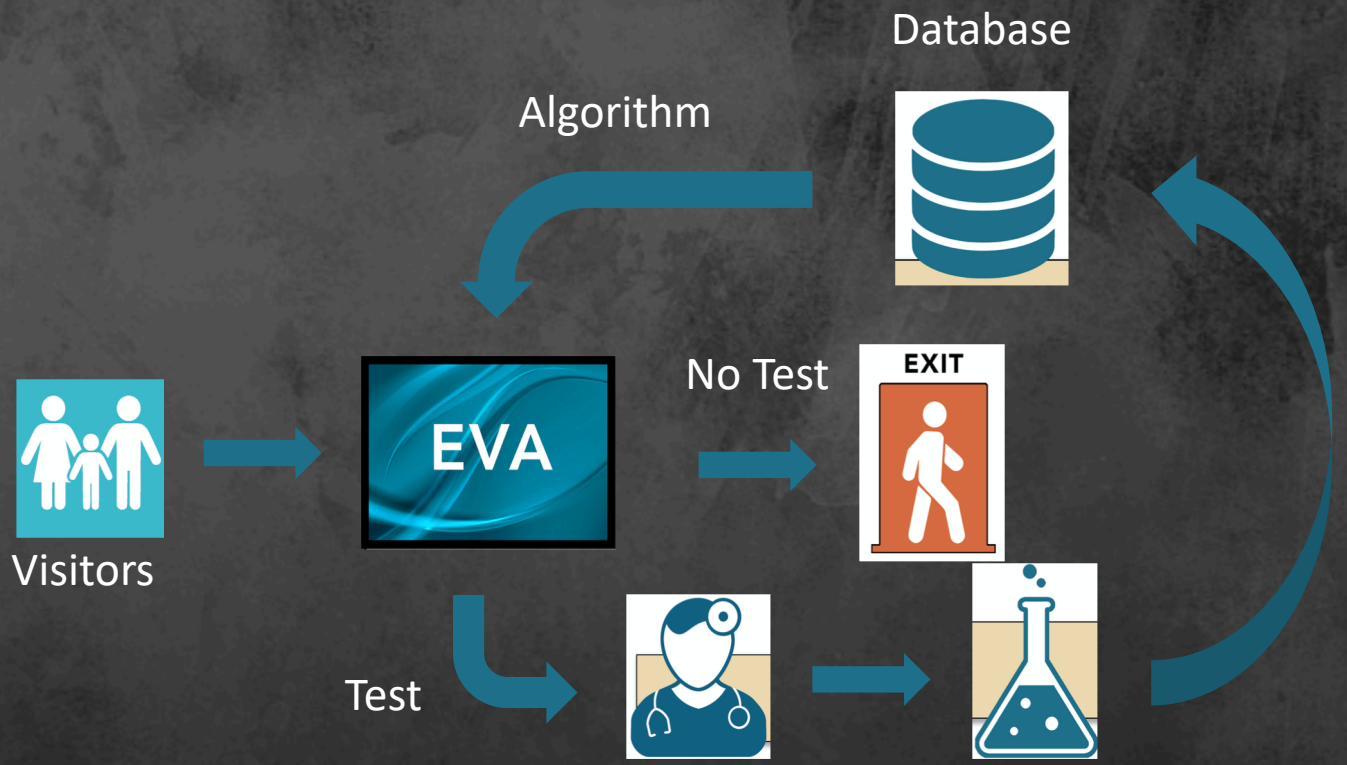
- Transfer learning reduces variance of estimators so that bandit strategy does not *always* explore

# Targeted COVID-19 Testing

30-100k daily tourist arrivals to Greece in Summer 2020

Only 6-8k daily tests

**Goal:** allocate limited tests to identify most positive cases at the border




*Bastani, Drakopoulos, Gupta, Vlachogiannis, Hadjicristodoulou, Lagiou, Magiorkinis, Paraskevis & Tsiodras, "Efficient and Targeted COVID-19 Border Testing via Reinforcement Learning"* ***Nature*** *(2021)*

# Objective

- Type $k$ passenger has test outcome $X_k \sim \text{Bernoulli}(R_k(t))$
- Decision: # passengers to test $N_{k,e}(t)$ of type $k$ at entry $e$ on day $t$
- Maximize expected # infections caught

$$\sum_{t=1}^{T} \sum_{k \in K_t} \sum_{e=1}^{E} N_{k,e}(t) R_k(t)$$

- Our estimate of $\hat{r}_k(t)$ depends on recent test allocations

# Exploration-Exploitation Tradeoff

- **Exploration:** Test "enough" passengers from each type $k$ to estimate whether risky

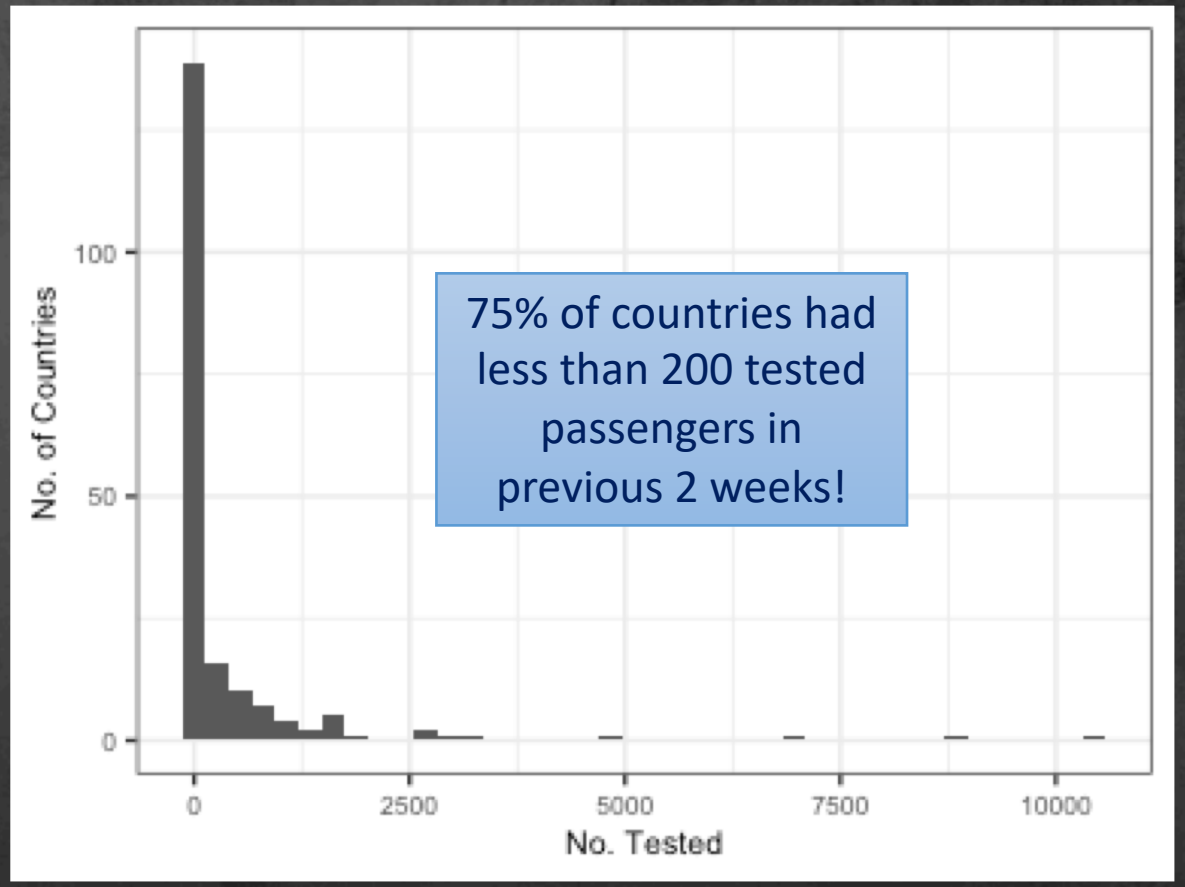
- **Exploitation:** Use these estimates to test arrivals from risky types & quarantine the most positive cases

- Popular solutions: UCB (Lai & Robbins '85, Auer '02), Thompson Sampling (Thompson '33), Gittins index (Gittins '79), …
- Also address batching, delayed feedback, nonstationarity & constraints

# Small Data

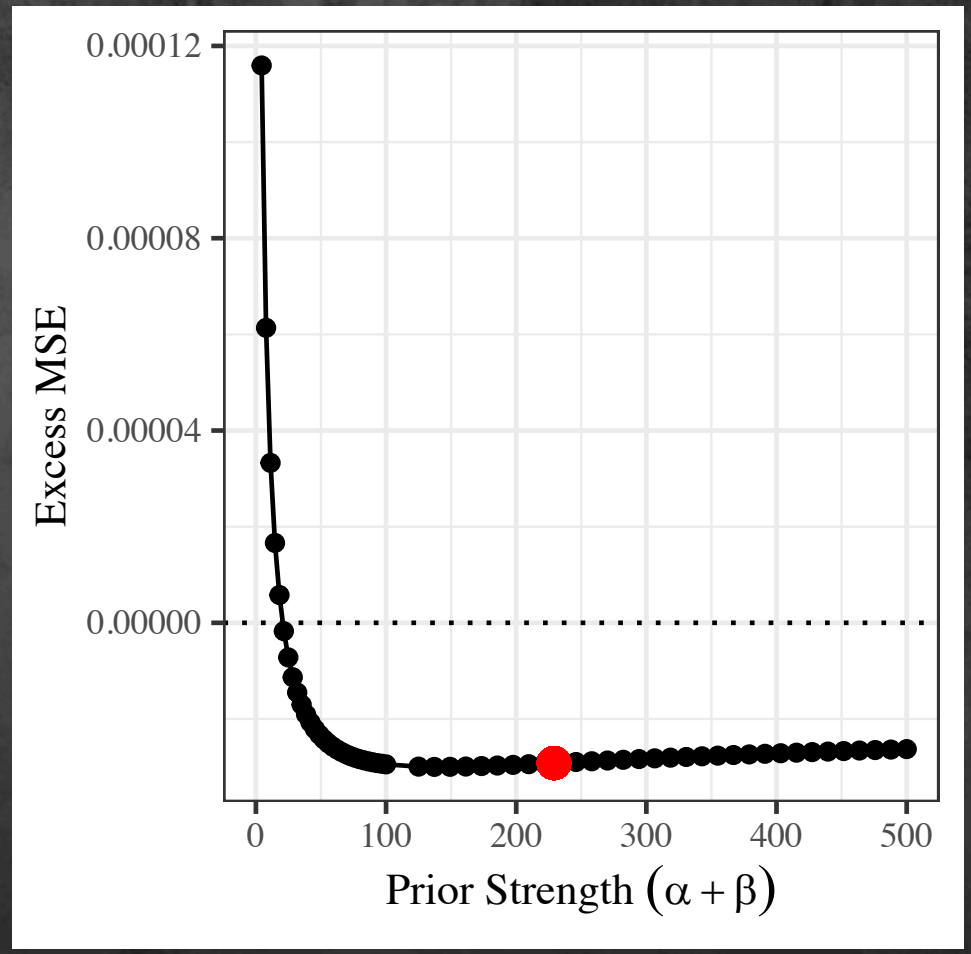
- Imbalanced data: ~1/500 test positive
- High-dimensional features: origin city/country

**Two weeks of Testing Data 1 Oct 2020**



# Transfer Learning

- Zero estimator beats no transfer learning
- Heuristic leveraging LASSO for dimension reduction + empirical bayes to learn across types
- Caught 1.85x more cases than random testing



# Question

- How to learn across many simultaneous, heterogeneous contextual bandits in moderate/high dimension?
  - Prior work on multitask bandits uses ridge regularization (Cesa-Bianchi et al ‘13, Soare et al ‘14, Gentile et al ‘14, Deshmukh et al, ’17) & shared Bayesian prior (Cella et al ‘20, Bastani et al ‘21, Kveton et al ‘21)
  - Does not improve (& sometimes worsens) regret bounds in $d, T$

- Reasonable assumption to improve performance?
- Special attention to “data-poor” instances

# Many bandits

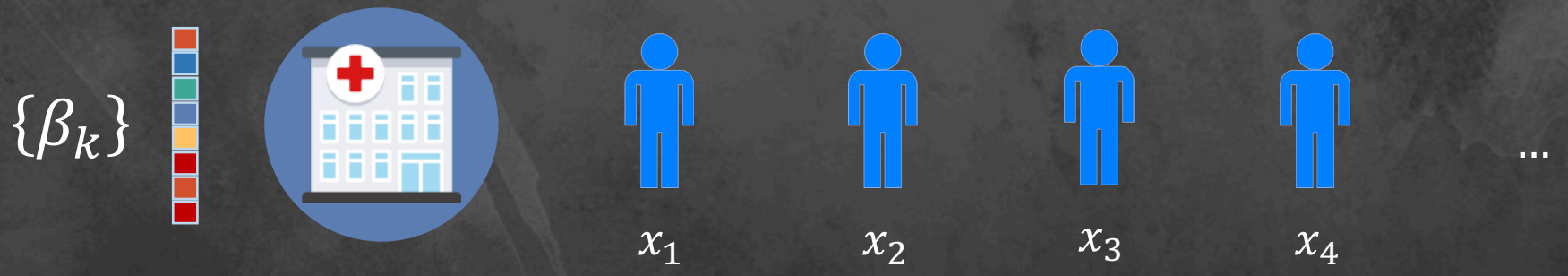
- Simultaneous, heterogeneous problems


*Demand prediction for dynamic pricing or inventory management*


*Patient health risk prediction for care management*

# Formulation: Single Bandit

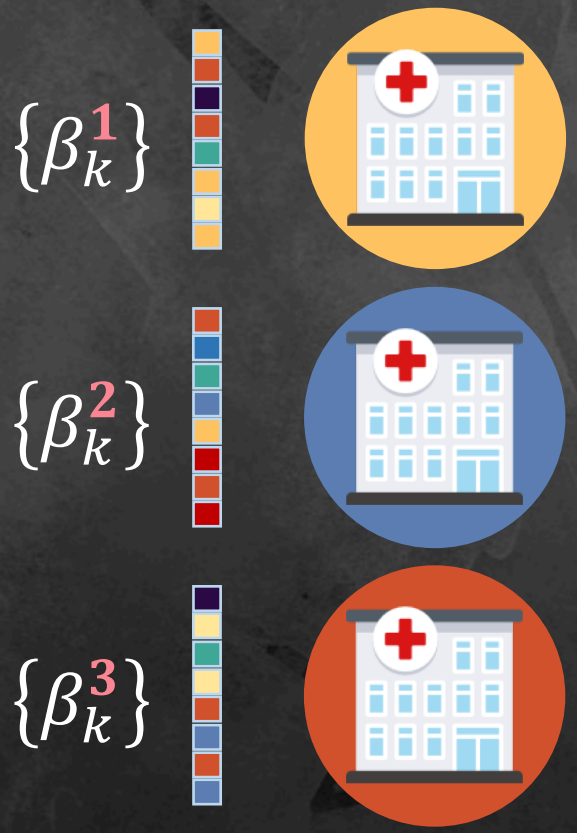
- Observe context $x_t$ at each time step $t$
- Choose an arm $k \in \{1, \dots K\}$ with unknown parameter $\beta_k$
- Observe reward $y_t = x_t^T \beta_k + \epsilon_t$



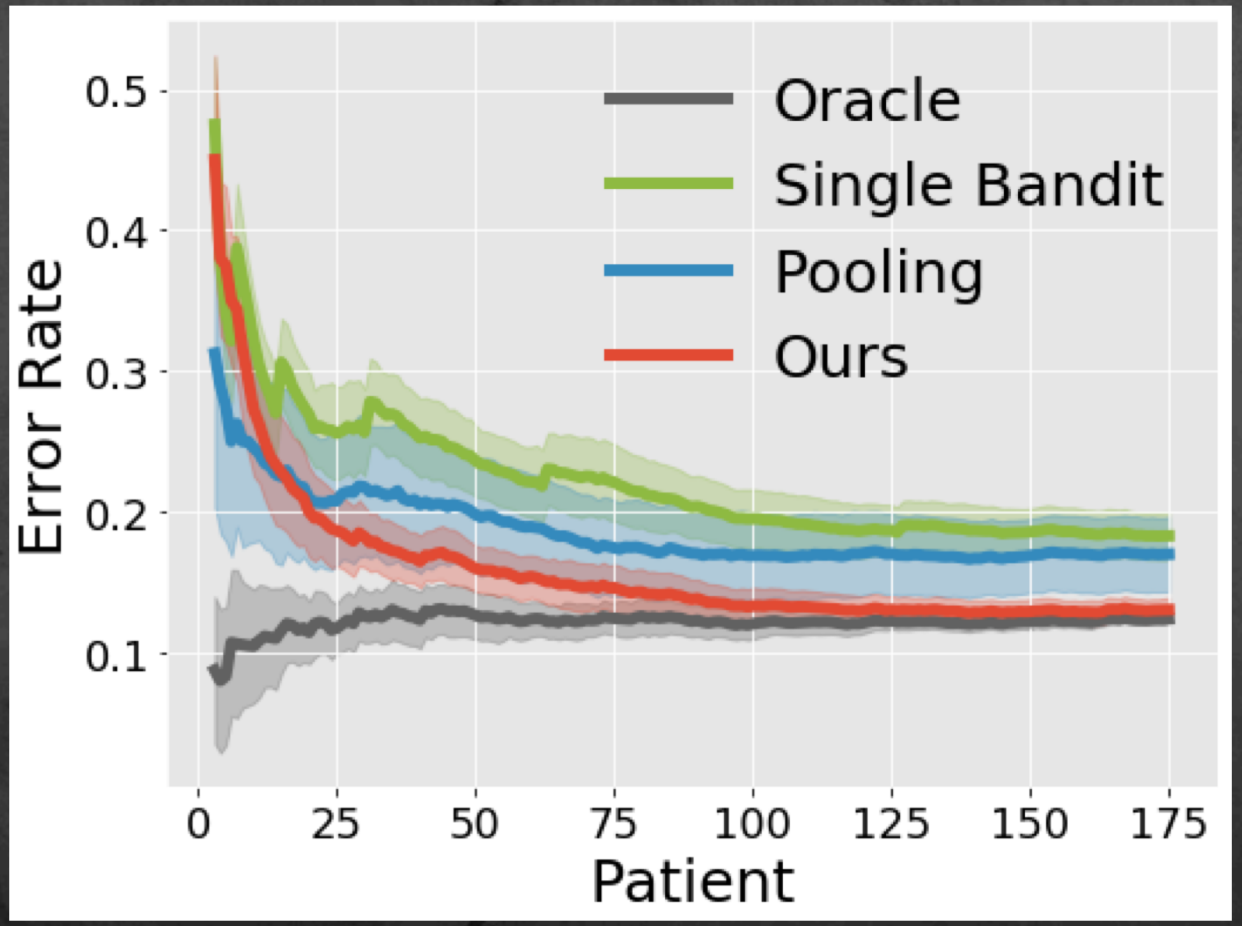

- Devise policy $\pi$ that minimizes regret over time horizon $T$
  - Thompson 1933, Lai and Robbins 1985, Auer 2002, Langford and Zhang 2008, Rusmevichientong and Tsitskilis 2010, Abbasi-Yadkori et al. 2012, Goldenshluger and Zeevi 2013, Agrawal and Goyal 2013, Russo and Van Roy 2014 …

# Formulation: Many Bandits

- $N$ bandit instances, contexts arrive at each w.p. $\{p_1, \dots, p_N\}$
- Each instance $j$ has unknown arm parameters $\left\{\beta_k^j\right\}$

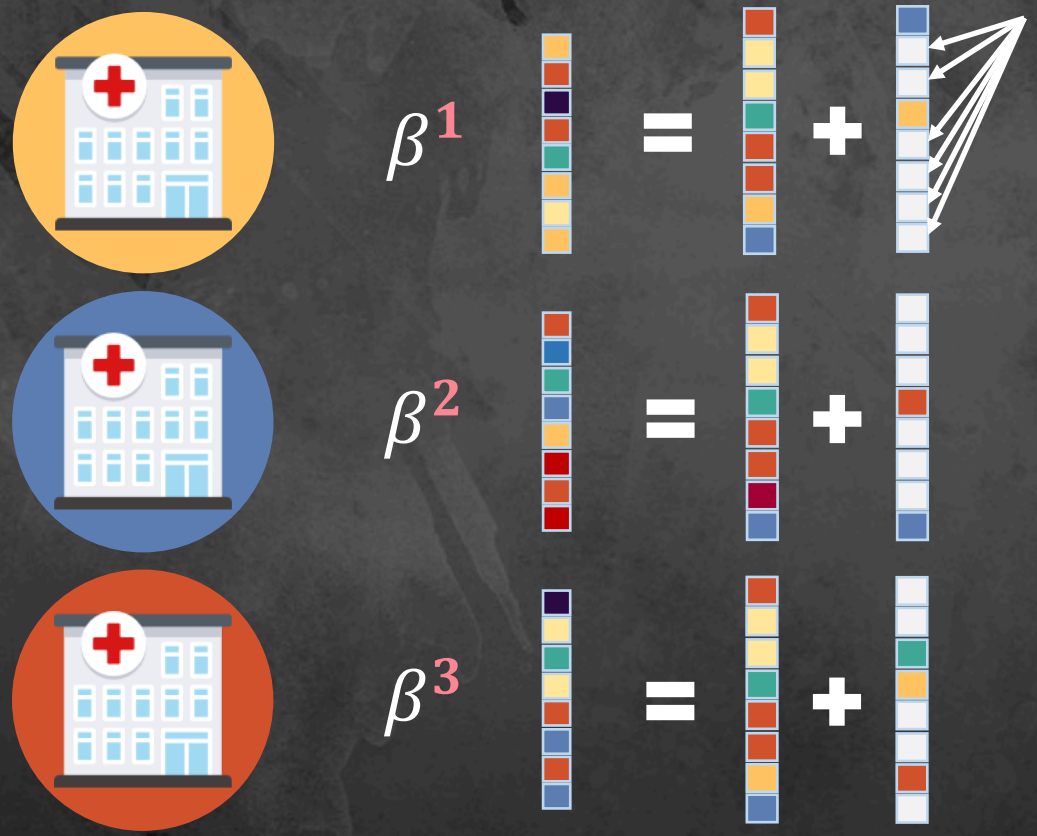


# Diabetes Diagnosis Data




- Claims data across 13 hospitals
- Predict diabetes diagnosis on next visit
- No sharing: high variance
- Pooling: biased

# Sparse Differences




White squares are zero

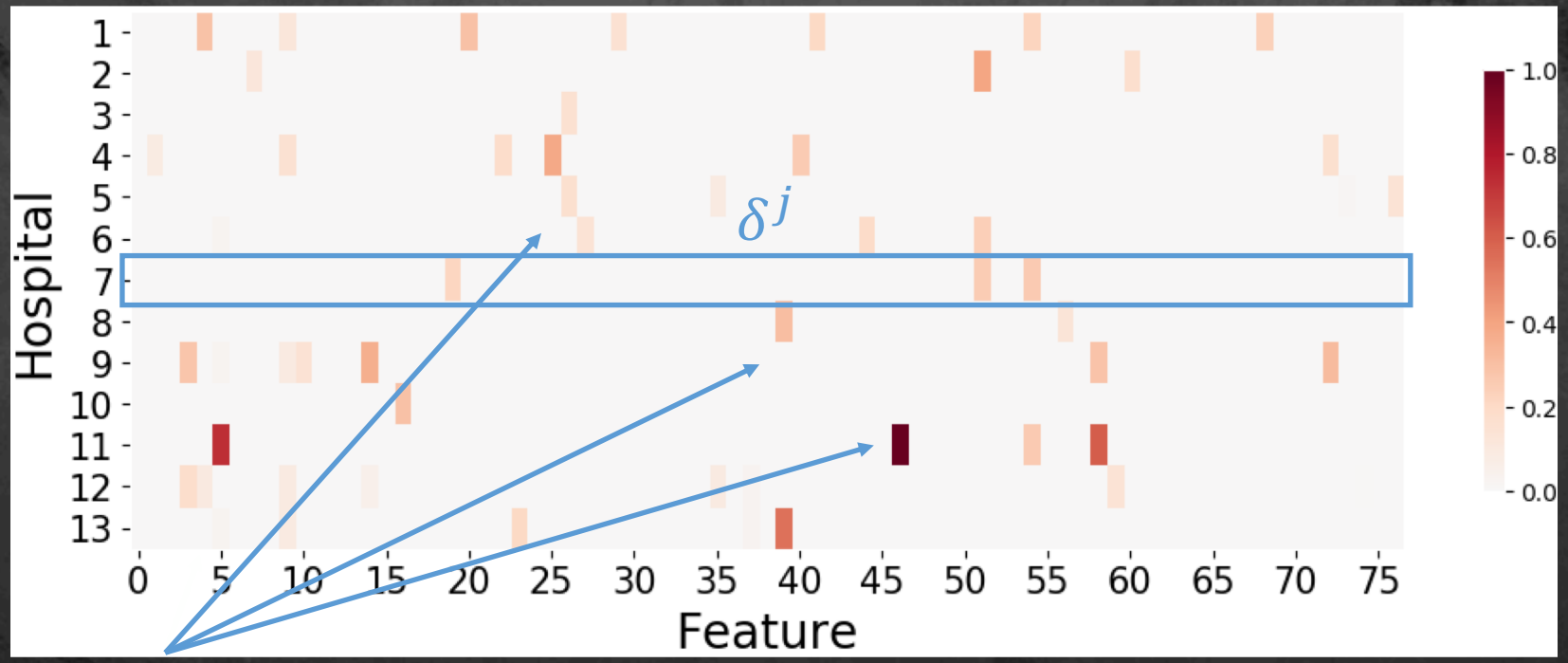
$$\beta^j = \underbrace{\beta^0}_{\text{shared}} + \underbrace{\delta^j}_{\text{instance-specific}}$$

where $\|\delta^j\|_0 \leq s$

and $s \ll d$

# Sparse Differences

- Systematic heterogeneity can be captured by a few features

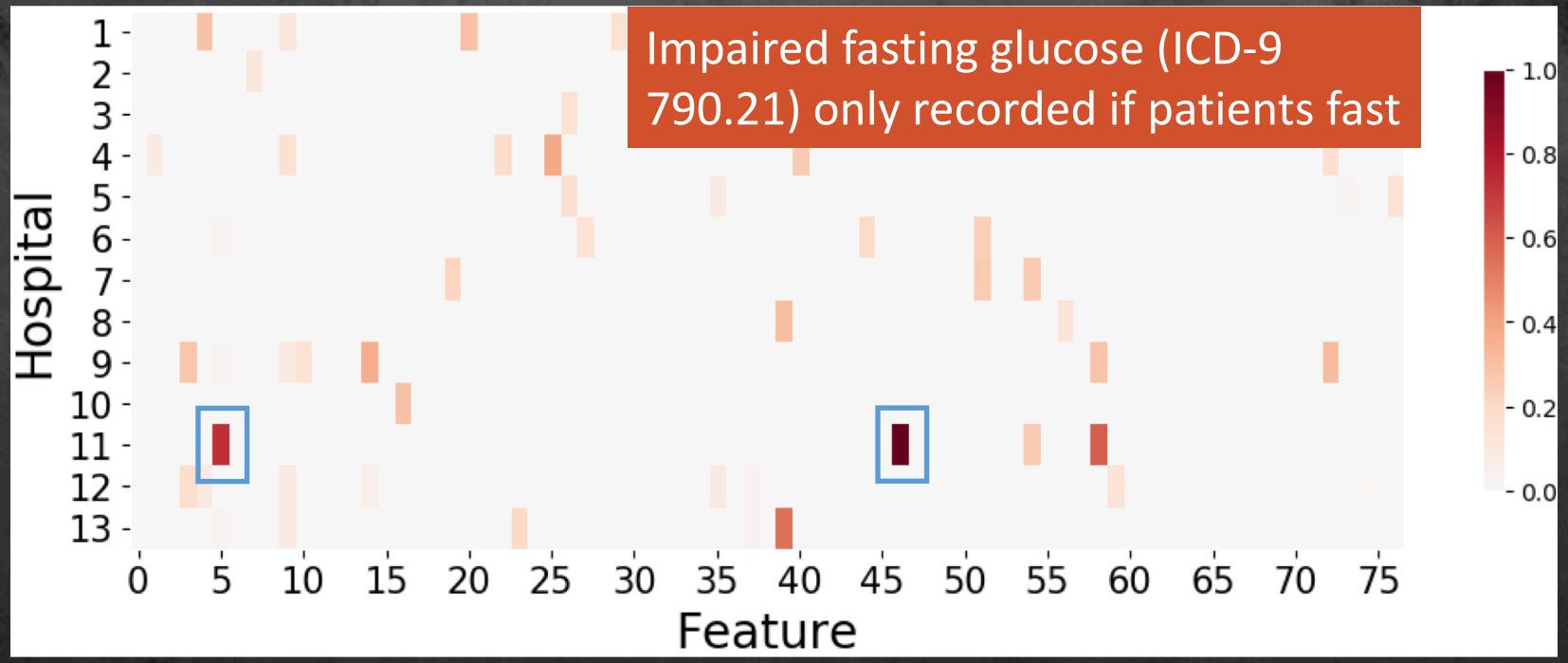



Colored squares are significant

*Bastani, "Predicting with Proxies: Transfer Learning in High Dimension,"* ***Management Science*** *(2021)*

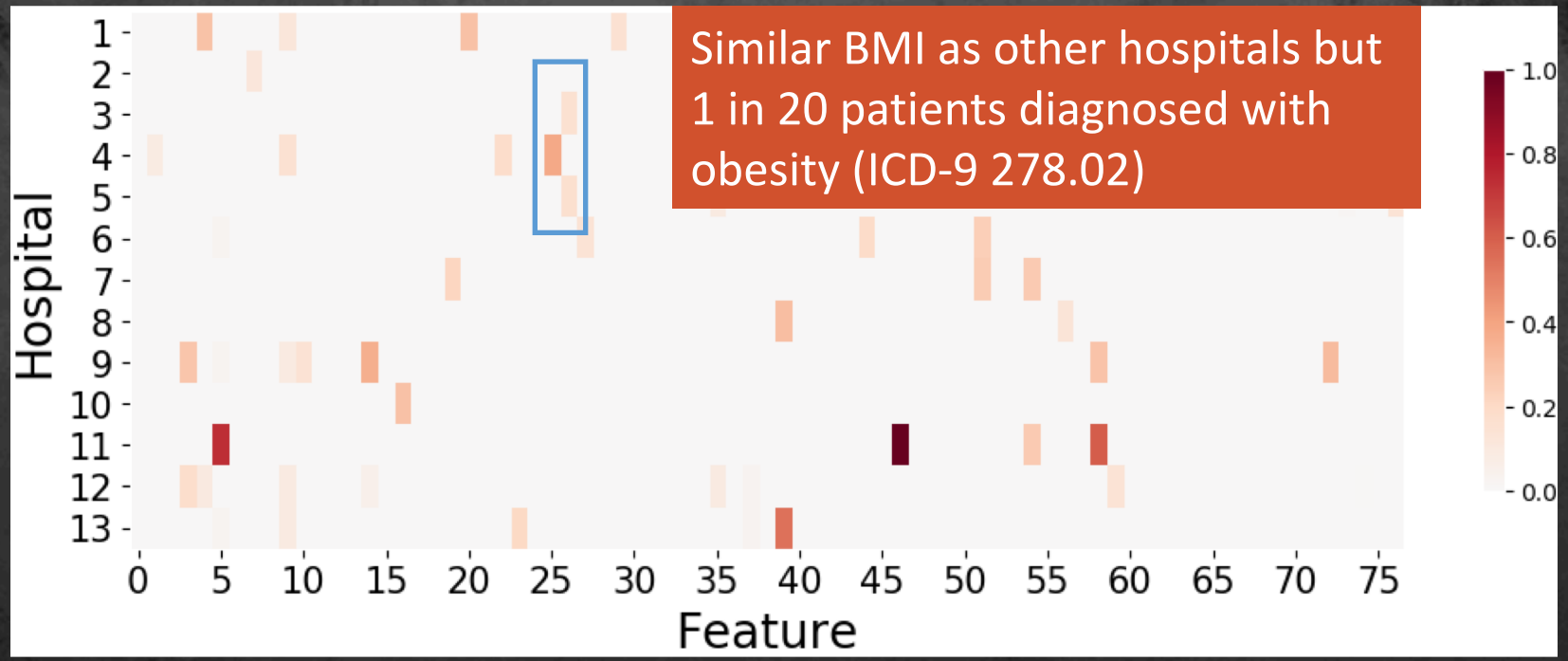
# Mechanisms

- Patient population differences or biases in measurement



# Mechanisms

- Patient population differences or biases in measurement



# Related Literature

- Multitask/transfer/meta-learning: Caruana ‘97, Pan et al. ‘10, Finn et al ’17
  - Ridge (Evgeniou and Pontil ‘04), kernel ridge (Evgeniou et al ‘05), shared Bayesian prior (Raina et al ‘06, Gupta and Kallus ‘21), data pooling (Ben-David et al ‘10, Crammer et al ‘08)
  - Do not improve performance bounds beyond constants (Hanneke and Kpotufe ’20)
- Introduce new multitask estimator
  - Combines robust statistics (Rousseuw ‘91, Lugosi and Mendelson ‘20) and LASSO (Candes and Tao ‘07, Bicket et al ‘09, Buhlmann and Van de Geer ‘11)
- Embed into multitask bandits
  - Ridge regularization (Cesa-Bianchi et al ‘13, Soare et al ‘14, Gentile et al ‘14, Deshmukh et al, ’17) & shared Bayesian prior (Cella et al ‘20, Bastani et al ‘21, Kveton et al ‘21)
  - Do not improve regret bounds beyond constants

# Static Problem

- Data across $N$ "neighboring" instances at sparsity level $s$
- Each instance $j$ has $n_j$ observations given by $\left(X^j, Y^j\right)$

$$Y^j = X^j\left(\beta^0 + \delta^j\right) + \epsilon^j$$

- Goal: estimate $\beta^j = \beta^0 + \delta^j$

**Standard Case**

$n_j \sim n_i$ : hospital $j$ has similar data size as other $i$'s

**Data-poor Case**

$n_j \sim \frac{n_i}{d^2}$ : hospital $j$ has much smaller data than other $i$'s

# Simple Baselines

- No transfer learning: train independent OLS $\hat{\beta}_{ind}^{j}$ on $\left(X^{j}, Y^{j}\right)$
  - Unbiased but high variance when $n_j$ is small

# Error Bounds

$$\sup_{G} \mathbb{E}\left[\left\|\beta^j - \hat{\beta}^j\right\|_1\right]$$

Omitting constants, log terms

| Estimator | Estimation Error | | Bound Type |
|---|---|---|---|
| | *Standard Regime* | *Data-Poor Regime* | |
| Independent $\hat{\beta}^j_{ind}$ | $\frac{d}{\sqrt{n_j}}$ | $\frac{d}{\sqrt{n_j}}$ | Lower |

# Simple Baselines

- Estimate shared model $\beta_0$ by:
  - Averaging independent estimators (Dobriban and Sheng 2021):

$$\hat{\beta}_{avg}^{j} = \frac{1}{N}\sum_{i\in[N]}\hat{\beta}_{j}^{ind}$$

  - Pooling data (Crammer et al. 2008, Ben-David et al. 2010):

$$\hat{\beta}_{j}^{pool} = \left(\sum_{i\in[N]} X^{i\top}X^{i}\right)^{-1}\left(\sum_{i\in[N]} X^{i\top}Y^{i}\right)$$

  - Low variance but high bias

# Error Bounds

$\sup_{G} \mathbb{E}\left[\left\|\beta^j - \hat{\beta}^j\right\|_1\right]$
Omitting constants, log terms

| Estimator | Estimation Error | | Bound Type |
|---|---|---|---|
| | *Standard Regime* | *Data-Poor Regime* | |
| Independent $\hat{\beta}^j_{ind}$ | $\frac{d}{\sqrt{n_j}}$ | $\frac{d}{\sqrt{n_j}}$ | Lower |
| Averaging $\hat{\beta}^j_{avg}$ | $\left\|\delta^j\right\|_1 + \frac{d}{\sqrt{Nn_j}}$ | $\left\|\delta^j\right\|_1 + \frac{1}{\sqrt{Nn_j}}$ | Lower |
| Data Pooling $\hat{\beta}^j_{pool}$ | $\left\|\delta^j\right\|_1 + \frac{d}{\sqrt{Nn_j}}$ | $\left\|\delta^j\right\|_1 + \frac{1}{\sqrt{Nn_j}}$ | Lower |

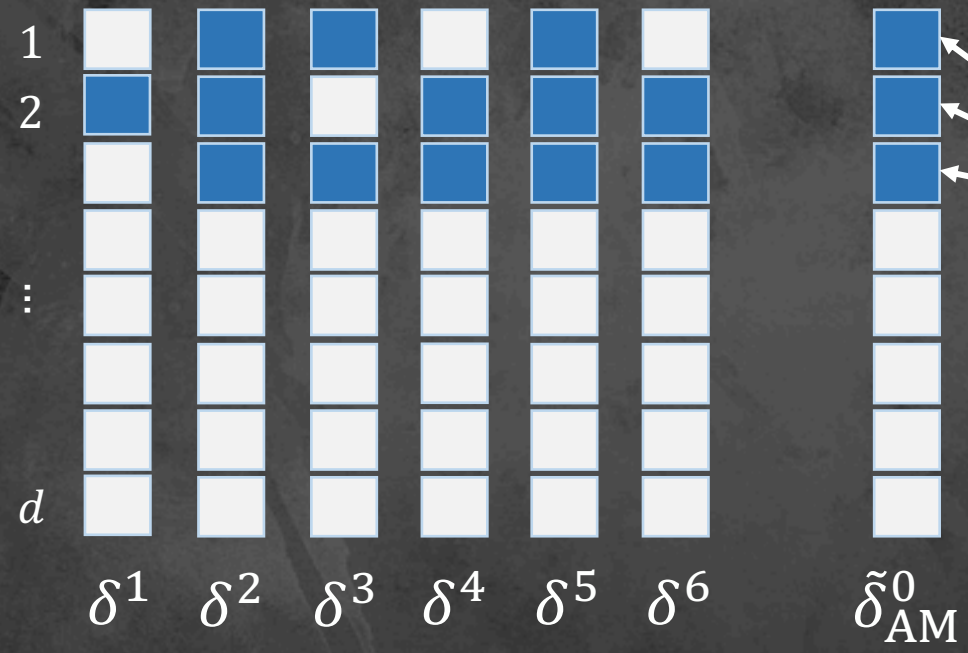
# Averaging Multitask

- Attempt to “debias” the shared model estimate

- Step 1: $\hat{\beta}_{AM}^{0} = \frac{1}{N}\sum_{i\in[N]} \hat{\beta}_{j}^{ind}$
- Step 2: $\hat{\beta}_{AM}^{j} = \arg\min_{\beta} \left\{ \frac{1}{\mathrm{n}_j} \left\| X^j\beta - Y^j \right\|_2^2 + \lambda_j \left\| \beta - \hat{\beta}_{AM}^{0} \right\|_1 \right\}$

- Step 1 converges to $\tilde{\beta}_{AM}^{0} = \beta^0 + \frac{1}{N}\sum_i \delta^i = \beta^0 + \tilde{\delta}_{AM}^{0}$

# Averaging Multitask

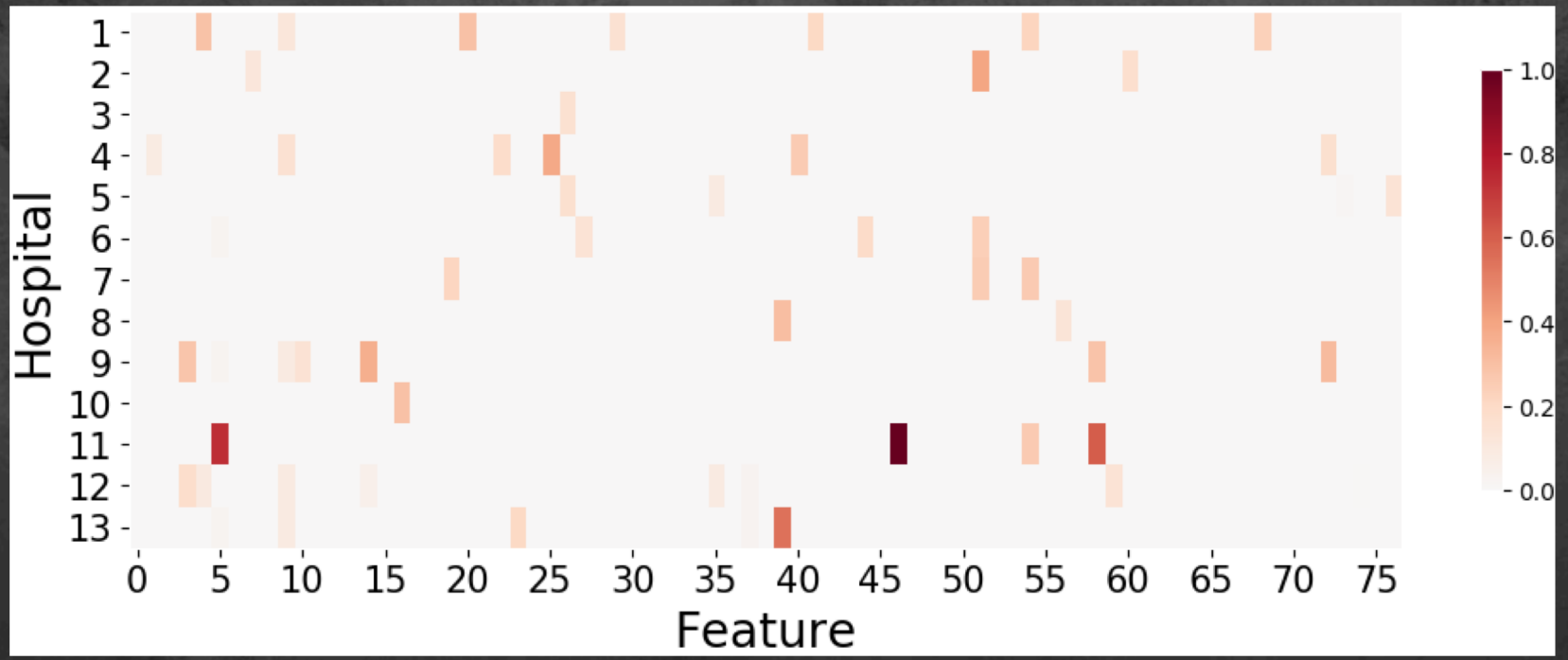



$\|\tilde{\beta}^0_{AM} - \beta^j\|_0 \leq 2s$

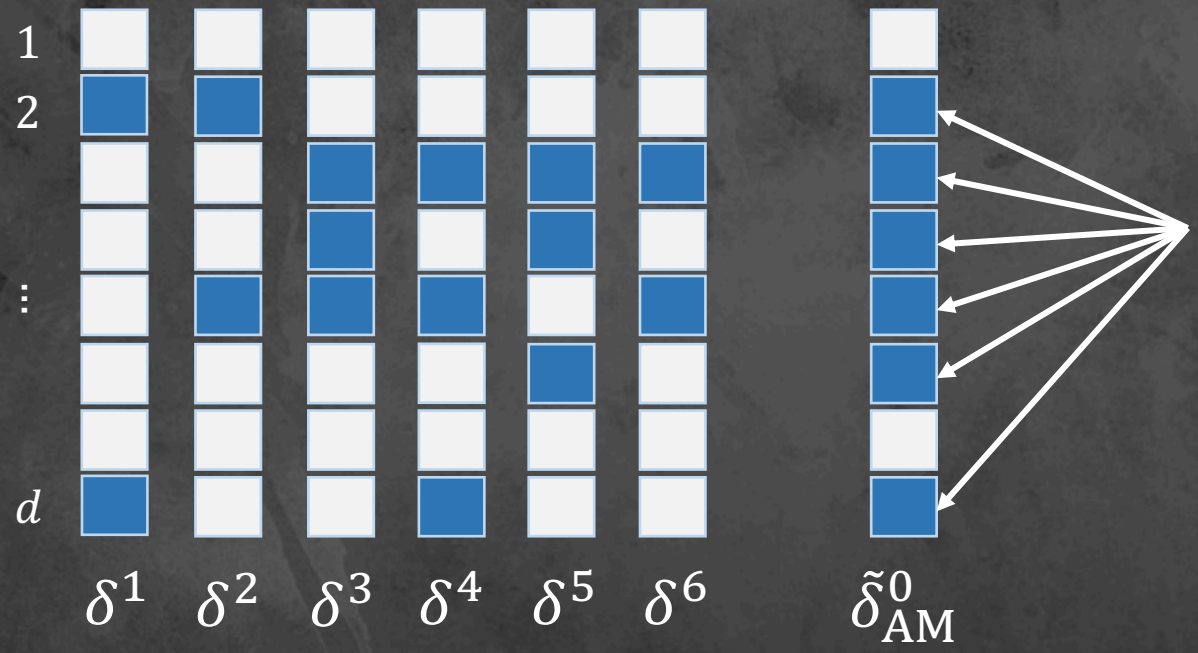
Can use LASSO to de-bias in Step 2

- Blue squares are nonzero

# Recall...


Hospital
1
2
3
4
5
6
7
8
9
10
11
12
13
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
Feature
1.0
0.8
0.6
0.4
0.2
0.0

# Averaging Multitask




$\|\tilde{\beta}^0_{AM} - \beta^j\|_0 \gg s$

Cannot use LASSO to de-bias in Step 2

- Blue squares are nonzero

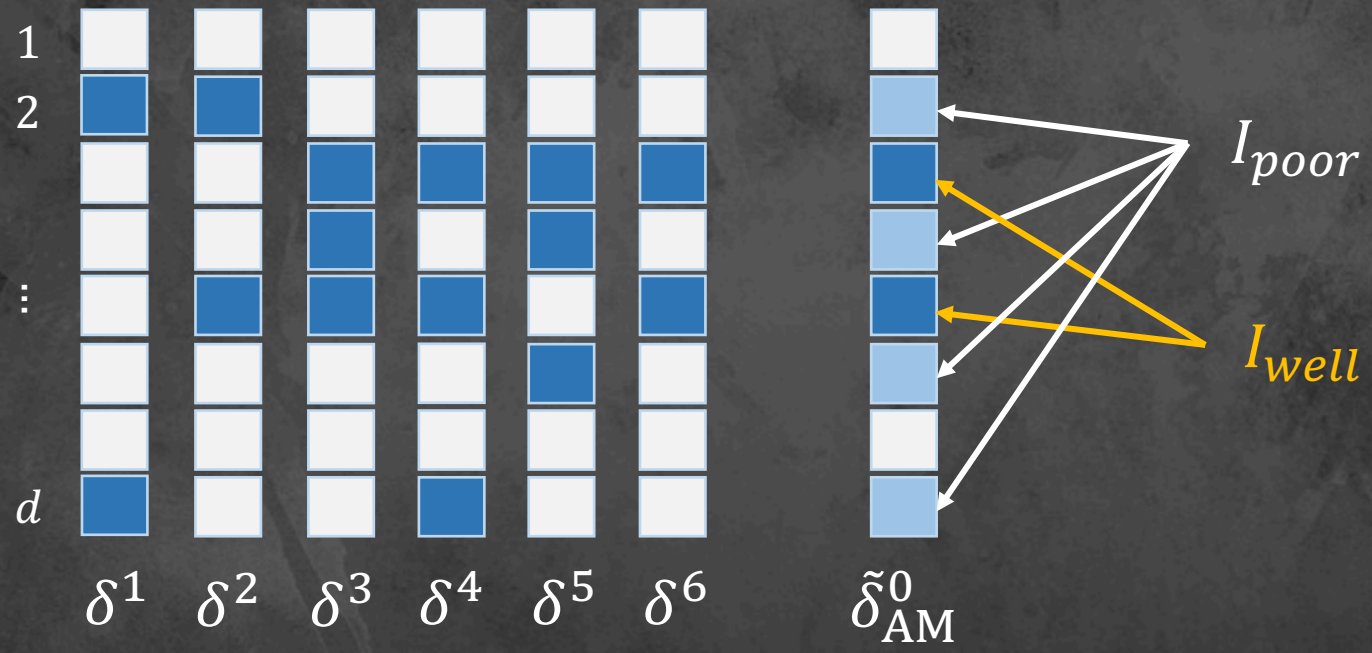
# Error Bounds

$\sup_{G} \mathbb{E}\left[\left\|\beta^j - \hat{\beta}^j\right\|_1\right]$

Omitting constants, log terms

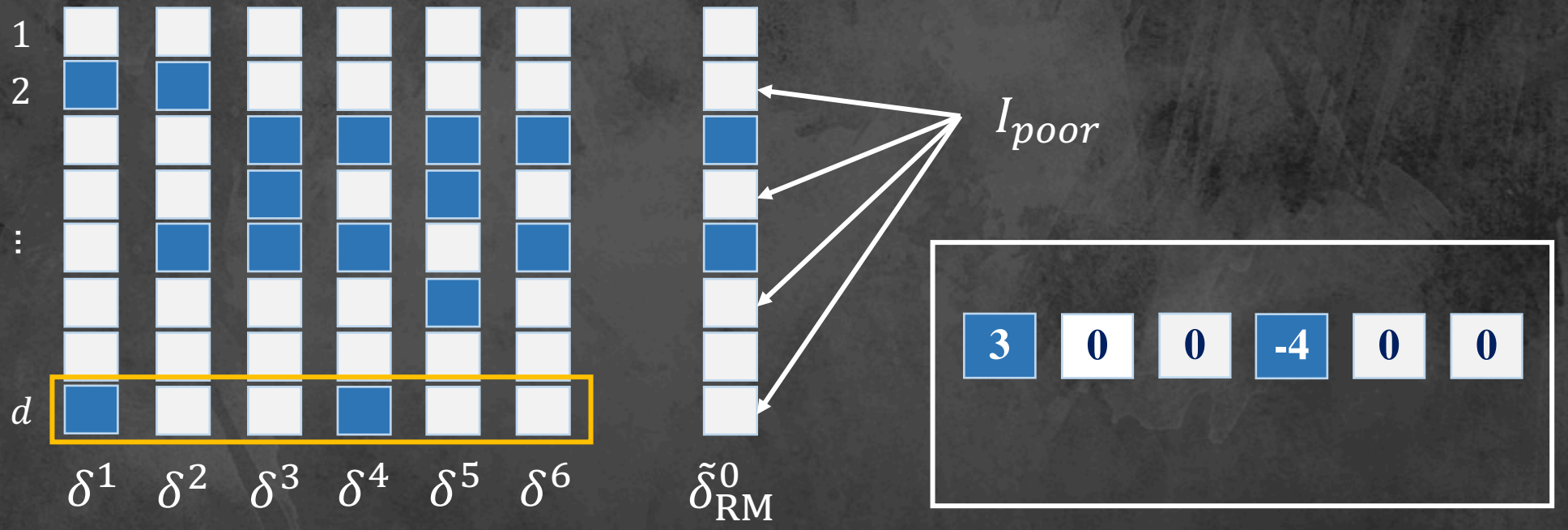
| Estimator | Estimation Error | | Bound Type |
|---|---|---|---|
| | *Standard Regime* | *Data-Poor Regime* | |
| Independent $\hat{\beta}^j_{ind}$ | $\frac{d}{\sqrt{n_j}}$ | $\frac{d}{\sqrt{n_j}}$ | Lower |
| Averaging $\hat{\beta}^j_{avg}$ | $\left\|\delta^j\right\|_1 + \frac{d}{\sqrt{Nn_j}}$ | $\left\|\delta^j\right\|_1 + \frac{1}{\sqrt{Nn_j}}$ | Lower |
| Data Pooling $\hat{\beta}^j_{pool}$ | $\left\|\delta^j\right\|_1 + \frac{d}{\sqrt{Nn_j}}$ | $\left\|\delta^j\right\|_1 + \frac{1}{\sqrt{Nn_j}}$ | Lower |
| Avg Multitask $\hat{\beta}^j_{AM}$ | $\min\frac{\{Ns, d\}}{\sqrt{n_j}} + \frac{d}{\sqrt{Nn_j}}$ | $\min\frac{\{Ns, d\}}{\sqrt{n_j}}$ | Lower |

# Robust Multitask



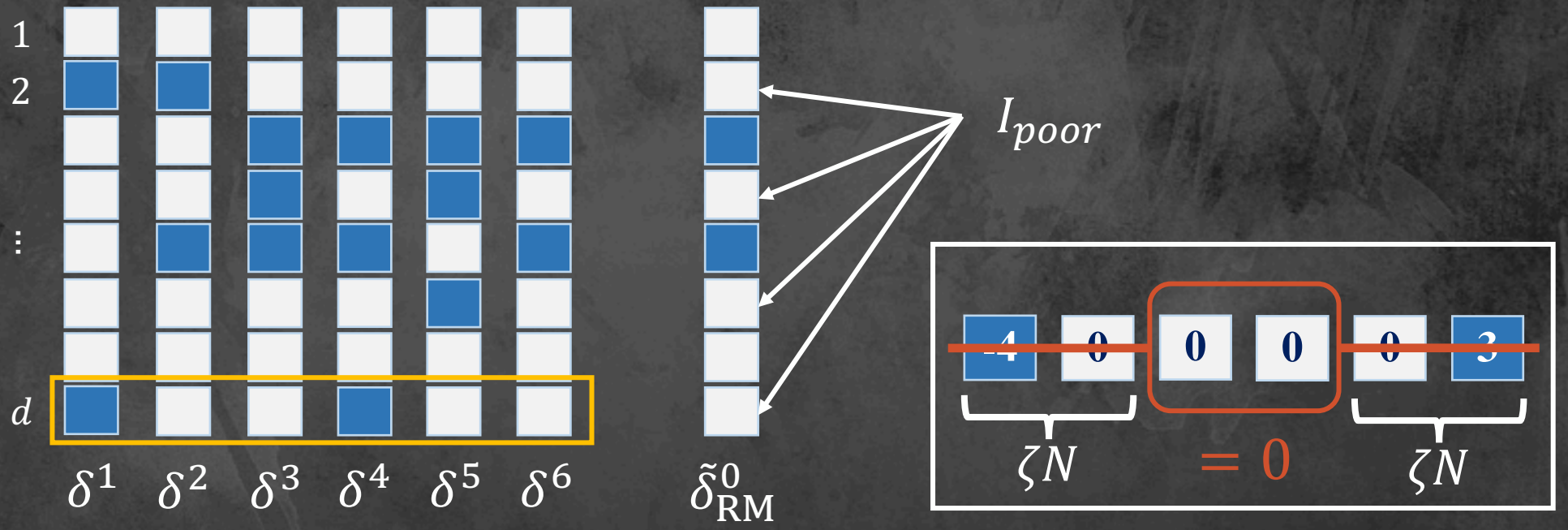

- Two subsets of components:
  - $I_{poor}$: few bias terms ($\zeta N$) have support
  - $I_{well}$: many bias terms have support; by pigeonhole, $|I_{well}| \leq \frac{Ns}{N\zeta} = O(s)$

# Robust Multitask



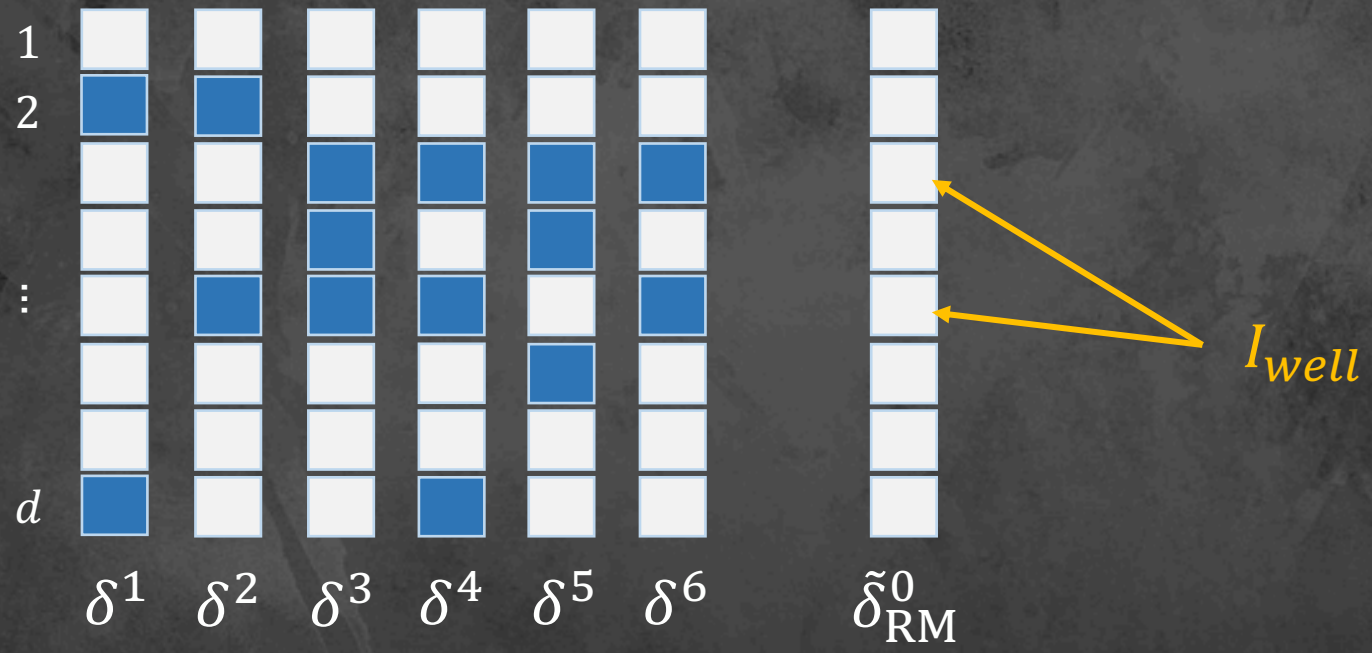

- Step 1: $\hat{\beta}^0_{RM,(i)} = \mathrm{TrimmedMean}\left(\left\{\hat{\beta}^j_{ind,(i)}\right\}_{j=1}^{N}, \zeta + \eta\right)$
  - Converges to $\beta^0_{(i)}$ for $i \in I_{poor}$

# Robust Multitask




- Step 1: $\hat{\beta}^0_{RM,(i)} = \mathrm{TrimmedMean}\left(\left\{\hat{\beta}^j_{ind,(i)}\right\}_{j=1}^{N}, \zeta + \eta\right)$
  - Converges to $\beta^0_{(i)}$ for $i \in I_{poor}$

# Robust Multitask




- Step 2: $\hat{\beta}^j = \min_{\beta} \left\{ \frac{1}{n_j} \left\| X^j \beta - Y^j \right\|_2^2 + \lambda_j \left\| \beta - \hat{\beta}^j_{RM} \right\|_1 \right\}$

# Robust Multitask

- $\beta^j - \hat{\beta}^0_{RM} = \underbrace{(\beta^j - \beta^0)}_{s\text{-sparse}} + \underbrace{(\beta^0 - \tilde{\beta}^0_{RM})}_{O(s/\zeta)\text{-sparse}} + \underbrace{(\tilde{\beta}^0_{RM} - \hat{\beta}^0_{RM})}_{\text{not sparse but small}}$

- Estimated shared model $\tilde{\beta}^0_{RM}$ is robust to corruptions in $I_{poor}$
- Components in $I_{well}$ can be debiased by LASSO

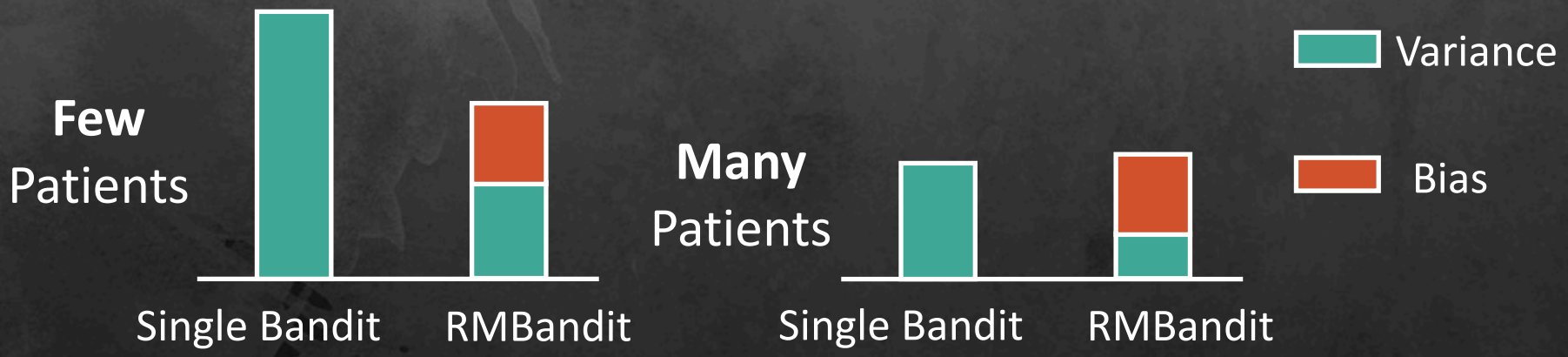
# Error Bounds

$\sup_{G} \mathbb{E}\left[\left\|\beta^j - \hat{\beta}^j\right\|_1\right]$
Omitting constants, log terms

| Estimator | Estimation Error | | Bound Type |
|---|---|---|---|
| | *Standard Regime* | *Data-Poor Regime* | |
| Independent $\hat{\beta}^j_{ind}$ | $\frac{d}{\sqrt{n_j}}$ | $\frac{d}{\sqrt{n_j}}$ | Lower |
| Averaging $\hat{\beta}^j_{avg}$ | $\left\|\delta^j\right\|_1 + \frac{d}{\sqrt{Nn_j}}$ | $\left\|\delta^j\right\|_1 + \frac{1}{\sqrt{Nn_j}}$ | Lower |
| Data Pooling $\hat{\beta}^j_{pool}$ | $\left\|\delta^j\right\|_1 + \frac{d}{\sqrt{Nn_j}}$ | $\left\|\delta^j\right\|_1 + \frac{1}{\sqrt{Nn_j}}$ | Lower |
| Avg Multitask $\hat{\beta}^j_{AM}$ | $\min\frac{\{Ns, d\}}{\sqrt{n_j}} + \frac{d}{\sqrt{Nn_j}}$ | $\min\frac{\{Ns, d\}}{\sqrt{n_j}}$ | Lower |
| **Robust Multitask $\widehat{\boldsymbol{\beta}}^{\boldsymbol{j}}_{\boldsymbol{RM}}$** | $\frac{\sqrt{\boldsymbol{sd}}}{\sqrt{\boldsymbol{n_j}}} + \frac{\boldsymbol{d}}{\sqrt{\boldsymbol{Nn_j}}}$ | $\frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n_j}}}$ | **Upper** |

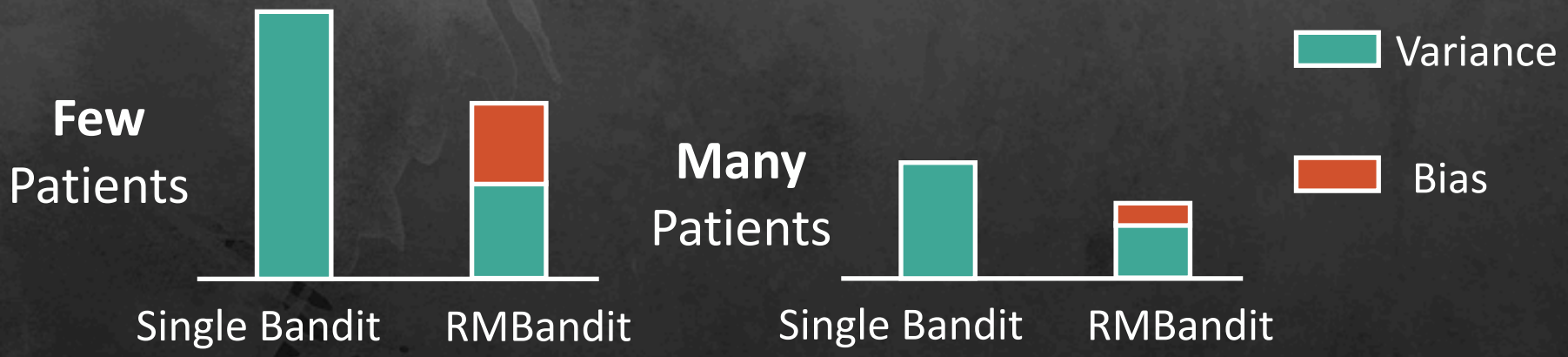
# Robust Multitask Bandit

- Embed robust multitask estimator into linear contextual bandit
- Build on high-dimensional bandits (Bastani & Bayati ’20)
  - Multitask estimator induces correlations across instances, but trimmed mean (Step 1) requires conditional independence → introduce batching strategy
  - Derive trimming path of $\zeta$ over time for bias-variance tradeoff



# Robust Multitask Bandit

- Embed robust multitask estimator into linear contextual bandit
- Build on high-dimensional bandits (Bastani & Bayati ’20)
  - Multitask estimator induces correlations across instances, but trimmed mean (Step 1) requires conditional independence → introduce batching strategy
  - Derive trimming path of $\zeta$ over time for bias-variance tradeoff



# Regret Bounds

| Estimator | Cumulative Regret | |
|---|---|---|
| | *Standard Instance* | *Data-Poor Instance* |
| **Single Bandit** | $d^2 \log^{\frac{3}{2}} d \log T$ | $d^2 \log^{\frac{3}{2}} d \log T$ |
| **RM Bandit** | $\boldsymbol{sd \log d \log^2 T}$ | $\boldsymbol{s^2 \log^2 dT}$ |

- Standard: Improved by factor of $d$, worse by $\log T$
- Data-poor: Exponential improvement in $d$, worse by $\log T$

# Remarks

- Federated approach: only need to share aggregate regression parameters across instances
- Fairness: exponential improvement for data-poor instances
  - Transfer learning significantly reduces “price of fairness” (Bertsimas et al ‘11)
- Production constraints: only need batched offline updates to models

# Diabetes Diagnosis

- 4121 patients across 13 hospitals
- Mean - 317 patients, median - 301 patients
- 76 features

**Demographics:** age, gender, BMI, …

**Patient conditions:** hypertension, obesity, osteoarthritis, insomnia, vitamin D deficiency, …

**Medications:** Prednisone, Ibuprofen, Ambien, …

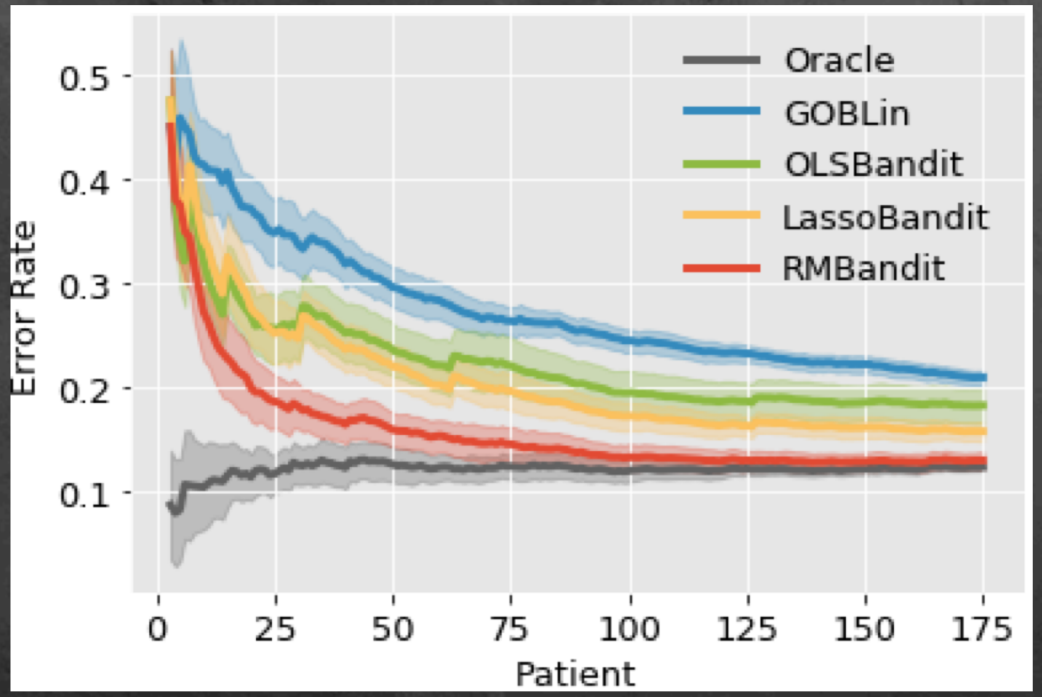
# Diabetes Diagnosis

- 2 arms: Assign treatment or not
- Binary reward
- Compare to:
  - Oracle: estimated true model using all data
  - OLS Bandit (Goldenshluger and Zeevi ’13)
  - Lasso Bandit (Bastani and Bayati ’21)
  - GOBLin (Cesa-Bianchi et al ’13)

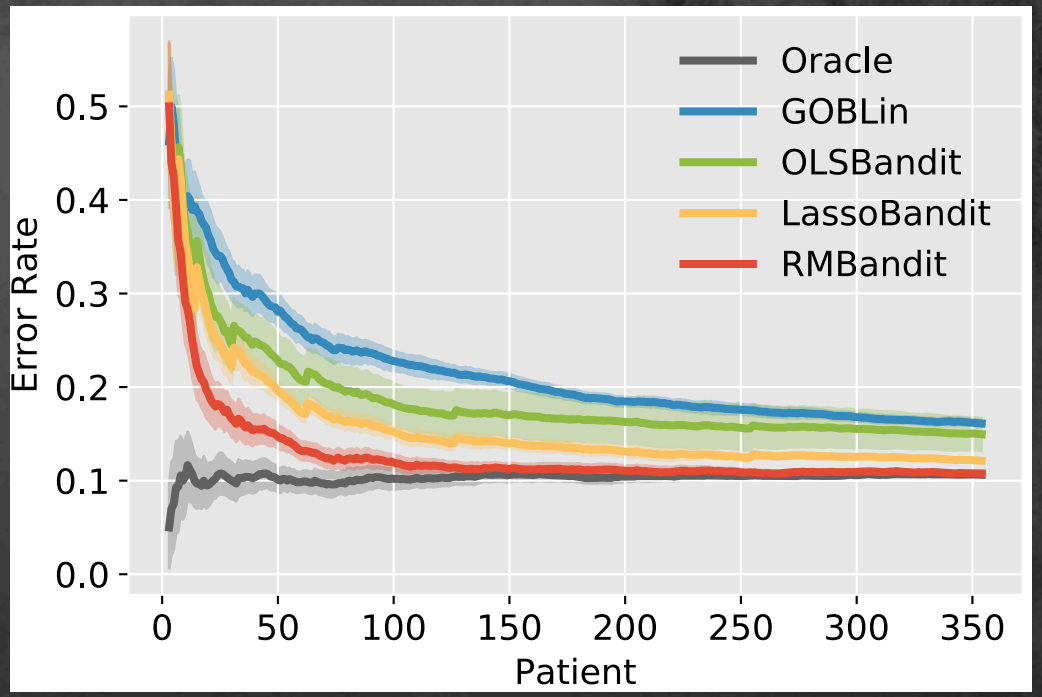
# Diabetes Diagnosis

~30% fewer errors than not using cross-hospital information

**Hospital with 176 patients**




**Hospital with 355 patients**



# Dynamic Pricing

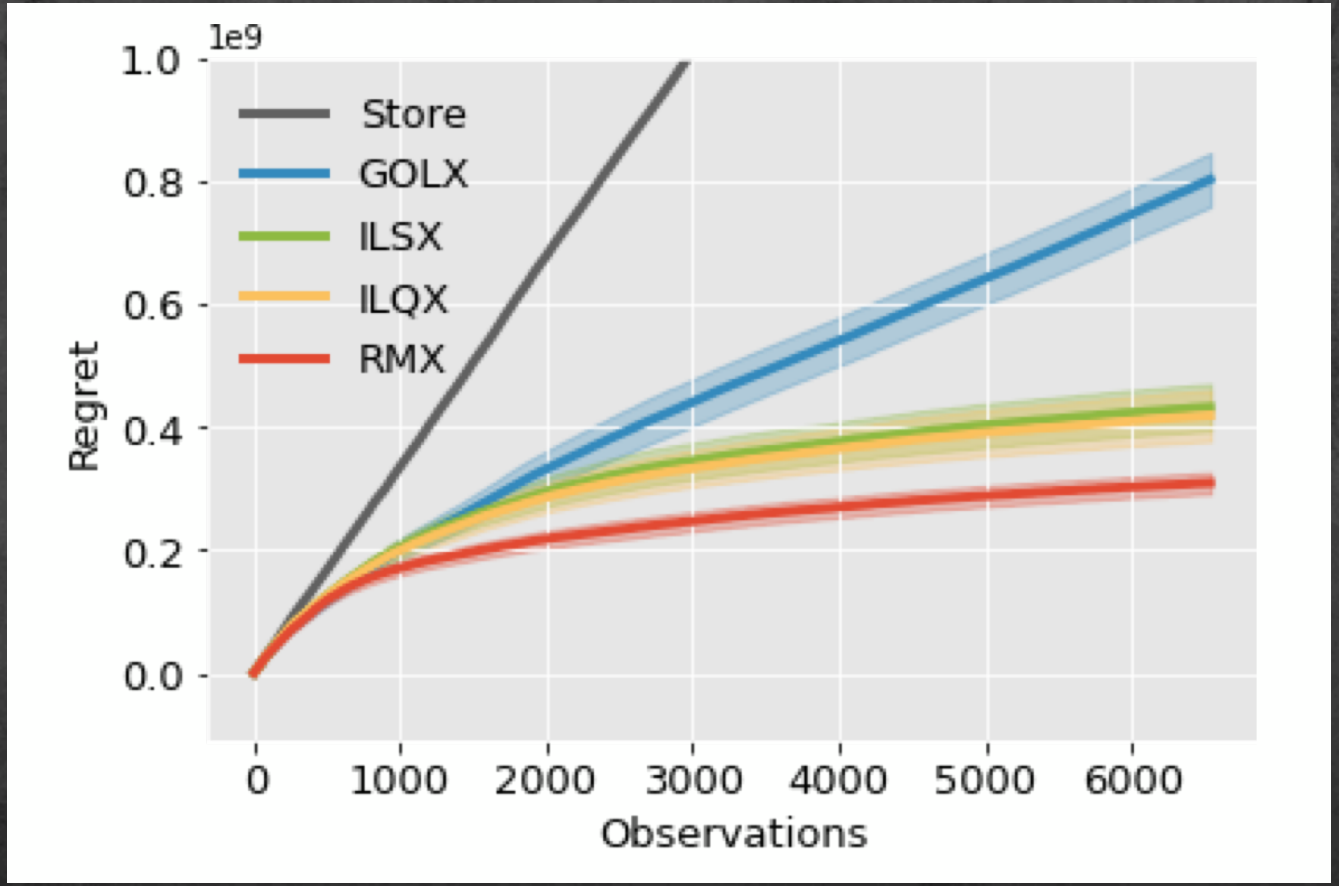
- Meal delivery company, 7 stores
- Weekly sales of 51 plans over 145 weeks
- Mean - 6747 week-plan observations
- 18 features
  - Promotions: email promotions, homepage featured, …
  - Cuisine type: Thai, Italian, …
  - Food category: Pasta, sandwich, salad, …

# Dynamic Pricing

- Linear demand function
- Our adaptation: RMX
- Compare to
  - ILSX (Ban and Keskin ’21): ~OLS Bandit
  - ILQX (Ban and Keskin ’21): ~Lasso Bandit
  - GOLX: GOBLin adapted to pricing setting
  - Store’s policy: Observed from data

# Dynamic Pricing

~6% revenue increase than not using cross-store information



# Summary

- Transfer learn effectively across simultaneous, heterogeneous bandits

- Novel robust multitask estimator
- Improve regret bounds in context dimension $d$
  - Exponential improvement for data-poor instances

- Related: combining short- & long-term outcomes in clinical trials (Anderer, Bastani & Silberholz; MS ‘21), meta Thompson Sampling (Bastani, Simchi-Levi & Zhu; MS ‘21)

# Thank you!

Questions? hamsab@wharton.upenn.edu