Decision-Aware Reinforcement Learning

Hamsa Bastani

Wharton School, University of Pennsylvania

Joint work with

Osbert Bastani, Angel Chung, Vahid Rostami, Johnna Sundberg, Musa Komeh, Ashley Schmidt & Lawrence Sandi

Learning and Optimization

- **Goal:** Given response y (e.g., today's demand), compute decision z (e.g., inventory to order) to minimize a known **decision loss** ℓ:

$$z^*(y) = \arg\min_z \ell(z; y)$$

- **Problem:** Optimization parameters y are unknown

- **Strategy:** Predict y based on covariates x (e.g., yesterday's demand)
  - For problems with state, equivalent to model-based offline RL

Learning and Optimization

- **Training phase:** Given examples $\{(x_i, y_i^*)\}$, train a function f_θ to predict y given x:

$$\hat{\theta} = \arg\min_{\theta} \sum_i \tilde{\ell}(f_\theta(x_i); y_i^*)$$

- **Testing phase:** Given a new x, form prediction $f_\theta(x)$ ("predict") and choose decision $z^*(\hat{y})$ ("optimize")
- **Key question:** What prediction loss $\tilde{\ell}(\hat{y}; y^*)$ to use in training?

Learning and Optimization

- **Decision-blind prediction loss:** Use a standard loss such as MSE:

$$\tilde{\ell}_{\mathrm{MSE}}(\hat{y}; y^*) = (\hat{y} - y^*)^2$$

- **Decision-aware prediction loss:** Use the decision loss

$$\tilde{\ell}(\hat{y}; y^*) = \ell(z^*(\hat{y}); y^*) = \ell(\arg\min_z \ell(z; \hat{y}); y^*)$$

- **Problem:** How to compute $\tilde{\ell}$?

Prior Work

- [Elmachtoub & Grigas (2021)] Approximate decision-aware prediction loss $\tilde{\ell}$ in the case where:
  - Decision loss ℓ is a linear program
  - Learning restricted to objective (constraint parameters are known)
  - Prediction function $f_\theta(x) = \theta^\top x$ is linear
- [Wilder et al (2019a/b)] Heuristics to learn a surrogate loss

- We are interested in:
  - General decision loss (nonlinear, unknown constraint parameters)
  - *Any* prediction functions (e.g., f_θ is a random forest)

Approximating the Loss

- We can Taylor expand in $\hat{y}$ around y^* (works well if $\hat{y} \approx y^*$)

$$\ell(z^*(\hat{y}); y^*) \approx \ell(z^*(y^*); y^*) + \nabla_z \ell(z^*(y^*); y^*)^\top \nabla_y z^*(y^*)(\hat{y} - y^*)$$

- First term (optimal performance) is constant and can be ignored
- Accounts for:
  - Effect of prediction on decision
  - Effect of decision on decision loss

Training Objective

- Prediction model objective is

$$\arg\min_{\theta} \sum_{i} \underbrace{\nabla_z \ell(z^*(y_i^*); y_i^*)^\top \nabla_y z^*(y_i^*)}_{\text{constant in } \theta} (f_\theta(x_i) - y_i^*)$$

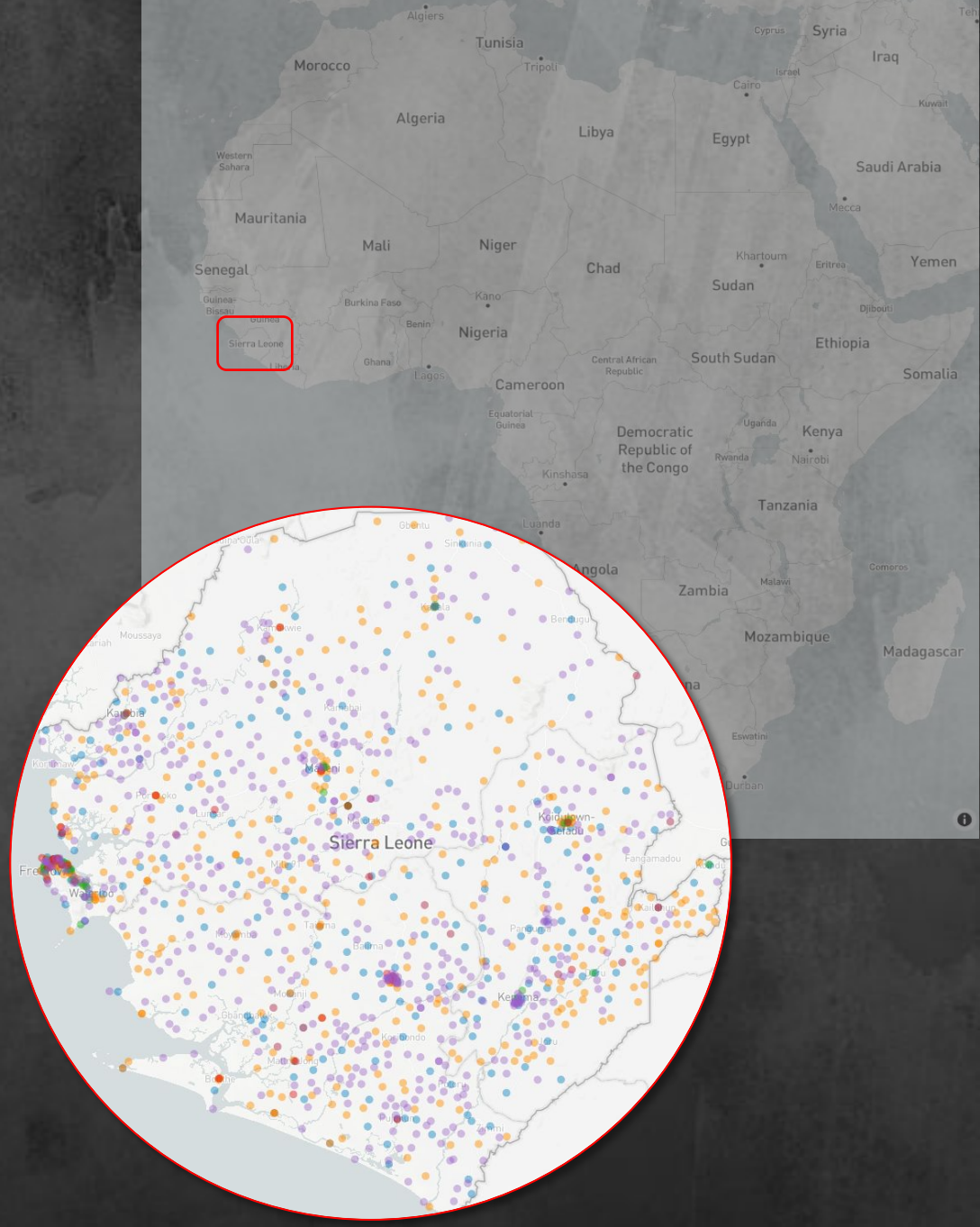
- Can be interpreted as re-weighting training examples (x_i, y_i^*)
- Compute gradient through OPT objective and OPT decision
  - Can be computed efficiently (Amos & Kolter, 2017)

Global Health Allocations

1260 health facilities in Sierra Leone

- Severe mismatch & inequality in health supply chains
- 42% of essential medicine needs currently *not* fulfilled

Goal: allocate limited inventory to facilities with highest forecasted need

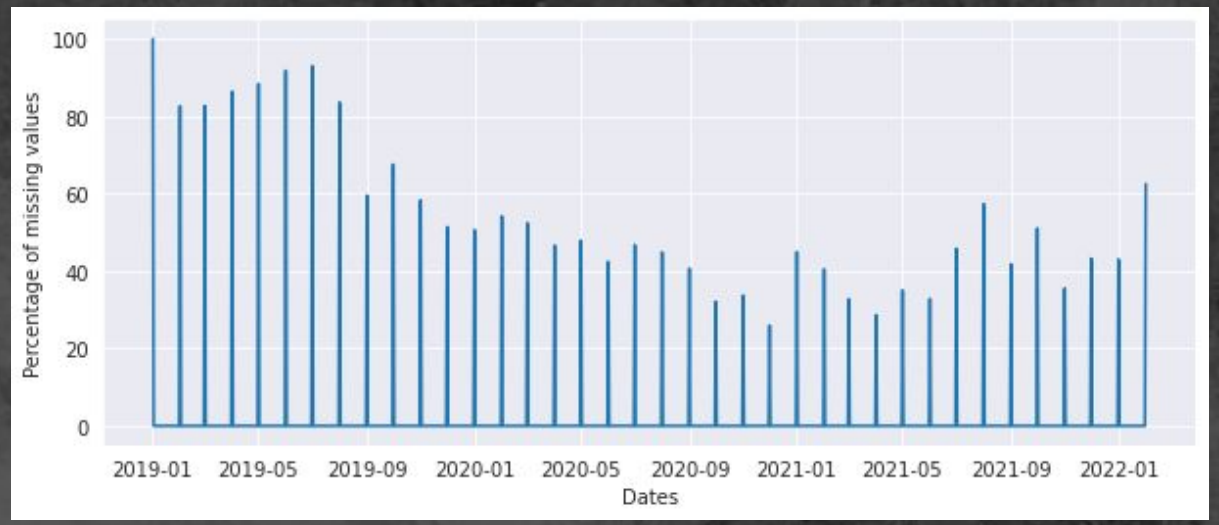


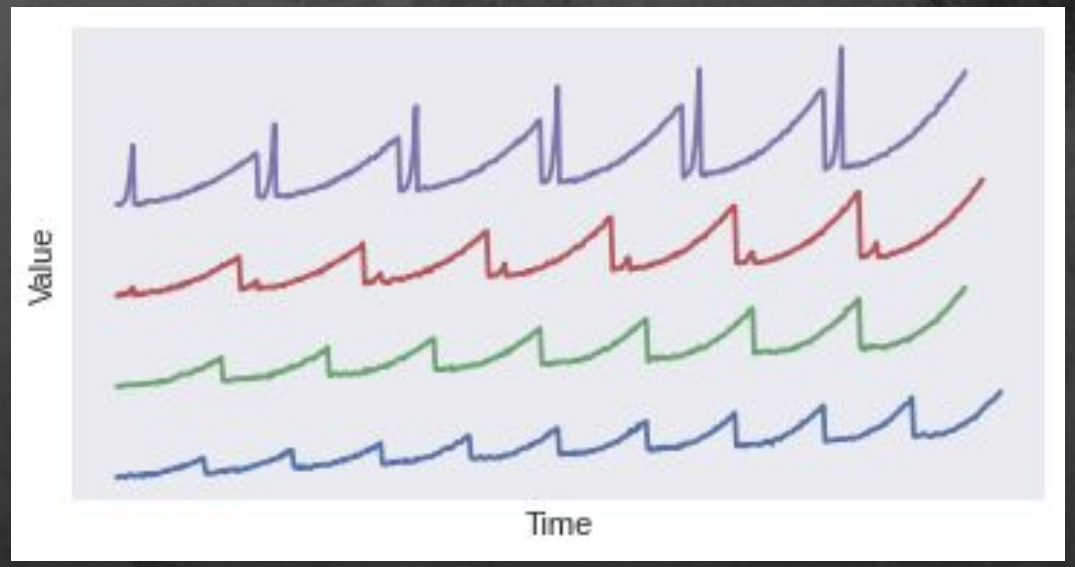
Focal Essential Meds

- Child Health <5 years of age
  - Amoxicillin 250mg, Dispersible, Tab
  - Oral Rehydration Salts (ORS), Sachet (correlation to zinc)
  - Zinc Sulphate 20mg, Tab (correlation to ORS)
- Maternal Health
  - Oxytocin 10IU, Inj, Amp
  - Magnesium Sulphate 50%, Inj, 10ml, Amp
- Family Planning (adolescent health, women of child bearing age)
  - Depot Medroxyprogestrone Acetate (Depo-Provera) 150 mg/ml, Pdr for Inj
  - Ethinylestradiol & Levonorgestrel (Microgynon 30) 30mcg & 150mcg, Tab
  - Jadelle- Levonorgestrel two rod 150mg, implant

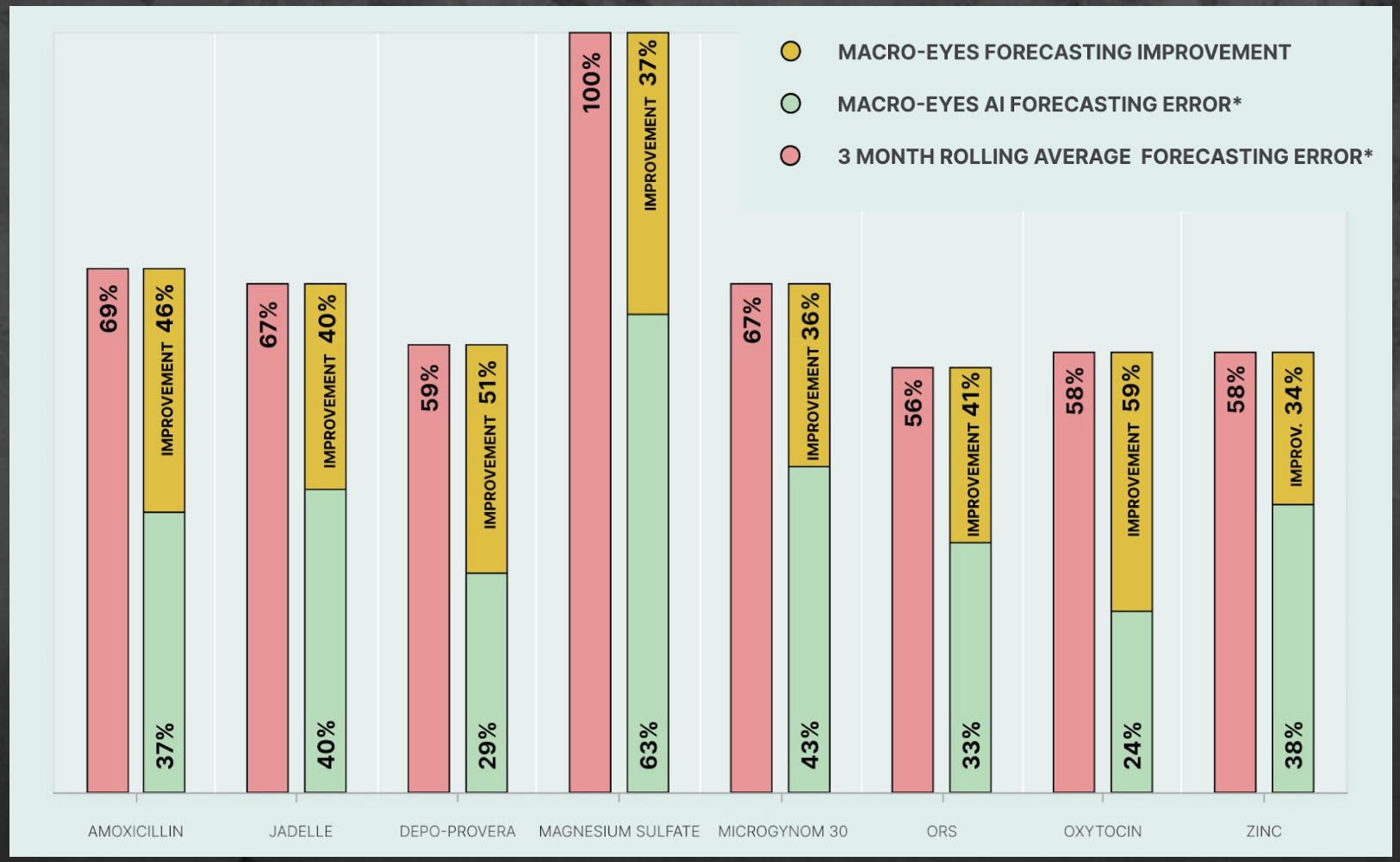
Data

- Dhis2, msupply forms
  - Widely used in Mozambique, Cote d'Ivoire, Rwanda, DRC, Chad, etc
- Significant % of missing values □ imputation
- 1000s of separate time series □ meta-learning
  - Leverage cross-product, cross-facility correlations
- Random forest "meta-model"





Out-of-Sample Results



Improve demand forecasts by **34-59%** on held out test set month

Stochastic Optimization

- **Decision:** allocations a_n^* across N facilities
- **Objective:** minimize cost of **unmet demand** at each location

$$\ell_n = \max\{\xi_n - s_n - a_n, 0\}$$

  - Current inventory s_n, demand ξ_n
- **Constraints:** fixed budget b, each district cannot hold more than its capacity c_n
- **Predictions:** draw random demands $\xi_i^{(k)}$ at each facility based on estimated distribution

$$\begin{aligned} a^* = \arg\min_{a \in \mathbb{R}^N_{\geq 0}} \quad & \sum_{k=1}^{K} \sum_{n=1}^{N} \ell_n^{(k)} \\ \text{subj. to} \quad & \sum_{n=1}^{N} a_n \leq b \\ & \ell^{(k)} \geq \xi^{(k)} - s - a \\ & \ell^{(k)} \geq 0 \\ & s + a \leq c \end{aligned}$$

* Efficient linear program with sample average approximation

For this LP…

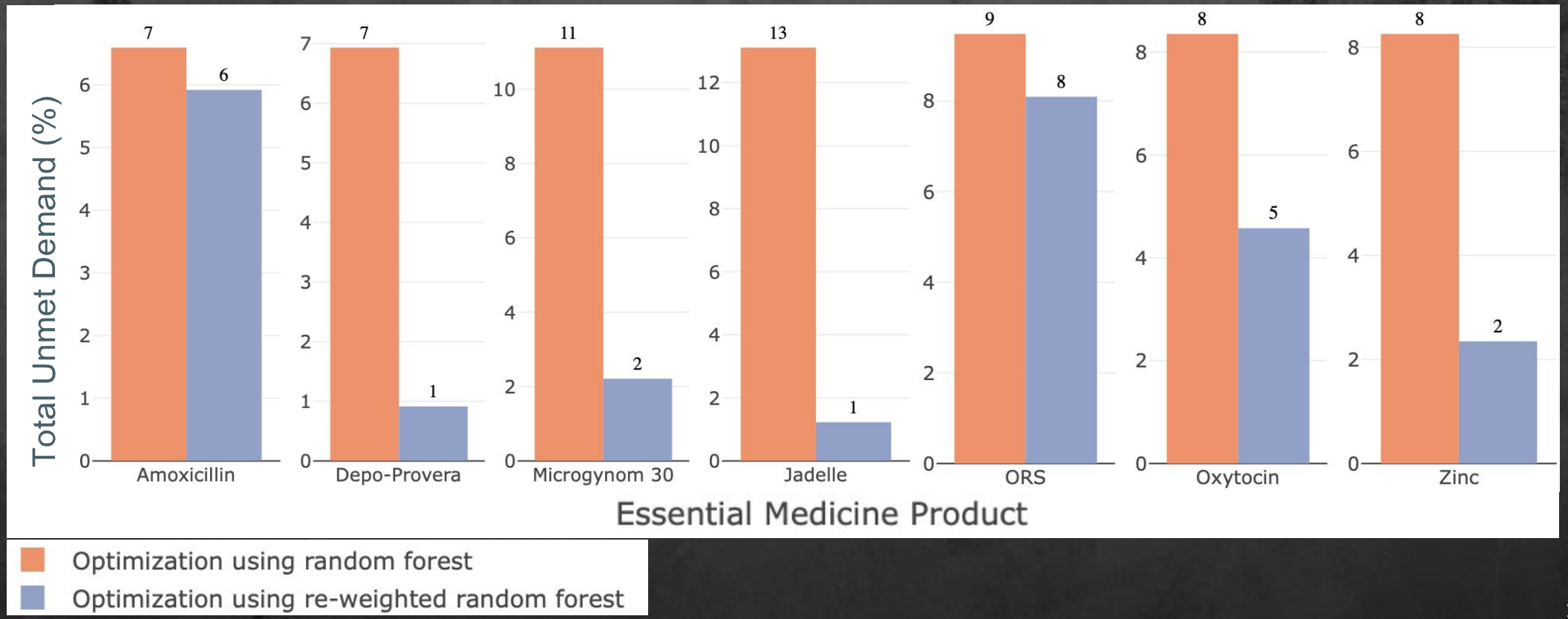
- Prediction model objective is approximately

$$\arg\min_{\theta} \sum_{k=1}^{K} \sum_{n=1}^{N} \mathbb{I}\left(\xi_n^{(k)} \geq s_n + a_n\right) \cdot \left| f_\theta(x_n) - \xi_n^{(k)} \right|$$

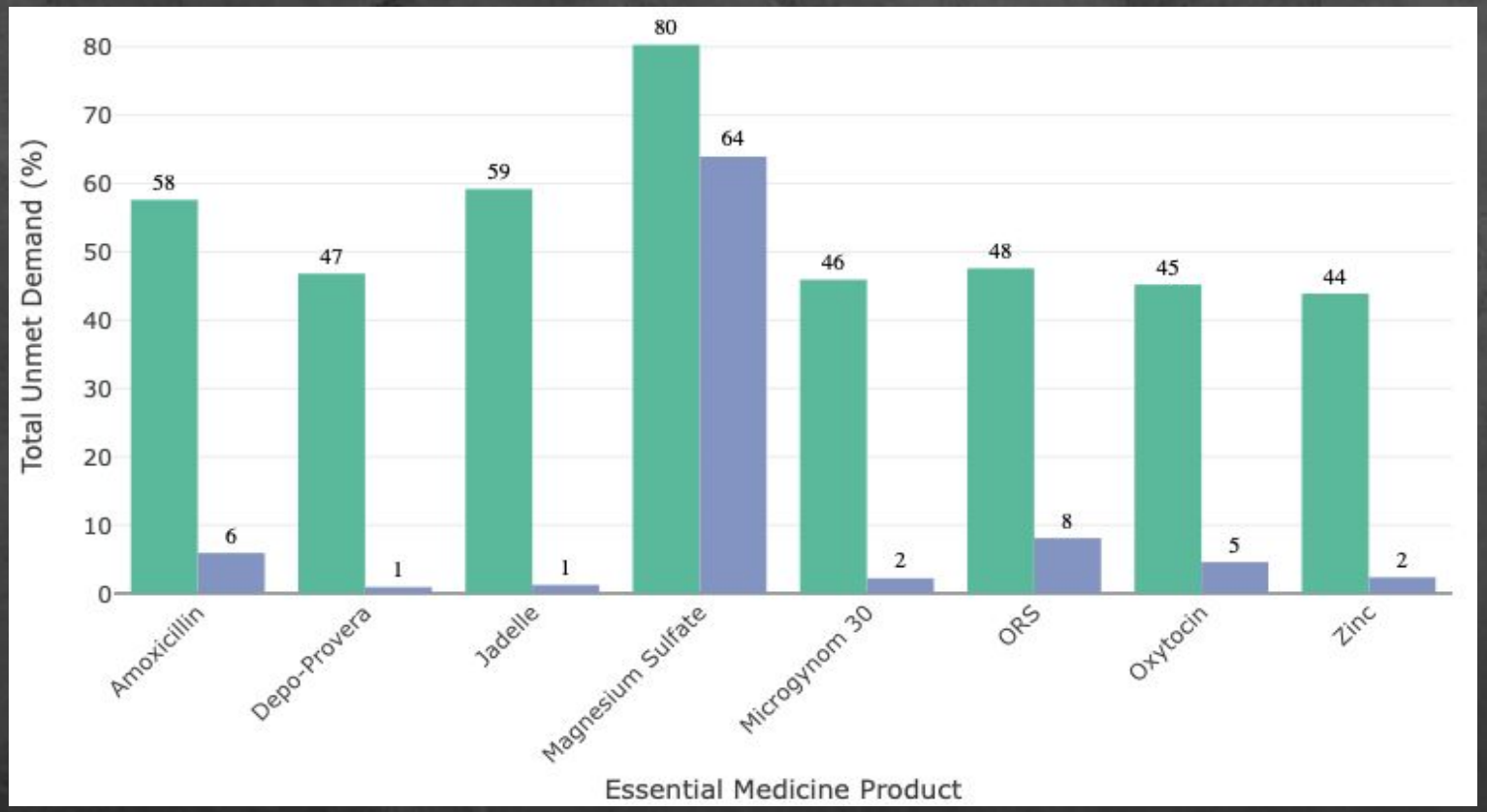
- i.e., we up-weight training examples with unmet demand

Decision-Aware vs. Decision-Blind

** Compare unmet demand of decision-blind vs decision-aware random forest + LP for a fixed budget on a held-out test set month*



End-to-End Relative to Current Approach



Maximum allocation for each product is based on the # of total stock allocated from the Excel tool received for Quarter 1 2022

*% of unmet demand= (unmet demand/actual demand)*100*

Promising end-to-end improvements using AI/OR over current system in Sierra Leone

Reduce **20%-98%** unmet demand for focal essential medicines

NMSA
NATIONAL MEDICAL SUPPLIES AGENCY

MACRO-EYES

Penn

Thank you!

Questions? hamsab@wharton.upenn.edu