Decision-Aware Learning for Global Health Supply Chains

Hamsa Bastani[1,2]

with Osbert Bastani[1], Angel Chung[1], Vahid Rostami[2], Johnna Sundberg[2], Musa Komeh[2], Ashley Schmidt[2] & Lawrence Sandi[3]

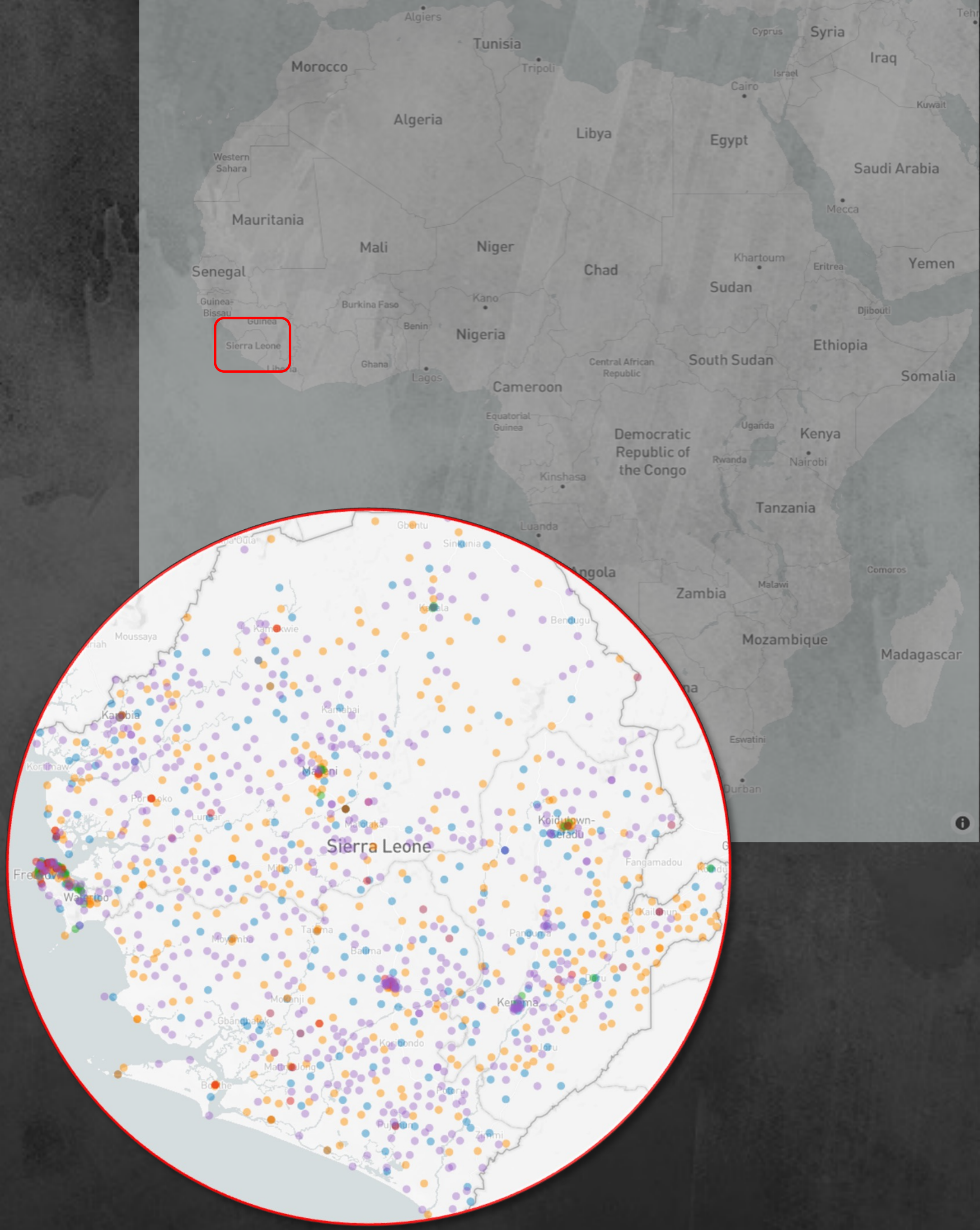
[1]Penn, [2]Macro-Eyes, [3]NMSA

Global Health Allocations

Distribute limited inventory of 100+ essential medicines across 1000+ health facilities in Sierra Leone

Highly uncertain demand, limited budgets → 42% of needs *not* fulfilled

Goal: use AI + OR to do better



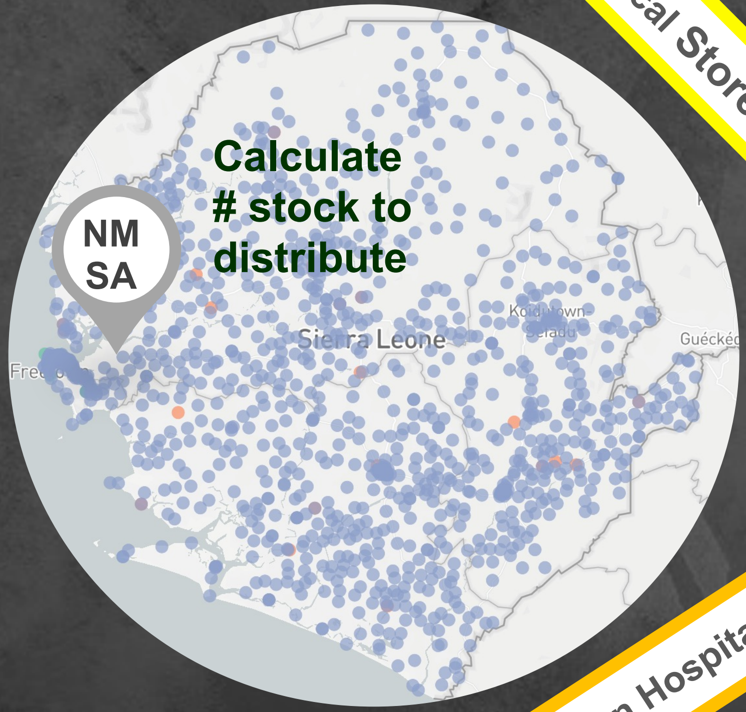
Current Approach

- Health facilities “request” 3-month rolling average of demand
- Complex Excel allocation tool (32 tabs)
  - Apply rationing parameters (differs based on product, facility type)
  - Population/poverty modifiers
  - Prioritize hospitals over other types of health facilities

Code	Level of Care	Essential	m Type	Item	On request forms	Include in Distribution ?	Total Request (Units)	DMS Request Total (Units)	District Hospital Request Total (Units)	WA Hospital Request Total (Units)
10000093	ALL	X	FHC	Albendazole 400mg, Tab	Yes	Yes	1,690,967	1,384,400	243,600	62,967
10000100	Hospital & CHC		FHC	Amoxicillin & Clavulanic Acid (Co-Amoxiclav) 500mg & 125mg, Tab	Yes	Yes	1,871,735	1,430,620	166,300	274,815
10000095	ALL	X	FHC	Amoxicillin 250mg, Dispersible, Tab	Yes	Yes	7,733,420	6,077,450	1,148,436	507,534
10000450	Hospital & CHC	X	FHC	Ampicillin 500mg, Pdr for IM/IV, Inj, Vial	Yes	Yes	1,472,982	839,070	403,755	230,157
10000703	ALL		FHC	Apron, Plastic, Disposable, Pcs	Yes	Yes	711,933	241,000	123,633	347,300
10000014	Hospital & CHC		FHC	Atropine Sulphate 1mg/ml, Inj, 1ml, Amp	Yes	No	78,678	27,110	40,851	10,717
10000684	Hospital Only		FHC	Bags, Blood, Collecting, 450ml, Pcs	Yes	Yes	68,300	-	58,850	9,450
10000520	ALL		FHC	Cannula, IV, 18G, Short, Sterile, Disposable, Pcs	Yes	yes	532,114	371,300	121,450	39,364

Code	DMS Request Total	DMS - Bo	DMS - Bombali	DMS - Bonthe	DMS - Kailahun	DMS - Kambia	DMS - Kenema	DMS - Koinadugu	DMS - Kono	DMS - Moyamba	DMS - Port Loko	DMS - Pujehun	DMS - Tonkolili	DMS - Western Area	DMS - Falaba	DMS - Karene	District Hospital Request To	Hosp - Bo	Hosp - Bombali Makeni
10000093	1,384,400	na	83,300	-	67,100	119,000	181,500	65,500	230,000	na	141,600	88,900	56,100	278,800	62,000	10,600	243,600	4,100	9,800
10000100	1,430,620	na	69,360	-	25,500	77,500	168,000	61,000	87,000	na	197,100	17,760	178,500	399,000	62,000	87,900	166,300	na	na
10000095	6,077,450	na	408,000	298,000	234,000	225,000	882,000	360,000	600,000	na	503,000	707,200	287,500	1,069,000	327,000	176,750	1,148,436	45,000	50,000
10000450	839,070	na	27,000	6,020	45,000	20,700	540,800	14,950	33,400	na	57,600	31,000	14,800	23,500	15,500	8,800	403,755	30,000	27,000
10000703	241,000	na	70,400	-	29,100	1,000	56,400	5,800	-	na	24,300	19,200	11,900	-	12,300	10,600	123,633	5,200	6,800
10000014	27,110	na	1,350	-	4,500	750	-	2,950	5,500	na	800	7,680	-	1,730	650	1,200	40,851	na	na
10000684	-	na	na	na	na	na	na	na	na	na	na	na	na	na	na	na	58,850	na	na
10000520	371,300	na	3,990	17,800	25,600	6,910	7,650	5,400	14,000	na	12,150	192,000	17,100	60,500	4,200	4,000	121,450	na	na

Step 1: Stock ->
3 national facility types



Medical Stores

District Hospital

Western Hospital

Step 2: 1st allocation
(National -> District)

15
Districts

18
District
locations

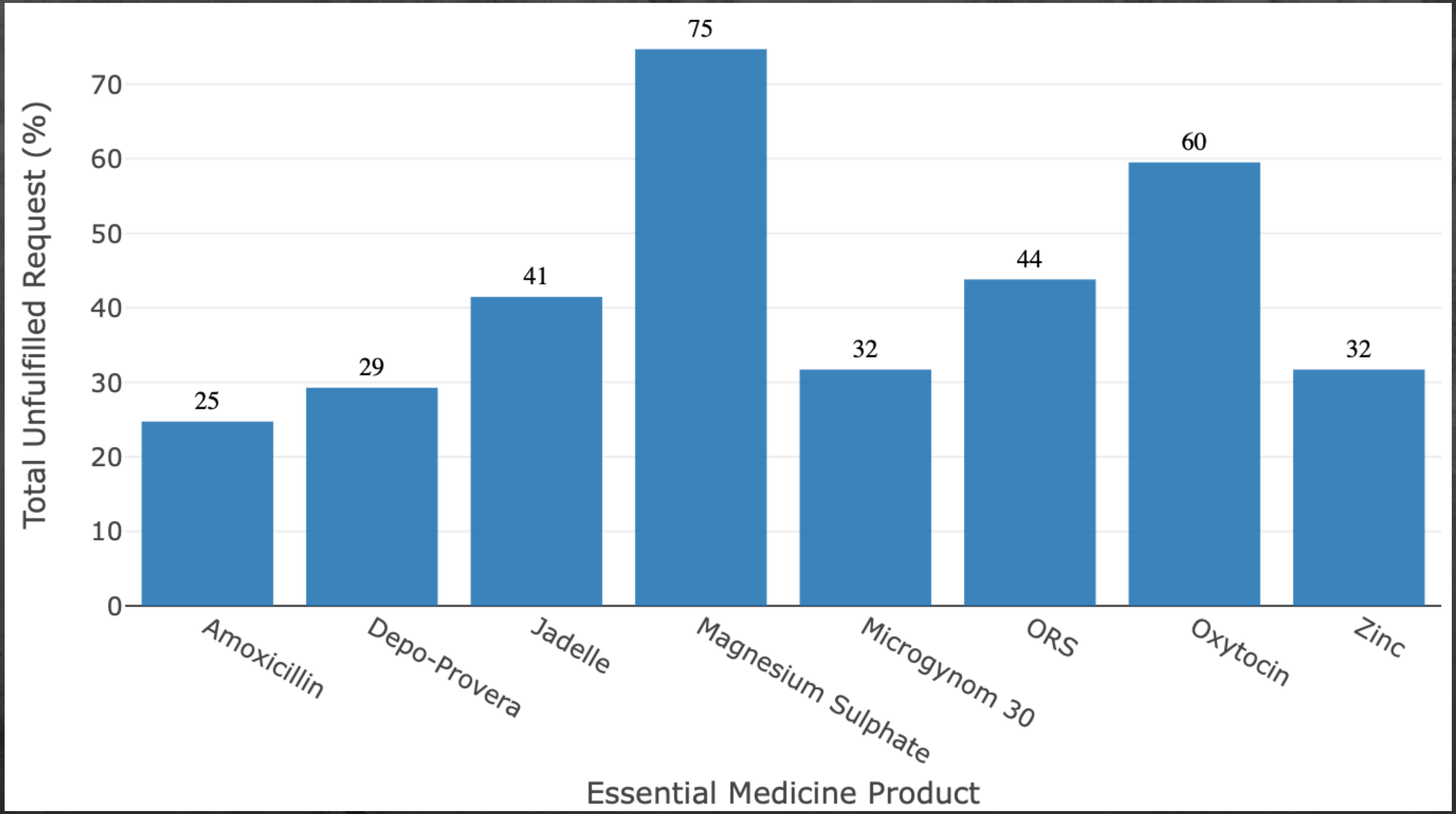
11
Districts
locations

Step 3: 2nd allocation

- Based on outstanding request & available stock
- District will distribute to local facilities within the location by themselves
  - Medical stores: around 1,200 local facilities
  - District hospital: 28 hospitals
  - Western hospital: 18 hospitals

- - - -> Distributed parameters are based on: # request, available stock, level of care, government rationing parameter

How does it work?



Avg unfulfilled demand across the country is **42%**!

Predict-then-Optimize?

- Step 1: Train demand prediction model
- Step 2: Optimize allocations based on predictions
- Active area of research: **end-to-end learning** with constrained optimization [Kotary, Fioretto, Van Hentenryck, Wilder (2021)]

Learning and Optimization

- **Goal:** Given response y (e.g., today's demand), compute decision z (e.g., inventory to allocate) 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)
  - Can be complex

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 $\tilde{\ell}$?

Interested in…

- **Generality:** able to interface with complex data science + optimization pipelines
- **Computational Tractability**
- **Principled:** approximates optimal decision loss

Prior Work - I

- [Bertsimas & Kallus (2018)] Cluster “nearby” observations (using CART or LOESS) to estimate conditional distribution $y|x$ in SAA
  - Not optimal: predictive model is the same regardless of opt problem
  - Not computationally tractable for complex predictive models

- [Kallus & Mao (2022)] Specialized approach for random forests + SAA
  - Not computationally tractable: every tree split requires re-solving optimization problem
  - Not general: strategy specific to tree models and one unknown parameter vector per optimization

Prior Work - II

- [Elmachtoub & Grigas (2021)] Approximate decision-aware prediction loss $\tilde{\ell}$ when ℓ is a LP and prediction $f_\theta(x) = \theta^\top x$ is linear
  - Not general: strategy specific to linear models and known constraints

- [Wilder et al (2019a/b)] Backpropagate through decision loss
  - Not general: prediction function $f_\theta(x)$ must be differentiable
  - Not computationally tractable: need to solve optimization problem at *every* gradient step

Our Strategy

- 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

Our Strategy

- 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)

Our Strategy

- Step 1: train arbitrary decision-blind model $f_\theta(x)$
- Step 2: compute gradients through arbitrary optimization problem to obtain training data weights $\{w_i\}$
- Step 3: re-train model $f'_\theta(x)$ with weighted training data $\{x_i, w_i, y_i\}$
- Step 4: run optimization problem with plug-in estimates

Our Strategy

- **Generality:** only requires re-weighting observations in *any* data science pipeline
  - Can directly use any off-the-shelf ML package, no re-implementation
- **Computational Tractability:** only requires training predictive model and solving optimization problem 2x
- **Principled:** directly approximates optimal decision loss

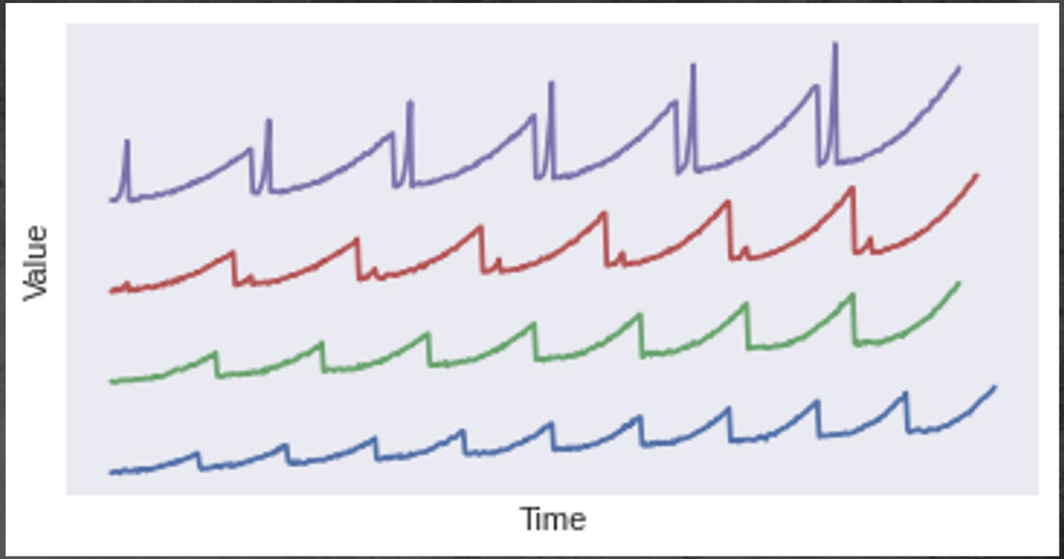
Allocating Essential Meds

Products

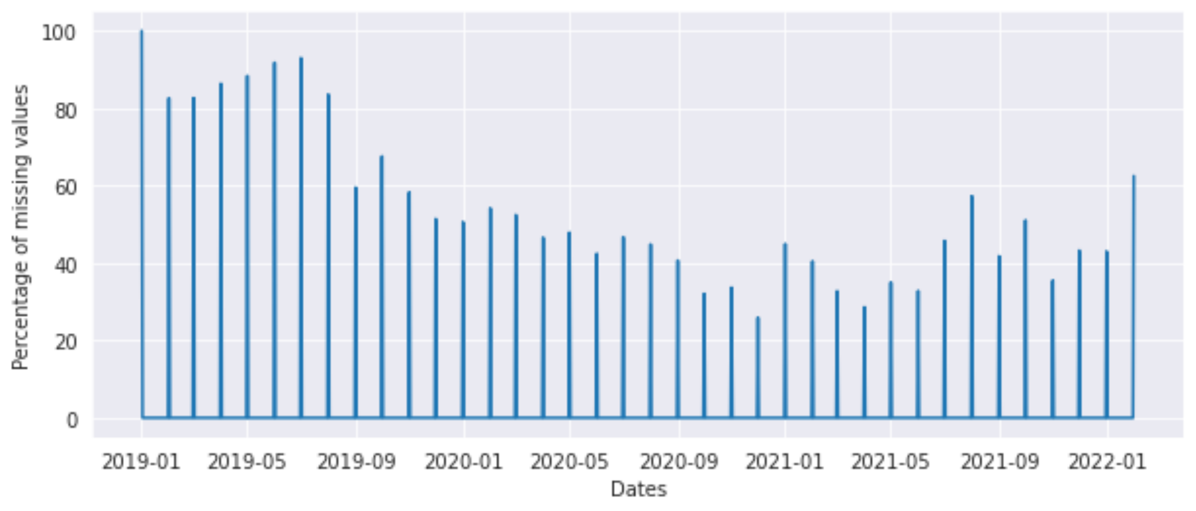
- Child Health <5 years of age
  - Amoxicillin 250mg, Dispersible, Tab
  - Oral Rehydration Salts (ORS), Sachet (correlation to zinc)
  - Zinc Sulphate 20mg, Tab (correlation to ORS)
- Maternal Health
  - Oxytocin 10IU, Inj, Amp
  - Magnesium Sulphate 50%, Inj, 10ml, Amp
- Family Planning (adolescent 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

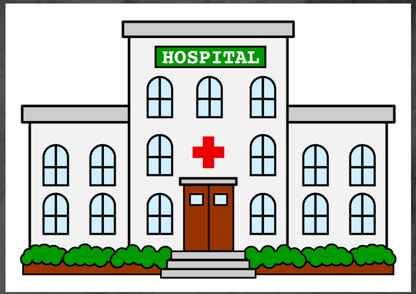
- dhis2 forms from Jan 2019 to July 2021 (31 months)
- Significant # of missing or unreliable values



- 9,000+ separate time series, each with only ~18 observations on average
- Standard time series forecasting does very poorly



Meta-Learning

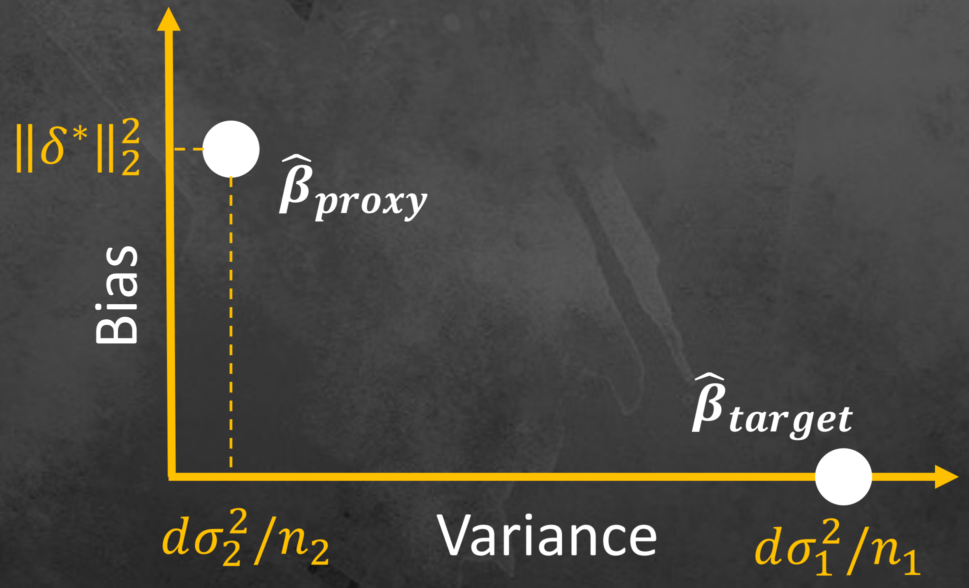


$D_{1,1}(1, \ldots, t)$	$D_{1,2}(1, \ldots, t)$	$\cdots$	$D_{1,N}(1, \ldots, t)$
$\vdots$	$\ddots$		$\vdots$
$D_{K,1}(1, \ldots, t)$	$D_{K,2}(1, \ldots, t)$	$\cdots$	$D_{K,N}(1, \ldots, t)$

Proxy data from other facilities & products!

Bias-Variance Tradeoff

- Train on proxy data: biased predictions, but low variance
- Train on target data: unbiased predictions, but high variance

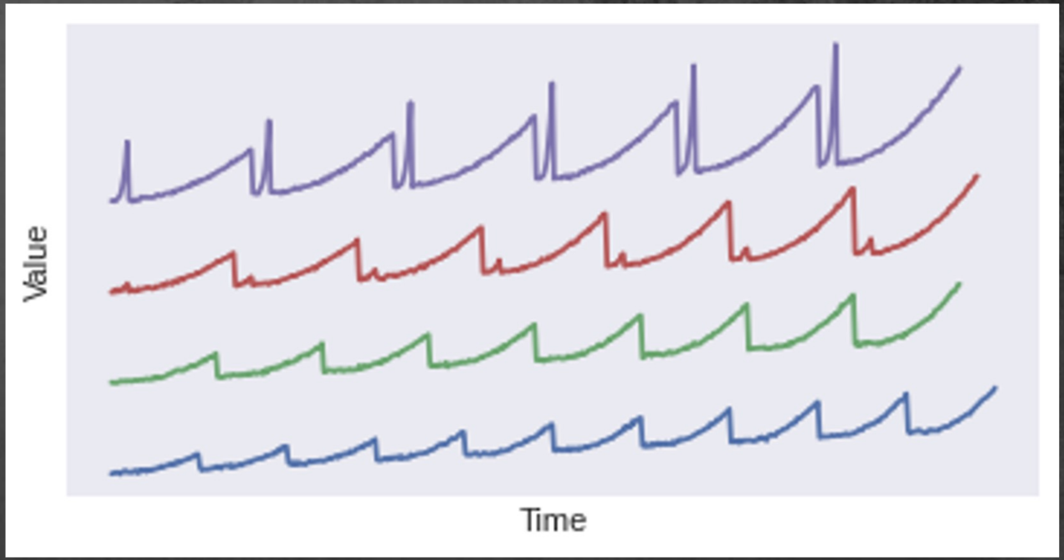


Meta-learning: combine both data sources – systematically accounting for uncertainty in proxy – to improve predictions

Bastani, Predicting with Proxies: Transfer Learning in High Dimension, Mgmt Sci (2020)

Meta-Learning

- Leverage cross-product, cross-facility correlations
- Data from other facilities / products act as "proxy" data to reduce variance at some cost of bias [Bastani (2021), Bastani, Simchi-Levi & Zhu (2022), Xu & Bastani (2022)]



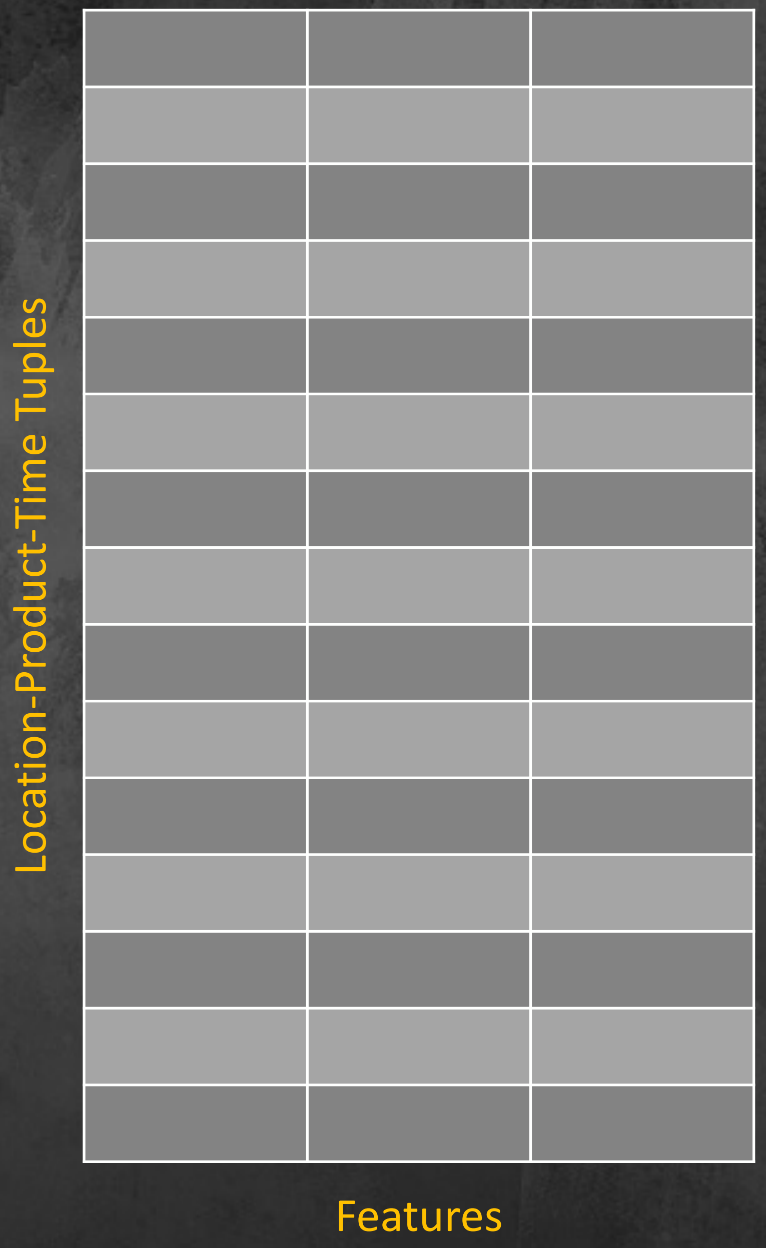
- Random forest "meta-model" forecasts jointly across all 8 products and ~1200 facilities

Prediction Setup

Hand-Engineered Features:

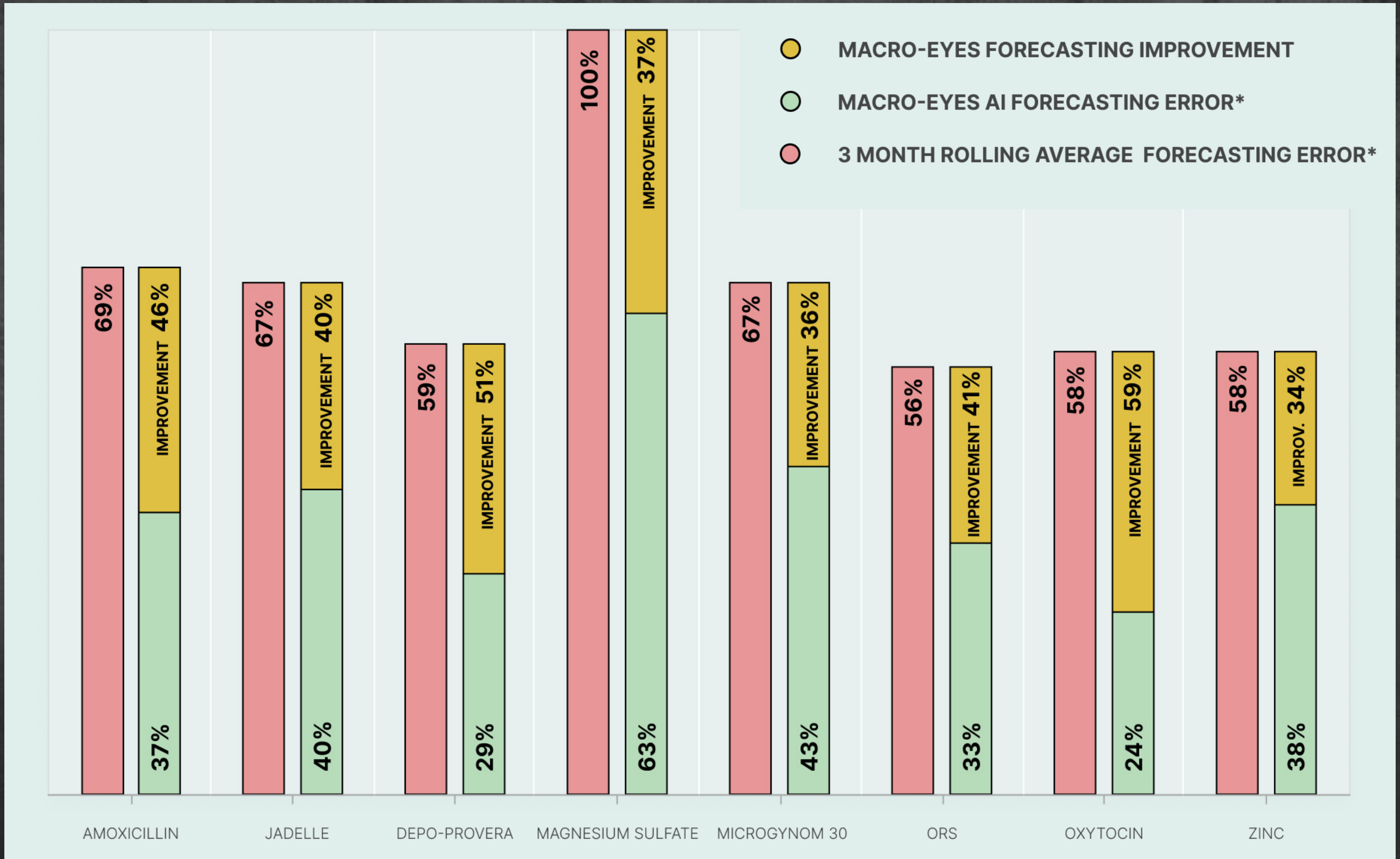
- lagged demands for each product in that facility for last 10 months
- month, year fixed effects
- rolling average of last 2/3/4/5/6/8/10 months + variance of last 3/6 months
- facility region, type

Outcome: demand for product-facility at time t



Forecasting error on March 2021 relative to 3-month rolling average

Out-of-Sample Results



Decision-blind random forest improves demand forecasts by **34-59%** on held out test set month

Bare Bones Stochastic Optimization

- **Goal:** allocations a_n^* across all N districts that minimizes cost
- **Objective:** 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

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

* Efficient linear program with sample average approximation

Gradient of the LP Solution

- Perturb demand to avoid degeneracy: $\xi_n + \eta_n$, where $\eta_n \sim N(0, \sigma^2)$

$$\nabla_{\xi_m} \left\{ \arg\min_a \sum_{n=1}^{N} \mathbb{E}_{\eta_n}[\max\{\xi_n + \eta_n - s_n - a_n, 0\}] \ \ \text{subj. to} \ \sum_{n=1}^{N} a_n = b \right\}$$

- Lagrangian:

$$L(a, \lambda) = \sum_{n=1}^{N} \mathbb{E}_{\eta_n}[\max\{\xi_n + \eta_n - s_n - a_n, 0\}] + \lambda \left(b - \sum_{n=1}^{N} a_n \right)$$

Gradient of the LP Solution

- The first-order condition is

$$0 = -\nabla_{a_n} L\big(a^*(\xi), \lambda^*(\xi)\big) = \underbrace{\mathbb{P}_{\eta_n}[s_n + a_n^*(\xi) \le \xi_n + \eta_n]}_{\text{CDF of } \eta_n} + \underbrace{\lambda^*(\xi)}_{\text{constant}}$$

- Taking the gradient with respect to ξ_m yields an equation involving $\nabla_{\xi_m} a_n^*(\xi)$, solve to obtain

$$\nabla_{\xi_m} a_n^*(\xi) = \delta_{m,n} + O\left(\frac{1}{N}\right)$$

Gradient of the LP Objective

$$\nabla_{\xi_m}\left\{\min_a \sum_{n=1}^{N} \mathbb{E}_{\eta_n}[\max\{\xi_n + \eta_n - s_n - a_n, 0\}] \ \text{ subj. to } \sum_{n=1}^{N} a_n = b\right\}$$

- Equivalently, $\nabla_{\xi_m} L\big(a^*(\xi), \lambda^*(\xi)\big)$, yielding

$$\nabla_{\xi_m} L\big(a^*(\xi), \lambda^*(\xi)\big) = \mathbb{P}_{\eta_m}[s_m + a_m \leq \xi_m + \eta_m] \approx \mathbb{I}[s_m + a_m \leq \xi_m]$$

Now that we have the gradients…

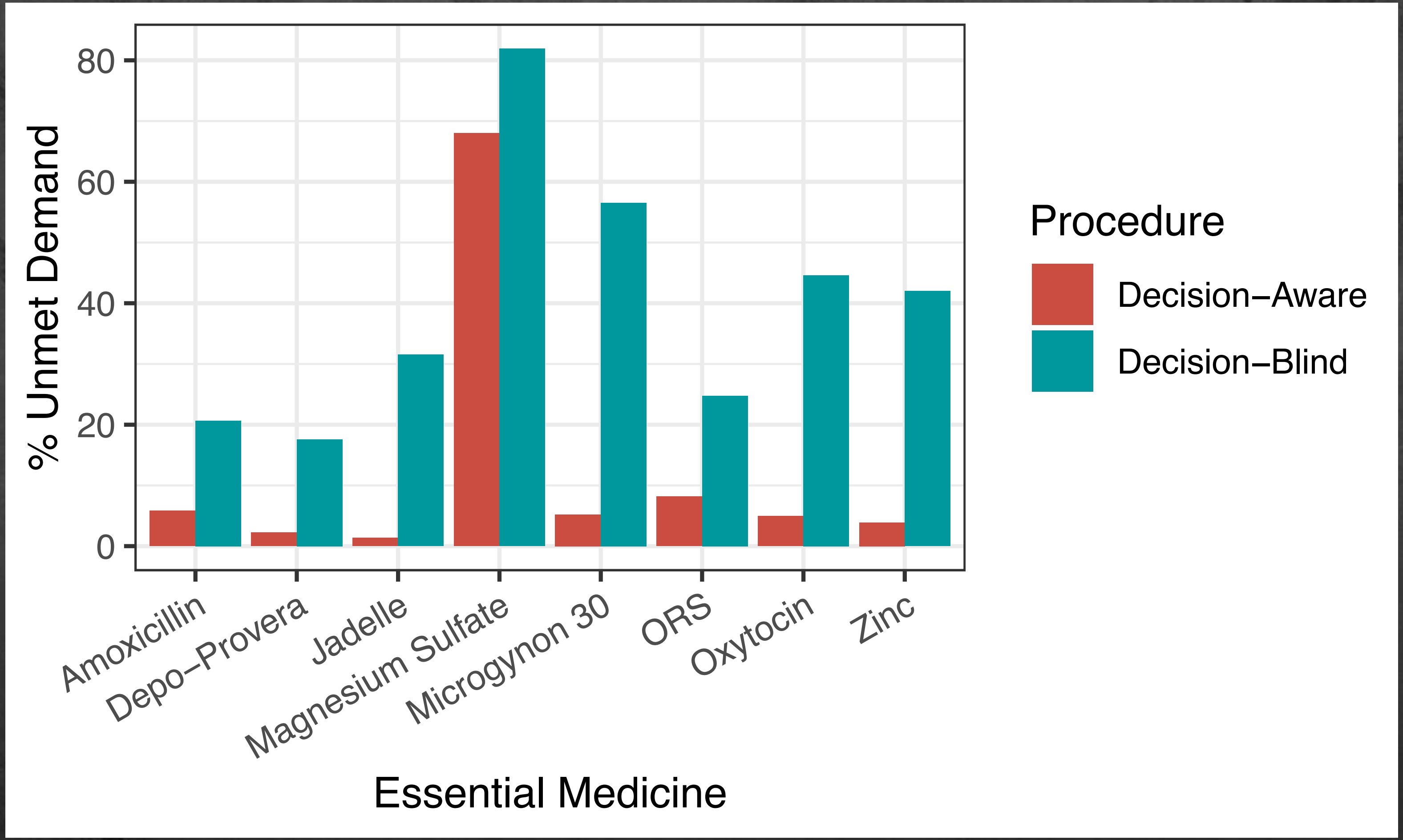
- Use gradients to obtain approximate predictive model objective:

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

- i.e., we up-weight training examples with unmet demand
- Classic ML “spends” capacity on predicting at facilities with low stockout likelihood; we focus on facilities that are relevant to the OPT objective

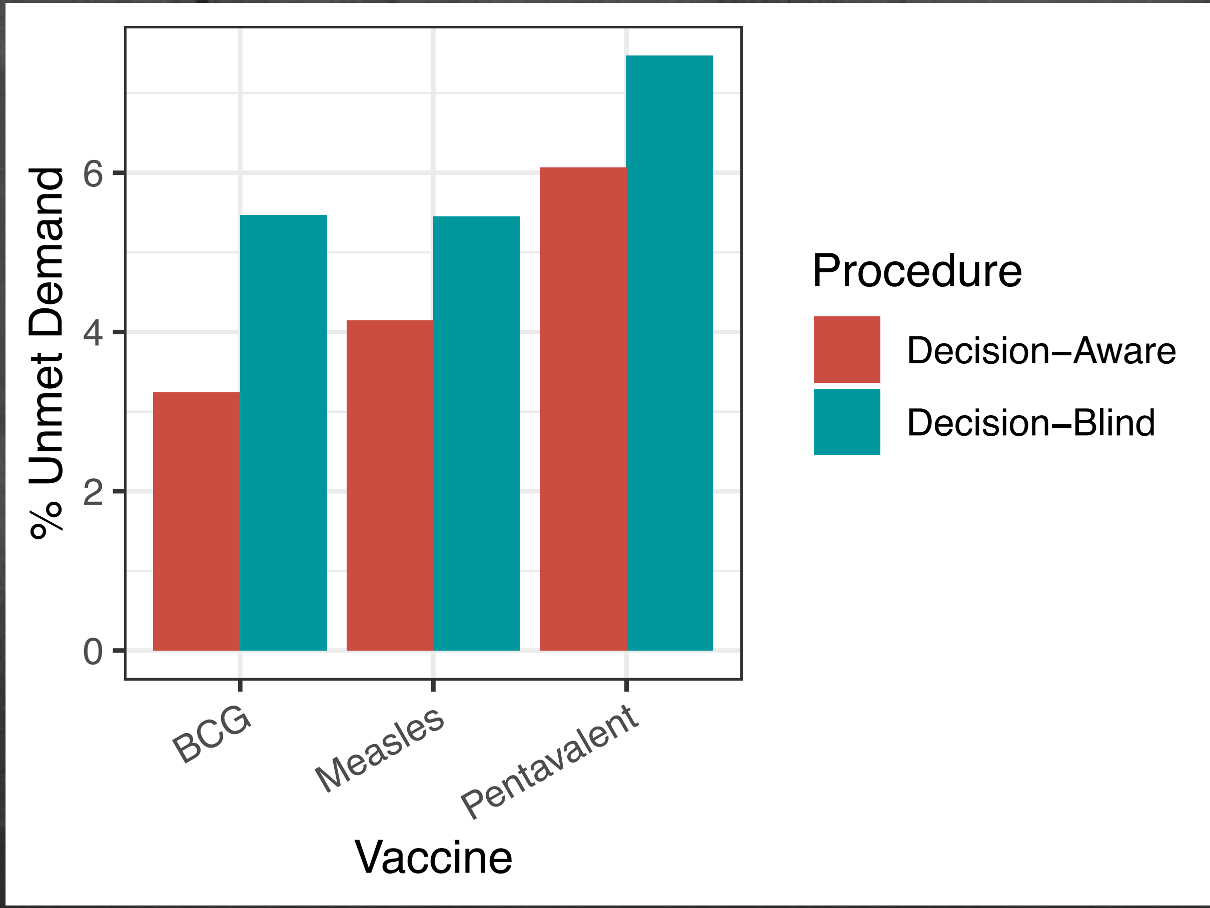
Out-of-Sample Results

** Compare unmet demand for a fixed budget on a held-out test set month*



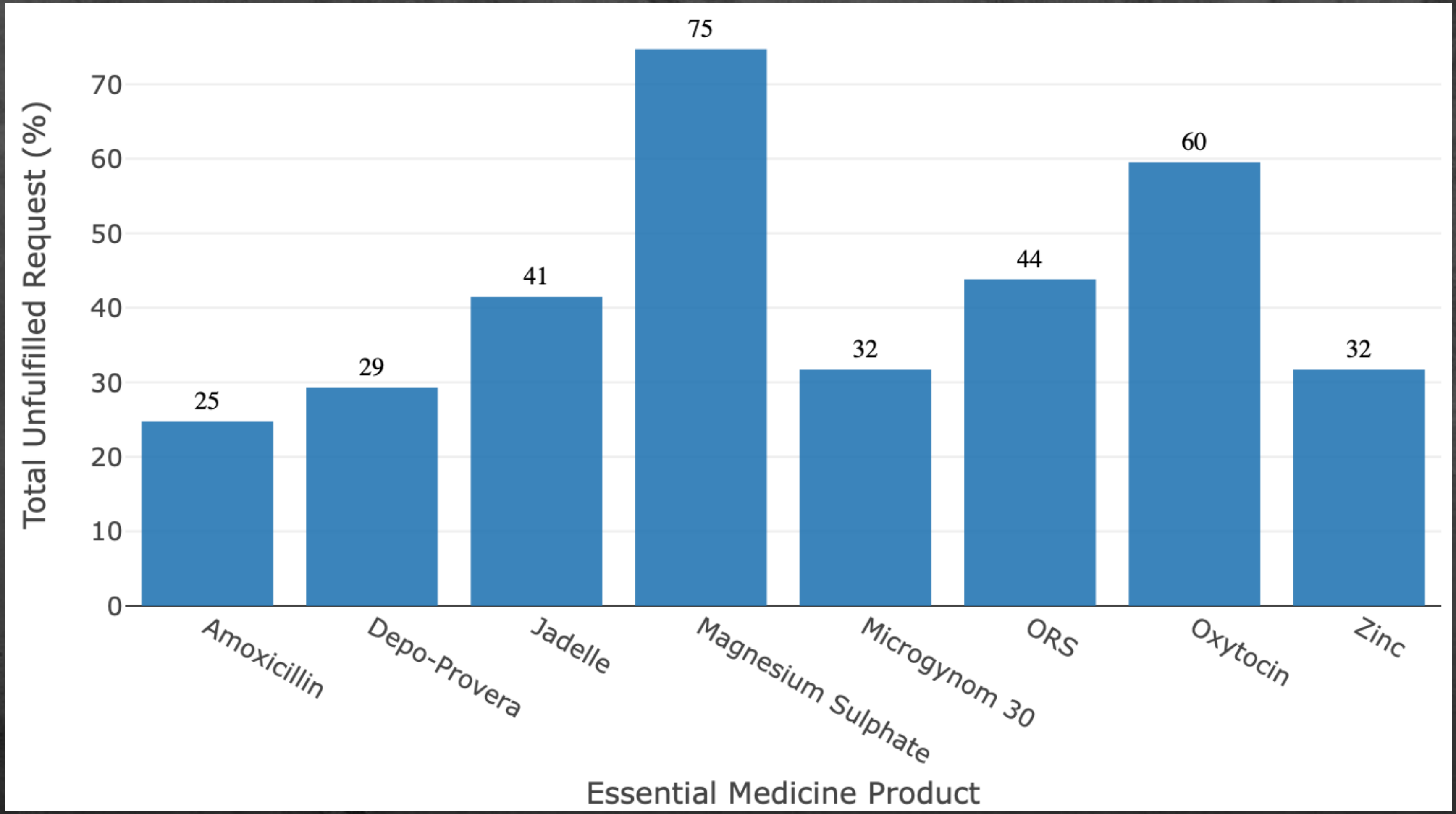
Out-of-Sample Results

** Compare unmet demand for a fixed budget on a held-out test set month*



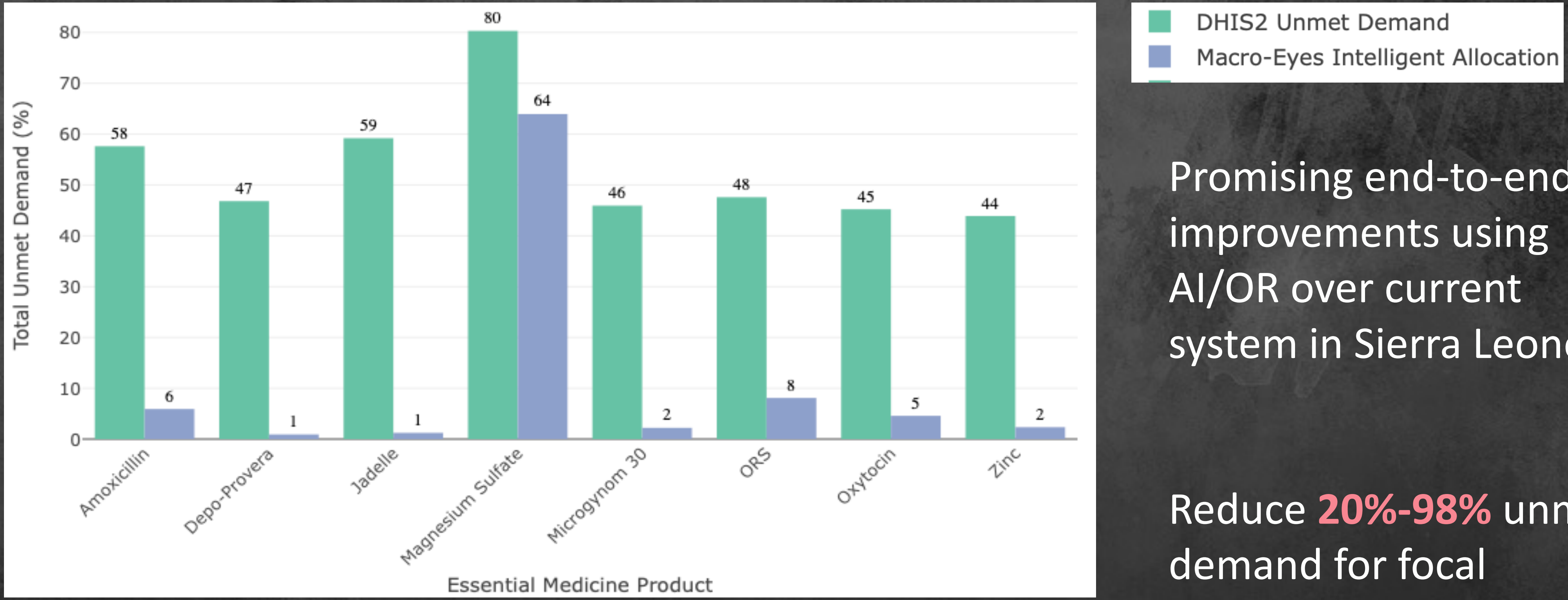
End-to-End Results

Recall…



Avg unfulfilled demand across the country is **42%**!

End-to-End Relative to Current Approach



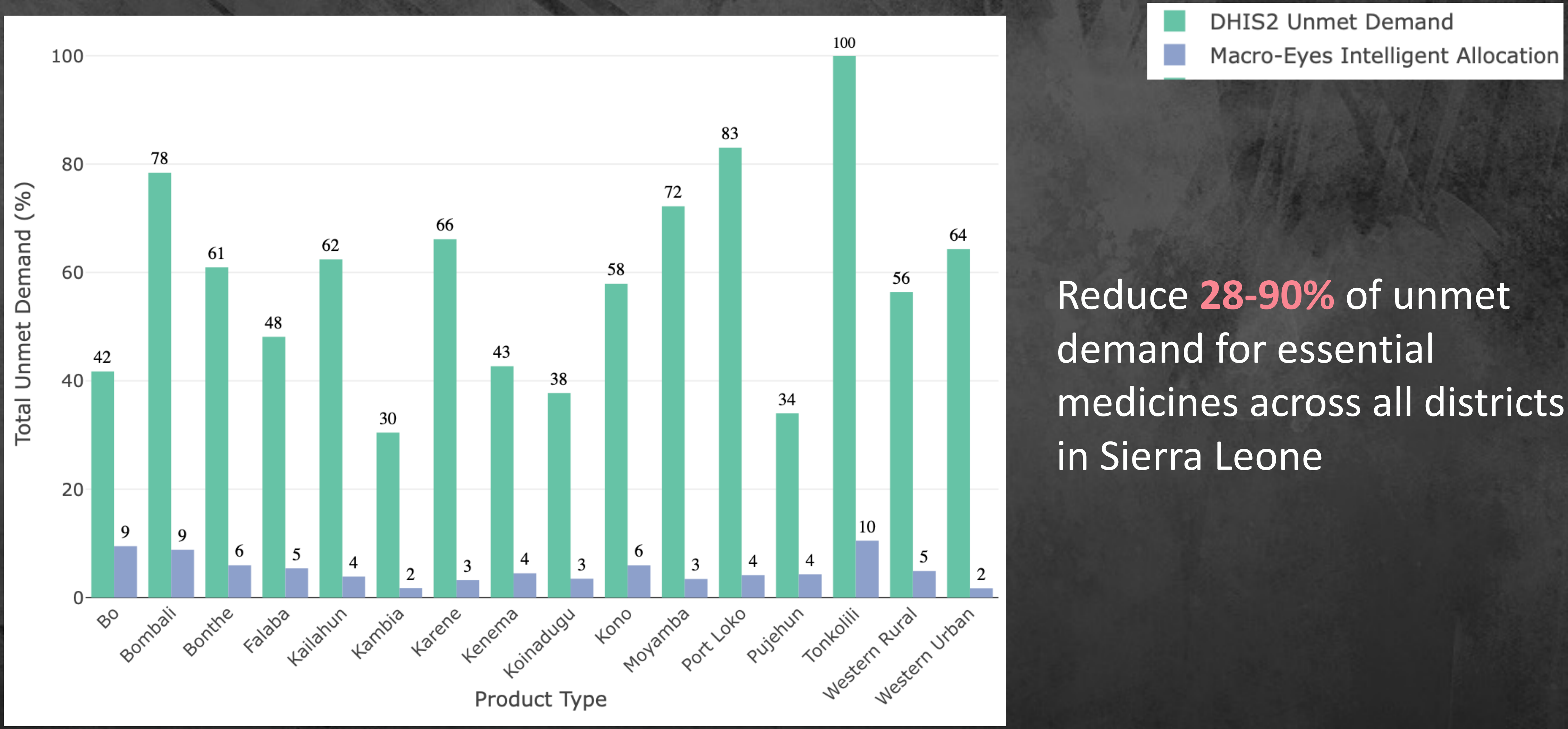
Maximum allocation for each product is based on the # of total stock allocated from the Excel tool received for Quarter 1 2022

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

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

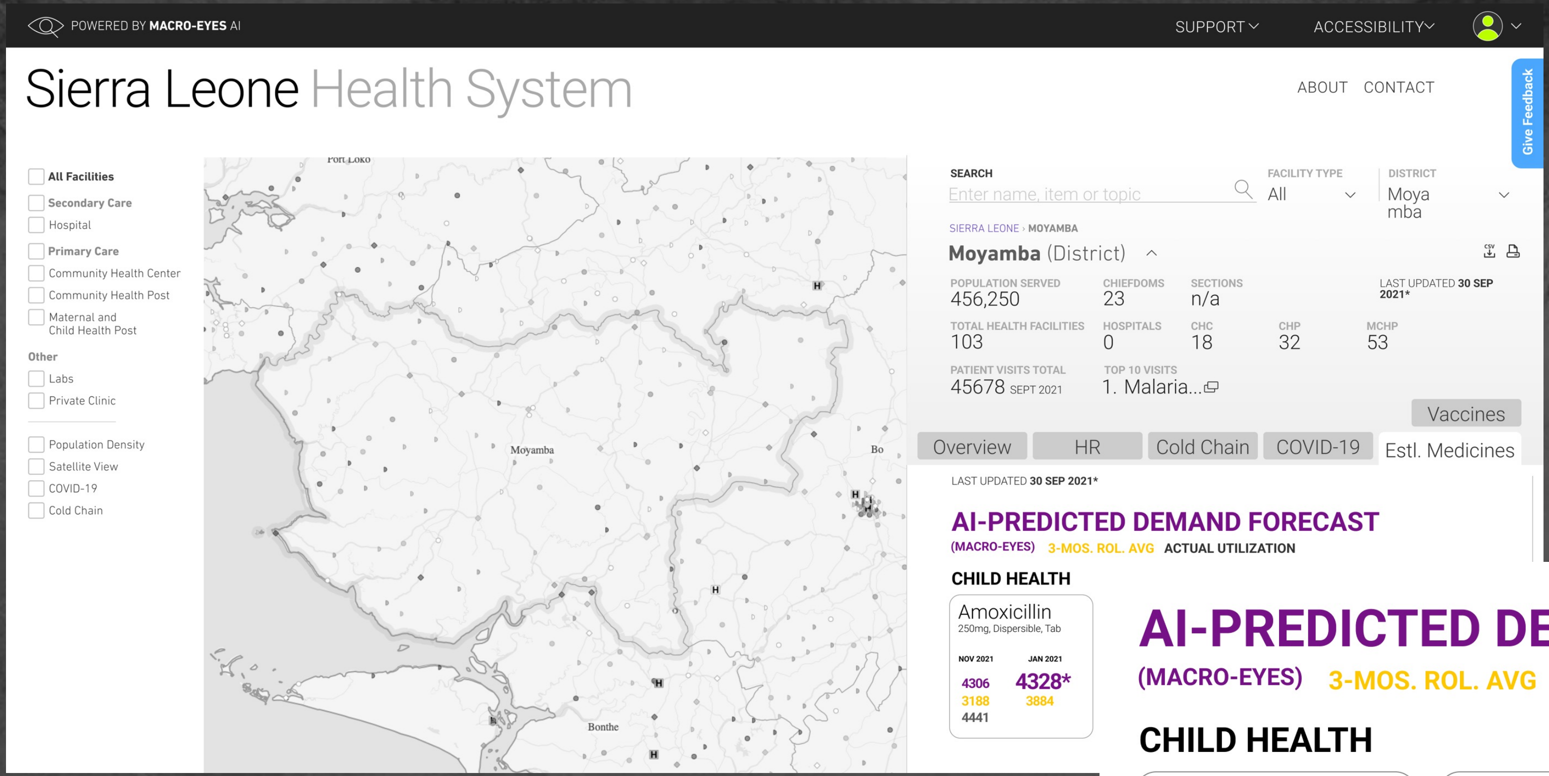
Reduce **20%-98%** unmet demand for focal essential medicines

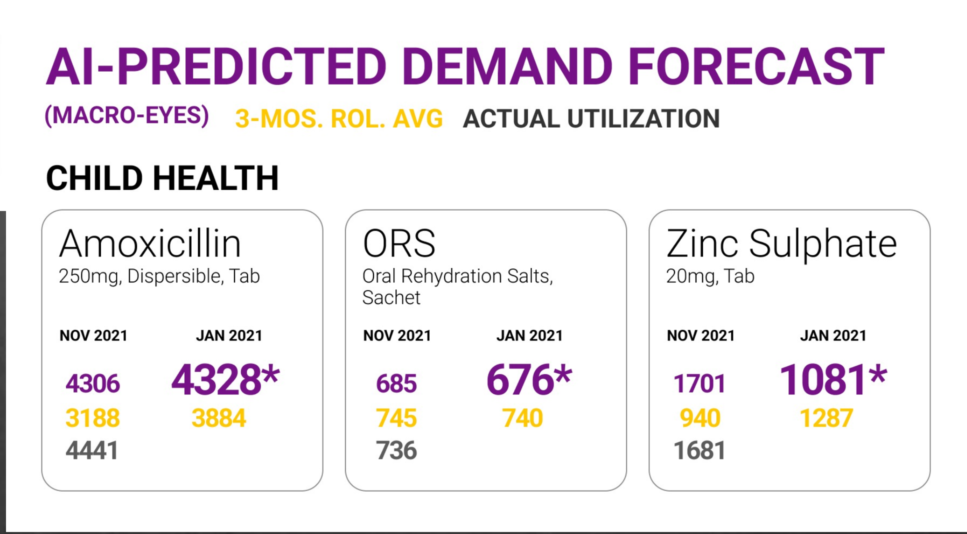
Improvement by District

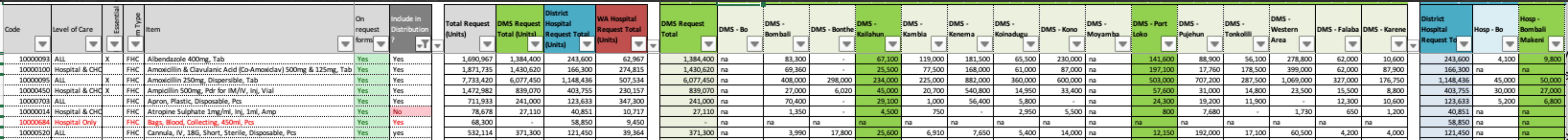


Reduce **28-90%** of unmet demand for essential medicines across all districts in Sierra Leone

Goal to deploy in other countries like Mozambique, Cote d'Ivoire, Rwanda, Democratic Republic of Congo that use the same dhis2 forms







Code	Level of Care	Essential	m Type	Item	On request forms	Include in Distribution ?	Total Request (Units)	DMS Request Total (Units)	District Hospital Request Total (Units)	WA Hospital Request Total (Units)
10000093	ALL	X	FHC	Albendazole 400mg, Tab	Yes	Yes	1,690,967	1,384,400	243,600	62,967
10000100	Hospital & CHC		FHC	Amoxicillin & Clavulanic Acid (Co-Amoxiclav) 500mg & 125mg, Tab	Yes	Yes	1,871,735	1,430,620	166,300	274,815
10000095	ALL	X	FHC	Amoxicillin 250mg, Dispersible, Tab	Yes	Yes	7,733,420	6,077,450	1,148,436	507,534
10000450	Hospital & CHC	X	FHC	Ampicillin 500mg, Pdr for IM/IV, Inj, Vial	Yes	Yes	1,472,982	839,070	403,755	230,157
10000703	ALL		FHC	Apron, Plastic, Disposable, Pcs	Yes	Yes	711,933	241,000	123,633	347,300
10000014	Hospital & CHC		FHC	Atropine Sulphate 1mg/ml, Inj, 1ml, Amp	Yes	No	78,678	27,110	40,851	10,717
10000684	Hospital Only		FHC	Bags, Blood, Collecting, 450ml, Pcs	Yes	Yes	68,300	-	58,850	9,450
10000520	ALL		FHC	Cannula, IV, 18G, Short, Sterile, Disposable, Pcs	Yes	yes	532,114	371,300	121,450	39,364

Code	DMS Request Total	DMS - Bo	DMS - Bombali	DMS - Bonthe	DMS - Kailahun	DMS - Kambia	DMS - Kenema	DMS - Koinadugu	DMS - Kono	DMS - Moyamba
10000093	1,384,400	na	83,300	-	67,100	119,000	181,500	65,500	230,000	na
10000100	1,430,620	na	69,360	-	25,500	77,500	168,000	61,000	87,000	na
10000095	6,077,450	na	408,000	298,000	234,000	225,000	882,000	360,000	600,000	na
10000450	839,070	na	27,000	6,020	45,000	20,700	540,800	14,950	33,400	na
10000703	241,000	na	70,400	-	29,100	1,000	56,400	5,800	-	na
10000014	27,110	na	1,350	-	4,500	750	-	2,950	5,500	na
10000684	-	na	na	na	na	na	na	na	na	na
10000520	371,300	na	3,990	17,800	25,600	6,910	7,650	5,400	14,000	na

Code	DMS - Port Loko	DMS - Pujehun	DMS - Tonkolili	DMS - Western Area	DMS - Falaba	DMS - Karene	District Hospital Request To	Hosp - Bo	Hosp - Bombali Makeni
10000093	141,600	88,900	56,100	278,800	62,000	10,600	243,600	4,100	9,800
10000100	197,100	17,760	178,500	399,000	62,000	87,900	166,300	na	na
10000095	503,000	707,200	287,500	1,069,000	327,000	176,750	1,148,436	45,000	50,000
10000450	57,600	31,000	14,800	23,500	15,500	8,800	403,755	30,000	27,000
10000703	24,300	19,200	11,900	-	12,300	10,600	123,633	5,200	6,800
10000014	800	7,680	-	1,730	650	1,200	40,851	na	na
10000684	na	na	na	na	na	na	58,850	na	na
10000520	12,150	192,000	17,100	60,500	4,200	4,000	121,450	na	na

Many exciting questions…

- How to target collecting missing information that is most likely to improve decision-aware learning?
- Forward-looking approach: decision-aware reinforcement learning?
- Demand forecasts used for other downstream decisions → decision-aware learning with only a prior on decisions?
- Human-AI interface: help non-technical decision-makers incorporate sudden changes like flooded roads, missing trucks, etc

NMSA
NATIONAL MEDICAL SUPPLIES AGENCY

MACRO-EYES

Penn

Thank you!

Questions? hamsab@wharton.upenn.edu