Optimizing Health Coverage in Ethiopia: A Learning-augmented Approach and Persistent Proportionality Under an Online Budget

Davin Choo[1*], Yohai Trabelsi[1*], Fentabil Getnet[2], Samson Warkaye Lamma[2], Wondesen Nigatu[3], Kasahun Sime[3], Lisa Matay[4], Milind Tambe[1†], Stéphane Verguet[4†]

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University
[2]National Data Management and Analytics Center for Health, Ethiopian Public Health Institute
[3]Primary Healthcare and Community Engagement Lead Executive Office, Ministry of Health, Ethiopia
[4]Department of Global Health and Population, Harvard T.H. Chan School of Public Health

Abstract

As part of nationwide efforts aligned with the United Nations' Sustainable Development Goal 3 on Universal Health Coverage, Ethiopia's Ministry of Health is strengthening health posts to expand access to essential healthcare services. However, only a fraction of this health system strengthening effort can be implemented each year due to limited budgets and other competing priorities, thus the need for an optimization framework to guide prioritization across the regions of Ethiopia. In this paper, we develop a tool, Health Access Resource Planner (HARP), based on a principled decision-support optimization framework for sequential facility planning that aims to maximize population coverage under budget uncertainty while satisfying region-specific proportionality targets at *every* time step. We then propose two algorithms: (i) a learning-augmented approach that improves upon expert recommendations at any single-step; and (ii) a greedy algorithm for multi-step planning, both with strong worst-case approximation estimation. In collaboration with the Ethiopian Public Health Institute and Ministry of Health, we demonstrated the empirical efficacy of our method on three regions across various planning scenarios.

Code — https://github.com/yohayt/OHCE/
Extended version — https://arxiv.org/pdf/2509.00135

1 Introduction

Ethiopia, the second most populous country in Africa, is home to over 130 million people across 12 regional states and two chartered cities, further divided into zones, woredas (districts), and kebeles (the smallest administrative units) (UNFPA 2022; Worldometers 2022). Roughly 76% of the population lives in rural areas, where access to healthcare is limited compared to urban settings (Hendrix et al. 2023).

The Ethiopian healthcare system follows a three-tier structure comprising Primary Health Care Units (PHCUs), general hospitals, and specialized hospitals. PHCUs, responsible for 70-80% of essential health services, have been expanded under the Ministry of Health's (MOH) flagship Health Extension Program (HEP), launched in 2003/04 (Wang et al. 2016). Unfortunately, persistent challenges remain as the evolution of the HEP is not aligned with the ever increasing demand for comprehensive essential healthcare services (health promotion-rehabilitative care) at the health post. This resulted in inadequate staffing and equipment, and declining and/or stagnating performance in maternal and child health indicators (Teklu et al. 2020).

To address these gaps, the MOH introduced the HEP Optimization Roadmap (2020–2035) (Ministry of Health [Ethiopia] 2020). A central reform is the reclassification of health posts into three categories: mixed, basic, and comprehensive. Comprehensive health posts are designed to deliver a broader suite of services, including childbirth, postnatal care, and chronic disease treatment, and require substantial investments in infrastructure, staffing, and medical supplies.

The roadmap calls for the construction (or upgrading) of over 2,000 comprehensive health posts, with an estimated cost of USD 300k–400k per facility excluding staffing and equipment. As of now, 260 are operational and 183 are under construction. While initial decisions used WHO's AccessMod platform (https://www.accessmod.org/), they did not fully incorporate region-specific prioritization preferences (see below) or account for geographic and demographic factors known to drive service disparities (Getnet et al. 2025).

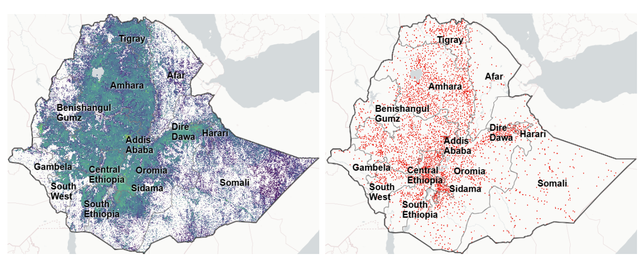

Figure 1: Map of Ethiopia overlaid with 2026 population estimates (log scale). Left: Brighter yellow areas indicate higher population density. Right: Locations of primary hospitals, health centers, and comprehensive health posts that can provide essential health services are highlighted in red.

Given limited resources and growing health demands,

[*]These authors contributed equally.
[†]These authors jointly supervised this work.

there is a pressing need for a strategic, data-driven approach to guide equitable facility placement (e.g., where to build the next comprehensive health post given limited annual budget to augment the existing network of healthcare facilities). This requires integrating diverse datasets, including geolocation of existing facilities, high-resolution population estimates, topographic and official district administrative shapefiles, and health service coverage indicators like skilled birth attendance and postnatal care. Planning is further complicated by uncertainty in budgets and population forecasts, and the need to ensure fairness at every decision point. Different regions may also impose varying distributional preferences: some may prioritize population density while others may target poorer districts.

Our contributions. In collaboration with the Ethiopian Public Health Institute and Ministry of Health, we formalize the problem and develop a Health Access Resource Planner (HARP) tool to help regional planners navigate these challenges and customize allocation strategies to local priorities.

1. In Section 2, we propose and study the new abstract problem of maximizing a non-decreasing submodular function under online partition and global proportional constraints (MOPGP), which captures the real-world challenge of prioritizing health facility upgrades in Ethiopia. Crucially, our framework accommodates region-specific distributional goals by encoding proportionality constraints, making it a flexible, shared decision-support tool for planners with possibly differing regional priorities.
2. We design algorithms to solve MOPGP with provable worst-case guarantees. These include a learning-augmented algorithm that integrates expert-provided selections, supporting regional planning workflows that start with partial plans and seek principled refinement.
3. In Section 5, and in the extended version, we evaluate the HARP tool on some regions of Ethiopia using two possible prioritization rules. We show that our approach successfully prioritizes coverage according to the specified criteria while maintaining reasonable overall coverage.

HARP enables regional planners to solve their specific instance of MOPGP using a unified, principled approach. While our empirical results offer insight into practical performance, we emphasize that all findings are intended as a proof of concept and not official policy recommendations.

2 Problem Formulation

Notation. We use lowercase letters for scalars, boldface for sets and sequences, and calligraphic letters for higher-order objects (e.g., sets of sets, graphs, matroids). For a set $\mathbf{A}$, let $|\mathbf{A}|$ denote its cardinality, and $2^{\mathbf{A}}$ its power set. The disjoint union of disjoint sets $\mathbf{A}$ and $\mathbf{B}$ is denoted $\mathbf{A} \uplus \mathbf{B}$. We use $\mathbb{R}_{\geq 0}$ for non-negative reals, $\mathbb{N}$ for natural numbers, $\mathbb{N}_{>0} = \mathbb{N} \setminus \{0\}$, and $[n] = \{1, \ldots, n\}$ for any $n \in \mathbb{N}_{>0}$. For any sequence $\mathbf{s} \in [r]^b$ and type $q \in [r]$, let $\#(q, \mathbf{s})$ denote the number of times q appear in $\mathbf{s}$, e.g., for $\mathbf{s} = (1, 2, 1, 3) \in [3]^4$, we have $\#(1, \mathbf{s}) = 2$, $\#(2, \mathbf{s}) = 1$, and $\#(3, \mathbf{s}) = 1$.

In the following, we will formally define MOPGP in full generality and then contextualize to how it assists in facility location within any particular region of Ethiopia.

We model MOPGP as a sequential subset selection task, where the goal is to maximize a non-decreasing submodular function (see Section 3 for a formal definition) subject to two constraints: (i) an *online budget constraint* modeling incremental budget arrivals, and (ii) a *global proportional constraint* enforcing type-level proportionality. Let $\mathbf{V} = \mathbf{V}^{(1)} \uplus \ldots \uplus \mathbf{V}^{(h)}$ denote the ground set partitioned over h rounds. Each element $e \in \mathbf{V}$ is associated with a type $\mathrm{type}(e) \in [r]$, inducing an second partitioning $\mathbf{V} = \mathbf{T}_1 \uplus \ldots \uplus \mathbf{T}_r$, where $\mathbf{T}_q = \{e \in \mathbf{V} : \mathrm{type}(e) = q\}$. At each round $t \in [h]$ (a year), a budget $b^{(t)} \leq |\mathbf{V}^{(t)}|$ is revealed and we must irrevocably select a subset $\mathbf{S}^{(t)} \subseteq \mathbf{V}^{(t)}$ of size $|\mathbf{S}^{(t)}| \leq b^{(t)}$. We define the cumulative budget and selection up to round t as $b^{(1:t)} = \sum_{\tau=1}^{t} b^{(\tau)}$ and $\mathbf{S}^{(1:t)} = \uplus_{\tau=1}^{t} \mathbf{S}^{(\tau)}$ respectively. The objective is to maximize $f : 2^{\mathbf{V}} \to \mathbb{R}$, a non-decreasing submodular function over the final selection $\mathbf{S} = \mathbf{S}^{(1:h)} \subseteq \mathbf{V}$, subject to the follow two constraints:

1. **Online budget constraint**: $|\mathbf{S}^{(t)}| \leq b^{(t)}$ for all $t \in [h]$.
2. **Global proportional constraint**: For given proportions $p_1, \ldots, p_r \in [0, 1]$ with $\sum_{q=1}^{r} p_q \leq 1$, the selection should satisfy $|\mathbf{S} \cap \mathbf{T}_q| \geq p_q \cdot |\mathbf{S}|$ for all types $q \in [r]$.

Unfortunately, it may not be feasible to always meet proportional targets. For instance, if $b^{(1)} = 1$ and we require each of $r \geq 2$ types to have at least $\frac{1}{10r}$ proportion, then no single-element selection in the first time step can meet this condition. As such, we define the notion of *satisfaction ratio* as $\alpha_q(\mathbf{S}) = \frac{|\mathbf{S} \cap \mathbf{T}_q|}{p_q \cdot |\mathbf{S}|}$ where $\alpha_{\min}(\mathbf{S}) = \min_{q \in [r]} \alpha_q(\mathbf{S})$ and define the best possible such ratio under budget b as $\beta(b) = \max_{\mathbf{A} \subseteq \mathbf{V}, |\mathbf{A}| = b} \alpha_{\min}(\mathbf{A})$. Note that the proportional constraints on all types *can* be satisfied when $\beta(b) \geq 1$.

To ensure long-term proportional balance, we insist that any feasible solution is a type feasible selection.

Definition 1 (Type feasible selection)**.** A subset $\mathbf{S}^{(1:h)} \subseteq \mathbf{V}$ of size $b^{(1:h)}$ is *type feasible* if, for all time steps $t \in [h]$:

1. $\alpha_{\min}(\mathbf{S}^{(1:t)}) = \beta(b^{(1:t)})$
2. Among all such sets, it minimizes $|\mathbf{Q}(\mathbf{S}^{(1:t)})|$ where

$$\mathbf{Q}(\mathbf{S}^{(1:t)}) = \Big\{ q \in [r] : \alpha_q(\mathbf{S}^{(1:t)}) = \beta(|\mathbf{S}^{(1:t)}|) \text{ and } \mathbf{V}^{(t)} \cap \mathbf{T}_q \neq \emptyset \Big\}$$

Intuitively, the set $\mathbf{Q}(\mathbf{S}^{(1:t)})$ contains all types with the minimum satisfaction ratio. Since improving the satisfaction ratio $\alpha_{\min}(\cdot)$ requires one to select at least one at least element of each type in $\mathbf{Q}(\cdot)$, minimizing $|\mathbf{Q}(\cdot)|$ serves to hedge against potential budget uncertainty.

We are now ready to formally define MOPGP.

Definition 2 (MOPGP)**.** Given a non-decreasing submodular function $f : 2^{\mathbf{V}} \to \mathbb{R}$, time horizon $h \in \mathbb{N}$, number of types $r \in \mathbb{N}_{>0}$, ground set $\mathbf{V} = \mathbf{V}^{(1)} \uplus \ldots \uplus \mathbf{V}^{(h)} = \mathbf{T}_1 \uplus \ldots \uplus \mathbf{T}_r$, online budgets $\mathbf{b} = \{b^{(1)}, \ldots, b^{(h)}\}$, and type proportions $\mathbf{p} = \{p_1, \ldots, p_r\}$, the MOPGP$(f, h, r, \mathbf{V}, \mathbf{b}, \mathbf{p})$

problem seeks to compute a type feasible selection $\mathbf{S} \subseteq \mathbf{V}$ such that $|\mathbf{S}^{(t)}| \leq b^{(t)}$ for all time steps $t \in [h]$.

Observe that MOPGP is a generalization of the well-studied NP-hard maximum coverage problem for a suitably defined set function f when $t = 1$, $r = 1$, and $p_1 = 0$ (i.e., no proportionality constraints). As such, one cannot expect to be able to compute an optimum selection $\mathbf{S}^* \subseteq \mathbf{V}$ in polynomial time in general. Instead, we will design and analyze efficient approximation algorithms that can compute some selection $\mathbf{S} \subseteq \mathbf{V}$ such that $f(\mathbf{S}) \geq \alpha \cdot f(\mathbf{S}^*)$ for some approximation ratio $0 \leq \alpha \leq 1$.

Contextualizing MOPGP **to the Ethiopian setting.** For our problem, we use 1km-by-1km grid population forecasts for each region in Ethiopia (WorldPop 2025) and geographical information to compute travel time (Weiss et al. 2020).

Let $\mathbf{U}$ denote the space of grid cells considered. For each year $t \in [h]$, the set $\mathbf{V}^{(t)}$ correspond to all grid cells in the region for which we can build a comprehensive health post while types correspond to the districts within the different regions in Ethiopia. Note that $\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(h)}$ are all defined over the same set of grid cells $\mathbf{U}$ but correspond to building a facility at different time steps. For any facility located at grid cell $c \in \mathbf{U}$, let $\text{COVERED}(c) \subseteq \mathbf{U}$ be the set of grid cells that are reachable from c within 2-hours of travel time, which depends on the topological characteristics of the terrain and may vary across locations and directions. For a set of facility locations (i.e., subset of grid cells) $\mathbf{S}$, we write $\text{COVERED}(\mathbf{S}) = \bigcup_{c \in \mathbf{S}} \text{COVERED}(c) \subseteq \mathbf{U}$ to mean the union of the covered cells.

As the planning process occurs over a multi-year horizon where facilities continue to provide coverage in all future time steps once it has been built, we define the objective function f is the number of people who can access a comprehensive health post within a walking time of at most two hours under the proposed selection. To be precise,

$$f(\mathbf{S}) = \sum_{t=1}^{h} \sum_{c \in \text{COVERED}(\mathbf{S}^{(1:t)})} w_c^{(t)} \tag{1}$$

where $w_c^{(t)}$ is the population forecast for year t for grid cell c. We show in the extended version that the objective in Eq (1) is submodular. Table 1 summarizes the mapping while Example 3 illustrates our formulation visually.

Example 3. Consider the 5×5 population grid in fig. 2 over $h = 2$ time steps, with $b^{(1)} = 2$ and $b^{(2)} = 1$. Suppose a facility at any cell provides coverage for any cell that is of Manhattan distance 1 from it. If we build two health posts at coordinate $\mathbf{S}^{(1)} = \{(2, 2), (4, 3)\}$ at the first time step and another at $\mathbf{S}^{(2)} = \{(2, 4)\}$ at the second time step, then we see that $f(\mathbf{S}) = f(\mathbf{S}^{(1)} \uplus \mathbf{S}^{(2)}) = 110$, where we obtain a coverage of 46 from Year 1 and 64 in Year 2. Observe that the population can change over time and health posts continue to provide coverage in future time steps once built.

3 Key Concepts and Related Work

After relating our formulation of MOPGP to the known concepts of submodular optimization and matroid theory, we review related work on optimization under these structures, highlighting limitations of existing methods in addressing the sequential and proportional constraints of MOPGP. Finally, we briefly introduce learning-augmented algorithms. For more related work, see the extended version.

Parameters for a region	Component in MOPGP
5-year planning horizon	$h = 5$ time steps
Districts within a region	r types
Annual building budget	Online budgets $b^{(t)}$
Location, at year t	The set $\mathbf{V}^{(t)}$
Posts built for year t	$\mathbf{S}^{(t)} \subseteq \mathbf{V}^{(t)}$ with $\lvert\mathbf{S}^{(t)}\rvert \leq b^{(t)}$
2-hour accessibility goal	Defines $c \in \text{COVERED}(\mathbf{S}^{(1:t)})$
Forecast for cell c in year t	Population weight $w_c^{(t)}$
Coverage objective	f defined in Eq (1)
District prioritization	Proportions $p_1, \ldots, p_r$
Persistent proportionality	Type feasibility in Definition 1

Table 1: Mapping the Ethiopian context to MOPGP.

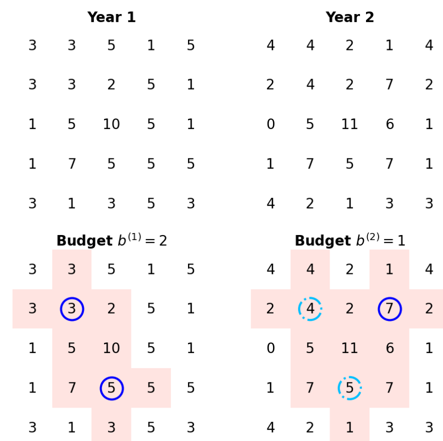

Figure 2: An example illustrating MOPGP notation defined in Section 2. Health post locations are indicated by blue circles. Covered cells are colored. Observe that the two facilities built in Year 1 continue to provide coverage in Year 2.

Submodular functions. Let $\mathbf{V}$ be a finite ground set (e.g., the pairs (location, year) for building health posts) and $f : 2^{\mathbf{V}} \to \mathbb{R}$ be a set function, e.g., a function that assigns a coverage value to any given set of locations where health posts are built. We say that f is non-decreasing if $f(\mathbf{A}) \leq f(\mathbf{B})$ for all $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{V}$, and f is submodular if for all $\mathbf{A} \subseteq \mathbf{B} \subseteq \mathbf{V}$ and $e \in \mathbf{V} \setminus \mathbf{B}$, $f(\mathbf{A} \cup \{e\}) - f(\mathbf{A}) \geq f(\mathbf{B} \cup \{e\}) - f(\mathbf{B})$. It is known that a greedy selection gives a $1 - 1/e$ approximation for maximizing non-decreasing submodular functions under a cardinality constraint (Nemhauser, Wolsey, and Fisher 1978). For notational convenience, we will write the marginal gain function g with respect to f as $g(\mathbf{A}, \mathbf{B}) = f(\mathbf{A} \uplus \mathbf{B}) - f(\mathbf{A})$ for any two disjoint subsets $\mathbf{A}, \mathbf{B} \subseteq \mathbf{V}$.

Matroids. Matroids encompass a wide range of constraints that enable efficient optimization algorithms that admit provable approximation guarantees. In this work, we are particularly interested in the special class of partition matroid (Recski 1973) which has applications like load balancing, scheduling, and constrained subset selection.

Definition 4 (Partition matroid)**.** Let $\mathbf{V} = \mathbf{V}_1 \uplus \ldots \uplus \mathbf{V}_r$ be a partition of $\mathbf{V}$ and $x_1, \ldots, x_r \in \mathbb{N}$ be upper bounds. Then, the independence set $\mathcal{I}$ of a partition matroid $\mathcal{M} = (\mathbf{V}, \mathcal{I})$ is defined as $\mathcal{I} = \{\mathbf{S} \subseteq \mathbf{V} : |\mathbf{S} \cap \mathbf{V}_q| \leq x_q, \forall q \in [r]\}$.

In our Ethiopian setting, budget and proportionality constraints each induce a partition matroid over the set of candidate facilities, e.g., budget at time step t imposes the constraint that any valid selection $\mathbf{S}$ satisfies $|\mathbf{S}^{(t)} \cap \mathbf{V}^{(t)}| \leq b^{(t)}$.

Online maximization under matroid constraints. A closely related problem to MOPGP is that of Problem (1.6) of (Fisher, Nemhauser, and Wolsey 1978), which we present using our notation in Definition 5.

Definition 5 (Maximizing a non-decreasing submodular function subject to k matroid intersections over h rounds)**.** Let $\mathbf{V} = \mathbf{V}^{(1)} \uplus \ldots \uplus \mathbf{V}^{(h)}$ be a set of h disjoint ground sets, $f : 2^{\mathbf{V}} \to \mathbb{R}$ be a non-decreasing submodular set function, and $\mathcal{M}_1 = (\mathbf{V}, \mathcal{I}_1), \ldots, \mathcal{M}_k = (\mathbf{V}, \mathcal{I}_k)$ be k matroids over $\mathbf{V}$. The goal is to find a subset $\mathbf{S} \subseteq \mathbf{V}$ that maximizes $f(\mathbf{S})$ subject to $\mathbf{S} \cap \mathbf{V}^{(t)} \in \mathcal{I}_1 \cap \ldots \cap \mathcal{I}_k$ for all $t \in [h]$.

If budget uncertainty were absent and only final-round proportional constraints mattered, the problem could be modeled using two partition matroids ($k = 2$), allowing the local greedy algorithm of (Fisher, Nemhauser, and Wolsey 1978) to achieve a $1/(k+1) = 1/3$ approximation. However, MOPGP differs fundamentally from Definition 5 in that matroidal constraints – driven by online budget arrivals – are revealed incrementally over time. This renders direct application of their method infeasible.

For the single-matroid ($k = 1$) case, (Calinescu et al. 2011) achieve a $1 - 1/e$ approximation via continuous greedy methods, while (Goundan and Schulz 2007) show that local greedy retains a $1/(\alpha + 1)$ approximation even when optimizing a proxy function g that approximates the true objective f within a factor of $\alpha \geq 1$.

To address evolving constraints, one may consider online optimization frameworks. The submodular secretary problem (Babaioff, Immorlica, and Kleinberg 2007) assumes fixed matroids and randomly arriving elements, whereas our setting reverses this: the ground set is fixed, but feasibility constraints (budgets) are revealed over time. Recent works of (Cristi et al. 2024; Santiago, Sergeev, and Zenklusen 2025) explored online feasibility arrival but focused on linear objectives and do not support global, type-based fairness constraints.

Finally, fairness-aware submodular optimization (Celis et al. 2019; Tsang et al. 2019) has addressed proportionality constraints, but these methods typically require full input access and use continuous relaxations or local exchanges, making them unsuitable for our setting of irrevocable, sequential decisions under uncertainty.

Representation-constrained coverage and online constraints Representation constraints have been studied in coverage problems. For example, (Asudeh et al. 2023) consider a max-k-coverage problem with group fairness constraints but enforce rigid group-wise coverage parity, which can lead to infeasibility. In contrast, our model allows flexible type-level bounds, maintaining feasibility and realism. Recently, ((Cristi et al. 2024); (Santiago, Sergeev, and Zenklusen 2025)) explored online problems where constraints, rather than elements, arrive over time, similar to our evolving feasibility model. However, they focus on maximizing simple linear objectives and do not directly extend to submodular objectives or type-based distributional constraints.

Learning-augmented algorithms. In practice, regional governments may already have existing facility selection plans (for any single fixed time step with budget b) that are constructed using domain knowledge, heuristics, or political considerations. Such an advice selection often encodes valuable local preferences but may not be optimal under the formal objective. Learning-augmented algorithms provide a formal framework to improve decision quality by incorporating advice or predictions, while maintaining worst-case guarantees when the advice is poor (Lykouris and Vassilvitskii 2021). The goal is to design algorithms that are both *consistent* (optimal when the advice is perfect) and *robust* (competitive with the best advice-free baseline), ideally with performance degrading gracefully as advice quality worsens. See (Mitzenmacher and Vassilvitskii 2022) for a survey. Closest to us are the works of (Liu and Srinivas 2024; Agarwal and Balkanski 2024), which aim to improve update times in dynamic settings. In contrast, we aim to improve the approximation ratio achievable by any polynomial-time algorithm in a static selection setting.

4 Algorithmic Contributions

We present a learning-augmented[1] approach for optimizing facility locations at any single time step in Section 4.1 and analyze a greedy approach for multi-step planning under uncertainty in Section 4.2. While our learning-augmented approach applies at every decision point, we only provide guarantees for a single fixed time step.[2] In Section 4.3, we highlight the importance of Definition 1 by showing that one cannot achieve non-trivial guarantees for MOPGP under online budgets without additional structure on the problem instance. All proofs are deferred to the extended version.

4.1 Learning-augmented single-step planning

Our goal in this section is to design a learning-augmented algorithm that uses such advice to produce selections from $\mathbf{V}$ that are provably no worse and potentially better. Here, we present the algorithm for the simple case of $r = 1$ type (Algorithm 1) and show how to extend it to the more general partition matroid setting with $r \geq 1$ in the extended version.

[1]While "learning-augmented" originated from using machine-learned predictions, the framework more broadly encompasses any source of advice, such as expert-curated selections in our setting.

[2]Single-step guarantees offer interpretable, modular assurances and are especially relevant in dynamic policy settings, where decisions are made incrementally over multiple time steps.

A key advantage of the learning-augmented approach is its practicality in real-world decision-making environments. Rather than discarding or overriding expert input, our method builds on it — treating existing recommendations as a starting point and refining them in a principled way. This makes it particularly attractive to policymakers and planners, as it respects domain expertise accumulated over years of fieldwork, local knowledge, and institutional memory. Importantly, our approach offers formal performance guarantees: it matches the expert plan when it is already optimal, and improves upon it when possible, without ever doing worse than the best advice-free alternative. This balance between respecting human judgment and augmenting it with algorithmic rigor makes the method especially compelling in collaborative, policy-driven settings.

At the high-level, our algorithmic approach is to augment subsets of the advice selection with greedy selections[3] For any subset $\mathbf{A}' \subseteq \mathbf{A}$, we fill the remaining $b - |\mathbf{A}'|$ slots by greedily selecting elements from $\mathbf{V} \setminus \mathbf{A}'$ to maximize marginal gain with respect to the non-decreasing submodular set function $g(\cdot, \mathbf{A}') = f(\cdot \cup \mathbf{A}') - f(\mathbf{A}')$. That is, we greedily choose elements conditioned on already chosen $\mathbf{A}'$.

Algorithm 1: Learning-augmented single-step of MOPGP

Input: Elements $\mathbf{V}$, non-decreasing submodular set function $f : \mathbf{V} \to \mathbb{R}$ to maximize for, budget $b \geq 1$, advice selection $\mathbf{A} \subseteq \mathbf{V}$ of size $|\mathbf{A}| = b$

Output: A selection $\mathbf{U} \subseteq \mathbf{V}$

Define subsets $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_b$ of $\mathbf{A}$ where $|\mathbf{A}_i| = i$ for sizes $i \in \{0, 1, \ldots, b\}$ ▷ $\mathbf{A}_0 = \emptyset$ and $\mathbf{A}_b = \mathbf{A}$

for $i \in \{0, 1, \ldots, b\}$ **do**

  Initialize $\mathbf{B}_i = \emptyset$

  while $|\mathbf{B}_i| < b - i$ **do**

    Add $e^* = \operatorname{argmax}_{e \in \mathbf{V} \setminus (\mathbf{A}_i \cup \mathbf{B}_i)} g(\mathbf{A}_i \cup \mathbf{B}_i, \{e\})$ to $\mathbf{B}_i$

  Define $\mathbf{U}_i = \mathbf{A}_i \uplus \mathbf{B}_i$ ▷ $|\mathbf{U}_i| = |\mathbf{A}_i| + |\mathbf{B}_i| = b$

return $\mathbf{U}_{i^*}$, where $i^* = \operatorname{argmax}_{i \in \{0,1,\ldots,b\}} f(\mathbf{U}_i)$

Theorem 6. *Consider Algorithm 1 and let* OPT $\subseteq \mathbf{V}$ *be an optimal selection. Then,*

$$f(\mathbf{U}) \geq \max_{i \in \{0,1,\ldots,b\}} f(\mathbf{A}_i) + \frac{1 - \frac{1}{e}}{\left\lceil \frac{|\mathrm{OPT} \setminus \mathbf{A}_i|}{b-i} \right\rceil} \cdot g(\mathrm{OPT}, \mathbf{A}_i)$$

Algorithm 1 is essentially an algorithm for size-constrained non-decreasing submodular maximization. Observe that $\mathbf{U}_0$ is the advice-free greedy selection $\mathbf{S}$, and $\mathbf{U}_b$ is the advice selection $\mathbf{A}$, so we always have $f(\mathbf{U}_{i^*}) \geq \max\{f(\mathbf{U}_0), f(\mathbf{A})\}$ and thus the output $\mathbf{U}_{i^*} \subseteq \mathbf{V}$ of our learning-augmented algorithm is at least $(1 - \frac{1}{e})$-robust and 1-consistent. This is because $f(\mathbf{U}_{i^*}) \geq f(\mathbf{U}_0) \geq (1 - \frac{1}{e}) \cdot f(\mathrm{OPT})$ by submodularity of f, and $f(\mathbf{U}_{i^*}) = f(\mathrm{OPT})$ when $f(\mathbf{A}) = f(\mathrm{OPT})$. Our approach of theorem 6 naturally extends to $r \geq 1$ types for type-decomposable objective functions, but with the base guarantee of $f(\mathbf{U}_0) \geq (1 - \frac{1}{e}) \cdot f(\mathrm{OPT})$ degrading to $f(\mathbf{U}_0) \geq \frac{1}{2} \cdot f(\mathrm{OPT})$ due to matroidal constraints; see the extended version for details.

[3] MOPGP generalizes the NP-hard problem of maximum coverage, and greedy selections attain tight approximation guarantees.

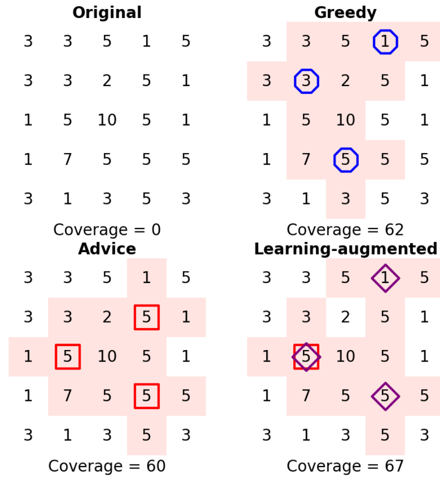

Figure 3: Using the facility at $(2, 1)$ as the partial advice, Algorithm 1 produces a selection with improved coverage. Note: The top-left cell is coordinate $(0, 0)$.

As a remark, we point out that there is nothing inherently special about the choice of $\mathbf{A}_0, \ldots, \mathbf{A}_b$. From the proof of theorem 6 in the extended version, one can see that Algorithm 1 can be modified to work with any polynomial number of subsets of $\mathbf{A}$ while still being polynomial time.

Example 7. fig. 3 shows three possible selections of $b = 3$ facilities on the first grid of Example 3. As before, suppose a facility at any cell provides coverage for any cell that is of Manhattan distance 1 from it. Greedily selecting based on marginal gain produces the selection $\mathbf{G} = [(3, 2), (0, 3), (1, 1)]$ with $f(\mathbf{G}) = 62$. Meanwhile, an advice selection of $\mathbf{A} = [(2, 1), (1, 3), (3, 3)]$ has $f(\mathbf{A}) = 60$. Running Algorithm 1 with the advice facility at coordinate $(2, 1)$ produces the selection $\mathbf{U} = [(2, 1), (3, 3), (0, 3)]$ with $f(\mathbf{U}) = 67$, which also happens to be an optimum selection.

4.2 Multi-step planning under budget uncertainty

We now show that a variant of the local greedy algorithm from (Fisher, Nemhauser, and Wolsey 1978), adapted to satisfy type feasibility, is both practical and provably robust. A key difficulty in solving MOPGP lies in the abstract nature of type feasibility: Definition 1 specifies high-level desiderata but does not provide a directly actionable definition. While the objective f is non-decreasing and submodular, the proportional constraints across types introduce combinatorial complexity that does not decompose neatly.

To make the notion of type feasibility operational, we introduce a fixed tie-breaking rule $\sigma : [r] \to [r]$ over types and restrict our attention to a structured subclass of solutions, which we call σ-type feasible selections.

Definition 8 (σ-type feasible selection)**.** Let $\sigma : [r] \to [r]$ be an arbitrary total preference ordering over types. A subset $\mathbf{S} \subseteq \mathbf{V}$ of size b is said to be σ-type feasible if it is type feasible *and* for $\mathbf{Q}(\mathbf{S}^{(1:t)})$, favors types according to σ: if

$\sigma(q) < \sigma(q')$, then type q is preferred over q'.

This reduction justifies restricting our attention to the class of σ-type feasible solutions: they are not only easier to operationalize algorithmically, but also provably approximate any type-feasible solution up to a factor that depends on the granularity of type availability.

Our next result tells us that any optimum σ-type feasible selection OPT_σ is competitive against any optimum type feasible selection OPT if at least k elements of each type is chosen per round when applying MULTISTEPPLANNING.

Theorem 9. *Let* OPT *be an optimal type-feasible selection and let* OPT_σ *be an optimal σ-type feasible selection. Fix an integer $k \in \mathbb{N}_{>0}$. If $|\mathrm{OPT}_\sigma \cap \mathbf{V}^{(t)} \cap \mathbf{T}_q| \geq k$ for all $t \in [h]$ and $q \in [r]$, then $f(\mathrm{OPT}) \leq \frac{k+1}{k} \cdot f(\mathrm{OPT}_\sigma)$.*

Note that we typically have $k \geq 1$: the Ethiopian Ministry of Health is currently building $2,000$ comprehensive health posts across ∼670 rural districts (OCHA 2023), and we expect more facilities to be built in the future.

Our algorithm MULTISTEPPLANNING is presented in Algorithm 2. At each round $t \in [h]$, MULTISTEPPLANNING determines how many comprehensive health posts $x_q^{(t)} \in \mathbb{N}$ from each district $q \in [r]$ should be selected (line 3), based on the desired global proportion and elements selected so far in $\mathbf{S}^{(1:t-1)}$. Note that $\sum_{q \in [r]} x_q^{(t)} = b^{(t)} \leq |\mathbf{V}^{(t)}|$. These type-specific quotas define a partition matroid (line 4), over which the algorithm performs greedy selection (lines 5-7) by iteratively adding the element with highest marginal gain while respecting the matroid. theorem 10 shows the formal guarantees of MULTISTEPPLANNING.

Algorithm 2: MULTISTEPPLANNING

Input: An MOPGP$(f, h, r, \mathbf{V}, \mathbf{b}, \mathbf{p})$ instance and a tie-breaking ordering $\sigma : [r] \to [r]$

Output: A σ-type feasible selection $\mathbf{S} = \mathbf{S}^{(1:h)}$

1: Define $\mathbf{S}^{(1)} = \ldots = \mathbf{S}^{(h)} = \emptyset$ ▷ Initialize selections

2: **for** $t = 1, \ldots, h$ **do**

3: For $q \in [r]$, define $\#(q, \mathbf{s}(b^{(1:0)}, b^{(1:0)})) = 0$ and

$$x_q^{(t)} = \#(q, \mathbf{s}(b^{(1:t)}, b^{(1:t)})) - \#(q, \mathbf{s}(b^{(1:t-1)}, b^{(1:t-1)}))$$

4: Let $\mathcal{M}^{(t)} = (\mathbf{V}^{(t)}, \mathcal{I}^{(t)})$ be the partition matroid at time t, where $\mathcal{I}^{(t)} = \{\mathbf{S} \subseteq \mathbf{V}^{(t)} : |\mathbf{S} \cap \mathbf{T}_q| \leq x_q^{(t)}\}$.

5: **for** $b^{(t)}$ times **do**

6: Let $\mathbf{S} = \bigcup_{\tau=1}^{t} \mathbf{S}^{(\tau)}$ be the selection so far

7: Add $e^* = \underset{\substack{e \in \mathbf{V}^{(t)} \setminus \mathbf{S} \\ \mathbf{S}^{(t)} \cup \{e\} \in \mathcal{I}^{(t)}}}{\operatorname{argmax}} g(\mathbf{S}, \{e\})$ to $\mathbf{S}^{(t)}$

8: Output $\mathbf{S}^{(1)} \uplus \ldots \uplus \mathbf{S}^{(h)}$

Theorem 10. *Fix an arbitrary tie-breaking ordering $\sigma : [r] \to [r]$. Assuming the objective function f can be evaluated in constant time given any subset,* MULTISTEPPLANNING *runs in $\mathcal{O}(\sum_{t=1}^{h} b^{(t)} \cdot (r + |\mathbf{V}^{(t)}|))$ time and outputs $\mathbf{S} = \uplus_{t=1}^{h} \mathbf{S}^{(t)}$ such that*

1. *For all $t \in [h]$, $\mathbf{S}^{(t)} \subseteq \mathbf{V}^{(t)}$ and $|\mathbf{S}^{(t)}| = b^{(t)}$.*
2. *The output $\mathbf{S}$ is a σ-type feasible selection.*
3. *For all $t \in [h]$, we have*

$$f(\mathbf{S}^{(1)} \uplus \ldots \uplus \mathbf{S}^{(t)}) \geq \frac{1}{2} \cdot f(\mathrm{OPT}_\sigma(b^{(1)}, \ldots, b^{(t)}))$$

where $\mathrm{OPT}_\sigma(b^{(1)}, \ldots, b^{(t)})$ is an optimum σ-type feasible selection restricted to budgets $b^{(1)}, \ldots, b^{(t)}$.

4.3 Our theoretical bounds are tight in general

Proposition 11 shows that *any* solution to MOPGP violating Definition 1 is brittle against future budget uncertainties while Proposition 12 shows that the approximation ratio terms for MULTISTEPPLANNING are essentially tight.

Proposition 11. *Let* ALG *be any deterministic algorithm to* MOPGP *aiming to produce selections $\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(h)}$ that maximizes $\beta(b^{(1:t)})$ at each time step $t \in [h]$. If* ALG *does not minimize $|\mathbf{Q}(\mathbf{S}^{(t)})|$ for each $t \in [h]$, then there exists instances where it fails to minimize $\beta(b^{(1:t)})$ at each $t \in [h]$.*

Proposition 12. *There exists instances where* MULTISTEPPLANNING *achieves an approximation ratio of $\leq \frac{1}{2}$ due to budget uncertainties and $\leq \frac{k}{k+1}$ due to σ-ordering.*

5 Experiments

We empirically evaluate the performance of the individual components of our proposed HARP tool. See the extended version for additional experimental details.

5.1 Experimental setup

Data. To model facility coverage, we use point-to-point walking distance estimates between 1km-by-1km grid cells across Ethiopia, based on the global friction surface from (Weiss et al. 2020), and population forecasts from WorldPop projections for the years 2026–2030 (WorldPop 2025).[4] We are also given a set of existing facilities $\mathbf{S}_{\text{existing}}$ and available budgets were chosen based on expert consultation.

Regions analyzed. We focus on rural regions where access remains limited, i.e., major cities such as Addis Ababa are excluded. We evaluate our algorithms on the regions of Afar, Benishangul Gumuz, and Somali in this section, and on the Sidama region in the extended version. This way we reflect Ethiopia's two dominant rural livelihood types.[5]

There are around 30 districts in Afar and Sidama, around 20 in Benishangul Gumuz, and around 70 districts in Somali. A larger budget was considered for Somali as it is much larger than the other regions.

Distributional Policies (DP). To ensure equitable allocation, each region may define a distributional policy by specifying target proportions $p_1, \ldots, p_r$ across r districts. These type-level constraints shape the facility selection process by prioritizing areas with greater unmet needs. In our experiments, we evaluate three policy scenarios:

---

[4]Population forecasts are inherently uncertain and the true objective function f is not directly accessible. In the extended version, we show that the impact on facility coverage is not substantial.

[5]In this section, we focus on less populated regions to better highlight the spatial coverage gains achieved by our HARP.

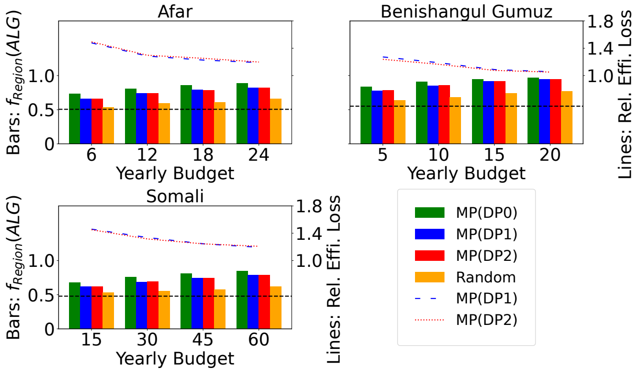

Figure 4: Coverage under varying annual budgets across the three regions of Afar, Benishangul Gumuz, and Somali. Bars show total population coverage by each policy, with the black line indicating existing coverage before any yearly budget was spent to build additional health posts. The dashed lines above indicate relative efficiency loss from enforcing proportional constraints for DP1 and DP2, compared to the unconstrained baseline (DP0).

DP0 No distributional constraint (i.e., $p_1 = \ldots = p_r = 0$), corresponding to unconstrained greedy selection.

DP1 Proportions p_q are set based on unassisted home birth rates, favoring districts with poor maternal care access.

DP2 Proportions p_q are set by the inverse of postnatal care coverage, favoring districts lacking early childhood services.

Both DP1 and DP2 target existing healthcare service gaps that can be provided by comprehensive health posts. While our framework supports multiple active constraints simultaneously, we focus on one distributional policy at a time to preserve theoretical guarantees and interpretability. In our experiments below, we write MP(DPi), for $i \in \{0, 1, 2\}$, to denote the output of Algorithm 2 applied with the proportional constraints imposed by each policy.

5.2 Experiment 1: Impact of budget on coverage

In fig. 4, the bars show how total coverage improves under different allocation strategies and budget levels over a five-year planning horizon. As all regions are large relative to the available budget, coverage increases approximately linearly with budget in all cases.

For $i \in \{1, 2\}$ in Afar, Benishangul Gumuz, and Somali, the dashed and dotted lines in fig. 4 show the ratios

$$\frac{f_{\text{region}}(\text{MP(DP0)} \cup \mathbf{S}_{\text{existing}}) - f_{\text{region}}(\mathbf{S}_{\text{existing}})}{f_{\text{region}}(\text{MP(DP}i) \cup \mathbf{S}_{\text{existing}}) - f_{\text{region}}(\mathbf{S}_{\text{existing}})},$$

quantifying the trade-off incurred by distributional constraints, where f is the coverage function defined in Eq (1).

Note that we subtracted away the coverage from existing health posts $\mathbf{S}_{\text{existing}}$ to measure the relative gain in coverage as existing health posts are assumed to remain operational throughout. This ratio is similar to the notion of price of fairness (Caragiannis et al. 2012): a ratio close to 1 indicates that enforcing proportional fairness incurs minimal efficiency loss, while higher values reflect stronger trade-offs. As expected, these ratios improve with increasing budgets, suggesting that the proportional constraints come at a lower cost when more resources are available.

5.3 Experiment 2: Evaluating district-level equity

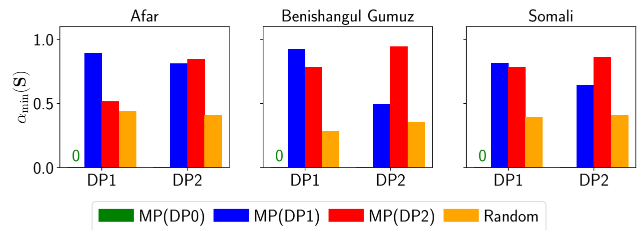

Figure 5: Min. satisfaction ratio $\alpha_{\min}$ attained under each policy, with proportions defined by DP1 or DP2. Over five-year horizon, annual budgets of 12 facilities were allocated for Afar, 10 for Benishangul Gumuz, and 30 for Somali.

In Experiment 1, we demonstrated that imposing proportional constraints does not significantly degrade coverage. We now show that these constraints improve adherence to the intended proportions. To this end, we examine the minimum satisfaction ratio, $\alpha_{\min}(\mathbf{S})$, achieved under each policy. As shown in fig. 5, MP(DP1) and MP(DP2) outperform all other methods in the DP1 and DP2 metrics respectively, while MP(DP0) consistently attains $\alpha_{\min} = 0$. Interestingly, in the Afar region, MP(DP1) yields significant gain under DP1 than MP(DP2) does under DP2, whereas the reverse trend is observed in the Benishangul Gumuz and Somali regions.

5.4 Experiment 3: High quality advice can help

To evaluate the practical utility of our learning-augmented approach, we conduct a *retrospective analysis* using the set of comprehensive health posts $\mathbf{A}$ that were previously constructed by regional planners: we simulate the original decision-making scenario in which $|\mathbf{A}|$ health posts must be selected from scratch. See the extended version for details.

6 Conclusion

In partnership with Ethiopia's Ministry of Health and Public Health Institute, we developed a decision-support tool (HARP) to guide health facility planning. HARP allows regional planners to solve their own instance of MOPGP within a unified, principled framework that models the planning task as a constrained submodular optimization with online budget and global proportionality constraints. On the technical front, we are the first to integrate learning-augmented algorithms with multi-step submodular optimization, enabling principled refinement of expert plans with worst-case guarantees. Using real-world health and demographic data from Ethiopia, we show that our methods are effective in supporting equitable, data-driven infrastructure decisions across diverse regional contexts.

References

Agarwal, A.; and Balkanski, E. 2024. Learning-Augmented Dynamic Submodular Maximization. In *Neural Information Processing Systems (NeurIPS)*, 14148–14176.

Asudeh, A.; Berger-Wolf, T.; DasGupta, B.; and Sidiropoulos, A. 2023. Maximizing Coverage While Ensuring Fairness: A Tale of Conflicting Objectives. *Algorithmica*, 85: 1287–1331.

Babaioff, M.; Immorlica, N.; and Kleinberg, R. 2007. Matroids, Secretary Problems, and Online Mechanisms. In *Symposium on Discrete Algorithms (SODA)*, 434–443.

Calinescu, G.; Chekuri, C.; Pal, M.; and Vondrák, J. 2011. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6): 1740–1766.

Caragiannis, I.; Kaklamanis, C.; Kanellopoulos, P.; and Kyropoulou, M. 2012. The Efficiency of Fair Division. *Theory of Computing Systems*, 50: 589–610.

Celis, L. E.; Huang, L.; Keswani, V.; and Vishnoi, N. K. 2019. Classification with Fairness Constraints: A Meta-Algorithm with Provable Guarantees. In *Conference on Fairness, Accountability, and Transparency (FAT)*, 319–328.

Cristi, A.; Dütting, P.; Kleinberg, R.; and Leme, R. P. 2024. Online Matroid Embeddings. *arXiv preprint arXiv:2407.10316*.

Fisher, M. L.; Nemhauser, G. L.; and Wolsey, L. A. 1978. *An analysis of approximations for maximizing submodular set functions – II*, volume 8. Springer.

Getnet, F.; Forzy, T.; Tesfaye, L.; Misganaw, A.; Memirie, S. T.; Geremew, S.; Berheto, T. M.; Wendrad, N.; Yihun, B. Y.; Mirutse, M. K.; and et al. 2025. Inequalities in tuberculosis control in Ethiopia: A district-level distributional modelling analysis. *Tropical Medicine & International Health*, 30(1): 31–42.

Goundan, P. R.; and Schulz, A. S. 2007. Revisiting the Greedy Approach to Submodular Set Function Maximization. *Optimization Online*.

Hendrix, N.; Warkaye, S.; Tesfaye, L.; Woldekidan, M. A.; Arja, A.; Sato, R.; Memirie, S. T.; Mirkuzie, A. H.; Getnet, F.; and Verguet, S. 2023. Estimated travel time and staffing constraints to accessing the Ethiopian health care system: A two-step floating catchment area analysis. *Journal of Global Health*, 13: 04008.

Liu, Q. C.; and Srinivas, V. 2024. The Predicted-Updates Dynamic Model: Offline, Incremental, and Decremental to Fully Dynamic Transformations. In *Conference on Learning Theory (COLT)*, 3582–3641. Proceedings of Machine Learning Research (PMLR).

Lykouris, T.; and Vassilvitskii, S. 2021. Competitive Caching with Machine Learned Advice. *Journal of the ACM (JACM)*, 68(4): 1–25.

Ministry of Health [Ethiopia]. 2020. Realizing Universal Health Coverage through Primary Healthcare: A Roadmap for Optimizing the Ethiopian Health Extension Program (2020–2035).

Mitzenmacher, M.; and Vassilvitskii, S. 2022. Algorithms with Predictions. *Communications of the ACM*, 65(7): 33–35.

Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An analysis of approximations for maximizing submodular set functions – I. *Mathematical Programming*, 14: 265–294.

OCHA, U. N. 2023. Ethiopia - Subnational Administrative Boundaries. Available at https://data.humdata.org/dataset/cod-ab-eth.

Recski, A. 1973. On partitional matroids with applications. *Coll. Math. Soc. J. Bolyai*, 10: 1169–1179.

Santiago, R.; Sergeev, I.; and Zenklusen, R. 2025. Constant-competitiveness for random assignment Matroid secretary without knowing the Matroid. *Mathematical Programming*, 210(1): 815–846.

Teklu, A. M.; Alemayehu, Y. K.; Medhin, G.; et al. 2020. National Assessment of the Ethiopian Health Extension Program: Final Master Report. Technical report, MERQ Consultancy PLC, Addis Ababa, Ethiopia.

Tsang, A.; Wilder, B.; Rice, E.; Tambe, M.; and Zick, Y. 2019. Group-Fairness in Influence Maximization. *International Joint Conference on Artificial Intelligence (IJCAI)*.

UNFPA. 2022. Ethiopia population, 2022. Accessed via UNFPA resources.

Wang, H.; Tesfaye, R.; Ramana, G. N. V.; and Chekagn, C. T. 2016. *Ethiopia Health Extension Program: An Institutionalized Community Approach for Universal Health Coverage*. World Bank.

Weiss, D. J.; Nelson, A.; Vargas-Ruiz, C. A.; Gligorić, K.; Bavadekar, S.; Gabrilovich, E.; Bertozzi-Villa, A.; Rozier, J.; Gibson, H. S.; Shekel, T.; and et al. 2020. Global maps of travel time to healthcare facilities. *Nature Medicine*, 26: 1835–1838.

Worldometers. 2022. Ethiopia Population (2022) – Worldometer. https://www.worldometers.info/world-population/ethiopia-population/. Accessed July 2025.

WorldPop. 2025. WorldPop Population Counts 2015–2030: Ethiopia — Spatial Distribution of Population. https://data.humdata.org/dataset/worldpop-population-counts-2015-2030-eth. Constrained estimates of total population per 100m grid cell in Ethiopia for 2015–2030, Geotiff format, CC BY4.0 license.