Visit Allocation Problems in Multi-Service Settings

Breugem, Thomas; Wolters, Tim Sergio; Van Wassenhove, Luk N.

Publication date:
2023

Document Version
Early version, also known as pre-print

CentER

Discussion Paper

No. 2023-004

VISIT ALLOCATION PROBLEMS IN MULTI-SERVICE SETTINGS: POLICIES AND WORST-CASE BOUNDS

By

Thomas Breugem, Tim Sergio Wolter,
Luk N. Van Wassenhove

17 January 2023

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

Visit Allocation Problems in Multi-Service Settings: Policies and Worst-Case Bounds

Thomas Breugem

Department of Management, Tilburg University, t.breugem@tilburguniversity.edu

Tim Sergio Wolter

Technology and Operations Management, INSEAD, timsergio.wolter@insead.edu

Luk N. Van Wassenhove

Technology and Operations Management, INSEAD, luk.van-wassenhove@insead.edu

January 16, 2023

Abstract

Problem definition: We consider a resource allocation problem faced by health and humanitarian organizations deploying mobile outreach teams to serve marginalized communities. These teams can provide a single service or an assortment of services during each visit. Combining services is likely to increase operational efficiency but decrease the relative benefit per service per visit, as operations are no longer tailored to a single service. The aim of this study is to analyze this benefit-efficiency trade-off. **Academic/practical relevance:** Increased operational efficiency will enable organizations to serve more people using fewer resources. This is important given the increasing funding gap organizations are facing. Our work adds to the literature on resource allocation problems and visit allocation problems specifically, where the focus has been primarily on single services. **Methodology:** We analyze a general visit allocation problem incorporating demand distribution (where to go) and return time (how frequently to go). We derive analytical bounds for the benefit-efficiency trade-off, and propose visit allocation policies with worst-case optimality guarantees. **Results:** Our results show the benefit-efficiency trade-off can be assessed based on high level parameters. We show demand alignment is a key driver of this trade-off. We apply our results to Praesens Care, a social enterprise start-up developing mobile diagnostic laboratories, and verify our insights using real-world data. **Managerial Implications:** Our research contributes to the discussion on innovation and increased efficiency in health and humanitarian aid delivery by quantifying operational trade-offs in offering assortments of services. Specifically, our results help assess the potential of integrated models for health and humanitarian aid delivery and provide organizations with easy-to-implement methods to determine close-to-optimal visiting policies. Importantly, our methods remain applicable in case of limited data, making them suitable for strategic decision-making.

Keywords: Resource Allocation, Health-Delivery Optimization, Visit Allocation, Mobile Lab Deployment, Worst-Case Analysis

1 Introduction

Visit allocation problems consider the deployment of mobile units to a set of locations to maximize the benefit of a given service. They are a type of resource allocation problem frequently studied in the context of outreach programs, where teams travel to rural areas to provide health and humanitarian services (see McCoy and Lee [2014], de Vries et al. [2021a], Alban et al. [2021], Breugem and Van Wassenhove [2022], among others). The key decisions are which locations to visit and how often, or with which frequency these locations should be visited. In particular, an optimal policy depends on the demand distribution (where to go) and the sensitivity to waiting/return time (how frequently to go) for a service.

Recent technological advancements increase the potential for integrating services, that is, a single mobile unit offering multiple services during a single visit. This ideally allows services to be provided more efficiently and hence enable organizations to serve more people using fewer resources. This is especially important in light of the increasing funding gap that non-governmental organizations are facing (Besiou and Van Wassenhove [2020]). However, different services have different characteristics. For example, reproductive health services are likely less urgent than diagnostics for highly infectious diseases or diseases with a short time until severe illness. These characteristics directly influence the optimal visit allocation: low urgency allows for visiting many different locations routinely, whereas high urgency necessitates focusing on a few locations with highest demand. Hence, different services will have different optimal visit allocation policies. This highlights an important trade-off between benefit and efficiency in multi-service settings: Combining services during a single visit is likely to increase operational efficiency, because integration allows to decrease the total number of visits, but decrease the relative benefit per service per visit, as operations are no longer tailored to a single service. This benefit-efficiency trade-off is not well-understood, and, as a result, it is not clear what the potential benefit of integration is nor what best visiting policies are given multiple services. In this research, we aim to fill this gap and analyze the benefit-efficiency trade-off for visit allocation in multi-service settings.

This research is motivated by Praesens Care, a social enterprise start-up producing mobile diagnostic laboratories (mobile labs) used predominantly in low- and middle-income countries in Africa.[1] Mobile labs are a promising approach to improving access to diagnostics in low- and middle-income countries, their potential recognized by large donors such as the European Union.[2] The newest generation of mobile labs have similar diagnostic capacities as stationary labs. Specifically, Praesens Care's mobile labs have a modular design which allow a single lab to provide state-of-the-art diagnostic services for a large number of pathogens. Furthermore, the labs have the additional benefit of being agile and quickly deployed in case of emergencies. In general, their mobility allows for pooling of resources and equipment in situations that do not require permanent availability of testing facilities.

Praesens Care's mobile labs have been successfully used in outbreak response for Ebola, and were an integral part of the response during the 2017 and 2018 Dengue outbreaks in Senegal. Having deployed

[1] For more information, see: `https://praesens.care/`

[2] See: `https://ec.europa.eu/international-partnerships/stories/eu-funded-mobile-labs-help-tackle-coronavirus-africa_en`.

their labs primarily in emergency settings, they face a number of strategic questions regarding non-emergency deployment. Specifically, what are the best use cases for the mobile lab? How can the mobile lab systematically (rather than ad-hoc) generate impact, and how much impact? And how does this depend on the offered services and operational context? Using mobile labs for non-emergency deployment is an important opportunity, since emergency deployment typically only happens a fraction of the year. Fundamental to answering these questions is assessing the benefit generated from the lab's ability to offer multiple services during a single visit.

The contributions of this paper are twofold. Firstly, we derive analytical bounds for the benefit-efficiency trade-off for visit allocation problems in multi-service settings. We model the visit allocation problem as a resource allocation problem with single resource constraint and assume the benefit function at each location for each service is proportional to a general non-decreasing concave function depending on the service characteristics. This set-up is general and can also be applied to other practical problems. Secondly, we use our analytical results to derive visit allocation policies maximizing the worst-case relative benefit from offering an assortment of services. We show these policies can be efficiently determined by solving a linear program. Our results show alignment in demand is the key driver of the benefit-efficiency trade-off. Our methods remain applicable in case of limited data, making them suitable for decision-making in cases where data is scarce. This makes them particularly suitable for strategic decision-making, where impact estimation (e.g., to convince potential donors) often precedes large-scale data collection. Furthermore, additional information can be incorporated when available. We conclude with a computational study based on Praesens Care's recent operations, and show how our results help assess the benefit-efficiency trade-off and lead to near-optimal policies.

The remainder of this paper is organized as follows. In Section 2, we discuss related work and position our analysis in the literature. In Section 3, we formalize the problem setting, and we provide our main results in Section 4. In Section 5, we apply our results to a case study based on Praesens Care. Section 6 summarizes our main results and conclusions.

2 Related Work

Our work contributes to three streams of literature: visit allocation problems in health and humanitarian operations, robust optimization and related concepts simultaneous optimization and approximate resource allocation, and the analysis of worst-case bounds in resource allocation, specifically in the context of fairness.

Visit allocation problems have received considerable attention in recent years. McCoy and Lee [2014] consider the impact of fairness requirements in outreach visit planning. The proposed model is applied to data from Riders for Health, a non-governmental organization operating throughout Africa. de Vries et al. [2021a] and Alban et al. [2021] consider planning outreach visits at MSI Reproductive Choices (MSI), a major non-governmental organization operating in 37 countries. de Vries et al. [2021a] determine optimal and heuristic visit policies for maximizing client volume, which are shown to perform well via a case study at MSI. Alban et al. [2021] consider a similar problem but with sigmodal, rather than concave, objective functions, to model service adoption. Finally, de Vries et al.

[2021b] consider active case finding for Human African Trypanosomiasis (HAT). The authors propose seven heuristics (three based on optimization, four on intuitive decision rules) and analyze their performance. Our work relates closely to abovementioned papers. However, while their focus is primarily on modeling benefit and optimizing visits for a single service, we focus on visit allocation when offering multiple services and the benefit-efficiency trade-off this entails.

Relevant work in the health delivery domain more generally includes Deo et al. [2013], on optimizing follow-up testing for patients that need treatment over a longer period, McCoy and Johnson [2014], on the impact of adherence (continuing treatment) on budget allocation decisions over multiple periods, Deo and Sohoni [2015], on the trade-off between centralized diagnostic networks and point-of-care testing, Deo et al. [2015], on modeling HIV screening, testing, and care, and Jónasson et al. [2017], on improving supply chains for Early Infant Diagnosis for HIV. Also in this research, the focus is typically on improving operations for a single service. We refer to Jónasson et al. [2022] for a general overview of Operations Management/Research literature focused on social impact, among which health delivery optimization, in low- and middle-income country settings.

Robust optimization considers optimizing over uncertain parameters confined to a given uncertainty set (see Ben-Tal et al. [2009] and Gorissen et al. [2015] for overviews). One key advantage of this approach is that it requires limited data and/or assumptions on underlying distributions. The goal is to find a solution that performs well (best) for all possible parameter combinations in the uncertainty set. Our approach is similar in that we aim to find visit allocation policies that perform well for multiple services, although these are assumed to be known rather than uncertain. One key difference with most robust optimization literature is our focus on the shape of the objective function, rather than uncertainty in (linear) constraints. In this light, the recent work of Chen et al. [2022] is particularly relevant. The authors consider a two-stage resource allocation problem among multiple regions, similar in structure to the model in this paper, where the cost function for each region is unknown, albeit assumed to be monotonic. Given limited data on the shape of the cost functions, a new type of uncertainty set based on statistic goodness-of-fit tests for monotonic functions is proposed to provide a tractable robust formulation of the problem that converges to the 'true' problem when more data becomes available. Our underlying problem is conceptually similar but our approach differs by focusing on analytical insights, thereby imposing additional structural assumptions on the problem, rather than data-driven methods to model the objective functions.

Our analytical approach relates closely to the work of Goel and Estrin [2005] and Goel and Meyerson [2006] on simultaneous optimization and Breugem et al. [2022] on approximate resource allocation problems. Simultaneous optimization considers selecting a single resource allocation that is guaranteed to be within a certain factor of the optimum for a given set of utility functions. Goel and Meyerson [2006] consider a general allocation problem and similar utility functions for all players, while Goel and Estrin [2005] allow non-identical utility functions (assumed to be scalar multiples of a single underlying function, similar to our approach) in the context of aggregation trees. The approximate resource allocation problem introduced in Breugem et al. [2022] provides a framework for analyzing decision making when one relies on approximations of the utility functions. The authors provide worst-case bounds for the utility derived from resource allocations when arbitrarily differences in utility functions are allowed.

Finally, our approach bears resemblance to other work on worst-case bounds in resource allocation, specifically work on the price of fairness (see, e.g., Bertsimas et al. [2011], Caragiannis et al. [2012], Bertsimas et al. [2012], Gur et al. [2021], Breugem and Van Wassenhove [2022]). The latter considers the worst-case loss resulting from imposing fairness constraints (e.g., max-min fairness, proportional fairness, or outcome constraints). Besides the difference in context (that is, fairness instead of multi-service settings), our analysis differs from this research by focusing on the relative benefit reduction for a given resource allocation (visit allocation policy) rather than the reduction from imposing additional constraints (fairness requirements).

Summarizing, there exists substantial literature on visit allocation problems in health delivery, and optimization in health delivery more generally. However, the focus is typically on modeling benefit and optimizing operations for a single service rather than multiple services simultaneously. In this research, we aim to bridge this gap by analyzing the benefit-efficiency trade-off and worst-case optimal policies for visit allocation problems in multi-service settings, thereby also extending the literature on worst-case analysis in resource allocation.

3 Problem Formulation

In this section, we formalize the visit allocation problem and the assumptions on the benefit functions. We formulate our terminology in terms of lab visit allocation, but note that the model and assumptions generalize across contexts.

3.1 Preliminaries

We consider scheduling mobile lab visits across n locations during a given planning period. Without loss of generality, we assume the length of the planning period is standardized to one. Hence, instead of allocating visits, we allocate a fraction of the visit capacity over the period. For example, in case of weekly visits over a 12 week planning period, allocating a fraction 0.5 of the visit capacity is equivalent to scheduling 6 visits. Let m be the number of services offered by the lab.

Given location i, for $i = 1, \dots, n$, and service j, for $j = 1 \dots, m$, we assume the benefit generated from site visits can be expressed as follows:

Assumption 1. *Consider a given location i, for $i = 1, \dots, n$, and service j, for $j = 1 \dots, m$. The benefit of allocating a fraction $v \in [0, 1]$ of the visit capacity to location i can be written as:*

$$\lambda_{ij} f_j(v), \tag{1}$$

with λ_{ij} a non-negative demand parameter and $f_j : [0, 1] \to [0, 1]$ the average benefit per time period per person as a function of v. The function f_j is a non-decreasing concave and continuous function satisfying $f_j(0) = 0$ and $f_j(1) = 1$.

Assumption 1 states f_j is a non-decreasing concave and continuous function of the allocated visit capacity. Furthermore, Assumption 1 states the benefit can be decomposed into two parts: one non-linear part specific to the service and the same for all locations (up to the allocated visits), and a demand parameter unique to each location-service pair. Intuitively, this means that heterogeneity in locations can be fully captured by differences in demand. The assumption of concavity and a fixed non-linear underlying function are common in visit allocation problems within the outreach context (e.g., McCoy and Lee [2014], de Vries et al. [2021a], Breugem and Van Wassenhove [2022]). The assumption of concavity is common for resource allocation in general (see, for example, the discussion in Bertsimas et al. [2011]). Cases where Assumption 1 does not necessarily hold include long term visit allocation with adaptation dynamics [Alban et al., 2021].

We show in the appendix that Assumption 1 holds for a wide variety of services for which the benefit depends solely on waiting time. That is, where the benefit a person derives can be written as $b(w)$, with w the waiting time until service for that person and $b(w)$ a non-negative and non-increasing function. In this case, the 'concavity' of the benefit function captures how sensitive the benefit of the service is to waiting time: the more linear the benefit function, the more sensitive the service is to waiting time. To illustrate this, consider the function:

$$b_{w_1,w_2}(w) = \begin{cases} 1 & \text{if } w \leq w_1 \\ \frac{w_2 - w}{w_2 - w_1} & \text{if } w_1 \leq w \leq w_2 \\ 0 & \text{if } w \geq w_2. \end{cases} \quad (2)$$

This function indicates a person gets full benefit if the waiting time is below w_1, and zero benefit if the waiting time exceeds w_2. In between, benefit decreases linearly. The smaller w_1 and w_2, the more urgent the service. The benefit functions $b_{1,2}(w)$ and $b_{4,6}(w)$ are shown in Figure 1 (left). The resulting benefit as function of the fraction of visit capacity is shown in Figure 1 (right). Note that a fraction v of the visit capacity is equivalent to a return time, and hence waiting time, of $1/v$ weeks. We observe the function for $b_{4,6}$ is substantially more concave compared to $b_{1,2}$. This is because the former allows for longer waiting times, and hence more benefit in case of a small number of visits.

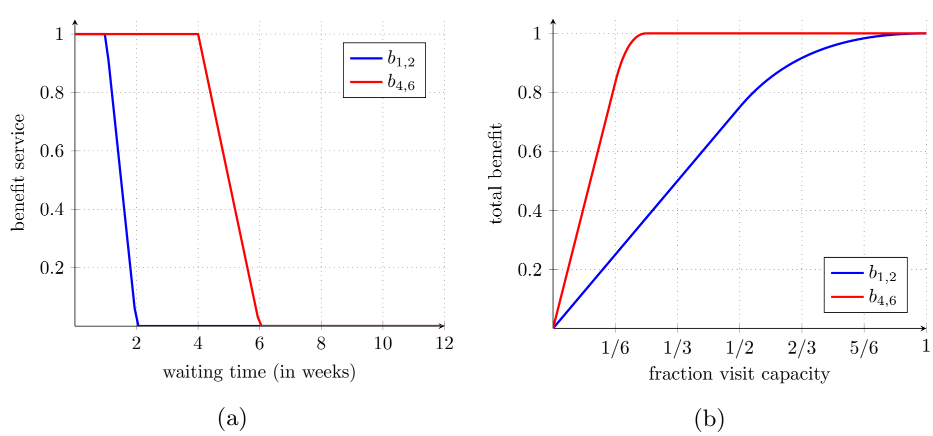


Figure 1: Example of benefit function given waiting time (left) and the resulting total benefit as a function of the fraction of visit capacity (right).

3.2 Mathematical Model

Under Assumption 1, we can model the problem of allocating site visits to maximize overall benefit as a resource allocation problem. Our model resembles the ones in McCoy and Lee [2014], Alban et al. [2021], and de Vries et al. [2021a], extended to multiple services.

We introduce decision variables $x_i \geq 0$, for $i = 1, \ldots, n$, indicating the fraction of visiting capacity allocated to location i. Let λ_{ij}, for $i = 1, \ldots, n$, and $j = 1, \ldots, m$, and f_j, for $j = 1, \ldots, m$, denote the demand parameters and the average benefit function, according to Assumption 1, for service $j = 1, \ldots, m$. With slight abuse of notation, we define λ_j as the n-dimensional vector with demand parameters λ_{ij}, for each service $j = 1 \ldots, m$. For a given service j, the problem of allocating mobile lab visits to maximize benefit can now be formulated as follows:

$$
\begin{aligned}
\mathrm{OPT}(\lambda_j, f_j) = \max \quad & \sum_{i=1}^{n} \lambda_{ij} f_j(x_i) && && \text{(3a)} \\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i \leq 1 && && \text{(3b)} \\
& x_i \leq 1 && i = 1, \ldots, n && \text{(3c)} \\
& x_i \geq 0 && i = 1, \ldots, n. && \text{(3d)}
\end{aligned}
$$

The Objective (3a) expresses we maximize total benefit. Constraint (3b) enforces the maximum number of visits. Finally, Constraints (3c) and (3d) specify the domain of the decision variables. We define $\mathrm{OPT}(\lambda_j, f_j)$ as the optimal value to (3). For a given allocation x, define $\mathrm{VALUE}(x, \lambda_j, f_j)$ as the objective value attained by x for service j:

$$\mathrm{VALUE}(x, \lambda_j, f_j) = \sum_{i=1}^{n} \lambda_{ij} f_j(x_i). \tag{4}$$

We are interested in the relative benefit ratio:

$$\mathrm{RATIO}(x, \lambda_j, f_j) = \frac{\mathrm{VALUE}(x, \lambda_j, f_j)}{\mathrm{OPT}(\lambda_j, f_j)}, \tag{5}$$

which captures the relative performance of x compared to the optimal solution for service j.

Using definition (5), we define the visit allocation problem given multiple services as the problem of maximizing the minimum benefit ratio over all services. We consider a visiting capacity $\gamma \geq 1$ for the labs offering multiple services (i.e., the number of mobile labs offering multiple services). The problem can then be formulated as follows:

$$\max \quad z \tag{6a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i \leq \gamma \tag{6b}$$

$$z \leq \text{RATIO}(x, \lambda_j, f_j) \qquad j = 1, \ldots, m \tag{6c}$$

$$x_i \leq 1 \qquad i = 1, \ldots, n \tag{6d}$$

$$x_i \geq 0 \qquad i = 1, \ldots, n \tag{6e}$$

$$z \geq 0. \tag{6f}$$

The Objective (6a) together with Constraints (6c) expresses we maximize the minimum relative benefit. Constraint (6b) enforces the visit capacity. Finally, Constraints (6d)–(6f) specify the domain of the decision variables.

Given (6), we define the benefit-efficiency trade-off as the optimal value to (6) as a function of γ, i.e., the relative benefit from offering multiple services at γ capacity compared to the benefit from m dedicated units of capacity (that is, one lab for each service). Note that when $\gamma = 1$, this relative benefit is by definition at most one. One key question is the lowest γ (highest level of efficiency) for which offering the m services simultaneously is guaranteed to provide the same benefit as operating m dedicated labs, one for each service. Below this level, there is a necessary trade-off between the efficiency from offering multiple services simultaneously and the relative benefit from dedicated labs with optimal visit policies tailored to each service. We quantify this level, and the trade-off below this level, in the next section.

4 Analysis

In this section, we present our main results. In Section 4.1, we consider the case of arbitrary demand, and in Section 4.2, we consider the case of two demand types (high/low) to derive further analytical insight. All proofs are omitted and presented in the appendix.

4.1 General Case

For a given n-dimensional demand vector v, define $\text{SUM}(v, \alpha)$ as the sum of the α largest elements of vector v:

$$\text{SUM}(v, \alpha) = \max \quad \sum_{i=1}^{n} v_i y_i \tag{7a}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} y_i \leq \alpha \tag{7b}$$

$$y_i \in [0, 1] \qquad i = 1, \ldots, n. \tag{7c}$$

We are now able to prove the main result:

Theorem 1. *Consider service j with n-dimensional demand vector λ_j and benefit function f_j, satisfying Assumption 1. Let non-negative scalars ℓ_j and u_j be such that* $\min\{\ell_j z, 1\} \leq f(z) \leq \min\{u_j z, 1\}$. *It holds that:*

$$\text{RATIO}(x, \lambda_j, f_j) \geq \text{BOUND}(x, \lambda_j, \ell_j, u_j), \tag{8}$$

where :

$$\text{BOUND}(x, \lambda_j, \ell_j, u_j) = \min_{\alpha \in [\ell_j, u_j]} \frac{\sum_{i=1}^{n} \lambda_{ij} \min\{\alpha x_i, 1\}}{\text{SUM}(\lambda_j, \alpha)}. \tag{9}$$

Furthermore, this bound is tight for all x.

Theorem 1 shows computing a tight lower bound on the ratio boils down to maximizing over a family of functions of remarkably simple structure, that is, of the form $\min\{\alpha z, 1\}$. The parameters ℓ_j and u_j in Theorem 1 provide a lower and upper bound on f_j. In particular, they bound the 'steepness' of the benefit function. High ℓ_j means it is likely not beneficial to visit a single location many times whereas low u_j implies the opposite. Note that $\ell_j = 1$ trivially holds for any function satisfying Assumption 1. Furthermore, it is not difficult to show that $\alpha \leq n$ always hold in the minimization problem in (8). This implies one can always set the upper bound $u_j = n$ in case no information is available. As such, imposing bounds ℓ_j and u_j allows for incorporating additional information on the benefit function, but does not restrict the generality of our results in any way. The following result states the case of no information explicitly:

Corollary 1. *Consider service j with n-dimensional demand vector λ_j and benefit function f_j, satisfying Assumption 1. It holds that:*

$$\text{RATIO}(x, \lambda_j, f_j) \geq \text{BOUND}(x, \lambda_j), \tag{10}$$

where:

$$\text{BOUND}(x, \lambda_j) = \min_{\alpha \in [1, n]} \frac{\sum_{i=1}^{n} \lambda_{ij} \min\{\alpha x_i, 1\}}{\text{SUM}(\lambda_j, \alpha)}. \tag{11}$$

Furthermore, this bound is tight for all x.

One can show only the values ℓ_j and u_j, and all integers in between, have to be considered for α in the minimization problem in (9). Given bounds ℓ_j and u_j for each service $j = 1, \ldots, m$, this means (6) can be reformulated into a tractable linear program to compute a tight lower bound on the relative benefit. Define p_j as the number of possible values for α given ℓ_j and u_j (that is, values ℓ_j and u_j, and all integers in between), and let α_{jh}, for $h = 1, \ldots, p_j$, denote these values. The linear program reads as follows:

$$\begin{aligned}
\max \quad & z && && \text{(12a)}\\
\text{s.t.} \quad & \sum_{i=1}^{n} x_i \leq \gamma && && \text{(12b)}\\
& z \leq \frac{\sum_{i=1}^{n} \lambda_{ij} y_{ijh}}{\text{SUM}(\lambda_j, \alpha_{jh})} && j = 1, \ldots, m,\ h = 1, \ldots, p_j && \text{(12c)}\\
& y_{ijh} \leq \alpha_{jh} x_i && i = 1, \ldots, n,\ j = 1, \ldots, m,\ h = 1, \ldots, p_j && \text{(12d)}\\
& y_{ijh} \leq 1 && i = 1, \ldots, n,\ j = 1, \ldots, m,\ h = 1, \ldots, p_j && \text{(12e)}\\
& y_{ijh} \geq 0 && i = 1, \ldots, n,\ j = 1, \ldots, m,\ h = 1, \ldots, p_j && \text{(12f)}\\
& x_i \geq 0 && i = 1, \ldots, n && \text{(12g)}\\
& z \geq 0. && && \text{(12h)}
\end{aligned}$$

Variables x_i and z are similar to Problem (6). Variables y_{ijh} are used to linearize the expression for the upper bound (9), via Equations (12d). Problem (12) has $\mathcal{O}\left(mn^2\right)$ variables and constraints, meaning it can be efficiently solved for practically-sized instances to compute the maximal worst-case benefit *and* obtain an allocation achieving this.

4.2 Low/High Demand Case

To gain further insight, we consider the special case of two demand types: for a given service, each location either has high demand, reflected by demand parameter $\mu \geq 1$, or low demand, reflected by a demand parameter standardized to one. This resembles the case where one has to balance visits between, e.g., large (urban) areas and small (rural) ones. We analyze the lowest γ for which offering multiple services during each visit is guaranteed to provide the same benefit as m dedicated labs, that is, when the increased number of visits one can provide by offering services simultaneously offsets the benefit loss from being unable to tailor the visits to a single service. We refer to this as the *dominating capacity* $\hat{\gamma}$, as it leads to at least as much benefit for each service as compared to dedicating one lab to each service separately.

For tractability reasons, we focus in this section on the case of no additional information on the benefit functions (i.e., $\ell_j = 1$ and $u_j = n$ for all j). Note this represents the worst-case, hence all bounds remain valid in case of additional information. We first consider the case where services have similar demand (up to scalar multiplication), after which we consider the case of limited (or no) alignment in demand.

4.2.1 Aligned Demand

Consider m services, for $j = 1, \ldots, m$, with n-dimensional demand vector λ_j. Assume each λ_j is proportional to the n-dimensional demand vector λ, i.e., are scalar multiples of this demand vector. For example, λ represents the population density. We have the following result:

Theorem 2. *Consider m services, for $j = 1, \ldots, m$, with n-dimensional demand vector λ_j propor-*

tional to λ, and benefit function f_j, satisfying Assumption 1. Assume λ satisfies $\lambda_i = \mu$ for the first k elements and $\lambda_i = 1$ for the remaining $n-k$ elements. The dominating capacity, denoted by $\hat{\gamma}(n, k, \mu)$, is bounded by:

$$\hat{\gamma}(n, k, \mu) \leq 1 + \left(1 - \frac{k}{n}\right)\left(1 - \frac{1}{\mu}\right). \tag{13}$$

Furthermore, this bound is tight.

Theorem 2 provides an easy to compute expression for an upper bound on the dominating capacity. Furthermore, this bound is tight, i.e., for any combination of parameters n, k, and μ there is an instance achieving this bound. One key insight from Theorem 2 is that the bound never exceeds two. That is, *offering multiple services at capacity two (i.e., two labs) guarantees a benefit ratio of at least one for any collection of services, provided demand is aligned.* Figure 3 shows the dominating capacity as a function of the demand parameter μ in the case of two demand types with a single high demand location ($k = 1$) and $n - 1$ low demand locations.

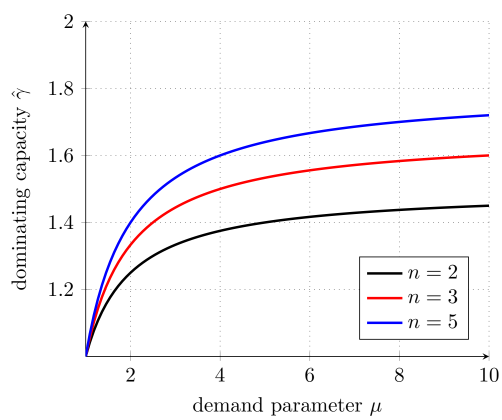


Figure 2: The dominating capacity $\hat{\gamma}$ as a function of the demand parameter μ in the case of two demand types with a single high demand location ($k = 1$) and $n - 1$ low demand locations.

The proof of Theorem 2 also provides an analytical bound on the benefit-efficiency trade-off, i.e., the worst-case benefit ratio given γ, and the allocation maximizing this ratio:

Corollary 2. *Consider m services, for $j = 1, \ldots, m$, with n-dimensional demand vector λ_j proportional to λ, and benefit function f_j, satisfying Assumption 1. Assume λ satisfies $\lambda_i = \mu$ for the first k elements and $\lambda_i = 1$ for the remaining $n - k$ elements. Let $z(\gamma)$ denote the optimal value to (6) given these m services. We have:*

$$z(\gamma) \geq \begin{cases} 1 & \text{if } \gamma \geq \hat{\gamma}(n, k, \mu) \\ \frac{\mu(n\gamma + k\mu - k)}{k(\mu-1)^2 + n(2-\mu)} & \text{otherwise.} \end{cases} \tag{14}$$

Furthermore, this bound is tight. The solution guaranteeing this bound allocates r resources uniformly to the k locations with demand μ and $\gamma - r$ uniformly to the remaining $n - k$ locations, where r is *given by:*

$$r = \begin{cases} \gamma - 1 + \frac{k}{n} & \text{if } \gamma \geq \hat{\gamma}(n,k,\mu) \\ 1 - \frac{\mu(2-\gamma)(n-k)}{\mu k^2 + (2\mu-1)(n-k)} & \text{otherwise.} \end{cases} \tag{15}$$

In general, the optimal allocation strikes a balance between a stationary solution ($r = \gamma$) and a completely mobile solution ($r/k = \gamma/n$). We remark that in case of no information on the benefit functions (i.e., $\ell_j = 1$ and $u_j = n$ for all j), the benefit ratio is at most one, because the benefit is naturally bounded by $\sum_{i=1}^{n} \lambda_i$, which is equal to the optimal benefit that can be achieved when setting $\alpha = n$ in Theorem 1. In Section 5, we also observe cases where the ratio can exceed one when additional information on the benefit functions is incorporated.

Figure 1 shows the bound on the benefit ratio (left) as a function of the demand parameter μ in the case of two demand types with a single high demand location and $n-1$ low demand locations and unit capacity for the mobile lab ($\gamma = 1$) for varying n, together with the bound on the benefit ratio for varying capacity levels γ for $n = 5$ (right). Figure 1 shows a single mobile lab can already achieve a high performance (well above 70% for all depicted n) over all benefit functions. It also shows the ratio improves substantially in the capacity. For $\gamma = 1.5$, the ratio is guaranteed to be one for $\mu \leq 8/3$, in line with Theorem 2.

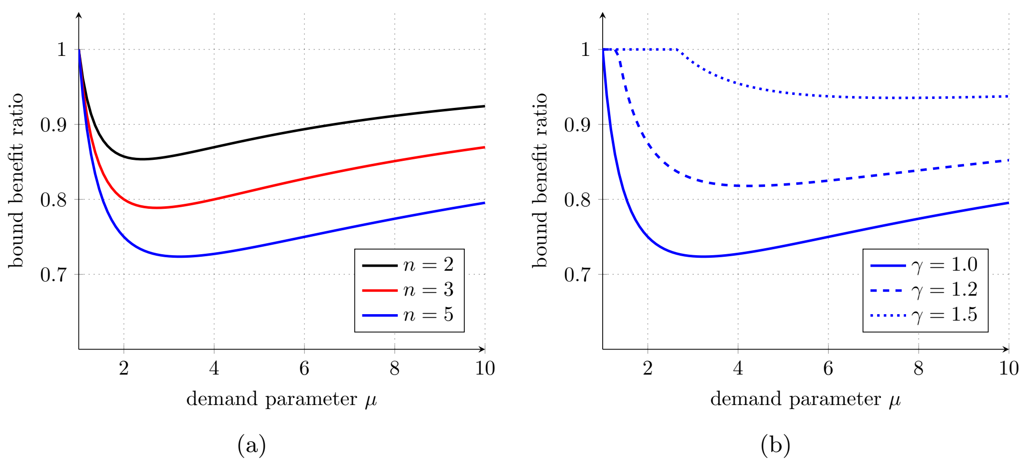


Figure 3: The bound on the benefit ratio (left) as a function of demand parameter μ in case of two demand types with a single high demand location ($k = 1$) and $n-1$ low demand locations for varying n, together with the bound on the benefit ratio (right) for different levels of capacity γ for $n = 5$.

4.2.2 Non-Aligned Demand

Next, we consider the impact of the alignment of demand. In particular, we assume again m services all with demand parameter μ at k locations. However, we assume limited overlap between demand, modeled via parameter h, indicating the number of locations with high demand μ for at least one service. Note that, by definition, we have $k \leq h \leq \min\{mk, n\}$. This parameter allows us to analyze the benefit from completely aligned demand ($h = k$) to completely non-aligned demand

($h = \min\{mk, n\}$), and all cases in between.

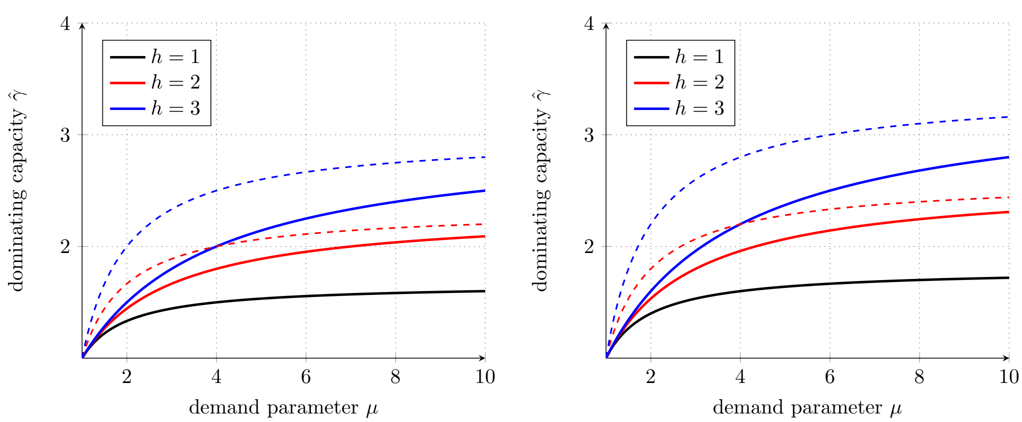
We have the following result:

Theorem 3. *Consider m services, for $j = 1, \ldots, m$, with n-dimensional demand vectors λ_j such that $\lambda_{ij} = \mu$ for k elements and $\lambda_{ij} = 1$ for the remaining $n-k$ elements, and benefit function f_j, satisfying Assumption 1. Let h indicate the number of locations with demand μ for at least one service. The dominating capacity, denoted by $\hat{\gamma}(n, k, h, \mu)$, is bounded by:*

$$\hat{\gamma}(n, k, h, \mu) \leq 1 + \frac{h}{n} \frac{(n-k)(\mu-1)}{k(\mu-1)+h}. \tag{16}$$

This bound is tight when $h = k$ or $h = \min\{mk, n\}$.

Theorem 3 incorporates misalingment of demand (via h) into an expression for the upper bound on the dominating capacity. Similar to Theorem 3, this expression is easy to compute. Figure 4 shows The dominating capacity as a function of the demand parameter μ in the case of two demand types with a single high demand location ($k = 1$) for $n = 3$, respectively $n = 5$, and a varying number of locations with high demand for at least one service (h).



(a) $n = 3$

(b) $n = 5$

Figure 4: The dominating capacity $\hat{\gamma}$ as a function of the demand parameter μ in the case of two demand types with a single high demand location ($k = 1$) and $n = 3$ low demand locations (left), respectively $n = 5$ low demand locations (right), for a varying number of locations with high demand for at least one service (h). The dashed lines indicate the bound obtained using the simplified expression (17) for the bound on the dominating capacity.

Similar to Corollary 2, the proof of Theorem 3 provides an analytical expression for the allocation corresponding to the bound. These expressions are presented in the proofs in appendix. Theorem 3 shows the dominating capacity increases less than linear in h. For h/k relatively close to one, a conservative, yet reasonable ballpark estimate in line with Theorem 2 is given by:

$$\hat{\gamma}(n, k, h, \mu) \leq 1 + \frac{h}{k} \left(1 - \frac{k}{n}\right) \left(1 - \frac{1}{\mu}\right). \tag{17}$$

The dashed lines in Figure 4 shows the bounds obtained using this expression. Note that for $h = k$, that is, $h = 1$ in Figure 4, the expression coincides with the bound of Theorem 2. Theorem 3 and expression (17) highlight demand alignment is an import driver of the benefit ratio. This is illustrated in Figure 4, where the dominating capacity increases substantially in h.

5 An Application in Mobile Laboratory Deployment

We apply our methods to a case study in Senegal based on Praesens Care. In Section 5.1 we describe the data and modeling assumptions. The computational results are presented in Section 5.2.

5.1 Data and Modeling Assumptions

We consider scheduling lab visits for the seven locations part of Praesens Care's pilot study over a period of 52 weeks. We consider a single lab and assume the lab stays at least one week at a location if visiting (to account for set-up time, among other things). This means we schedule 52 visits in total.

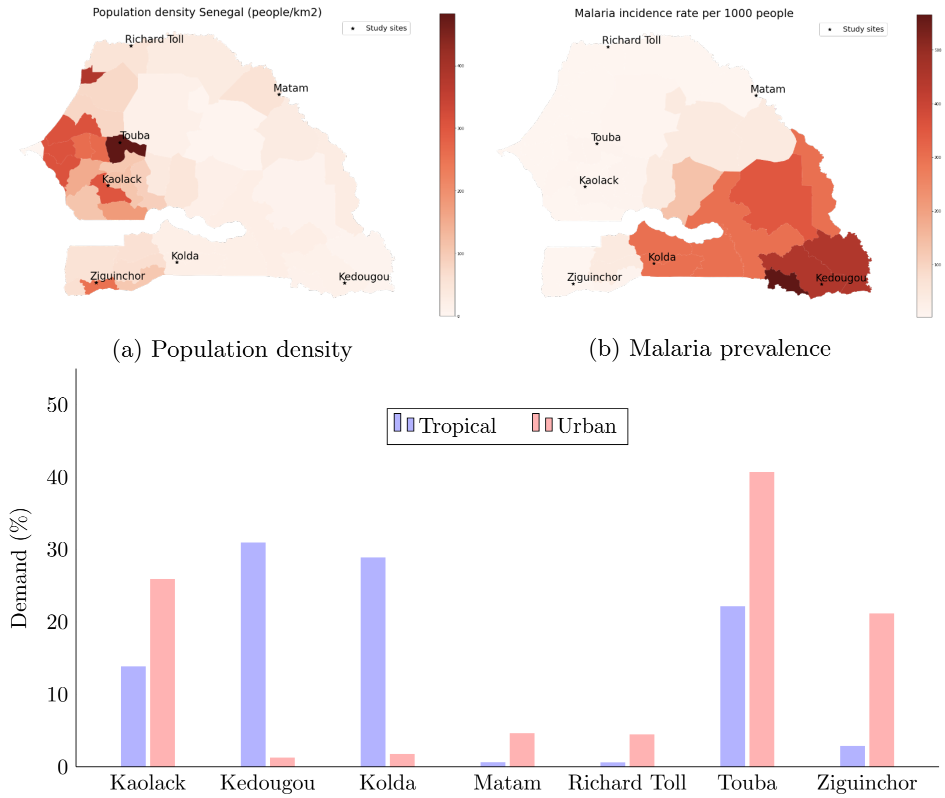
The locations represent diverse areas of the country, see Figure 5. We consider two demand distributions. The 'Urban' demand distribution, based on population density[3] (Figure 5a) representative for e.g., Tuberculosis and COVID-19 demand distributions, and the 'Tropical' demand distribution, taking Malaria incidence rates[4] (Figure 5b) as a representative sample. Urban demand is highest in the West of the country, especially around Touba. In contrast, Tropical demand is highest in the South East, especially around Kedougou.

Figure 5c shows the demand distribution for the seven locations. The tropical demand distribution is highest in Kedougou and Kolda, two locations in the South with low population density and hence (relatively) low urban demand. Koalack and Touba are densely populated areas and have the highest urban demand. They also have substantial tropical demand. In light of our analysis in Section 4, this implies that there is some alignment in the tropical and urban demand patterns. However, there is also misalignment for some locations. Ziguinchor, for example, is a clear example of high urban demand, but very low tropical demand.

Next to demand distribution, we consider the acuteness of the service. Acute services are for diseases that require immediate diagnosis and treatment. This includes Malaria, COVID-19, Influenza, and most arboviruses. Routine services are for (regular) treatment that is less time sensitive. Routine services are diagnostics and treatment for HIV/AIDS, Tuberculosis, diabetes, Hepatitis, among others. An essential difference between acute and routine services is the time a patient can wait for diagnosis. For example, visiting once every month might be sufficient for patients requiring routine Tuberculosis testing, but would imply diagnosis comes too late for many patients suffering from Malaria of COVID-19 diagnosis. Hence, the benefit generated from site visits depends on the time sensitivity (acute or routine) of the service.

[3]Obtained from: https://www.ansd.sn/index.php?option=com_regions&view=regions&Itemid=213, published by Agence Nationale de la Statistique et de la Démographie.

[4]Obtained from the U.S. President's malaria initiative Senegal operational plan, retrieved from https://www.pmi.gov/resources/malaria-operational-plans-mops/.



(a) Population density

(b) Malaria prevalence

(c) Demand distribution

Figure 5: Map of Senegal with the seven pilot locations.

To analyze the impact of acuteness, we consider three services with different levels of acuteness. We model each service via a waiting time function of the form (2). The three levels are High ($w_1 = 1, w_2 = 1$), Medium ($w_1 = 1, w_2 = 2$) , and Low ($w_1 = 4, w_2 = 8$) acuteness. That is, high acuteness implies benefit is generated only when a person is diagnosed within the same week of falling ill, medium urgency means there is still benefit when this person needs to wait for an additional week, while low urgency would still lead to benefit for waiting times up to two months.

5.2 Computational Results

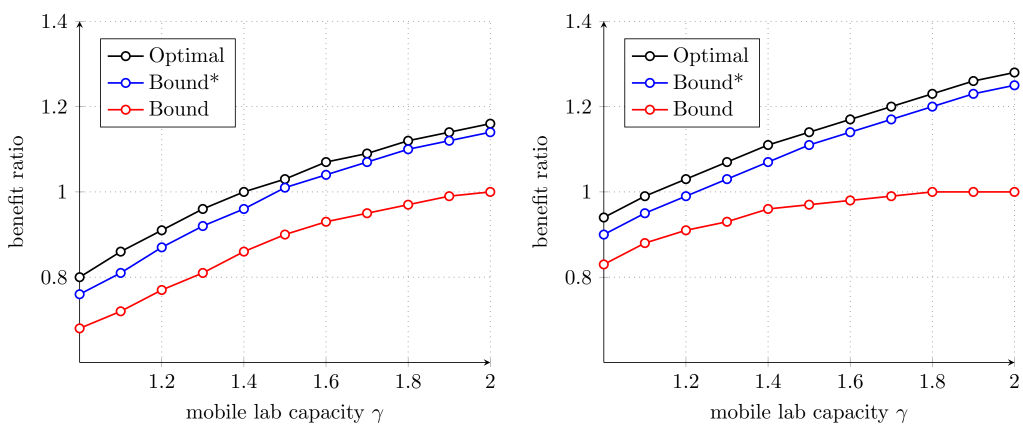
We first focus on the case $\gamma = 1$ to assess the benefit of introducing a single mobile lab offering multiple services. Table 1 shows the maximum benefit ratio for all combinations of demand types and acuteness levels. That is, row 'Medium Tropical' and column 'High Urban' means we consider the joint visit allocation problem for a service with tropical demand pattern and medium acuteness and a service with urban demand pattern and high acuteness. Table 1 also shows the lower bound on the benefit ratio with and without information on the lower and upper bounds on the benefit functions (denoted Bound* and Bound, respectively), obtained from (12). For a given function with pair w_1 and w_2, one can readily show lower and upper bound parameters ℓ and u are given by $\ell = w_1$ and $u = (w_1 + w_2)/2$. In case of no information on w_1 and w_2, we simply set $\ell = 1$ and $u = n$.

		High Urban			Medium Urban			Low Urban		
		Benefit	Bound*	Bound	Benefit	Bound*	Bound	Benefit	Bound*	Bound
High	Tropical	78.0	78.0	67.3	74.4	74.4	67.3	74.2	67.9	67.3
	Urban	100.0	100.0	82.5	93.0	91.8	82.5	86.6	84.1	82.5
Medium	Tropical	88.5	87.8	67.3	78.5	75.4	67.3	82.3	67.9	67.3
	Urban	93.0	91.8	82.5	100.0	91.8	82.5	92.8	84.1	82.5
Low	Tropical	78.4	75.7	67.3	79.7	75.7	67.3	94.9	85.9	67.3
	Urban	86.6	84.1	82.5	92.8	84.1	82.5	100.0	97.0	82.5

Table 1: The maximum relative benefit ratio for all combinations of demand types and acuteness levels, and the lower bound on the benefit with and without information on the lower and upper bounds on the benefit functions (denoted Bound* and Bound, respectively).

The results in Table 1 confirm demand alignment is a key driver of the benefit that can be achieved. In all cases, there is a substantial difference in benefit when considering two urban services versus one urban and one tropical service. The difference depends on the acuteness of the services. In case both services are low acuteness, the benefit is high (94.9%). In all other cases, it is substantially lower (almost always below 80%). When demand is aligned, the benefit is substantially higher. This is also reflected in the bounds: in case of no information (i.e., for *all* combinations of services), aligned demand has a benefit of at least 82.5%, whereas non-aligned can be as low as 67.3%. We also observe the bound with information provides a good estimate in most cases. The difference is largest in combinations with less acute services.

Next, we analyze the increase in benefit for $\gamma \geq 1$. Figure 6 shows maximum achievable benefit, and the lower bound with and without information, as function of the capacity γ for demand patterns Tropical/Urban (left) and Urban/Urban (right) For the sake of exposition, the curves represent the average computed over all possible service combinations.



(a) Tropical/Urban Demand

(b) Urban/Urban Demand

Figure 6: The maximum achievable benefit, and lower bounds with and without information, as function of the capacity γ for demand patterns Tropical/Urban (left) and Urban/Urban (right).

Figure 6 is in line with the analysis in Section 4. We observe substantially less than two units (1.4, respectively 1.1) of capacity are necessary to achieve the same performance as two dedicated labs. Furthermore, the benefit ratio quickly exceeds one. There is a substantial difference between the two demand patterns, in line with Theorem 3. Note that the bound without information stays below one, in line with Theorem 2. In general, the bound with information provides a very close estimate of the optimal benefit for a given capacity. The bound without information is more conservative, but still provides useful insight. For example, it shows that even without further information on the benefit functions, a capacity of 1.5, 25% lower compared to dedicated operations, will lead to a maximum benefit reduction of as little as 3%. This conservative estimate provides valuable information for decision-making. The realized benefit will be as least as good: a capacity of 1.5 will lead to an increase in benefit of 14%, as Figure 6 shows.

Finally, we consider the performance of the allocation obtained from solving (12). Table 2 also shows the bound with and without information for the found visit allocation. Again, we show the averages for each combination of demand type over all possible service combinations. As benchmarks, we consider three intuitive policies: a stationary policy, in which the lab stays put in the location with most demand, a completely mobile policy, in which the lab visits all locations equally, and a proportional allocation, which allocates visits proportional to the demand parameters.

	Tropical/Urban			Urban/Urban		
	Loss	Bound*	Bound	Loss	Bound*	Bound
Optimal	79.9	68.5	47.7	94.0	84.8	65.0
Best Bound*	75.8	75.5	51.6	91.3	89.9	67.3
Best Bound	69.5	67.6	67.3	85.6	82.5	82.5
Proportional	62.4	57.0	51.9	76.7	71.6	69.1
Stationary	44.0	43.4	22.2	61.7	59.5	40.9
Mobile	45.7	40.2	34.9	43.2	37.9	35.0

Table 2: The benefit and worst-case bounds (without and without information) for different allocation policies. The results shows the averages for each combination of demand type over all possible service combinations.

The results in Table 2 shows the best bound allocation policy easily outperforms all other policies and is close to optimal (on average) in case of information. Without information, the performance worsens, but the solution remains of high quality compared to the benchmark policies, where we see only the proportional policy achieves a reasonable performance. Table 2 also shows there is a certain limit to the value the worst-case bound can provide in determining a good allocation. The optimal allocation performs substantially worse in terms of the bound with information compared to the worst-case bound minimizing policy Best Bound*, and, similarly, both the optimal allocation and Best Bound* perform substantially worse in terms of the bound without information compared to the policy Best Bound.

6 Conclusion

In this paper, we considered a resource allocation problem faced by health and humanitarian organizations deploying mobile outreach teams to serve marginalized communities. In such programs, combining services is likely to increase operational efficiency but decrease the relative benefit per service per visit, as operations are no longer tailored to a single service. We analyzed a general visit allocation problem incorporating demand distribution (where to go) and return time (how frequently to go) to analyze this benefit-efficiency trade-off. In doing so, we derived analytical bounds and proposed visit allocation policies with worst-case optimality guarantees. We focused specifically on the dominating capacity, defined as the lowest capacity (i.e., highest efficiency) for which offering services simultaneously is guaranteed to provide the same benefit as operating dedicated labs. The set-up of our analysis is general and can also be applied to other practical problems.

Our results showed the benefit-efficiency trade-off and dominating capacity can be assessed based on high level parameters. In particular, we showed demand alignment is a key driver of both. We considered both general demand distributions and the special case of two types of demand, where locations either have high demand (e.g., urban) or low demand (e.g., rural). Among other results, we showed that for the case of two demand types, two units of mobile capacity guarantee a benefit ratio of at least one for any collection of services in case of aligned demand. We also provided expressions to compute how the dominating capacity increases in case of non-aligned demand.

We applied our results to a case study based on Praesens Care, a social enterprise start-up developing mobile diagnostic laboratories, and verified our insights using real-world data. In line with our analytical results, our experiments showed demand alignment is a key driver of benefit. We also observed that the derived bounds provided good estimates of the benefit and dominating capacity, especially when information on the benefit functions was incorporated. This also shows the potential of incorporating expert judgements or 'guestimates', into the bounds to estimate potential benefit. Finally, our results showed that the allocation policy derived from the analytical results performed well and easily outperformed all benchmark policies. Hence, our results can also help in operational decision-making, providing close-to-optimal policies with guaranteed minimum benefit.

We propose different avenues for further research. Firstly, our analysis did not incorporate the testing capacity of the lab. In case capacity is insufficient to cover demand for all services, prioritization is necessary in service delivery. This will likely change the benefit-efficiency trade-off. Secondly, interesting work remains to be done in modeling the benefit of services. This includes varying benefit functions for locations (beyond scalar multiplication) and benefit functions explicitly accounting for a patient's long term treatment (e.g., by means of follow-up visits). Since this would make the problem highly complex, new types of analysis are likely necessary. it would be interesting to determine which of our results remain true in these contexts. Finally, our analysis considers a stylized allocation problem to derive general insights. It would be interesting to apply our methods to different cases in health and humanitarian service delivery in practice.

A Generality Assumption 1

Assumption 1 holds for a wide variety of services for which the benefit depends solely on waiting time. That is, where the benefit a person derives can be written as $b(w)$, with w the waiting time until service for that person and $b(w)$ a non-negative and non-increasing function.

Given a benefit function $b(w)$, consider a given location and assume an exogenous (i.e., independent of lab visits) and steady demand rate λ per time period. Suppose the lab visits k times during time period T, with inter-arrival times δ_i, for $i = 1, \ldots, k$. Within the above framework, the problem of timing visits to maximize overall benefit can be written as:

$$\max \quad \sum_{i=1}^{k} \int_0^{\delta_i} \lambda b(w) \mathrm{d}w \tag{18a}$$

$$\text{s.t.} \quad \sum_{i=1}^{k} \delta_i = T \tag{18b}$$

$$\delta_i \geq 0 \quad i = 1, \ldots, k. \tag{18c}$$

Here, the Objective (18a) represents the benefit over time period T given interarrival times δ_i, and Constraint (18b) ensure the interarrival times sum to T. We have the following result:

Theorem 4. *Let $b : \mathbb{R}_+ \to \mathbb{R}_+$ be non-increasing. Define $v = k/T$ as the fraction of visit capacity allocated. The optimal solution to* (18) *is given by $\delta_i = 1/v = T/k$, for $i = 1, \ldots, v$. The resulting objective value is $\lambda f(v)$, with:*

$$f(v) = v \int_0^{\frac{1}{v}} b(w) \mathrm{d}w, \tag{19}$$

is concave and non-decreasing in v. The function $f : [0, 1] \mapsto [0, 1]$ represents the average benefit per time period per patient.

Proof. Without loss of generality, we assume $\lambda = 1$. The optimality of setting $\delta_i = T/k$ follows directly from the fact that $\int_0^{\delta_i} b(w) \mathrm{d}w$ is concave in δ_i (because $b(w)$ is non-increasing in w). The resulting objective value is given by:

$$\frac{k}{T} \int_0^{\frac{T}{k}} b(w) \mathrm{d}w = v \int_0^{\frac{1}{v}} b(w) \mathrm{d}w \tag{20}$$

We now show this function is concave in k, and hence in $v = k/T$. Consider k_1, k_2, and the convex combination $\theta k_1 + (1 - \theta) k_2$, for some $\theta \in [0, 1]$. To prove concavity we need:

$$\frac{\theta k_1 + (1-\theta) k_2}{T} \int_0^{\frac{T}{\theta k_1 + (1-\theta) k_2}} b(w) \mathrm{d}w \geq \frac{\theta k_1}{T} \int_0^{\frac{T}{k_1}} b(w) \mathrm{d}w + \frac{(1-\theta) k_2}{T} \int_0^{\frac{T}{k_2}} b(w) \mathrm{d}w. \tag{21}$$

Note the above is equivalent to:

$$\frac{\theta k_1 + (1-\theta)k_2}{T} \int_{\frac{T}{k_2}}^{\frac{T}{\theta k_1 + (1-\theta)k_2}} b(w)\mathrm{d}w \geq \frac{\theta k_1}{T} \int_{\frac{T}{k_2}}^{\frac{T}{k_1}} b(w)\mathrm{d}w. \tag{22}$$

Observe that:

$$\frac{\theta k_1 + (1-\theta)k_2}{\theta k_1} \left[\frac{T}{\theta k_1 + (1-\theta)k_2} - \frac{T}{k_2} \right] = \frac{T}{\theta k_1} \left[\frac{k_2 - \theta k_1 - (1-\theta)k_2}{k_2} \right] = \frac{T}{k_1} - \frac{T}{k_2}. \tag{23}$$

As such, we can write (22) in the form:

$$M \int_{\frac{T}{k_2}}^{\frac{T}{k_2}+\alpha} b(w)\mathrm{d}w \geq \int_{\frac{T}{k_2}}^{\frac{T}{k_2}+M\alpha} b(w)\mathrm{d}w, \tag{24}$$

with $M = \frac{\theta k_1 + (1-\theta)v_2}{\theta k_1}$ and $\alpha = \frac{T}{\theta k_1 + (1-\theta)k_2} - \frac{T}{k_2}$. This inequality clearly holds (because b is non-increasing), hence the result follows.

□

B Proofs Section 4

B.1 Proof Theorem 1

To prove Theorem 1, we first prove the following lemma:

Lemma 1. *Consider service j with n-dimensional demand vector λ_j and benefit function f_j, satisfying Assumption 1. Let non-negative scalars ℓ_j and u_j be such that* $\min\{\ell_j z, 1\} \leq f(z) \leq \min\{u_j z, 1\}$. *There exists a benefit function g of the form:*

$$g(z) = \min\{u_j z, \beta + \phi z, 1\}, \tag{25}$$

satisfying $g(z) \geq \min\{\ell_j z, 1\}$*, such that:*

$$\mathrm{RATIO}(x, \lambda_j, f_j) \geq \mathrm{RATIO}(x, \lambda_j, g) \tag{26}$$

Proof. For notational convenience, we drop the subscript j throughout the proof. Let function g minimize $\mathrm{RATIO}(x, \lambda_j, \cdot)$, and satisfy Assumption 1 and $\min\{\ell z, 1\} \leq g(z) \leq \min\{uz, 1\}$. Throughout we will refer to such a function as a *minimizing function.* Let y be a solution achieving $\mathrm{OPT}(\lambda, g)$. Without loss of generality, assume the values λ_i values are descending in i. By the structure of the objective (3a), it is then clear that also the values y_i can also assumed to be descending.

For notational convenience, define $y_0 = 1/\ell$ and $y_{n+1} = 0$. Note no benefit maximizing solution will have values exceeding $1/\ell$, hence the ordering of y_i values is preserved. First, we claim g can be assumed to be piece-wise linear with potential break points only at values y_i for $i = 0, \ldots, n$. To see this, suppose g does not satisfy this condition. We define the piece-wise linear function h by

$h(y_i) = g(y_i)$ for $i = 0, \dots, n+1$,

$$h(z) = h(y_{i+1}) + \frac{h(y_i) - h(y_{i+1})}{y_i - y_{i+1}}(z - y_{i+1}), \tag{27}$$

for $z \in [y_{i+1}, y_i]$ and $i = 0, \dots, n$, and finally $h(z) = 1$ for $z \geq y_0$. Note h is concave, and satisfies Assumption (1) and $\min\{\ell z, 1\} \leq h(z) \leq \min\{uz, 1\}$. Since g is concave and $h(y_i) = g(y_i)$ we have $h \leq g$ and $\sum_{i=1}^n \lambda_i h(y_i) = \sum_{i=1}^n \lambda_i g(y_i)$. It follows $\text{RATIO}(x, \lambda, h) \leq \text{RATIO}(x, \lambda, g)$, hence the assumption of piece-wise linearity is without loss of generality.

Next we show $g(z)$ can assumed to be of the form:

$$g(z) = \min\{uz, \beta + \phi z, 1\}. \tag{28}$$

for some non-negative scalars ϕ and β such that $g(z) \geq \min\{\ell z, 1\}$. Note that because g is piece-wise linear, it can be fully defined as the minimum of affine functions. To prove (28), it is sufficient to show that for any break point of g, i.e., any point y_i for which:

$$g(y_i) > g(y_{i+1}) + \frac{g(y_{i-1}) - g(y_{i+1})}{y_{i-1} - y_{i+1}}(y_i - y_{i+1}), \tag{29}$$

it must hold that $g(z) = \min\{uz, 1\}$. We can assume there is at least one break point, otherwise (28) clearly holds by setting $\phi = 0$ and $\beta = \ell$. Without loss of generality, assume g has the minimal number of break points out of all minimizing functions.

Let y_k, with $1 \leq k \leq n$, be a break point for which $g(y_k) < \min\{uy_k, 1\}$ (note y_0 and y_{n+1} will by definition satisfy $g(y_i) = \min\{u_j y_i, 1\}$). Furthermore, let indices $1 \leq e, f \leq n$ be the largest, respectively, smallest, indices such that y_e and y_f are break points and $e < k < f$. If no such index e, respectively f, exists, we set $e = 0$, respectively $f = n + 1$. Consider the function $\bar{g}_\delta$:

$$\bar{g}_\delta(z) = \begin{cases} g(y_f) + \left[\frac{g(y_k)+\delta-g(y_f)}{y_k-y_f}\right](z - y_f) & \text{if } z \in [y_f, y_k], \\ g(y_k) + \delta + \left[\frac{g(y_e)-g(y_k)-\delta}{y_e-y_k}\right](z - y_k) & \text{if } z \in (y_k, y_e], \\ g(z) & \text{otherwise.} \end{cases} \tag{30}$$

Essentially, $\bar{g}_\delta$ changes $g(y_k)$ to $g(y_k) + \delta$, while preserving concavity. The latter can be ensured by enforcing lower and upper bounds $\underline{\delta}$ and $\bar{\delta}$. It is clear the lower bound follows directly from (29):

$$\underline{\delta} = g(y_f) - \frac{g(y_e) - g(y_f)}{y_e - y_f}(y_k - y_f) - g(y_k). \tag{31}$$

The existence of an upper bound $\bar{\delta} > 0$ follows from the fact that $g(y_k) < \min\{u_j y_k, 1\}$, by assumption, and the way e and f are selected.

The difference between $\bar{g}_\delta$ and g is given by:

$$\Delta_\delta(z) = \bar{g}_\delta(z) - g(z) = \begin{cases} \left[\frac{\delta}{y_k - y_f}\right](z - y_f) & \text{if } z \in [y_f, y_k] \\ \left[\frac{\delta}{y_e - y_k}\right](y_e - z) & \text{if } z \in (y_k, y_e] \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

We get:

$$\min_{\delta \in [\underline{\delta}, \bar{\delta}]} \text{RATIO}_j(x, \lambda_j, \bar{g}_\delta) = \min_{\delta \in [\underline{\delta}, \bar{\delta}]} \frac{\text{VALUE}(x, \lambda, \bar{g}_\delta)}{\text{OPT}(\lambda, \bar{g}_\delta)} \tag{33}$$

$$\leq \min_{\delta \in [\underline{\delta}, \bar{\delta}]} \frac{\sum_{i=1}^n \lambda_i \bar{g}_\delta(x_i)}{\sum_{i=1}^n \lambda_i \bar{g}_\delta(y_i)} \tag{34}$$

$$= \min_{\delta \in [\underline{\delta}, \bar{\delta}]} \frac{\sum_{i=1}^n \lambda_i g(x_i) + \sum_{i=1}^n \lambda_i \Delta_\delta(x_i)}{\sum_{i=1}^n \lambda_i g(y_i) + \sum_{i=1}^n \lambda_i \Delta_\delta(y_i)} \tag{35}$$

$$\leq \frac{\sum_{i=1}^n \lambda_i g(x_i)}{\sum_{i=1}^n \lambda_{ij} g(y_i)} \tag{36}$$

$$= \frac{\text{VALUE}(x, \lambda, g)}{\text{OPT}(\lambda, g)} \tag{37}$$

$$= \text{RATIO}(x, \lambda, g). \tag{38}$$

Here, we use $\text{OPT}(\lambda, \bar{g}_\delta) \geq \sum_{i=1}^n \lambda_i \bar{g}_\delta(y_i)$, by definition for inequality (34). For inequality (36), we note that Δ_δ is linear in δ and hence the function in (35) is strictly non-increasing or strictly non-decreasing in δ. Hence, inequality (36) follows by picking δ equal to $\underline{\delta}$ (< 0) or $\bar{\delta}$ (> 0), accordingly. If the inequality is strict, it follows g is not a minimizing function. If (36) holds with equality, we set δ equal to $\underline{\delta}$ to obtain the function $\bar{g}_{\underline{\delta}}$ with one less break point (by definition of $\underline{\delta}$) and $\text{RATIO}(x, \lambda, \bar{g}_{\underline{\delta}}) \leq \text{RATIO}(x, \lambda, g)$, contradicting g is a minimizing function with minimal number of break points. As such, it must hold that $g(y_k) = \min\{uy_k, 1\}$, and hence g of the form (28). ☐

Following Lemma 1, we prove the following result:

Lemma 2. *Consider a service with with n-dimensional demand vector λ and benefit function g of the form:*

$$g(z) = \min\{uz, \beta + \phi z, 1\}. \tag{39}$$

Let scalars ℓ and u be such that $\min\{\ell z, 1\} \leq g(z) \leq \min\{uz, 1\}$. *We have:*

$$\text{RATIO}(x, \lambda, g) \geq \min_{\alpha \in [\ell, u]} \text{RATIO}(x, \lambda, \min\{\alpha z, 1\}). \tag{40}$$

Proof. The result is equivalent to claiming one can assume $\beta = 0$ in (39) for a minimizing function. Assume $\beta \in (0, 1)$, clearly $\beta = 1$ can be excluded since $\phi \geq 0$. Without loss of generality, assume g is the minimizing function of the form (28) with the largest β.

Consider adjusting β by an additive factor δ, giving the function:

$$\bar{g}_\delta(z) = \min\{uz, \beta + \delta + \phi z, 1\}, \tag{41}$$

with $|\delta| \leq \varepsilon$ sufficiently small.

First, consider VALUE(x, λ, g). Let J be all indices i for which $g(x_i) = \beta + \phi x_i$. If $\delta < 0$, we have:

$$\text{VALUE}(x, \lambda, g) - \text{VALUE}(x, \lambda, \bar{g}_\delta) = \sum_{i \in J} \lambda_i \delta. \tag{42}$$

If $\delta > 0$, we have:

$$\text{VALUE}(x, \lambda, \bar{g}_\delta) - \text{VALUE}(x, \lambda, g) \leq \sum_{i \in J} \lambda_i \delta. \tag{43}$$

Next, consider OPT(λ, g). We partition the indices i three sets. Let K_1 be all i such that $g(y_i) = u_j y_i$, K_3 all remaining for which $g(y_i) = 1$, and K_2 all indices not in K_1 nor K_2.

Suppose $\delta > 0$. Note for any $i \in K_3$ we can decrease y_i by δ/ϕ while ensuring $f(y_i) = 1$ remains true. We can transfer this amount uniformly to the values in K_1. This means we get:

$$\begin{aligned} \text{OPT}(\lambda, \bar{g}_\delta) - \text{OPT}(\lambda, g) &\geq \text{VALUE}(y, \lambda, \bar{g}_\delta) - \text{VALUE}(y, \lambda, g) && (44) \\ &= \sum_{i \in K_2} \lambda_i \delta + \frac{|K_3|}{|K_1|} \sum_{i \in K_1} \lambda_i u(\delta/\phi). && (45) \end{aligned}$$

Next, suppose $\delta < 0$. Note for any $i \in K_3$ we can increase y_i by δ/ϕ to ensure $f(y_i) = 1$ remains true. We can compensate for this by uniformly decreasing the values in K_1. Note there is always one index in K_1, otherwise there should be no break point. Note the slope we have when decreasing is at most u. This means we get:

$$\begin{aligned} \text{OPT}(\lambda, g) - \text{OPT}(\lambda, \bar{g}_\delta) &\leq \text{VALUE}(y, \lambda, g) - \text{VALUE}(y, \lambda, \bar{g}_\delta) && (46) \\ &\leq \sum_{i \in K_2} \lambda_i \delta + \frac{|K_3|}{|K_1|} \sum_{i \in K_1} \lambda_i u(\delta/\phi). && (47) \end{aligned}$$

It follows we have

$$\begin{aligned} \min_{\delta \in [-\varepsilon, \varepsilon]} \text{RATIO}(x, \lambda, \bar{g}_\delta) &= \min_{\delta \in [-\varepsilon, \varepsilon]} \frac{\text{VALUE}(x, \lambda, \bar{g}_\delta)}{\text{OPT}(\lambda, \bar{g}_\delta)} && (48) \\ &= \min_{\delta \in [-\varepsilon, \varepsilon]} \frac{\text{VALUE}(x, \lambda, g) + \text{VALUE}(x, \lambda, \bar{g}_\delta) - \text{VALUE}_j(x, \lambda, g)}{\text{OPT}(\lambda, g) + \text{OPT}(\lambda, \bar{g}_\delta) - \text{OPT}(\lambda, g)} && (49) \\ &\leq \min_{\delta \in [-\varepsilon, \varepsilon]} \frac{\text{VALUE}(x, \lambda, g) + \sum_{i \in J} \lambda_i \delta}{\text{OPT}(\lambda, g) + \sum_{i \in K_2} \lambda_i \delta + \frac{|K_3|}{|K_1|} \sum_{i \in K_1} \lambda_i u(\delta/\phi)} && (50) \\ &\leq \frac{\text{VALUE}(x, \lambda, g)}{\text{OPT}(\lambda, g)}. && (51) \end{aligned}$$

The expression in (50) is monotonic in δ. If (50) is increasing (decreasing) in δ we set $\delta = -\varepsilon$ ($\delta = \varepsilon$) to obtain a function with lower benefit ratio, contradicting g is minimizing. In case inequality (51) is not tight, i.e., expression (50) is constant in δ, we can increase β while keeping the ratio the same, contradicting g is the minimizing function with largest β. It follows any worst-case function can be assumed to have $\beta = 0$. □

The main result now follows:

Theorem 1. *Consider service j with n-dimensional demand vector λ_j and benefit function f_j, satisfying Assumption 1. Let non-negative scalars ℓ_j and u_j be such that* $\min\{\ell_j z, 1\} \leq f(z) \leq \min\{u_j z, 1\}$*. It holds that:*

$$\text{RATIO}(x, \lambda_j, f_j) \geq \text{BOUND}(x, \lambda_j, \ell_j, u_j), \tag{8}$$

where :

$$\text{BOUND}(x, \lambda_j, \ell_j, u_j) = \min_{\alpha \in [\ell_j, u_j]} \frac{\sum_{i=1}^{n} \lambda_{ij} \min\{\alpha x_i, 1\}}{\text{SUM}(\lambda_j, \alpha)}. \tag{9}$$

Furthermore, this bound is tight for all x.

Proof. From Lemma 2 it follows the family of worst-case functions is given by $g(z) = \min\{\alpha z, 1\}$ for $\alpha \in [\ell, u]$. For these functions, it is readily seen that:

$$\text{OPT}(\lambda, g) = \text{SUM}(\lambda, \alpha). \tag{52}$$

Taking into account the different bounds for the services $j = 1, \ldots, m$, the final result follows. Tightness follows from noting that the lower bound equals the ratio for any function of the form $\min\{\alpha z, 1\}$. □

Corollary 1 directly follows from Theorem 1 by setting $\ell = 1$ and $u = n$.

B.2 Proof Theorem 2

Theorem 2. *Consider m services, for $j = 1, \ldots, m$, with n-dimensional demand vector λ_j proportional to λ, and benefit function f_j, satisfying Assumption 1. Assume λ satisfies $\lambda_i = \mu$ for the first k elements and $\lambda_i = 1$ for the remaining $n - k$ elements. The dominating capacity, denoted by $\hat{\gamma}(n, k, \mu)$, is bounded by:*

$$\hat{\gamma}(n, k, \mu) \leq 1 + \left(1 - \frac{k}{n}\right)\left(1 - \frac{1}{\mu}\right). \tag{13}$$

Furthermore, this bound is tight.

Proof. From Corollary 1, we have:

$$\text{RATIO}(x, \lambda_j, f_j) \geq \text{BOUND}(x, \lambda_j) = \min_{\alpha \in [1,n]} \frac{\sum_{i=1}^{n} \lambda_i \min\{\alpha x_i, 1\}}{\text{SUM}(\lambda, \alpha)}, \tag{53}$$

where in the last equality we use that λ_j is proportional to λ. We can therefore focus on the minimization problem in (53).

Clearly, any worst-case optimal allocation will give the same amount of resources to the first k and remaining $n-k$ locations. Let r denote the amount of resources allocated to the first k locations and q to the second, with $r + q \leq \gamma$. For given α, the expression for the right-hand side is given by:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{k}, 1\right\} + (n-k)\min\left\{\frac{\alpha q}{n-k}, 1\right\}}{\alpha\lambda}, \tag{54}$$

when $\alpha < k$, and:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{k}, 1\right\} + (n-k)\min\left\{\frac{\alpha q}{n-k}, 1\right\}}{k\lambda + \alpha - k}, \tag{55}$$

when $\alpha \geq k$.

First, consider (54). Clearly, this expression is non-decreasing in α. Hence, we get:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{k}, 1\right\} + (n-k)\min\left\{\frac{\alpha q}{n-k}, 1\right\}}{\alpha\lambda} \geq \frac{k\lambda \min\{r, 1\} + (n-k)\min\left\{\frac{kq}{n-k}, 1\right\}}{k\lambda}, \tag{56}$$

for all $\alpha \leq k$.

Next, consider (55). Clearly $r/k \geq q/(n-k)$ for any worst-case optimal allocation. Consider $\alpha \leq k/r$. If $r \geq 1$, we get expression (56). Otherwise, consider $\alpha > k$ and $\alpha \leq k/r$. We get:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{k}, 1\right\} + (n-k)\min\left\{\frac{\alpha q}{n-k}, 1\right\}}{k\lambda + \alpha - k} = \frac{\alpha\lambda r + \alpha q}{k\lambda + \alpha - k}. \tag{57}$$

This expression is readily seen to be non-decreasing in α, hence the lowest is achieved for $\alpha = k$, giving again expression (56).

Finally, consider the case $\alpha > \max\{k, k/r\}$. We get the expression:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{k}, 1\right\} + (n-k)\min\left\{\frac{\alpha q}{n-k}, 1\right\}}{k\lambda + \alpha - k} = \frac{k\lambda + (n-k)\min\left\{\frac{\alpha q}{n-k}, 1\right\}}{k\lambda + \alpha - k} \tag{58}$$

This expression is readily shown to be non-increasing for $\alpha \geq (n-k)/q$, and monotonic for $\alpha < (n-k)/q$. It follows the expression attains it minimum over $\alpha \geq \max\{k, k/r\}$ at either $\alpha = \max\{k, k/r\}$ or $\alpha = n$. Combining all results, we obtain that the minimization problem in (53) is equivalent to a minimization problem over two affine functions:

$$\min \left\{ \frac{k\lambda \min\{r, 1\} + (n-k) \min\left\{\frac{kq}{n-k}, 1\right\}}{k\lambda}, \frac{k\lambda + (n-k)\min\left\{\frac{nq}{n-k}, 1\right\}}{k\lambda + n - k} \right\}. \tag{59}$$

It remains to determine the r value maximizing this bound. Clearly, it is without loss of optimality to set $r \leq 1$ and $q \leq (n-k)/k$. This simplifies the expression to:

$$\min \left\{ \frac{\lambda r + q}{\lambda}, \frac{k\lambda + (n-k)\min\left\{\frac{nq}{n-k}, 1\right\}}{k\lambda + n - k} \right\}. \tag{60}$$

Suppose:

$$\frac{\lambda - \gamma}{\lambda - 1} \leq \gamma - 1 + \frac{k}{n}. \tag{61}$$

In this case, we can set $r = \min\{\gamma - 1 + \frac{k}{n}, 1\}$ and $q = (n-k)/n$ to obtain a fraction of one while satisfying $r + q \leq \gamma$. Otherwise, we have $q \leq (n-k)/n$ and $q = \gamma - r$ and can solve:

$$\min \left\{ \frac{\lambda r + \gamma - r}{\lambda}, \frac{k\lambda + n(\gamma - r)}{k\lambda + n - k} \right\}. \tag{62}$$

for r to obtain:

$$r = \frac{k\lambda^2 + \gamma(\lambda - 1)(n-k)}{k\lambda^2 + (2\lambda - 1)(n-k)}, \tag{63}$$

which gives the expression for the ratio:

$$\frac{\mu(n\gamma + k\mu - k)}{k(\mu - 1)^2 + n(2 - \mu)}. \tag{64}$$

The result now follows by noting that expression (61) is equivalent to:

$$\gamma \geq 1 + \left(1 - \frac{k}{n}\right)\left(1 - \frac{1}{\lambda}\right). \tag{65}$$

□

B.3 Proof Theorem 3

Theorem 3. *Consider m services, for $j = 1, \ldots, m$, with n-dimensional demand vectors λ_j such that $\lambda_{ij} = \mu$ for k elements and $\lambda_{ij} = 1$ for the remaining $n-k$ elements, and benefit function f_j, satisfying Assumption 1. Let h indicate the number of locations with demand μ for at least one service. The dominating capacity, denoted by $\hat{\gamma}(n, k, h, \mu)$, is bounded by:*

$$\hat{\gamma}(n, k, h, \mu) \leq 1 + \frac{h}{n} \frac{(n-k)(\mu - 1)}{k(\mu - 1) + h}. \tag{16}$$

This bound is tight when $h = k$ or $h = \min\{mk, n\}$.

Proof. We proceed similar to the proof of Theorem 2, now taking limited overlap into account.

We consider an allocation that allocates r resources to all the h allocations that have high demand for some j. Let q denote the amount allocated to the remaining $n-h$ locations, with $r + q \leq \gamma$. For given α and j, the expression for the minimization problem on the right-hand side in (53) is given by:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{h}, 1\right\} + (h-k)\min\left\{\frac{\alpha r}{h}, 1\right\} + (n-h)\min\left\{\frac{\alpha q}{n-h}, 1\right\}}{\alpha\lambda}, \tag{66}$$

when $\alpha < k$, and:

$$\frac{k\lambda \min\left\{\frac{\alpha r}{h}, 1\right\} + (h-k)\min\left\{\frac{\alpha r}{h}, 1\right\} + (n-h)\min\left\{\frac{\alpha q}{n-h}, 1\right\}}{k\lambda + \alpha - k}, \tag{67}$$

when $\alpha \geq k$.

First, consider (66). Clearly, this expression is non-decreasing in α. Hence, we get it is bounded by:

$$\frac{k\lambda \min\left\{\frac{kr}{h}, 1\right\} + (h-k)\min\left\{\frac{kr}{h}, 1\right\} + (n-h)\min\left\{\frac{kq}{n-h}, 1\right\}}{k\lambda}, \tag{68}$$

for all $\alpha \leq k$.

Next, consider (67). Clearly $r/h \geq q/(n-h)$ for any ratio-maximizing allocation. Consider $\alpha \leq h/r$. If $r \geq h/k$, we get expression (68). Otherwise, consider $\alpha > k$ and $\alpha \leq h/r$. We get that the expression is equal to:

$$\frac{\alpha\lambda r(k/h) + \alpha r(f-h)/f + \alpha q}{k\lambda + \alpha - k}. \tag{69}$$

This expression is readily seen to be non-decreasing in α, hence the lowest is achieved for $\alpha = k$, giving again expression (68).

Finally, consider the case $\alpha > \max\{k, h/r\}$. We get the expression:

$$\frac{k\lambda + h - k + (n-k)\min\left\{\frac{\alpha q}{n-h}, 1\right\}}{k\lambda + \alpha - k} \tag{70}$$

This expression is readily shown to be non-increasing for $\alpha \geq (n-h)/q$, and monotonic for $\alpha < (n-h)/q$. It follows the expression attains it minimum over $\alpha \geq \max\{k, h/r\}$ at either $\alpha = \max\{k, h/r\}$ or $\alpha = n$. Combining all results, we obtain that minimization problem in (53) is equivalent to:

$$\min\left\{\frac{k\lambda \min\left\{\frac{kr}{h}, 1\right\} + (h-k)\min\left\{\frac{kr}{h}, 1\right\} + (n-h)\min\left\{\frac{kq}{n-h}, 1\right\}}{k\lambda}, \frac{k\lambda + h - k + (n-k)\min\left\{\frac{nq}{n-h}, 1\right\}}{k\lambda + n - k}\right\}. \tag{71}$$

It remains to determine the r value maximizing this bound. Clearly, it is without loss of optimality to set $r \leq h/k$ and $q \leq (n-h)/k$. This simplifies the expression to:

$$\min \left\{ \frac{\lambda r(k/h) + r(h-k)/h + q}{\lambda}, \frac{k\lambda + h - k + (n-h)\min\left\{\frac{nq}{n-h}, 1\right\}}{k\lambda + n - k} \right\}. \tag{72}$$

Suppose:

$$\frac{h}{k}\frac{\lambda - \gamma}{\lambda - 1} \leq \gamma - 1 + \frac{h}{n}. \tag{73}$$

In this case, we can set $r = \min\{\gamma - 1 + \frac{h}{n}, 1\}$ and $q = (n-h)/n$ to obtain a fraction of one while satisfying $r + q \leq \gamma$. Otherwise, we can solve for r, similar to the proof of Theorem 2. The result now follows by noting that expression (73) is equivalent to:

$$\gamma \geq 1 + \frac{h}{n}\frac{(n-k)(\lambda-1)}{k(\lambda-1)+h} \tag{74}$$

Note that if $h = k$, this reduced to the bound of Theorem 2. Tightness follows by noting that the assumption of allocation resources uniformly over the h locations is clearly optimal for the worst-case allocation when $h = k$ or $h = \min\{mk, n\}$. □

References

A. Alban, P. Blaettchen, H. de Vries, and L. N. Van Wassenhove. Resource allocation with sigmoidal demands: Mobile healthcare units and service adoption. *Manufacturing & Service Operations Management*, 2021.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*, volume 28. Princeton university press, 2009.

D. Bertsimas, V. F. Farias, and N. Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.

D. Bertsimas, V. F. Farias, and N. Trichakis. On the efficiency-fairness trade-off. *Management Science*, 58(12):2234–2250, 2012.

M. Besiou and L. N Van Wassenhove. Humanitarian operations: A world of opportunity for relevant and impactful research. *Manufacturing & Service Operations Management*, 22(1):135–145, 2020.

T. Breugem and L. N. Van Wassenhove. The price of imposing vertical equity through asymmetric outcome constraints. *Management Science*, 2022.

T. Breugem, T. Dollevoet, and D. Huisman. Is equality always desirable? analyzing the trade-off between fairness and attractiveness in crew rostering. *Management Science*, 68(4):2619–2641, 2022.

I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou. The efficiency of fair division. *Theory of Computing Systems*, 50(4):589–610, 2012.

Y. Chen, N. Marković, I. O. Ryzhov, and P. Schonfeld. Data-driven robust resource allocation with monotonic cost functions. *Operations Research*, 70(1):73–94, 2022.

H. de Vries, L. E. Swinkels, and L. N. Van Wassenhove. Site visit frequency policies for mobile family planning services. *Production and Operations Management*, 30(12):4522–4540, 2021a.

H. de Vries, J. van de Klundert, and A. Wagelmans. Toward elimination of infectious diseases with mobile screening teams: Hat in the drc. *Production and Operations Management*, 30(10):3408–3428, 2021b.

S. Deo and M. Sohoni. Optimal decentralization of early infant diagnosis of hiv in resource-limited settings. *Manufacturing & Service Operations Management*, 17(2):191–207, 2015.

S. Deo, S. Iravani, T. Jiang, K. Smilowitz, and S. Samuelson. Improving health outcomes through better capacity allocation in a community-based chronic care model. *Operations Research*, 61(6):1277–1294, 2013.

S. Deo, K. Rajaram, S. Rath, U. S. Karmarkar, and M. B. Goetz. Planning for hiv screening, testing, and care at the veterans health administration. *Operations research*, 63(2):287–304, 2015.

A. Goel and D. Estrin. Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk. *Algorithmica*, 43(1-2):5–15, 2005.

A. Goel and A. Meyerson. Simultaneous optimization via approximate majorization for concave profits or convex costs. *Algorithmica*, 44(4):301–323, 2006.

B. L. Gorissen, İ. Yanıkoğlu, and D. den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.

Y. Gur, D. Iancu, and X. Warnes. Value loss in allocation systems with provider guarantees. *Management Science*, 67(6):3757–3784, 2021.

J. O. Jónasson, S. Deo, and J. Gallien. Improving hiv early infant diagnosis supply chains in sub-saharan africa: Models and application to mozambique. *Operations Research*, 65(6):1479–1493, 2017.

J. O. Jónasson, K. Ramdas, and A. Sungu. Social impact operations at the global base of the pyramid. *Production and Operations Management*, 2022.

J. McCoy and H. Lee. Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management*, 23(6):965–977, 2014.

J. H. McCoy and E. M. Johnson. Clinic capacity management: Planning treatment programs that incorporate adherence. *Production and Operations Management*, 23(1):1–18, 2014.