Resource Allocation with Sigmoidal Demands: Mobile Healthcare Units and Service Adoption

Andres Alban, Philippe Blaettchen, Luk N. Van Wassenhove
Technology and Operations Management area, INSEAD, Fontainebleau, France 77305

Harwin de Vries
Department of Technology and Operations Management, Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands 3062PA

Achieving broad access to health services (a target within the sustainable development goals) requires reaching rural populations. Mobile healthcare units (MHUs) visit remote sites to offer health services to these populations. However, limited exposure, health literacy, and trust can lead to sigmoidal (S-shaped) adoption dynamics, presenting a difficult obstacle in allocating limited MHU resources. It is tempting to allocate resources in line with current demand, as seen in practice. However, to maximize access in the long term, this may be far from optimal, and insights into allocation decisions are limited.

We present a formal model of the allocation of MHU resources, i.e., the frequency of visits to each site, to maximize long-term uptake of preventative health services. We formulate the problem as the optimization of a sum of sigmoidal functions. While the problem is NP-hard, we provide closed-form solutions to particular cases of the model that elucidate insights into the optimal allocation. For example, more visits should generally be allocated to sites where the cumulative demand potential is higher and, counterintuitively, often those where demand is currently lower. To apply our insights in practice, we propose a practical method for estimating our model's parameters from pre-existing data. Our estimation approach achieves better predictions than standard methods. Finally, we demonstrate the potential of our approach by applying our methods to family planning MHUs in Uganda. In particular, we show that operationalizable heuristic allocations, grounded in our insights, outperform allocations based on current demand.

Key words: humanitarian logistics; diffusion of innovations; sigmoidal programming; gradient boosting

History: Published in *Manufacturing & Service Operations Management*: doi.org/10.1287/msom.2021.1020.

1. Introduction

Achieving broad and affordable access to health services is necessary to fulfill the Sustainable Development Goals (SDGs) (Keskinocak and Savva 2020). Currently, half the world's population lacks access to essential services (WHO and The World Bank 2017), even though universal health coverage is a target of *SDG 3: Good Health and Well-Being*. This inaccessibility can have important negative consequences for other SDGs by reducing productivity (Stenberg et al. 2014) and educational attainment (Jukes et al. 2007). Broadening access to health services requires strengthening delivery to rural areas, where approximately half of humanity lives (The World Bank 2018), but rural populations' access tends to be significantly worse than urban populations', leading to stark differences in health outcomes, especially in developing countries (Strasser et al. 2016). Aside from physical barriers—geographical isolation (Crouse et al. 2010), a lack of affordable transportation (Hodgson et al. 1998), the concentration of resources within cities (Strasser 2003)—lower educational levels (van Maarseveen 2020) and health literacy (Aljassim and Ostini 2020), and a lack of exposure to and trust in services and providers (Wickstrom et al. 2013) effectively reduce rural populations' access to health services. We summarize the latter as informational barriers to access.

In this paper, we analyze the resource allocation problem faced by humanitarian organizations attempting to broaden access of rural populations to health services with the help of mobile healthcare units (MHUs). Also referred to as outreach teams, MHUs establish temporary service delivery sites near rural populations. They can provide a resource-efficient alternative to stationary clinics and hospitals that may be prohibitively costly due to low population density (Doerner et al. 2007). However, even if MHUs can overcome physical barriers to access, infrequent contact with health services and service providers does not necessarily alleviate the informational barriers. This is particularly true in the case of health services for which the need is not immediately obvious, such as preventative care, as opposed to evident needs such as emergency care (Berkman et al. 2011). Nonetheless, in the absence of stationary providers, MHUs play an active role also in delivering services with high informational barriers, such as vaccination (Vaahtera et al. 2000), cancer screening (Mauad et al. 2011), HIV counseling (Mabuto et al. 2014) and family planning (Wickstrom et al. 2013). The gradual spread of information through word-of-mouth within rural communities driven by previous adopters can overcome informational barriers and enhance adoption (Neke et al. 2018). MHUs may support it by actively building trust (Aung et al. 2015).

Humanitarian organizations that allocate MHU resources to improve access to such services face two interrelated challenges: (i) understanding the complex adoption dynamics driven by word-of-mouth effects, often in an environment with limited data and poor computational infrastructure (Ergun et al. 2014, Besiou and Van Wassenhove 2020), and (ii) strategically allocating scarce resources, taking into account slow build-up for adoption demand and eventual saturation. Moreover, in the humanitarian context, planning solutions need to be interpretable to verify adherence to standards of practice and improve trust and applicability (Gralla and Goentzel 2018).

Our work aims at helping humanitarian organizations overcome these challenges. We provide a formal model for allocating available MHU resources for organizations aiming to broaden access to health services whose adoption is hindered by informational barriers. We adapt the Bass (1969) model to reflect the resulting adoption dynamics and predict the adoption levels engendered by potential allocations. We choose this model for its flexibility, interpretability, and established performance in practice. Our allocation model maximizes adoption as derived from the Bass model, rather than incorporating also returning clients. This choice is appropriate when informational barriers have a distinct impact on initial adoption, and intervals between subsequent service demands are long enough that repeat clients can be served effectively with a baseline visit frequency, as in the examples above. Moreover, we abstract from the shorter-term scheduling of MHUs to focus on the complexities introduced by long-term changes in adoption patterns. We identify solutions that enable organizations to improve resource allocations in two ways. First, we derive operationalizable insights that can help guide long-term decision-making, especially in environments with little data. For example, counter to intuition, more resources should be devoted to those sites within a prioritized group with lower current adopter demand to anticipate saturation effects. Second, we provide interpretable and implementable procedures to predict adopter demands and to allocate MHU resources by relying on an approach with low requirements for data collection and computation.

Through a collaboration with MSI Reproductive Choices (henceforth MSI), we have had access to data, documentation, and interactions with experts on the operations of family planning MHUs. MSI is a non-governmental organization that provides family planning services in 37 countries across the world. We consider the problem of planning 25 MHUs that MSI employs to serve rural populations in Uganda. We apply our model to this setting and evaluate the benefits of our insights compared to current practice. MSI is an ideal testing ground for our research because family planning services are a prime example of the use of MHUs and play a central role in the sustainable development agenda. In particular, unwanted pregnancies negatively impact the health of mothers and children (Conde-Agudelo et al. 2012), making family planning essential to achieve *SDG 3: Good Health and Well-Being*. It is estimated that each dollar spent on improving access to modern contraceptive services saves \$1.4 in maternal and child health care (Singh and Darroch 2012). Family planning services are also essential for progress towards other SDGs (Starbird et al. 2016). For example, adolescent pregnancies have large lifetime opportunity costs in the form of foregone work and education opportunities (Narita and Diaz 2016). Hence, they can form an obstacle to achieving *SDG 1: No Poverty* and *SDG 4: Quality Education*. The ability to freely choose whether and at what age to be pregnant is essential to empower women (Plambeck and Ramdas 2020). Thus, *SDG 5: Gender Equality* explicitly requires universal access to sexual and reproductive health.

The remainder of the paper is structured as follows. In Section 2, we discuss how our work contributes to the operations management literature. In Section 3, we develop a model to optimally allocate MHU resources to sites, given complex adoption dynamics at each site. We analyze this model in Section 4 before developing a procedure to predict demand in Section 5. We exemplify our approach in Section 6 with the case of MSI Uganda before concluding in Section 7.

2. Literature review

We contribute to three streams of literature: (i) team deployment in the humanitarian sector, (ii) diffusion of innovations, and (iii) resource allocation with sigmoidal objectives.

2.1. Team deployment in the humanitarian sector

The routing and scheduling of MHU visits in developing countries have been studied as part of the vast literature on vehicle routing and covering tour problems (see, e.g., Laporte 1992), but with modified objectives for the humanitarian context (Hodgson et al. 1998, Doerner et al. 2007). A similar problem also arises in the disaster relief literature (de la Torre et al. 2012, Ferrer et al. 2018). These works are concerned with finding the shortest path over a network of sites while satisfying constraints on the visited sites. While such decisions are relevant for planning visits in the short-term, we focus on the long-term resource allocation to a large number of sites.

In this vein, McCoy and Lee (2014) describe a problem in which resources, in the form of visit capacities, are allocated to sites with the objective of providing equitable and efficient service, using the concept of α-fairness. Our objective function allows us to incorporate a measure of fairness by weighing sites differently. We also include lower bounds on the allocated resources to guarantee a certain level of service. de Vries et al. (2021b) proposes a model for epidemic control where limited resources are allocated to screening the populations at different sites. Their goal is to minimize disease transmission, which means that demand follows different dynamics than the adoption of health services we consider. Most closely related to our work, de Vries et al. (2021a) propose a model for the resource allocation problem of family planning MHUs. Their objective is to maximize the number of clients reached during visits, including new adopters and existing clients, assuming that demand depends on visit frequencies. They observe in their data set that demand has a concave (logarithmic or square root) relationship with the time between two visits. As in de Vries et al. (2021a), we build simplified policies based on our model's optimal solution, considering that a limited number of different frequency choices is more practical. Unlike the authors, however, we focus on the longer-term objective of broadly enabling health service adoption, taking into account that adoption follows complex dynamics driven by word-of-mouth and saturation effects. Based on our model, we are able to derive novel managerial insights into the resource allocation problem for organizations that are concerned foremost with broadening access. While health service access for existing clients is important, we ensure a sufficient level of service through a constraint. We do not consider the effects of visit frequencies on demand, as we observe that these are weaker for adopters than existing clients—existing client demand may accumulate when team visits are infrequent. Also in the context of family planning services, Van Rijn et al. (2020) consider assigning MHUs to sets of sites. In particular, they analyze the performance of decentralized approaches to this assignment problem. In this paper, we focus on allocating resources in the form of visits to sites. This planning task can occur either globally, before sites are assigned to mobile units, or locally, to plan individual MHUs' resource allocations.

2.2. Diffusion of innovations

Diffusion of innovations is a large area of research within the behavioral sciences (Rogers 2010). The Bass (1969) model was developed to quantify the behavioral concepts in the diffusion of durable goods. It relies on three parameters with a clear interpretation: innovation coefficient, imitation coefficient, and market potential. The result is a sigmoidal (S-shaped) function of the cumulative number of adopters of an innovation over time, henceforth termed Bass function. In addition to simplicity and interpretability, the Bass model has been shown to perform well in forecasting sales of innovations (cf. Bass 1980). Therefore, we implement the Bass model to capture the demand for MHU services, where each visit to a site corresponds to a step forward in the Bass function.

To estimate the model's parameters, Bass (1969) proposed ordinary least squares. The resulting estimates may be biased and unstable, so several other procedures have been proposed, including non-linear least-squares and maximum likelihood methods (Meade and Islam 2006). These techniques also tend to suffer from estimation issues, mainly when the inflection point of cumulative demand has not yet been observed (Boswijk and Franses 2005). This problem comes up, e.g., when sales in new markets or new products need to be estimated with little data. Data from old markets is only an imperfect proxy because the diffusion dynamics vary due to differing product characteristics, economic and cultural conditions, and marketing measures (Peres et al. 2010).

Three general approaches use data from previous markets to predict adoption dynamics in new markets: (i) a hierarchical Bayesian approach implements priors based on previous markets and updates as sales data is observed (e.g., Lenk and Rao 1990); (ii) a comparative approach estimates coefficients for markets with a long history and then makes predictions based on market characteristics (e.g., Lee et al. 2014); and (iii) a regression approach replaces coefficients with functions of product, market, and marketing characteristics within the estimation process and can make predictions for new markets solely based on characteristics observed early on (e.g., Gatignon et al. 1989). We estimate demands at geographically isolated sites, which relates to estimating demand for the same product in independent markets with similar characteristics. Both the first and second approaches require some sites with a long demand history. In our case, we have many sites with few observations that can, nonetheless, provide valuable information, rendering the third approach ideal. We extend this approach with machine learning tools.

2.3. Resource allocation with sigmoidal objective

We formulate the resource allocation problem as a knapsack problem with a sum of Bass functions as the objective, each representing the demand for health service adoption at a site. We analyze the more general case where the objective is a sum of sigmoidal functions. Several solution procedures have been proposed in the literature. Ağrali and Geunes (2009) present a dynamic programming approach that finds an approximate solution in pseudo-polynomial time. Srivastava and Bullo (2014) propose an algorithm that uses Lagrangian relaxation to find a solution in polynomial time within a constant factor of the optimal solution. Udell and Boyd (2014) propose a branch and bound algorithm that is guaranteed to converge to the optimal solution, albeit in exponential time. We do not aim at improving algorithmic solutions but rather at providing structural results and insights that help organizations allocate their MHU resources. In doing so, we consider simplified versions of the general model by adding additional assumptions. Ginsberg (1974) provides results for the case when all sigmoidal functions are the same. We provide stronger results by making the additional assumption that the sigmoidal functions are point-symmetric around the inflection point (an assumption satisfied by the Bass function and many other functions used in practice). Moreover, we provide results when the demand functions are all shifted forms of the same sigmoidal function and derive insight into more complicated settings using comparative statics.

3. A model of resource allocation under sigmoidal demands

We now formalize the resource allocation problem of an organization that provides well-defined health services and deploys mobile healthcare units to reach clients in remote areas. Let u denote the number of MHUs that visit the set of sites $N = \{1, \ldots, n\}$. The organization decides how many visits each site receives during a planning horizon of H units of time, denoted with $\pi = (\pi_i)_{i \in N}$. To generate insights with our model, in line with McCoy and Lee (2014) and de Vries et al. (2021b), we do not constrain π_i to integer values, nor do we prescribe the exact scheduling and routing.

As outlined in the introduction, we assume that the organization's ultimate objective is to broaden access to its health services, expressed by the number of new service adopters during a visit. To capture the total demand of new adopters during all visits to a site $i \in N$ during the planning horizon, we use the function $\psi_i(\pi_i)$. We will discuss the dynamics underlying this function in the next subsection. Because the impact of providing access may differ between sites, we use $\beta_i \in [0, 1]$ to weight adopter demands. For example, MSI generates more impact when serving younger and poorer populations. Thus, the objective is to maximize *effective adopter demand* $\Psi(\pi) = \sum_{i \in N} \beta_i \psi_i(\pi_i)$, subject to a number of constraints, which we outline next.

First, the total number of visits is limited by available MHU capacity. We assume the total visit time, including travel, is constant across sites and normalize this to one unit of time. For example, in the context of the application we discuss in Section 6, MHUs usually visit one site per day. With this normalization, $\Pi \stackrel{\text{Def.}}{=} uH$ is the total capacity. The organization may also have a limited budget for performing its services. We assume such a constraint is not binding because the costs of running an MHU (e.g., labor, vehicles) are comparatively high. Hence, budget constraints affect the number of teams rather than the number of clients served by each team and are reflected in Π.

Our discussions with MSI support this assumption. Second, a site can only be visited effectively by one MHU at any given time, meaning that the planning horizon H limits the number of visits. Other constraints arise due to the need to prepare and announce visits. An upper bound on the number of visits T, with $T \le H$, allows us to combine these considerations. Finally, the organization may impose a minimum number of visits to each site F to fulfill ethical and medical requirements regarding the repetition or discontinuation of its services. The organization solves

$$\max_{\pi} \sum_{i \in N} \beta_i \psi_i(\pi_i) \tag{1a}$$

$$\text{s.t.} \quad \sum_{i \in N} \pi_i \le \Pi \tag{1b}$$

$$F \le \pi_i \le T, \ \forall i \in N. \tag{1c}$$

In practice, sites may be assigned to specific MHUs. This can be captured by optimizing each mobile unit's allocation individually, defining N as the set of sites served by the MHU.

3.1. Specifying the adoption dynamics—the Bass diffusion model

The adoption of health services provided by MHUs follows complex dynamics. Potential clients may be unaware of a service, unconvinced of its benefits, or mistrust the provider. The community can then be an important driver of adoption by showcasing the service's benefits and safety.

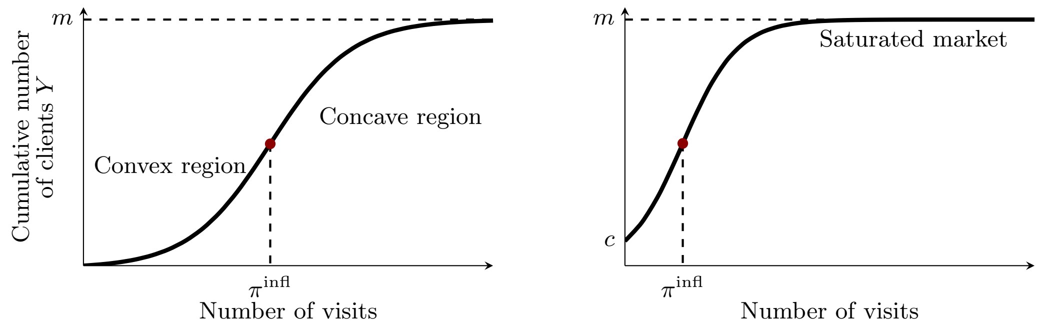
The Bass (1969) model captures exactly such dynamics while being sufficiently flexible to represent demand in many practical applications. The model has three parameters: (i) the innovation coefficient reflects the share of *innovators* in the population who demand a product or service even when nobody else uses it; (ii) the imitation coefficient reflects the share of *imitators* who demand a product or service only if they observe others using it (when the innovation coefficient is low, but the imitation coefficient is high, word-of-mouth effects play a major role in building demand); (iii) the market potential signifies the level at which demand *saturates*.

We assume that adopter demand faced by MHUs at site i follows the diffusion process defined by the Bass model. Specifically, after π_i visits to the site, the cumulative number of clients is

$$Y_i(\pi_i) = m_i \frac{1 - \theta_i e^{-(p_i+q_i)\pi_i}}{1 + \frac{q_i}{p_i}\theta_i e^{-(p_i+q_i)\pi_i}}. \tag{2}$$

Here, $p_i \ge 0$ and $q_i \ge 0$ are the innovation and imitation coefficients, respectively, in units of 1/visit. The market potential, the maximum cumulative number of clients, is denoted by $m_i > 0$, i.e., $\lim_{\pi_i \to \infty} Y_i(\pi_i) = m_i$. The parameter $\theta_i = p_i(m_i - c_i)/(p_i m_i + q_i c_i)$ is used to simplify notation, where $c_i = Y_i(0) \ge 0$ denotes the cumulative number of clients at the start of the planning horizon. If $c_i = 0$, then $\theta_i = 1$ so that (2) simplifies to the exact expression in Bass (1969). The cumulative number of clients in the Bass diffusion model is exemplified in Figure 1.

Figure 1 **The Bass diffusion model**



(a) No existing clients (b) Existing clients, higher q

While the Bass model was derived for the adoption of durable goods in continuous time, clients in the present context can only receive services when an MHU visits. Thus, we discretize time into intervals of unit length. Taking a long-term view on allocating resources, we do not prescribe an exact visit schedule. Hence, we assume that a site's adopters are given by the next step in the discretized Bass function whenever a team visits. Implicit here is that inter-arrival times do not affect adopter demand. de Vries et al. (2021a) indicate that client demand is an increasing and concave function of inter-arrival time. However, the need to prepare visits lower-bounds the inter-arrival time and limits this effect. As a result, adopter demand during visit $k \in \mathbb{N}$ is $Y_i(k) - Y_i(k-1)$, and total demand during the planning horizon (the objective function) can be rewritten as $\psi_i(\pi_i) = Y_i(\pi_i) - c_i$. Alternatively, one can think of π_i as the cumulative marketing effort in a generalized Bass model (Bass et al. 1994), which would lead to the same expression for the demand function.

The above model assumes a fixed market potential m_i. In practice, however, the long-run market potential may be affected by population growth. If the population growth rates are similar for all sites, the effect on the allocation is negligible. Appendix E provides evidence that MSI's sites in Uganda have similar growth rates.

Finally, we point out that the Bass (1969) model assumes no repeated services and captures only the demand generated by new adopters of the service. Existing clients may require service again from time to time. While this impacts the number of clients served, it does not affect the percentage of the population that serves as a driver for word-of-mouth effects and, thus, the adoption by new clients. Constraint (1c) ensures that sites receive a minimum number of visits. If this is sufficient for existing clients to obtain services at a reasonable frequency, removing their demands from consideration is in line with the stated objective of broadening access to health services.

With the above considerations, we can reformulate the organization's problem as follows:

$$\begin{aligned} \max_{\pi} \quad & \sum_{i \in N} \beta_i (Y_i(\pi_i) - c_i) \\ \text{s.t.} \quad & \text{(1b) and (1c).} \end{aligned} \tag{3}$$

4. Optimal allocation of visits to sites

In this section, we analyze Problem (1) to characterize the optimal allocation. We consider special cases in Sections 4.1 and 4.2 that allow us to gain intuition about the optimal allocation and provide actionable insights for the decision-maker. Section 4.3 provides an algorithm to approximate the solution to (1) in the general case. In Section 4.4, we present comparative statics results that yield further understanding of how the optimal allocation depends on the adoption dynamics and site composition. We summarize our insights in Section 4.5. All proofs are found in Appendix B.

While our focus is on the Bass function, we present results for a more general set of demand functions that includes the Bass function as a special case. Namely, we assume throughout the analysis that adopter demand fulfills Assumption 1:

Assumption 1 **(Sigmoidal demand function)**. *The demand function $\psi_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ for any site is twice continuously differentiable, strictly increasing, and there is a unique inflection point $\pi_i^{infl} \geq 0$, such that $\psi_i''(\pi_i) > 0$ for $\pi_i < \pi_i^{infl}$, and $\psi_i''(\pi_i) < 0$ for $\pi_i > \pi_i^{infl}$.*

Before moving to the main results, we show in Proposition 1 that the Bass function in (2) satisfies Assumption 1. The proposition further shows that the Bass function is point-symmetric around the inflection point, a property that we will assume for certain results.

Proposition 1. *The function $\tilde{\psi}_i(\pi_i) \overset{Def.}{=} Y_i(\pi_i) - c_i$ is twice continuously differentiable, strictly increasing, and has the unique inflection point $\pi_i^{infl} = \max\{\log(\theta_i q_i/p_i)/(p_i + q_i),\ 0\}$ such that $\tilde{\psi}_i''(\pi_i) > 0$ if $\pi_i < \pi_i^{infl}$ and $\tilde{\psi}_i''(\pi_i) < 0$ if $\pi_i > \pi_i^{infl}$. In addition, $\tilde{\psi}_i(\pi_i)$ is point-symmetric around π_i^{infl}, that is $\tilde{\psi}_i(\pi_i^{infl} + \epsilon) - \tilde{\psi}_i(\pi_i^{infl}) = \tilde{\psi}_i(\pi_i^{infl}) - \tilde{\psi}_i(\pi_i^{infl} - \epsilon)$ $\forall \epsilon \in \left[0, \pi_i^{infl}\right]$.*

With the assumptions of our model in hand, we state Proposition 2, which formalizes a guarantee that all capacity is used in (1):

Proposition 2. *A maximizer $\pi^* = (\pi_i^*)_{i \in N}$ of (1) exists and either satisfies the capacity constraint with equality, $\sum_{i \in N} \pi_i^* = \Pi$, or all the upper bounds with equality, $\pi_i^* = T,\ \forall i \in N$.*

If the allocation $\pi_i = T,\ \forall i \in N$, is feasible, this also trivially constitutes the optimum. This is the case if and only if $nT \leq \Pi$. In the remainder, we assume $nT > \Pi$ and that (1b) is replaced by an equality constraint in (1), which is without loss of generality according to Proposition 2. Any optimal solution must then fulfill the following set of Karush-Kuhn-Tucker (KKT) conditions:

$$
\begin{aligned}
(\pi_i - T)(\pi_i - F)(\beta_i \psi_i'(\pi_i) - \lambda) = 0,\ \forall i \in N, \\
(\pi_i - T)(\beta_i \psi_i'(\pi_i) - \lambda) \geq 0,\ \forall i \in N, \\
(\pi_i - F)(\beta_i \psi_i'(\pi_i) - \lambda) \geq 0,\ \forall i \in N,
\end{aligned}
\tag{4}
$$

where λ is the Lagrange multiplier of the equality constraint. The KKT conditions imply that either $\pi_i = T$, $\pi_i = F$, or π_i satisfies

$$\beta_i \psi_i'(\pi_i) = \lambda. \tag{5}$$

For a given λ, there are at most two solutions to (5): one in the concave region ($\pi_i \geq \pi_i^{\text{infl}}$) and one in the convex region ($\pi_i \leq \pi_i^{\text{infl}}$). We use $\hat{\pi}_i(\lambda) \geq \pi^{\text{infl}}$ and $\tilde{\pi}_i(\lambda) \leq \pi^{\text{infl}}$ to denote the two potential solutions to (5) for a site i. If no solution exists, we define $\hat{\pi}_i(\lambda) = \pi^{\text{infl}}$ and $\tilde{\pi}_i(\lambda) = \pi^{\text{infl}}$, respectively.

For any π and λ, we thus denote four mutually exclusive sets, L, G, E_1, and E_2:

$$\begin{aligned}
L(\pi,\lambda) &= \{i \in N : \pi_i = F\},\\
G(\pi,\lambda) &= \{i \in N : \pi_i = T\},\\
E_1(\pi,\lambda) &= \{i \in N \backslash (L \cup G) : \pi_i = \hat{\pi}_i(\lambda)\}, \text{ and}\\
E_2(\pi,\lambda) &= \{i \in N \backslash (L \cup G \cup E_1) : \pi_i = \tilde{\pi}_i(\lambda)\}.
\end{aligned} \tag{6}$$

The sets $L(\pi,\lambda)$ and $G(\pi,\lambda)$ are the sites for which the inequality constraints (lower and upper bound, respectively) are tight. The sets $E_1(\pi,\lambda)$ and $E_2(\pi,\lambda)$ are the sites for which equation (5) is satisfied. For all sites $i \in E_1(\pi,\lambda)$, $\psi''(\pi_i) \leq 0$, and we say that i is in the concave region (see Figure 1a). Similarly, for all sites $i \in E_2(\pi,\lambda)$, $\psi''(\pi_i) > 0$, and we say that i is in the convex region.

Using the above first-order necessary conditions, as well as second-order necessary conditions introduced in the proof, we show that any optimal solution can be partitioned into these four sets, with some additional properties:

Proposition 3. *Let π^* be an optimal solution to (1). Then there exists a λ^*, as well as mutually exclusive sets $L^* = L(\pi^*,\lambda^*)$, $G^* = G(\pi^*,\lambda^*)$, $E_1^* = E_1(\pi^*,\lambda^*)$, and $E_2^* = E_2(\pi^*,\lambda^*)$, such that*

(a) $N = L^* \cup G^* \cup E_1^* \cup E_2^*$ (the sets are also exhaustive);

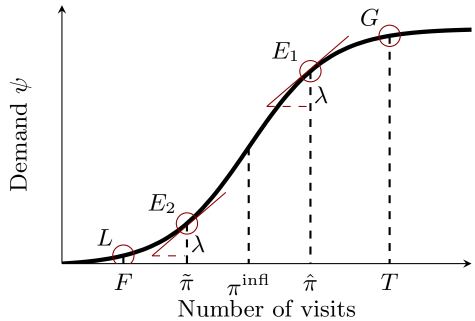
(b) if $i \in L^*$, then $\beta_i \psi_i'(F) \leq \lambda^*$;

(c) if $i \in G^*$, then $\beta_i \psi_i'(T) \geq \lambda^*$;

(d) $|E_2^*| \leq 1$ and, if $i \in E_2^*$ and $j \in E_1^*$, then $\beta_i \psi_i''(\pi_i^*) \leq -\beta_j \psi_j''(\pi_j^*)$.

We provide a complete proof in Appendix B. Parts (a)-(c) follow from the first-order conditions, which require that the derivatives of all sites in the interior (in E_1^* or E_2^*) are the same and equal to the Lagrange multiplier. The intuition for this result is the following. Assume, to the contrary, that two sites i and j are in the interior of the solution and have different derivatives. Say, for example, that the derivative is smaller for site i. Then one can obtain an improved feasible solution by adding a small enough ε to π_j and subtracting ε from π_i. Thus, the problem can be reduced to finding the optimal λ^* and deciding which sites belong to which set in (6). Figure 2 illustrates the four possible visit numbers of an optimal allocation for a fixed λ.

Figure 2 **The allocation of visits to sites for a given λ and the corresponding set memberships**



The assumption that ψ_i is continuously differentiable allows us to show that part (d) is a necessary condition, compared to Theorem 3 of Ağrali and Geunes (2009) that shows the existence of a solution with such a property when assuming Lipschitz continuity. Interestingly, the second-order conditions do not guarantee that E_2^* is empty, but instead that it contains at most one site. In Sections 4.1 and 4.2, we show that, under some additional conditions, there exists a solution where E_2^* is empty unless E_1^* is empty, and we use this fact to derive solutions to simplified models.

4.1. Optimal allocation for homogeneous sites

In this subsection, we assume that sites are homogeneous: $\psi_i = \psi$ and $\beta_i = \beta\ \forall i \in N$. This is a useful assumption when there is little historical data to go by, and populations around the different sites are not too varied in terms of their size, as well as economic and social conditions.

Ginsberg (1974) provides an analysis of this case. We additionally assume that the cumulative demand function ψ is point-symmetric, as applies to the Bass function (cf. Proposition 1). The additional assumption allows us to show stronger results:

Proposition 4. *Assume a homogeneous set of sites N with $\psi_i = \psi$ for all $i \in N$, and let ψ be point-symmetric around the inflection point π^{infl}. Moreover, assume a homogeneous weighting $\beta_i = \beta$. Then, it is optimal to either*

(i) select any x^* sites that are visited equally often with $\pi_i = F + (\Pi - Fn)/x^*$, while the remaining $n - x^*$ sites are visited with $\pi_i = F$, where $x^* = \arg\max_{x\in\{1,\dots,n\}} x\left[\psi\left(F + \frac{\Pi - Fn}{x}\right) - \psi(F)\right]$; or

(ii) select $x^{**} = \lfloor(\Pi - nF)/(T - F)\rfloor$ sites that are visited equally often with $\pi_i = T$, while the rest is visited with $\pi_i = F$, except for one site that is visited with $\pi_i = \Pi - x^{**}T - (n - x^{**} - 1)F$.

Under these conditions, the problem is reduced to optimizing a single variable: x^*, the number of sites to be visited equally often with a relatively high frequency. We refer to such sites as *prioritized*. The remaining sites are visited only the minimum number of times ($\pi_i^* = F$).

This result uncovers a problem of equity even for identical sites. Some sites receive the bare minimum of resources, while other sites receive a large fraction. Therefore, it is critical to set the lower bound F appropriately such that every site receives a satisfactory level of service.

Before turning to the analysis of other cases, we briefly discuss the effect that the underlying adoption dynamics have on the number of prioritized sites x^*. For this purpose, we assume that T is sufficiently large such that Case (i) above applies, that ψ follows the Bass function, and that x^* can take non-integer values. The effect of a small change in a parameter of the Bass function can then be identified easily through implicit derivation. An increase in c can be interpreted as shortening the convex part of the Bass function. This means that the number of adopters per visit begins to saturate more quickly, and, thus, it is beneficial to choose a larger set of sites. When m increases, assuming $c>0$, the opposite happens—the convex part is lengthened, and x^* decreases. An increase in p implies that innovation is more prominent and that adopter demand grows more quickly but saturates faster. Early on, it may thus be optimal to visit a smaller number of prioritized sites x^* more frequently. However, if cumulative demand is already close to the inflection point, an increase in p leads to an increase in x^*. An increase in q has more complex effects tied to c: the benefits of increasing q scale with the number of existing clients because they enable imitation, but saturation effects act as a counterweight and also scale with existing clients.

4.2. Optimal allocation for homogeneous sites with different starting points

Next, we assume a set of sites with the same impact multipliers and point-symmetric demand functions but a different number of existing clients. This is a useful assumption when populations around sites are not too different, but sites have previously been served with varying frequencies.

Let ψ be the demand function that all of the sites follow and let π_i^0 be the number of previous visits to the site, that is, $\psi(\pi_i^0)=c_i$. Then, $\psi_i(\pi_i)=\psi(\pi_i^0+\pi_i)-\psi(\pi_i^0)$. Moreover, let π^{infl} be the inflection point of ψ and $\hat{\pi}(\lambda)$ the point that satisfies the KKT condition in (5) within the concave region of ψ. Let $C_1=\{i\in N:\pi_i^0+F\geq\pi^{\text{infl}}\}$ (resp. $C_2=\{i\in N:\pi_i^0+F<\pi^{\text{infl}}\}$) be the set of sites for which the starting point, i.e., previous visits plus minimum number of visits F, is strictly greater (resp. lower) than the inflection point in ψ.

Proposition 5. *Assume $\psi_i(\pi_i)=\psi(\pi_i^0+\pi_i)-\psi(\pi_i^0)$ for all $i\in N$, and let ψ be point-symmetric around the inflection point π^{infl}. There exists an optimal solution π^* to (1) such that*

(a) $E_2^*=\emptyset$ or $E_1^*=\emptyset$;

(b) if $i,j\in C_2$, $i\in L^*$, and $\pi_i^0>\pi_j^0$, then $j\in L^*$;

(c) if $i,j\in C_1$, $i\in L^*$, and $\pi_i^0<\pi_j^0$, then $j\in L^*$; and

(d) if $i\in E_1^*$ and $j\notin L^*$, then $\pi_j^0+\pi_j^*\leq\pi_i^0+\pi_i^*$.

Proposition 5 allows us to find the optimal allocation through a greedy search by sequentially taking sites out of the set L. This idea is formalized in Corollary 1:

Corollary 1. *Assume a set of sites as in Proposition 5. Wlog, reorder the sites such that* *(a) if $j \in C_1, i \in C_2$, then $i < j$;*

(b) if $i, j \in C_2$, $\pi_i^0 \geq \pi_j^0$, then $i < j$; and

(c) if $i, j \in C_1$, $\pi_i^0 \leq \pi_j^0$, then $i < j$.

There exists $l^* \in \{0, \dots, |C_2|\}$ and λ^* such that an optimal allocation π^* exists that satisfies either

$$\pi_i^* = \begin{cases} \min\left\{\left(\hat{\pi}(\lambda^*) - \pi_i^0\right), T\right\}, & \text{if } i \leq l^* \\ F, & \text{if } i \in C_2 \text{ and } i > l^* \\ \min\left\{\max\{\hat{\pi}(\lambda^*) - \pi_i^0, F\}, T\right\}, & \text{if } i \in C_1 \end{cases} \quad \text{for all } i \in N, \text{ or} \tag{7}$$

$$\pi_i^* = \begin{cases} \Pi, & \text{if } i = 1 \\ F, & \text{if } i > 1 \end{cases} \quad \text{for all } i \in N. \tag{8}$$

We propose an algorithm in Appendix A that finds an optimal allocation through a search for l^* and λ^*, requiring a constant number of line searches only. Here, we provide an overview. For any l sites with a starting point in the convex part (sites in C_2), the procedure computes the best allocation when those sites are prioritized. The prioritized sites are those with the highest current demand (adopters per visit). For each choice of l, some sites are also prioritized with a starting point in the concave part. However, this is limited by all prioritized sites having the same number of adopters per visit at the end of the planning horizon. As a result, the prioritized sites from C_1 are those with the lowest current demand. Finally, we note that prioritized sites receive more resources the lower their cumulative demand level and other sites receive only the minimum number of visits.

4.3. Optimal allocation for heterogeneous sites

When sites are heterogeneous, the allocation problem is NP-hard (Ağrali and Geunes 2009, Theorem 1). The algorithm of Srivastava and Bullo (2014) is a heuristic approach for solving the general allocation problem that links to the intuitive development we have done here. The algorithm uses a relaxation of (1) to search for the Lagrange multiplier and then allocate resources given this multiplier. We apply this algorithm in Section 6.3 to obtain a proxy of the optimal allocation.

4.4. The impact of adoption dynamics on the optimal allocation

Without a closed-form solution, it is difficult to intuit how differences in the underlying adoption dynamics affect the optimal allocation: should a site with more imitators be visited more often, or one with more innovators? What does it mean for the visit frequency at one site when the number of existing clients at another site is large? We can provide intuition into how an optimal allocation is affected by small differences between the adoption dynamics at sites. This can also be useful when an organization aims to affect a site's parameters, for example, through marketing.

We analyze the effects of a small change in a parameter s that affects one site's demand function. Suppose that π^* is an optimal allocation and, without loss of generality, assume that the inequality constraints are not binding for sites $i = 1, \dots, \hat{n}$, that is $F < \pi_i^* < T$, but binding for all other sites.

Assume further that the change in s does not impact whether any inequality constraints are binding. Recall the first-order conditions $\beta_i \psi_i'(\pi_i^*) = \lambda^* \ \forall i = 1, \ldots, \hat{n}$, and the equality constraint $\sum_{i=1}^{\hat{n}} \pi_i^* = \Pi - \sum_{i=\hat{n}+1}^{n} \pi_i^*$. By taking derivatives with respect to s on both sides, we obtain $\beta_i \frac{\partial \psi_i'}{\partial s}(\pi_i^*) + \beta_i \psi_i''(\pi_i^*) \frac{\partial \pi_i^*}{\partial s} = \frac{\partial \lambda^*}{\partial s}$ and $\sum_{i=1}^{\hat{n}} \frac{\partial \pi_i^*}{\partial s} = 0$ (because of our assumptions, the derivatives for sites $i = \hat{n}+1, \ldots, n$ are zero). Solving the system of equations for $\partial \pi_i^* / \partial s$ and $\partial \lambda^* / \partial s$, we have

$$\frac{\partial \pi_i^*}{\partial s} = \frac{-\frac{1}{\psi_i''(\pi_i^*)} \frac{\partial \psi_i'}{\partial s}(\pi_i^*) \sum_{j \neq i, j \leq \hat{n}} \frac{1}{\beta_j \psi_j''(\pi_j^*)} + \sum_{j \neq i, j \leq \hat{n}} \frac{1}{\beta_i \psi_i''(\pi_i^*) \psi_j''(\pi_j^*)} \frac{\partial \psi_j'}{\partial s}(\pi_j^*)}{\sum_{j \leq \hat{n}} \frac{1}{\beta_i \psi_j''(\pi_j^*)}}, \text{ and} \tag{9}$$

$$\frac{\partial \lambda^*}{\partial s} = \frac{\sum_{j \leq \hat{n}} \frac{1}{\psi_j''(\pi_j^*)} \frac{\partial \psi_j'}{\partial s}(\pi_j^*)}{\sum_{j \leq \hat{n}} \frac{1}{\beta_j \psi_j''(\pi_j^*)}}. \tag{10}$$

Suppose that s only affects Site 1 and all sites $i < \hat{n}$ are in E_1^*, i.e., $\psi_i''(\pi_i^*) < 0$ (this is not a strong condition as at most one site is in E_2^*, see Proposition 3). Then, (9) and (10) simplify to

$$\frac{\partial \pi_1^*}{\partial s} = \underbrace{-\frac{\sum_{j \neq 1, j \leq \hat{n}} \frac{1}{\beta_j \psi_j''(\pi_j^*)}}{\sum_{j \leq \hat{n}} \frac{1}{\beta_j \psi_j''(\pi_j^*)}} \frac{1}{\psi_1''(\pi_1^*)}}_{>0} \frac{\partial \psi_1'}{\partial s}(\pi_1^*), \qquad \frac{\partial \pi_i^*}{\partial s} = \underbrace{\frac{\frac{1}{\beta_i \psi_i''(\pi_i^*) \psi_1''(\pi_1^*)}}{\sum_{j \leq \hat{n}} \frac{1}{\beta_j \psi_j''(\pi_j^*)}}}_{<0} \frac{\partial \psi_1'}{\partial s}(\pi_1^*), \ \forall i = 2, \ldots, \hat{n}, \text{ and}$$

$$\frac{\partial \lambda^*}{\partial s} = \underbrace{\frac{\frac{1}{\psi_1''(\pi_1^*)}}{\sum_{j \leq \hat{n}} \frac{1}{\beta_j \psi_j''(\pi_j^*)}}}_{>0} \frac{\partial \psi_1'}{\partial s}(\pi_1^*).$$

Thus, the sign of the first-order change is fully determined by $\partial \psi_1'(\pi_1^*)/\partial s$. If $\partial \psi_1'(\pi_1^*)/\partial s > 0$, then π_1^* increases, while π_i^* decreases for $i = 2, \ldots, \hat{n}$. Interestingly, the shadow price of the capacity constraint, λ^*, increases if and only if π_1^* increases.

Suppose now that $\psi_1(\pi_1)$ is given by the Bass function $Y_1(\pi_1) - c_1$. We can easily compute derivatives with respect to the relevant parameters p_1, q_1, m_1, and c_1 to obtain the direction of change. For reasons of space, we provide the formal results in Appendix C. We show that an increase in the market potential of Site 1 always increases the optimal number of visits to that site while decreasing the optimal number of visits to other sites. This is intuitive because a larger pool of potential clients results in more adopters per visit. An increase in the number of existing clients reduces the optimal number of visits to Site 1 while increasing the number of visits to other sites. This is because, with more existing clients, there are fewer potential ones.

The results for p_1 and q_1 are more nuanced. Intuitively, p_1 and q_1 represent rates of diffusion: p_1 for innovators and q_1 for imitators. Larger values imply faster diffusion. It may be optimal to increase the number of visits to take advantage of this, or to decrease the number of visits because the same number of adopters can be reached with fewer visits. In fact, if π_1^* is small, $\partial Y_1'(\pi_1^*)/\partial s > 0$

for $s \in \{p_1, q_1\}$. For larger values of π_1^*, when adopter demand at the site is close to saturation, $\partial Y_1'(\pi_1^*)/\partial s < 0$. The point where the sign of $\partial Y_1'(\pi_1^*)/\partial p_1$ changes is always in the concave region, within the assumed range of π_1^*. The point where the sign of $\partial Y_1'(\pi_1^*)/\partial q_1$ changes is in the concave region if and only if $q_1 > p_1$, which is commonly observed in practice (see, e.g., Bass 1969).

4.5. Summary of managerial insights from the analysis

We summarize here the key insights for decision-makers, derived through the analyses in the preceding sections. These insights may be used to develop a heuristic approach to the allocation problem, especially in settings with poor historical data. We provide examples of heuristic approaches in Section 6.3 and compare them to both myopic and optimal allocations.

There are three intuitively appealing approaches to allocating MHU resources that can be observed, at least anecdotally, in practice. One is to allocate resources equally. The second is to reserve most resources for a small group of sites with very high current (i.e., per-visit) adopter demand. The third is to allocate resources in proportion to the current adopter demand. The optimal allocation runs counter to these intuitive approaches but incorporates some of their elements. As opposed to spreading resources equally, it is optimal to prioritize a subset of sites. Rather than prioritizing only the top demand sites, this subset should be large enough to anticipate saturation effects. While sites are more likely to be prioritized in the optimal allocation when current adopter demand is high, this holds only for less developed sites, i.e., where the pool of existing clients is relatively small compared to the potential. Prioritization amongst more developed sites may, in fact, follow a reverse order. A more general yardstick for prioritization is the level of cumulative adopter demand. When allocating resources between prioritized sites, as opposed to apportioning them proportional to current demand, resources should be allocated such that the number of adopters per visit at the end of the planning horizon is equal across sites. Therefore, sites with higher adoption potential, which generally have lower current demand, should receive more visits.

These two interlinked decisions—which sites to prioritize and how to allocate resources between prioritized sites—are influenced by the underlying adoption dynamics. In particular, more sites should be prioritized when market potentials are smaller, there are more existing clients, or diffusion rates are higher (both in the form of innovation and imitation). In terms of the allocated resources, a particular site should be favored if it has higher market potential or fewer existing clients, ceteris paribus. Similarly, higher diffusion rates at a site should mean that more resources are allocated to that site unless saturation effects are already noticeable.

5. A procedure for predicting adopter demands

In order to allocate resources optimally, we need to be able to predict demand at all sites. Here, we outline a method that combines the Bass diffusion model's predictive power and machine learning. We assume that the adoption dynamics at each site follow a Bass function and use a combination of data gathered in the field (visits) and publicly available data about the sites (location, demographics, economy, . . .) to estimate the parameters with machine learning methods.

In the context of MHUs, there are many concurrent markets for the same health services, and adoption proceeds at each site largely independently because potential clients interact mostly with people served at the same site. At the same time, while sites vary in the composition of their population, there are many economic and cultural commonalities between populations, which means that the underlying dynamics (specifically, the imitation and innovation coefficients) are likely to be linked. Therefore, we take the approach of estimating the Bass parameters as a function of site characteristics, thus pooling the demand data from all sites to get better estimates of the Bass model for each site, especially for sites with little demand data available.

We address two additional issues relevant to the context of MHUs. First, there may be existing clients that are not registered but contribute to the diffusion of health services through word-of-mouth effects. Such a discrepancy may be due to a lack of quality data gathered in the field, e.g., many NGOs have only started gathering data at scale in recent years to increase efficiency and accountability, leading to missing data on existing clients. Moreover, while the sites visited by MHUs are typically isolated, some migration may occur and provide a baseline of clients that have already obtained the service. Therefore, our approach is able to estimate existing clients at each site, c_i, from incomplete data on existing clients as well as site characteristics.

Second, while data gathered in the field may be limited, many data sources are becoming publicly available that allow us to enhance predictions. In particular, we can estimate the market potential from the population in proximity to a site. Our approach estimates m_i as a fraction of the population living in the *catchment area* of site i, where the fraction of potential clients in the population is based on the experience from related settings and the demographic composition around the site. The catchment area varies across sites, specifically between rural and urban areas (Doerner et al. 2007). However, it is likely that at sites with similar characteristics, people are willing to walk similar distances, even if the population residing within that distance varies. Hence, we obtain population data at different distances from a site, using the WorldPop project (Lloyd et al. 2019). We represent the population in the catchment area as a function of the radius r_i that people living close to site i are willing to travel to obtain service. The parameter r_i is then estimated as part of the suggested procedure. The market potential m_i is obtained by multiplying the population in the catchment area by the fraction of the population that are potential clients.

We use gradient boosting, a state-of-the-art machine learning method (Natekin 2020), to estimate each site's Bass parameters. With gradient boosting, a function of the difference between actual and predicted outcomes (the loss function) is minimized iteratively. In each iteration, pseudo-residuals of the predicted outcomes are regressed on observables using a weak learner, often a regression tree model. The prediction model is then a weighted combination of the previous iteration's model and the model generated by the weak learner. We modify the standard approach in two ways: while adopter demand for health services is the outcome of interest and the basis of the loss function, we simultaneously use four regression trees in order to estimate p_i, q_i, r_i, and c_i as a function of site characteristics, instead of estimating demand directly. We further add a penalty term to constrain the model parameters within the relevant space of the Bass model.

Combining these aspects, we propose the following procedure for predicting the functions $Y_i(\pi_i)$:

1. Determine the market potential of a site as a function of the site-specific radius, say $m_i(r_i)$, using the population density around the site multiplied by the fraction of potential clients.
2. Define a functional form of the innovation and imitation coefficients, the site-specific radius, as well as the number of existing clients as functions of the site characteristics X_i, say $p(X_i)$, $q(X_i)$, $r(X_i)$, and $c(X_i)$, to be learned from the data. We use a sum of weighted regression trees. We also explore a linear form in Section 6.2.
3. Define a loss function. We use the mean squared error or quadratic loss function: $\sum_{i=1}^{n}\sum_{j=1}^{v_i}\left[y_{i,j} - Y\left(j; p(X_i), q(X_i), m_i\left(r(X_i)\right), c(X_i)\right)\right]^2$, where Y is the number of adopters according to the Bass model at the j-th visit, given the model parameters at site i, $y_{i,j}$ is the number of adopters observed, and v_i is the total number of visits.
4. Estimate $\hat{p}(X_i)$, $\hat{q}(X_i)$, $\hat{r}(X_i)$, and $\hat{c}(X_i)$ with a modified gradient boosting algorithm, in which each iteration simultaneously updates the four models. If we use a linear functional form, we can estimate the parameters directly by minimizing the quadratic loss function over the coefficients, i.e., we obtain a non-linear least-squares estimator.
5. Predict $Y_i(\pi_i)$ based on Equation (2) with $Y_i(\pi_i) = Y\left(\pi_i; \hat{p}(X_i), \hat{q}(X_i), m_i\left(\hat{r}(X_i)\right), \hat{c}(X_i)\right)$.

The predictive power of our approach far outperforms that of a non-linear least-squares approach (the mean squared error is approximately than a third). In fact, the approach has nearly the same predictive power as a pure "black-box" gradient boosting approach. We implement this procedure in the R programming language and make it available together with synthetic data at **blinded for peer review**.

Experiments with synthetic data allow us to investigate the procedure's accuracy. We observe that predictions are highly accurate and that parameter estimates predict the inflection point of the cumulative demand curve correctly. Estimates of q_i, r_i, and m_i are accurate. Small inaccuracies are possible in p_i and c_i estimates because nearly equivalent demand curves can be generated by compensating lower (higher) values of p_i with higher (lower) values of c_i.

6. Application to MSI Uganda

In this section, we apply our model to MSI Uganda, an organization that faces the resource allocation problem discussed throughout this paper. We give some background on the organization in Section 6.1 before applying our demand prediction techniques in Section 6.2. We use the model parameters estimated to illustrate the impact of an optimal resource allocation in Section 6.3 and compare the optimal allocation to current implementations, a myopic policy, as well as heuristics that make use of our analytical insights.

6.1. MSI Uganda's family planning MHUs

Context. Family planning services are an important contributor to *SDG 3: Good Health and Well-Being*. Reasons include the mental (Groupo Médico and Global Doctors for Choice Network 2011) and physical (Conde-Agudelo et al. 2012) adverse effects of unwanted pregnancies and quick successions of childbearing, as well as the 25 million unsafe abortions yearly caused by undesired pregnancies and resulting in 14–40 thousand deaths (WHO 2019b,a). Apart from adversely impacting health, a lack of family planning is a major barrier to *SDG 1: No Poverty*, *SDG 4: Quality Education*, and *SDG 5: Gender Equality*.

Organization. MSI Reproductive Choices (MSI) tries to address the need for family planning services of underserved populations worldwide. In Uganda, MSI has been active for over 30 years. Today, it is one of the largest providers of sexual and reproductive healthcare in the country. MSI focuses on reaching predominantly young and poor clients within hard-to-reach rural locations and in urban slums. To serve remote areas, the organization employs 31 MHUs in this country (MSI Reproductive Choices 2021b).

We focus on visits by the 25 standard MHUs (in contrast to specialized teams, which we exclude for consistency). Each unit operates with a driver and two to four family planning providers (doctors, nurses) and is responsible for visiting a mostly exclusive set of villages. When visiting a village, MHUs usually set themselves up at local health centers, where potential clients can approach them for services. Such a visit is commonly announced multiple weeks in advance.

Currently, resource allocation decisions do not systematically consider the trade-offs arising from adoption dynamics and the need to develop demand at sites. In fact, MSI's guidelines state explicitly that the frequency of visits should be based on current demand (MSI Reproductive Choices 2016).

Visit data. Between May 2015 and November 2019, standard teams provided contraceptive services to 542,845 clients at 1,237 sites with known geolocations. We observe the acting team, basic demographics and previous contraceptive status of the clients, and the services performed. In some cases, we observe additional information, for example, about the perception of marketing tools. We focus exclusively on client interactions regarding long-term contraceptive methods, i.e., intrauterine devices and implants. There are two main reasons for this. First, MSI shows an increasing focus on expanding choice of contraceptive methods to include long-term methods, due to their effectiveness and convenience (Duvall et al. 2014), and three-quarters of MSI clients worldwide now use long-term methods (MSI Reproductive Choices 2020). In Uganda, long-term methods account for 90% of client interactions and complement the short-term methods predominantly provided in the public health sector. MHUs are usually deployed to a site only when there is no other provider of long-term methods. Second, beyond MSI, long-term methods are seen as a cornerstone in scaling up access to family planning in the developing world (WHO 2012, Shoupe 2016).

Client interactions concerning long-term methods fall into one of two categories: the client may already be using such a method and require removal with or without replacement (existing clients), or seek usage for the first time (adopters—MSI uses the term also for clients who are not using a contraceptive method at the time of visit but may have used one in the past. If a client previously used a long-term method, we consider them an existing client for this paper). We focus our attention on adopter demand to reflect MSI's ambition of broadening access to modern contraceptive methods (MSI Reproductive Choices 2021a). Existing clients are significantly less hampered by informational barriers to access. Their physical barriers to access can be overcome if their demands, e.g., for removals or replacements, can realistically be met with a baseline visit frequency.

Demographic data. Uganda is subdivided administratively into districts, counties, sub-counties, parishes, and villages. As of 2014, the time of the most recent publicly available census, there were 7,557 parishes with an average population of 4,583. The census data is available at the level of parishes and contains statistics on both individuals and households, which we use as site characteristics, in addition to the data provided by MSI. There are other influences on demand, e.g., religious and ethnic diversity in different regions that influence social norms around family planning or variations in exposure to public promotional activities. While the census does not capture these directly, it provides a wealth of information.

In order to estimate the population in the catchment area of a site, we rely on the Uganda 2020 dataset from the WorldPop project (WorldPop et al. 2018), which provides population estimates at a resolution of 3 arc (approximately 100×100 meters at the equator). Such a granular breakdown allows us to define a site's population (i.e., the population in the catchment area) as a function of the radius that potential clients are willing to travel. To speed up computation, we define $m_i(r_i)$ as third-degree polynomials that fit the WorldPop data with an adjusted $R^2 \geq 0.97$. To derive the market potential as a fraction of the site's population, we make two calculations. First, we note that among all observed MSI clients, 98.86% are female, and 99.59% are within the reproductive age (15–49 years old, see UN 2019). Hence, we multiply the total population with the fraction of females within reproductive age in the area around the site, as found from the census data. Second, based on Track20's classification of the stages of adoption of modern contraceptive methods (reported in Feyisetan et al. 2017), we assume that the prevalence of such methods reaches saturation at approximately 80%, the highest contraceptive prevalence rate among developed countries. We thus multiply the number from the previous step by 80%.

Impact of service delivery at a site. MSI's impact from providing contraceptive methods depends on the age and economic condition of clients. We compute poverty and age scores for each site and multiply those to derive the impact of serving clients at a specific site. For the poverty score, we use district-level GDP per capita data that we regress on the district-level census information. We then use the parish-level census information to predict GDP per capita around the sites. The scores are normalized, such that the site with the lowest predicted GDP per capita has a score of 1. To compute the age score, we calculate the fraction of young persons (15–24 years old, see UN 2020) out of the population in reproductive age. Again, the score is normalized so that the site with the highest fraction of young persons has a score of 1. In summary, $\beta_i = (\text{age score})_i(\text{poverty score})_i$.

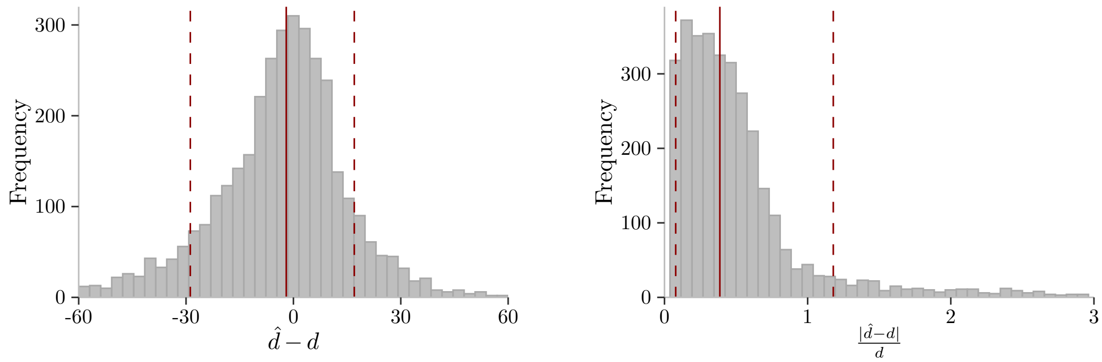
In Appendix D, we describe in detail the process of cleaning and matching between the different data sets and the auxiliary data sets used for this purpose.

6.2. Adopter demand prediction results

We aim to obtain accurate predictions of how demand evolves in the long term to incorporate demand build-up and saturation effects into the planning procedure. For this purpose, we assume that contraceptive service adoption by MSI's target population follows a Bass model. Multiple considerations drive the choice of this model. First, social interactions have been established widely as a key driver for the adoption of contraceptive methods (see, e.g., Bongaarts and Watkins 1996). Second, we are interested in long-term demand trends of entire sites rather than short-term effects or individual behaviors. Given these two requirements, the healthcare modeling literature suggests a system dynamics approach (Barton et al. 2004). The Bass model is likely the most used system dynamics model to tackle questions of adoption. It is supported by well-established diffusion theory and can represent a wide range of adoption dynamics.

Based on the assumed Bass dynamics, we predict adopter demands at different sites with the estimation procedure outlined in Section 5. The observation period is limited, making it necessary to evaluate our approach's predictive performance in the short term. For this purpose, we split the visit data into a training (May 11, 2015, until November 29, 2018; 12,876 visits) and a test period (November 30, 2018, until November 29, 2019; 3,518 visits). This means that we use the last full year of observations from MSI (21.46% of visits) as test data. We choose a date-specific cut-off to simulate the prediction problem faced by MSI when planning its resource allocation based on the information obtained from previous visits. In Figure 3a, we present a histogram of the deviation between predicted and actual adopter demands in the test period. Figure 3b contains a histogram of the absolute difference as a percentage of actual demands.

Figure 3 **Difference between predicted and actual demands**



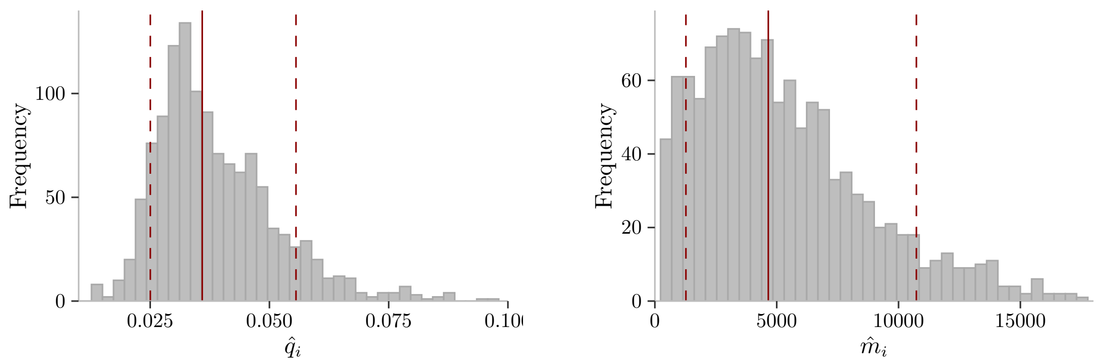
(a) Predicted minus actual adopter demand (b) Absolute difference over actual adopter demand

Note. Actual and predicted demands are denoted with d and $\hat{d}$. The red lines represent 10%, 50%, and 90% percentiles, from left to right. Panel (a) (resp. Panel (b)) trims 2.0% (resp. 2.9%) of observations from the test data for clarity.

We can use a model-independent state-of-the-art approach to derive an upper bound for prediction accuracy in the short term. In particular, we implement a standard gradient boosting approach, using the R-package gbm (Ridgeway 2019). The mean squared error in the test data using this approach is 16% less than in our model-based approach. This indicates that the Bass model may not capture some demand drivers. However, we point out that the limited test period puts more emphasis on short-term variations in demand levels compared to the long-term variations captured by the Bass model. To see this, note that the Bass estimates indicate that the adopters observed in our data correspond only to 8% (resp. 6%) of the market potential of the mean (resp. median) site. This reduces to 3% (resp. 2%) when considering only the test period. That our approach achieves comparable predictive accuracy to a state-of-the-art approach without model constraints provides important evidence for our assumption. At the same time, it is unlikely that a model-independent approach will appropriately capture longer-term trends in the adopter demand.

Our model-based estimation approach further provides us with insights into the drivers of adoption, as the procedure outputs $\hat{p}_i$, $\hat{q}_i$, $\hat{m}_i$, and $\hat{c}_i$ for each site. Interestingly, we find that $\hat{p}_i$ is consistently estimated very low (median[IQR] $= 1.23 \cdot 10^{-9}[9.97 \cdot 10^{-10}$–$1.76 \cdot 10^{-9}]$) compared to $\hat{q}_i$ (median[IQR] $= 3.59 \cdot 10^{-2}[2.99 \cdot 10^{-2}$–$4.57 \cdot 10^{-2}]$). In comparison, the meta-analysis of Van den Bulte and Stremersch (2004) finds that the imitation coefficient is on average two orders of magnitude larger than the innovation coefficient. Thus, the data suggests that most adoption activity is driven by imitation, in line with MHU experiences about the value of community spread (Wickstrom et al. 2013). While there is little variance in $\hat{p}_i$ between different sites, $\hat{q}_i$ varies strongly (see Figure 4a) and dominates the diffusion process. We also present the estimated market potential $\hat{m}_i$ in Figure 4b. Our estimation procedure adds a penalty term that ensures the radius is not estimated much higher than 7km, a reasonable walking distance for health services (see Doerner et al. 2007), and the radius is consistently estimated close to this bound. However, $\hat{m}_i$ has significant variance due to the variation in population density at the sites. The estimates of $\hat{c}_i$ are 8.2%[5.2%–13.9%] of $\hat{m}_i$ for the median[IQR] site. This suggests that a majority of sites are well below saturation.

Figure 4 Coefficient estimates



(a) Imitation coefficient

(b) Market potential

Note. The red lines represent the 10%, 50%, and 90% percentiles, from left to right. Panel (b) trims 1.9% for clarity.

As the procedure uses regression trees within the gradient boosting framework, it allows us to infer the importance of regressors in estimating different parameters. We highlight that these connections are not causal. However, they may be useful by providing a sanity check and giving a starting point for better understanding the impact that visits and marketing activities have on sites. For example, the estimate $\hat{c}_i$ is determined largely by the number of existing clients observed during each visit, but also the average number of children of adopters has some influence. The estimate $\hat{p}_i$, indicating the strength of innovation, is influenced by the number of MSI clients for other services than the ones considered here, i.e., shorter-term methods. This gives some credence to the assumption that experience with modern contraceptive methods affects the willingness to adopt long-term methods, even if there is no prior community experience with the latter. Finally, we highlight that the percentage of young people with work has a strong effect on the radius of the catchment area and, thus, $\hat{m}_i$, lending further confidence in the method and potentially providing insights into the optimal spacing between sites.

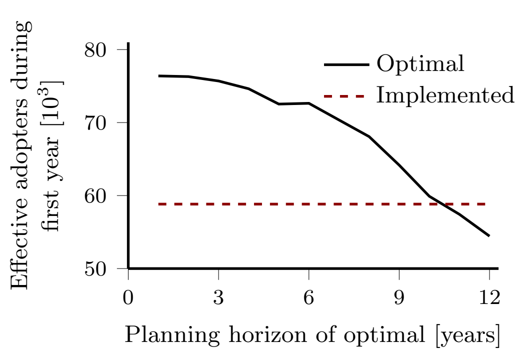
6.3. Resource allocation to MSI's sites

We study the performance of an optimal allocation based on the estimated model parameters. To identify an optimal allocation, we employ the algorithm of Srivastava and Bullo (2014) with a subsequent local search. We first compare it against the allocation implemented by MSI. We then develop operationalizable heuristics based on our managerial insights and compare those to the optimal allocation. As a measure of comparison, we use effective adopters, i.e., the number of adopters weighted by the impact generated at the site where they demand service. We implement computations in the Julia programming language and make them available together with synthetic data at **blinded for peer review**.

6.3.1. Optimal vs. implemented allocation. Aside from the Bass model parameters estimated in Section 6.2, the optimal allocation is based on four parameters: the planning horizon H, the total visit capacity Π_H during the planning horizon, as well as the minimum and the maximum number of visits to each site, F_H and T_H. We derive these from the test period, i.e., the last full year of observations in the data. In particular, there are 1229 sites and 3509 visits, so for a planning horizon of $H = 1$ year, we specify that $\Pi_1 = 3509$. MSI requires several weeks to prepare a visit, so we constrain the maximum number of visits to $T_1 = 12$ per year. Moreover, to ensure a certain level of service, we constrain the minimum number of visits to $F_1 = 1$ per year. For any other planning horizon, H, we scale the constraints accordingly: $\Pi_H = \Pi_1 H$, $T_H = T_1 H$, and $F_H = F_1 H$.

A key characteristic of the optimal allocation, compared to current practice, is that it recognizes saturation effects and the resulting need to build up adopter demand at otherwise low-demand sites. Hence, when the planning horizon is long enough, current demand is sacrificed to achieve high demand in the future. We show this in Figure 5. Here, we assume a planning horizon $H = 1, \ldots, 12$, but measure the effective adopter numbers during the first year. For this purpose, we take the actual observations from the test period whenever a visit occurred, or we impute the numbers based on our estimated demand model. As a counterfactual, we use the allocation implemented in practice and, thus, the actual adopters observed during the test period. With a planning horizon of $H = 1$ year and the same total number of visits, the optimal allocation outperforms the implemented allocation. As the planning horizon increases, however, current adopter demand is strategically sacrificed for the benefit of future adopter demand, and the optimal allocation's first-year performance decreases.

Figure 5 Effective adopters during the first year in the optimal compared to the actual allocation



6.3.2. Optimal vs. heuristic allocations. The benefits of our approach become apparent only when we also consider effective adopters in later years. Unfortunately, we do not observe a counterfactual in the data because the test period is only one year. Thus, we simulate allocations based on the estimated demand functions. We compare our approach to several heuristic allocations and display the results in Figure 6a. A myopic allocation assigns the maximum number of visits to the sites for which the number of effective adopters is currently highest and the minimum number of visits to all other sites. It is comparable to the optimal allocation for short planning horizons. However, as the planning horizon increases, we also observe what we would expect from the implemented allocation in the longer term: saturation of adopter demand. The myopic allocation has to be adapted dynamically to counter this effect partially—we say "Myopic x" if the allocation is adapted to current demand every x years. However, even with yearly adaptations, it does not match the performance of the optimal allocation in the long term.

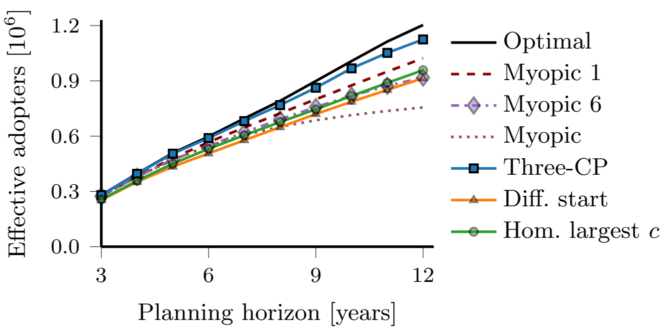
The remaining heuristics make use of the insights summarized in Section 4.5 to develop practical allocations that avoid saturation without dynamic adjustments. Noting that E_2^* has at most one member, we split the sites into the three sets L^*, G^*, and E_1^*, defined in (6). We ignore heterogeneity between the sites in E_1^* and distribute resources equally among them to simplify implementation. This is in line with de Vries et al. (2021a), who suggest splitting sites into few groups with identical visit frequencies per group, and we thus use their naming, *Three-Category Policy*. The heuristic, as defined here, still requires the computation of the optimal allocation. However, it simplifies implementation drastically, and Figure 6a shows that this is without a significant loss in performance.

Next, the *Different Start* heuristic uses the algorithm developed in Section 4.2, based on the average values of the estimates $\hat{p}_i$, $\hat{q}_i$, and $\hat{m}_i$, but with the individual estimates $\hat{c}_i$ for the number of existing clients. This is appropriate when the adoption dynamics and market potentials at different sites are highly similar, but the visit histories differ significantly.

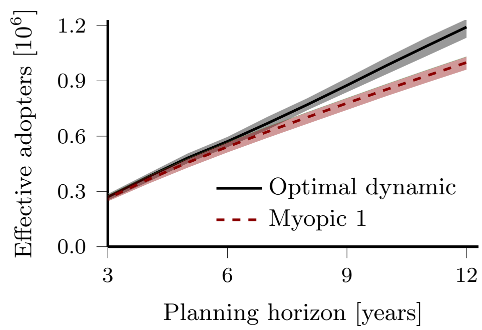
The final heuristic makes use of the algorithm developed in Section 4.1 to determine how many sites should be prioritized. The algorithm assumes entirely homogeneous sites, so we take averages of $\hat{p}_i$, $\hat{q}_i$, $\hat{m}_i$, and $\hat{c}_i$. As sites are not actually homogeneous, we employ the following approach to choose which of the sites to prioritize, termed *Homogeneous, Largest c*. The sites with the largest values of $\hat{c}_i$ are prioritized as long as they have not yet reached the inflection point in the diffusion process. We exclude sites beyond the inflection point to avoid saturated markets. Using $\hat{c}_i$ for prioritization captures the results from Section 4.2. Visits are equally distributed among the prioritized sites, so the complexity of implementing the allocation is much reduced.

The latter three heuristics indicate the usefulness of the insights derived in Section 4. Even without frequent changes to the visit schedule and when based on the assumption of homogeneity, they largely avoid saturation. Except for the *Three-Category Policy*, the heuristics are simpler to compute than the optimal allocation. At the same time, except for *Different Start*, the heuristics are easy to implement. We believe that regional managers can further improve performance by combining insight-based heuristics with information available on the ground.

Figure 6 Performance throughout the planning horizon of the optimal and heuristic allocations



(a) Parameter certainty



(b) Uncertainty sequentially resolved

Note. Panel (a) assumes known Bass model parameter values, and Panel (b) assumes uncertain parameters.

6.3.3. Uncertainty in parameter estimates. To investigate the effect of uncertainty in Bass parameter estimates, we simulate two dynamic heuristics. To this end, we perform 50 replications of the following procedure. We sample the true parameters by scaling each of the original estimates by independent random variables uniformly distributed between 0.5 and 1.5. When allocations are adjusted each year, we assume that estimates approach the true value. In particular, we assume that the difference between a parameter's true value and the estimated value is proportional to 1 over the square root of the number of visits to the site. The Myopic 1 allocation dedicates all resources, barring constraints, to the sites with the highest estimated current demand at the beginning of the year. Another heuristic, based on the optimal allocation ("Optimal dynamic"), dynamically updates the Bass estimates every year and reoptimizes the allocation for the remaining planning horizon. Figure 6b shows the range of effective adopters across replications for both allocations. When adapting allocations dynamically, the uncertainty does not significantly affect performance, and our approach continues to outperform the myopic one in the long term.

7. Conclusions

A central objective within the sustainable development agenda is to provide broad and affordable access to health services worldwide. Nearly half of humanity lives in rural areas, leading to difficulties in expanding access. One promising approach is the use of mobile healthcare units that establish temporary delivery sites close to rural populations. Different types of organizations employ MHUs for a range of health services, each facing complex planning challenges. In this paper, we focus on organizations that provide specialized health services for which access is limited both by physical barriers and informational barriers, for example, HIV counseling, disease screening, vaccination, or family planning. The long-term allocation of resources has traditionally been based on the implicit assumption of constant demands. This ignores the presence of informational barriers, which make community interaction and trust-building essential, leading to complex adoption dynamics. We analyze the resource allocation problem when such adoption dynamics are made explicit.

Our resource allocation model for MHUs allows us to derive several insights. First, we find that it is optimal to prioritize a subset of sites while allocating the minimum necessary resources to all remaining sites (recall that all sites are guaranteed an acceptable level of service). Thus, distributing resources equally among all sites can be far from optimal, but focusing on too few sites may precipitate saturation effects. High current adopter demand can be a good criterion for prioritizing amongst underdeveloped sites, but, in general, cumulative adopter demand should be preferred as a metric. Second, among the prioritized sites, those with lower cumulative demand should be visited more often because word-of-mouth effects can lead to demand snowballing. Thus, for prioritized sites that are not yet experiencing saturation, more resources should be allocated to those where current demand is lower. This is in contrast with the practice of allocating more resources to sites with higher current demand. Third, our results confirm the intuitive effect of diffusion characteristics on the optimal allocation. Sites with higher diffusion rates through innovation or imitation, larger market potential, or fewer existing clients should generally receive more resources.

Our application to MHUs delivering family planning services in Uganda demonstrates the potential benefits of accounting for the future evolution of adopter demand when allocating resources. It shows how to predict future demands without additional data-gathering efforts and how to make decisions based on these predictions. Moreover, we find evidence in the data suggesting that imitation dominates the diffusion of family planning services and that the willingness to use long-term methods is associated with the prevalence of short-term methods in the community.

Naturally, our work has limitations. In particular, we make two assumptions about the organization employing MHUs. First, it aims to broaden access to a health service for which there are significant informational barriers, in addition to the physical ones. Second, existing clients requiring repeated services do not face those informational barriers, and a baseline frequency of visits can guarantee their access to services. This excludes, for example, organizations that provide primary care through MHUs. As with any model, we also need to make simplifying assumptions about a complex reality. In particular, we only consider the (long-term) allocation of visits to sites, but not their exact scheduling. Thus, we assume that there is no effect of the inter-arrival time on the number of adopters. Recent research indicates that there is a positive relationship between client volume and inter-arrival time (de Vries et al. 2021a), even though we verify that this is lower for service adopters than existing clients. We note that MHUs, whose resources are allocated based on our model, can still react to this effect by adjusting their final visit schedule. In particular, they may space out visits to a site when current demand is high (in the intermediate part of the Bass curve) and increase the frequency of visits when demand is low (when demand is initially building up). We believe that further research should integrate the long-term and short-term perspectives while also considering other aspects of the visit schedule affecting demand, for example, its regularity. Such research must overcome the dilemma of leaving enough flexibility to adjust schedules in the face of uncertainties while scheduling sufficiently in advance to incorporate the long-term demand dynamics described here.

We further assume the allocation of visits to be the only decision variable. In practice, organizations may directly influence the adoption dynamics through marketing or community engagement activities. To do so systematically, however, requires an in-depth understanding of these activities' effects. We provide an initial step in this direction. In terms of the estimation procedure, we point out the limited time during which we observe client interactions. We hope to further validate our approach in the future using longer time series. Finally, relying mainly on publicly available data introduces some limitations on the granularity, accuracy, and timeliness of estimates. However, it contributes to achieving a practicable and replicable approach to long-term planning of limited organizational resources.

Our framework allows for an effective planning procedure that may contribute to broadening access to health services and, thus, advance the sustainable development agenda. It also provides a foundation to discuss the importance of long-term planning in this context. This can be of particular importance because donors often push for short-term results with their funding decisions.

References

Ağrali S, Geunes J (2009) Solving knapsack problems with s-curve return functions. *European Journal of Operational Research* 193(2):605–615.

Aljassim N, Ostini R (2020) Health literacy in rural and urban populations: A systematic review. *Patient Education and Counseling* 103(10):2142–2454.

Aung KK, Hill C, Bennet J, Song Z, Oriol NE (2015) The emerging business models and value proposition of mobile health clinics. *American Journal of Accountable Care* 3(4):36.

Barton P, Bryan S, Robinson S (2004) Modelling in the economic evaluation of health care: Selecting the appropriate approach. *Journal of health services research & policy* 9(2):110–118.

Bass FM (1969) A new product growth for model consumer durables. *Management Science* 15(5):215–227.

Bass FM (1980) The relationship between diffusion rates, experience curves, and demand elasticities for consumer durable technological innovations. *The Journal of Business* S51–S67.

Bass FM, Krishnan TV, Jain DC (1994) Why the Bass model fits without decision variables. *Marketing Science* 13(3):203–223.

Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K (2011) Low health literacy and health outcomes: an updated systematic review. *Annals of internal medicine* 155(2):97–107.

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

Bongaarts J, Watkins SC (1996) Social interactions and contemporary fertility transitions. *Population and Development Review* 639–682.

Boswijk HP, Franses PH (2005) On the econometrics of the Bass diffusion model. *Journal of Business & Economic Statistics* 23(3):255–268.

Conde-Agudelo A, Rosas-Bermudez A, Castaño F, Norton MH (2012) Effects of birth spacing on maternal, perinatal, infant, and child health: A systematic review of causal mechanisms. *Studies in Family Planning* 43(2):93–114.

Crouse HL, Macias CG, Cruz AT, Wilson KA, Torrey SB (2010) Utilization of a mobile medical van for delivering pediatric care in the bateys of the Dominican Republic. *International Journal of Emergency Medicine* 3(4):227.

de la Torre LE, Dolinskaya IS, Smilowitz KR (2012) Disaster relief routing: Integrating research and practice. *Socio-Economic Planning Sciences* 46(1):88–97.

de Vries H, Swinkels LE, Van Wassenhove LN (2021a) Policies for mobile family planning services. *Production and Operations Management,* forthcoming.

de Vries H, van de Klundert J, Wagelmans A (2021b) Towards elimination of infectious diseases with mobile screening teams: HAT in the DRC. *Production and Operations Management,* forthcoming.

Doerner K, Focke A, Gutjahr WJ (2007) Multicriteria tour planning for mobile healthcare facilities in a developing country. *European Journal of Operational Research* 179(3):1078–1096.

Duvall S, Thurston S, Weinberger M, Nuccio O, Fuchs-Montgomery N (2014) Scaling up delivery of contraceptive implants in sub-Saharan Africa: Operational experiences of Marie Stopes International. *Global Health: Science and Practice* 2(1):72–92.

Ergun Ö, Gui L, Heier Stamm JL, Keskinocak P, Swann J (2014) Improving humanitarian operations through technology-enabled collaboration. *Production and Operations Management* 23(6):1002–1014.

Ferrer JM, Martín-Campo FJ, Ortuño MT, Pedraza-Martínez AJ, Tirado G, Vitoriano B (2018) Multi-criteria optimization for last mile distribution of disaster relief aid: Test cases and applications. *European Journal of Operational Research* 269(2):501–515.

Feyisetan B, Adetunji J, Starbird E (2017) Stages in the adoption of modern contraceptive methods: Do the growth patterns in developing countries follow the s-curve model? Accessed May 1, 2021, https://tinyurl.com/yylols98.

Gatignon H, Eliashberg J, Robertson TS (1989) Modeling multinational diffusion patterns: An efficient methodology. *Marketing Science* 8(3):231–247.

Ginsberg W (1974) The multiplant firm with increasing returns to scale. *Journal of Economic Theory* 9(3):283–292.

Gralla E, Goentzel J (2018) Humanitarian transportation planning: Evaluation of practice-based heuristics and recommendations for improvement. *European Journal of Operational Research* 269(2):436–450.

Groupo Médico, Global Doctors for Choice Network (2011) Unwanted pregnancy forced continuation of pregnancy and effects on mental health. Accessed May 1, 2021, https://tinyurl.com/y3upratk.

Hodgson MJ, Laporte G, Semet F (1998) A covering tour model for planning mobile health care facilities in Suhum District, Ghana. *Journal of Regional Science* 38(4):621–638.

Jukes MC, Drake LJ, Bundy DA (2007) *School Health, Nutrition and Education for All: Levelling the Playing Field* (CABI).

Keskinocak P, Savva N (2020) A review of the healthcare-management (modeling) literature published in Manufacturing & Service Operations Management. *Manufacturing & Service Operations Management* 22(1):59–72.

Laporte G (1992) The vehicle routing problem: An overview of exact and approximate algorithms. *European Journal of Operational Research* 59(3):345–358.

Lee H, Kim SG, Park Hw, Kang P (2014) Pre-launch new product demand forecasting using the Bass model: A statistical and machine learning-based approach. *Technological Forecasting and Social Change* 86:49–64.

Lenk PJ, Rao AG (1990) New models from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Science* 9(1):42–53.

Lloyd CT, Chamberlain H, Kerr D, Yetman G, Pistolesi L, Stevens FR, Gaughan AE, Nieves JJ, Hornby G, MacManus K, et al. (2019) Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data* 3(2):108–139.

Luenberger DG, Ye Y (1984) *Linear and Nonlinear Programming* (Springer).

Mabuto T, Latka MH, Kuwane B, Churchyard GJ, Charalambous S, Hoffmann CJ (2014) Four models of HIV counseling and testing: Utilization and test results in South Africa. *PloS One* 9(7):e102267.

Mauad EC, Silva TB, Latorre MR, Vieira RA, Haikel RL, Vazquez VL, Longatto-Filho A (2011) Opportunistic screening for skin cancer using a mobile unit in Brazil. *BMC Dermatology* 11(1):12.

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

Meade N, Islam T (2006) Modelling and forecasting the diffusion of innovation—a 25-year review. *International Journal of Forecasting* 22(3):519–545.

MSI Reproductive Choices (2016) Chapter 4: Operational excellence. *Outreach Channel Success Model*.

MSI Reproductive Choices (2020) Global impact report 2019. Accessed May 1, 2021, https://tinyurl.com/yy4xgcfx.

MSI Reproductive Choices (2021a) About us—Our ambition. Accessed May 1, 2021, https://tinyurl.com/yxuxfvea.

MSI Reproductive Choices (2021b) Uganda. Accessed May 1, 2021, https://tinyurl.com/y34cddpa.

Narita R, Diaz MDM (2016) Teenage motherhood, education, and labor market outcomes of the mother: Evidence from Brazilian data. *EconomiA* 17(2):238 – 252.

Natekin A (2020) Open machine learning course mlcourse.ai: Topic 10. Gradient Boosting. Accessed May 1, 2021, https://tinyurl.com/y4zwkyx5.

Neke NM, Gadau G, Wasem J (2018) Policy makers' perspective on the provision of maternal health services via mobile health clinics in Tanzania—findings from key informant interviews. *PLoS ONE* 13(9):e0203588.

Peres R, Muller E, Mahajan V (2010) Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* 27(2):91–106.

Plambeck EL, Ramdas K (2020) Alleviating poverty by empowering women through business model innovation: Manufacturing & Service Operations Management insights and opportunities. *Manufacturing & Service Operations Management* 22(1):123–134.

Prussing JE (1986) The principal minor test for semidefinite matrices. *Journal of Guidance, Control, and Dynamics* 9(1):121–122.

Ridgeway G (2019) Generalized boosted models: A guide to the gbm package. Accessed May 1, 2021, https://tinyurl.com/y5alu3nk.

Rogers EM (2010) *Diffusion of Innovations* (Simon and Schuster).

Shoupe D (2016) LARC methods: Entering a new age of contraception and reproductive health. *Contraception and Reproductive Medicine* 1(4).

Singh S, Darroch JE (2012) Adding it up: Costs and benefits of contraceptive services—estimates for 2012. Accessed May 1, 2021, https://tinyurl.com/y2sp4kv2.

Srivastava V, Bullo F (2014) Knapsack problems with sigmoid utilities: Approximation algorithms via hybrid optimization. *European Journal of Operational Research* 236(2):488–498.

Starbird E, Norton M, Marcus R (2016) Investing in family planning: Key to achieving the Sustainable Development Goals. *Global Health: Science and Practice* 4(2):191–210.

Stenberg K, Axelson H, Sheehan P, Anderson I, Gülmezoglu AM, Temmerman M, Mason E, Friedman HS, Bhutta ZA, Lawn JE, et al. (2014) Advancing social and economic development by investing in women's and children's health: A new global investment framework. *The Lancet* 383(9925):1333–1354.

Strasser R (2003) Rural health around the world: Challenges and solutions. *Family Practice* 20(4):457–463.

Strasser R, Kam SM, Regalado SM (2016) Rural health care access and policy in developing countries. *Annual Revue of Public Health* 37:395–412.

The World Bank (2018) Rural population (% of total population). Accessed May 1, 2021, https://tinyurl.com/y6clzxp3.

Udell M, Boyd S (2014) Maximizing a sum of sigmoids, https://tinyurl.com/yxfhjplf.

UN (2019) Global progress in satisfying the need for family planning. Accessed May 1, 2021, https://tinyurl.com/y6pzpypc.

UN (2020) Youth. Accessed May 1, 2021, https://tinyurl.com/y6nyla2h.

Vaahtera M, Kulmala T, Maleta K, Cullinan T, Salin ML, Ashorn P (2000) Childhood immunization in rural Malawi: Time of administration and predictors of non-compliance. *Annals of Tropical Paediatrics* 20(4):305–312.

Van den Bulte C, Stremersch S (2004) Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science* 23(4):530–544.

van Maarseveen R (2020) The urban rural-education gap: Do cities indeed make us smarter? CPB Discussion Paper. Accessed May 1, 2021, https://tinyurl.com/yxn6y7kb.

Van Rijn L, de Vries H, Van Wassenhove LN (2020) Decentralized site assignment for family planning outreach teams, unpublished manuscript.

Wang X, Rafa M, Moyer JD, Li J, Scheer J, Sutton P (2019) Estimation and mapping of sub-national GDP in Uganda using NPP-VIIRS imagery. *Remote Sensing* 11(2):163.

WHO (2012) From evidence to policy: Expanding access to family planning. Accessed May 1, 2021, https://tinyurl.com/y4nhc2ou.

WHO (2019a) Maternal mortality. Accessed May 1, 2021, https://tinyurl.com/y6kqygtv.

WHO (2019b) Preventing unsafe abortion. Accessed May 1, 2021, https://tinyurl.com/y3y4ctwo.

WHO, The World Bank (2017) Tracking universal health coverage: 2017 global monitoring report. Accessed May 1, 2021, https://tinyurl.com/y3ek945p.

Wickstrom J, Yanulis J, Van Lith L, Jones B (2013) Approaches to mobile outreach services for family planning: A descriptive inquiry in Malawi, Nepal, and Tanzania. Accessed May 1, 2021, https://tinyurl.com/cfswmz5q.

WorldPop, CIESIN, Columbia University (2018) The spatial distribution of population in 2020, Uganda. Accessed May 1, 2021, https://tinyurl.com/y4qy8hox.

Appendix A: Algorithm for identifying the solution in the case of Section 4.2

Define the capacity used by an allocation as described in Corollary 1: $P(l,\lambda) = \sum_{i=1}^{l} \min\{(\hat{\pi}(\lambda) - \pi_i^0), T\} + \sum_{i=l+1}^{|C_2|} F + \sum_{i \in C_1} \min\{\max\{\hat{\pi}(\lambda) - \pi_i^0, F\}, T\}$. Any optimal allocation satisfies $P(l^*, \lambda^*) = \Pi$. For a fixed l, $P(l,\lambda)$ is decreasing in λ, and it is increasing in l for a fixed λ. Hence, we can define the following algorithm for finding the optimal allocation for sites with homogeneous and symmetric demand functions but different starting points.

Data: Set of sites N ordered as described in Corollary 1
Result: π^*
for l **in** $0, \ldots, |C_2|$ **do**
  if $\exists \lambda$ *such that* $P(l,\lambda) = \Pi$ **then**
    $\lambda^* \leftarrow$ solution of $P(l,\lambda) = \Pi$;
    $\pi(l) \leftarrow$ optimal allocation given l and λ^* in (7);
    $\Psi(l) \leftarrow$ objective value obtained by allocation $\pi(l)$;
  else
    $\Psi(l) \leftarrow 0$;
  end
end
$\pi(|C_2|+1) \leftarrow$ optimal allocation given in (8);
$\Psi(|C_2|+1) \leftarrow$ objective value obtained by allocation $\pi(|C_2|+1)$;
$l^* \leftarrow \arg\max_l \Psi(l)$;
$\pi^* \leftarrow \pi(l^*)$;

Appendix B: Proofs of mathematical claims in the main text

Proof of Proposition 1. Simple calculus shows that $\tilde{\psi}_i'' = Y_i''(\pi_i) = -\frac{m_i p_i \theta_i (p_i+q_i)^3 e^{(p_i+q_i)\pi_i}\left(p_i e^{(p_i+q_i)\pi_i} - q_i\theta_i\right)}{\left(p_i e^{(p_i+q_i)\pi_i} + q\theta_i\right)^3}$. The sign of $Y_i''(\pi_i)$ is determined by the sign of $p_i e^{(p_i+q_i)\pi_i} - q_i\theta_i$, which has a unique root at $\log(\theta_i q_i/p_i)$ and is positive for $\pi_i < \log(\theta_i q_i/p_i)$ and negative for $\pi_i > \log(\theta_i q_i/p_i)$. If $\log(\theta_i q_i/p_i) < 0$, then Y_i is concave so we set $\pi_i^{\text{infl}} = \max\{\log(\theta_i q_i/p_i), 0\}$. By substituting the expression for $\pi_i^{\text{infl}} \pm \epsilon$ into Y_i and some algebra, it follows that Y_i is symmetric around π_i^{infl}. The term c_i cancels out.

Proof of Proposition 2. First, a solution π^* exists because the objective function is bounded and the domain is compact. Now, suppose that for such a solution, $\sum_{i \in N} \pi_i^* < \Pi$ and there exists a j such that $\pi_j^* < T$. Let $\epsilon = \min\{\Pi - \sum_{i \in N} \pi_i^*\ ,\ T - \pi_j^*\}$. Consider the feasible allocation $\pi' = (\pi_1^*, \ldots, \pi_j^* + \epsilon, \ldots, \pi_n^*)$. Because ψ is a strictly increasing function, we obtain $\sum_{i \in N} \beta_i \psi_i(\pi_i^*) - \sum_{i \in N} \beta_i \psi_i(\pi_i') = \beta_j \psi_j(\pi_j^*) - \beta_j \psi_j(\pi_j^* + \epsilon) < 0$, which contradicts the assumption that π^* is optimal. □

Proof of Proposition 3. The existence of a λ^* such that (a), (b), and (c) follows directly from the KKT conditions in (5). To prove (d), we use the second-order conditions. The second-order necessary conditions require that the Hessian of the Lagrangian evaluated at a local maximum is negative semi-definite on the hyperplane tangent to the active constraints (Luenberger and Ye 1984, section 11.8). Because the constraints are linear, the Hessian $H(\pi)$ of the Lagrangian is the Hessian of the objective function, which is diagonal with entries $\beta_i \psi_i''(\pi_i)$. To test whether this matrix is negative semi-definite on the hyperplane of active constraints, assume, without loss of generality, that the inequality constraints are tight for the set of sites such that $j > m$

for some $m \leq n$, i.e., $\pi_j = F$ or $\pi_j = T$, while for the set of sites $i \leq m$, the inequality constraints are not binding, i.e., $F < \pi_i < T$.

To define the hyperplane tangent to the active constraints, first note that there are $n - m + 1$ active constraints, so the hyperplane is $(m-1)$-dimensional. For $i < m$, consider the vectors A_i, where all entries are zeros, except for entry i that is a 1, and entry m that is a -1. For a solution $\pi' = \pi + \epsilon A_i$, the inequality constraints are still active because the components $m+1, \ldots, n$ are zeros, and the equality constraint is still satisfied because any move of an additional ϵ units in the i-th direction is balanced by the $-\epsilon$ units moved in the m-th direction. Moreover, the vectors A_i are linearly independent, so the matrix $A = (A_1 \ A_2 \ \cdots \ A_{m-1})$ defines the hyperplane tangent to the active constraints: $\{y \in \mathbb{R}^{m-1} : Ay = 0\}$. Thus, the second-order necessary conditions reduce to $A^\top H(\pi) A$ being negative semi-definite or equivalently $-A^\top H(\pi) A$ being positive semi-definite (see Luenberger and Ye 1984, sections 11.5-11.6 for an overview of this approach). E.g., for $m = 3$,

$$-A^\top H(\pi) A = \begin{pmatrix} -\beta_1 \psi_1''(\pi_1) - \beta_3 \psi_3''(\pi_3) & -\beta_3 \psi_3''(\pi_3) \\ -\beta_3 \psi_3''(\pi_3) & -\beta_2 \psi_2''(\pi_2) - \beta_3 \psi_3''(\pi_3) \end{pmatrix}.$$

A necessary condition for positive semi-definiteness is that all principal minors (not only the leading ones) are non-negative (Prussing 1986). The $m-2$ and $m-3$ order principal minors, i.e., the determinants of submatrices that delete all but one and two rows and columns, give us for $i, j, k \in \{1, \ldots, m\}$ the following necessary conditions:

$$-\beta_i \psi_i''(\pi_i) - \beta_j \psi_j''(\pi_j) \geq 0 \tag{11a}$$

$$-\beta_i \psi_i''(\pi_i) - \beta_k \psi_k''(\pi_k) \geq 0 \tag{11b}$$

$$-\beta_j \psi_j''(\pi_j) - \beta_k \psi_k''(\pi_k) \geq 0 \tag{11c}$$

$$\beta_i \beta_j \psi_i''(\pi_i) \psi_j''(\pi_j) + \beta_i \beta_k \psi_i''(\pi_i) \psi_k''(\pi_k) + \beta_j \beta_k \psi_j''(\pi_j) \psi_k''(\pi_k) \geq 0. \tag{11d}$$

We do not require the conditions for higher-order principal minors for the theory developed here.

From the conditions in (11), we get that $\beta_i \psi_i''(\pi_i) + \beta_j \psi_j''(\pi_j) \leq 0$ for any set of sites $i, j \in E_1(\pi, \lambda) \cup E_2(\pi, \lambda)$, $i \neq j$. It follows that, for an optimal solution π^*, no more than one of $\psi_i''(\pi_i^*)$ for $i \in E_1^* \cup E_2^*$ can be positive, i.e., only one site can be a member of E_2^*. In addition, the member of E_2^* needs to have the smallest second derivative in absolute value. $\square$

Proof of Proposition 4. We first show that there exists an optimal allocation π^* such that either $E_1^* = \emptyset$ or $E_2^* = \emptyset$ with a symmetric demand function. First, let π^* be a solution to (1) such that $j \in E_2^*$ (there is at most one member of E_2 by Proposition 3). If E_1^* contains more than one element, we can show that this solution is dominated by an allocation where all or some capacity of site j is allocated to sites in E_1^*. Let π^{**} be such that $\pi_j^{**} = \pi_j^* - \epsilon$ and $\pi_i^{**} = \pi_i^* + \epsilon / |E_1^*| \; \forall i \in E_1^*$, where $|E_1^*|$ is the number of elements in E_1^* and $\epsilon > 0$ is small enough not to make any new inequality constraints binding. Because of the symmetry, it can easily be shown that $\Psi(\pi^{**}) > \Psi(\pi^*)$, implying that if $E_2 \neq \emptyset$ then E_1^* can have at most one member. If k denotes a unique element of E_1^*, any solution that transfers some capacity from site j to k (or vice-versa) attains the same objective value. Now, let π^{**} be such that either $\pi_k^{**} = T$ or $\pi_j^{**} = F$: $\pi_j^{**} = \pi_j^* - \min\{T - \pi_k^*, \, \pi_j^* - F\}$ and $\pi_k^{**} = \pi_k^* + \min\{T - \pi_k^*, \, \pi_j^* - F\}$. It follows that we can transform any optimal solution to one where either $E_1^* = \emptyset$ or $E_2^* = \emptyset$. Moreover, from part (c) in Proposition 3 follows that either $E_1^* = \emptyset$ or $G^* = \emptyset$. Thus, when considering all the possible allocations into the four sets, the two alternatives in the proposition will always yield an optimal solution. In (i), we have E_1^* and L^* non-empty, while E_2^* and G^* are empty. In (ii), we have E_1^* empty and at most one site in E_2^*. The expressions for x^* and x^{**} are such that they maximize the respective configuration. $\square$

Proof of Proposition 5. (a) follows from the same argument used in the proof of Proposition 4: any optimal allocation with E_1^* and E_2^* nonempty can be transformed into an allocation with at least the same rewards and one of the two being empty.

To prove (b), let π^* be a solution such that $i \in L^*$ and $j \notin L^*$. For such an optimal solution to exist, the following condition has to hold: $\pi_j^0 + \pi_j^* \geq \pi_i^0 + \pi_i^*$. This condition ensures that site j is further along the curve; otherwise, because $i, j \in C_2$, we could transfer $\epsilon > 0$ units from site j to i and obtain a better allocation. Now, consider the alternative allocation π^{**} such that $\pi_j^{**} = \pi_i^0 - \pi_j^0 + \pi_i^*$, $\pi_i^{**} = \pi_j^0 - \pi_i^0$, and $\pi_k^{**} = \pi_k^*$ for all $k \neq i, j$. It is easy to check that $\psi_i(\pi_i^{**}) = \psi_j(\pi_j^*)$ and $\psi_j(\pi_j^{**}) = \psi_i(\pi_i^*)$ which implies that π^{**} attains the same objective as π^*. Thus, π^{**} is also optimal, and we have shown existence. The same argument does not hold if $\pi_i^0 < \pi_j^0$ because the transformation would violate the constraints.

To prove (c), let π^* be an optimal solution such that $i \in L^*$ and $j \notin L^*$. Consider the alternative allocation π^{**} such that $\pi_j^{**} = F$ and $\pi_i^{**} = \pi_j^*$ and $\pi_k^{**} = \pi_k^*$ for all $k \neq i, j$. Because $i, j \in C_1$, we can easily check that π^{**} obtains a higher reward. Thus, we reach a contradiction.

To prove (d), suppose to the contrary that π^* is an optimal solution such that $i \in E_1^*$, $j \notin L^*$, and $\pi_j^0 + \pi_j^* > \pi_i^0 + \pi_i^*$. Consider the alternative allocation π^{**} such that $\pi_j^{**} = \pi_j^* - \epsilon$ and $\pi_i^{**} = \pi_i^* + \epsilon$ for $0 < \epsilon < \pi_j^* - \pi_i^*$, and $\pi_k^{**} = \pi_k^*$ for all $k \neq i, j$. Since $\pi_j^0 + \pi_j^*$ and $\pi_i^0 + \pi_i^*$ lie in the concave region, allocation π^{**} dominates π^*. Because π^* is an optimal allocation, we reach a contradiction. $\square$

Proof of Corollary 1. Given the constructed order of set N, Proposition 5 guarantees the existence of an optimal allocation π^* so that there exist k^* and l^* such that $\pi_i^* > F$ if $i \in \{1, \ldots, l^*\} \cup \{|C_2|+1, \ldots, |C_2|+k^*\}$, and $\pi_i = F$ otherwise. Moreover, either $|E_2^*| = 1$ and $|E_1^*| = 0$, or $|E_2^*| = 0$. For the former, (8) describes the only possible optimal allocation. For the latter, we show that (7) is the optimal allocation with Lagrange multiplier λ^*. The KKT conditions require that all members of the set E_1^* satisfy $\pi_i^* = \hat{\pi}(\lambda^*) - \pi_i^0$ for some λ^*. Moreover, part (d) of Proposition 5 shows that for all sites $i \in G^*$, we have that $T \leq \hat{\pi}(\lambda^*) - \pi_i^0$. Thus, for all $i \leq l^*$, we have that $\pi_i^* = \min\{\hat{\pi}(\lambda^*) - \pi_i^0, T\}$. Similarly, for sites $i \in C_1$, $i \leq |C_2| + k^*$, we have that $\pi_i^* = \min\{\hat{\pi}(\lambda^*) - \pi_i^0, T\}$. In addition, for $i > |C_2| + k^*$, we know from Proposition 3 that $\beta\psi_i(F) \leq \lambda^*$ and, as $i \in C_1$, this implies that $\hat{\pi}(\lambda^*) - \pi_i^0 \leq F$. Therefore, for any $i \in C_1$, $\pi_i^* = \min\{\max\{\hat{\pi}(\lambda^*) - \pi_i^0, F\}, T\}$. $\square$

Appendix C: Formal results related to Section 4.4

PROPOSITION 6. $\frac{\partial Y_i'(\pi_i^*)}{\partial m_i} > 0$.

Proof. It is easy to see that $\frac{\partial Y_i'(\pi_i^*)}{\partial m_i} = \frac{e^{(p_i+q_i)\pi_i^*}(p_i+q_i)^2 K}{\left[(m_i-c_i)q_i+(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\right]^3}$, with a K such that $\lim_{\pi_i^* \to 0} K = m_i(p_i+q_i)(m_i^2p_i+c_i^2q_i) > 0$ and $\frac{\partial K}{\partial \pi_i^*} = e^{(p_i+q_i)\pi_i^*}(p_i+q_i)(m_ip_i+c_iq_i)(m_i^2p_i+2m_ic_iq_i-c_i^2q_i)$. Note that $m_i^2p_i+2m_ic_iq_i-c_i^2q_i$ is concave in c_i, with the maximum at $c_i = m_i$. Here, it takes the value $m_i^2(p_i+q_i) > 0$. It follows that K is increasing in π_i^* and, because $K > 0$ holds for $\pi_i^* = 0$, it is always positive. $\square$

PROPOSITION 7. *$\frac{\partial Y_i'(\pi_i^*)}{\partial c_i} > 0$ if $\pi_i^* < \pi_i^{infl}$ and $\frac{\partial Y_i'(\pi_i^*)}{\partial c_i} < 0$ if $\pi_i^* > \pi_i^{infl}$.*

Proof. Assume that $\pi_i^{\text{infl}} > 0$. With some algebra, it follows that $-\frac{\partial Y_i'(\pi_i^*)}{\partial c_i} = \frac{e^{(p_i+q_i)\pi_i^*} m_i^2 (p_i+q_i)^3 \left[-(m_i-c_i)q_i+(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\right]}{\left[(m_i-c_i)q_i+(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\right]^3}$. But $-(m_i - c_i)q_i + (m_ip_i + c_iq_i)e^{(p_i+q_i)\pi_i^*} > 0 \Leftrightarrow \pi_i^* > \frac{\log\left(\frac{(m_i-c_i)q_i}{m_ip_i+c_iq_i}\right)}{p_i+q_i} \Leftrightarrow \pi_i^* > \pi_i^{\text{infl}}$. If $\pi_i^{\text{infl}} = 0$, the result follows directly. □

PROPOSITION 8. *$\frac{\partial Y_i'(\pi_i^*)}{\partial p_i} > 0$ if $\pi_i^* < \tilde{\pi}_i$ and $\frac{\partial Y_i'(\pi_i^*)}{\partial p_i} < 0$ if $\pi_i^* > \tilde{\pi}_i$ for some $\tilde{\pi}_i > \pi_i^{infl}$.*

Proof. It is easy to see that $\frac{\partial Y_i'(\pi_i^*)}{\partial p_i} = \frac{e^{(p_i+q_i)\pi_i^*} m_i(m_i-c_i)(p_i+q_i)K}{\left[(m_i-c_i)q_i+(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\right]^3}$, with a K such that $\lim_{\pi_i^*\to 0} K = m_i^2(p_i+q_i)^2 > 0$, $\lim_{\pi_i^*\to\pi_i^{\text{infl}}} = 4(m_i-c_i)q_i(m_ip_i+c_iq_i) > 0$ (assuming $\pi_i^{\text{infl}} > 0$), $\lim_{\pi_i^*\to\infty} = -\infty$, and $\frac{\partial K}{\partial \pi_i^*} = -(p_i+q_i)(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\left[(m_i-c_i)\left(e^{(p_i+q_i)\pi_i^*}-1\right) + (p_i+q_i)\pi_i^*(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\right] < 0$. □

PROPOSITION 9. *$\frac{\partial Y_i'(\pi_i^*)}{\partial q_i} > 0$ if $\pi_i^* < \tilde{\pi}_i$ and $\frac{\partial Y_i'(\pi_i^*)}{\partial q_i} < 0$ if $\pi_i^* > \tilde{\pi}_i$ for some $\tilde{\pi}_i$, where $\tilde{\pi}_i \geq \pi_i^{infl}$ if $q_i \geq p_i$ or $\pi_i^{infl} = 0$, and $\tilde{\pi}_i < \pi_i^{infl}$ otherwise.*

Proof. With some algebra, it follows that $\frac{\partial Y_i'(\pi_i^*)}{\partial q_i} = \frac{e^{(p_i+q_i)\pi_i^*} m_i(m_i-c_i)(p_i+q_i)K}{\left[(m_i-c_i)q_i+(m_ip_i+c_iq_i)e^{(p_i+q_i)\pi_i^*}\right]^3}$, with a K such that $\lim_{\pi_i^*\to 0} K = m_ic_i(p_i+q_i)^2 > 0$, $\lim_{\pi_i^*\to\infty} = -\infty$, and $\frac{\partial^2 K}{\partial \pi_i^{*2}} = -e^{(p_i+q_i)\pi_i^*}(p_i+q_i)^3(m_ip_i + c_iq_i)\left[c_i + \pi_i^*(m_ip_i+c_iq_i)\right] < 0$. Moreover, it holds that $\lim_{\pi_i^*\to 0}\frac{\partial K}{\partial \pi_i^*} = -(m_i-c_i)(p_i+q_i)^2(m_ip_i+c_iq_i)$, so it must hold that $\frac{\partial K}{\partial \pi_i^*} < 0$ everywhere. Additionally, assuming $\pi_i^{\text{infl}} > 0$, $\lim_{\pi_i^*\to\pi_i^{\text{infl}}} K = 2(m_i-c_i)(q_i-p_i)(m_ip_i + c_iq_i)$, which is non-negative if and only if $q_i \geq p_i$. □

PROPOSITION 10. *Suppose that the inequality constraints are not binding for sites $1,\ldots,\hat{n}$ and that all sites $i < \hat{n}$ are members of E_1^*. Then, $\frac{\partial \pi_1^*}{\partial \beta_1} > 0$, $\frac{\partial \pi_i^*}{\partial \beta_1} > 0$ for $j \leq \hat{n}, j \neq i$, and $\frac{\partial \lambda^*}{\partial \beta_1} > 0$.*

Proof. Take derivatives with respect to β_1 on either side of the $\hat{n}+1$ equations $\beta_i\psi_i'(\pi_i^*) = \lambda^*\ \forall i = 1,\ldots,\hat{n}$, $\sum_{i=1}^{\hat{n}}\pi_i^* = \Pi - \sum_{i=\hat{n}+1}^{n}\pi_i^*$. The resulting system of equations is

$$\psi_1'(\pi_1^*) + \beta_1\psi_1''(\pi_1^*)\frac{\partial \pi_1^*}{\partial \beta_1} = \frac{\partial \lambda^*}{\partial \beta_1}, \beta_i\psi_i''(\pi_i^*)\frac{\partial \pi_i^*}{\partial \beta_1} = \frac{\partial \lambda^*}{\partial \beta_1}\ \forall i = 2,\ldots,\hat{n}, \sum_{i=1}^{\hat{n}}\frac{\partial \pi_i^*}{\partial \beta_1} = 0.$$

Solving this, we arrive at $\frac{\partial \pi_1^*}{\partial \beta_1} = -\frac{\sum_{j\neq 1, j\leq\hat{n}}\frac{1}{\beta_j\psi_j''(\pi_j^*)}}{\sum_{j\leq\hat{n}}\frac{1}{\beta_j\psi_j''(\pi_j^*)}}\frac{1}{\beta_1\psi_1''(\pi_1^*)}\psi_1'(\pi_1^*) > 0$, $\frac{\partial \pi_i^*}{\partial \beta_1} = \frac{\frac{1}{\beta_i\psi_i''(\pi_i^*)\beta_1\psi_1''(\pi_1^*)}}{\sum_{j\leq\hat{n}}\frac{1}{\beta_j\psi_j''(\pi_j^*)}}\psi_1'(\pi_1^*)$, for $i = 2,\ldots,\hat{n} < 0$, $\frac{\partial \lambda^*}{\partial \beta_1} = \frac{\frac{1}{\beta_1\psi_1''(\pi_1^*)}}{\sum_{j\leq\hat{n}}\frac{1}{\beta_j\psi_j''(\pi_j^*)}}\psi_1'(\pi_1^*) > 0$. □

Appendix D: Detailed description of the data matching process

Due to frequent changes to administrative regions at all levels and possible inaccuracies in the data, we combine multiple sources to establish a robust matching between the sites and the corresponding census and population information. We first enumerate the data sets used, the variables within those data sets required for matching and providing population and census information, as well as their source:

[1] MSI sites. Data: site name, team, district. Source: MSI internal

[2] Facility information. Data: site name, coordinates, parish, sub-county, county, district. Source: http://catalog.data.ug/dataset/health-centres-uganda

[3] Site information. Data: site name, coordinates. Source: OpenStreetMaps, verified with MSI

[4] Village shapes 2011. Data: shape, village, parish, sub-county, county, district. Source: http://catalog.data.ug/dataset/cc5656d0-13e8-4540-adcc-7d21378c75c1

[5] Sub-county shapes 2014. Data: shape, sub-county, county, district. Source: http://catalog.data.ug/dataset/uganda-subcounties-2014

[6] Uganda census 2014. Data: parish, sub-county, county, district, census information. Source: http://ghdx.healthdata.org/record/uganda-population-and-housing-census-2014

[7] Uganda population 2020 (predicted). Data: squares, population. Source: https://www.worldpop.org/geodata/summary?id=6446

We next describe the procedure. We attempt to match on complete names. If there are no matches, the closest strings (using the Levenshtein distance) are identified and manually controlled.

1. Clean [1].
   (a) Change of district to pre-2016 name (23 of Uganda's 135 districts were created during the span of observations by splitting off from extant districts, so we use the old districts)
   (b) Sites without district information: add district if sites can be matched based on name and team
2. Extend the list given in [2] with sites found in OpenStreetMaps or Google Maps. Match [2] and [3] to [1] based on site name and manually correct any cases without full matching.
3. Match [4] to [1].
   (a) Match if site name and village match, and districts identical.
   (b) Coordinates in [1]: only accept matches if coordinates within village shape or less than 15km from centroid. For multi-matches, accept that with smallest distance to village centroid.
   (c) No coordinates in [1]: if match unique, accept and use the village centroid as coordinates.
4. Match [5] to [1], if coordinates within sub-county shape.
5. Match [6] to [1].
   (a) Match on locality data from [2] if available and unique (in order of priority, at parish, sub-county, county, or district level).
   (b) Match on locality data from [3] if available and unique. If unique match with [4], use parish level.
   (c) Match census information at lowest level (90.97% / 5.52% / 0.13% of sites matched at parish / sub-county / county level). Remaining 3.38% without reliable location information dropped.
6. Match [7] to site clusters (referred to as "sites" throughout the paper).
   (a) All sites within a radius of 5 km from another site belong to the same cluster.
   (b) Record population of all squares in [7] within 8 km from any site in a cluster and corresponding smallest distance to the site ("radius").
   (c) For radii between 1 and 8 km, use a third-degree polynomial to estimate a cluster's population. Average (resp. minimum) adjusted R^2 is 0.9993 (resp. 0.9722).
   (d) Remove clusters with total population at radius 5 km above 95% quantile (focus on remote clients).

7. Compute cluster impact values.
   (a) For each cluster, compute a poverty score
      i. Regress district-level log GDP per capita (from Wang et al. 2019) on combinations of six district-level census variables. The best model achieves adjusted R^2 of 0.65.
      ii. Using site-specific census data, predict GDP per capita for each site's surrounding area.
      iii. Normalize prediction to $[0, 1]$ poverty score: poverty $\text{score}_i = \frac{\min_j \log \text{GDP per capita}_j}{\log \text{GDP per capita}_i}$.
   (b) For each cluster, compute an age score.
      i. Using site-specific census data, compute fraction of young persons (15–24 years old) out of population in reproductive age (15–49 years old).
      ii. Normalize to $[0, 1]$ age score: age $\text{score}_i = \frac{\min_j 1-\text{fraction young}_j}{1-\text{fraction young}_i}$.
   (c) Compute cluster impact $\beta_i = (\text{age score})_i(\text{poverty score})_i$.

Appendix E: Analysis of demographic changes

We verify that the population mix at a site remains relatively constant. We do not observe variation over time in the census data, as we only make use of the 2014 census, and the previous one was taken in 2002, judged to be too long ago for relevance. However, we observe the percentage of young clients, which we take as a proxy for the percentage of young persons in the census population, one of the components of the impact parameter β_i. In particular, the percentage of young clients per visit exhibits a standard deviation of 0.13 (resp. 0.12) for the average (resp. median) site. If the percentage of young clients were entirely random, the standard deviation should be approximately 0.5. For the percentage of young clients within the adopter population, the standard deviation is 0.16 (resp. 0.15) on average (resp. in the median).

We also verify that population growth is relatively constant between sites. We compare the relative population growth at each site between 2015 and 2020, using the WorldPop database, described in Section 5. The coefficient of variation of the empirical distribution is 0.55, so the variation can be considered low.