

Accepted Manuscript

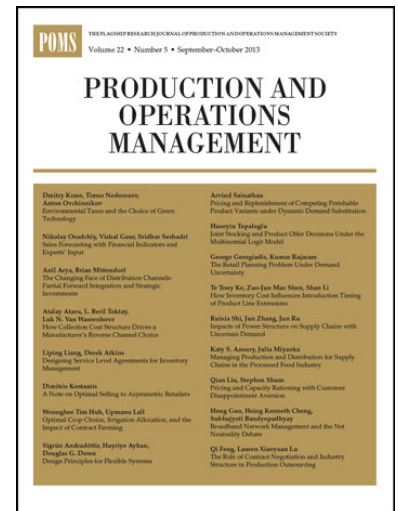
Title: Site Visit Frequency Policies for Mobile Family Planning Services

Authors: Harwin De Vries, Lisa E. Swinkels, Luk N. Van Wassenhove

DOI: <https://doi.org/doi:10.1111/poms.13484>

Reference: POMS 13484

To appear in: Production and Operations Management



Please cite this article as: De Vries Harwin.,et.al., Site Visit Frequency Policies for Mobile Family Planning Services. *Production and Operations Management* (2021), <https://doi.org/doi:10.1111/poms.13484>

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/poms.13484

Site Visit Frequency Policies for Mobile Family Planning Services

Harwin De Vries ^{*1}, Lisa E. Swinkels², and Luk N. Van Wassenhove²

¹Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands, harwin.devries@rsm.nl

²Technology and Operations Management, INSEAD, Boulevard de Constance, 77300 Fontainebleau, France

May 22, 2021

Abstract

Improving access to family planning services is key to achieving many of the United Nations sustainable development goals. To scale up access in remote areas and urban slums, many developing countries deploy mobile family planning teams that visit “outreach sites” several times per year. Visit frequencies have a significant effect on the total number of clients served and hence the impact of the outreach program. Using a large dataset of visits in Madagascar, Uganda and Zimbabwe, our study models the relationship between the number of clients seen during a visit and the time since the last visit and uses this model to analyse the characteristics of optimal frequencies. We use the latter to develop simple frequency policies for practical use, prove bounds on the worst-case optimality gap, and test the impact of the policies with a simulation model. Our main finding is that despite the complexity of the frequency optimisation problem, simple policies yield near-optimal results. This holds even when few data are available and when the relationship between client volume and the time since the last visit is misspecified or substantially biased. The simulation for Uganda shows a potential increase in client numbers of between 7% and 10%, which corresponds to more than 12,000 additional families to whom family planning services could be provided. Our results can assist policymakers in determining when to start data-driven frequency determination and which policies to implement.

Keywords: mobile health units; family planning; visit frequency; simple policies; worst-case bound; simulation

History: Received: May 2020; accepted: April 2021 by Subodha Kumar after two revisions.

*Corresponding author.

¹
This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/poms.13484](https://doi.org/10.1111/poms.13484)

This article is protected by copyright. All rights reserved

1 Introduction

Access to family planning plays a crucial role in achieving many of the United Nations (UN) sustainable development goals (Starbird et al., 2016), notably goal 3 (good health and well-being) by preventing maternal deaths and controlling the spacing of pregnancies. Universal access to contraception is estimated to reduce unintended pregnancies by 75%, maternal deaths by 25% (Darroch et al., 2017), and infant mortality by 10% (Cleland et al., 2006). Additionally, family planning allows women to postpone the birth of their first child and advance their education, which aids goals 4 (quality education), 1 (no poverty) and 5 (gender equality). For example, if adolescent girls in Brazil and India could postpone childbearing until their early twenties, the economic productivity of those countries would increase by more than US\$3.5 billion and US\$7.7 billion respectively (UNFPA, 2014). The United Nations estimates that, for “every dollar spent in family planning, between two and six dollars can be saved in interventions aimed at achieving other development goals” (UN Population Division, 2009).

Despite these compelling facts, for at least one out of five women¹ in Sub-Saharan Africa the need for family planning goes unmet (United Nations, 2015)². Many live in rural areas and urban slums, where access to family planning is limited or non-existent (Eva and Ngo, 2010; Solo and Bruce, 2010). Mobile (i.e., traveling) outreach teams are crucial to scale up access to family planning in these areas. Yet with a widening gap in funding of family planning organisations (UNFPA, 2018), mobile services have to *reach more people with fewer resources*.

Mobile teams, consisting of doctors and nurses, are often the only provider of family planning services and have to allocate time to the sites in the catchment area. Naturally, they tend to visit sites that attract many clients more often than those that attract relatively few. Marie Stopes International (MSI), an NGO with 500 mobile family planning teams, has therefore introduced the guideline that (besides other factors) “The level of client demand [...] should be used to determine the optimal frequency of visits” (MSI, 2016). Though this guideline is intuitive, it is loosely specified. Our interviews with outreach leads from five African countries show that country programs interpret these guidelines in different ways, and our data analyses reveal that adherence to them is low. This, as our study shows, has a negative impact on client volumes per outreach day, used in this study as the metric of effectiveness.

Developing more firmly specified policies is a necessary but difficult task. Policies not only

¹The data concern married or in-union women only.

²Women are defined to have an unmet need for family planning if they indicate that they want to stop or delay childbearing but are not using any method of contraception.

depend on client numbers but *affect* client numbers: More frequent site visits may lead to fewer clients per visit. In the absence of research on this effect, however, this has so far not been factored into decisions on visit frequency. Our study thus focuses on the Frequency Determination Problem (FDP):

Given a team, a set of sites, a lower and an upper bound on the frequency of visits to each site, and a certain number of service delivery days per month, what frequencies should be assigned to each site so as to maximise the effectiveness of the visits?

We use a large dataset from MSI to fit a fixed-effects model that captures the link between effectiveness and visit frequency. We then mathematically model FDP and analyse this model to obtain characteristics of optimal frequencies. Though the latter can be efficiently calculated by algorithms, there are strong arguments for basing frequency decisions on simple policies. We therefore use the characteristics of optimal frequencies to develop practical frequency policies and test them using a simulation of MSI's mobile teams in Uganda. Our findings were discussed with staff at MSI headquarters in London and with the Marie Stopes Uganda team.

The main finding is that, despite the complexity of the frequency optimisation problem, simple visit frequency policies yield near-optimal results. This holds even when little data is available, parameter estimates are substantially biased, and the objective function is misspecified, and even though several simplifying assumptions are made. Based on the data for Marie Stopes Uganda, we estimate that putting these policies in place could increase client numbers by between 7% and 10%. If this were representative of MSI as a whole, an increase of 7% would correspond to at least 175,000 clients globally³. In a context of severely reduced funding, such an increase in effectiveness would go some way to achieving the UN sustainable development goals.

Our paper is organised as follows. Section 2 gives further details on the way mobile teams operate and the FDP. Section 3 discusses the relevant literature. In Section 4, we develop an econometric model for client numbers to examine the effect of visit frequencies on the number of clients per visit. In Section 5, we formally model and analyse the FDP and propose exact solution methods and policies. Section 6 presents numerical results on the effectiveness of the policies for Marie Stopes Uganda.

2 Problem Description

This section describes the frequency determination problem (FDP) faced by outreach teams. This was obtained through discussions with outreach leads from five countries in which MSI

³This number is based on 500 teams who work at least 200 days a year and see on average 25 clients per day.

operates, the international outreach lead and the directors of the research team.

Each outreach team has a fixed set of sites which it visits regularly to deliver family planning services, determined in cooperation with the government. These are often small healthcare facilities that do not provide (long-term) family planning methods. They also use pop-up tents. Teams provide short-term and long-term family planning methods as well as counselling, treatment of side effects, and implant and intrauterine device (IUD) removal. Some also offer permanent methods of contraception (i.e., sterilisation).

Each team has a limited number of days per month to visit sites. Nearly all site visits last a full day. Sometimes two neighbouring sites can be visited on the same day. In most countries the visits are day trips that start and end at the team's base, but they may camp on site to reach remote locations on multi-day trips. Before the visit, marketing or "demand generation" is done through radio, posters or community health workers to encourage people to come. During the visit, drivers sometimes tour the village with a megaphone to announce that the team is on site. Since demand generation typically occurs locally and travel times between sites in rural areas are typically long, clients rarely come from different sites.

Client volumes per site visit are determined by factors such as the intensity of demand generation, staffing, the service package, whether a market-day or vaccination campaign is taking place, the time of arrival/departure, and visit frequencies. We focus on optimising the latter. MSI recognises its importance in its guideline "The level of client demand [...] should be used to determine the optimal frequency of visits" (MSI, 2016). Similarly, USAID's "handbook for program planners" for mobile outreach services mentions selecting "frequency of mobile outreach services" as the second program decision (after selecting sites to be visited) and states that the optimal frequency "will depend on demand for services" (Solo and Bruce, 2010).

Visit frequencies determine the "return time", or the time between visits. They should take into account ethical and medical considerations, since it is neither fair nor medically advisable to have long return times. Clients may want to have a contraceptive device replaced or removed or require treatment for a side effect. MSI has set an interval between visits of no more than six months. They can be flexible on return times only if government providers can assure removal of contraceptive devices and provide access to short-term methods such as condoms. If not, they aim for a return time of three months at most. Our analyses assume a six-month upper bound. There is a minimum interval between visits as it takes about a month to generate sufficient demand and coordinate the visit with those managing the local facilities.

These bounds leave room to optimise visit frequencies. Although it makes sense to visit sites that attract many clients more often, this has not yet been translated into a concrete policy. Should a site that attracts twice as many clients be visited twice as often? What would higher visit frequencies look like? How far should it depend on historic frequencies? Questions such as these are presently unanswered.

Perhaps unsurprisingly, many teams link their return times to demand in a relatively ad-hoc way. Data from Marie Stopes Uganda show that for those sites with return times below the 6-month upper bound, there is little to no alignment with client numbers (as illustrated in Figure 1). There are quite a few sites with a long return time but a large number of clients. Conversely, many sites have low average client numbers but are visited frequently. Shifting resources from sites with short return times and low average client numbers to those with longer return times and higher numbers would seem likely to improve effectiveness (measured as the average client volume per outreach day).

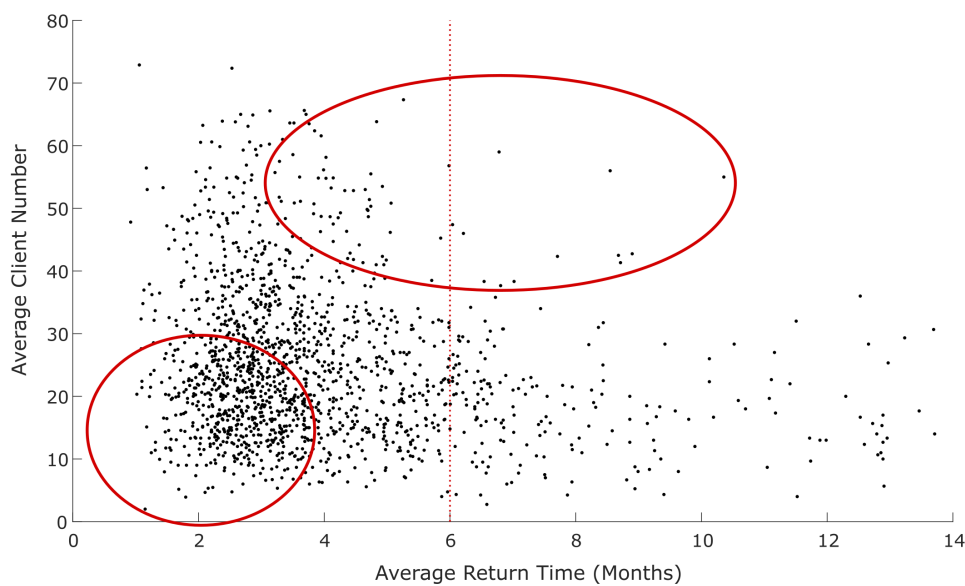


Figure 1: Current alignment of return time with demand. Each dot represents a site. The bottom left circle highlights sites with a low client volume and low average return time. The other circle highlights sites with a high client volume and high average return time. Shifting resources from the former to the latter would seem likely to improve effectiveness.

To optimise frequency policies it is key to understand how return times affect client volumes. Our conversations with MSI staff suggest that client numbers tend to increase with an increase in return time, which was attributed to more time for demand-generation, word-of-mouth effects, and existing clients demanding contraceptive device renewals or removals. However, the strength of this effect may be limited. For example, adequate demand-generation can result in substantial

client volumes irrespective of the return time. Long return times were said to reduce client satisfaction and induce a loss of trust in the teams, which also suggests that the effect diminishes at some point. More specific evidence on the effect of return time on client numbers is lacking (an issue addressed in Section 4).

The effectiveness of simple policies as opposed to advanced algorithms is also unclear. Despite the complexity of the FDP, organisations such as MSI have good reason to prefer simple policies, the main one being that they fit with the prevailing decentralised decision-making culture. Another is that advanced algorithms push up the costs of implementation, roll-out, and training (De Vries and Van Wassenhove, 2020). Organisations also value flexibility, i.e., policies that simply recommend a frequency and leave the teams to decide on day-to-day planning. For this reason, the policies proposed in this paper only recommend frequencies, allowing each team to incorporate local knowledge into its decision making. For example, MSI team leaders know when it is market day in a village, which tends to lead to higher numbers of clients. According to Peter Schaffler, MSI's outreach lead, "Team leaders are also extremely aware of road conditions and incorporate this knowledge into their routing decisions" (personal communication, May 1, 2018). Section 6 analyses the optimality gap of visit frequency policies.

A third question considered is how the effectiveness of a policy depends on the amount of data available. MSI started systematic data collection in Uganda in 2015. Other countries started more recently, so few data points are available. This can be perceived as a barrier to data-driven decision making. We analyse the impact of data quantity in Section 6.

Based on discussions with MSI, we have chosen to analyse policies in terms of numbers of clients per outreach day. One alternative measure would be to use the total couple years of protection (CYPs) provided per outreach day. The CYPs of a particular contraceptive method is the number of years that it protects a couple from an unwanted pregnancy. Another measure could be weighted client numbers depending on specific target groups: the young, the poor, and clients who have no other way to access family planning. This would take into account that seeing 20 clients in a relatively wealthy, well-served site is not the same as seeing the same number of clients at a poorer, remote, underserved site. The main reason for our choice (clients per outreach day) is that it fits the way outreach teams presently make decisions. The policies we encountered during our interviews were all expressed in terms of client numbers. MSI Madagascar, for example, aims to carry out monthly visits for sites that attract on average more than 20 clients per visit. Maximising client volumes also fits well with maximising CYPs and

weighted client volumes, since differences between sites in the same catchment area in terms of either CYPs or the percentage of clients in each target group are small. In Section 6.2, we examine how the performance of our policies is affected by this choice.

3 Literature

Three streams of literature are closely related to our work: mobile (family planning) outreach, global health operations management, and related scheduling problems. The FDP most closely resembles a knapsack problem (KP) with a separable concave non-linear objective function and linear constraints (see Ibaraki and Katoh, 1988, for an overview of solution algorithms). In particular, the FDP can be modelled as a KP where all items have the same weight, the objective value is a non-linear function of the number of a given item included (i.e., the visit frequency), and this number is bounded from above and below. The specific objective function of the FDP is what makes this problem distinctive.

Mobile (Family Planning) Outreach. Access to a health service in hard-to-reach areas can be enhanced by a.o. scaling up telehealth/eHealth programmes (see, e.g., Medlife in India), optimizing the match of supply and demand (see, e.g., platforms such as StanPlus), expanding and strengthening the network of *static health units*, and deploying mobile health units (MHUs). The latter are the primary source of healthcare services for millions of people (Oriol et al., 2009), both in high-income areas and low- and middle-income countries. They deliver primary care, maternal and newborn care, family planning, tuberculosis screening and treatment, HIV screening and treatment, cancer screening, dental services, treatment of childhood pneumonia, and many other services (Khanna and Narula, 2016). Several studies have shown MHUs to be capable of improving health, reducing treatment delay, improving treatment adherence, decreasing mortality, and decreasing healthcare costs (Khanna and Narula, 2016).

Mobile family planning units have been shown to significantly affect contraceptive use. Using data from Zambia, White and Speizer (2007) estimate that the modern contraceptive prevalence rate in rural areas would increase by 5.9 percentage points if all women had at least one outreach visit. Joshi and Schultz (2013) study the impact of introducing outreach services in villages in Bangladesh, and show substantial benefits in terms of birth spacing and fertility rates in comparison to “control villages”. Similarly, Lutalo et al. (2010) use a randomised controlled trial in Uganda to show that introducing outreach services increases the use of hormonal contraceptives and decreases pregnancy rates. Outreach has also been shown to increase adoption of long-term family planning methods and reach poor and underserved populations (Ngo et al., 2017; Noccio

and Reichwein, 2013).

Access to and impact of family planning is commonly measured through indicators such as (modern) contraceptive prevalence, unmet need, the number of unintended pregnancies, unsafe abortions, and maternal deaths averted, and adult birth rates (cf. FP2020, 2020). Since it is hard and time-consuming to measure site-level changes in these indicators, mobile outreach programs commonly use service statistics such as the number of (young) clients served, the number of adopters (i.e., clients who (re)started using contraception), and CYPs (cf. Ngo et al., 2017). Weinberger et al. (2013) propose a model to translate these statistics into an estimated impact on contraceptive prevalence.

Though a significant body of research has covered *impact* of mobile family planning, little is known about *impact drivers*. This paper contributes to the literature on mobile (family planning) outreach by exploring how visit frequencies affect client volumes and how to incorporate this in visit frequency policies.

Global Health Operations Management. Our work also contributes to the growing stream of research that uses OR/MS knowledge, models, and techniques to address global health challenges (Kraiselburd and Yadav, 2013). One branch considers enhancing access to *health products* by studying, e.g., root causes of stockouts (Karamshetty et al., 2021; Karimi et al., 2021), supply network (re)design (Vledder et al., 2019), inventory policy design (Leung et al., 2016), funding systems (Taylor and Xiao, 2014; Gallien et al., 2017), inventory allocation (Natarajan and Swaminathan, 2017), distribution (Parvin et al., 2018) and procurement (Martin et al., 2020). A second considers improving access to *health services*. Examples include studies of network design for clinics (De Vries et al., 2020), community healthcare (Cherkesly et al., 2019), and diagnostics (Jónasson et al., 2017; Deo and Sohoni, 2015), incentive design for patients and providers (Mehrotra and Natarajan, 2020), and patient enrolment/ capacity allocation (McCoy and Eric Johnson, 2014),

We identified five OR/MS papers that specifically consider the deployment of mobile healthcare units when capacity is restricted. Deo et al. (2013) study mobile asthma care units that visit schools, taking the timing of visits as given, and optimise the scheduling of patients to maximise health gains. Our work is similar in that we also consider (a proxy of) outcomes, but differs in that we consider optimising the frequency of visits. Hodgson et al. (1998) and Doerner et al. (2007) do consider the scheduling of site visits, and model this as a covering tour problem (CTP). The aim is to construct a tour through a subset of locations in a network, subject to the constraint

that a pre-specified percentage of the population (or the locations considered) must be within a given distance from a location in the tour. There are two structural differences between the CTP and the FDP. First, a CTP does not consider optimising visit frequencies, nor how demand is dependent on frequency. Second, the objective of a CTP is to minimise travel time, whereas we aim to maximise the number of clients reached in a setting where outreach teams use one-day trips to visit sites.

McCoy and Lee (2014) consider the allocation of motorbike visits to outreach sites, with the objective being to maximise effectiveness and fairness. Each site visit is said to satisfy a certain “need.” Effectiveness is defined as the total “need” satisfied. Fairness is incorporated by assuming diminishing utility of additional visits to a site. The authors quantify these objectives as a weighted sum of the number of visits to each site, raised to the power of some constant. As detailed later, the functional form of our objective function is different and fits with the data from MSI. Unlike them, we also consider lower and upper bounds on the number of visits. We focus on policies and insight into the effectiveness of family planning operations, whereas they focus mainly on analytical insights into optimal allocation decisions.

De Vries et al. (2021) analyse the scheduling problem for mobile teams in medical screening. The problem clusters villages that can be easily incorporated into a daily schedule in terms of the travelling distance between them. It then assigns each mobile team to one cluster in each planning period. A key difference is that they consider how to optimise schedules (i.e., which sites to visit in which month), whereas our work considers the more tactical question of how to design frequency policies. They consider a curve describing how health evolves between visits; their objective is to minimise the total expected disease burden over the planning horizon, i.e., the area under the curve. In contrast, we consider the specific value of the curve for each visit, namely the number of clients. As a consequence, the functional form of the objective function differs. As in this paper, De Vries et al. (2021) consider both optimal solution approaches and planning rules.

Related Scheduling Problems. Two other problems are mathematically closely related to the FDP. Chemical product line scheduling problems deal with one or multiple parallel machines that process multiple products (see Floudas and Lin (2004) for a review). The problem is to determine the frequency of production for each product and the length of each production period so as to minimise inventory and set-up costs. Despite obvious differences between the concepts, this can be seen as similar to the FDP, if machines are seen as teams and products as sites.

Frequency decisions affect the objective function, as applies to the FDP. The corresponding functional form of the objective function is, however, substantially different.

Maintenance scheduling problems (MSPs) consider a fixed number of machines to be maintained and a maximum number of maintenance visits in a given period of time (see Nicolai and Dekker (2008) for a review). If we substitute machines for sites and maintenance visits for site visits, the similarities are clear. However, the objective function is substantially different. In addition, our focus is on the development of simple policies, whereas the MSP literature typically considers exact approaches.

4 Data Description and Analysis

This section describes the data used in this paper and analyses the effect of return time (the time since the last visit) on client numbers. The results are used in the modelling of the FDP in Section 5 and in the case study in Section 6.

4.1 Data Description

We use data on outreach visits from MSI. MSI is exceptional in the sense that it is one of the few humanitarian organisations to systematically capture data on operations and use it to improve policies, guidelines, and operations.

We use data from Marie Stopes Uganda from May 2015 to September 2017, from Marie Stopes Madagascar from January 2016 to March 2018, and from Marie Stopes Zimbabwe from January 2017 to September 2017. These countries were chosen by MSI because the data were of good quality. The start dates correspond to the start of systematic data gathering and the end dates to the time that access to the data was terminated. The cleaned data sets for Uganda, Madagascar and Zimbabwe include 10,293 visits to 1,581 sites, 10,498 visits to 1,794 sites and 732 visits to 243 sites respectively. For each site visit, the data sets show the name of the site, the date, the team, and a list of clients. For each client, they list among others the age bracket, number of children, and the family planning method. We use this data to determine the total client volume and the return time in months for each site visit. Descriptive statistics are presented in Table 1.

For each site, the first visit included in the data is excluded, because the return time is unknown. We also exclude all sites that have only one visit for which return time is known. This excludes 23%, 46% and 71% of the sites for Uganda, Madagascar and Zimbabwe respectively. With the Uganda data, we exclude visits where multiple teams go to the same site. These visits represent special events, such as youth focus days, which result in unusually high numbers of clients. Some exceptionally high numbers are also observed in Zimbabwe. Based on our conversations

with MSI staff, we set the cut-off value for the number of clients that can realistically be seen by one team to 120. We exclude the 3.5% of visits that exceed this threshold. From the Madagascar data we exclude the 0.4% of visits where no family planning methods are provided, because these are special events such as AIDS prevention or vaccination campaigns.

Table 1: Descriptive statistics of the datasets

Variable	Madagascar			Uganda			Zimbabwe		
	Avg	Med	Std	Avg	Med	Std	Avg	Med	Std
# visits per site:	6.8	6.0	4.3	7.5	7.0	3.9	4.0	4.0	1.6
Client volume per visit:	23.9	22.0	12.1	27.0	23.0	18.1	40.7	35.0	26.3
Return time:	2.7	2.1	2.2	3.1	2.5	2.4	2.0	2.0	1.1

4.2 Effect of Return Time on Client Numbers

In this section we aim to model the effect of the return time for visit v to site s , RT_{vs} , on the client number for visit v to site s , CL_{vs} . Specifically, we want to discover the shape and magnitude of the effect and the distribution of the error terms. We use the following general function to model CL_{vs} :

$$CL_{vs} = m_s f(RT_{vs}) + \varepsilon_{vs} \quad (1)$$

Here ε_{vs} denotes an error term, which captures visit-specific factors for which accurate data are lacking (e.g., weather, marketing efforts, and the start time of a visit). Parameter m_s represents the expected client volume obtained in site s for some “baseline return time” which we set at one month (without loss of generality; we return to this in Online Appendix C). Function $f(RT_{vs})$ represents a return-time-dependent multiplier for the baseline client volume. For example, $f(3)=1.2$ expresses that, when the return time is three months, the expected client volume is 20% higher than the expected client volume for the baseline return time.

As stressed during conversations with MSI outreach leads for Madagascar, Tanzania and Sierra Leone and with HQ staff (see Section 2), client numbers are determined by many factors that are not related to return time, such as the population density of the area and marketing efforts. We therefore split $f(RT_{vs})$ into a constant part and a return-time-dependent part:

$$CL_{vs} = m_s(\alpha + \beta g(RT_{vs})) + \varepsilon_{vs} \quad (2)$$

Here, α and β are non-negative constants. We estimate both parameters for Uganda, Madagascar and Zimbabwe. The larger β , the more client numbers grow in return time, possibly

because of word of mouth, community health workers, or renewals. α will indicate the percentage of clients reached when $g(RT_{vs}) = 0$. We consider three functional forms for $g(RT_{vs})$: RT_{vs} , $\log(RT_{vs})$ and $\sqrt{RT_{vs}}$. In line with input obtained from MSI staff, all forms can model client numbers as a strictly increasing function in return time. The logarithmic and square root functions model this increase to be diminishing over time.

The formulation reflects the implicit assumption that client numbers for different sites are mutually independent. As the distances between sites are large and use of health services tends to drop off sharply the farther clients have to travel, this assumption is realistic (see, for example, Tanser et al. (2006)). Another implicit assumption is that the effect of return time is proportional to m_s . The results presented in Online Appendix B suggest that this assumption is realistic.

Appendix A describes how we fit function (2) and how we deal with endogeneity (site characteristics affect return times) and heteroskedasticity (the variation in client volumes is larger for sites that attract more clients). In short, to handle endogeneity in return times, we first estimate β by relating within-site variations in return times to within-site variations in client volumes. Plugging this estimate into function (2), we estimate α and m_s using weighted least squares (WLS) with weights m_s to reduce heterogeneity. Since fitting m_s requires solving a complex non-linear, non-convex optimisation problem, we test two heuristic procedures in Online Appendix C, including using average historical client volumes (multiplied by a constant) as an estimate. We show that different procedures have hardly any effect on the resulting estimates and we therefore use the average historical client volumes as estimates for m_s in the next sections. Section 6.2 shows that the main insights we present later are insensitive to biased parameter estimates.

4.3 Results

Estimates for α and β . Table 2 describes the estimated values of α and β for each function $g(\cdot)$. The effect of return time on client numbers is significant and positive for all the countries considered. Surprisingly, the strength of this effect appears to be rather limited. For example, the “logarithmic model” estimates that increasing the return time from one month to six months increases the expected client volume by 6%, 35% and 27% for Madagascar, Uganda and Zimbabwe respectively. In Uganda and Zimbabwe the client numbers “build up” more strongly over time than in Madagascar. This could be because in Uganda and Zimbabwe more of the clients come to see the outreach team because of word of mouth or encouragement from community health workers, which takes time. In all three countries, however, a significant part of the client volume

is not related to demand that “builds up” over time. The MSI staff to whom we presented these results, clarified that there is a large unmet need in many sites, so attracting clients is largely a matter of reaching this population through proper “demand generation.” These results have important implications for visit frequencies, as we explore in Section 6.

Table 2: Estimated WLS parameters (standard error between parentheses), root mean squared error of calibration (RMSEC), mean absolute error (MAE), and root mean squared error of prediction (RMSEP) for multiple functional forms

Country	Parameter	$g(RT_{vs})$		
		$\log(RT_{vs})$	$\sqrt{RT_{vs}}$	RT_{vs}
Madagascar	$\hat{\alpha}$	0.9811 (0.0034)	0.9127 (0.0034)	0.9631 (0.0034)
	$\hat{\beta}$	0.0306 (0.0045)	0.0582 (0.0076)	0.0142 (0.0021)
	RMSEC	1.7158	1.7158	1.7158
	MAE	6.4980	6.4950	6.4958
	RMSEP	2.3811	2.3855	2.3873
Uganda	$\hat{\alpha}$	0.8475 (0.0042)	0.6506 (0.0042)	0.8651 (0.0042)
	$\hat{\beta}$	0.1665 (0.0082)	0.2104 (0.0103)	0.0444 (0.0025)
	RMSEC	2.1978	2.2012	2.2018
	MAE	8.9229	8.9361	8.9415
	RMSEP	2.8513	2.8530	2.8434
Zimbabwe	$\hat{\alpha}$	0.9580 (0.0131)	0.4661 (0.0132)	0.6878 (0.0133)
	$\hat{\beta}$	0.1424 (0.0164)	0.4174 (0.0392)	0.1699 (0.0162)
	RMSEC	2.2565	2.2815	2.2868
	MAE	11.0476	11.3363	11.4061
	RMSEP	3.7574	3.8280	3.9022

All parameters are significant at a 1% significance level.

Comparison of functions $g(RT_{vs})$. The root mean square errors of calibration (RMSEC) presented in Table 2 are based on the residuals resulting from the weighted least squares regression, calculated as in (3). We note that this metric does not differ greatly between the different functional forms $g(RT_{vs})$. For Madagascar, the three models perform almost equally well. The RMSECs corresponding to the “logarithmic model” are slightly lower for Uganda and Zimbabwe. We observe the same when considering the mean *absolute* (unweighted) error (MAE), calculated as the absolute difference between predicted and actual client numbers (i.e., calculated as $\sqrt{m_s}e_{vs}$, with e_{vs} as defined in (3)). The table also presents the (weighted) root mean square error of *prediction* (RMSEP) based on out-of-sample predictions. Specifically, the models were calibrated using all data except the last visit for each site. The fitted model was subsequently used to predict the client volume for the last visit. The RMSEP for the logarithmic model is lowest for each country, suggesting that this function provides the best overall fit among the data sets considered.

$$e_{vs} = \frac{CL_{vs}}{\sqrt{m_s}} - \sqrt{m_s} \cdot (\hat{\alpha} + \hat{\beta} \cdot g(RT_{vs})) \quad (3)$$

Error term distribution. To simulate the performance of various visit frequency policies (see Section 6), insight into the distribution of e_{vs} is required. Using the distribution fitter of MATLAB version R2015b, we find that the shifted log normal distribution fits the residuals e_{vs} most closely. Figure 2 shows the fit on the residuals for Uganda for $g(RT_{vs}) = \log(RT_{vs})$. The calibrated parameter values are: $\mu = 1.955$, $\sigma = 0.310$, and shift = -7.4 .

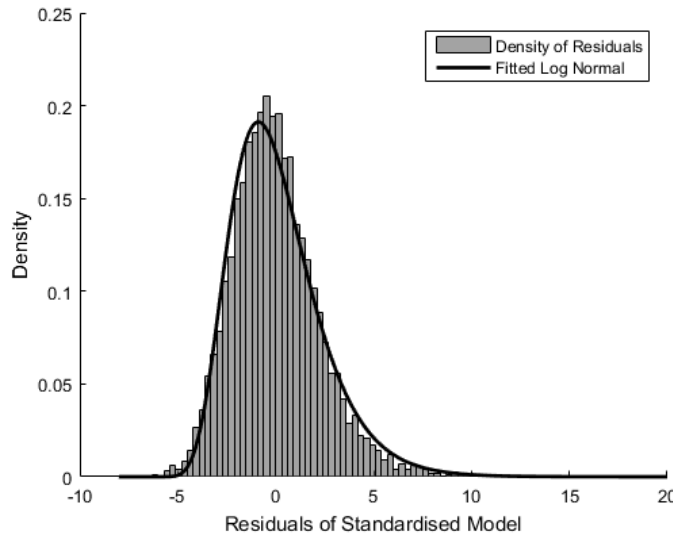


Figure 2: Shifted log normal distribution fitted to errors for Uganda

We can summarise our findings as follows. First, the results indicate that the effect of return time on client numbers is statistically significant and positive for all the countries considered. Second, return time has a relatively limited effect on client numbers. Finally, we showed the error terms to approximately follow a log-normal distribution. The relations identified in this section will be used in the modeling of the FDP in Section 5 and in the case study in Section 6.

5 Model and Solution Methods

This section has three objectives: to derive insights on determinants of optimal visit frequencies, to develop methods to determine optimal frequencies, and to explain various (heuristic) policies for choosing visit frequencies. In Section 5.1 we formally model the FDP. Section 5.2 develops algorithms to solve the model and Section 5.3 analyses the resulting frequencies. In Section 5.4 we use insights from the algorithms developed in Section 5.3 to develop several frequency policies.

5.1 Model

The model considers one outreach team and a set of outreach sites \mathcal{S} for which it is responsible. Decision variable y_s represents the number of full-day visits to site s per month. The assumption of full day visits is generally met, because travel times in developing countries are long. Two small sites close together that are always combined in a single visit can be incorporated into the model as one artificial site. We optimise over y_s only and leave out the exact planning of which site to visit on which day. We make this simplification because our work is about the strategic/tactical decision to choose a policy for visit frequencies. Furthermore, this modeling choice suits the purpose of this research, which is to gain insights into optimal solutions. These insights can provide decision makers with intuition for (near) optimal decisions. Such intuitions are more difficult to obtain from the output of an optimisation run of a complex model.

We assume that the y_s visits per month are evenly spread over time such that the return time in months, RT_s , is $1/y_s$. Marie Stopes International is increasingly focusing on enhancing schedule regularity, as evidenced, for example, by this statement from MSI's "Success Model" handbook: *A regular schedule increases the community's trust in the care that they will receive, especially for those who rely on MSI for the provision of STMs [short term methods] and follow up care.* It also considers implementing tools and the policies presented in this work, which would likely also decrease variations in return times. There will nevertheless remain reasons to choose a return time that differs from $1/y_s$. Examples include bad road conditions, team vacations, and non-availability of the site. Section 6.2 therefore examines how our results are affected when return times are stochastic. Since return times can be formulated in days, y_s is in practice close to continuous. We model y_s as such. Parameter D represents the total number of full day site visits that can be scheduled in a month. In line with Section 4, we represent the relationship between the expected number of clients at a site s and y_s by function $m_s \cdot (\alpha + \beta \cdot g(1/y_s))$ with a site-specific multiplier m_s . The number of visits to site s is constrained by an upper bound, UB_s , and a lower bound, LB_s . As explained in Section 2, these bounds ensure that the return time is large enough to adequately prepare for the visit and small enough to be medically and ethically acceptable. For MSI, $LB_s = 1/3$ (once per 3 months) when government providers can assure removals of contraceptive devices and provide access to short-term methods such as condoms, and $LB_s = 1/6$ otherwise.

The FDP can then be modeled as in model (M1). The objective (4a) maximises the total expected number of clients reached in a month. Capacity constraint (4b) restricts the total

number of visits, and (4c) and (4d) are boundary constraints.

$$(M1) \quad \text{maximise} \quad \sum_{s \in \mathcal{S}} y_s m_s (\alpha + \beta g(1/y_s)) \quad (4a)$$

$$\text{subject to} \quad \sum_{s \in \mathcal{S}} y_s \leq D \quad (4b)$$

$$y_s \leq UB_s \quad \forall s \in \mathcal{S} \quad (4c)$$

$$y_s \geq LB_s \quad \forall s \in \mathcal{S} \quad (4d)$$

Problem (M1) is a non-linear simple allocation problem with boundary constraints (Derman, 1959). Throughout this paper we consider the typical case when $\sum_{s \in \mathcal{S}} LB_s < D$ and $\sum_{s \in \mathcal{S}} UB_s > D$. Otherwise, the problem is infeasible or trivial. In the following sections, we analyse this model for different functions $g(1/y_s)$.

5.2 Solution Algorithms

In this section we explain the algorithms used to solve (M1) to optimality for multiple functions $g(1/y_s)$ and analyse the characteristics of optimal frequencies of each functional form. Section 5.2.1 considers the simplest case, where client numbers are independent of return time: $g(1/y_s) = 0$. This may be realistic when services are relatively unknown (i.e., there is no natural build-up of demand) and where site-specific factors such as population size and efforts to market the visit determine the number of clients. Section 5.2.2 considers the simplest increasing effect in which client numbers are linearly dependent on return time, i.e., $g(1/y_s) = 1/y_s$. Section 5.2.3 considers the nonlinear dependencies $g(1/y_s) = \log(1/y_s)$ and $g(1/y_s) = \sqrt{1/y_s}$.

5.2.1 Client Numbers Independent of Return Time

When client numbers are independent of return time and $g(1/y_s) = 0$, the problem becomes a knapsack problem with bounds. For this problem, the optimal solution can easily be found by a greedy algorithm (Kellerer et al., 2004). Online Appendix D describes this algorithm in the notation of this paper.

In short, we initially set all visit frequencies at the lower bound. That leaves $D - \sum_{s \in \mathcal{S}} LB_s$ days to allocate. In optimality, we assign as many days as possible to the sites with the highest m_s . We thus increase the visit frequencies of the ‘unsaturated’ sites for which m_s is highest to their upper bound until all available days are used. Consequently, at most one site will be assigned a frequency not equal to the lower or upper bound.

5.2.2 Client Numbers Linearly Dependent on Return Time

If client numbers are linearly dependent on return time, i.e., $g(1/y_s) = 1/y_s$, we can rewrite objective function (4a) as

$$\sum_{s \in \mathcal{S}} y_s m_s (\alpha + \beta \frac{1}{y_s}) = \sum_{s \in \mathcal{S}} \alpha m_s y_s + \sum_{s \in \mathcal{S}} \beta m_s$$

The first term equals the objective function for constant demand. The second term is constant and can be discarded in optimisation. What remains is an objective function of the form considered in Section 5.2.1. The same algorithm can thus be applied to find the optimal solution. There will therefore still be at most one site with an optimal frequency not equal to the lower or upper bound.

5.2.3 Client Numbers Non-Linearly Dependent on Return Time

Under a logarithmic dependency $g(1/y_s) = \log(1/y_s)$ and a square root dependency $g(1/y_s) = \sqrt{1/y_s}$ of client numbers on return time, the problem (M1) becomes non-linear and cannot be solved trivially.

Applying Lagrange relaxation leads to $y_s^{\log^*}(\lambda)$ and $y_s^{\text{sqrt}^*}(\lambda)$ as the optimal frequency for a given non-negative Lagrange multiplier λ for $g(1/y_s) = \log(1/y_s)$ and $g(1/y_s) = \sqrt{1/y_s}$ respectively.

$$y_s^{\log^*}(\lambda) = \max \left\{ LB_s, \min \left\{ UB_s, e^{\frac{m_s(\alpha-\beta)-\lambda}{m_s\beta}} \right\} \right\} \quad (5)$$

$$y_s^{\text{sqrt}^*}(\lambda) = \max \left\{ LB_s, \min \left\{ UB_s, \left(\frac{m_s\beta}{2(\lambda - m_s\alpha)} \right)^2 \right\} \right\} \quad (6)$$

We use the iterative procedure from Bretthauer and Shetty (1995) to compute λ^* , the optimal value of λ . Online Appendix D describes this algorithm in more detail. The optimal frequencies are then $y_s^{\log^*}(\lambda^*)$ and $y_s^{\text{sqrt}^*}(\lambda^*)$, which are not dichotomous. They can be both on the bounds or in between the bounds.

5.3 Numerical Insights into Optimal Visit Frequencies

The algorithms in Sections 5.2.1 and 5.2.2 provide a valuable insight: There is at most one site with an optimal frequency different from the lower or upper bound if $g(1/y_s) = 0$ or $g(1/y_s) = 1/y_s$. Under a non-linear dependency, such insights cannot be derived directly from the algorithm referred to in Section 5.2.3. In the remainder of this section, we use that algorithm to solve several small problem instances to gain insight into the characteristics of an optimal solution.

In particular, we assess to what extent optimal solutions resemble lower and upper bound frequencies and how this depends on context. Such insights are crucial for the development of frequency policies. We specifically consider three contextual factors. The first is the magnitude of the effect of return time on client numbers, i.e., the value of the parameter β . We compare optimal frequencies for the values of β fitted for Uganda and Madagascar in Section 4, since these values were the highest and lowest respectively. Second, we examine how frequencies are affected by the level of variation in m_s . We consider three instances. Instance 1 has 40 sites with $m_s = 2.5, 5, \dots, 100$. Instances 2 and 3 have the same number of sites but $m_s = 31, 32, \dots, 70$ and $m_s = 40.5, 41, \dots, 60$ respectively. Third, to study the effect of capacity on optimal frequencies, we consider what happens when $D = 20$ (240 days p. year) and $D = 16\frac{2}{3}$ (200 days p. year). Furthermore, we set $LB_s = 1/6$ and $UB_s = 1$ for all sites s , which corresponds to return times of between one and six months.

We compute the optimal frequencies for each instance using $g(1/y_s) = \log(1/y_s)$. Table 3 shows, for each instance, the number of frequencies on the lower bound, the upper bound, and in between. The results are similar for $g(1/y_s) = \sqrt{1/y_s}$, and the corresponding table is provided in Online Appendix D.

Table 3: Frequencies of Sites in Several Instances under $g(RT_{vs}) = \log(RT_{vs})$, $LB_s = LB$, and $UB_s = UB$

	Capacity	<i>Parameters for Uganda</i>			<i>Parameters for Madagascar</i>		
		<i># sites with frequency=</i>	<i>Between</i>	<i>UB</i>	<i># sites with frequency=</i>	<i>Between</i>	<i>UB</i>
		<i>LB</i>	<i>LB and UB</i>		<i>LB</i>	<i>LB and UB</i>	
Instance 1	$16\frac{2}{3}$	18	20	2	27	2	11
	20	15	17	8	23	2	15
Instance 2	$16\frac{2}{3}$	8	32	0	26	3	11
	20	5	35	0	22	3	15
Instance 3	$16\frac{2}{3}$	0	40	0	24	6	10
	20	0	40	0	20	6	14

Table 3 shows that for several sites the frequencies are *not* set to the bounds. This happens more frequently in Uganda than in Madagascar. The reason is that the value of β is higher for Uganda. The higher the value of β , the greater the increase in client numbers on the return visit. The model then “deviates” more from the constant model, where it is optimal to set frequencies to the bounds. Additionally, the numbers show that a larger variation in m_s , as in Instance 1, results in more frequencies on the bounds. To explain the rationale, let us consider a situation in which there is just enough capacity to serve one of two sites, s_1 and s_2 , with the upper bound

frequency and the other with the lower bound frequency. If baseline client volume $m_{s_1} = 20$ and $m_{s_2} = 80$, then it is intuitively clear that serving s_1 with the lower bound frequency and s_2 with the upper bound frequency is optimal. However when $m_{s_1} = 50$ and $m_{s_2} = 51$, assigning the lower bound frequency to s_1 and the upper bound frequency to s_2 is not optimal. The reason lies in the concave shape of $g(RT_{vs})$: Increasing the frequency of s_1 by Δ would lead to a relatively large increase in client volume, and decreasing the frequency of s_2 by Δ would lead to a relatively small decrease in client volume.

In Online Appendix E we also show that frequencies that are not on the bounds are about equally distributed across the interval $[LB, UB]$. That is, they are not concentrated around the bounds. This suggests that it is beneficial to develop frequency policies which allow for at least one frequency value between the bounds.

Finally, the number of sites for which the assigned frequency is not on the upper or lower bound does not seem to depend strongly on capacity. In all instances, a decrease in capacity of 40 days makes hardly any change to this. In short, under a logarithmic effect of return time on client numbers, the optimal frequencies seem to differ most strongly from the bounds when the variation in m_s is small and the effect of return time on client numbers is large.

5.4 Policies

In this section, we use insights from Section 5.3 to develop frequency policies. All our data-driven policies rely on an estimate of the site-specific multiplier m_s . We consider two ways to estimate this, which define two types of policy. The first type estimates m_s as the historical client average. This is simple and close to current practice, where teams tend to consider past client numbers informally – i.e., they do not correct for the impact that return times had on historical client numbers. We refer to these policies as ‘ignore return time’ (IRT) policies. The second type of policy does make this correction and we refer to these policies as ‘correct for return time’ (CRT) policies. CRT policies estimate m_s by \hat{m}_s as defined in (7). Here, \mathcal{V}_s is the set of the V_s previous visits to site s .

$$\hat{m}_s = \frac{1}{V_s} \sum_{v \in \mathcal{V}_s} \frac{CL_{vs}}{\hat{\alpha} + \hat{\beta} \cdot g(RT_{vs})} \quad (7)$$

This is an unbiased estimator that assumes client numbers CL_{vs} are determined by (2)⁴ CRT policies are considered more complex, because they require estimates for α and β and more complex calculations.

⁴This follows when dividing both sides from (2) by $(\alpha + \beta g(RT_{vs}))$, subtracting ε_{vs} from both sides, and taking the expectation on both sides.

Given a choice on how to estimate m_s , various frequency policies are possible. We next explain four of them in increasing order of complexity: the equal policy, two-category policy, three-category policy, and infinite category policy. The two-, three-, and infinite category policies can be used in combination with both estimates of m_s . For ease of exposition, we consider the case when all sites have the same lower bound $LB_s = LB$ and upper bound $UB_s = UB$. The policies can easily be adjusted to the case when bounds differ per site.

Equal Policy (EP). In this baseline policy, each team divides its monthly capacity of D days equally across all sites for which it is responsible. Each site is assigned the same frequency $D/|\mathcal{S}|$. This fractional frequency is transformed into return times.

Two-Category Policy (TwoCP). The previous section shows that where client numbers are not dependent on return time or have a linear connection to it, all frequencies are set to the bounds with at most one exception per team. Our numerical experiments show that many frequencies are also set to the bound when client numbers depend non-linearly on return time. This leads to the hypothesis that a TwoCP – a policy that specifies two frequency categories – could perform in a close to optimal way. In such a policy, sites are assigned to either the low-frequency category (LFC), i.e., $y_s = LB$, or the high-frequency category (HFC), i.e., $y_s = UB$. The policy initially assigns all sites to the LFC and reassigns them to the HFC in decreasing order of estimated m_s for as long as capacity allows. This may leave some spare capacity after the last reassignment. Assuming the upper bound is a hard bound, we shall increase the frequency of sites in the LFC to slightly above the lower bound so all capacity is fully utilised.

Three-Category Policy (Three-CP). Under non-linear dependencies of client numbers on return time, the optimal frequencies can deviate from the bounds, as shown in Section 5.3. It might therefore be desirable to include a middle-frequency category (MFC). Section 5.3 showed that the *number* of sites for which the optimal frequency is not on the bounds is strongly dependent on α and β and the variation of m_s over the sites. It is possible to optimise the size of the MFC and its corresponding frequency, but this would be a study in its own right. We therefore fix the frequencies corresponding to the three categories, and suggest that a national or team guideline should be issued on the percentage of sites to allocate to the MFC.

Such guideline on the percentage of sites in MFC can, for instance, be developed by solving model (3C), which models the problem of allocating sites to frequency categories. Here we define $\mathcal{C} = \{\text{LFC}, \text{MFC}, \text{HFC}\}$ as the set of categories and y_c as the frequency corresponding to category $c \in \mathcal{C}$. Furthermore, a_{sc} represents the expected number of clients for site s if this site

is in category c , which we estimate through formula (2). Binary variable x_{sc} denotes whether site s is in category c . In the case of team-specific guidelines on the percentage of sites in the MFC, model (3C) can be solved for each team. For national guidelines, organisations can solve (3C) for each team and use the aggregated percentage.

$$(3C) \quad \text{maximise} \quad \sum_{s \in \mathcal{S}} \sum_{c \in \mathcal{C}} y_c a_{sc} x_{sc} \quad (8a)$$

$$\text{subject to} \quad \sum_{c \in \mathcal{C}} x_{sc} = 1 \quad \forall s \in \mathcal{S} \quad (8b)$$

$$\sum_{s \in \mathcal{S}} \sum_{c \in \mathcal{C}} y_c x_{sc} \leq D \quad (8c)$$

$$x_{sc} \in \mathbb{B} \quad \forall s \in \mathcal{S}, c \in \mathcal{C} \quad (8d)$$

Given the percentage of sites per category, the assignment of sites to the HFC and then to the MFC and LFC is done in decreasing order of estimated m_s . Implementation is discussed in more detail in Section 6.

Infinite Category Policy (InfCP). The InfCP applies the exact algorithms from Section 5.2 using estimates for m_s . This policy can recommend all frequencies between the lower and upper bounds. It is of course more difficult to implement.

5.5 Optimality Gaps

We now analytically prove a bound on the optimality gap when using two frequency categories instead of infinitely many and then generalize this to the case when k frequency categories are distinguished. We specifically consider the realistic case when $g(RT_s)$ is concave and first derive the following lemma (see Appendix B for the proof):

Lemma 1. *Objective function 4a is concave in y_s for any function $g(RT_s)$ that is concave in RT_s .*

In addition, we consider the realistic scenario when visiting a site more frequently never leads to fewer clients in that site *in total* and state without proof (it follows when applying the product rule) that:

Lemma 2. *Objective function 4a is non-decreasing in y_s if and only if $\left(\alpha + \beta g\left(\frac{1}{y_s}\right)\right) \geq \frac{d}{dy_s} \beta g\left(\frac{1}{y_s}\right)$.*

We note that, for the fitted values of α and β presented in Section 4.3, this condition is always met.

Let us write the objective function again as $\sum_s m_s f(y_s)$, let $f'(y)$ denote the derivative of function $f(y)$, and let $\tilde{y}(\delta) = \{y : f'(y) = \delta\}$ – the point where the derivative equals δ . Furthermore, let n_c denote the optimal number of sites to be assigned to frequency category $c \in \{1, \dots, k\}$, y_c the corresponding frequency with $y_1 = LB$ and $y_k = UB$, and \mathcal{S}_c the set of sites assigned to category c . Assume that this leaves no spare capacity. Assume further that the decision maker knows the exact value of m_s and assigns the n_1 sites with the lowest value of m_s to the lowest frequency category, the n_2 remaining sites with the lowest value of m_s to the second lowest frequency category, etc. Proposition 3 states the corresponding optimality gap for the two-category policy.

Proposition 3. *The optimality gap of TwoCP, measured as the absolute difference with the optimal InfCP solution value, is bounded from above by $\sum_{s \in \mathcal{S}} m_s \Delta$, where Δ is calculated as $f(\tilde{y}(\delta)) + (y_{HFC} - \tilde{y}(\delta))\delta - f(y_{HFC})$ and $\delta = \frac{f(y_{HFC}) - f(y_{LFC})}{y_{HFC} - y_{LFC}}$.*

We refer to Appendix B for the proof, which utilizes the two conditions stated in Lemmas 1 and 2. Specifically, we replace the objective function by a linear “outer approximation”. The optimal solution is the same as the TwoCP solution and its value provides an upperbound on the optimal solution value. Taking the difference between this value and the “real” value of the TwoCP solution yields the optimality gap. We apply the same approach to generalize the result to the case when there are k frequency categories.

Proposition 4. *The optimality gap of a policy that distinguishes k frequency categories, measured as the absolute difference with the optimal InfCP solution value, is bounded from above by $\sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}_c} m_s \max_{c \in \mathcal{C}} \{\Delta_c\}$, where, Δ_c is calculated as $f(\tilde{y}(\delta_c)) + (y_c - \tilde{y}(\delta_c))\delta_c - f(y_c)$ and $\delta_c = \frac{f(y_{c+1}) - f(y_c)}{y_{c+1} - y_c}$ for $c < k$ and $\delta_k = \delta_{k-1}$.*

For the case study presented in the next section, with $f(y_s) = y_s(\alpha + \beta \log(1/y_s))$, $\alpha = 0.8475$ and $\beta = 0.1665$ (i.e., the fitted function for Uganda), we derive that this corresponds to a worst-case optimality gap of only 7.7% for TwoCP and 4.0% for ThreeCP, measured as a percentage of the InfCP solution value. This worst-case gap is not tight and is hence likely even smaller in reality. This is a very important result, since a large worst-case gap would likely be seen as a barrier in practice.

We do note that these results are obtained for the case when the decision maker knows the exact value of m_s , whereas in practice it must be estimated. The next section therefore examines the optimality gap numerically.

6 Case Study: Outreach Teams of Marie Stopes Uganda

In this section we study the performance of the policies proposed in Section 5.4, using Marie Stopes Uganda as a case study. We use real data from 10,293 visits to 1,581 sites in Uganda. We first explain the simulation model before presenting the results.

6.1 Simulation Model

The simulation model uses function (2) with the parameters α , β , m_s and the error distribution found in Section 4.3 to simulate client numbers. It assumes a logarithmic dependency of client numbers on return times, because Section 4.3 suggests that this functional form provides the best fit. Our simulations follow the outline depicted in Figure 3. They start by simulating n historical data points which the policies take as input. We change the value of n in our experiments to assess the impact of the amount of data available. For each historical data point, we first randomly draw a historical return time from our data for the corresponding site and use the error distribution fitted in Section 4.3 to draw an error. Function (2) transforms these return times and errors into simulated client volumes. In Step 2, the policies use these simulated client numbers to estimate m_s for each site s . Such estimates can differ from the value of m_s used by the simulation model – which we refer to as the “actual” value of m_s – due to stochasticity and the estimation method used by the policy. We refer to these cases as imperfect information, because m_s is estimated. For comparison reasons, we also consider the case of perfect information in which the actual value of m_s is known. In Step 3, the policies subsequently use these estimates to determine frequencies. Finally, we use function (2) in Step 4 to compute for each policy the expected average number of clients per visit when the frequencies recommended by the policy are implemented. To obtain a precise estimate of expected average client numbers, we repeat these steps 100 times. Appendix C explains each of the four steps in more detail.

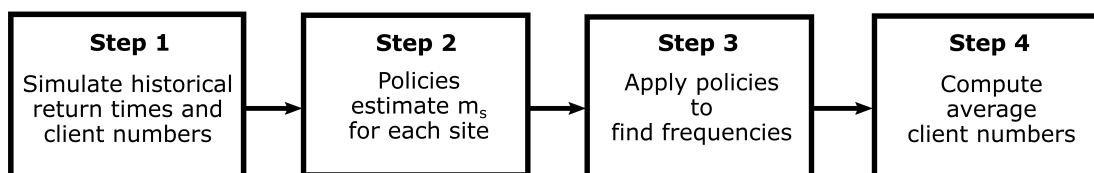


Figure 3: Overview of simulation

6.2 Simulation Results

We first assess the trade-off between the complexity and effectiveness of policies when historical return times are used in Step 1. Second, we analyse how the effectiveness of the policies depends on the amount of historical data available for each site. Third, we analyse the sensitivity of results to different simulated return times.

Comparison of Performance of Policies. Figure 4 shows the average number of clients per visit that can be served when applying the various policies for $n = 5$ as well as the current⁵ average number in the data. The policies that assign different frequencies to different sites clearly perform better than either the equal policy (EP) or the current situation. Interestingly, the simplest differentiation policy, the two-category policy with a simple estimate for m_s (TwoCP-IRT), already yields over 7% more clients than the current situation. Moving from two to three categories leads to another 1% increase. The effect of more complex policies is relatively small. Moving from three to an infinite number of categories results in a further increase of less than 0.4%. The increase in client volume achieved by correcting for historical return times when categorising sites :

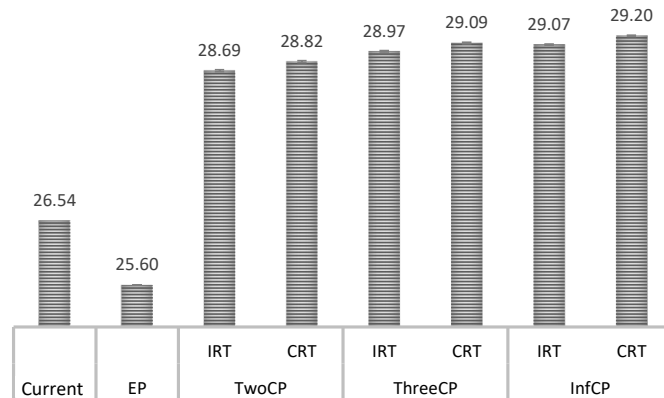


Figure 4: Average number of clients per visit of policies for $n = 5$, including 95% confidence interval (width < 0.03 for each policy)

Although the increases in client numbers might seem small, even a small increase per visit mean that, over the course of a year an enormous number of additional families can be provided with family planning services. With $D = 20$ and 24 teams, the increase in client numbers when moving from the current situation to TwoCP-IRT corresponds to an increase of over 12,000 clients a year in Uganda alone.

Impact of Number of Observations. Many countries currently have little historical data.

In Uganda, Zimbabwe, and Madagascar, for example, 36%, 91%, and 47% of the sites have five

⁵This current average client number per visit is the average across all client numbers in the *cleaned* data set. The official constraints in the current situation are somewhat less restrictive than those used in the models; in practice frequencies are currently not always within the bounds.

or less data points, respectively. Lack of data may induce reluctance to implement the proposed data-driven methodology. We therefore examine the impact of n , the number of observations available to estimate m_s , on the effectiveness of policies. Figure 5 plots the average client number for $n \in \{1, 2, 5, 10, 20, 50\}$. Equal polices are not data-driven and therefore not influenced by limited data availability. For all other policies, effectiveness increases with n , but this increase is small when n is large. With $n = 50$, all policies perform to within 0.6% of the number of clients obtained when there is perfect information.

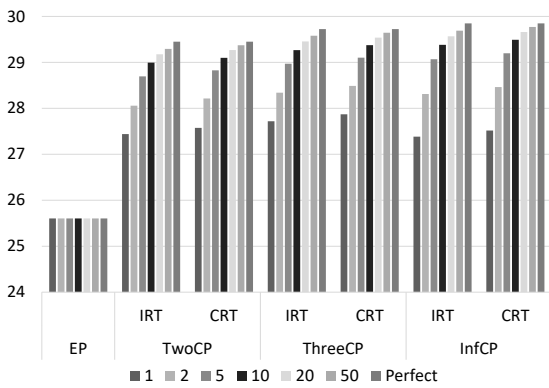


Figure 5: Average number of clients per visit for multiple values of n and all policies

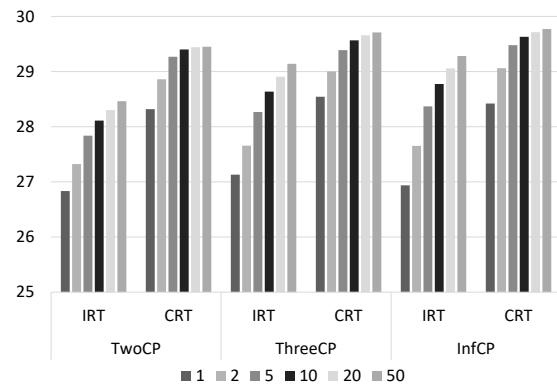


Figure 6: Average number of clients per visit for multiple values of n and all policies when using policy return times in Step 1

Even with $n = 1$, all policies outperform the EP by at least 3%. The ThreeCP consistently performs around 1% better than the TwoCP. It is interesting that the InfCP performs worse than the TwoCP and ThreeCP if $n = 1$. The policy appears to assign frequencies that are too specific, based on too little information. The InfCP only outperforms the ThreeCP when $n \geq 5$. With perfect information, the TwoCP results in 11% more clients than is currently the case, the ThreeCP in close to 1% more clients than the TwoCP, and the InfCP in less than 0.5% more than the ThreeCP. The effect of correcting for return time diminishes with n from around 0.5% when $n = 1$ to around 0.25% when $n = 50$.

The importance of correcting historical client volume for return times. We previously used historical return times in the first step of the simulation. Once policies are in use, the return times will be very different. We therefore also examine how the policies perform when in use. Instead of sampling from actual historical return times in step 1 of the simulation model, we consider the case when return times were optimised by the policy itself⁶. All other simulation steps remain the same. Figure 6 shows the resulting average client numbers per visit.

⁶That is, the return times suggested by the policy when the actual value of m_s is known.

When $n = 1$, the policies that include return times now reach over 5% more clients than their counterparts that ignore return times. Even for $n = 50$, the TwoCP-CRT, ThreeCP-CRT and InfCP-CRT perform better than their counterparts by 3.5%, 2.0% and 1.7% respectively.

These results can be explained by the observation that the variation in client numbers will decrease when mobile teams start using frequency policies. This happens because sites with a high m_s are visited often, resulting in relatively low client numbers, and the opposite happens for sites with a low m_s . This leads to a bias in m_s estimates used in IRT policies: When \hat{m}_s is based solely on client numbers, the baseline client volume m_s for sites that are visited very frequently will be underestimated and the volume for sites that are visited much less often will be overestimated. The importance of correcting historical client volume for return times therefore increases when frequencies are already data-driven.

6.3 Sensitivity Analyses

Our analyses and policies implicitly make several simplifying assumptions. This section summarises the results of extensive sensitivity analyses presented in Online Appendix G.

Return Time Stochasticity. Our evaluation of frequency policies implicitly assumes that each site is visited with a fixed time interval. As explained in Section 5.1, return times are stochastic in reality, which may bias our evaluation. Our results, however, show that this bias is very small and that it hardly affects the optimality gap⁷.

Client Characteristics. Family planning organisations perceive a.o. young clients and adopters as key target groups while our policies assign equal weight to each client. Our results show that optimality gaps are hardly affected when the “real” weight of young clients and adopters is significantly higher. The explanation is that differences in the percentage of young clients and adopters among a given team’s sites are typically minor.

Biased Progression Rates. The strength of demand progression, captured by parameter β , has to be estimated from historical data. Our analyses show that the policies are robust to significant overestimation of β . The explanation is that the policies set many frequencies to the bounds, which is optimal when $\beta = 0$. The policies are also insensitive to substantial *underestimations* of β . TwoCP appears more sensitive than the other policies. The reason is that, when β is large, it becomes relatively attractive to set many frequencies between the bounds (see Section 5.3).

⁷I.e., the gap with the exact InfCP solution value, measured as a percentage of the latter

Heterogeneous Progression Rates. Our models implicitly assume that parameter β is the same across all sites. Due to differences in client characteristics, however, demand progression may differ per site. We therefore examine the performance of our policies when demand progression is represented by site-specific constant β_s . As expected, optimality gaps increase with the amount of heterogeneity in β_s , but this increase is rather minor, even when the level of heterogeneity is substantial. The reason appears to be that, for many sites with a high (low) value baseline client volume, even a substantially biased estimate of m_s correctly identifies the site as one to be assigned to the high frequency category (low frequency category).

Functional Form of $g(RT_s)$. Section 4.3 suggests minor differences in the fit of various functional forms for the relationship between client volumes per visit and return time. Our results show that the performance of our policies is hardly impacted when the “real” functional form differs from the assumed one (logarithmic). The explanation appears to be that, for the fitted values of α and β , the “distance” between the logarithmic function and the other functions is small.

7 Conclusion and Discussion

Mobile outreach teams are crucial to scale up access to family planning in rural areas and urban slums. Gaps in funding mean that mobile teams have to *reach more people with fewer resources*, without sacrificing the quality of the service. This paper contributes to this endeavour by studying the frequency determination problem (FDP) or how to determine site visit frequencies for mobile outreach teams in order to maximise client volume per outreach day. Visit frequencies have an upper and lower bound based on medical, ethical and planning concerns and are further constrained by the number of days each mobile team is available for site visits. We specifically consider the FDP in the context of family planning and use data from Marie Stopes International (MSI).

The complexity of the FDP lies in how the frequency of site visits affects client numbers, specifically through the return time (i.e., the time since the last visit). We use MSI data for three different countries to analyse the strength of this effect. We find that the client numbers are higher when the return time is longer, but that the magnitude of this effect differs across the countries considered. For example, we estimate that increasing the return time from one month to six months increases the expected client volume by 6%, 35% and 27% for Madagascar, Uganda and Zimbabwe respectively.

We examine how optimal frequencies depend on the shape and magnitude of the effect of

return time on client numbers. When client numbers are non-linearly dependent on return times (we consider a logarithmic and square root shape), the optimal frequencies can be largely on the bounds, depending on various factors. They deviate widely from the bounds when the effect of return time on client numbers is large – i.e., when demand progresses strongly over time – such that the objective function differs most strongly from the linear case. Deviations from bounds are also more frequent when there is only a small variation in the so-called baseline client volume – the expected client volume for a given baseline return time. In extreme cases, when all sites have the same baseline client volume or when demand increases very strongly with the return time, it can be optimal to set none of the frequencies to the bounds. These results have important implications for outreach program managers. First, they show that in none of the realistic functional forms and parameter settings considered it is optimal to only have one frequency category (i.e., to assign the same frequency to each site). Second, they indicate when it is beneficial to have three or more frequency categories: when demand progresses strongly over time and when there are large variations in baseline client volumes across sites.

We used these insights to develop simple frequency policies that use only two (TwoCP) or three (ThreeCP) different frequencies and present analytical results on their worst-case performance for the case when decision makers know the exact value of the baseline value of each site. For our case study on Marie Stopes Uganda, they reveal a worst-case optimality gap of 7.7% for TwoCP and 4.0% for ThreeCP. This is an important finding, as a large optimality cap could be seen as a barrier to adoption of the policies.

We also tested the policies' effectiveness in an extensive simulation analysis for Marie Stopes Uganda. The main takeaway for outreach program managers is that the effectiveness of mobile teams can increase substantially when visit frequencies are based on historical client numbers. We estimate that the expected number of clients can increase by over 7% compared to the current number when using policies that base visit frequencies on historic client numbers. An increase on this scale would correspond to more than 12,000 additional families that could be provided with family planning services over a year in Uganda alone. Though reaching this number does not require dropping sites altogether, it does require decreasing visit frequencies for sites with few clients – a difficult decision to make.

Interestingly, we find that simple planning policies yield solutions that are close to those of exact methods. For example, the results from a policy with two possible frequencies are less than 2% below those obtained using exact methods. This is important, because such a policy fits

well with the decentralised decision making culture in humanitarian organisations and because the implementation, training, and maintenance costs of advanced tools are high (De Vries and Van Wassenhove, 2020). A simple policy might thus be more cost-effective. It would also allow local contextual knowledge and experience to be incorporated into the decisions, and help to safeguard the teams' autonomy.

A second finding is that the effectiveness of policies depends strongly on the variation in historical return times. We analysed policies that correct historical client volumes for return times, and found that when there is little variation in historical return times, these perform less than 0.5% better than policies that do not correct for this. The latter are easier to implement as they only require the outreach team to extract the average historical client volume (i.e, no additional calculations). In addition, such policy resembles the informal decision rules observed in practice. This suggests that the “costs” of correcting historical client volumes for return times outweigh the benefits when there is little variation in historical return times. However, when the variation increases, which happens once data-driven frequency policies are implemented, correcting for return times can increase expected client numbers by more than 5%⁸. This, then, calls upon outreach program managers to develop ways to facilitate such correction. It could be done by developing a simple spreadsheet model, integrating this in existing software, or through a correction table. Additionally, awareness of this matter can be raised in training sessions.

Third, we find that data-driven frequency determination performs well even when little historical data is available. Even if data on *one* site visit is available for each site, all data-driven policies outperform current practice. This has important implications for the uptake of this approach to determine visit frequencies, because limited data availability is often seen as a barrier in practice.

As explained, a strength of our policies is that they are easy to adopt and require little data. Implicitly, they thereby make several simplifying assumptions: 1) the strength of demand progression (captured by β) is the same across sites, 2) young clients and adopters have the same weight as other clients, and 3) return times are constant. Our sensitivity analyses show that the performance of the policies is hardly affected by these assumptions. We also show that performance is robust to over- or underestimating demand progression and misspecifying the functional form of the relationship between return time and client volume. These findings are important for practitioners, as they reduce/take away potential barriers for implementation.

⁸N.B., these percentages were obtained for a specific organisation and country and might differ in other situations or contexts.

A limitation of this paper is that we assume that the part of the client numbers *not* affected by return time will be constant over time, regardless of the policy. This is realistic in the short and medium term, but in the long term there may be fewer clients in need of family planning around a site that is visited often, because the needs of many have been met in earlier visits. Reversely, the very young populations in developing countries could lead to large cohorts entering their reproductive lives and hence increase the need. Other aspects of a policy besides visit frequency can influence long-term client numbers. One is trust in mobile teams. In all the policies considered in this study, a site is visited with the same frequency over time. The resulting predictability of future visits increases trust in the organisation and thus client numbers. Future research is needed to analyse the magnitude of such long-term effects. Follow-up research is also needed to study the relationship between return times and health or inconvenience. Our model captures this relationship in a dichotomous manner: a return time is either medically and ethically acceptable or not. Future empirical research could develop more precise models and subsequently test the performance of guidelines that solely consider client volumes.

Teams may have reasons to deviate from the recommended return times on some occasions. The weather, accessibility, security, a market day or vaccination day, for example, may prompt them to visit a site sooner or later than recommended by the policy. Visits also need to be planned with the outreach sites (often local health facilities), which can impose additional constraints. This clearly has an impact on performance. In future research, we hope to pilot our policies and assess them empirically, thereby closing the modelling cycle. If the pilot confirms our findings and leads to a global roll-out of the policy, it could substantially increase the number of woman with access to family planning. As mentioned, one dollar invested in family planning is estimated to save between two and six dollars in interventions aimed at meeting other SDGs. Despite convincing statistics, however, there continue to be major funding gaps. Our study exemplifies how knowledge and tools from the operations management discipline can help to scale up access without additional investments, and can thus aid progress towards many of the UN sustainable development goals.

Acknowledgements

This work was supported by grants from Sorbonne Université, Kurt Björklund MBA'96J Research Fund, and The Patrick Cescau/Unilever Endowed Fund. We would like to thank all staff from Marie Stopes for their support and close cooperation on this research. We also thank the anonymous reviewers and the editor for the helpful comments on earlier drafts of the manuscript.

References

- Boyd, S., Boyd, S.P., Vandenberghe, L., 2004. Convex optimization. Cambridge university press.
- Bretthauer, K., Shetty, B., 1995. The nonlinear resource allocation problem. *Operations Research* 43, 670–683.
- Cherkesly, M., Rancourt, M.È., Smilowitz, K.R., 2019. Community healthcare network in underserved areas: Design, mathematical models, and analysis. *Production and Operations Management* 28, 1716–1734.
- Cleland, J., Bernstein, S., Ezeh, A., Faundes, A., Glasier, A., Innis, J., 2006. Family planning: the unfinished agenda. *The Lancet* 368, 1810–1827.
- Darroch, J., Sully, E., Biddlecom, A., 2017. Adding it up: Investing in contraception and maternal and newborn health, 2017—supplementary tables. New York, NY: The Guttmacher Institute .
- De Vries, H., Van de Klundert, J.J., Wagelmans, A.P., 2020. The roadside healthcare facility location problem: A managerial network design challenge. *Production and Operations Management* 29, 1165–1187.
- De Vries, H., Van Wassenhove, L.N., 2020. Do optimization models for humanitarian operations need a paradigm shift? *Production and Operations Management* 29, 55–61.
- De Vries, H., Wagelmans, A., Van de Klundert, J., 2021. Towards elimination of infectious diseases with mobile screening teams: HAT in the DRC. *Production and Operations Management* .
- Deo, S., Iravani, S., Jiang, T., Smilowitz, K., Samuelson, S., 2013. Improving access to community-based chronic care through improved capacity allocation. *Operations Research* 61, 1259–1462.
- Deo, S., Sohoni, M., 2015. Optimal decentralization of early infant diagnosis of HIV in resource-limited settings. *Manufacturing & Service Operations Management* 17, 191–207.
- Derman, C., 1959. A simple allocation problem. *Management Science* 5, 453–459.
- Doerner, K., Focke, A., Gutjahr, W., 2007. Multicriteria tour planning for mobile healthcare facilities in a developing country. *European Journal of Operational Research* 179, 1078–1096.

- Eva, G., Ngo, T., 2010. MSI mobile outreach services: retrospective evaluations from Ethiopia, Myanmar, Pakistan, Sierra Leone, and Viet Nam. London: Marie Stopes International.
- Floudas, C., Lin, X., 2004. Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review. *Computers & Chemical Engineering* 28, 2109–2129.
- FP2020, 2020. FP2020 core indicators. http://www.track20.org/pages/data_analysis/core_indicators/overview.php. Accessed 2020-10-27.
- Gallien, J., Rashkova, I., Atun, R., Yadav, P., 2017. National drug stockout risks and the global fund disbursement process for procurement. *Production and Operations Management* 26, 997–1014.
- Hodgson, M., Laporte, G., Semet, F., 1998. A covering tour model for planning mobile health care facilities in Suhum District, Ghana. *Journal of Regional Science* 38, 621–638.
- Ibaraki, T., Katoh, N., 1988. *Resource Allocation Problems*. MIT Press, Cambridge, MA.
- Jónasson, J.O., Deo, S., Gallien, J., 2017. Improving HIV early infant diagnosis supply chains in sub-Saharan Africa: Models and application to Mozambique. *Operations Research* 65, 1479–1493.
- Joshi, S., Schultz, T.P., 2013. Family planning and women’s and children’s health: Long-term consequences of an outreach program in Matlab, Bangladesh. *Demography* 50, 149–180.
- Karamshetty, V., De Vries, H., Van Wassenhove, L.N., Dewilde, S., Minnaard, W., Ongarora, D., Abuga, K., Yadav, P., 2021. Inventory management practices in private healthcare facilities in Nairobi county. *Production and Operations Management* .
- Karimi, A., Mishra, A., Natarajan, K.V., Sinha, K.K., 2021. Managing commodity stock-outs in public health supply chains in developing countries: An empirical analysis. *Production and Operations Management* .
- Kellerer, H., Pferschy, U., Pisinger, D., 2004. Basic algorithmic concepts, in: *Knapsack problems*. Springer, pp. 15–42.
- Khanna, A.B., Narula, S.A., 2016. Mobile health units: mobilizing healthcare to reach unreachable. *International Journal of Healthcare Management* 9, 58–66.

- Kraiselburd, S., Yadav, P., 2013. Supply chains and global health: an imperative for bringing operations management scholarship into action. *Production and Operations Management* 22, 377–381.
- Leung, N.H.Z., Chen, A., Yadav, P., Gallien, J., 2016. The impact of inventory management on stock-outs of essential drugs in sub-Saharan Africa: secondary analysis of a field experiment in Zambia. *PloS One* 11, e0156026.
- Lutalo, T., Kigozi, G., Kimera, E., Serwadda, D., Wawer, M.J., Zabin, L.S., Gray, R.H., 2010. A randomized community trial of enhanced family planning outreach in Rakai, Uganda. *Studies in Family Planning* 41, 55–60.
- Martin, P., Gupta, D., Natarajan, K.V., 2020. Vaccine procurement contracts for developing countries. *Production and Operations Management* 29, 2601–2620.
- McCoy, J., Eric Johnson, M., 2014. Clinic capacity management: Planning treatment programs that incorporate adherence. *Production and Operations Management* 23, 1–18.
- McCoy, J., Lee, H., 2014. Using fairness models to improve equity in health delivery fleet management. *Production and Operations Management* 23, 965–977.
- Mehrotra, M., Natarajan, K.V., 2020. Value of combining patient and provider incentives in humanitarian health care service programs. *Production and Operations Management* 29, 571–594.
- MSI, 2016. Outreach Success Model V1.0, Chapter 4: Operational Excellence. Technical Report. London: Marie Stopes International.
- Natarajan, K.V., Swaminathan, J.M., 2017. Multi-treatment inventory allocation in humanitarian health settings under funding constraints. *Production and Operations Management* 26, 1015–1034.
- Ngo, T., Nuccio, O., Pereira, S., Footman, K., Reiss, K., 2017. Evaluating a LARC expansion program in 14 sub-Saharan African countries: a service delivery model for meeting FP2020 goals. *Maternal and Child Health Journal* 21, 1734–1743.
- Nicolai, R., Dekker, R., 2008. Optimal maintenance of multi-component systems: a review, in: *Complex system maintenance handbook*. Springer, pp. 263–286.

Noccio, O., Reichwein, B., 2013. Understanding clients and achieving FP2020 goals: Exit interviews help deliver client-centred services. London: Marie Stopes International.

Oriol, N.E., Cote, P.J., Vavasis, A.P., Bennet, J., DeLorenzo, D., Blanc, P., Kohane, I., 2009. Calculating the return on investment of mobile healthcare. *BMC Medicine* 7, 27.

Parvin, H., Beygi, S., Helm, J.E., Larson, P.S., Van Oyen, M.P., 2018. Distribution of medication considering information, transshipment, and clustering: Malaria in Malawi. *Production and Operations Management* 27, 774–797.

Solo, J., Bruce, L., 2010. Expanding contraceptive choice to the underserved through mobile outreach service delivery: A handbook for program planners. Washington: USAID .

Starbird, E., Norton, M., Marcus, R., 2016. Investing in family planning: key to achieving the sustainable development goals. *Global Health: Science and Practice* 4, 191–210.

Tanser, F., Gijsbertsen, B., Herbst, K., 2006. Modelling and understanding primary health care accessibility and utilization in rural South Africa: an exploration using a geographical information system. *Social Science & Medicine* 63, 691–705.

Taylor, T.A., Xiao, W., 2014. Subsidizing the distribution channel: Donor funding to improve the availability of malaria drugs. *Management Science* 60, 2461–2477.

UN Population Division, 2009. What would it take to accelerate fertility decline in the least developed countries? Technical Report. URL: <http://www.upc.apgdm.org/en/content/what-would-it-take-accelerate-fertility-decline-least-developed-countries>.

UNFPA, 2014. Population and poverty. Technical Report. URL: <http://www.unfpa.org/resources/population-and-poverty>.

UNFPA, 2018. Consolidating UNFPA Funding Architecture. Technical Report. URL: https://www.unfpa.org/sites/default/files/event-pdf/UNFPA_Thematic_Funds_-_Funding_Gaps.pdf.

United Nations, D., 2015. Trends in Contraceptive Use Worldwide 2015. Technical Report. URL: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesapd_report_2015_trends_contraceptive_use.pdf.

Vledder, M., Friedman, J., Sjöblom, M., Brown, T., Yadav, P., 2019. Improving supply chain for essential drugs in low-income countries: results from a large scale randomized experiment in Zambia. *Health Systems & Reform* 5, 158–177.

Weinberger, M.B., Fry, K., Boler, T., Hopkins, K., 2013. Estimating the contribution of a service delivery organisation to the national modern contraceptive prevalence rate: Marie Stopes International’s Impact 2 model. *BMC Public Health* 13, S5.

White, J.S., Speizer, I.S., 2007. Can family planning outreach bridge the urban-rural divide in Zambia? *BMC Health Services Research* 7, 143.

A Estimating α and β

In this appendix, we explain how we estimate α and β and how we deal with heteroskedasticity. The estimation of β is complicated by the endogeneity of return time that occurs if frequencies are already partly determined by client numbers. To account for this effect, we subtract from model (2) the average of model (2) over all visits $v \in \mathcal{V}_s$. The resulting so-called ‘fixed-effects model’ (9) exploits the within-site variation in return times and client numbers and discards the variation across sites, because the latter can be confounded with an effect of client numbers on return times. We estimate β by $\hat{\beta}$, the OLS estimator of β obtained from model (9).

$$CL_{vs} - \frac{1}{V_s} \sum_{v \in \mathcal{V}_s} CL_{vs} = m_s \cdot \beta \cdot \left(g(RT_{vs}) - \frac{1}{V_s} \sum_{v \in \mathcal{V}_s} g(RT_{vs}) \right) + \eta_{vs} \quad (9)$$

Parameter α is canceled out in (9). $\hat{\alpha}$, the OLS estimator α , is computed by plugging $\hat{\beta}$ into model (2) and performing OLS. The corresponding residuals e_{vs} are computed in (10).

$$e_{vs} = CL_{vs} - m_s \cdot \left(\hat{\alpha} + \hat{\beta} \cdot g(RT_{vs}) \right) \quad (10)$$

We acknowledge that our model focuses on just one determinant of client volumes and thereby excludes many others. However, we do not believe this leads to an omitted variable bias. Variations in return times are in part due to a lack of systematic planning, in part due to availability constraints for the facility where the site visit takes place, and in part due to local circumstances such as ‘the roads are going to be bad or good’ or ‘there is going to be a market day’ or ‘there is going to be an under five day’ (Peter Schaffler, May 1, 2018). Throughout our years of collaboration with MSI, we have not come across any anecdotal evidence/ expert

opinion/ mechanism suggesting that one of these factors is both correlated with the difference in return time and the difference in client volume. In Online Appendix A, we show that our data is heteroskedastic, and motivate our choice to account for that by dividing both the dependent and the independent variables by $\sqrt{m_s}$.

B Proofs on Optimality Gaps

Lemma 1. *Objective function 4a is concave in y_s for any function $g(RT_s)$ that is concave in RT_s .*

Proof. Let us rewrite objective function 4a as $\sum_{s \in \mathcal{S}} y_s G_s(\frac{1}{y_s})$. The result follows from the observations that $y_s G_s(\frac{1}{y_s})$ represents the so-called *perspective* of function $G_s(t)$, that $G_s(t)$ is concave in t by assumption (because $g(RT_s)$ is), and that the perspective of a concave function is concave as well (Boyd et al., 2004). □

Proposition 3. *The optimality gap of TwoCP, measured as the absolute difference with the optimal InfCP solution value, is bounded from above by $\sum_{s \in \mathcal{S}} m_s \Delta$, where Δ is calculated as $f(\tilde{y}(\delta)) + (y_{HFC} - \tilde{y}(\delta)) \delta - f(y_{HFC})$ and $\delta = \frac{f(y_{HFC}) - f(y_{LFC})}{y_{HFC} - y_{LFC}}$.*

Proof. Consider the TwoCP solution, which assigns the n_{HFC} sites with the highest value of m_s to *HFC* and the others to *LFC*. Clearly, the corresponding solution value provides a lower bound on the optimal solution value, and is given by $\sum_{s \in \mathcal{S}_{LFC}} m_s f(y_{LFC}) + \sum_{s \in \mathcal{S}_{HFC}} m_s f(y_{HFC})$.

Now consider the function $\bar{f}(y) = f(\tilde{y}(\delta)) + (y - \tilde{y}(\delta)) \delta$. This function represents the tangent of f at point $\tilde{y}(\delta)$. The concavity of f implies that $\bar{f}(y) \geq f(y)$. Replacing objective function 4a by *maximise* $\sum_{s \in \mathcal{S}} m_s \bar{f}(y_s)$ therefore yields an upper bound on the optimal solution value. The corresponding problem is trivial: the linear, non-decreasing objective function makes it optimal to greedily assign the n_{HFC} sites with the highest value of m_s to *HFC* (see Section 5.2.2) yielding solution value $\sum_{s \in \mathcal{S}_{LFC}} m_s \bar{f}(y_{LFC}) + \sum_{s \in \mathcal{S}_{HFC}} m_s \bar{f}(y_{HFC})$.

Before we finish our proof, we note that:

$$\bar{f}(y_{HFC}) - f(y_{HFC}) = \bar{f}(y_{LFC}) + \delta(y_{HFC} - y_{LFC}) - (f(y_{LFC}) + \delta(y_{HFC} - y_{LFC})) \quad (11)$$

$$= \bar{f}(y_{LFC}) - f(y_{LFC}) \quad (12)$$

$$= f(\tilde{y}(\delta)) + (y_{HFC} - \tilde{y}(\delta)) \delta - f(y_{HFC}) \quad (13)$$

$$= \Delta \quad (14)$$

Using these results, we now finish our proof:

$$\sum_{s \in \mathcal{S}_{LFC}} m_s (f(y_{LFC}) - \bar{f}(y_{LFC})) + \sum_{s \in \mathcal{S}_{HFC}} m_s (f(y_{HFC}) - \bar{f}(y_{LFC})) \quad (15)$$

$$= \sum_{s \in \mathcal{S}_{LFC}} m_s \Delta + \sum_{s \in \mathcal{S}_{HFC}} m_s \Delta \quad (16)$$

$$= \sum_{s \in \mathcal{S}} m_s \Delta \quad (17)$$

□

Proposition 4. *The optimality gap of a policy that distinguishes k frequency categories, measured as the absolute difference with the optimal InfCP solution value, is bounded from above by*

$\sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}_c} m_s \max_{c \in \mathcal{C}} \{\Delta_c\}$, where, Δ_c is calculated as $f(\tilde{y}(\delta_c)) + (y_c - \tilde{y}(\delta_c)) \delta_c - f(y_c)$ and $\delta_c = \frac{f(y_{c+1}) - f(y_c)}{y_{c+1} - y_c}$ for $c < k$ and $\delta_k = \delta_{k-1}$.

Proof. Consider the solution of the policy that distinguishes k frequency categories. It assigns the n_k sites with the highest value of m_s to category k , the n_{k-1} with the highest value of m_s among the remaining sites with to category $k - 1$, etc. The corresponding solution value provides a lower bound on the optimal solution value, and is given by $\sum_{s \in \mathcal{C}} \sum_{s \in \mathcal{S}_c} m_s f(y_c)$.

Now consider piece-wise linear function $\underline{f}(y)$ with breakpoints $(y_c, f(y_c))$. This function hence has $k - 1$ line segments. Let δ_c represent the slope of segment c , calculated as $\frac{f(y_{c+1}) - f(y_c)}{y_{c+1} - y_c}$ for $c < k$. The line segment itself is represented by $\underline{f}_c(y) = f(y_c) + \delta_c(y - y_c)$ for $y \in [y_c, y_{c+1}]$. The concavity of $f(y)$ implies that $f(y) \geq \underline{f}(y)$. We now construct function $\bar{f}(y) \geq f(y)$ by “shifting” function $\underline{f}(y)$ up by some constant Δ : $\bar{f}(y) = \underline{f}(y) + \Delta$. We represent the “shifted” line segments by $\bar{f}_c(y) = \underline{f}_c(y) + \Delta$. Note that line segment c needs to be shifted up by at least Δ_c to ensure that $\bar{f}_c(y) \geq f(y)$, where $\Delta_c = f(\tilde{y}(\delta_c)) + (y_c - \tilde{y}(\delta_c)) \delta_c - f(y_c)$

Hence, setting $\Delta = \max_{c \in \mathcal{C}} \{\Delta_c\}$ ensures that $\bar{f}(y) \geq f(y)$. Replacing objective function 4a by *maximise* $\sum_{s \in \mathcal{S}} m_s \bar{f}(y_s)$ therefore yields an upper bound on the optimal solution value. The corresponding problem is trivial. Since $f(y)$ is assumed to be non-decreasing in y_s , it is optimal to assign the n_k sites with the highest value of m_s to category k , the n_{k-1} with the highest value of m_s among the remaining sites with to category $k - 1$, etc. This yields solution value $\sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}_c} m_s \bar{f}(y_c) = \sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}_c} m_s (f(y_c) + \Delta)$. The difference with the lower bound solution hence amounts to $\sum_{c \in \mathcal{C}} \sum_{s \in \mathcal{S}_c} m_s \Delta$, which completes the proof. □

C Simulation Details

Step 1: Simulate historical return times and client numbers. In step 1 we simulate client numbers using Equation (2) with the actual m_s and the $\hat{\alpha}$, $\hat{\beta}$ and estimated error distribution found for Uganda in Section 4. We simulate client numbers for n visits per site. We initially randomly draw return times from the set of historical return times for that site.

Step 2: Estimate m_s for each site. In the perfect information case, we use the actual m_s for each site. In the imperfect information case, we estimate m_s as the average simulated client number for IRT policies and by Equation (7) for CRT policies.

Step 3: Apply policies to find frequencies. The policies explained in Section 5.4 take the estimated values of m_s and return the recommended visit frequency per site. We next state the parameter choices and then the implementation details for ThreeCP. As explained in Section 2, we use an upper bound of one visit per month and a lower bound of one visit per six months. We set D equal to 20 days per month.

We next explain the implementation of the ThreeCP. For the middle category we choose a visit frequency of one visit per 3 months: $MB = 1/3$. This value is chosen because access to short-term family planning methods (such as condoms and birth control pills) should ideally be provided every three months. We solve model (3C) to derive a national guideline for the percentage of sites in the MFC. For a given team, we operationalise this guideline as follows: Let x be the number of sites this team should assign to the MFC, according to the national guideline. As this number may be fractional and may not utilise all the capacity, we propose rounding this number to the closest multiple of five. Starting from the allocation of the TwoCP, the policy then moves the site from the UFC with the lowest estimated value of m_s to the LFC (which frees up a capacity of $UB - LB = 5/6$ days per month) and then moves the five sites from the LFC with the highest estimated value of m_s to the MFC (which costs $MB - LB = 1/6$ days per site per month, thus $5/6$ days in total).

Step 4: Compute expected client numbers. For the case of imperfect information, we simulate client numbers, estimate m_s and compute frequencies. We then compute the expected client number per month per site, ECL_s , by multiplying the frequency with the expected client number per visit from (2). The total client number per month is then computed by summing across all sites, $\sum_{s \in \mathcal{S}} ECL_s$. We repeat this process 100 times and average this total over all 100 times. Dividing this average by the total number of monthly visits (i.e., $D = 20$ times the number of teams) then yields the average client number per visit.