A Fair Incentive Scheme for Community Health Workers

Avinandan Bose[1], Tracey Li[2], Arunesh Sinha[3], Tien Mai[4]

[1]University of Washington
[2]D-Tree International
[3]Rutgers University
[4]Singapore Management University
avibose@cs.washington.edu, tli@d-tree.org, arunesh.sinha@rutgers.edu, atmai@smu.edu.sg

Abstract

Community health workers (CHWs) play a crucial role in the last mile delivery of essential health services to under-served populations in low-income countries. Many non-governmental organizations (NGOs) provide training and support to enable CHWs to deliver health services to their communities, with no charge to the recipients of the services. This includes monetary compensation for the work that CHWs perform, which is broken down into a series of well-defined tasks. In this work, we partner with a NGO D-Tree International to design a fair monetary compensation scheme for tasks performed by CHWs in the semi-autonomous region of Zanzibar in Tanzania, Africa. In consultation with stakeholders, we interpret fairness as the equal opportunity to earn, which means that each CHW has the opportunity to earn roughly the same total payment over a given T month period, if the CHW reacts to the incentive scheme almost rationally. We model this problem as a reward design problem for a Markov Decision Process (MDP) formulation for the CHWs' earning. There is a need for the mechanism to be simple so that it is understood by the CHWs, thus, we explore linear and piecewise linear rewards in the CHWs' measured units of work. We solve this design problem via a novel policy-reward gradient result. Our experiments using two real world parameters from the ground provide evidence of reasonable incentive output by our scheme.

1 Introduction

Community health workers (CHWs) play a crucial role in the delivery of essential health services to under-served populations in low-income countries. Many non-governmental organizations (NGOs) provide training and support to CHWs, as well as monetary compensation for the work they perform (Kok et al. 2015), in order that CHWs can serve their local communities free of charge. In this work, we partner with a NGO D-Tree International to design a fair monetary compensation scheme for tasks performed by 2200 CHWs in Zanzibar in Tanzania, Africa serving about 37,000 pregnant women and 200,000 children. In particular, the tasks performed by CHWs in this work are to enroll clients (pregnant women and young children) and visit enrolled clients at specified frequencies. Each CHW is assigned to a catchment area within which he or she conducts their work. The catchment areas vary in terms of population size and demographic distribution.

As part of a broader study, via informal discussion and feedback we gained an understanding of how the CHWs perceive their current monetary payment, as well as other aspects of their general working conditions. The main observations related to monetary compensation, were that (1) monetary compensation plays a significant role in motivating CHWs to complete their tasks, and (2) CHWs are aware that they don't always have the opportunity to earn as much as some of their colleagues. This second point can be attributed to geographic and temporal variation of population demographics which means that some CHWs sometimes have the opportunity to enroll and visit more clients than other CHWs and therefore have the opportunity to earn more. However, consultation with CHWs and other relevant stakeholders indicates that the optimal situation, from the CHWs' perspective, is for all CHWs to have *equal opportunity to earn*. We interpret this requirement to mean that each CHW must have the opportunity to earn roughly the same total payment over a given T month period, if the CHW reacts to the incentive scheme almost rationally. We allow some bounded rational behavior as the ground situation is challenging and perfect rationality is often not observed, especially in low-income settings, for a number of reasons (Matúŝ and Martonĉik 2017). This mechanism design problem has unique technical characteristics, as we describe next.

Our *first technical contribution* is a Markov Decision Process (MDP) formulation for the problem where states track the current number of enrolled clients n_t. The CHW chooses the number of clients to enroll ($x_{e,t}$) and number of visits to complete ($x_{s,t}$) in time step t. We design an immediate reward scheme that pays $f_e(x_{e,t})$ and $f_s(x_{s,t})$ where f_e, f_s are parameterized by θ_e, θ_s respectively. There is a need for the mechanism to be simple so that it is understood by the CHWs, hence we explore two classes of functions for f_e, f_s: linear and piecewise linear. The CHWs respond to the reward scheme, but as observed from past performance, the response is not always rational. Thus, assuming a bounded rational (quantal responding (McFadden 1976)) myopic response, we formulate the whole problem as a T horizon MDP, in which the designer chooses the immediate reward and the CHWs choose a policy in response. However, the fairness criteria to optimize is that the long term return V_θ

for every CHW is to be close to the budget B allocated per CHW for T months irrespective of the initial enrollment n_0; this is quite different from usual notions of maximizing revenue in mechanism design or maximizing value in MDPs. Further, in this formulation, we choose a myopic CHW but show that for our problem a rational myopic CHW does also maximize the long term reward (Theorem 1), thus, myopicity is not a restriction. We also propose an alternate and more standard value maximization formulation, but find instability issues with this formulation in experiments.

As our *second contribution*, in order to optimize the fairness criteria using a gradient based method, we compute a novel policy-reward gradient (Theorem 2). This is needed as the aforementioned MDP has design parameters $\theta = (\theta_e, \theta_s)$ that determine both the immediate reward and the policy, hence standard policy gradient theorem (Sutton et al. 1999) does not directly apply. We solve our mechanism design formulation using this policy-reward gradient within a stochastic gradient set-up with prioritized sampling.

Finally, we experiment with real world data from the field, obtained from the CHW program in a region in East Africa. We find that we are able to meet our equal opportunity of payment criteria closely. Further, we analyze a few aspects of our problem, including convergence and robustness for different parameters. Overall, we (as well as the NGO) find the linear incentive function to be more satisfactory. We hope our work leads to increased CHW work satisfaction which translates to better health delivery.

2 Related Work

In mechanism design literature on incentives, there is a lot of work on incentivizing crowd workers (Prelec 2004; Witkowski and Parkes 2012; Dasgupta and Ghosh 2013; Yin, Chen, and Sun 2013; Radanovic and Faltings 2013; Waggoner and Chen 2014; Yin and Chen 2015; Radanovic, Faltings, and Jurca 2016; Elmalech et al. 2017). However, this line of work has many differences from our work: (1) generally work in crowd-souring incentive design focuses on truthful task completion, which is different from our problem, (2) a crowd-sourced task can be redone and performed by multiple workers (used for audits in some papers), which is not possible in our setting, and (3) CHW tasks involve physical travel and real-life interpersonal interaction, while crowd-sourced tasks are mostly on a computer. A notion of fairness based on ability and different task evaluation criteria (when task completion is evaluated by peers) is also considered in this line of work (Kamar and Horvitz 2012; Goel and Faltings 2019). Our fairness criteria is different as evaluation of quality of work is quantitatively known in terms of enrollments and service visits, but rather we want to equalize the opportunity to earn. Bounded rationality has also been studied in mechanism design (Zhang 2018; Hu, Zhang, and Li 2019) and decision-making (Bose, Sinha, and Mai 2022), something we also consider. Incentives for ensuring desired outcomes in sharing economy (e.g., bike sharing) (Singla et al. 2015; Zhou et al. 2019) are further away from our work. Even further away, there are reward shaping methods for reinforcement learning (RL), but these work design rewards for faster or successful convergence of RL (Ng, Harada, and Russell 1999; Sutton and Barto 2018). In contrast, our work is a mechanism design problem where the designer has a different objective to optimize.

There are several papers in the social sciences domain that have evaluated CHW programs. A randomized control trial (RCT) in Zambia (Shen et al. 2017) tested the effect of an incentive scheme where rewards were based on performance, and performance evaluation was subjectively done by supervisors, thus, the design followed is more subjective than objectively data driven. Nonetheless, the results showed that performance based incentives led to increased job satisfaction and decreased attrition. In another work (Ormel et al. 2019) conducted in Ethiopia, Kenya and Malawi, health workers expressed that when their expectations regarding incentives were not met then this negatively affected their motivation. Expectation gaps took various forms: lower than expected financial incentives, later than expected payments, fewer than expected material incentives and job enablers and *unequally distributed incentives* across groups of CHWs. CHWs perceived the way material incentives were distributed as unfair, due to favoritism and lack of transparency. Our scheme aims to be completely transparent and also simple to explain. Other studies have similar claims (Glenton et al. 2010; Furth and Crigler 2012). Our feedback results in the Experiments section further supports these claims. Also, from our domain knowledge and from other works, we know that CHWs become demotivated when payments are late, and when there are fewer than expected job enablers, e.g., inadequate supervision. Thus, an efficient and supportive environment is also important on top of the monetary payments.

3 Problem Formulation

Informal Description

The problem at hand involves a CHW enrolling and providing services, via household visits, to clients. The description below focuses on one type of client: pregnant women. We use pregnant women as the example in our description, although there are similar tasks related to enrolling children who are under five years old. These tasks are common for CHWs in many low-income countries. Ideally, pregnant women should be enrolled at the start of their pregnancies in order to receive the necessary health care in the early stages of pregnancy, but it can be very difficult for CHWs to do this in cultures where pregnancies are kept secret. However, CHWs should be able to enroll most of the pregnant women by the time they are visibly pregnant. Similar considerations hold for enrolling children, where it can be difficult to locate newborn babies, but nearly all older children can relatively easily be enrolled. Besides enrollments, CHWs also visit the enrolled clients' houses when they are due for a visit (see our problem specific schedule in experiments) in order to provide the necessary health services.

Formal Description

First, we describe the evolution of the number of pregnant women, which is a stochastic process that is independent of CHWs' actions. Let N_t denote the number of pregnant women at the end of month t. It is assumed that in month t $c_{e,t} \geq 0$ (random variable) women become pregnant and $c_{d,t} \geq 0$ (random variable) women give birth (and thus, are no more pregnant). Hence, the total number of pregnant women at the end of month t is $N_t = N_{t-1} + c_{e,t} - c_{d,t}$. Further, as the total number of pregnant women stays roughly the same over time, we impose $\mathbb{E}[c_{e,t}] = \mathbb{E}[c_{d,t}] = c$ for all t. It is assumed that N_0 is the same for all CHWs and also $N_t \leq 2N_0$ (as the total number remains roughly same). Also, as a simplifying assumption, among the enrolled clients, a fraction α are scheduled to receive a visit every month. The parameters $c_{e,t}$ and $c_{d,t}$ are calculated from census estimates and the fact that it should be possible for most women to be enrolled by the time they approach childbirth.

More formally, we can model the whole process as a T horizon undiscounted MDP for the CHW (S, A, r, P). Here, the state $(n_t, N_t) \in S$ where $S = \{0, 1, \ldots, 2N_0\} \times \{0, \ldots, 2N_0\}$ tracks the number of enrolled clients and the total number of pregnant women at the end of month t. The action contained in space A is state dependent and described next (reward r and transition P is described after that).

CHW action: Let $x_{e,t}, x_{s,t}$ be the two non-negative integer variables that track the enrollment and health visits in month t. The enrollment $x_{e,t}$ possible in a month is upper bounded by K, because of physical limitations on how many houses can be scouted for finding clients. K is larger than any possible value of $c_{e,t}$ or $c_{d,t}$, that is, $K > c_{e,t}$ and $K > c_{d,t}$. Further, the enrollment in month t is also upper bounded by the maximum number of unenrolled pregnant women: $N_{t-1} - n_{t-1} + c_{e,t}$. Hence, $0 \leq x_{e,t} \leq \min(K, N_{t-1} - n_{t-1} + c_{e,t})$. Also, $0 \leq x_{s,t} \leq \lfloor \alpha n_{t-1} \rfloor$. We use $A(t)$ to denote allowable actions in month t, which depends on the state (n_{t-1}, N_{t-1}).

Reward and policy: We want to design an incentive scheme f_e, f_s as a function of $x_{e,t}, x_{s,t}$. We assume f_e, f_s are parameterized by θ_e, θ_s respectively. Thus, the reward for the CHW in month t is $r_\theta(x_{e,t}, x_{s,t}) = f_e(x_{e,t}) + f_s(x_{s,t})$, where $\theta = (\theta_e, \theta_s)$. Note that it is a deliberate choice (as a designer) to make the reward independent of the state, since we concluded after deliberation with our NGO partners that state dependent rewards are overly complicated for CHWs and other stakeholders to interpret. Further, it is natural to force f_e, f_s to be *monotonic* in its argument, that is, an increase in the amount of work done by a CHW leads to an increase in payment.

It is also known that in general and in the context of community health workers (Agarwal et al. 2021), that humans often choose options with bounded rationality. Thus, we adopt the quantal response (QR) model, a very well known and often used model (McFadden 1976; McKelvey and Palfrey 1995; Sinha et al. 2018), to model the CHW's choice making process. In particular, following myopic QR response, the CHW will choose $x_{e,t}, x_{s,t}$ in month t with probability

$$\pi_\theta(x_{e,t}, x_{s,t}) = \frac{e^{r_\theta(x_{e,t}, x_{s,t})/\mu}}{\sum_{(x'_{e,t}, x'_{s,t}) \in A(t)} e^{r_\theta(x'_{e,t}, x'_{s,t})/\mu}},$$

where μ is a rationality parameter (0 for rational, ∞ for completely irrational). Note that we have fixed the policy π_θ of the CHW (given rewards), that is, the MDP is played with the given policy above. The subscript θ emphasizes the dependence of π_θ on the reward parameter θ.

Transition: The number of enrolled clients at the end of month t is $n_t = n_{t-1} + x_{e,t} - c_{d,t}$. The number of pregnant women at the end of month t is $N_t = N_{t-1} + c_{e,t} - c_{d,t}$. Note the transitions are stochastic since $c_{e,t}, c_{d,t}$ are random variables.

Long term reward: It can be inferred from the description above that the policy π_θ is stationary, given θ. Given, π_θ and a fixed N_0, we can define the value function as a function of n_0: $V_\theta(n_0)$ as:

$$V_\theta(n_0) = \mathbb{E}_{\substack{(x_{e,t}, x_{s,t}) \sim \pi_\theta \\ n_t = n_{t-1} + x_{e,t} - c_{d,t} \\ N_t = N_{t-1} + c_{e,t} - c_{d,t}}} \Big[\sum_{t=1}^{T} r_\theta(x_{e,t}, x_{s,t}) \mid n_0 \Big]$$

where the expectation is w.r.t. the stochastic policy π_θ and the distribution of $c_{e,t}, c_{d,t}$.

The mechanism design problem: Formally, our problem is to choose θ such that $|V_\theta(n_0) - B| \leq \epsilon$ for all $n_0 \in I$, where B is per CHW budget, ϵ is a small constant, and I are possible initial number of enrollments (initial states). This formally models the fairness criteria stated in the introduction. One way to achieve the above is to perform:

$$\min_\theta \sum_{n_0 \in I} (V_\theta(n_0) - B)^2 \tag{1}$$

An Alternate Formulation

Another, somewhat indirect, way to formulate the same problem is to incorporate the leftover budget in the state and when the accumulated reward exceeds B then a designed large loss is set as the reward. Formally, we can model the above as a MDP $\{S, A, P, r\}$, where the states are one among $S = \{0, 1, \ldots, N\} \times \{0, \ldots, 2N_0\} \times \{0, 1, \ldots, B\}$. The actions are same as earlier $x_{e,t}, x_{s,t}$ in a month t with given restrictions as $A(t)$. Given action, $x_{e,t}, x_{s,t}$, the transition from $(n_{t-1}, N_{t-1}, b_{t-1})$ to the next state $(n_{t-1} + x_{e,t} - c_{d,t}, N_{t-1} + c_{e,t} - c_{d,t}, b_{t-1} - f_e(x_{e,t}) - f_s(x_{s,t}))$ happens with probability given by the distribution of $c_{e,t}, c_{d,t}$. The immediate reward now depends on the state, in particular on the component b, $r_\theta(b_{t-1}, x_{e,t}, x_{s,t}) = f_e(x_{e,t}) + f_s(x_{s,t})$ for $b_{t-1} \geq f_e(x_{e,t}) + f_s(x_{s,t})$ and $r_\theta(b_{t-1}, x_{e,t}, x_{s,t}) = -M$ for $b_{t-1} < f_e(x_{e,t}) + f_s(x_{s,t})$, where M is a large number. Given, stationary π_θ we can define the value function: $V_\theta(n_0)$ as:

$$V_\theta(n_0, B) = \mathbb{E}_{\substack{(x_{e,t}, x_{s,t}) \sim \pi_\theta \\ n_t = n_{t-1} + x_{e,t} - c_{d,t} \\ N_t = N_{t-1} + c_{e,t} - c_{d,t}}} \Big[\sum_{t=1}^{T} r_\theta(b_{t-1}, x_{e,t}, x_{s,t}) \mid n_0 \Big]$$

where the expectation is w.r.t. the probability distribution over actions and the distribution of $c_{e,t}, c_{d,t}$.

In this new formulation, we can aim to maximize $V_\theta(n_0, B)$ for all $n_0 \in I$. We can replace for all $n_0 \in I$ with a uniform distribution $U(I)$ over allowed initial values of n_0 in I. Note that this formulation results in a larger state space, but has a standard value maximization objective. This formulation is indirect as we need to carefully evaluate the optimized θ so that the budget is not exceeded for any start state in I (as we only maximize starting with the uniform distribution over I). Also, the penalty term $-M$ must be chosen carefully, as we show instability for certain choices of M in experiments. Overall, the instability and indirect approach makes this formulation unsuitable for operationalization by the NGO.

Properties of the Formulation

The reader might have observed that the CHWs, when rational ($\mu = 0$) play a myopic best response. One may wonder if this is too restrictive, as a rational CHW can be expected to maximize long term reward. The following result shows that a myopic best responding CHW does indeed maximize the long term reward also in our problem, hence the myopic behavior is not a restrictive assumption.

Theorem 1. *Given monotonic incentive functions, a myopic best responding CHW policy also maximizes the long term T step reward in our problem.*

Proof Sketch. The proof is by backward induction, where first the result is shown for T and then at any intermediate step it is shown that the myopic best response leads to higher (or same) enrollment and thus higher (or same) enrollment and service visit payments. Full proof is in appendix. □

4 Gradient-Based Solution Approach

The first choice in our solution approach is to choose the function class for f_e, f_s noting that these functions have to be strictly monotonic. We consider two function classes: (1) *linear function* $f_e(x) = \theta_e x, f_s(x) = \theta_s x$ where $\theta_e, \theta_s \geq 0$ and (2) *piecewise linear function*, we consider pieces spaced k apart in x and pieces indexed by $l \in \{0, 1, \ldots\}$ to get

$$f_e(x) = \theta_{e,l}(x - lk) + k \sum_{l'=0}^{l-1} \theta_{e,l'} \text{ for } (l+1)k \geq x > lk$$

and analogously for pieces n apart

$$f_s(x) = \theta_{s,l}(x - ln) + n \sum_{l'=0}^{l-1} \theta_{s,l} \text{ for } (l+1)n \geq x > ln,$$

where $\sum_{l'=0}^{-1}$ is considered as zero. We impose $\geq \theta_{e,l} \geq 0$ and $\theta_{s,l} \geq 0$ for all l. In the piecewise linear case, $\theta = (\theta_{e,0}, \ldots, \theta_{e,\lceil K/k \rceil}, \theta_{s,0}, \ldots, \theta_{s,\lceil \alpha N/n \rceil})$. These different types of utilities clearly satisfy monotonicity.

Importantly, these function forms provide a simplicity to the resulting incentive scheme that makes it easily interpretable to CHWs, as per the specification of the NGO. In fact, the NGO has a preference for the purely linear functions due to the benefits conferred by simplicity; here we explore both options for a broader applicability. Next, we compute the gradient of V_θ to aid in a gradient based solution for both formulations stated in the previous section.

Gradient of V_θ: We start by simplifying some notation to use standard notation in reinforcement learning and prove a general result. Thus, we will drop some arguments and use the terms V_θ where n_0 is implicit but not written, a_t for $(x_{e,t}, x_{s,t})$, s_t instead of state (n_{t-1}, N_{t-1}) at time step t, and $r_{\theta,t}$ for $r_\theta(\cdot)$ (in either of the two formulations). Also, we use $\tau \sim \pi_\theta$ to denote sampled trajectories, where the randomness arises from the environment (specifically from $c_{i,t}$) and from the stochastic policy π_θ. A trajectory τ is a sequence of $(s_t, a_t, r_{\theta,t})$.

We seek a gradient of V_θ w.r.t. θ, but we cannot use the standard policy gradient theorem (Sutton et al. 1999) as the immediate reward depends on θ. Hence, we prove:

Theorem 2. *For a trajectory τ, let $G_t = \sum_{t'=t}^{T} r_{\theta,t'}$. Then,*

$$\nabla_\theta V_\theta = \mathbb{E}_{\tau \sim \pi_\theta}\Big[\sum_{t=1}^{T} G_t \nabla_\theta \log \pi_\theta(a_t|s_t)\big) + \sum_{t=1}^{T} \nabla_\theta r_{\theta,t}\Big]$$

Proof is in appendix. In particular, for our problem the above general result takes a very convenient form:

Corollary 1. *For our problem, in both formulations $\nabla_\theta \log \pi_\theta(x_{e,t}, x_{s,t})$ (with state implicit) takes the form*

$$\frac{1}{\mu}\big(\nabla_\theta r_{\theta,t} - \sum_{x'_{e,t}, x'_{s,t}} \pi_\theta(x'_{e,t}, x'_{s,t}) \nabla_\theta r_{\theta,t}\big)$$

For linear utilities, given the two dimensional $\theta = (\theta_e, \theta_s)$, we get $\nabla_\theta r_{\theta,t} = (x_{e,t}, x_{s,t})^T$. For piecewise linear utilities, using $(\cdot)_{e,l}$ and $(\cdot)_{s,l}$ to denote the components of gradient w.r.t. $\theta = (\theta_{e,0}, \ldots, \theta_{e,\lceil K/k \rceil}, \theta_{s,0}, \ldots, \theta_{s,\lceil \alpha N/n \rceil})$ we get

$$(\nabla_\theta r_{\theta,t})_{e,l} = \begin{cases} k & x_{e,t} > (l+1)k \\ x_{e,t} - lk & (l+1)k \geq x_{e,t} > lk \\ 0 & lk \geq x_{e,t} \end{cases}$$

$$(\nabla_\theta r_{\theta,t})_{s,l} = \begin{cases} n & x_{s,t} > (l+1)n \\ x_{s,t} - ln & (l+1)n \geq x_{s,t} > ln \\ 0 & ln \geq x_{s,t} \end{cases}$$

Given the gradients above, we use stochastic gradient descent to optimize the objective in Equation 1 using a simulator (included in code). We use stochastic version where we sample from the set I, as the set I can be large (this is the same as minibatch training for neural networks). Further, in the sampling process we use *prioritized sampling* and choose trajectories with initial enrollments n_0 with probability proportional to $(V_\theta(n_0) - B)^2$, that is, higher loss terms are chosen with higher probability to be optimized. The alternative formulation is also trained using standard minibatch training, but as stated earlier the choice of $-M$ is difficult to make and often results in instability.

5 Experiments

As stated earlier, we have conducted this study in partnership with D-Tree International, who support the government of Zanzibar to run Zanzibar's CHW program. The CHW program comprises over 2000 CHWs, who provide coverage all over the region. Each CHW is provided with, and trained to use, a smartphone that has a specially designed app installed on it that functions as a client management and decision support tool. CHWs enter data about their clients into the app, which then automatically schedules the required visits for each client and prompts the CHW to complete those visits

Parameters	Value pregnant women	Value child
N_0	24	124
$\mathbb{E}[c_{e,t}] = \mathbb{E}[c_{d,t}]$	2	2
K	8	16
α	1/3	1/5
B (assumed)	100	100
T	6	6

Table 1: Parameter values per CHW

at the necessary time. For pregnant women, the visitation scheduled is set to one visit per trimester of the pregnancy (which means a total of 3 visits in a full 9 month pregnancy, assuming the client is enrolled early on in her pregnancy). Thus, for pregnant women the fraction $\alpha = 1/3$. The visitation schedule for children is determined by the child's age, with more frequent visits scheduled for younger children: 6 visits in year 0-1, 3 visits in year 1-2, 1 visit every year in years 2-5. Assuming the same number of children in each year age group, averaged over months, this gives $\alpha = 6/12 \times 1/5 + 3/12 \times 1/5 + 3 \times 1/12 \times 1/5 = 1/5$. Next, the population sizes and other parameters differ slightly for each CHW, we take an average of these and present the various parameter values per CHW in Table 1.

Feedback about Incentives

Here, we first present the current incentives scheme and feedback from CHWs. The current incentive scheme is very simple and does not differentiate between pregnant women and children. Payments are currently calculated and made on a monthly basis. For enrollments, each CHW is paid a fixed amount for each client (either pregnant woman or child) that he or she enrolls, up to a maximum of 4 per month. The CHW does not receive payment for any additional client enrollments. For visits (either pregnant woman or child), CHWs do not receive payment unless they complete at least 5 visits in a month. From then, they receive a fixed amount for completing 5-11 visits, receive double that amount for completing 12-15 visits, and receive the maximum amount if they complete 16 or more visits. This scheme has led to an imbalance in the proportion of all pregnant women that are served, versus the proportion of children that are served, because it is generally easier for CHWs to enroll children than pregnant women due to the culture of secrecy around pregnancy, described earlier. In order to achieve more equal coverage of the two client groups, the NGO is therefore looking to redesign the incentive scheme in order to specify separate targets and payments for pregnant women and children.

Throughout the program, CHWs and other stakeholders have provided feedback about the challenges and working conditions of CHWs. Here, we summarize the collective responses that are relevant to the topic of incentives: (1) Several CHWs described how some potential clients refuse to be enrolled in the program, which is due to a number of social and cultural reasons. This means that although a CHW may be able to successfully locate an unenrolled pregnant woman or child, the CHW cannot complete the enrollment and thus

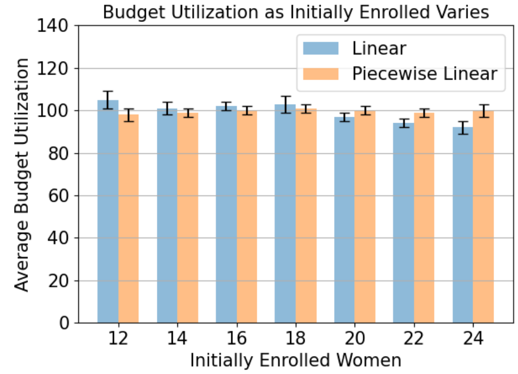

Figure 1: Payment (Budget used) for varying initial enrollment for pregnant women

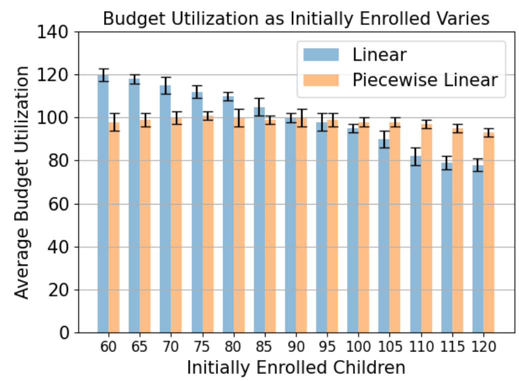

Figure 2: Payment (Budget used) for varying initial enrollment for children

won't be compensated for their effort. (2) CHWs generally understand that if they enroll more clients, they will have more visits scheduled and therefore a higher chance of earning the maximum amount. (3) Despite understanding the previous point, many CHWs aren't aware that different clients are associated to different visitation schedules, and are therefore surprised that they are not often scheduled to visit many of their clients. This is particularly common when CHWs enroll a lot of older children, who are scheduled to only receive one visit per year. (4) CHWs indicate that although it is not their sole motivation, monetary payment is very important to them and they are dissatisfied when payments are delayed or are less than what they expected. (5) In addition to monetary payments, CHWs are also motivated by social rewards, such as recognition or appreciation from their supervisors and communities. (6) Many CHWs and other stakeholders express the opinion that all CHWs should be able to earn the same maximum amount each month.

In response to this feedback, the NGO has been working with stakeholders to identify how to make it easier for CHWs to enroll clients, particularly pregnant women, and to better inform stakeholders about the visitation schedule. We

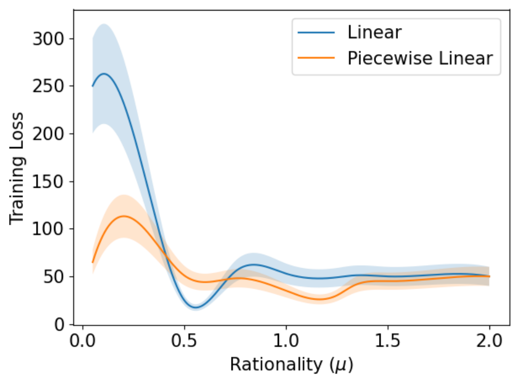

Figure 3: Training loss for varying μ (women)

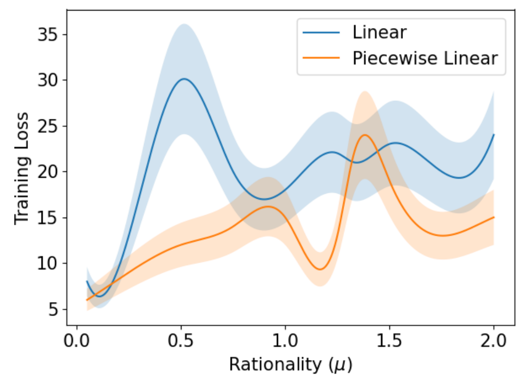

Figure 4: Training loss for varying μ (children)

note that the feedback obtained is consistent with what has previously been reported in literature, particularly in terms of the importance (and, in fact, necessity) of both monetary and non-monetary rewards (Kok et al. 2015), and that CHWs' expectations about payments need to be considered as much as the absolute payment value (Ormel et al. 2019).

Incentive Design Results

We present results along various dimensions below. Additionally, we also verify the robustness of our solution by performing sensitivity analysis. All our experiments were run on a 2.1 GHz CPU with 128GB RAM. All the datapoints in the results are averaged over 1000 test runs. For our results, unless we vary a parameter, we fix the parameter values to the default numbers shown in Table 1 and the default for μ is 1.33, a typical value reported in literature (Yang et al. 2011; Ge and Godager 2021). For training, we use five random restarts and choose the one with lowest loss. The minibatch size is 512. For piecewise linear function, piece lengths k, n are chosen as four. More set-up details are in appendix.

Performance with varying initial enrollment n_0: We first verify how closely our main criteria of roughly equal payment is satisfied. We call this the budget utilization, which we want to be close to $B = 100$ as specified in Equation 1. Figures 1 and 2 show the results for pregnant women and children respectively with varying number of initial enrollments. We vary the initial enrollment from 50% to 100% of N_0 as a pre-service survey ensures that at least 50% of clients are always enrolled. The results are obtained using the learned parameters. The results show very close adherence to the required criteria, and better adherence using the higher capacity function class of piecewise linear functions.

Converged loss value with varying rationality μ: We vary the rationality parameter and plot the *converged value* of the loss for pregnant women and children as shown in Figures 3 and 4; the plot is for 1000 instances of training (and as stated best of five runs in each training instance). It can be seen that piecewise linearity generally results in lower loss, especially for lower μ. At lower μ the policy is less stochastic and a more complex function than for higher μ, which might be the reason for the better loss of piecewise linear functions.

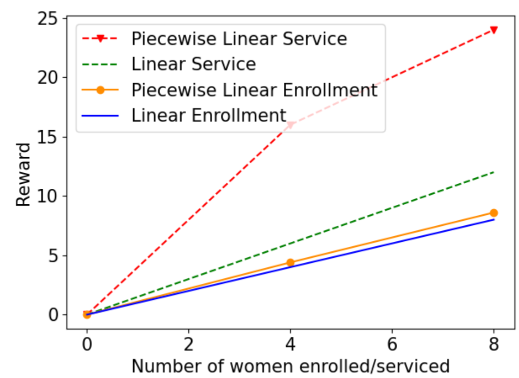

Figure 5: Incentive function for pregnant women

Learned model: Next, we show the learned incentive functions for enrollments and service visits in Figures 5 and 6 for pregnant women and children, respectively. It can be seen the piecewise linear function is not simple to interpret by CHWs and not simple to explain to CHWs. This supports our recommendation for using linear functions in practice.

Sensitivity Analysis: Next, we explore the robustness of our model when the assumed parameter values do not match the actual ones. Two of our parameters are susceptible to this: μ and α. We plot the budget utilization of the learned model using the default $\mu = 1.33$ tested with varying (in test simulator) actual value of μ of CHW in Figures 7 and 8 for pregnant women and children respectively. The robustness to uncertain μ for both linear and piecewise linear incentive function is clear from these figures.

Next, we plot the budget utilization of the learned model using the default α tested with varying actual value of α of CHW in Figures 9 and 10 for pregnant women and children respectively. For pregnant women, both linear and piecewise linear are somewhat equally robust, but for children the linear function is slightly more robust. We interpret this as a simpler model (linear) generalizing better to uncertainty than the piecewise linear model that gets excessively tuned to the default α value. This further supports our recommen-

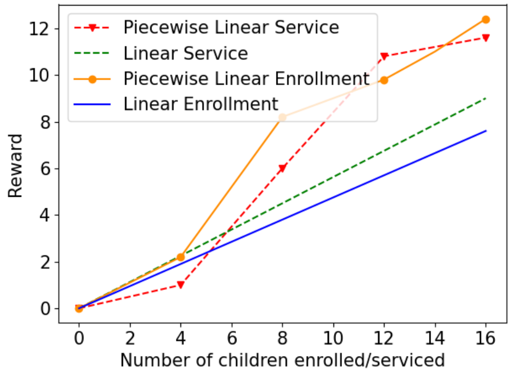

Figure 6: Incentive function for children

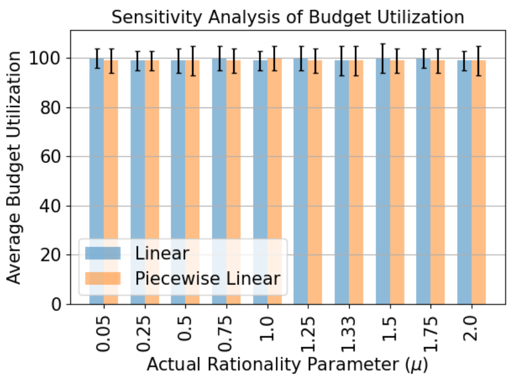

Figure 7: Budget used when trained with $\mu = 1.33$ for varying actual μ for pregnant women task

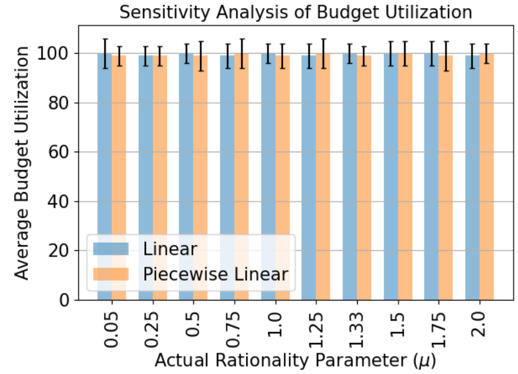

Figure 8: Budget used when trained with $\mu = 1.33$ for varying actual μ for children task

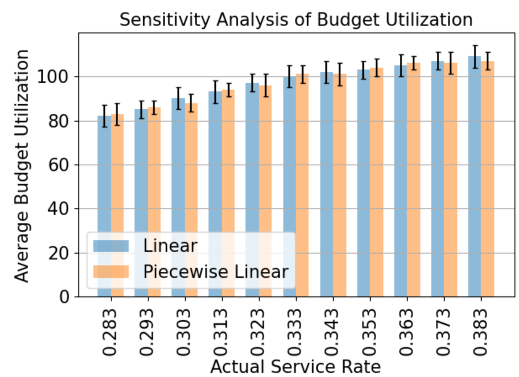

Figure 9: Budget used when trained with $\alpha = 0.33$ for varying actual α for pregnant women task

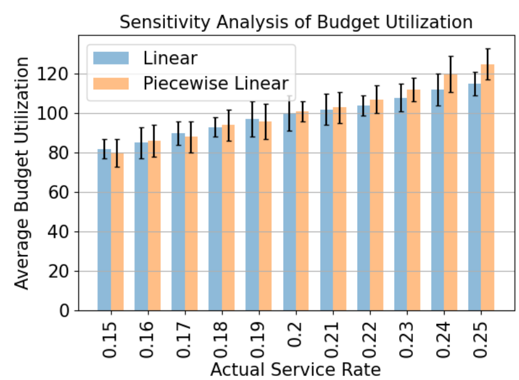

Figure 10: Budget used when trained with $\alpha = 0.2$ for varying actual α for children task

dation for using linear functions in practice.

Alternative formulation: We observe that although M can be tuned to get a mean payment around B in the alternative formulation, the variance is very high. Due to space constraints, we put figures showing the high variance for the budget utilization for different values of M in appendix.

6 Limitations and Conclusion

Our designed monetary incentive scheme will address the issue of fairness, which is a concern raised by many stakeholders, by providing CHWs with equal opportunity to earn. This should lead to increasing CHWs' satisfaction and motivation. However, it is well known that non-monetary incentives are also very important motivators for CHWs, as documented by many studies (Glenton et al. 2010; Singh et al. 2015; World Health Organization et al. 2018). Even in our own feedback from CHWs, CHWs expressed a desire for more social recognition and status, such as having ID cards or uniforms, receiving certificates, and having awards ceremonies to provide them with more motivation. Thus, we point out that a fair monetary incentive scheme must be complemented by a fair non-monetary incentive scheme and an appropriately supportive environment in order to ensure an improvement in CHWs' performance and therefore improved health care delivery amongst under-served populations. We aim for our incentives scheme to be deployed and tested by our partner NGO, with a hope to provide improved health services to hundreds of thousands of beneficiaries.

Acknowledgments

This work was done when all authors other than Tracey Li were at the Singapore Management University. This work was partially supported by Google Research's AI for Social Good Faculty Award (2020) for Arunesh Sinha. D-tree's portion of the work was funded by a Google AI for Social Good award.

References

Agarwal, S.; Abuya, T.; Kintu, R.; Mwanga, D.; Obadha, M.; Pandya, S.; and Warren, C. E. 2021. Understanding community health worker incentive preferences in Uganda using a discrete choice experiment. *Journal of global health*, 11.

Bose, A.; Sinha, A.; and Mai, T. A. 2022. Scalable Distributional Robustness in a Class of Non-Convex Optimization with Guarantees. In *Advances in Neural Information Processing Systems*.

Dasgupta, A.; and Ghosh, A. 2013. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, 319–330.

Elmalech, A.; David, E.; Sarne, D.; and Hajaj, C. 2017. Enhancing crowdworkers' vigilance. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Furth, R.; and Crigler, L. 2012. Improving CHW program functionality, performance, and engagement: operations research results from Zambia. *Research and evaluation report. USAID health care improvement project*.

Ge, G.; and Godager, G. 2021. Predicting strategic medical choices: An application of a quantal response equilibrium choice model. *Journal of choice modelling*, 39: 100282.

Glenton, C.; Scheel, I. B.; Pradhan, S.; Lewin, S.; Hodgins, S.; and Shrestha, V. 2010. The female community health volunteer programme in Nepal: decision makers' perceptions of volunteerism, payment and other incentives. *Social science & medicine*, 70(12): 1920–1927.

Goel, N.; and Faltings, B. 2019. Deep bayesian trust: A dominant and fair incentive mechanism for crowd. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 1996–2003.

Hu, Z.; Zhang, J.; and Li, Z. 2019. General robustness evaluation of incentive mechanism against bounded rationality using continuum-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 6070–6078.

Kamar, E.; and Horvitz, E. 2012. Incentives for truthful reporting in crowdsourcing. In *AAMAS*, volume 12, 1329–1330.

Kok, M. C.; Dieleman, M.; Taegtmeyer, M.; Broerse, J. E.; Kane, S. S.; Ormel, H.; Tijm, M. M.; and De Koning, K. A. 2015. Which intervention design factors influence performance of community health workers in low-and middle-income countries? A systematic review. *Health policy and planning*, 30(9): 1207–1227.

Matúŝ, A.; and Martonĉik, M. 2017. A review of consequences of poverty on economic decision-making: A hypothesized model of a cognitive mechanism. *Frontiers in psychology*, 8: 1784.

McFadden, D. L. 1976. Quantal choice analaysis: A survey. *Annals of Economic and Social Measurement, Volume 5, number 4*, 363–390.

McKelvey, R. D.; and Palfrey, T. R. 1995. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1): 6–38.

Ng, A. Y.; Harada, D.; and Russell, S. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. In *Icml*, volume 99, 278–287.

Ormel, H.; Kok, M.; Kane, S.; Ahmed, R.; Chikaphupha, K.; Rashid, S. F.; Gemechu, D.; Otiso, L.; Sidat, M.; Theobald, S.; et al. 2019. Salaried and voluntary community health workers: exploring how incentives and expectation gaps influence motivation. *Human resources for health*, 17(1): 1–12.

Prelec, D. 2004. A Bayesian truth serum for subjective data. *science*, 306(5695): 462–466.

Radanovic, G.; and Faltings, B. 2013. A robust bayesian truth serum for non-binary signals. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Radanovic, G.; Faltings, B.; and Jurca, R. 2016. Incentives for effort in crowdsourcing using the peer truth serum. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(4): 1–28.

Shen, G. C.; Nguyen, H. T. H.; Das, A.; Sachingongu, N.; Chansa, C.; Qamruddin, J.; and Friedman, J. 2017. Incentives to change: effects of performance-based financing on health workers in Zambia. *Human resources for health*, 15(1): 1–15.

Singh, D.; Negin, J.; Otim, M.; Orach, C. G.; and Cumming, R. 2015. The effect of payment and incentives on motivation and focus of community health workers: five case studies from low-and middle-income countries. *Human resources for health*, 13(1): 1–12.

Singla, A.; Santoni, M.; Bartók, G.; Mukerji, P.; Meenen, M.; and Krause, A. 2015. Incentivizing users for balancing bike sharing systems. In *Twenty-Ninth AAAI conference on artificial intelligence*.

Sinha, A.; Fang, F.; An, B.; Kiekintveld, C.; and Tambe, M. 2018. Stackelberg security games: Looking beyond a decade of success. In *IJCAI*.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.

Sutton, R. S.; McAllester, D.; Singh, S.; and Mansour, Y. 1999. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12.

Waggoner, B.; and Chen, Y. 2014. Output agreement mechanisms and common knowledge. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

Witkowski, J.; and Parkes, D. C. 2012. A robust bayesian truth serum for small populations. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

World Health Organization; et al. 2018. *WHO guideline on health policy and system support to optimize community health worker programmes*. World Health Organization.

Yang, R.; Kiekintveld, C.; Ordonez, F.; Tambe, M.; and John, R. 2011. Improving resource allocation strategy against human adversaries in security games. In *Twenty-Second International Joint Conference on Artificial Intelligence*.

Yin, M.; and Chen, Y. 2015. Bonus or not? learn to reward in crowdsourcing. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Yin, M.; Chen, Y.; and Sun, Y.-A. 2013. The effects of performance-contingent financial incentives in online labor markets. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Zhang, L. 2018. *Bounded Rationality and Mechanism Design*. Ph.D. thesis, The Ohio State University.

Zhou, P.; Wei, X.; Wang, C.; and Yang, Y. 2019. Explore truthful incentives for tasks with heterogenous levels of difficulty in the sharing economy. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*.

A Proofs

Proof of Theorem 1

Proof. The CHWs' long term enrollment payoff is higher for higher number of total enrollment in T steps (due to strict monotonic f_e). The myopic best response for enrollment is to enroll $\min(K, N_{t-1} - n_{t-1} + c_{e,t})$ and for visits to visit all $\lfloor \alpha n_{t-1} \rfloor$ clients. We consider an optimal policy π^* maximizing long term reward and show that π^* must be playing myopic best response. We prove this by backward induction. Fix the sequence of realizations $c_{e,1}, \dots$ and $c_{d,1}, \dots$ (thus, these are not random variables here). We can fix these as these quantities do not depend on the policy.

First, π^* must play a best response in the last time step T otherwise the policy that plays π^* till $T-1$ followed by best response achieves higher overall reward due to higher enrollment and/or higher service visit. Next, assume π^* plays myopic best response from step $i+1$ onwards and is the optimal policy with the least time step $i+1$ where it plays a myopic best response. Suppose at step i, policy π^* plays non-best myopic response. Let π' be exact same policy as π^* except it plays myopic best response at step i. We show π' provides higher or same long term reward, contradicting optimality of π^* with the least time step $i+1$ for myopic best response. Thus, π^* must be playing myopic best response for all i from the start.

Let E' be total number of enrollments by π' in and after step i, and same quantity is E^* for π^*. As the same state (n_{i-1}, N_{i-1}) is reached following both π' and π^*, after step i we will have $n'_i \geq n^*_i$ due to higher (or same) enrollment from π'. Since both follow best response from step $i+1$ onwards, they both enroll K in each time step till π' first reaches the upper bound (since $n'_i > n^*_i$), from where on it enrolls $c_{e,t}$. This maximizes the enrollment π' can have after i which is $N_{i-1} - n_{i-1} + \sum_{t=i}^{T} c_{e,t}$; π^* can also have maximum enrollments up to $N_{i-1} - n_{i-1} + \sum_{t=i}^{T} c_{e,t}$. Thus, $E' \geq E^*$ If π' does not reach the upper bound then π^* also does not, and both enroll K in each step from $i+1$ onwards. But, due to higher (or same) enrollment in step i by π' (and since n_{i-1} is same), again we have $E' \geq E^*$. Thus, π' fetches higher enrollment incentive in step i and beyond.

Moreover, due to the lead $n'_i \geq n^*_i$, $n'_j \geq n^*_j$ for all $j \geq i+1$. As π' visits the maximum possible at each time step in i and beyond, this number of visits in every time step is also more than that of π^*. Thus, π' fetches higher visit incentive in step i and beyond.

The above is true for any realization of the $c_{e,t}, c_{d,t}$ random variables, thus, it is also true in expectation over the distribution of these variables. $\square$

Proof of Theorem 2

Proof. We use $\pi_\theta(\tau)$ to indicate probability of trajectory τ that accounts for both randomized policy and transition uncertainty. We do the steps below:

$$\nabla_\theta V_\theta = \nabla_\theta \mathbb{E}_{\tau \sim \pi_\theta}[\sum_{t=1}^{T} r_{\theta,t}]$$

Using τ_t to denote truncated trajectory till time step t and noting that distribution of $r_{\theta,t}$ depends on τ_t only

$$= \sum_{t=1}^{T} \nabla_\theta \mathbb{E}_{\tau_t \sim \pi_\theta}[r_{\theta,t}]$$

Using $\nabla_\theta \pi_\theta(\tau_t) r_{\theta,t} = \pi_\theta(\tau_t) r_{\theta,t} \nabla_\theta \log(\pi_\theta(\tau_t) r_{\theta,t})$, we get

$$= \sum_{t=1}^{T} \mathbb{E}_{\tau_t \sim \pi_\theta}[r_{\theta,t} \nabla_\theta \log(\pi_\theta(\tau_t) r_{\theta,t})]$$

Using $\log(\pi_\theta(\tau_t) r_{\theta,t}) = \log(\pi_\theta(\tau_t)) + \log(r_{\theta,t})$ and $\log(\pi_\theta(\tau_t)) = \sum_{t'=1}^{t} \log \pi_\theta(a_{t'}|s_{t'}) + K$, where K independent of θ we get

$$= \sum_{t=1}^{T} \mathbb{E}_{\tau_t \sim \pi_\theta}\Big[r_{\theta,t}\big(\sum_{t'=1}^{t} \nabla_\theta \log \pi_\theta(a_{t'}|s_{t'})\big) + r_{\theta,t} \nabla_\theta \log r_{\theta,t}\Big]$$

Using $r_{\theta,t} \nabla_\theta \log r_{\theta,t} = \nabla_\theta r_{\theta,t}$ and pulling summation inside $\mathbb{E}$

$$= \mathbb{E}_{\tau \sim \pi_\theta}\Big[\sum_{t=1}^{T} r_{\theta,t}\big(\sum_{t'=1}^{t} \nabla_\theta \log \pi_\theta(a_{t'}|s_{t'})\big) + \sum_{t=1}^{T} \nabla_\theta r_{\theta,t}\Big]$$

Using fact that $\sum_{t=1}^{T} a_t \sum_{t'=1}^{t} b_{t'} = \sum_{t=1}^{T} b_t \sum_{t'=t}^{T} a_{t'}$, we get

$$= \mathbb{E}_{\tau \sim \pi_\theta}\Big[\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_t|s_t)(\sum_{t'=t}^{T} r_{\theta,t'}) + \sum_{t=1}^{T} \nabla_\theta r_{\theta,t}\Big]$$

$$= \mathbb{E}_{\tau \sim \pi_\theta}\Big[\sum_{t=1}^{T} G_t \nabla_\theta \log \pi_\theta(a_t|s_t) + \sum_{t=1}^{T} \nabla_\theta r_{\theta,t}\Big]$$

$\square$

B Details and Additional Results

Training and simulation details: Algorithm 1 present our simulation and training steps in pseudocode.

Alternate formulation figures: Figures 11 and 12 show the results for pregnant women and children respectively, for the alternate formulation. As can be seen, the error bars are much larger than the direct formulation. The x-axis values are chosen to show those values of M where the mean payment value (budget used) is close to 100, which happens for different values for pregnant women and children.

Learned parameters as service fraction α varies: We show the values of learned parameters for *linear function* with varying α for pregnant women and children in Figures 13 and 14 respectively. It can be seen that service incentives decrease with higher fraction α, likely because services for a higher proportion of enrolled beneficiaries results in higher payment from services visits and hence lower enrollment payments are sufficient.

Experiment set-up: We use $\mathbb{E}[c_{e,t}] = \mathbb{E}[c_{d,t}] = 2$, as stated in Table 1. The distribution of both $c_{e,t}$ and $c_{d,t}$ is assumed to be uniform over $\{1, 2, 3\}$.

Algorithm 1: Simulation and Training

Learnable Parameters : θ_s (Service) and θ_e (Enrollment)
Input : Initial eligible population : N_0, Monthly change : $\mathbb{E}[c_{e,t}], \mathbb{E}[c_{e,t}]$, Maximum Enrollment : K, Service Rate : α, Budget : B, Time Period : T, Reward Function : r_θ, Rationality Parameter : μ
for each step $1 \leq t' \leq$training epochs **do**
  Sample initially enrolled $n_0 \sim$Uniform$(\lfloor N_0/2 \rfloor, \dots, N_0)$
  for $1 \leq t \leq T$ **do**
    Compute eligible actions $A(t) = \{0, \dots, \min(K, N_{t-1} - n_{t-1} + c_{e,t})\} \times \{0, \dots, \lfloor \alpha n_{t-1} \rfloor\}$
    Sample actions $x_{e,t}, x_{s,t} \sim \pi_\theta(x_{e,t}, x_{s,t}) = \frac{e^{r_\theta(x_{e,t}, x_{s,t})/\mu}}{\sum_{(x'_{e,t}, x'_{s,t}) \in A(t)} e^{r_\theta(x'_{e,t}, x'_{s,t})/\mu}}$
    Sample $c_{d,t} \sim Uniform(\mathbb{E}[c_{d,t}] - 1, \mathbb{E}[c_{d,t}], \mathbb{E}[c_{d,t}] + 1)$ and $c_{e,t} \sim Uniform(\mathbb{E}[c_{e,t}] - 1, \mathbb{E}[c_{e,t}], \mathbb{E}[c_{e,t}] + 1)$
    $n_t = n_{t-1} + x_{e,t} - c_{d,t}$
    $N_t = N_{t-1} + c_{e,t} - c_{d,t}$
  end for
  Store action history $\{(x_{s,1}, x_{e,1}), \dots, (x_{s,T}, x_{e,T})\}$ in replay buffer $\mathcal{D}$
  if $t' \% update\_frequency = 0$ **then**
    Sample a minibatch of experiences $d_1, \dots$ from $\mathcal{D}$. Update θ_s, θ_e by minimizing $\sum_i (V_\theta(d_i) - B)^2$
  end if
end for

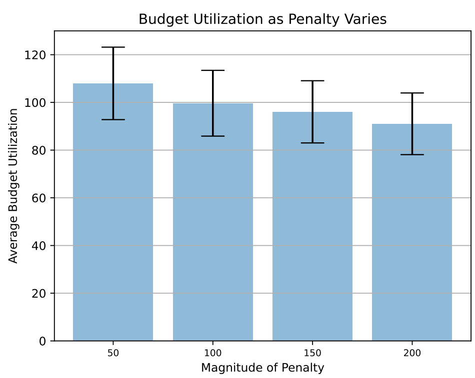

Figure 11: Budget Utilized as the Penalty term M of the alternate approach is varied for pregnant women task.

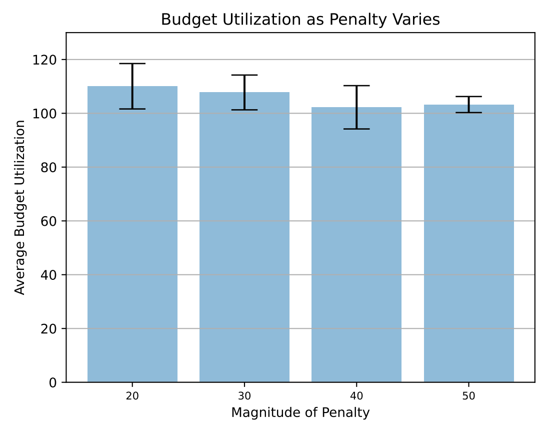

Figure 12: Budget Utilized as the Penalty term M of the alternate approach is varied for pregnant women task.

Data source: This work is part of a broader multi-year study. The broad statistical data and the conclusion inferred from informal group discussions were already known prior to the start of this incentive paper.

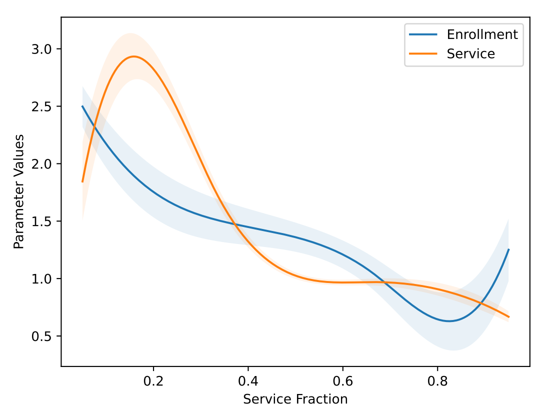

Figure 13: Parameter Values as service fraction is varied for pregnant women task.

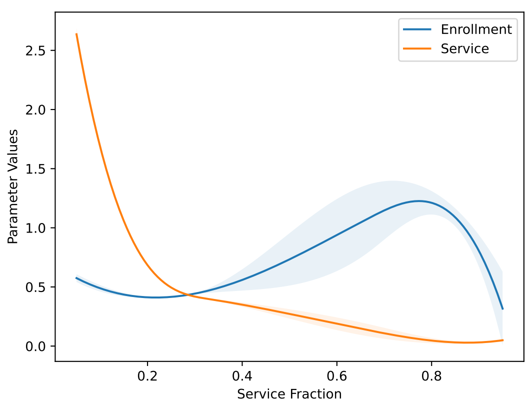

Figure 14: Parameter Values as service fraction is varied for children task.