

SMaRT: Online Reusable Resource Assignment and an Application to Mediation in the Kenyan Judiciary

Shafkat Farabi¹, Didac Marti Pinto², Wei Lu³, Manuel Ramos-Maqueda³, Sanmay Das¹, Antoine Deeb³ and Anja Sautmann³

¹Virginia Tech

²University College London

³The World Bank

mfarabi@vt.edu, didac.marti.pinto.19@ucl.ac.uk, wlu4@worldbank.org,
mramosmaqueda@worldbank.org, sanmay@vt.edu, adeeb1@worldbank.org,
asautmann@worldbank.org

Abstract

Motivated by the problem of assigning mediators to cases in the Kenyan judicial system, we study an online resource allocation problem where incoming tasks (cases) must be immediately assigned to available, capacity-constrained resources (mediators). The resources differ in their quality, which may need to be learned. In addition, resources can only be assigned to a subset of tasks that overlaps to varying degrees with the subset of tasks other resources can be assigned to. The objective is to maximize task completion while satisfying soft capacity constraints across all the resources. The scale of the real-world problem poses substantial challenges, since there are over 2000 mediators, and a multitude of combinations of geographic locations (87) and case types (12) that each mediator is qualified to work on. Together, these features—unknown quality of new resources (newly onboarded mediators), soft capacity constraints (due to the mandate to assign cases without delay), and high-dimensional state space—make existing scheduling and resource allocation algorithms either inapplicable or inefficient. We formalize the problem in a tractable manner, using a quadratic program formulation for assignment and a multi-agent bandit style framework for learning. We demonstrate the key properties and advantages of our new algorithm, SMaRT (Selecting Mediators that are Right for the Task), compared with baselines on some stylized instances of the mediator allocation problem. We then turn to considering its application to real-world data on cases and mediators from the Kenyan Judiciary. SMaRT outperforms baselines and allows for controlling the tradeoff between the strictness of the capacity constraints and overall case resolution rates, both in situations where mediator quality is known beforehand and when the problem is bandit-like in that learning is part of the problem definition. On the strength of these results, we plan to conduct a randomized controlled trial in which we deploy SMaRT

in the Judiciary’s mediation management system.

1 Introduction

As part of its Court Annexed Mediation (CAM) system, the Judiciary of Kenya diverts civil cases to court appointed mediators before they are tried in front of a judge. If a case fails to reach a settlement in mediation, it is escalated to a judicial trial. These mediators are accredited to operate in specific *court stations* (serving different geographic regions) and for specific *case types* (for example, family matters or business disputes). In addition to differences in which cases mediators are accredited to handle, there is also strong evidence that mediators differ in their ability to help parties reach a successful resolution. However, for reasons of time constraints, diminishing performance and equity, it is considered desirable to limit the number of cases each mediator is handling concurrently.

Once a case arrives into the system, a CAM officer is responsible for finding a suitable mediator who is accredited for the relevant court station and case type, and who is not over-committed in terms of the number of cases they are handling. The CAM officer communicates with the mediator, and, if the mediator is willing, assigns them the case. Until recently, this process was manual, which is both complicated and prone to inefficiencies. CAM officers must perform a complex, dynamic task that involves assessing mediator quality in addition to managing load across mediators. Moreover, any individual CAM officer may not account for the effects of their assignment decisions on overall mediation capacity, as mediators often work in more than one court station. Analysis of historical data and continual monitoring have established that manual assignment is inefficient (barely outperforming round-robin allocation in terms of resolution rates) and unequal (in terms of cases awarded across mediators conditioned on past performance), not to mention time consuming.¹

For the reasons above, the Judiciary is interested in automating the process of case-mediator matching. If successful, an automated assignment tool can significantly reduce the number of cases that are returned to the court for a judicial trial, in-

¹Source: Private communications.

creasing the successful resolution of cases between the parties and reducing the strain and backlog in the Judiciary.

The first challenge in designing such a tool is to determine how good a mediator is at reaching case settlements. We quantify a given mediator’s contributions to successful resolution by adapting value-added models, typically used to measure the performance of teachers, to our setting. Continual entry and exit of mediators also means that the tool needs to balance exploring new mediators and exploiting identified better mediators, as in the Multi-Armed Bandit (MAB) literature [Slivkins, 2019]. To accommodate uncertainty and learning, we maintain Gaussian belief distributions over mediator value added, centered around our point estimate. MABs and their variants have, of course, been researched extensively and used in many real-world applications such as online recommendation systems, ad-auctions, etc. [Bouneffouf *et al.*, 2020]. Mediator assignment resembles the Bandits-with-Knapsacks (BwK) model of Badanidiyuru *et al.* [2018], but with a few key differences. Notably, in the BwK model, there is a strict capacity limit for each resource over the entire course of time. In mediator assignment, this constraint is (1) dynamic, in that resources replenish – when a mediator finishes with a case, their capacity is freed up – and (2) soft, because mediators can temporarily go over capacity (and in manual assignment often do), although this is considered undesirable by the Judiciary.

The second major challenge is that of online assignment across mediators with different accreditation sets (i.e., court stations and case types they are accredited for) and values added. The closest existing work that addresses a similar challenge is the OM-RR-PA online assignment of reusable resources algorithm [Dong *et al.*, 2021]. Again, there are several key differences with our setting. First, OM-RR-PA operates with hard capacity constraints on each reusable resource and allows delayed assignments of arriving tasks to avoid overload of resources. Second, it assumes that the time a resource will be in use once allocated is deterministic and known. Finally, OM-RR-PA does not address learning, and including the complexities of learning when mediators are accredited across many different possible subsets of tasks would be non-trivial in that setting.

In summary, mediator assignment requires a new approach, which is the goal of this paper. We address all the challenges above with our proposed algorithm SMaRT (Selecting Mediators that are Right for the Task). SMaRT assignment decides about mediator assignments online. It uses econometric estimation of the value added of a mediator (the percentage point increase in the base probability of case resolution if that mediator is assigned to the case), supported by Bayesian posterior updates. A quadratic programming (QP) formulation penalizes over-capacity allocations. The relative weight given to case resolution rates vs. capacity constraint violations can be tweaked by tuning a parameter.

After validating the behavior of SMaRT in two stylized simulations, we turn to our main empirical analysis, which uses real data on case arrivals and mediators in the Kenyan Judiciary. We demonstrate the benefits of SMaRT relative to several sensible baselines, including round-robin allocation and capacitated versions of greedy and Thompson Sampling. SMaRT typically outperforms baselines in maximizing agree-

ment rates while maintaining reasonably low mediator overload rates. This validation provides justification for a pilot deployment of SMaRT in Kenya in the near future.

2 Background and Related Work

Prior work links judicial inefficiency to weaker economic performance [Kapopoulos and Rizos, 2024] and highlights substantial scope for improving court operations [Ferro *et al.*, 2020; Ippoliti and Tria, 2020]. In Kenya, active case management has been associated with improved judicial efficiency and reduced crime rates [Odhiambo, 2016]. Nonetheless, civil cases remain severely backlogged. To address this, the Kenyan Judiciary introduced Court-Annexed Mediation (CAM) to reduce the time and cost of dispute resolution. Court trials often span years and incur substantial legal and administrative costs; mediation offers a faster, more flexible, and less adversarial alternative. Under CAM, civil cases are screened and referred to mediation, with court-appointed mediators facilitating negotiation between parties within a 70-day window. Cases that reach agreement are resolved; others proceed to trial. A key challenge to scaling CAM lies in mediator assignment: historical data indicate that mediators persistently achieving low case agreement rates continue to receive cases, and many mediated disputes ultimately return to court. We therefore ask how an algorithmic approach that leverages case-level and mediator-level data can optimally assign mediators to incoming cases to improve overall case outcomes.

Multi-armed bandits (MAB) are a standard framework for exploration–exploitation problems where the value of different options is incompletely known [Slivkins, 2019]. Bandits with Knapsacks (BwK) extend this framework to settings with resource constraints, where each arm pull yields a reward and incurs stochastic resource costs subject to global budgets [Badanidiyuru *et al.*, 2018]. While conceptually related to our mediator load constraints, key differences outlined previously prevent direct application of BwK methods to our setting. Bernasconi *et al.* [2024] extend BwK to Bandits with Replenishable Knapsacks, which handle deterministically replenishing resources but cannot accommodate the constraint violations that arise because cases must be assigned a mediator on arrival. Agrawal and Devanur [2014] in turn generalize BwK by allowing penalties and diminishing returns, constraining the time-averaged outcome vector to lie in a given convex set and defining total reward as a concave, Lipschitz-continuous function of the outcome vector. However, their formulation does not support replenishable resources.

MABs are widely used to learn agent quality in online resource allocation and multi-agent systems [Tran-Thanh *et al.*, 2014; Ho *et al.*, 2014; Liu and Chen, 2017]. MABs are employed for adaptive worker selection under budget and load constraints in crowdsourcing [Rangi and Franceschetti, 2018] and for online matching of compute to incoming jobs [Xu *et al.*, 2020]. MABs have also been deployed in socially impactful domains, including health information delivery in low-resource settings [Mate *et al.*, 2020; Mate *et al.*, 2022; Dasgupta *et al.*, 2025; Verma *et al.*, 2023], maternal healthcare [Boehmer *et al.*, 2025; Wang *et al.*, 2023; Mate *et al.*, 2021], public health adherence [Liang *et al.*, 2025a], environmen-

tal monitoring [Martin and Johnson, 2020], corporate social responsibility [Ron *et al.*, 2021], and disaster management [Liang *et al.*, 2025b].

The other thread of relevant literature is on modeling online bipartite matching problems with linear programs (LP) to design allocation algorithms. This general technique has been applied to several domains, including online ad auctions [Ho *et al.*, 2013; Devanur and Hayes, 2009], crowd-sourced task assignment [Ho and Vaughan, 2012], and organ transplantation [Li *et al.*, 2019]. These works typically assume that agents are impatient and resources disposable. Dong *et al.* [2021] adapt reusable resources with patient agents in OM-RR-PA. Similarly, we model mediator allocation to cases using a bipartite graph and apply a quadratic programming framework for optimization. Mediator assignment must deal with several novel issues that OM-RR-PA does not confront, including stochastic resolution times, and soft capacity constraints. Dickerson *et al.* [2021] study online bipartite matching with impatient agents and reusable resources under stochastic replenishment. Their framework does not permit the soft capacity constraint violations allowed in our setting and assumes each resource serves at most one agent at a time, whereas our application requires simultaneous multiple uses. The full-information mediator allocation problem can be formulated as an MDP [Puterman, 1994], but this does not accommodate online learning of mediator VAs and scales poorly to real world scenarios due to state-space explosion. Constraint programming approaches (e.g., RCPSP [Cavalcante *et al.*, 2013]) can capture mediator capacity and licensing constraints, but cannot capture case resolution rate optimization and immediate case assignments.

3 SMaRT Assignment

3.1 Court Annexed Mediation (CAM) in Kenya

In the CAM process, a CAM officer assigns a mediator to each incoming case. Mediators are accredited only for specific case types and court stations, so assignments are drawn from the pool of mediators eligible for the case’s (type, station) pair. The CAM secretariat prefers that mediators handle no more than three concurrent cases.

Until recently, CAM officers navigated these constraints manually, supported by a digital mediation platform called ‘Cadaster’ that was developed with the support of the World Bank and tracks mediator accreditation and cases as well as mediator assignments and past case outcomes. Our value-added estimates (see below) suggest that even mediators that performed consistently below average were often assigned in preference to higher-performing mediators, and the 3-case capacity constraint was frequently needlessly violated. CAM is now piloting a “smart assignment” feature within Cadaster that automatically proposes the most suitable mediator for each case for CAM officer approval. The algorithm developed here is proposed as the back end of that feature.

3.2 Value Added Estimation

Inspired by the literature on the contribution of teachers to student learning outcomes, we model a mediator’s contribution to case resolution using a value-added model [Kane and

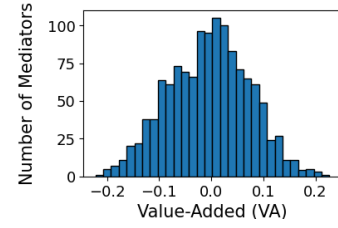


Figure 1: Histogram of empirical mediator VA estimates

Staiger, 2008]. Analysis of historical data shows that mediator skill is idiosyncratic and not predicted by individual level-observables.² To obtain a mediator’s Value Added (VA), we estimate the probability of case resolution net of mediator contribution p_i based on the regression $Y_{ij} = X_i\beta + \nu_j + \epsilon_{ij}$. $Y_{ij} = 1$ if case i assigned to mediator j reaches agreement, and 0 otherwise. X_i is a vector of controls including case type, court station, time period, and referral mode fixed effects, ν_j is a vector of mediator fixed effects, and ϵ_{ij} is the error term. We residualize outcomes as $\hat{r}_i = Y_{ij} - X_i\hat{\beta}$ and adapt shrinkage to estimate VA for mediators with at least 2 cases. For each mediator j with n_j cases, we split cases chronologically into two groups and compute average residuals $\bar{r}_j^1$ and $\bar{r}_j^2$. We estimate case-level variance σ_ϵ^2 using $\text{Var}(\hat{r}_i - \bar{r}_j^t)$, where $t = 1, 2$ indexes the (chronological) group, mediator-level variance σ_μ^2 via $\text{Cov}(\bar{r}_j^1, \bar{r}_j^2)$, and group-level variance σ_θ^2 as $\text{Var}(\hat{r}_i) - \sigma_\mu^2 - \sigma_\epsilon^2$. Mediator VA is then estimated as $\hat{\mu}_j = \frac{\lambda_j}{n_j} \sum_{i=1}^{n_j} \hat{r}_i$, with shrinkage factor $\lambda_j = \frac{\sigma_\mu^2}{\sigma_\mu^2 + h_j}$ and $h_j = \frac{1}{2} \left(\sigma_\theta^2 + \frac{\sigma_\epsilon^2}{0.5n_j} \right)$. From our estimate of σ_μ^2 , a one standard deviation increase in mediator VA corresponds to 13.36 percentage points increase in the likelihood of case resolution. A histogram of mediator VAs estimated using this method from historical case allocations (2016–2025) is shown in Figure 1.

Estimating Uncertainty and Making Efficient Updates Many bandit algorithms, such as Thompson Sampling, require both point estimates of VA and associated uncertainty in those estimates to balance between exploration and exploitation of arms. Moreover, the above VA estimation procedure is computationally expensive and cannot be run after each case resolution. To resolve this, we define a Gaussian belief over μ_j centered around the VA point estimate $\hat{\mu}_j$. Recall that the probability of successful resolution of case i with base resolution probability p_i assigned to mediator j with VA μ_j is given by $(p_i + \mu_j)$. We use an indicator random variable Y_{ij} to denote whether the case is successfully resolved. We place a prior $\mu_j \sim \mathcal{N}(0, \sigma_\mu^2)$ and update after observing Y_{ij} via $f(\mu_j | Y_{ij}) \propto (p_i + \mu_j)^{Y_{ij}} (1 - p_i - \mu_j)^{1 - Y_{ij}} \exp(-\mu_j^2 / 2\sigma_\mu^2)$. We further approximate the posterior with a Gaussian $\mathcal{N}(\mu_j | Y_{ij}, (\sigma_j | Y_{ij})^2)$ through moment matching. Upon observing an additional outcome $Y_{i'j}$, this posterior is used as the prior to update $f(\mu_j | Y_{ij}, Y_{i'j})$ sequentially. In practice, posterior means computed in this way initially deviate moderately

²Deciding a mediator’s case load *a priori* only based on such observables would also violate ethical standards of non-discrimination.

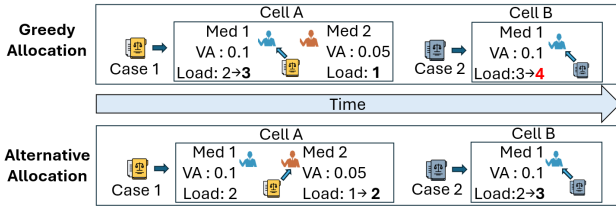


Figure 2: A hypothetical situation where holding back the best mediator is beneficial in the long run. Greedily assigning Case 1 in Cell A leads to overload when a new case arrives in Cell B. Arrows indicate caseload changes after assignment. Red caseload indicates overload.

from the explicit VA estimations above,³ but these deviations compound over multiple updates. We therefore recalibrate the posteriors weekly by repeating the full VA estimation procedure and setting the new prior at $\mathcal{N}(\hat{\mu}_j, (\sigma_j | Y_{ij}, Y_{i'j}, \dots)^2)$.

3.3 Assignment as a Quadratic Program

Our goal is to design an automated case allocation system that maximizes case resolution rate. Without capacity constraints, this would be a conventional MAB problem. Within each cell, the algorithm would simply identify the best arm (mediator), and then play that arm repeatedly. However, local greedy assignment does not translate to global optimality in the presence of capacity constraints and overlapping accreditation sets as outlined above. There can be value in “holding back” a high-performing mediator based on the “opportunity cost” of not being able to assign this mediator to a future case.

Abstracting from learning, Figure 2 illustrates a stylized scenario with mediators 1 and 2, where $\mu_1 > \mu_2$. Mediator 1 can serve both cells, while mediator 2 is restricted to Cell A. If a case arrives in Cell A followed by one in Cell B, a greedy policy assigns the first case to mediator 1, leading to overload for this mediator when the next case arrives in Cell B. Assigning the Cell A case to mediator 2 instead preserves mediator 1 for Cell B, improving capacity management at the cost of a lower expected agreement rate. By a similar logic, holding back high-performing mediators in one cell can prevent very low-performing assignments elsewhere, thereby increasing overall agreement rates.

Note that our algorithm must have tolerance for case overloads. With stochastic arrivals and a requirement to immediately assign cases, there is a non-zero probability of all mediators in a cell reaching capacity. To find optimal allocations, we model a relaxed version of the problem as a mathematical program. Our approach accommodates caseload constraint violations by introducing a penalty term in the objective. This term is quadratic to deter the best mediator from successive overloads, making it a quadratic program (QP).

In the QP, mediators and cases are indexed by u and v . U denotes the set of all mediators. V_r denotes the case(s) that have arrived prior to solving the QP. Future cases that may arrive are represented with shadow cases V_{sh} sampled from a distribution. Mediator VA is μ_u , and case fixed effects are p_v , interpreted as the probability of resolution with a zero-VA mediator. Mediator caseload and capacity are given by $L(u)$

and $C(u)$. The allocation state is represented by a bipartite graph $G_s = (U \cup V, E)$, where $(u, v) \in E$ indicates that mediator u is accredited for case v . In the QP, x_{uv} denotes the (possibly fractional) allocation of case v to mediator u . The model runs over a time horizon T with timesteps t . $t_a(v)$ denotes arrival time for case v . We solve this QP with small time horizon ($T=10$) and assume case mediation durations outlast this short time horizon.

$$\begin{aligned} \max_{x, \xi} \quad & \sum_{e \in E} x_e (\mu_u + p_v) - \lambda \sum_{u \in U} \xi_u^2 \\ \text{s.t.} \quad & \sum_{e \in E(v)} x_e = 1 \quad \forall v \in V_r \end{aligned} \quad (C1)$$

$$\sum_{e \in E(v)} x_e \leq 1 \quad \forall v \in V_{sh} \quad (C2)$$

$$\begin{aligned} L(u) + \sum_{e \in E(u)} x_e \mathbb{1}[t_a(v) \leq t] \\ \leq C(u) + \xi_u \quad \forall u \in U, \forall t \in [T] \end{aligned} \quad (C3)$$

$$0 \leq x_e \leq 1 \quad \forall e \in E \quad (C4)$$

$$\xi_u \geq 0 \quad \forall u \in U \quad (C5)$$

C1 enforces all real cases to be fully allocated, while C2 allows shadow cases to go unassigned. C3 imposes a soft capacity constraint ensuring mediators remain within capacity, with violations captured by the slack penalty term ξ_u . These violations are penalized quadratically in the objective to trade off expected case resolution against mediator overload. The parameter λ controls the severity of overload penalty. A discussion of the QP solvability and time complexity can be found in Appendix 6.1 in the full version [Farabi *et al.*, 2026].

3.4 SmaRT Assignment Algorithm

Now we bring everything together in SmaRT (Algorithm 1). Records of unresolved cases (L) and mediation outcome history (H) for each mediator are maintained and updated as the algorithm executes. Upon case arrival, shadow cases are sampled and a state graph (G_s) is constructed with mediators, cases, and accreditation. The QP is then solved using Gurobi. This solution consists of fractional assignments of cases to mediators. Next, the algorithm looks at all the fractional assignments to mediators for the arriving case, and identifies the one ($\hat{u}$) with highest fractional assignment to the arriving case v_r . SmaRT allocates the case to this mediator.

Bandit Learning If the true mediator VAs are known, then the QP reflects an approximation of the optimal assignment. Later we test this assumption in situations where we give

Algorithm 1 SmaRT

- 1: **Input:** λ , arriving case v_r , mediators U , history H , caseloads L , horizon T
 - 2: $\{\mu_u\}_{u \in U} \leftarrow \text{CalculateVA}(H)$
 - 3: $V \leftarrow \{v_r\} \cup \text{SampleCases}(0, T)$
 - 4: $E \leftarrow \{(u, v) : v \in V, u \in \text{FindAccreditedMeds}(v)\}$
 - 5: $G_s \leftarrow \text{InitializeStateGraph}(V, E)$
 - 6: $QP \leftarrow \text{InitializeQP}(G_s, L, \lambda)$
 - 7: $x \leftarrow \text{Solve}(QP)$
 - 8: $\hat{u} \leftarrow \arg \max_{u: (u, v_r) \in E} x_{u, v_r}$
 - 9: $\text{Allocate}(v_r, \hat{u})$
 - 10: $H \leftarrow H \cup \{(v_r, \hat{u})\}; L[\hat{u}] \leftarrow L[\hat{u}] + 1$
-

³See Appendix 6.4 and 6.5 in the full version [Farabi *et al.*, 2026]

Mediator	VA	Accreditations	
		Scenario 1	Scenario 2
Med-1	0.10	Cell A, Cell B	Cell A, Cell B
Med-2	0.05	Cell B	Cell A
Med-3	-0.10	Cell B	Cell B

Table 1: Mediator VAs and accreditations in two stylized scenarios.

SMaRT the true VAs (we refer to SMaRT with true VAs as SMaRT(GS) (for Gold Standard)). When the true VAs are unknown, we have to confront the exploration-exploitation dilemma. While some of this is mitigated in our setting by the mediator capacity constraints and the exploration they enforce, we also consider a variant of SMaRT that incentivizes exploration in a manner analogous to Thompson Sampling. The two versions of SMaRT we consider that incorporate learning differ only in how they implement *CalculateVA(H)*:

- SMaRT(E): As cases arrive, mediator VAs are defined to be the expectation (mean) of corresponding VA beliefs.
- SMaRT(S): As cases arrive, mediator VAs are sampled from corresponding VA beliefs, akin to Thompson sampling, encouraging exploration.

4 Results

4.1 Benchmark Policies

We compare SMaRT with the following allocation policies:

- Least Load: The accredited mediator with lowest caseload is assigned to an arriving case. Ties are broken randomly.
- Greedy*: Assign each arriving case to the accredited mediator with the highest mean VA belief. Mediators at capacity are set aside until all other accredited mediators are equally overloaded.
- TS*: Assign each arriving case to the accredited mediator with the highest sampled VA from the belief. Mediators at capacity are set aside until all other accredited mediators are equally overloaded.
- Upper Bound: To upper-bound agreement rates, we ignore capacity and always assign the highest-VA accredited mediator by true VA.

We define **Agreement Rate** as the share of arriving cases that end in agreement rather than return to court and **Overload-Case-Days/Med-Year** (OCDM) as $\frac{365}{mT} \sum_j \sum_d \max(OL_{j,d}, 0)$, which expresses the degree of overload an average mediator experiences. Here $OL_{j,d}$ is mediator j 's case overload on day d . T is the total runtime (in days), and m is the number of mediators. Both metrics are averaged across multiple (128) independent runs when reported in the following sections.

4.2 Stylized Scenario Analysis

To understand and validate the behavior of SMaRT, we crafted two small-scale scenarios with two cells and three mediators.

Mediator VAs and accreditations are shown in Table 1. Med-1 has the highest VA and is shared among cells A and B. Med-2 is slightly worse and depending on the scenario, is either restricted to cell A or B. Med-3 has the lowest VA and is restricted only to cell B. Cases arrive at a slightly higher rate at Cell A. The base p values are set to 0.5 for all cases.

Algorithm	λ	Known VA (SMaRT(GS))		Learning + Mean VA (SMaRT(E))		Learning + Sampled VA (SMaRT(S))	
		Agreement Rate	OCDM	Agreement Rate	OCDM	Agreement Rate	OCDM
Least Load	—	0.406	25.13	—	—	—	—
Greedy*	—	0.439	41.71	0.425	36.31	—	—
TS*	—	—	—	—	—	0.404	34.17
SMaRT	0.01	0.453	94.12	0.438	79.16	0.437	67.67
	0.05	0.449	54.34	0.430	46.95	0.428	45.09
	0.10	0.444	44.85	0.427	39.03	0.427	38.46
	0.50	0.439	38.53	0.422	33.04	0.423	32.00
Upper Bound	—	0.460	132.15	—	—	—	—

Table 2: Scenario 1 results under different VA information assumptions. As expected, mediator overload (OCDM) decreases as λ increases, but so does case agreement rate.

Table 2 shows performance results in terms of agreement rates and average OCDM in Scenario 1. The “Known VA” column isolates the tradeoff between agreement rate and overload (we see similar patterns in the learning variants). SMaRT behaves as expected. With lower λ s, it achieves better case agreement rates than Greedy*, at the cost of higher overload. As λ increases, SMaRT distributes caseload more evenly, with some decline in case agreement rate. Figure 3 visualizes this effect for SMaRT(GS) in individual runs of the simulation. At $\lambda = 0.01$, Med-1 takes on the bulk of cases at both court-stations, leading to significant overloads (red) while Med-3 is never used. As the overload penalty λ increases, SMaRT(GS) increases allocations to Med-2 and Med-3 to hold back Med-1 for cases in Cell A.

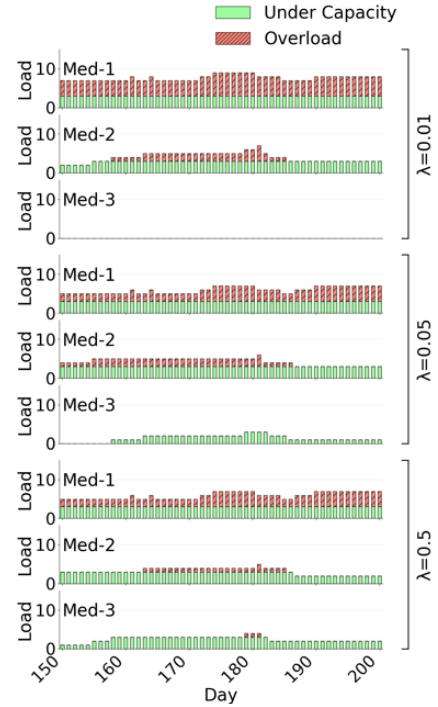


Figure 3: An example of case allocations (with known med VAs) in Scenario 1 by SMaRT. At lower λ , SMaRT favors overloading the best mediator (Med-1) to maximize case agreement rate. At higher λ , SMaRT sacrifices agreement rate and focuses more on efficient case-load management, evidenced especially by the reduced red bars on Med-1 and the increased allocations to Med-3.

Algorithm	λ	Known VA (SMaRT(GS))		Learning + Mean VA (SMaRT(E))		Learning + Sampled VA (SMaRT(S))	
		Agreement Rate	OCDM	Agreement Rate	OCDM	Agreement Rate	OCDM
Least Load	—	0.380	4.13	—	—	—	—
Greedy*	—	0.406	5.18	0.394	5.06	—	—
TS*	—	—	—	—	—	0.392	4.82
SMaRT	0.01	0.452	94.30	0.417	62.06	0.417	53.80
	0.05	0.440	43.90	0.403	25.34	0.407	29.29
	0.10	0.430	22.91	0.398	14.34	0.400	17.12
	0.50	0.410	5.43	0.394	5.54	0.394	5.67
Upper Bound	—	0.460	132.15	—	—	—	—

Table 3: Scenario 2 results under different VA information assumptions. Note the substantial gains in agreement rate at low λ s for SMaRT compared with Greedy* when VA is known. Similar to Scenario 1, Agreement Rate and OCDM decrease with λ , as intended.

In scenario 2, both cells now have access to alternatives to Med-1. Cell A can fall back to Med-2 with a net positive VA, while cell B has to fall back to low-performing Med-3. Holding Med-1 back in cell A can help avoid using Med-3 in cell B. We present simulation results from different settings in Table 3. Greedy* has a lower resolution rate but also less overload than in Scenario 1, because Med-3 now more often takes on cases in cell B. SMaRT(GS) with low λ almost exclusively uses Med-1 as in Scenario 1. Even at higher λ s, SMaRT(GS) achieves higher resolution rates than Greedy*, at the expense of higher overloads, because SMaRT(GS) finds it beneficial to overload Med-1 slightly in cell B before assigning cases to Med-3. The dynamics are similar with learning enabled.

Analysis of shadow prices As a sanity check, we analyze mediator “shadow prices,” given by the dual variables of the QP constraints. These prices are also useful for interpretability of SMaRT as well as for future primal-dual online variants of the algorithm. Figure 4 reports mean shadow prices under SMaRT. As expected, shadow prices increase with λ and correlate with mediator VAs. However, Med-1’s shadow price is much higher than Med-2’s, reflecting Med-1’s added value for Cell B. This produces the intended “holding back” effect: cases in Cell B become less likely to use Med-1 when they may be more valuable elsewhere. As λ increases, this effect strengthens. From an interpretability perspective, the “shadow price” of a mediator reflects their value when assigned to a different case. The algorithm tries to assign the best possible (highest-value generating) mediator to a case, net of the value of “holding them back” for a case where they would be even more valuable. Thus, SMaRT assigns mediators based not only on immediate VA but also on the opportunity cost of using them now, while allowing administrators to tune capacity

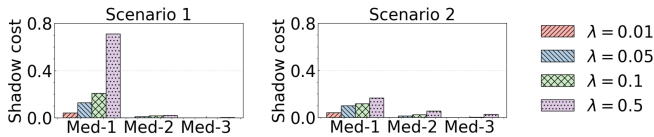


Figure 4: Mean shadow price by mediator during allocation by SMaRT on the stylized examples for varying λ . As λ increases, the shadow prices increase, demonstrating increased marginal benefit of holding them back for future cases. Note also Med-1’s relatively higher value in Scenario 1, where only they can cover cases in Cell A, versus Scenario 2.

Algorithm	λ	Agreement Rate	OCDM	Caseload Gini Index
Least Load	—	0.467	0.10	26.1%
Greedy*	—	0.569	0.32	74.6%
SMaRT(GS)	0.01	0.617	52.68	87.3%
	0.02	0.605	33.22	83.8%
	0.03	0.596	24.30	81.8%
	0.04	0.593	19.25	80.7%
	0.05	0.590	15.96	79.8%
	0.10	0.584	8.82	77.9%
	0.50	0.574	0.94	74.7%
Upper Bound	—	0.659	122.01	96.5%

Table 4: Simulations with true VA information available to the allocation algorithms on data collected from the judiciary. At $\lambda = 0.01$, SMaRT achieves an agreement rate close to the upper bound but with much lower aggregate overload (OCDM) and much higher equity across mediators (Gini Index). As λ increases, the agreement rate declines somewhat while overload and equity improve substantially.

violations through a single parameter.

4.3 Real Data

Simulation framework We now turn to our main experiment on real data. We use mediation data collected from the Kenyan Judiciary covering the period 2016–2025. The dataset includes 30,633 cases involving roughly 2,100 mediators across 13 case types and 87 court stations, with information on arrivals, assignments, conclusion dates, outcomes, mediator accreditations, and station preferences. Based on this data, we develop a case assignment simulation to benchmark assignment strategies including SMaRT(GS), SMaRT(E), SMaRT(S) and compare them. We use the Value Added model to compute the distribution of mediator VAs (μ), as well as baseline probabilities of case resolution (p). Mediator VAs are independently sampled from this distribution during simulations. We fit Poisson distributions for case arrival across (case-type, court-station) pairs on the data and sample case arrivals independently and identically. These distributions are also used to sample shadow cases within SMaRT. Case outcomes are sampled from Bernoulli($\Pi_{[0,1]}(p_{\text{case}} + \mu_{\text{med}})$). Survival analysis on the case mediation time indicates that duration is dependent on case-type and outcome. We fit log-normal models to the mediation duration conditional on case-type and outcome for sampling mediation duration in the simulation. For each setting, we average results over 128 seeded simulations. For a given seed, all inputs (case arrivals, mediator VAs, and accreditation patterns) are identical across methods. Since every arriving case must be assigned immediately, differences in performance reflect only how each method matches mediators to cases.

Results Simulation results under full information of mediator VAs are reported in Table 4. First, note that SMaRT(GS) with $\lambda = 0.01$ attains a successful resolution rate that is close to the best achievable— but with considerably better capacity management, since Upper Bound by design gives cases only to the best mediator for each (court station, case type) cell. Relative to Greedy*, this is nearly a 8% gain in performance, but it comes at the cost of substantially higher overload: OCDM rises to 52, compared to 0.32 under Greedy*. OCDM is an aggregate measure of our goal to minimize and equalize case

overload. We therefore also compute Gini Index which shows the level of inequality in case assignment across mediators. With a high case influx rate, lower overloads correlate with equity of caseload distribution, which makes Gini Index a good comparative measure. As λ increases, the agreement rate goes down but so do OCDM and the Gini Index, demonstrating the “tunability” of SMaRT.

We now benchmark performance when mediator VAs are unknown and must be learned from observed case outcomes. These results are presented in Table 5. We consider two different initializations of the learning process for the mediators. In one, we initialize all mediator beliefs as $\mathcal{N}(0, \sigma_\mu^2)$ – this is the “blank slate” version where we know nothing about the mediators in advance, which could be used in an environment where the idea of VA estimation is entirely new, or where all the mediators arrive at the same time. In the second, VA beliefs are instead initialized in a manner that replicates what we might expect to be able to do when we deploy this system in practice for the Kenyan Judiciary. We iterate through case records from the prior six years and sample each case outcome using its p -value and the assigned mediator’s VA. Using these outcomes, we compute an initial belief for each mediator VA. It is interesting to note that there is considerable variation in the number of cases mediators have mediated in the real data. As a result, the initial beliefs exhibit substantial variation in their uncertainty. Some highly experienced mediators have narrow beliefs concentrated near their true VA, whereas newer mediators have wider beliefs centered near 0.

First, consider the results from the blank slate experiment on the left panel of Table 5. Greedy* attains a higher agreement rate than TS*, while the latter is more equitable and leads to less OCDM, a byproduct of greater exploration. Both SMaRT(E) and SMaRT(S) behave similarly to how SMaRT(GS) behaves with known VAs: they deliver higher agreement rates at smaller λ s, and increasing equity and reduced OCDM as λ increases. At small λ s they outperform their relevant baselines (Greedy* for SMaRT(E) and TS* for SMaRT(S)) in terms of case agreement rates. In this experiment, TS* (and SMaRT(S)) are strictly worse than Greedy* (and SMaRT(E)). This is because capacity management itself serves to induce sufficient exploration. It is notable that SMaRT(E) consistently exceeds Greedy* in agreement rate.

When we turn to the “warm start” paradigm where initial beliefs are calculated from data (the right panel of Table 5), this picture changes. Now, TS* performs significantly better than Greedy* both in agreement rate and in managing overloads. The explanation for this lies in uneven knowledge regarding mediator VAs. Greedy* focuses on exploiting experienced mediators it already has accurate estimates of VA for, neglecting to explore less experienced mediators, and thus failing to identify inexperienced mediators with high VA. Conversely, TS* does explore these mediators and is able to exploit the high true VA mediators among them.

For the same reason, but exacerbated because of its primary focus on high agreement rates and more willingness to tolerate overload, with low λ , the behavior of SMaRT(E) is much poorer than in other experiments (and worse than Greedy*). Increasing λ in this case leads to benefits in not just equity across mediators but also in agreement rates by

Algorithm	λ	Initial belief set to $\mathcal{N}(0, \sigma_\mu^2)$			Initial belief calculated from data		
		Agreement Rate	OCDM	Caseload Gini Index	Agreement Rate	OCDM	Caseload Gini Index
Greedy*	–	0.489	0.44	68.4%	0.473	0.38	63.1%
SMaRT(E)	0.01	0.518	72.17	86.4%	0.448	37.22	74.0%
	0.05	0.499	24.69	76.1%	0.461	8.77	64.6%
	0.10	0.494	14.77	73.2%	0.462	4.25	62.9%
	0.50	0.491	4.20	69.6%	0.465	0.57	61.0%
TS*	–	0.484	0.34	53.7%	0.489	0.31	50.8%
SMaRT(S)	0.01	0.499	27.31	61.0%	0.495	9.31	52.3%
	0.05	0.488	8.08	53.1%	0.492	3.17	49.1%
	0.10	0.488	4.31	51.4%	0.490	1.48	47.3%
	0.50	0.484	0.54	49.5%	0.489	0.29	47.3%

Table 5: Simulations on data collected from the judiciary where VAs need to be learned during allocation. SMaRT(E) works well from a “blank slate” initialization, where all mediator VAs must be learned, since the capacity constraint leads to sufficient exploration. SMaRT(S) performs much better when VAs are initialized from real data, because it can better explore mediators on whom the algorithm has collected little data.

inducing exploration as a side effect. SMaRT(S) is by far the best performer, with the sampling inducing exploration even at low λ , and higher values of λ inducing more equity in allocation across mediators without significant reduction in case agreement performance in this case.

5 Conclusion

This paper stems from a six-year collaboration between the World Bank and the Kenyan Judiciary to modernize CAM through Cadaster, a digital mediation platform developed with substantial World Bank technical support to digitize workflows and identify bottlenecks in manual assignment. The paper describes the design of an algorithm –SMaRT– to confront a real challenge faced by the Kenyan Judiciary. SMaRT’s design is grounded in Judiciary operational needs, with the authors advising the IT team and regularly coordinating with the CAM Secretariat and communicating with court officers. While in theory the problem could map to a number of different algorithmic approaches in the literature (bandits, online matching for resource allocation, etc.), when we worked with stakeholders we discovered nuances that made direct application of existing approaches infeasible. In particular, both in practice and conceptually – since case assignment cannot be delayed – the capacity constraints on mediators are *de facto* soft constraints. This feature combined with the accreditation restrictions and the complex nature of learning when new mediators can enter the system raise a number of interesting challenges that necessitated the development of a new algorithm. We believe that there is a range of interesting theoretical and algorithmic work to be done in this space, as governments face similar allocation problems in non-market settings in many domains. SMaRT is queued to be Cadaster’s “smart assignment” feature backend for automated mediator recommendation to CAM officers. SMaRT has been approved for a year-long RCT with ethics approval; if it significantly improves agreement rates and mediator overload, the control group will be transitioned to SMaRT in consultation with the Judiciary.

Ethical Statement

This project is conducted in close collaboration with the Kenyan Judiciary and is designed to reflect the operational requirements of the governing agency. This study was approved by the National Commission For Science, Technology & Innovation of Kenya (NACOSTI) and the ethics review committee at Kenyatta University, Nairobi. The data used in our experiments were collected by the Kenyan Judiciary and provided to the authors in de-identified form, with all personal information about mediators and case parties removed to protect privacy. The de-identified data will be made publicly available in the World Bank microdata library (<https://microdata.worldbank.org>) and the code base will be available through the World Bank's reproducible research repository (<https://reproducibility.worldbank.org/>) after it has gone through mandatory external reproducibility verification.

Acknowledgments

This work is partially supported by NSF Award 2533162, the World Bank's Knowledge for Change Program, and by the World Bank's Global Program on GovTech and Public Sector Innovation through the GTGP Trust Fund. The authors thank our project partners, Hon. Caroline Kendagor, Hon. Grace Sitati, Hon. Moses Wanjala, Clifford Ogutu, Court-Annexed Mediation (CAM), and the Mediation Accreditation Committee (MAC).

References

- [Agrawal and Devanur, 2014] Shipra Agrawal and Nikhil R. Devanur. Bandits with Concave Rewards and Convex Knapsacks. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, page 989–1006, New York, NY, USA, 2014. Association for Computing Machinery.
- [Badanidiyuru *et al.*, 2018] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with Knapsacks. *J. ACM*, 65(3), March 2018.
- [Bernasconi *et al.*, 2024] Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. Bandits with Replenishable Knapsacks: the Best of both Worlds. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*.
- [Boehmer *et al.*, 2025] Niclas Boehmer, Yunfan Zhao, Guojun Xiong, Paula Rodriguez-Diaz, Paola Del Cueto Cibrian, Joseph Ngonzi, Adeline Boatin, and Milind Tambe. Optimizing Vital Sign Monitoring in Resource-Constrained Maternal Care: An RL-Based Restless Bandit Approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39 of AAAI'25, pages 28843–28849. AAAI Press, 2025.
- [Bouneffouf *et al.*, 2020] Djallel Bouneffouf, Irina Rish, and Charu Aggarwal. Survey on Applications of Multi-Armed and Contextual Bandits. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, page 1–8. IEEE Press, 2020.
- [Cavalcante *et al.*, 2013] Victor F. Cavalcante, Carlos H. Cardonha, and Ricardo G. Herrmann. A Resource Constrained Project Scheduling Problem with Bounded Multitasking. *IFAC Proceedings Volumes*, 46(24):433–437, 2013. 6th IFAC Conference on Management and Control of Production and Logistics.
- [Dasgupta *et al.*, 2025] Arpan Dasgupta, Gagan Jain, Arun Suggala, Karthikeyan Shanmugam, Milind Tambe, and Aparna Taneja. Bayesian Collaborative Bandits with Thompson Sampling for Improved Outreach in Maternal Health. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '25*, page 547–555, Richland, SC, 2025. International Foundation for Autonomous Agents and Multiagent Systems.
- [Devanur and Hayes, 2009] Nikhil R. Devanur and Thomas P. Hayes. The Adwords Problem: Online Keyword Matching with Budgeted Bidders under Random Permutations. In *Proceedings of the 10th ACM Conference on Electronic Commerce, EC '09*, page 71–78, New York, NY, USA, 2009. Association for Computing Machinery.
- [Dickerson *et al.*, 2021] John P. Dickerson, Karthik A. Sankararaman, Aravind Srinivasan, and Pan Xu. Allocation Problems in Ride-sharing Platforms: Online Matching with Offline Reusable Resources. *ACM Transactions on Economics and Computation*, 9(3), June 2021.
- [Dong *et al.*, 2021] Zehao Dong, Sanmay Das, Patrick Fowler, and Chien-Ju Ho. Efficient Nonmyopic Online Allocation of Scarce Reusable Resources. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '21*, page 447–455, Richland, SC, 2021. International Foundation for Autonomous Agents and Multiagent Systems.
- [Farabi *et al.*, 2026] Shafkat Farabi, Didac Marti Pinto, Wei Lu, Manuel Ramos-Maqueda, Sanmay Das, Antoine Deeb, and Anja Sautmann. SMaRT: Online reusable resource assignment and an application to mediation in the Kenyan judiciary, 2026. arXiv preprint arXiv:2602.18431.
- [Ferro *et al.*, 2020] Gustavo Ferro, Victoria Oubiña, and Carlos Romero. Benchmarking Labor Courts: An Efficiency Frontier Analysis. *International Journal for Court Administration*, 11(2):7, 2020.
- [Ho and Vaughan, 2012] Chien-Ju Ho and Jennifer Vaughan. Online Task Assignment in Crowdsourcing Markets. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, AAAI'12*, page 45–51. AAAI Press, 2012.
- [Ho *et al.*, 2013] Chien-Ju Ho, Shahin Jabbari, and Jennifer Wortman Vaughan. Adaptive Task Assignment for Crowdsourced Classification. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 534–542, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.
- [Ho *et al.*, 2014] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive Contract Design for Crowdsourcing Markets: Bandit Algorithms for Repeated Principal-Agent Problems. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation, EC '14*,

- page 359–376, New York, NY, USA, 2014. Association for Computing Machinery.
- [Ippoliti and Tria, 2020] R. Ippoliti and G. Tria. Efficiency of Judicial Systems: Model Definition and Output Estimation. *Journal of Applied Economics*, 23(1):385–408, 2020.
- [Kane and Staiger, 2008] Thomas J Kane and Douglas O Staiger. Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation. Working Paper 14607, National Bureau of Economic Research, December 2008.
- [Kapoor and Vaidya, 1986] S Kapoor and P M Vaidya. Fast Algorithms for Convex Quadratic Programming and Multicommodity Flows. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, STOC ’86, page 147–159, New York, NY, USA, 1986. Association for Computing Machinery.
- [Kapopoulos and Rizos, 2024] Panayotis Kapopoulos and Anastasios Rizos. Judicial Efficiency and Economic Growth: Evidence based on European Union data. *Scottish Journal of Political Economy*, 71(1):101–131, 2024.
- [Li et al., 2019] Zhuoshu Li, Kelsey Lieberman, William Macke, Sofia Carrillo, Chien-Ju Ho, Jason Wellen, and Sanmay Das. Incorporating Compatible Pairs in Kidney Exchange: A Dynamic Weighted Matching Model. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC ’19, page 349–367, New York, NY, USA, 2019. Association for Computing Machinery.
- [Liang et al., 2025a] Biyonka Liang, Lily Xu, Aparna Taneja, Milind Tambe, and Lucas Janson. Context in Public Health for Underserved Communities: A Bayesian Approach to Online Restless Bandits. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence*, volume 39 of AAAI’25, page 28195–28203. AAAI Press, 2025.
- [Liang et al., 2025b] Jun Liang, Zongjia Zhang, and Yanpeng Zhi. Multi-Armed Bandit Approaches for Location Planning with Dynamic Relief Supplies Allocation Under Disaster Uncertainty. *Smart Cities*, 8(1), 2025.
- [Liu and Chen, 2017] Yang Liu and Yiling Chen. Sequential Peer Prediction: Learning to Elicit Effort using Posted Prices. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, volume 31 of AAAI’17, page 607–613. AAAI Press, 2017.
- [Martin and Johnson, 2020] David M Martin and Fred A Johnson. A Multiarmed Bandit Approach to Adaptive Water Quality Management. *Integrated Environmental Assessment and Management*, 16(6):841–852, 2020.
- [Mate et al., 2020] Aditya Mate, Jackson A. Killian, Haifeng Xu, Andrew Perrault, and Milind Tambe. Collapsing Bandits and Their Application to Public Health Interventions. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS ’20, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [Mate et al., 2021] Aditya Mate, Andrew Perrault, and Milind Tambe. Risk-Aware Interventions in Public Health: Planning with Restless Multi-Armed Bandits. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS ’21, page 880–888, Richland, SC, 2021. International Foundation for Autonomous Agents and Multiagent Systems.
- [Mate et al., 2022] Aditya Mate, Lovish Madaan, Aparna Taneja, Neha Madhiwalla, Shresth Verma, Gargi Singh, Aparna Hegde, Pradeep Varakantham, and Milind Tambe. Field Study in Deploying Restless Multi-Armed Bandits: Assisting Non-profits in Improving Maternal and Child Health. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12017–12025, 2022.
- [Odhiambo, 2016] Oginga Jared Odhiambo. *Technical Efficiency of the Kenyan Judiciary: A Case of the Magistrates’ Courts*. Master’s thesis, University of Nairobi, 2016.
- [Puterman, 1994] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics. Wiley, 1994.
- [Rangi and Franceschetti, 2018] Anshuka Rangi and Massimo Franceschetti. Multi-Armed Bandit Algorithms for Crowdsourcing Systems with Online Estimation of Workers’ Ability. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS ’18, page 1345–1352, Richland, SC, 2018. International Foundation for Autonomous Agents and Multiagent Systems.
- [Ron et al., 2021] Tom Ron, Omer Ben-Porat, and Uri Shalit. Corporate Social Responsibility via Multi-Armed Bandits. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’21, page 26–40, New York, NY, USA, 2021. Association for Computing Machinery.
- [Slivkins, 2019] Aleksandrs Slivkins. Introduction to Multi-Armed Bandits. *Foundations and Trends in Machine Learning*, 12(1–2):1–286, November 2019.
- [Tran-Thanh et al., 2014] Long Tran-Thanh, Sebastian Stein, Alex Rogers, and Nicholas R Jennings. Efficient Crowdsourcing of Unknown Experts Using Bounded Multi-Armed Bandits. *Artificial Intelligence*, 214:89–111, 2014.
- [Verma et al., 2023] Shresth Verma, Aditya Mate, Kai Wang, Neha Madhiwalla, Aparna Hegde, Aparna Taneja, and Milind Tambe. Restless Multi-Armed Bandits for Maternal and Child Health: Results from Decision-Focused Learning. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS ’23, page 1312–1320, Richland, SC, 2023. International Foundation for Autonomous Agents and Multiagent Systems.
- [Wang et al., 2023] Kai Wang, Shresth Verma, Aditya Mate, Sanket Shah, Aparna Taneja, Neha Madhiwalla, Aparna Hegde, and Milind Tambe. Scalable Decision-Focused Learning in Restless Multi-Armed Bandits with Application to Maternal and Child Health. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12138–12146, 2023.
- [Xu et al., 2020] Huanle Xu, Yang Liu, Wing Cheong Lau, Tiantong Zeng, Jun Guo, and Alex X. Liu. Online resource allocation with machine variability: A bandit perspective.

IEEE/ACM Transactions on Networking, 28(5):2243–2256, 2020.

[Ye and Tse, 1989] Yinyu Ye and Edison Tse. An extension of Karmarkar projective algorithm for convex quadratic programming. *Math. Program.*, 44(2):157–179, July 1989.

6 Appendix

6.1 Solvability of the QP

We show that the quadratic program in Section 3.3 is a convex quadratic program and that, under the natural assumption that every real case has at least one eligible mediator, it admits an optimal solution.

Let $z = (x, \xi) \in \mathbb{R}^{|E|+|U|}$, where x denotes assignment variables and ξ denotes mediator-level overload slack variables. Let $c_e = \mu_u + p_v$ for $e = (u, v)$. The QP can be written as

$$\max_{x, \xi} c^\top x - \lambda \sum_{u \in U} \xi_u^2,$$

subject to constraints (C1)–(C5). Equivalently, multiplying the objective by -1 , this is the minimization problem

$$\min_{x, \xi} -c^\top x + \lambda \sum_{u \in U} \xi_u^2.$$

The Hessian of this minimization objective with respect to $z = (x, \xi)$ is block diagonal:

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & 2\lambda I_{|U|} \end{bmatrix}.$$

For any conformable vector $w = (w_x, w_\xi)$,

$$w^\top Q w = 2\lambda \|w_\xi\|_2^2 \geq 0$$

whenever $\lambda \geq 0$. Thus $Q \succeq 0$, and the minimization form has a convex quadratic objective. Since constraints (C1)–(C5) are all linear equalities or inequalities, the QP is a convex quadratic program.

It remains to establish feasibility and existence of an optimum. Define the assignment polytope

$$P = \left\{ x : \begin{array}{ll} \sum_{e \in E(v)} x_e = 1 & \forall v \in V_r, \\ \sum_{e \in E(v)} x_e \leq 1 & \forall v \in V_{sh}, \\ 0 \leq x_e \leq 1 & \forall e \in E \end{array} \right\}.$$

Assume that every real case $v \in V_r$ has at least one eligible mediator, i.e., $E(v) \neq \emptyset$. Then $P \neq \emptyset$: for each real case v , choose one incident edge $e_v \in E(v)$ and set $x_{e_v} = 1$; set all other real-case incident variables and all shadow-case variables to zero. This satisfies (C1), (C2), and (C4). Since P is defined by finitely many linear constraints and box constraints, it is compact.

For any fixed $x \in P$, define the smallest nonnegative slack satisfying the capacity constraints as

$$\xi_u^*(x) = \left[\max_{t \in [T]} \left\{ L(u) + \sum_{e=(u,v) \in E(u)} x_e \mathbb{1}[t_a(v) \leq t] - C(u) \right\} \right]_+.$$

where $[a]_+ = \max\{a, 0\}$. Then $(x, \xi^*(x))$ satisfies (C3) and (C5). Hence the full feasible set is nonempty.

Moreover, for any fixed $x \in P$, choosing any $\xi_u > \xi_u^*(x)$ cannot improve the objective when $\lambda > 0$, because the objective subtracts $\lambda \xi_u^2$. When $\lambda = 0$, the objective is independent of ξ , so $\xi^*(x)$ may still be chosen without loss of optimality.

Therefore, the QP is equivalent, for purposes of attaining an optimum, to maximizing the continuous function

$$\phi(x) = c^\top x - \lambda \sum_{u \in U} (\xi_u^*(x))^2$$

over the compact set P .

The function $\xi_u^*(x)$ is continuous because it is the positive part of the maximum of finitely many affine functions of x . Therefore $\phi(x)$ is continuous. By the Weierstrass extreme value theorem, ϕ attains a maximum on P . Let x^* be a maximizer and set $\xi^* = \xi^*(x^*)$. Then (x^*, ξ^*) is feasible for the original QP and attains the optimal objective value. Thus the QP admits at least one optimal solution.

6.2 Runtime Analysis of SMarT:

In SMarT, each arriving case is solved independently, so the runtime of a single QP depends on the number of sampled shadow cases and the number of eligible mediators for that arriving case, rather than on the total duration of the simulation. Since the QP is convex with positive semi definite Q , there exist algorithms that can find solutions in polynomial time [Ye and Tse, 1989], [Kapoor and Vaidya, 1986]. For example, standard interior-point analysis for convex QPs, with $n = |E| + |U|$ variables and $m = O(|V_r| + |V_{sh}| + |U|T + |E| + |U|)$ constraints, the worst-case per-iteration complexity is $O((n+m)^3)$. In practice, we solve the QP using Gurobi, which exploits sparsity and optimized barrier-based routines for continuous convex QPs, and we observe substantially faster empirical runtime than this generic worst-case bound. The total runtime of SMarT scales this polynomial complexity of solving each QP linearly in the number of cases assigned in total.

In our full-scale experiments presented in Tables 4 and 5, the median instance involved 52 total cases and 701 eligible mediators, yielding a median QP size of 3,632 variables and 14,326 constraints. The median solution time was 0.25 seconds, with the 95th and 99th percentiles at 1.23 and 2.68 seconds, respectively.

6.3 Linear Approximation of the QP

The QP in section 3.3 can be approximated using an LP in a small local region in the following manner:

LP-Approximation

$$\max_{x, \xi} \sum_{e \in E} x_e (\mu_u + p_v) - \lambda \sum_{u \in U} \xi_u \times L(u)$$

$$\text{s.t.} \quad \sum_{e \in E(v)} x_e = 1 \quad \forall v \in V_r \quad (\text{C6})$$

$$\sum_{e \in E(v)} x_e \leq 1 \quad \forall v \in V_{sh} \quad (\text{C7})$$

$$\begin{aligned} L(u) + \sum_{e \in E(u)} x_e \mathbb{1}[t_a(v) \leq t] \\ \leq C(u) + \xi_u \quad \forall u \in U, \forall t \in [T] \end{aligned} \quad (\text{C8})$$

$$0 \leq x_e \leq 1 \quad \forall e \in E \quad (\text{C9})$$

$$0 \leq \xi_u \leq \max(L(u) - C(u) + 1, 0) \quad \forall u \in U \quad (\text{C10})$$

Algorithm	λ	Known VA (SMaRT(GS))		Learning + Mean VA (SMaRT(E))		Learning + Sampled VA (SMaRT(S))	
		Agreement Rate	OCDM	Agreement Rate	OCDM	Agreement Rate	OCDM
Least Load	–	0.406	25.13	–	–	–	–
Greedy*	–	0.439	41.71	0.425	36.31	–	–
TS*	–	–	–	–	–	0.404	34.17
SMaRT	0.01	0.444	62.57	0.424	48.33	0.425	46.74
	0.05	0.438	40.54	0.418	34.92	0.423	36.05
	0.10	0.437	37.51	0.417	32.07	0.421	32.11
	0.50	0.436	36.83	0.417	31.00	0.420	30.90
Upper Bound	–	0.460	132.15	–	–	–	–

Table 6: Scenario 1 results with LP-Approximation.

Algorithm	λ	Known VA (SMaRT(GS))		Learning + Mean VA (SMaRT(E))		Learning + Sampled VA (SMaRT(S))	
		Agreement Rate	OCDM	Agreement Rate	OCDM	Agreement Rate	OCDM
Least Load	–	0.380	4.13	–	–	–	–
Greedy*	–	0.406	5.18	0.394	5.06	–	–
TS*	–	–	–	–	–	0.392	4.82
SMaRT	0.01	0.447	70.09	0.413	47.16	0.407	36.99
	0.05	0.429	27.58	0.403	18.89	0.402	21.03
	0.10	0.416	11.49	0.398	10.38	0.395	11.29
	0.50	0.413	5.53	0.397	5.79	0.394	5.42
Upper Bound	–	0.460	132.15	–	–	–	–

Table 7: Scenario 2 results with LP-Approximation.

Note the replacement of the quadratic term $\lambda \sum_{u \in U} \xi_u^2$ in the objective with $\lambda \sum_{u \in U} \xi_u \times L(u)$. Here $L(u)$ is the current case-load of mediator u at the time a new case arrives. For an optimal QP solution, $\xi_u = \max(0, L'(u) - C(u))$, where $L'(u)$ is the mediator case load in that optimal solution. It follows that $\sum_{u \in U} \xi_u^2 = \sum_{u \in U} \xi_u \times (L'(u) - C(u)) \approx \sum_{u \in U} \xi_u \times L'(u)$. If $L(u)$ is the caseload during the LP formulation, and if we enforce $\xi_u \leq L(u) - C(u) + 1$ in the LP, then we could approximate the penalty term as $\lambda \sum_{u \in U} L(u)$. This approximation is good only for values of ξ_u that are near $L(u)$. This could be useful when solving a QP is deemed computationally expensive.

We repeat all of the experiments presented in the paper with **LP-Approximation** and provide the results in Tables 6, 7, 8 and 9. Notice that the results are very similar to the same simulations with the QP presented in Tables 2, 3, 4 and 5. Much of the analysis and discussions regarding the results remain valid. However, when LP-approximation is used, we notice slight lower resolution rates at lower λ s and lower OCDM at higher λ s. Also, please note that ‘Least Load’, ‘Greedy*’, ‘TS*’, and ‘Upper Bound’ rows are verbatim copied from Tables 2, 3, 4 and 5 to 6, 7, 8 and 9 to help the reader compare SMaRT algorithms against baselines. Only the SMaRT(GS), SMaRT(E) and SMaRT(S) algorithms are impacted by using LP instead of the QP, and the corresponding rows are re-simulated and updated.

6.4 Estimating uncertainty of mediator Value-Added

In this section, we clarify the mathematics underlying the uncertainty estimation for mediator VAs. We denote the outcome of case i assigned to mediator j using the random variable Y_{ij} . $Y_{ij} = 1$ when the case is resolved successfully, and 0 otherwise. Also, let us recall the probability of successful resolution in this case would be given by $p_i + \mu_j$. In other words:

Algorithm	λ	Agreement Rate	OCDM	Caseload Gini Index
Least Load	–	0.467	0.10	26.1%
Greedy*	–	0.569	0.32	74.6%
SMaRT(GS)	0.01	0.615	49.37	87.1%
	0.02	0.600	24.45	82.1%
	0.03	0.589	12.10	78.7%
	0.04	0.581	6.05	76.6%
	0.05	0.578	3.16	75.4%
	0.10	0.571	0.39	74.0%
	0.50	0.570	0.31	73.9%
Upper Bound	–	0.659	122.01	96.5%

Table 8: Simulations with true VA information available to the allocation algorithms on data collected from the judiciary with LP-Approximation.

Algorithm	λ	Initial belief set to $\mathcal{N}(0, \sigma_\mu^2)$			Initial belief calculated from data		
		Agreement Rate	OCDM	Caseload Gini Index	Agreement Rate	OCDM	Caseload Gini Index
Greedy*	–	0.489	0.44	68.4%	0.473	0.38	63.1%
SMaRT(E)	0.01	0.514	70.00	87.0%	0.444	30.29	71.5%
	0.05	0.496	10.32	72.0%	0.465	0.70	61.2%
	0.10	0.490	1.22	68.3%	0.471	0.38	61.5%
	0.50	0.490	0.42	68.0%	0.471	0.37	61.9%
TS*	–	0.484	0.34	53.7%	0.489	0.31	50.8%
SMaRT(S)	0.01	0.496	21.69	57.5%	0.492	6.14	49.4%
	0.05	0.487	3.88	50.8%	0.489	0.82	47.5%
	0.10	0.485	0.74	49.7%	0.490	0.28	47.7%
	0.50	0.483	0.28	49.4%	0.491	0.25	48.0%

Table 9: Simulations on data collected from the judiciary where VAs need to be learned during allocation with LP-Approximation.

$$Pr[Y_{ij} | \mu_i] = \begin{cases} \mu_j + p_i, & \text{If } Y_{ij} = 1, \\ 1 - (\mu_j + p_i), & \text{If } Y_{ij} = 0 \end{cases}$$

$$\begin{aligned} f(\mu_j | Y_{ij} = 1) &= \frac{\Pr(Y_{ij} = 1 | \mu_j) f(\mu_j)}{\int_{-\infty}^{\infty} \Pr(Y_{ij} = 1 | \mu_j) f(\mu_j) d\mu_j} \\ &= \frac{(\mu_j + p_i) \frac{1}{\sqrt{2\pi}\sigma_\mu} \exp\left(-\frac{\mu_j^2}{2\sigma_\mu^2}\right)}{\int_{-\infty}^{\infty} (\mu_j + p_i) \frac{1}{\sqrt{2\pi}\sigma_\mu} \exp\left(-\frac{\mu_j^2}{2\sigma_\mu^2}\right) d\mu_j} \\ &\propto (\mu_j + p_i) \exp\left(-\frac{\mu_j^2}{2\sigma_\mu^2}\right). \end{aligned}$$

Computing the mediator VAs explicitly using methods outlined in section 3.2 provides us a global mediator level standard deviation σ_μ . We assume a prior $\mathcal{N}(0, \sigma_\mu)$ on the mediator VAs. With that assumption, we get:

$$f(\mu_j) = \frac{1}{\sqrt{2\pi}\sigma_\mu} e^{\left(-\frac{\mu_j^2}{2\sigma_\mu^2}\right)}$$

Now, if we make this assignment and observe the outcome Y_{ij} , then a Bayesian posterior can be used to compute:

$$f(\mu_i | Y_{ij}) = \frac{\Pr[Y_{ij} | \mu_j] f(\mu_j)}{\int_{-\infty}^{\infty} \Pr[Y_{ij} | \mu_j] f(\mu_j) d\mu_j} \quad (1)$$

We now simplify equation 1 on a case by case basis.

When $Y_{ij} = 1$:

Similarly, when $Y_{ij} = 0$, we can simplify to:

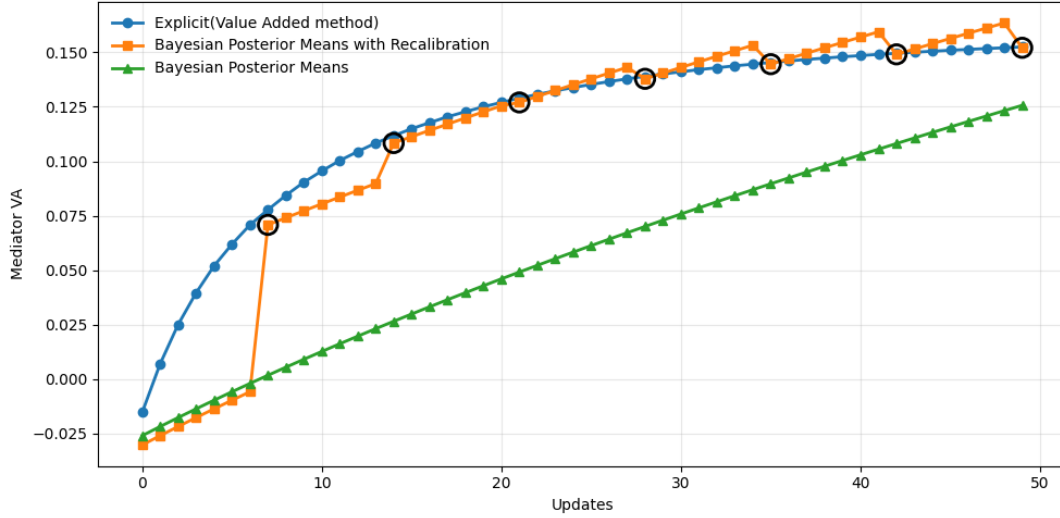


Figure 5: Trajectory of a single mediator's value-added (VA) estimate over 50 sequential successful case-outcome updates under three methods. The Bayesian posterior mean without recalibration (green) drifts away from the explicit VA estimate (blue), while periodic recalibration applied every seventh update (orange; circled points) limits this drift and keeps posterior estimates closely aligned with the explicit VA.

$$f(\mu_j | Y_{ij} = 0) \propto (1 - (\mu_j + p_i))e^{(-\frac{\mu_j^2}{2\sigma_\mu^2})}$$

Combining both, we can write:

$$f(\mu_j | Y_{ij}) \propto (\mu_j + p_i)^{Y_{ij}} (1 - (\mu_j + p_i))^{1-Y_{ij}} \exp(-\frac{\mu_j^2}{2\sigma_\mu^2}) \quad (2)$$

Now, for the next case i' being assigned to the same mediator j , we can use the posterior in Eq 2 as a prior and compute the posterior $f(\mu_j | Y_{ij}, Y_{i'j})$. However, we have found using $f(\mu_j | Y_{ij})$ directly as a prior makes the analysis complicated. As a remedy, we use gaussian moment matching to define the posterior $f(\mu_j | Y_{ij})$ to be $\mathcal{N}(\mu_j | Y_{ij}, (\sigma_j | Y_{ij})^2)$ where $\mu_j | Y_{ij} = E[f(\mu_j | Y_{ij})]$ and $(\sigma_j | Y_{ij})^2 = \text{Var}[f(\mu_j | Y_{ij})]$. Using this method, we use gaussian priors to compute posteriors as successive case outcomes become available.

Finally, the standard deviations associated with these posteriors are used as the uncertainty associated with mediator j 's VA.

6.5 Efficient VA updates and Re-calibration of the posteriors

Ideally during the simulation, mediator VAs should be calculated explicitly using the VA method outlined in section 3.2. This computation needs to be done every time an assigned case reaches conclusion. The explicit VA calculation method requires processing all case outcomes allocated to all mediator until that point. As a result, the explicit method is computationally prohibitive to repeat for every case allocation.

Computing the VA uncertainties using Bayesian posterior method can provide us a computationally inexpensive estimate of the mediator VAs in the form of posterior means. However,

Gaussian moment matching introduces inaccuracies in the posteriors which can add up to be significant. Therefore, these estimates are rather unreliable as a substitute for explicit VA computations.

We have found the hybrid approach of recalibrating the posterior means by resetting them to VAs computed explicitly to be a good trade off. In the simulation, we start off with explicit VA estimations. When cases are resolved, we perform the immediate VA updates using posterior means. And every 7 days, we explicitly re-compute mediator VAs using the Value Added method with updated case histories, and then recalibrate the Bayesian posteriors.

To illustrate the accuracy gains from recalibration, we track the estimated VA of a single mediator over 50 sequential case-outcome updates under each of the three methods. In this experiment, all 50 case outcomes are successful, and recalibration is performed every seventh update. Figure 5 shows the resulting trajectories. Without recalibration, the Bayesian posterior mean (green) can drift substantially away from the explicit VA estimate (blue). By contrast, periodic recalibration (orange) constrains this drift, keeping the posterior estimates within a tolerable range of the explicit VAs.