

DARTS

Dynamic Adaptive Rerandomization via Thompson Sampling

Targeting Prognostic Covariates in Budget-Constrained Sequential Experiments

Kat Husar and Alex Volfovsky

Experiments are everywhere

ONLINE PRODUCTS

Does a new ad headline work?

At Bing, a small change to how ad headlines were displayed became a controlled test with real users.

HEALTH POLICY

What does Medicaid change?

Oregon used a lottery to assign the chance to apply for Medicaid, testing effects on care, health, and financial strain.

DEVELOPMENT

Do cash transfers change lives?

PROGRESA randomized when localities started the program, testing effects on schooling, health, and nutrition.

Every experiment hides a few expensive decisions

which data are worth collecting

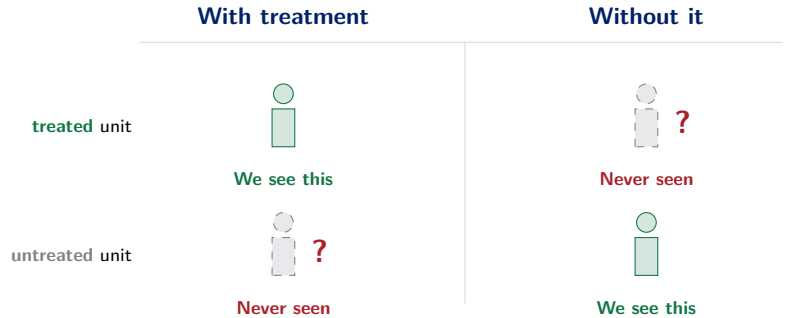
which people or places affect one another

how are resources shared between treatment groups

We often make assumptions **before the experiment begins**, and then treat our guesses as facts for the rest of the study.

What if the experiment could learn these while it runs?

What an experiment is for

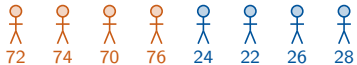


We want the average difference between the two columns, but half of every row is missing.

Randomization is the standard answer: assign treatment by chance, and the difference in the two averages estimates the effect.

Comparable on average — but you run *one* experiment

Eight patients enrol. Older patients recover more slowly.



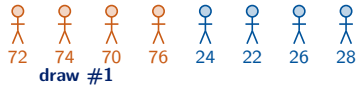
four in their 70s

four in their 20s

We flip fair coins. What can come out?

Comparable on average — but you run *one* experiment

Eight patients enrol. Older patients recover more slowly.

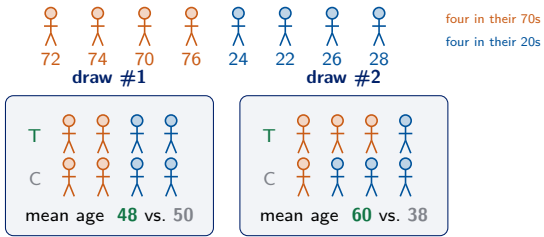


four in their 70s
four in their 20s

We flip fair coins. What can come out?

Comparable on average — but you run *one* experiment

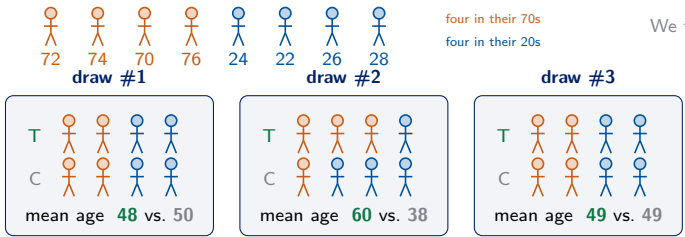
Eight patients enrol. Older patients recover more slowly.



We flip fair coins. What can come out?

Comparable on average — but you run *one* experiment

Eight patients enrol. Older patients recover more slowly.

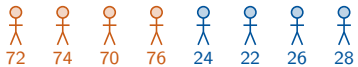


We flip fair coins. What can come out?

across *all* possible draws the two groups have the same average age — that is what randomization buys

Comparable on average — but you run *one* experiment

Eight patients enrol. Older patients recover more slowly.

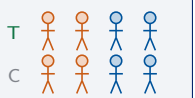


four in their 70s

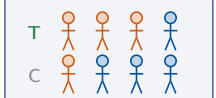
four in their 20s

We flip fair coins. What can come out?

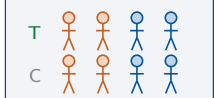
draw #1



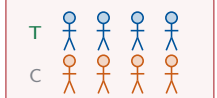
draw #2



draw #3



draw #4 — also legal!

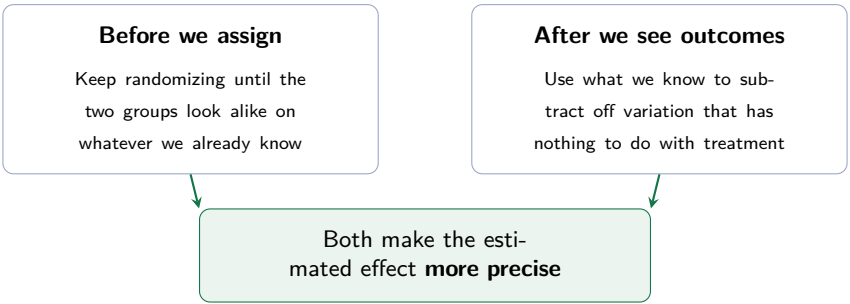


across *all* possible draws the two groups have the same average age — that is what randomization buys

This draw is just as likely and now the treatment effect is tangled up with **age**

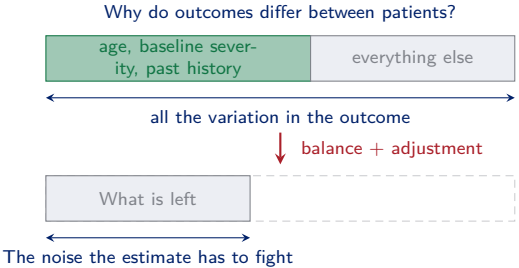
The guarantee is about the **procedure** on average, not about **your** single experiment.

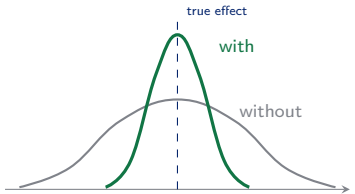
We get to choose how to use what we know



Same units, same estimand — a narrower answer.

Precision comes from knowing something in advance





How much the estimate moves from one experiment to the next

In the eight-patient example that slice was *age*. The more of the outcome our advance information explains, the more precision we get

But information is limited

A survey can only be so long

Every extra question costs you responses. You have to decide which questions make the cut.

Some data has to be bought

Vendors license records by the row. More variables per person means fewer people.

Clinical measurement is expensive

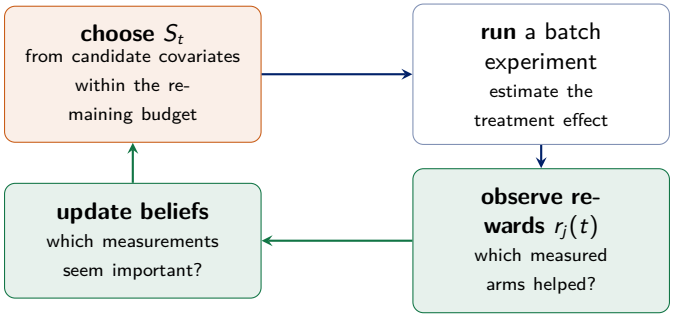
Assays, imaging and specialist reads are priced per patient, and they take time.

Sometimes the limit is physical

There is only so much blood you can draw, and only so much of a participant's attention.

A large pool of things we *could* measure, a hard limit on what we can *afford*.

So let the experiment find out



Choose what to measure, learn what helps, and update our beliefs about what matters for the next batch.

Formal setup

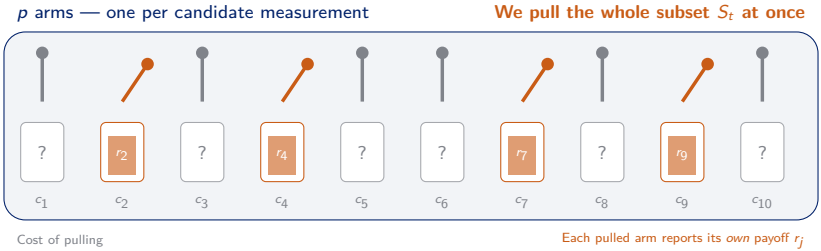
p	c_j	B	T	θ_j
things we could measure	cost of measuring j on a batch	fixed budget for the whole study	predetermined number of batches	how useful j is — **unknown**

In batch t we buy a subset S_t of the p candidates. Write $\theta_j \in [0, 1]$ for how useful covariate j tends to look — a stand-in for the share of outcome variation it explains, which is what actually buys precision. If we knew every θ_j, choosing what to buy would be an ordinary knapsack:

$$\max_{S_1,\ldots,S_T} \sum_{t=1}^{T} \sum_{j \in S_t} \theta_j \qquad \text{subject to} \qquad \sum_{t=1}^{T} \sum_{j \in S_t} c_j \;\leq\; B.$$

We estimate usefulness from earlier batches while the study spends its budget.

A combinatorial bandit with one budget



The action is a set. We choose S_t.

Feedback is per arm. Observe rewards separately for each arm.

Total budget. Spend it too early and it will be gone.

The action is a set, feedback arrives arm by arm, and one budget limits all the pulls.

Learning which measurements matter

1. Start with a prior

We do not observe θ_j, the true usefulness of arm j.

Use covariate predictiveness as a proxy.

$\mathrm{Beta}(1,1)$

flat prior on predictiveness

2. Measure a proxy

Fit LASSO on the *untreated* units.

A rescaled coefficient becomes the fractional reward:

$$r_j(t) = \frac{|\hat{\lambda}_j|}{\max_{k \in S_t} |\hat{\lambda}_k|}$$

3. Update the belief

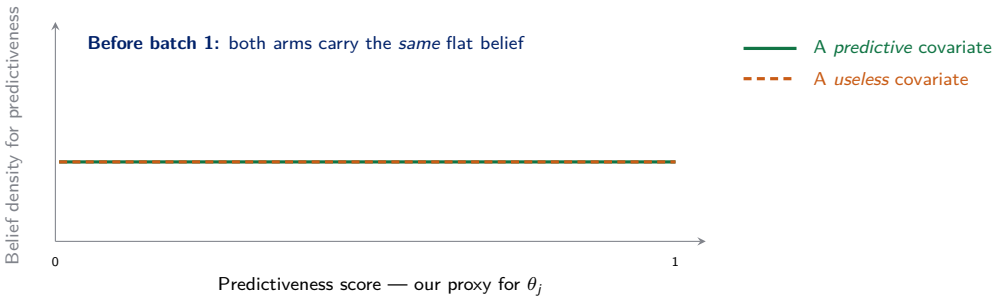
For a measured candidate $j \in S_t$:

$$\alpha_{j,t+1} = \alpha_{j,t} + r_j(t)$$
$$\beta_{j,t+1} = \beta_{j,t} + 1 - r_j(t)$$

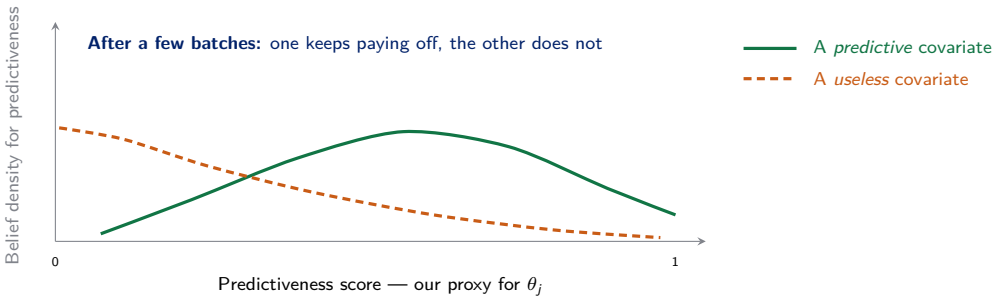
If $j \notin S_t$, nothing updates.

The reward is not θ_j itself: it is evidence about which covariates are worth measuring next.

The belief: Thompson sampling

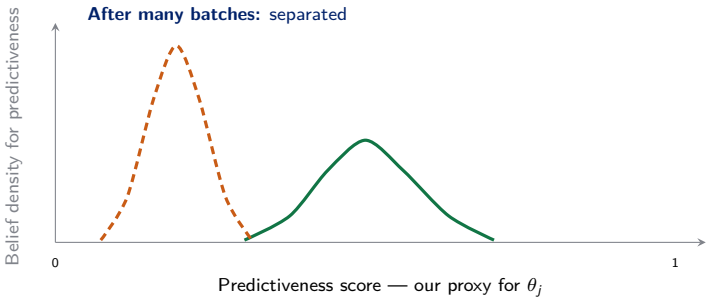


The belief: Thompson sampling



Only the arms we actually bought move — the rest keep the belief they had.

The belief: Thompson sampling



Only the arms we actually bought move — the rest keep the belief they had.

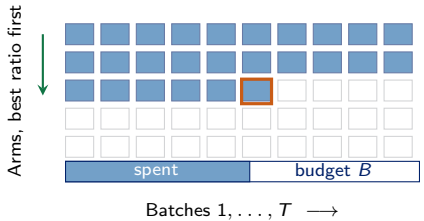
We **draw** from the posterior — this is **Thompson sampling**.

A wide belief sometimes draws high, so an uncertain arm still gets tried: exploration-exploitation trade off

How to choose S_t

Draw $\tilde{\theta}_j$ from the Beta posterior.

Buy in decreasing order of $\dfrac{\tilde{\theta}_j}{\tilde{c}_j^{\,\mathrm{eff}}}$ until the budget stops you



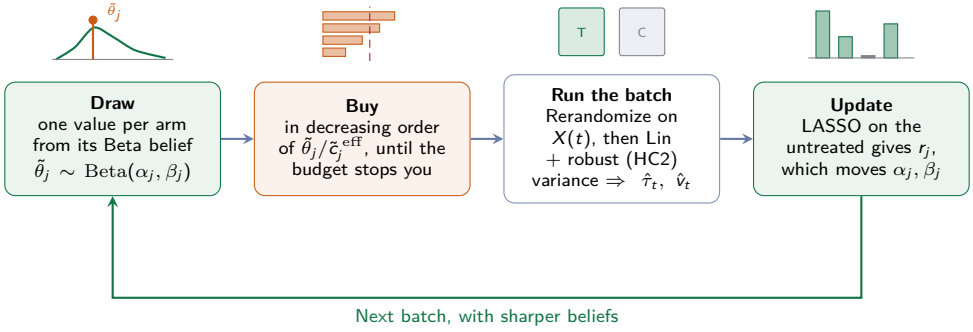
Blue = we bought that arm in that batch; orange = where the budget ran out.

The operational rule

Rank each measurement by expected usefulness divided by its current effective cost.

- A measurement that looks useful but is expensive becomes harder to buy early.
- The price rises when spending now would leave too little budget for later batches.

DARTS Algorithm



At the end we pool the batches by inverse-variance weighting: $\hat{\mu}_T = \sum_t \hat{v}_t^{-1} \hat{\tau}_t \big/ \sum_t \hat{v}_t^{-1}$ with $\hat{\sigma}_T^2 = \left(\sum_t \hat{v}_t^{-1}\right)^{-1}$ — better-balanced batches count for more.

What the theory guarantees

Theorem 1: unbiased within a batch

If the measurement choice uses only earlier batches, the randomized treatment contrast is exactly unbiased.

Theorem 2: asymptotic validity

The pooled estimator is asymptotically normal, with a conservative variance estimate. The resulting interval has at least nominal coverage.

Design variance does not increase

Rerandomization can only reduce the design variance relative to simple random assignment, under the stated conditions.

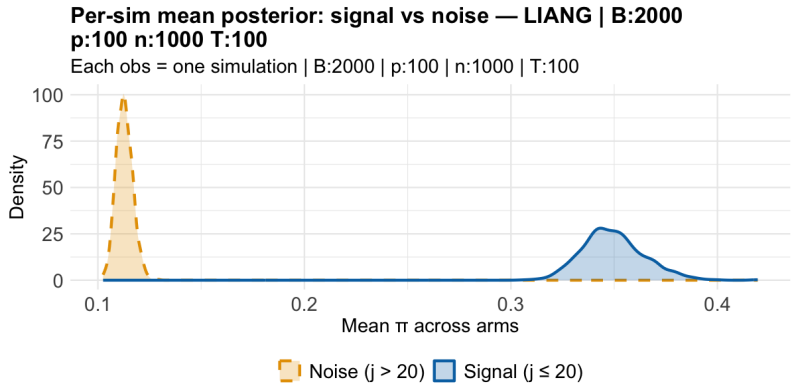
Theorem 3: sublinear regret

Posterior sampling learns which measurements matter. Its cumulative Bayes regret grows sublinearly in the total number of pulls.

The safeguards: choose measurements from the past, randomize treatment within the batch, and fix the horizon in advance.

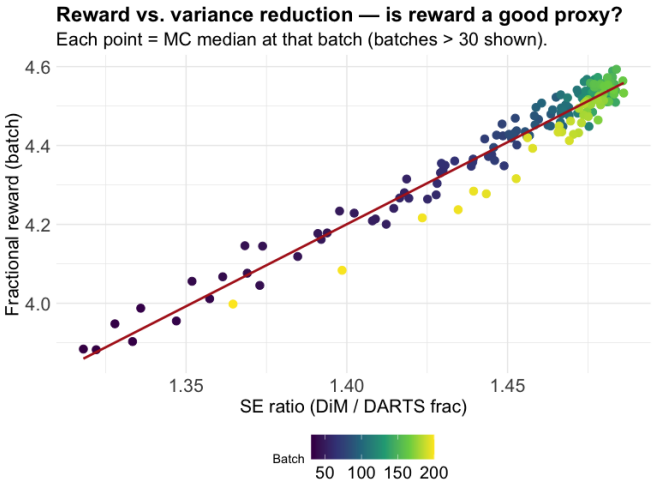
Does it find the useful measurements?

100 candidate covariates, 20 truly prognostic; correlated covariates, nonlinear outcome; 100 batches of 1000 units; equal costs, $B = 2000$; 1000 Monte Carlo replications.



Each arm's final belief, $\pi_j = \alpha_j/(\alpha_j + \beta_j)$ — how useful DARTS decided covariate j was.

Is the proxy tracking what we care about?



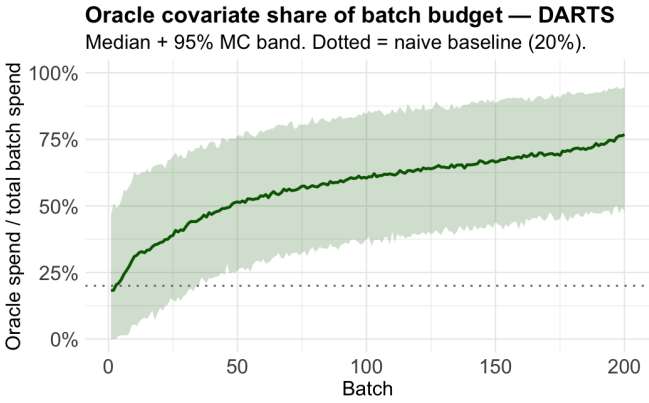
The reward is a proxy: LASSO coefficients, not variance reduction.

Here it is plotted against the precision gain it stands in for, over plain difference in means (DiM): $\hat{\sigma}_{\mathrm{DiM},t}/\hat{\sigma}_{\mathrm{DARTS},t}$.

$r \approx 0.98$ **across batches**

Darker points are earlier batches.

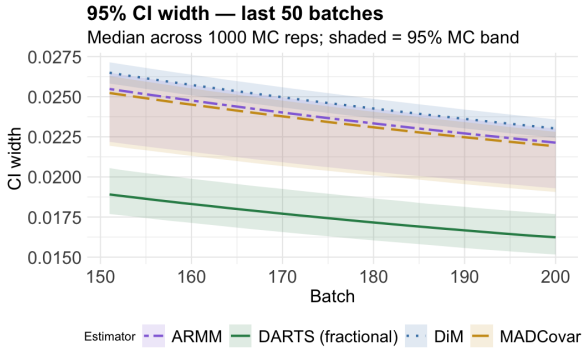
The budget migrates to the right measurements



Share of each batch's budget spent on the twenty covariates that actually matter.

Starts at the **20%** you get by buying at random; on average it climbs past 75%.

Does it give tighter intervals?



With a random twelve, careful design buys almost nothing. The gain comes from **learning**.

Limitations and possible extensions

Correlated arms

Measurements can overlap or substitute for one another. Learning each arm separately can waste pulls.

Nonlinear outcomes

A LASSO-style reward can miss interactions, thresholds, or a curved prognostic relationship.

Prices can move

Vendor prices, assay capacity, or measurement time may change after the study starts.

The paper's simulations include correlated arms and nonlinear outcomes. The current budget model treats costs as fixed, so changing prices are a natural robustness extension.

Future experiments need to coordinate

Experiments overlap in time

An earlier experiment can also change the environment that later experiments are trying to measure.

Resources are shared

Shipping capacity, support staff, inventory, or measurement budgets can be jointly scarce within and between experiments.

The population is not reset

Units used in one study may be unavailable for the next. The next sample is then selected by history, not drawn from a clean pool.

The design must adapt

Power calculations and assignment rules need to account for which units and resources earlier studies have already consumed.

A lot of interesting problems in coordinating **several experiments** that compete for the same units and capacity.

Thank you!

Appendix: Algorithm 1

Initialize $\alpha_j \leftarrow 1$, $\beta_j \leftarrow 1$, $\nu_j \leftarrow 1$ for $j \in [p]$; $\nu_{\text{budget}} \leftarrow 1$; $B_0 = \min(B, T)$; $\tilde{c}_j = c_j B_0 / B$; $\varepsilon = \sqrt{\log(p+1)/B_0}$.

For $t = 1, \dots, T$:

A. *Acquisition.* Sample $\tilde{\theta}_j \sim \mathrm{Beta}(\alpha_j, \beta_j)$; form $\tilde{c}_j^{\,\text{eff}} = \tilde{c}_j \nu_{\text{budget}} + \nu_j$;

$$S_t = \textsc{Greedy}\Big(\{\tilde{\theta}_j / \tilde{c}_j^{\,\text{eff}}\}_{j \in [p]},\ B_{t-1}\Big).$$

Then for $j \in S_t$: $\nu_j \leftarrow \nu_j(1+\varepsilon)$, $\nu_{\text{budget}} \leftarrow \nu_{\text{budget}}(1+\varepsilon)^{\tilde{c}_j}$, and $B_t = B_{t-1} - \sum_{j \in S_t} \tilde{c}_j$.

B. *Design.* Measure $X(t) \in \mathbb{R}^{n \times |S_t|}$; draw $\mathcal{K}$ candidate assignments; $z^\star = \arg\min_z d_M(z, X(t))$; set $Z(t)$ by a fair coin flip between $z^\star$ and $\mathbf{1} - z^\star$.

C. *Estimation.* $\hat{\tau}_t = \hat{\beta}_Z$ from $Y(t) \sim Z(t) \times X(t)$ with HC2 variance $\hat{v}_t$; then $\hat{\sigma}_t^2 = (\hat{\sigma}_{t-1}^{-2} + \hat{v}_t^{-1})^{-1}$ and $\hat{\mu}_t = \hat{\sigma}_t^2 \big(\hat{\mu}_{t-1}/\hat{\sigma}_{t-1}^2 + \hat{\tau}_t / \hat{v}_t\big)$.

D. *Learning.* LASSO on the controls of batch t; $r_j = |\hat{\lambda}_j| / \max_k |\hat{\lambda}_k|$; $\alpha_j \mathrel{+}= r_j$, $\beta_j \mathrel{+}= 1 - r_j$ for $j \in S_t$.

Return $\hat{\mu}_T$, $\hat{\sigma}_T$, and $\pi_j = \alpha_j / (\alpha_j + \beta_j)$.

Appendix: proof sketch of Theorem 1

Let $\mathcal{G}_t = \mathcal{F}_{t-1} \vee \sigma(X(t))$. Write $Y_i(1,t) = Y_i(0,t) + \tau_i(t)$ and decompose

$$\hat{\tau}_t^{\mathrm{DIM}} - \tau = \underbrace{\tfrac{1}{N_1}\sum_{Z_i=1} Y_i(0,t) - \tfrac{1}{N_0}\sum_{Z_i=0} Y_i(0,t)}_{\text{(I)}} + \underbrace{\tfrac{1}{N_1}\sum_{Z_i=1} \tau_i(t) - \tau}_{\text{(II)}}.$$

- Conditional on $\mathcal{G}_t$ the selected covariates *and* all potential outcomes are fixed — this is exactly where $\mathcal{F}_{t-1}$-measurability of S_t is used.
- Mirror symmetry gives $P_t(Z = z \mid \mathcal{G}_t) = P_t(Z = \mathbf{1} - z \mid \mathcal{G}_t)$, so treated and control labels are exchangeable and $\mathbb{E}[(\mathrm{I}) \mid \mathcal{G}_t] = 0$.
- The same symmetry gives $\mathbb{E}[(\mathrm{II}) \mid \mathcal{G}_t] = \bar{\tau}_t - \tau$; i.i.d. sampling of the batch, independent of $\mathcal{F}_{t-1}$, gives $\mathbb{E}[\bar{\tau}_t \mid \mathcal{F}_{t-1}] = \tau$.
- Iterated expectation closes it. □

The argument is *batch-local*: nothing requires S_t to be constant across batches. For the Lin-adjusted estimator the same decomposition leaves an $O(n^{-1})$ term, negligible relative to the standard error when $T = o(n)$.

Appendix: regularity conditions for Theorem 2

With $\sigma_t^2 = \mathrm{Var}(\hat{\tau}_t \mid \mathcal{G}_t)$, $\sigma_{t,\circ}^2$ the design-agnostic variance under simple random assignment, $\bar{w}_t = \sigma_{t,\circ}^{-2}$, $w_t = \hat{v}_t^{-1}$ and $e_t = \hat{\tau}_t - \tau$:

A1 $\max_{t \le T} |w_t / \bar{w}_t - 1| \xrightarrow{p} 0$ — relative weight consistency

A2 $\sum_t (w_t - \bar{w}_t) e_t = o_p\big(\sqrt{\sum_t \bar{w}_t^2 \sigma_t^2}\big)$ — estimated-weight remainder

A3 $\max_{t \le T} \bar{w}_t^2 \sigma_t^2 \big/ \sum_s \bar{w}_s^2 \sigma_s^2 \to 0$ — uniform asymptotic negligibility

A4 $\sum_t \bar{w}_t^2 \sigma_t^2 \big/ \sum_t \bar{w}_t \xrightarrow{p} \ell \in (0, 1]$

A5 $\sigma_t^2 \le \sigma_{t,\circ}^2$ for all t — rerandomization never inflates variance

A6 $\sum_t \bar{w}_t \, \mathbb{E}[e_t \mid \mathcal{G}_t] = o_p\big(\sqrt{\sum_t \bar{w}_t^2 \sigma_t^2}\big)$ — negligible centering bias

A7 $\sup_t \mathbb{E}[|e_t|^{2+\delta} \mid \mathcal{G}_t] \le M$ for some $\delta > 0$

A3 and A7 together give the conditional Lindeberg condition for the martingale difference array $\{\bar{w}_t(e_t - \mathbb{E}[e_t \mid \mathcal{G}_t])\}$; the martingale CLT then yields normality, and A4–A5 deliver $c^2 = \ell^{-1} \ge 1$. A5 imports the finite-population result of Li & Ding into the superpopulation framework as a primitive.

Appendix: the LP, its dual, and the shadow prices

The LP relaxation replaces $a_j(t) \in \{0,1\}$ with $\zeta_j \geq 0$:

$$\max_{\zeta \geq 0} \sum_j \zeta_j \theta_j^\star \quad \text{s.t.} \quad \sum_j \zeta_j M_{ij} \leq B', \quad i = 1, \ldots, p+1, \qquad B' = \min(B, T), \ \tilde{c}_j = c_j B'/B.$$

- $\mathrm{OPT}_{\mathrm{LP}} \geq \mathrm{OPT}_{\mathrm{IP}}$, so benchmarking against the LP is conservative (Das et al., Lemma 3).
- Dual shadow prices $\nu \in \mathbb{R}^{p+1}$ — one per arm plus one for the global budget — updated by multiplicative weights, for $j \in S_t$:

$$\nu_j \leftarrow \nu_j(1+\varepsilon), \qquad \nu_{p+1} \leftarrow \nu_{p+1}(1+\varepsilon)^{\tilde{c}_j}, \qquad \varepsilon = \sqrt{\tfrac{\log(p+1)}{B_0}}.$$

- **Why row by row:** with one budget constraint and box constraints $\zeta_j \leq T$, the LP optimum saturates arms in decreasing $\theta_j/\tilde{c}_j$ order.
- **Why not a per-batch budget B/T?** If the good set costs more than B/T it can never be bought, so its rewards are never observed, however much total budget remains (Das et al., Lemma 1).
- **Planning gap.** With known means the policy loses $C_D L(\theta^\star)(\sqrt{\log(p+1)/B'} + p/B')$, negligible once $B' \gg p$.

Appendix: Bayes risk proof sketch (Theorem 3)

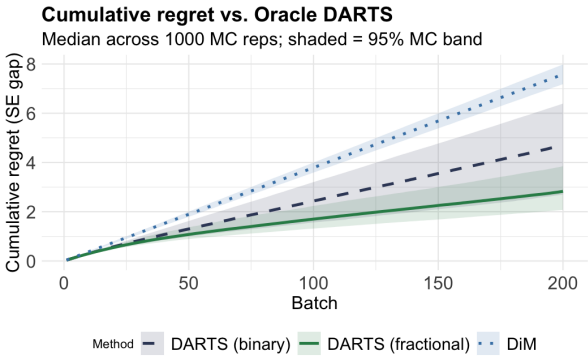
Russo–Van Roy: because the law of S_t matches the posterior law of $S_t^\star$, *any* $\mathcal{F}_{t-1}$-measurable UCB sequence U_t may be inserted for free,

$$\mathrm{BayesRisk} = \underbrace{\mathbb{E}\sum_t [U_t(S_t) - r_{\theta^\star}(S_t)]}_{\text{confidence width}} + \underbrace{\mathbb{E}\sum_t [r_{\theta^\star}(S_t^\star) - U_t(S_t^\star)]}_{\text{UCB failure}}.$$

- Semi-bandit feedback gives per-arm Hoeffding bounds with $U_{j,t} = \hat{\theta}_j(t) + \sqrt{\log(pT^2)/2n_j(t)}$; the failure term is $O(K^2/pT)$ and vanishes.
- Width term: $\sum_t \sum_{j \in S_t} 2\sqrt{\log(pT^2)/2n_j(t)}$. Swap the order of summation, apply the harmonic-sum bound, then Cauchy–Schwarz, to get $2\sqrt{2\log(pT^2)} \cdot \sqrt{p\,Q(T)}$.
- Two caps on total pulls $Q(T)$: at most K per round ($Q \le KT$), and at least $c_{\min}$ per pull ($Q \le B/c_{\min}$). Hence $Q(T) \le \min(KT, B/c_{\min})$. □

Fractional rewards (Corollary 2). The pseudo-posterior adds $K \sum_t \mathbb{E}[\varepsilon_t]$ with $\varepsilon_t = d_{\mathrm{TV}}(\tilde{\Pi}_t, \Pi_t \mid \mathcal{D}_{t-1})$ — small when rewards concentrate near 0 or 1. It depends on an unobservable TV distance, so it is indicative rather than operational. **Theorem 4** applies the gap-free lower bound of Kveton et al. on a uniform-cost instance, giving $\Omega(\sqrt{pQ})$.

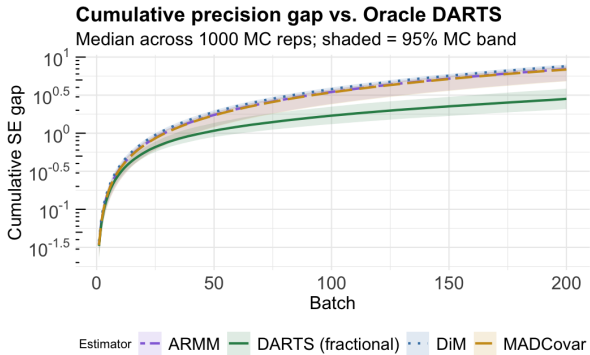
Appendix: fractional vs binary rewards



Cumulative regret relative to Oracle DARTS, 1000 replications.

Both reward schemes are sublinear. The fractional scheme converges faster, which is why it is the default even though the clean theory covers the binary case.

Appendix: cumulative precision gap vs. the baselines



Liang DGP, $n = 1000$, $T = 200$, $p = 100$, $B = 2000$, variable costs $c_j \sim \mathrm{Unif}(0, 2)$.

DARTS accumulates strictly less precision gap than ARMM and MADCovar throughout the horizon, with non-overlapping 95% bands.

Appendix: simulation design

Linear DGP. $X \sim \mathcal{N}(0, I_p)$, $Y_i(0) = X_i^\top \beta + 2X_{i,1}X_{i,5} + \varepsilon_i$, with $\beta_j \sim \mathcal{N}(2, 0.1)$ for $j \leq 10$, $\mathcal{N}(-2, 0.05)$ for $11 \leq j \leq 20$, $\mathcal{N}(0, 0.01)$ for $j > 20$ — noise covariates are weakly prognostic, not inert.

Liang DGP. $X_{ij} = (e_i + z_{ij})/2$ with $e_i, z_{ij} \overset{iid}{\sim} \mathcal{N}(0,1)$ — a shared random effect, so arms are correlated and the independence assumption is violated — scaled to $[0,1]$ and centred, then

$$Y_i(0) = \frac{10X_{i,2}}{1 + X_{i,1}^2} + 5\sin(X_{i,3} - X_{i,4}) + 2X_{i,5} + X_i^\top \beta + \varepsilon_i.$$

Both DGPs: $Y_i(1) = Y_i(0) + \tau$ with $\tau = 4$.

MC grid. $T = 100$, $n \in \{100, 1000\}$, $p \in \{100, 1000\}$, $B \in \{1000, 2000, 20000\}$, equal costs, 1000 replications.

Method comparison. $T = 200$, $n = 1000$, $p = 100$, $B = 2000$, $c_j \overset{iid}{\sim} \mathrm{Unif}(0, 2)$. ARMM and MADCovar each receive ≈ 12 randomly pre-selected covariates (about 3 of them truly prognostic); DARTS sees the full pool under the same budget.

Implementation. $\mathcal{K} = 5000$ candidate assignments; LASSO penalty by 5-fold CV (`lambda.min`) on controls; if LASSO selects nothing, rewards are 0 and the prior is unchanged; Moore–Penrose pseudoinverse for near-singular $\hat{\Sigma}(t)$. R 4.5.0, `glmnet`.

Appendix: rerandomization vs. rerandomization + adjustment

Setup	Method	Emp. SD	MSE	RE vs. DiM	% red.
LIANG, B=2000, p=100, n=100	DiM	0.0308	9×10^{-4}	1.00	0.0
	Rerand	0.0249	6×10^{-4}	1.54	34.9
	Rerand + adj	0.0218	5×10^{-4}	**1.99**	**49.8**
	Oracle	0.0153	2×10^{-4}	4.04	75.3
LIANG, B=2000, p=100, n=1000	DiM	0.0086	1×10^{-4}	1.00	0.0
	Rerand	0.0064	$< 10^{-4}$	1.79	44.0
	Rerand + adj	0.0057	$< 10^{-4}$	**2.32**	**56.9**
	Oracle	0.0045	$< 10^{-4}$	3.64	72.5
LINEAR, B=2000, p=100, n=1000	DiM	0.0581	0.0034	1.00	0.0
	Rerand	0.0518	0.0027	1.26	20.6
	Rerand + adj	0.0308	0.0009	**3.56**	**71.9**
	Oracle	0.0139	0.0002	17.59	94.3
LINEAR, B=2000, p=1000, n=100	DiM	0.1869	0.0349	**1.00**	0.0
	Rerand	0.1837	0.0337	1.04	3.5
	Rerand + adj	0.1898	0.0360	0.97	−3.1
	Oracle	0.0477	0.0023	15.38	93.5

Adjustment dominates except when p is large relative to n, where OLS runs out of degrees of freedom. 1000 replications.

Selected references: rerandomization and adjustment

Morgan, Kari Lock, and Donald B. Rubin. "Rerandomization to Improve Covariate Balance in Experiments," *The Annals of Statistics*, 2012, 40(2): 1263–1282.
https://arxiv.org/abs/1207.5625

Li, Xinran, Peng Ding, and Donald B. Rubin. "Asymptotic Theory of Rerandomization in Treatment-Control Experiments," *Proceedings of the National Academy of Sciences*, 2018, 115(37): 9157–9162.
https://arxiv.org/abs/1604.00698

Lin, Winston. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique," *The Annals of Applied Statistics*, 2013, 7(1): 295–318.
https://projecteuclid.org/journals/annals-of-applied-statistics/volume-7/issue-1/Agnostic-notes-on-regression-adjustments-to-experimental-data-Reexamining-Freedmans/10.1214/12-AOAS583.full

Zhou, X., Li, X., and Ding, P. "Sequential Rerandomization," *Biometrika*, 2018, 105(3): 745–752.
https://doi.org/10.1093/biomet/asy019

References for the opening examples

The Bing headline test. Ron Kohavi and Stefan Thomke, "The Surprising Power of Online Experiments," *Harvard Business Review*, September–October 2017. The 12% figure and the "months in a backlog" detail are both from this article.
https://hbr.org/2017/09/the-surprising-power-of-online-experiments

The Oregon Medicaid lottery. Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker and the Oregon Health Study Group, "The Oregon Health Insurance Experiment: Evidence from the First Year," NBER Working Paper 17190, July 2011; *Quarterly Journal of Economics* 127(3), 2012.
https://www.nber.org/papers/w17190 ▪ project page: https://www.nber.org/oregon

The PROGRESA rollout. The evaluation sample included 506 Mexican localities: 320 began the program immediately and 186 began about two years later. Baseline household data were collected before rollout. *Review of Economic Studies* 79(1), 2012; see also J-PAL's evaluation summary.
https://academic.oup.com/restud/article/79/1/37/1562110 ▪
https://www.povertyactionlab.org/evaluation/impact-progresa-health-mexico

The opening slide uses these three documented examples plus a general measurement-budget motivation.