

Experimental Methods: Lecture 1

Causal Inference and Potential Outcomes

Raymond Duch

April 30, 2020

Director CESS Nuffield/Santiago

Road Map to Lecture 1

- Potential outcomes and causal inference
- Average Treatment Effects (ATE)
- Alternative designs

Why Should We Do Experiments? I

- Lauren E. Young (2019) The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe. *APSR* 113(1):140–155.
- What is the effect of the emotion of fear on citizen dissent in autocracies?
- How would we demonstrate that the effect of fear on dissent is unconfounded by other variables?
 - Characteristics that induce emotions
 - New information about a threat

Why Should We Do Experiments? II

- Lab-in-the-field experiment in Harare, Masvingo and Manicaland provinces in Zimbabwe
- Random assignment to treatment or control: Affective emotional memory task (AEMT)
- Ethical implications

Why Should We Do Experiments? III

- Sample: 647 participants from six communities in Zimbabwe where the NGO Voice for Democracy (VfD) has a network of mobilizer and informants; and affected by state-sponsored violence
 - Treatment 1: Enumerator asks to describe a situation of fear *around politics and elections* [political fear]
 - Treatment 2: Enumerator asks to describe a situation of *general fears other than politics or elections* [general fear]
 - Control: Enumerator asks to describe a situation that makes them relaxed
- Outcome: Propensity to dissent: hypothetical (via index) and behavioral (via selection of political wristband)

Why Should We Do Experiments? IV

- Young (2019)

TABLE 3. The Fear Treatments Reduce Dissent

	Hypothetical		Behavioral	
	General fear (1)	Political fear (2)	General fear (3)	Political fear (4)
ATE ¹	-0.545	-0.773	-0.104	-0.189
SE ²	(0.077)	(0.080)	(0.050)	(0.053)
RI <i>p</i> - value ³	<0.001	<0.001	0.035	<0.001
<i>N</i>	484	486	329	326
Sample	All		Wristband ⁴	

¹The first row presents the estimated average treatment effects (ATEs) of the general and political fear treatments on the hypothetical measure of propensity to dissent in columns 1 and 2, and the behavioral measure in columns 3 and 4. ATEs are calculated based on assignment to treatment and weighted by inverse propensity scores by block.

²Robust standard errors (SEs) from linear regression analysis.

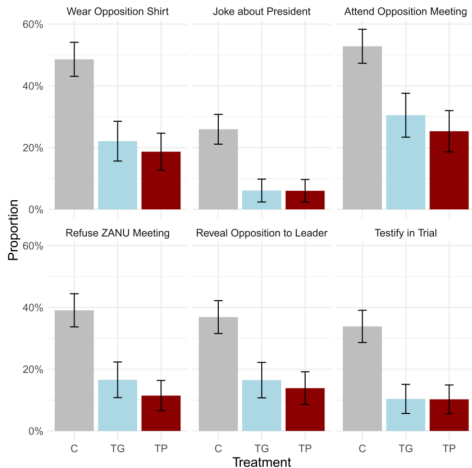
³The *p*-value is based on a two-tailed test using randomization inference.

⁴The estimate of the treatment effect on the wristband measure comes from the subset of the sample respondents who were offered a choice between two real wristbands. Results are similar for the full sample.

Why Should We Do Experiments? V

- Young (2019)

FIGURE 2. The Fear Treatments Cause Substantively Large Increases in the Proportion of Respondents Who Are Very Likely or Sure to Dissent During an Election Period



Defining Treatment

- The variable d_i indicates whether the i th subject is treated
- In the typical case of binary treatments, $d_i = 1$ means the i th subject receives the treatment
- $d_i = 0$ means the i th subject does not receive the treatment
- It is assumed that d_i is observed for every subject

Potential Outcomes

- Y_i : the potential outcome for subject i
- $Y_i(d_i)$: the outcome for subject i , written as a function of the treatment i received; it is generally the case that we observe only one of the potential outcomes for each i
- For the binary-valued treatment, there are two "potential outcomes":
 - $Y_i(1)$, the potential outcome for i conditional on i being treated
 - $Y_i(0)$, the potential outcome for i conditional on i not being treated

Potential Outcome Schedule

- "Hypothetical"
- Comprehensive list of potential outcomes for all subjects
- Rows of this schedule are indexed by i , and the columns are indexed by d
- Potential outcomes for the fifth subject may be found in adjacent columns of the fifth row

Treatments as random variables

- Note on notation: We distinguish between d_i , the treatment that a given subject actually receives, and D_i , the treatment that could be administered hypothetically.
- D_i is a random variable (the i th subject might be treated in one hypothetical study and not in another).
- $Y_i(0)|D_i = 1$: untreated potential outcome for subjects that would receive the treatment under a hypothetical random assignment.
- We use D_i when talking about the statistical properties of treatments.

Chattopadhyay & Duflo 2004

- Randomized policy experiment in India
- 1990s, one-third of village council heads reserved for women
- women.csv contains subset of data from West Bengal
- Gram Panchayat (GP) = level of government
- Analysis?
 - Was randomization implemented properly?
 - Conjecture: more drinking facilities under women
 - Conjecture: no effect on irrigation

Potential Outcomes Local Budget

	Budget share if village head is male	Budget share if village head is female	Treatment Effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	15	20	5

Potential Outcome Subgroup

- Sometimes useful to refer to potential outcomes for a subset of the subjects
- Expressions of the form $Y_i(d)|X = x$ denote potential outcomes when the condition $X = x$ holds
- For example, $Y_i(0)|d_i = 1$ refers to the untreated potential outcome for a subject who actually receives the treatment

Conditional potential outcomes

- $Y_i(0)|d_i = 1$: untreated potential outcome for subjects that receive treatment
- $Y_i(0)|d_i = 0$: untreated potential outcome for subjects that do not receive treatment
- $Y_i(1)|d_i = 1$: treated potential outcome for subjects that receive treatment
- $Y_i(1)|d_i = 0$: treated potential outcome for subjects that do not receive treatment

Individual Level Causal Effect

- For subject i , the effect of the treatment is conventionally defined as the difference between outcomes across the two potential outcomes:

$$\delta_i = Y_i(1) - Y_i(0)$$

- Alternatively:

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0))D_i$$

- Often referred to as the Rubin causal model; perhaps more appropriately, the Neyman-Holland-Rubin causal model
- **The Fundamental Problem of Causal Inference** only one of the two potential outcomes is realized, so that δ_i is typically non-operational

Realized Potential Outcomes

- Use lower-case letters for realization of the potential quantities (again, typically only one of the two potential outcomes is realized)
 1. $y_i(1)$, the outcome observed for i conditional on $d_i = 1$ (i is treated)
 2. $y_i(0)$, the outcome observed for i conditional on $d_i = 0$ (i is not treated)

The Fundamental Problem of Causal Inference

Table 1: Table 2.1, p35 Morgan and Winship, *Counterfactuals and Causal Inference*

Group	$Y_i(1)$	$Y_i(0)$
Treatment ($D_i = 1$)	Observable	Counterfactual
Treatment ($D_i = 0$)	Counterfactual	Observable

Observed Outcomes

- The connection between the observed outcome and the underlying potential outcome is given by the equation $Y_i = d_i Y_i(1) + (1 - d_i) Y_i(0)$
- This equation indicates that the $Y_i(1)$ are observed for subjects who are treated, and the $Y_i(0)$ are observed for subjects who are not treated
- For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, not both

Observed Outcomes Local Budget

	Budget share if village head is male	Budget share if village head is female
Village 1	?	15
Village 2	15	?
Village 3	20	?
Village 4	20	?
Village 5	10	?
Village 6	15	?
Village 7	?	30

Average Treatment Effect

- Average Treatment Effect:

$$E(\delta) = E[Y(1)] - E[Y(0)] = E[Y(1) - Y(0)]$$

- where the expectation is over a population, and so no subscript i
- This is operational, in that we can compute sample estimates of $E[Y(1)]$ and $E[Y(0)]$: e.g., the sample averages:

$$\hat{y}(1) = \frac{1}{n_1} \sum_{i:d_i=1} y_i(1) \text{ and } \frac{1}{n_0} \sum_{i:d_i=0} y_i(0)$$

- where n_1 and n_0 are the number of subjects in groups $d(1)$ and $d(0)$ respectively

Randomization Generates Unbiased Estimates of Average Treatment Effect

- Rubin (1974) calls this:

$$\begin{aligned}\hat{\delta} &= \hat{y}(1) - \hat{y}(0) \\ &= \hat{y}_d\end{aligned}$$

- Under certain circumstances, this is an unbiased estimate of the population average treatment effect δ
- Why? How?
- Nice, informal treatment in "Two Formal Benefits of Randomization"

Properties of Random Assignment

- Under equal probability random assignment, the conditional ATE among the treated is the same as the conditional ATE among the control group, which is therefore the same as the ATE
- The expected $Y_i(0)$ in the treatment group is the same as the expected $Y_i(0)$ in the control group
- When random assignment is not used (i.e., observational research), the unbiasedness of the difference-in-means estimator becomes a matter of conjecture

Potential Outcomes: Core Assumptions

- We assume that each subject has two potential outcomes $Y_i(1)$ if treated and $Y_i(0)$ if not treated
- Each potential outcome depends **solely** on whether the subject **itself** receives the treatment
- Potential outcomes respond only to the treatment and not to some other feature of the experiment - such as assignment or measurement

The Beauty of Randomization: Independence

- Treatment status is statistically independent of potential outcomes and background attributes \mathbf{X}

$$Y_i(0), Y_i(1), \mathbf{X} \perp\!\!\!\perp D_i$$

- If a subject is randomly assigned to treatment, knowing whether a subject is treated provides no information about the subject's potential outcomes, or background attributes.

Exclusion restriction

- Let $Y_i(z, d)$ be the potential outcome when $z_i = z$ and $d_i = d$ for $z \in (0, 1)$ and for $d \in (0, 1)$
- For example, if $z_i = 1$ and $d_i = 1$, the subject is assigned to the treatment group and receives the treatment
- Or $z_i = 1$ and $d_i = 0$ - subject is assigned treatment but does not receive treatment
- The exclusion restriction is that $Y_i(1, d) = Y_i(0, d)$ - subjects only respond to input from d_i
- The excludability assumption cannot be verified empirically because we never observe both and for the same subject

Classic Drug Experiment Example

- Treatment group receives a new drug
- Control group receives nothing
- Experiment confounds this treatment with receipt of a pill
- If patients respond to the pill rather than the pill's ingredients, excludability is violated
- Jeopardizes unbiasedness of the difference-in-means estimator

Non-interference

- Permits us to ignore the potential outcomes that would arise if subject i were affected by the treatment of other subjects
- Formally, we reduce the schedule of potential outcomes $Y_i(\mathbf{d})$, where \mathbf{d} describes all of the treatments administered to all subjects, to a much simpler schedule $Y_i(d)$, where d refers to the treatment administered to subject i .
- Implies that so long as a subject's treatment status remains constant, that subject's outcome is unaffected by the particular way in which treatments happened to be assigned to other subjects

Non-interference violated

- Police patrols displace crime from treated to untreated areas
- Non-interference violated if your estimand is following:
 - Average potential outcome when a block is treated minus average potential outcome when no blocks are treated
- If police patrols displace crime from treated to untreated areas, observed outcomes in control will not be potential outcomes when no treatment administered anywhere
- Estimated ATE will tend to exaggerate the true ATE

Core assumptions violated?

- Blinding: Researchers are not blinded to experimental assignment when measuring outcomes
- Attrition: Some of the subjects in the treatment group become discouraged and drop out of the study
- Compensatory behavior: Upon noticing that some subjects are excluded from a poverty aid treatment because they were assigned to the control group, an aid organization endeavors to treat those who were assigned to the control group
- Multiple outcomes: A weight loss intervention randomly assigns students who come to a cafeteria for lunch to a treatment consisting of small dishes and portions; outcomes are measured in terms of total calories consumed at the cafeteria during lunchtime

Estimation

Difference-in-means is an unbiased estimator of ATE

$$\begin{aligned} E\left[\frac{\sum_1^m Y_i}{m} - \frac{\sum_{m+1}^N Y_i}{N-m}\right] &= E\left[\frac{\sum_1^m Y_i}{m}\right] - E\left[\frac{\sum_{m+1}^N Y_i}{N-m}\right] \\ &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[\tau_i] = ATE \end{aligned}$$

Estimation

Difference-in-means estimator implemented via OLS

$$\begin{aligned}Y_i &= Y_i(0)(1 - d_i) + Y_i(1)d_i \\&= Y_i(0) + Y_i(1) - Y_i(0)d_i \\&= \mu_{Y(0)} + [\mu_{Y(1)} - \mu_{Y(0)}]d_i + Y_i(0) - \mu_{Y(0)} \\&\quad + [(Y_i(1) - \mu_{Y(1)}) - (Y_i(0) - \mu_{Y(0)})]d_i \\&= \alpha + \beta d_i + \epsilon_i,\end{aligned}$$

$\alpha = \mu_{Y(0)}$ (average of untreated potential outcomes for all N),

$Y_i(0)$ = untreated potential outcome

$\beta = \mu_{Y(1)} - \mu_{Y(0)}$

and ϵ_i comprises idiosyncratic variation in untreated responses plus idiosyncratic variation in treatment effects.

Chattopadhyay & Duflo 2004

- Randomized policy experiment in India
- 1990s, one-third of village council heads reserved for women
- women.csv contains subset of data from West Bengal
- Gram Panchayat (GP) = level of government
- Analysis?
 - Was randomization implemented properly?
 - Conjecture: more drinking facilities under women
 - Conjecture: no effect on irrigation

Name	Description
GP	An identifier for the Gram Panchayat (GP)
village	identifier for each village
reserved	binary variable indicating whether the GP was reserved for women leaders or not
female	binary variable indicating whether the GP had a female leader or not
irrigation	variable measuring the number of new or repaired irrigation facilities in the village since the reserve policy started
water	drinking-water facilities in the village since the reserve policy started

Table 4.6: *The Variable Names and Descriptions of the Women as Policy Makers Data.*

```
women <- read.csv("women.csv")
```

```
##proportion of female politicians in  
reserved GP vs. unreserved GP
```

```
mean(women$female[women$reserved] == 1)  
[1] 1
```

```
mean(women$female[women$reserved == 0])  
[1] 0.07476636
```

```
## drinking-water facilities  
mean(women$water[women$reserved == 1]) -  
      mean(women$water[women$reserved == 0])
```

```
## [1] 9.25223
```

```
## irrigation facilities  
mean(women4irrigation[women$reserved == 1]) -  
      mean(women$irrigation[women$reserved == 0])
```

```
## [1] -0.3693319
```

What is a treatment, after all?

Within-subject designs

- Subject i receives multiple treatments at multiple time points t
- Behavioral games; answering questions

Now potential outcomes can be written as $Y_{t-1,t,t+1}$ as a function of whether treatment is administered in the preceding, current, or next time period

ATE is now $E[Y_{010} - Y_{000}]$. Given no anticipation ($Y_{001} = Y_{000}$) and no persistence ($Y_{100} = Y_{000}$), the within-subject design identifies ATE

Within-subject designs

- Costs
 - Demand effects
 - Confounders from multiple treatments (assumptions stated above)
 - Complexity and validity: willingness to pay in Charness et al 2012
- Benefits
 - Internal validity unrelated to assignment mechanism
 - Statistical power
 - Proximity to theory
- Example: Bellemare and Shearer 2009

Between-subject designs

- Formal setup as all of part 1 of today's lecture
- Each individual exposed to 1 treatment
- ATE as difference in means of *different* groups of individuals
- Costs
 - Natural anchor with respect to (economic) decision-making?
 - Statistical power
- Example: Gneezy and List (2006)
- Overall: confounders \succ power

Within vs Between: Bellemare et al. 2016

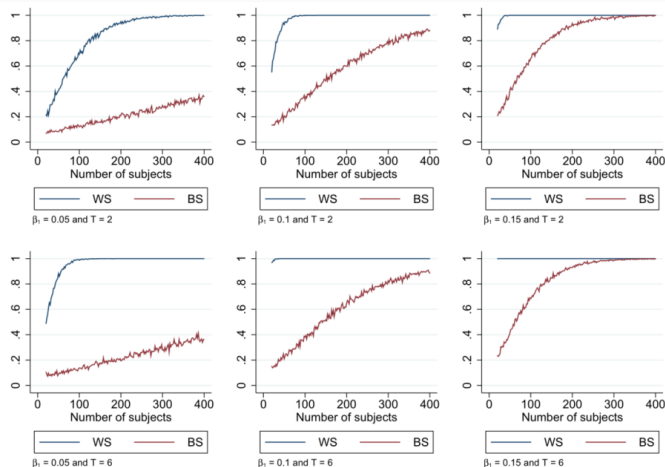


Fig. 2 Simulated statistical power of BS and WS designs with $T = 2$ and $T = 6$ for the high-noise scenario. Simulations based on values $\sigma_{\mu}^2 = 0.09$ and $\sigma_{\epsilon}^2 = 0.02$. Results for the BS design are computed by allocating the same number of subjects to control and treatment conditions for all periods. Results for the WS design are computed by assigning all subjects to the same number of control and treatment periods

Pre-treatment and post-treatment

- Q. Why would you want to collect the same information twice, pre-treatment and post-treatment? Do you gain anything?
- A. Yes, precision!
- Instead of having a single outcome measure Y_i , redefine as *change* from pre-test to post-test
- We compare 2 quantities:
 - $(Y_i - X_i)$ for $d_i = 1$
 - $(Y_i - X_i)$ for $d_i = 0$
 - *difference-in-differences* estimator

Pre-treatment and post-treatment

Is this estimator unbiased?

$$\begin{aligned}E(\widehat{ATE}) &= E[Y_i - X_i | D_i = 1] - E[Y_i - X_i | D_i = 0] \\&= E[Y_i | D_i = 1] - E[X_i | D_i = 1] - E[Y_i | D_i = 0] - E[X_i | D_i = 0] \\&= E[Y_i(1)] - E[Y_i(0)]\end{aligned}$$

In general, difference-in-means and difference-in-differences generate unbiased estimates – but what if we also care about sampling variability of this estimator?

Pre-treatment and post-treatment

In general, $SE(\widehat{ATE}') < SE(\widehat{ATE})$ if either holds

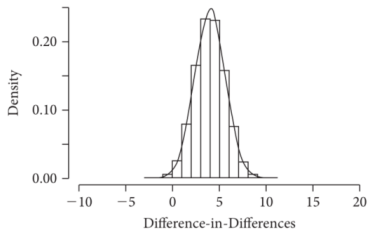
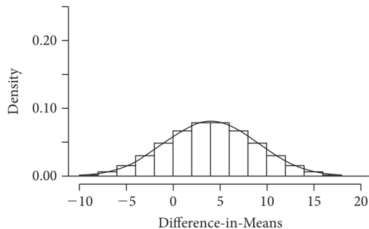
$$Cov(Y_i(0), X_i) + Cov(Y_i(1), X_i) > Var(X_i)$$

$$\frac{Cov(Y_i(0), X_i)}{Var(X_i)} + \frac{Cov(Y_i(1), X_i)}{Var(X_i)} > 1.$$

That is, *when a covariate X_i strongly predicts potential outcomes*

Pre-treatment and post-treatment

Sampling distribution of two estimators: difference-in-means and difference-in-differences



Multiple treatment arms

- Define a factorial experiment as an experiment involving factors 1 and 2, with factor 1 conditions being A and B, and factor 2 conditions being C and D and E
- Then, allocate subjects at random to every possible combination of experimental conditions
- $\{AC, AD, AE, BC, BD, BE\}$

Multiple treatment arms

From Rosen 2010

	Colin		Jose	
	Good grammar	Bad grammar	Good grammar	bad grammar
% Received reply	52	29	37	34
(N)	(100)	(100)	(100)	(100)

This design requires us to be especially careful with defining the causal estimand – what quantity are we interested in in this application?

Multiple treatment arms

Quiz: Why would these two models estimate the same quantities from the Rosen 2010 experiment?

$\{NG, HG, NB, HB\}$ are indicator variables for each of the 4 treatment groups

$J_i = 1$ if Jose Ramirez; $G_i = 1$ if good grammar

$$Y_i = b_1 CG + b_2 JG + b_3 CB + b_4 JB + u_i$$

$$Y_i = a + bJ_i + cG_i + d(J_i G_i) + u_i$$

What quantity in the table do each of the coefficients represent?

Multiple treatment arms

- A special case: conjoint experiments
- What components of the manipulation produce the observed effect?
- We can answer this with a 2x2 design as shown above, but sometimes things are multidimensional
- Subjects chose between J profiles, K times
 - Each profile has a set of L discretely valued attributes
 - D is total number of levels for attribute l

Hainmueller et al 2014

Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States.

	Immigrant 1	Immigrant 2
Prior Trips to the U.S.	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
Reason for Application	Reunite with family members already in U.S.	Reunite with family members already in U.S.
Country of Origin	Mexico	Iraq
Language Skills	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
Profession	Child care provider	Teacher
Job Experience	One to two years of job training and experience	Three to five years of job training and experience
Employment Plans	Does not have a contract with a U.S. employer but has done job interviews	Will look for work after arriving in the U.S.
Education Level	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
Gender	Female	Male

	Immigrant 1	Immigrant 2
If you had to choose between them, which of these two immigrants should be given priority to come to the United States to live?	<input type="radio"/>	<input type="radio"/>

Types of Sequential Randomised Experiments

- Non-adaptive - assignment probabilities fixed
- Treatment-adaptive - change based on number of subjects in treatment
- Covariate-adaptive - change based on covariate profiles of new and previous subjects
- Responsive-adaptive - change as function of previous units' outcomes

Treatment-adaptive designs

- *ATE* is not always quantity of interest
- Particularly online firms such as Google, Tiktok, FB, etc.
 - Randomly assign sampled users to different arms and dynamically re-orient sample based on which is more successful/more informative
 - Identify which of many will get the most clicks
- But also of interest to political scientists: Ballot initiatives and malfeasance information
- How do adaptive multi-arm trials work?

Table 2: Study One, Treatments and Outcome Measures

	Minimum Wage	Right to Work
Question Text	Imagine that the following ballot measure were up for a vote in your state. [ballot measure text]. If this measure were on the ballot in your state, would you vote in favor or against? [I would vote in favor of this measure; I would vote against this measure]	Imagine that the following ballot measure were up for a vote in your state. [ballot measure text]. If this measure were on the ballot in your state, would you vote in favor or against? [I would vote in favor of this measure; I would vote against this measure]
Proposal 1	The measure would: increase the minimum wage [from {current}] to {current + 1} per hour, adjusted annually for inflation, and provide that no more than \$3.02 per hour in tip income may be used to offset the minimum wage of employees who regularly receive tips.	The measure would [amend the State Constitution to]: prohibit, as a condition of employment, forced membership in a labor organization (union) or forced payments of dues or fees, in full or pro-rata ("fair-share"), to a union. The measure will also make any activity which violates employees' rights provided by the bill illegal and ineffective and allow legal remedies for anyone injured as a result of another person violating or threatening to violate those employees' rights. The measure will not apply to union agreements entered into before the effective date of the measure, unless those agreements are amended or renewed after the effective date of the measure.
Proposal 2	The measure would: raise the minimum wage [from {current}] to {current + 1} per hour effective September 30th, 2021. Each September 30th thereafter, minimum wage shall increase by \$1.00 per hour until the minimum wage reaches {current + 5} per hour on September 30th, 2026. From that point forward, future minimum wage increases shall revert to being adjusted annually for inflation starting September 30th, 2027.	The measure [reads / would amend the State Constitution to read]: The right of persons to work may not be denied or abridged on account of membership or nonmembership in any labor union or labor organization, and all contracts in negotiation or abrogation of such rights are hereby declared to be invalid, void, and unenforceable.
Proposal 3	The measure reads: Shall the minimum wage for adults over the age of 18 be raised [from {current}] to {current + 1} per hour by January 1, 2019?	The measure would [amend the State Constitution to]: ban any new employment contract that requires employee to resign from or belong to a union, pay union dues, or make other payment to a union. Required contributions to charity or other third party instead of payments to union are also banned. Employees must authorize payroll deduction to unions. Violations of the section is a misdemeanor.
Proposal 4	The measure would: raise the minimum wage [from {current}] to {current + 1} per hour worked if the employer provides health benefits, or {current + 2} per hour worked if the employer does not provide health benefits.	The measure [reads / would amend the State Constitution to read]: No person shall be deprived of life, liberty or property without due process of law. The right of persons to work shall not be denied or abridged on account of membership or nonmembership in any labor union, or labor organization.
Proposal 5	The measure would: raise the State minimum wage rate [from {current}] to at least {current + 1} per hour, and require annual increases in that rate if there are annual increases in the cost of living.	

Boldface text indicates randomly varied elements.

Treatment-adaptive designs

- When researchers are initially agnostic about the relative performance of the K arms, priors are distributed uniformly over parameter space, i.e. $Beta(1, 1)$
- In each t , treatment is randomly assigned according to probability of arms being best (= highest success rate)

$$P\left[\Theta_k = \max_k \{\Theta_1, \dots, \Theta_K\} | (X_1^{n_{1,t}}, \dots, X_K^{n_{K,t}})\right]$$

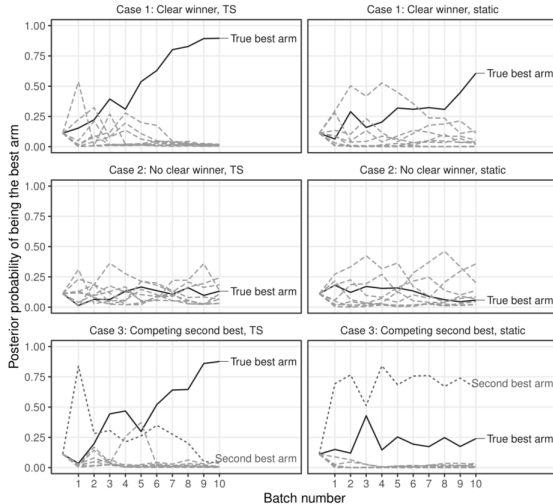
for K arms, vector of responses under treatment arm k observed up until and including t $X_k^{\{n_{k,t}\}}$ and Θ_k distributions of success rates

Treatment-adaptive designs

- Simulations to illustrate design and estimation
- Sample 100 observations for each of 10 periods, updating posterior probability of being best after each period, and assign treatment probabilities in the subsequent period accordingly
- In the first case, one arm has a true 0.20 probability of success, and the remaining 8 arms have a 0.10 probability of success

Treatment-adaptive designs

Figure 1: Simulated Posterior Probabilities Over Time, Thompson Sampling and Static Designs



Treatment-adaptive designs

Figure 3: Study One, Overtime Posterior Probabilities

