

Experimental Design as Market Design

Yusuke Narita

April 14, 2017

Golden Age of Randomized Experiments

Randomized Controlled Trials (RCT) or A/B tests

=Gold standard of evidence-based decision-making in medicine/policy/biz

RCTs are often with

- large # of subjects (esp. in med & biz)
- high-stakes or even life-or-death treatment (esp. in med & policy)
 - cancer drugs
 - basic income
 - health insurance



FINLAND: Basic Income experiment authorized by Parliament

December 18, 2016 ▾ Kate McFarland ▾ News News & events



My agenda: Design randomized experiments that respect subject welfare

My Agenda

Method: I propose an experimental design with these features:

- ① *Welfare*: It optimally assigns treatment to subjects with
 - better predicted treatment effects
 - stronger preferences for the treatment
- ② *Incentive*: It incentivizes subjects to honestly reveal preferences
- ③ *Information*: It randomly assigns treatment & generates at least as much causal info as vanilla RCT



Standard designs fail to satisfy (1) or (2)

Empirics: I also quantify points (1)-(3) by empirically applying the design to a water cleaning RCT in Kenya (Kremer et al. 11)

Experimental Design Setting

i_1, \dots, i_n : Subjects

t_0, t_1, \dots, t_m : Treatments

$c_t \in \mathbb{N}$: Treatment t 's capacity (with $\sum_t c_t = n$)

w_{it} : Subject i 's willingness to pay for treatment t with

$$w_{it} \geq w_{it'} \Leftrightarrow i \text{ weakly prefers } t \text{ over } t'$$

e_{ti} : t 's predicted treatment effect for subject i with

$$e_{ti} \geq e_{t'i} \Leftrightarrow t \text{ is predicted weakly more effective than } t' \text{ for } i$$

t_0 : Placebo or control with normalization $e_{t_0i} = w_{it_0} = 0$ for all i

Detail

Where Do WTP & Predicted Effects Come from?

Preferences or WTP w_{it} : A few possible sources:

- ① Designer may estimate it from choice data (e.g. Kremer et al. 11)
- ② Each subject i may self-report $(w_{it})_t$ (e.g. Ashraf et al. 10)

Predicted effect e_{ti} : A few possible sources:

- ① Designer may estimate it from prior data
 - Prior RCT (repeated RCT common in medical & biz)
 - Prior quasi-experimental or observational studies
- ② (Designer may ask experts to forecast it (e.g. DellaVigna-Pope 16))

Proposal 1/2

Definition (**Experiment-as-Market** a.k.a. EXaM)

- ① In computer, distribute common artificial budget b to every subject
- ② Find “price-discriminated market equilibrium”
i.e., feasible treatment assignment prob.s (p_{it}^*) & prices (π_{te}^*) with
 - Utility maximization s.t. budget constraint: For every subject i ,
 $(p_{it}^*)_t \in \operatorname{argmax}_{\text{feasible } (p_{it})_t} \sum_t p_{it} w_{it}$ s.t. $\sum_t p_{it} \pi_{te_i}^* \leq b$
(Ties are broken by uniformly mixing cheapest (p_{it})'s)
 - Effectiveness-discriminated treatment pricing: $\forall t \exists \alpha_t < 0 \& \beta_t \forall e$,
$$\pi_{te}^* = \alpha_t e + \beta_t$$
 - Meeting capacity constraint: $\sum_i p_{it}^* \leq c_t$ for every treatment t

Proposal 2/2

In general, however, it's possible $p_{it}^* = 0$ or 1 .

Definition (Experiment-as-Market a.k.a. EXaM)

- ① Compute p_{it}^* as in last slide
- ② Fix any $\epsilon \in [0, 1]$ as a lower bound on treatment prob.s and compute

$$p_{it}^*(\epsilon) \equiv (1 - q)p_{it}^* + qp_{it}^0$$

where

- $p_{it}^0 \equiv c_t/n$ is vanilla RCT's uniform treatment assignment prob.s
- $q \equiv \inf\{q \geq 0 | p_{it}^*(\epsilon) \in [\epsilon, 1 - \epsilon] \text{ for all } i, t\}$

- ③ Draw final treatment assignment from $(p_{it}^*(\epsilon))$.

Welfare: Theory of EXaM

Proposition (Randomized Controlled Welfare Property)

For any $\epsilon \geq 0$, EXaM $p_{it}^*(\epsilon)$ always exists & optimally assigns treatment to subjects with

- better predicted treatment effects or
- stronger preferences for the treatment

i.e., \nexists another assignment prob.s (p_{it}) with $p_{it} \in [\epsilon, 1 - \epsilon]$ for all i, t &

- $\sum_t p_{it} e_{ti} \geq \sum_t p_{it}^* e_{ti}$ (expected predicted effect) &
- $\sum_t p_{it} w_{it} \geq \sum_t p_{it}^* w_{it}$ (expected willingness to pay)

for all i with at least one strict inequality.

Incentive: Theory of EXaM 1/2

Proposition (Informal)

For any ϵ , EXaM is “almost incentive compatible” as for WTP reporting

Consider sequence of exp. design problems indexed by # of subjects n :

$$(i_1, \dots, i_n, t_0, t_1, \dots, t_m, (c_t^n), (w_{it}^n), (e_{ti}^n))_{n \in \mathbb{N}}$$

For n -th problem in the sequence, let

$$p_{it}^{*n}(w_i, F_w) \equiv \int p_{it}^{*n}(w_i, w_{-i}) \times \Pr(w_{-i} | w_{-i} \sim_{iid} F_w) dw_{-i},$$

where

- $p_{it}^{*n}(w_i, w_{-i})$: $\Pr(\text{EXaM assigns } i \text{ to } t \text{ when subjects report WTP } (w_i, w_{-i}) \text{ in } n\text{-th problem}).$
- $\Pr(w_{-i} | w_{-i} \sim_{iid} F_w)$: $\Pr(w_{-i} \text{ is realized from } n-1 \text{ independent identically distributed draws from distribution } F_w)$

Incentive: Theory of EXaM 2/2

Take its limit as # of subjects $n \rightarrow \infty$ to get

$$p_{it}^{*\infty}(w_i, F_w) \equiv \lim_{n \rightarrow \infty} p_{it}^{*n}(w_i, F_w),$$

the limit prob. that EXaM assigns i to t when i reports w_i & others' WTP is drawn from F_w .

Proposition (Asymptotic Incentive Compatibility of EXaM)

For any ϵ , EXaM is almost incentive compatible as for WTP reporting, i.e., for all F_w & w_i, w'_i ,

$$\underbrace{\sum_t p_{it}^{*\infty}(w_i, F_w) \times w_{it}}_{E(\text{true WTP}) \text{ for } i \text{ from truthtelling}} \geq \underbrace{\sum_t p_{it}^{*\infty}(w'_i, F_w) \times w_{it}}_{\dots \text{from lying}}$$

Information: Econometrics of EXaM 1/2

Definition (Data)

Each experimental design generates data $(t_i, y_i, w_{it}, e_{ti})$ where

- t_i : subject i 's assigned treatment
- y_i : i 's observed outcome

To ignore sampling randomness, consider limit as # of subjects $n \rightarrow \infty$

Proposition

If RCT's data identifies a causal effect, EXaM's data also identifies it.

Corollary

EXaM's data identifies Average Treatment Effect, Conditional ATE, etc.

Information: Econometrics of EXaM 2/2

Corollary

EXaM's data identifies Average Treatment Effect, Conditional ATE, etc.

Why? EXaM randomly assigns treatment conditional on $x_i \equiv (w_{it}, e_{ti})_t$

- ① Identify treatment t_1 's ATE relative to control t_0 conditional on x_i by

$$\underbrace{E(Y_i | i \text{ assigned } t_1, x_i)}_{\text{Observable}} - \underbrace{E(Y_i | i \text{ assigned } t_0, x_i)}_{\text{Observable}} \equiv CATE(x_i)$$

- ② Compute weighted avg of $CATE$ to get treatment effects of interest

e.g. Average Treatment Effect: $\int CATE(x_i) dF(x_i)$

Marginal TE at $w_{it} = w$: $\int CATE(x_i) dF(x_i | w_{it} = w)$

Also OK to condition on propensity score instead of possibly high dim. x_i

Comparison with Existing Designs

Proposition

None of followings satisfies the above welfare/incentive/info properties.

- Vanilla RCT

Designs respecting WTP or preferences:

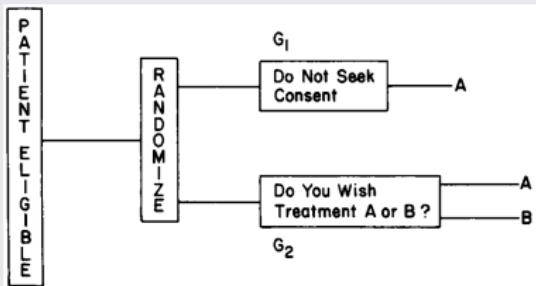
- “Consent Trial” (Zelen 79, Angrist-Imbens 91, some medical RCT)
- “Selective Trial” (Chassang-Miquel-Snowberg 12)
- “Thompson Sampling” (Thompson 33, some web A/B tests)

Designs respecting predicted effects:

- “Play-the-Winner Trial” (Wei-Durham 78, some medical RCT)
- “Response-Adaptive Biased Coin Design” (Eisele 94, medical RCT)
- “Empirical Welfare Maximization” (Manski 05, B-Dupas 12)

Comparison: Randomized Consent Trial (Zelen 79 NEJM)

Definition (RConsentT)



RCT with “one-sided non-compliance” intentionally introduced to respect pref.s

Later rediscovered & refined in econometrics
(Bloom 84 & Angrist-Imbens 91)

Application: abortion & depression RCT,
Job Training Partnership Act in the US

Proposition

RConsentT violates Randomized Controlled Welfare Property
(since it ignores WTP intensity & predicted effectiveness).

Empirical Application



Setting: Spring water cleaning RCT in Kenya
(Kremer et al. 11)

Embedding this setting into my model:
 t_1 vs t_0 : access to a cleaned spring vs no access

w_{it_1} : subject i 's willingness to pay for t_1 (access to a cleaned spring)

$e_{t_1 i}$: treatment effect of t_1 for subject i

What I Do: Implement & compare Experiment-as-Market vs RCT

Estimating Treatment Effects $e_{t_1 i}$ of Better Water

- ① Run OLS regression at (household, spring, trip) level:

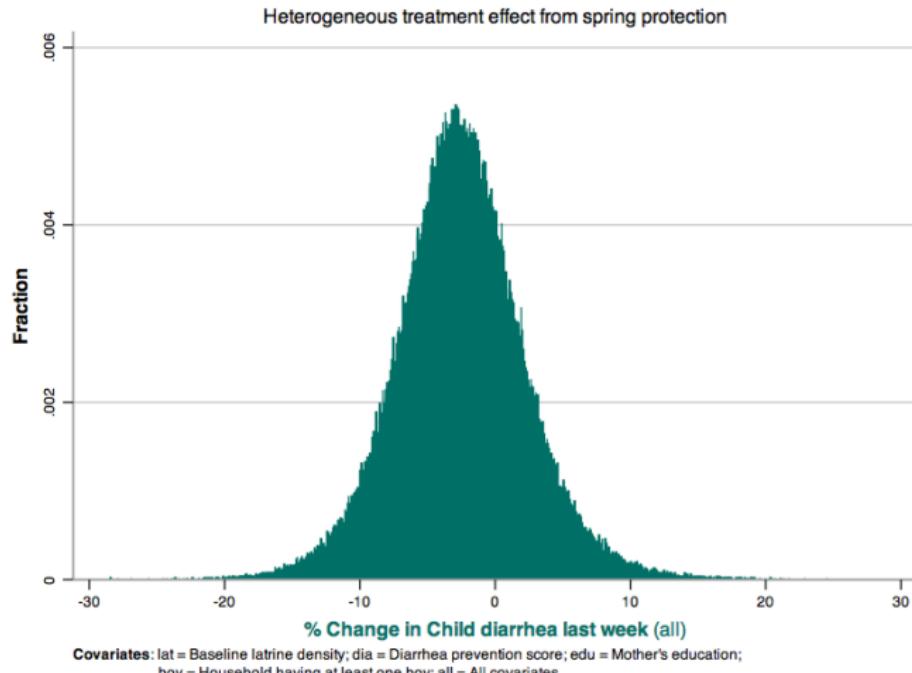
$$Y_{ijt} = (\phi_1 + \phi_2 X_i) T_{jt} + \alpha_i + \alpha_t + u_{ij} + \epsilon_{ijt}$$

where

- outcome Y_{ijt}
 $\equiv 1\{\text{HH } i \text{ drawing water from spring } j \text{ in trip } t \text{ has child with diarrhea}\}$
- treatment $T_{jt} \equiv 1\{\text{spring } j \text{ is cleaned up in trip } t\}$
- X_i : household i 's covariates (baseline latrine density, diarrhea prevention score, mother's years of educ, having boy)

- ② Estimate treatment effects $e_{t_1 i}$ by $\hat{\phi}_1 + \hat{\phi}_2 X_i$

Heterogeneity in Treatment Effects $e_{t_1 i}$



Estimating Willingness-to-pay w_{it_1} for Better Water

- ① Use households' spring choices to estimate mixed logit model:

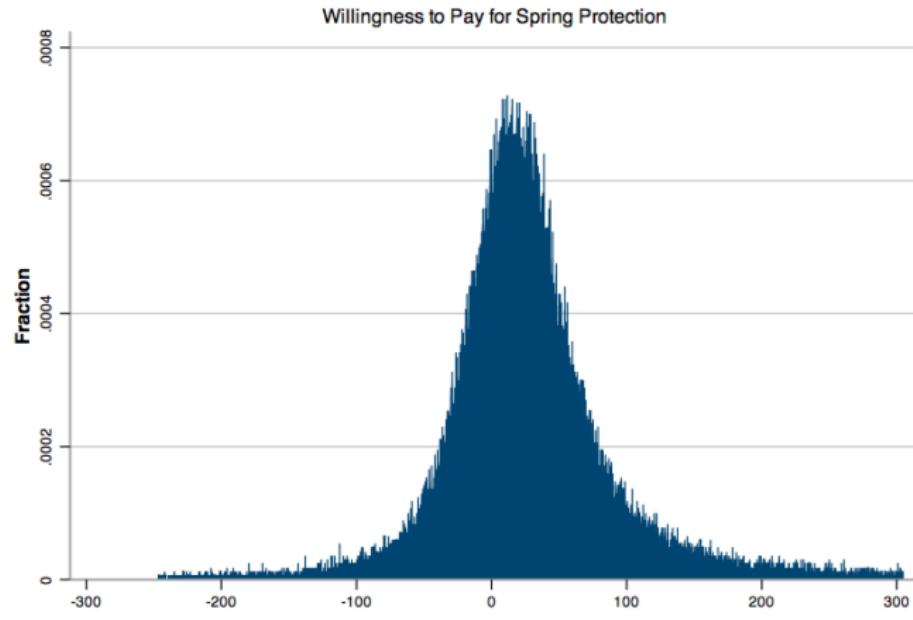
$$u_{ijt} = (\beta_i + \gamma_1 X_i) T_{jt} - C_i D_{ij} + Z_j + e_{ijt}$$

where

- D_{ij} : household i 's roundtrip distance to spring j
- β_i & C_i : normal & triangular (≥ 0), respectively.
- e_{ijt} : logit utility shocks

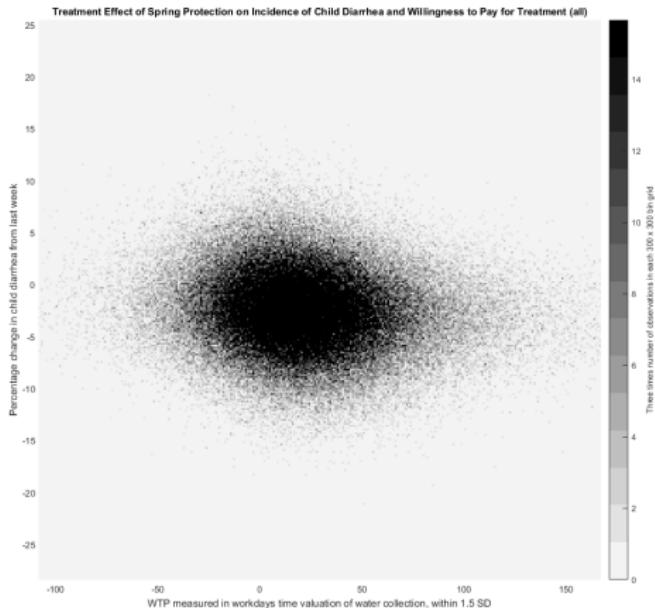
- ② Estimate WTP for treatment $w_{t_1 i}$ by $\hat{\beta}_i + \hat{\gamma}_1 X_i$

Heterogeneity in Willingness-to-pay w_{it_1}



Covariates: lat = Baseline latrine density; dia = Diarrhea prevention score; edu = Mother's education; boy = Household having at least one boy; all = All covariates. Figure exhibits $\mu \pm 3\sigma$.

Little Correlation/Selection between w_{it_1} & $e_{t_1 i}$



→ Need to respect both WTP w_{it_1} & treatment effects $e_{t_1 i}$

Comparing EXaM vs RCT

Simulate estimated models to compare EXaM & RCT w.r.t.

- ① *Welfare*: How much do they assign treatment to subjects with
 - better predicted treatment effects
 - stronger preferences for the treatment
- ② *Incentive*: How much do they incentivize honest pref. reporting?
- ③ *Information*: How much causal information do they produce?

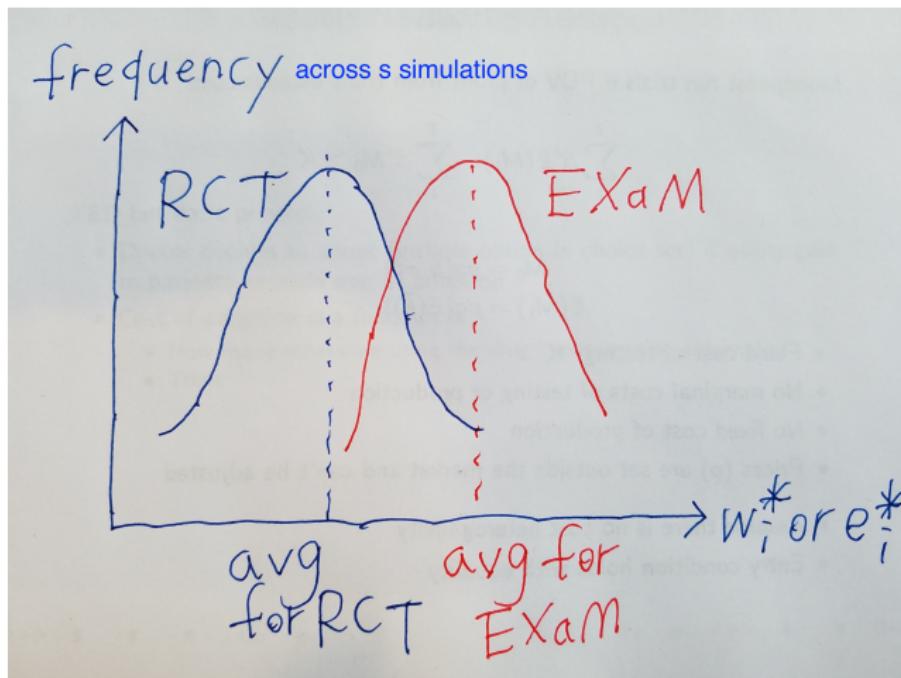
Throughout the analysis, fix the set of subjects & treatments. Set

- capacity $c_{t_1} = \#$ of subjects assigned to t_1 in Kremer et al.'s RCT
- capacity $c_{t_0} = (\text{total } \# \text{ of subjects}) - c_{t_1}$
- $\epsilon = 0.1$

EXaM vs RCT: Welfare 1/2

- 1) Repeat the following procedure for s times
 - 1.a) Simulate (e_{it_1}, w_{it_1}) from their estimated distributions
 - 1.b) Run EXaM on the simulated data in step a to get $p_{it}^*(\epsilon)$
 - 1.c) For each subject i , compute $w_i^* \equiv E(w_{it} \text{ from } i\text{'s assigned treatment } t \text{ under } p_{it}^*(\epsilon)) = \sum_t p_{it}^*(\epsilon)w_{it}$ $e_i^* \equiv E(e_{it} \text{ from } i\text{'s assigned treatment } t \text{ under } p_{it}^*(\epsilon)) = \sum_t p_{it}^*(\epsilon)e_{it}$
- 2) Draw the histogram of w_i^* & e_i^* over the s simulations & subjects

EXaM vs RCT: Welfare 2/2



EXaM improves on RCT in terms of welfare

EXaM vs RCT: Information 1/2

1) Repeat the following procedure for s times

1.a) Simulate (e_{it_1}, w_{it_1}) & run EXaM to get $p_{it_1}^*(\epsilon)$

1.b) Use $p_{it_1}^*(\epsilon)$ to draw a final deterministic treatment assignment

$$D_i \equiv 1\{i \text{ is assigned to } t_1\}$$

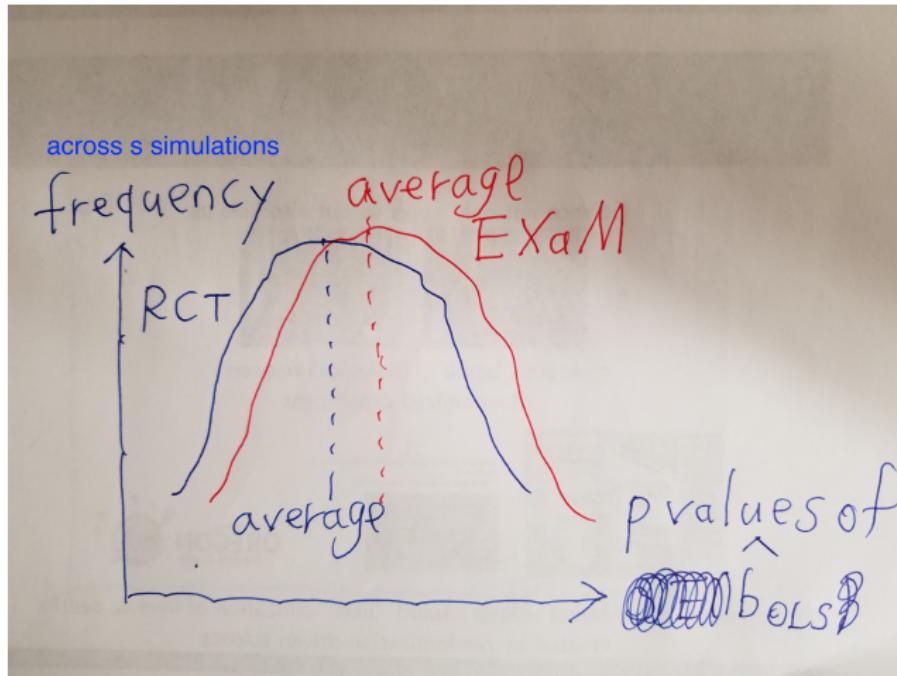
1.c) Draw outcome Y_i simulating Kremer et al's estimated OLS model

1.d) Estimate treatment effects by \hat{b}_{OLS} from

$$Y_i = a + bD_i + \underbrace{\sum_p c_p 1\{p_{it_1}^*(\epsilon) = p\}}_{\text{propensity score controls}} + e_i$$

2) Draw the histogram of p values of \hat{b}_{OLS} over the s simulations

EXaM vs RCT: Information 2/2

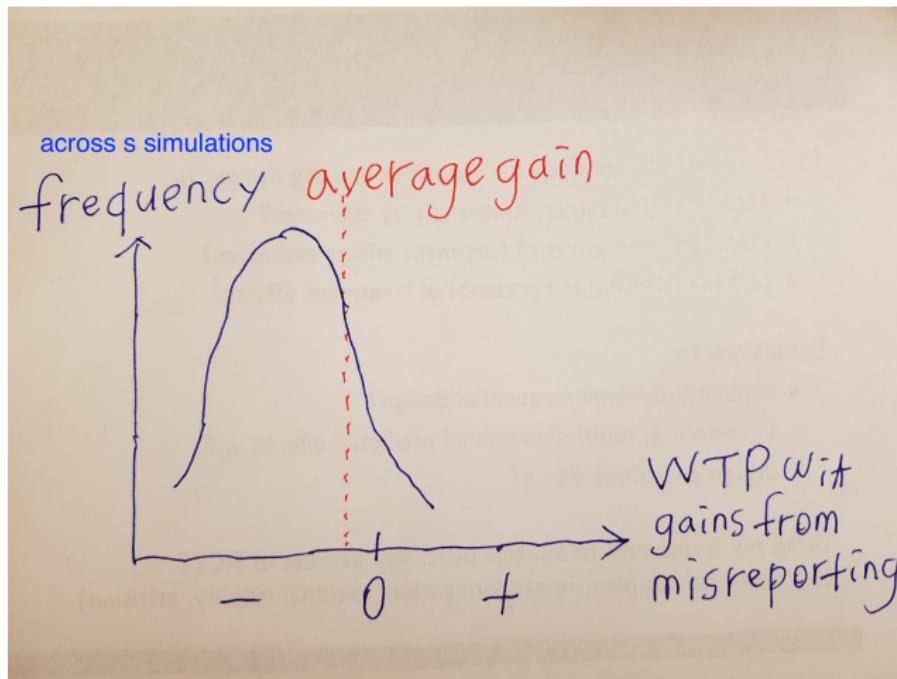


EXaM enables precise enough causal inference

Almost Incentive Compatibility of EXaM 1/2

- 1) Repeat the following procedure for s times
 - 1.a) Simulate (e_{it_1}, w_{it_1}) from their estimated distributions
 - 1.b) Run EXaM on the simulated data in step a
 - 1.c) Randomly pick one subject j & one WTP manipulation w'_{jt_1} by j
 - 1.d) Run EXaM on the simulated data with WTP manipulation w'_{jt_1}
 - 1.e) Compute WTP gain from manipulating reported WTP:
$$Gain \equiv E(w_{jt} \text{ from } j\text{'s assigned treatment } t | \text{ step d}) - E(\text{same} | \text{ step b})$$
- 2) Draw the histogram of $Gain$ over the s simulations

Almost Incentive Compatibility of EXaM 2/2



→ EXaM gives subjects little incentive for WTP misreporting

Future Directions

More statistical comparison of EAaM vs existing designs by...

- standard errors in treatment effects estimates?
- mean squared errors of treatment effects estimates?
- power of testing the presence of treatment effects?

Extensions to...

- explicitly dynamic/sequential design?
- uncertain & multi-dimensional predicted effects e_{ti} ?
- design of sample size n ?

Using my framework to analyze other key aspects of RCT?
(e.g. pre-analysis plan, re-randomization, external validity, attrition)

Stepping Back: The Science Unto Death

Science not only saves us but also kills us:



Atomic bomb

Biological weapon

Randomized experiment



FINLAND: Basic Income experiment authorized by Parliament

December 20, 2016 • Kata McFarland • News-Nova Scotia



Experiment-as-Market is step toward “best” allocation of lives & deaths created by randomization-driven science