Robust Randomized Experiments for Causal Effects Under Privacy

Manjusha Kancharla, Hyunseung Kang

Department of Statistics
University of Wisconsin - Madison

EuroCIM 2021, Theme 3: Mixed Topics
July 2, 2021

Randomized Control Trials (RCTs) and Data Privacy

- Randomized control trials
  - Gold standard to estimate average causal effects of a treatment/intervention (ATE).
  - Carry an axiomatic assumption that individuals freely share their response with the investigator.
- But, what if the response is sensitive in nature?
  - E.g., voting behaviour, alcohol consumption, mental health related.
  - Such responses should (ideally) be privatized.
- What if some responses (e.g., from online A/B tests) are protected by law?
  - GDPR (2016): EU law for online data privacy.

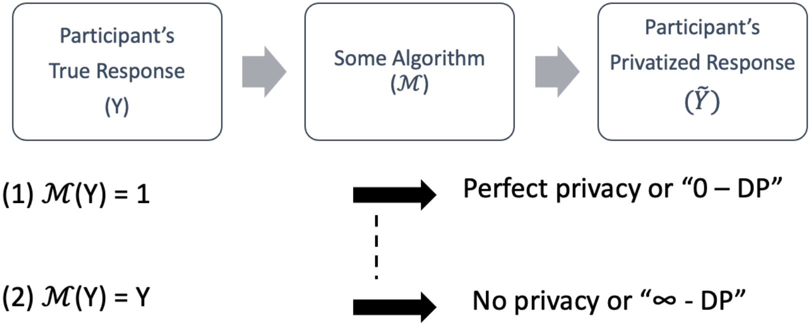
Can we estimate causal effects in an experiment WHILE protecting individual's data privacy, specifically their response to treatment?

Our Proposal

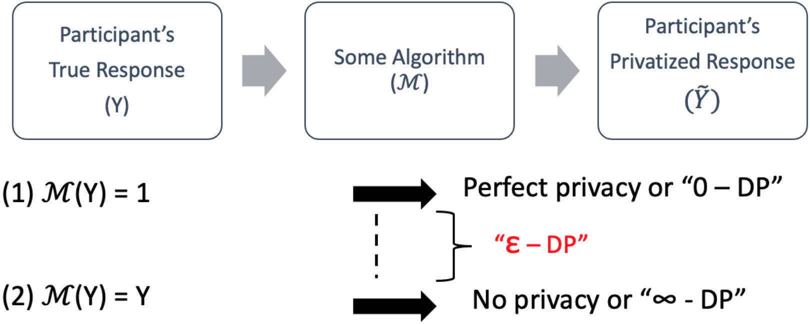
Robust, Private Randomized Control Trial (RP-RCT):

1. Guarantees individual's data privacy through Differential Privacy
2. Allows for estimation of causal effects

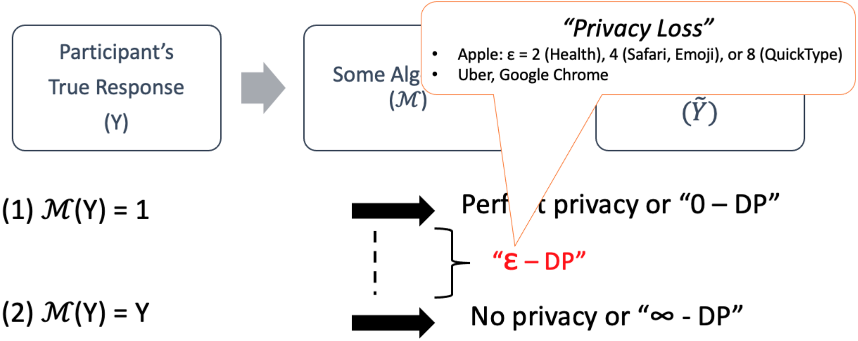
Differential Privacy (DP) (Dwork et al. (2006))



Differential Privacy (DP) (Dwork et al. (2006))



Differential Privacy (DP) (Dwork et al. (2006))



Example: Forced Randomized Response (FRR) (Warner (1965))

Response Prompt: “Did you pay attention in lecture today?”

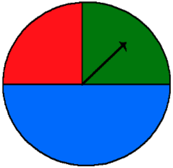
Bob’s True Response:
Y = “NO”

FRR Algorithm, $\mathcal{M}$(Y).

If spinner lands on :

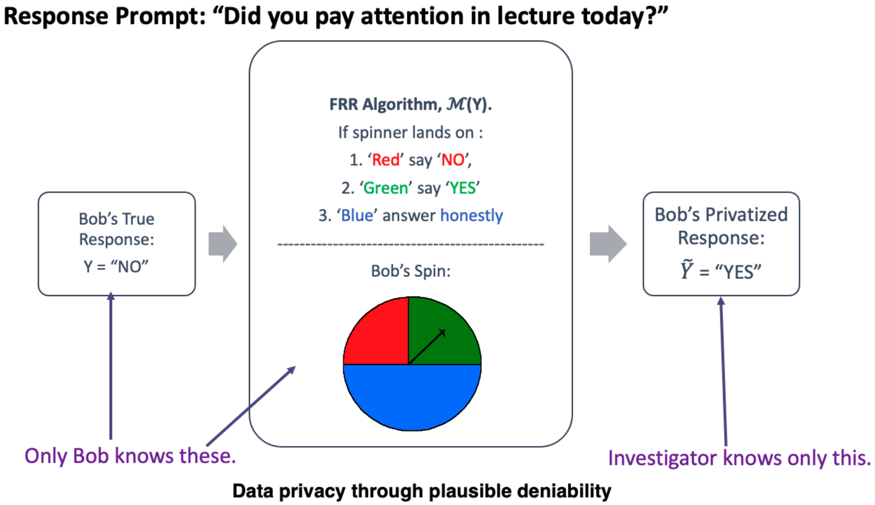
1. ‘Red’ say ‘NO’,
2. ‘Green’ say ‘YES’
3. ‘Blue’ answer honestly

Bob’s Spin:



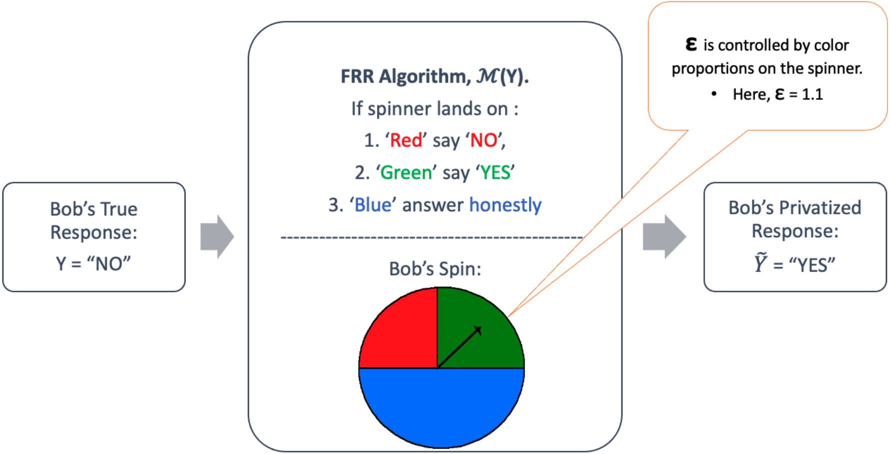
Bob’s Privatized Response:
$\tilde{Y}$ = “YES”

Example: Forced Randomized Response (FRR) (Warner (1965))



Example: Forced Randomized Response (FRR) (Warner (1965))

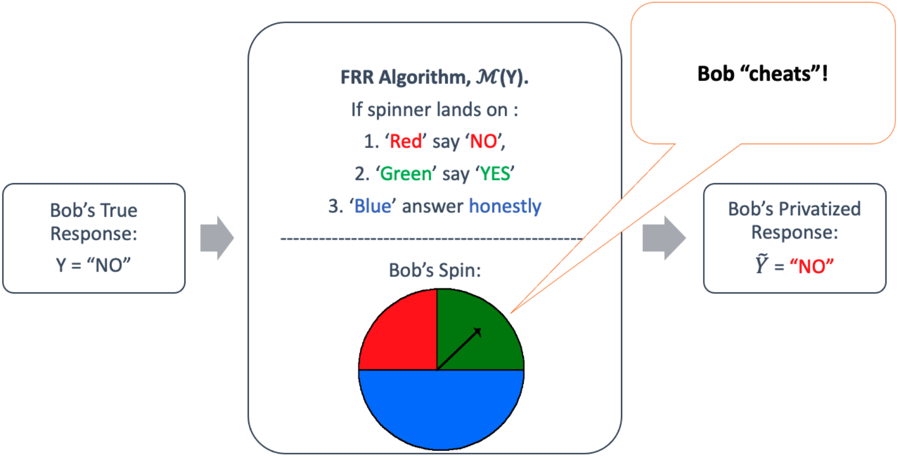
Response Prompt: "Did you pay attention in lecture today?"



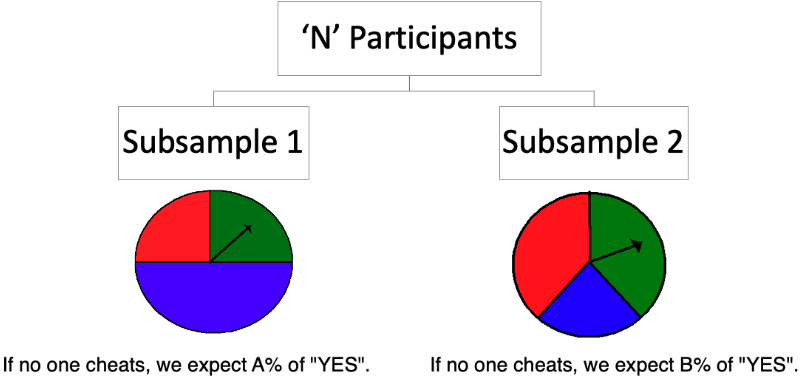

EVERYBODY LIES.
HOUSE

Non-Adherence (i.e., Cheating) in FRR

Response Prompt: “Did you pay attention in lecture today?”

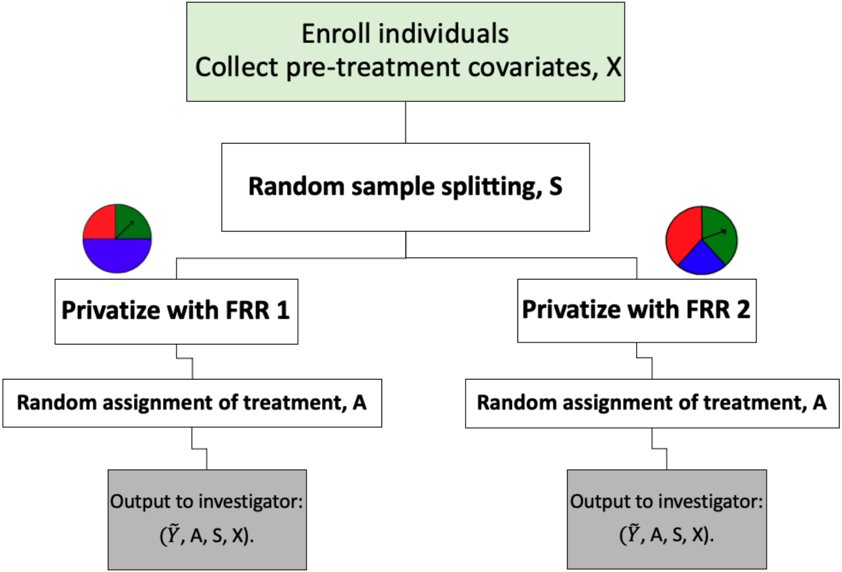


Detecting Proportion of Cheaters Via Sample Splitting and Mixed FRR (Clark and Desharnais (1998))

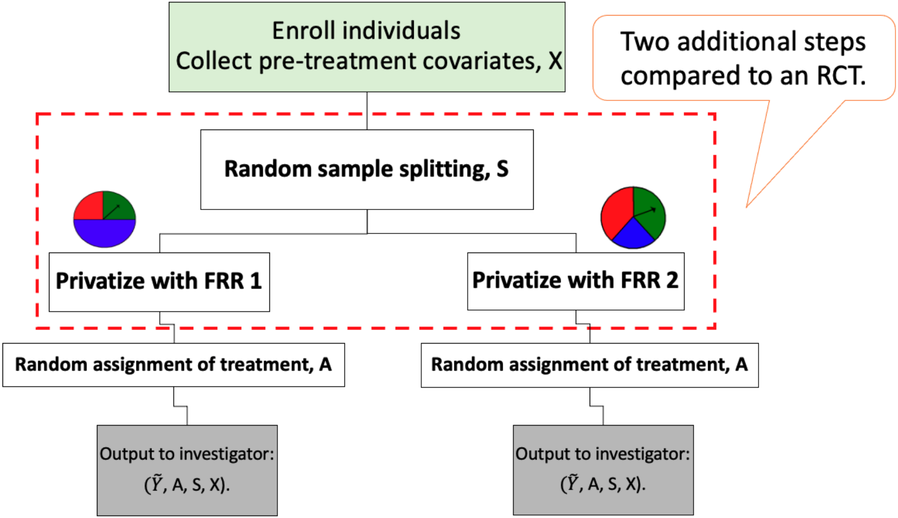


- Expected difference in % of "YES": $\Delta_E = A\% - B\%$.
- If observed difference $\Delta_O \neq \Delta_E$, there are cheaters!
- Key point: Sample splitting + mixed FRRs (i.e., spinners)

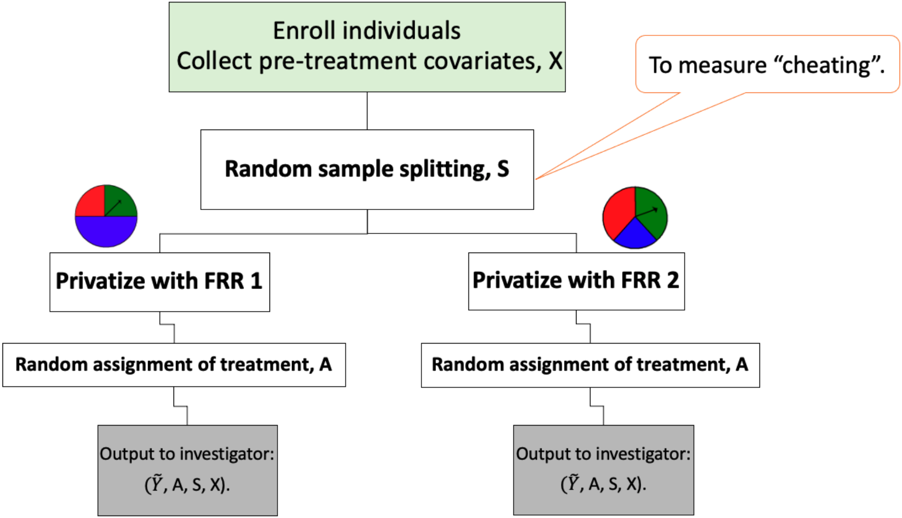
Our Proposal: Robust, Private RCT (RP-RCT)



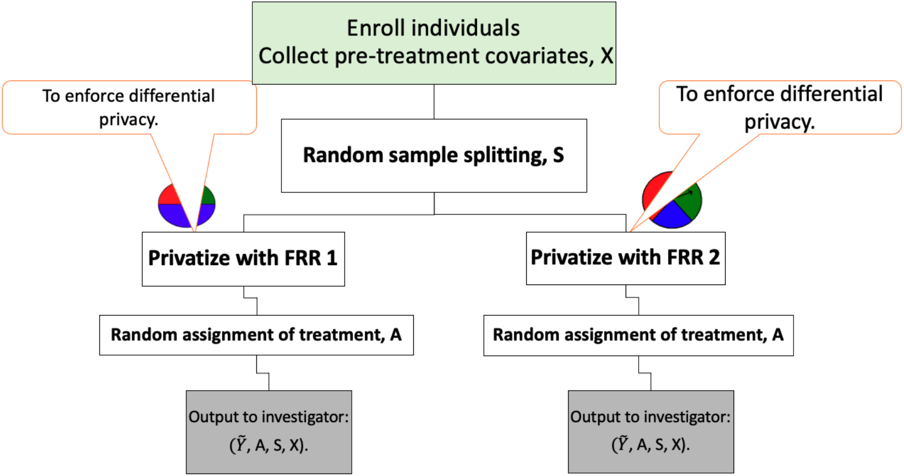
Our Proposal: Robust, Private RCT (RP-RCT)



Our Proposal: Robust, Private RCT (RP-RCT)



Our Proposal: Robust, Private RCT (RP-RCT)



Properties of RP-RCT

Theorem: Identification of ATE

Let λ be the proportion of non-cheaters in the study. Under RP-RCT and $0 < Pr(\lambda) \leq 1$, we can identify ATE among non-cheaters, i.e.,

$$E[Y_i(1) - Y_i(0) \mid \text{Non-Cheaters}] = \frac{E[\tilde{Y}_i \mid A_i = 1] - E[\tilde{Y}_i \mid A_i = 0]}{\lambda \times r_\epsilon},$$

where, λ can be estimated from privatized data and r_ϵ is the amount of privatization used.

- All honest: RP-RCT identifies population ATE.
- All cheaters: RP-RCT cannot identify ATE.
- Similar to LATE (Angrist, Imbens, and Rubin (1996)), we cannot identify who is a non-cheater from the data.
- Covariate-adjusted, doubly-robust estimation is possible.

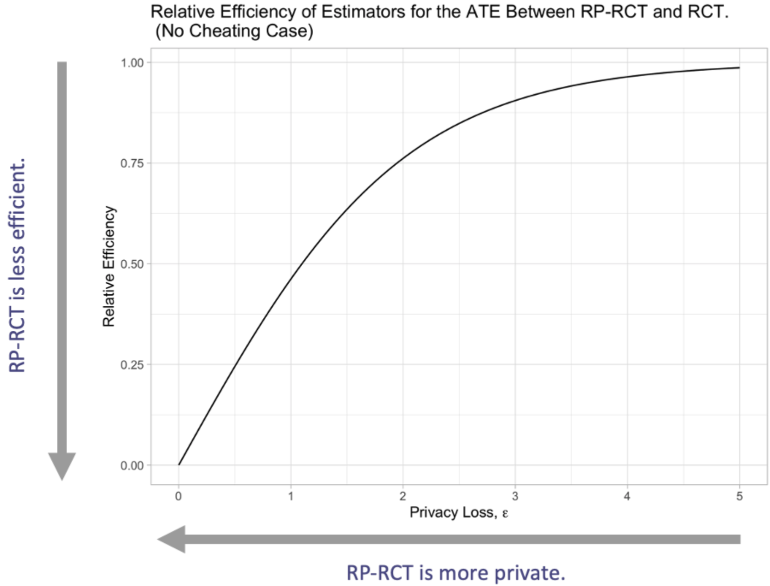
Properties of RP-RCT

Theorem: Differentially Privacy of RP-RCT.

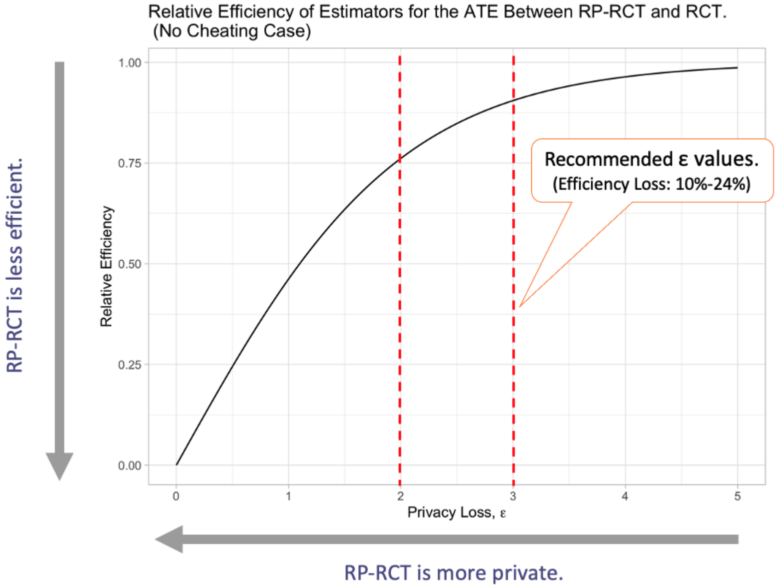
For any treatment arm, the response $\tilde{Y}_i$ from RP-RCT is $\epsilon-$differentially private.

- ϵ depends on two FRRs (i.e., spinners) in each subsample.
- ϵ, the acceptable privacy level, is chosen by investigator.

Efficiency – Privacy Tradeoff for RP-RCT (vs RCT)



Efficiency – Privacy Tradeoff for RP-RCT (vs RCT)



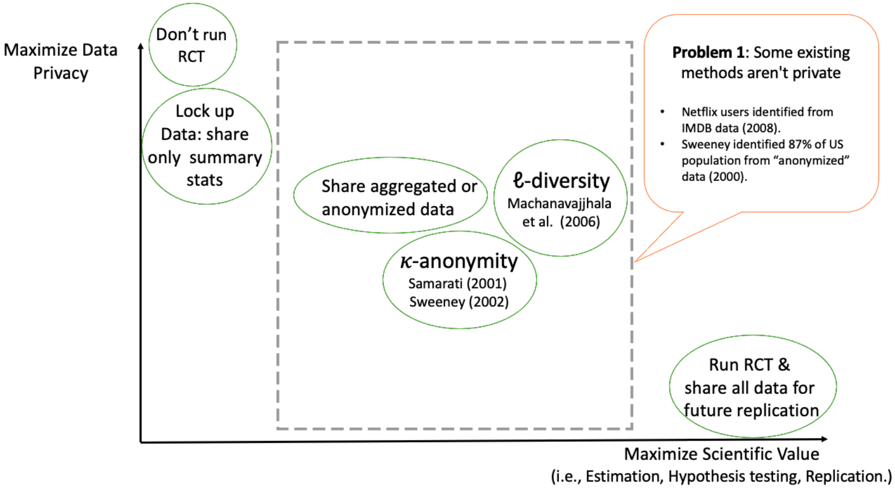
Future Work

Extend RP-RCT to

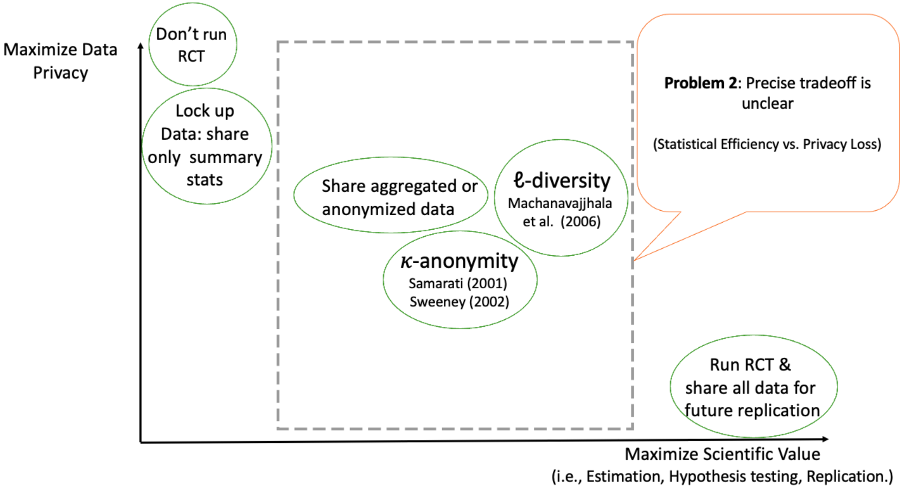
- Accommodate continuous but bounded responses to treatment
- Accommodate non-compliance
- Observational studies where treatment is also private

Thank you!

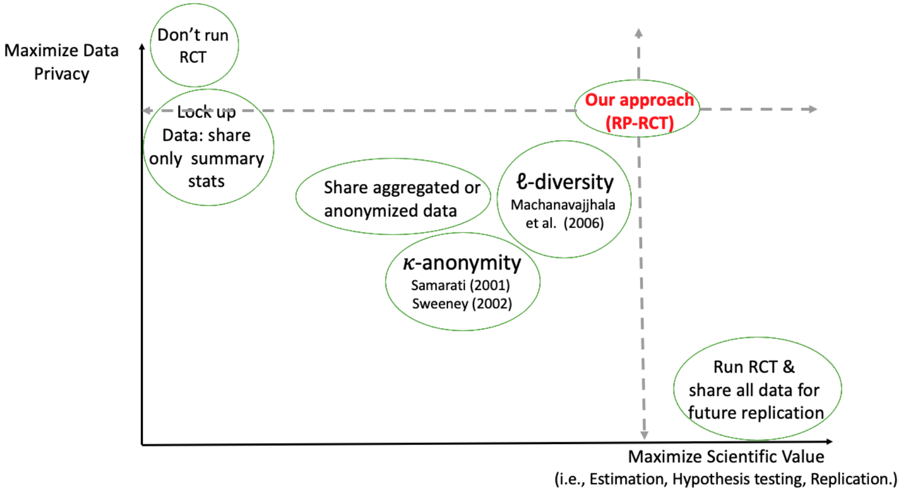
Current State of Data Privacy in RCTs



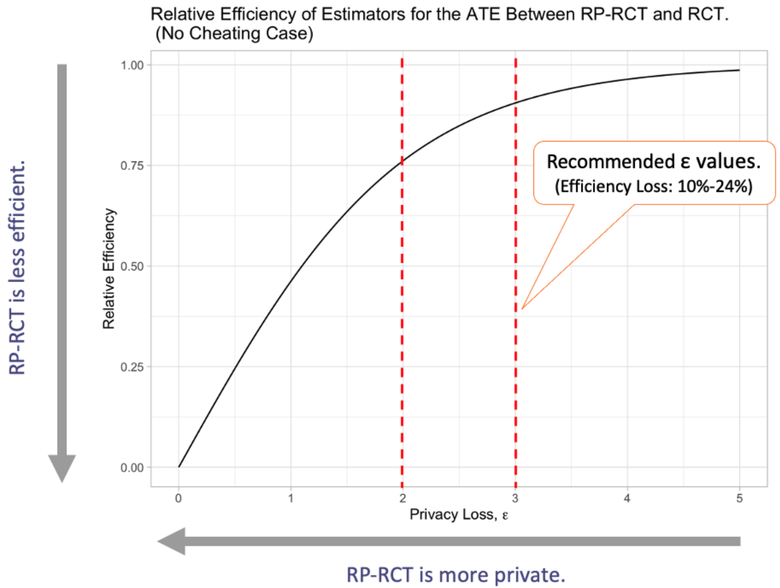
Current State of Data Privacy in RCTs



Current State of Data Privacy in RCTs



Efficiency – Privacy Tradeoff for RP-RCT



References

Angrist, Joshua D, Imbens Guido W., and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Clark, Stephen J., and Robert A. Desharnais. 1998. "Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model." *Psychological Methods* 3 (2): 160–68.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." In *Theory of Cryptography*, 265–84. Springer Berlin Heidelberg.

Warner, S. L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60: 63–69.