Background and Motivation
○○○○○

Results
○○○○○○○○○○

Power and Sample Size Analysis
○○○○○

Methods in Context
○○○○○○○

Robust Designs for Prospective Randomized Trials Surveying Sensitive Topics

Evan T. R. Rosenman, Rina Friedberg, and Mike Baiocchi

Harvard Data Science Initiative

August 8, 2022

Outline

A Cluster-Randomized Trial in Nairobi, Kenya

- Project emerged out of a **cluster-randomized trial** seeking to reduce sexual violence among adolescents in Nairobi, Kenya
  - 4,100 sixth grade girls
  - Randomized treatment among 94 schools
  - Two years of follow-up (2016 to 2018)

A Cluster-Randomized Trial in Nairobi, Kenya

- Project emerged out of a **cluster-randomized trial** seeking to reduce sexual violence among adolescents in Nairobi, Kenya
  - 4,100 sixth grade girls
  - Randomized treatment among 94 schools
  - Two years of follow-up (2016 to 2018)
- **Treatment:** ImPower
  - Locally designed 12-hour training program.
  - Empowerment, situational awareness, verbal skills, self-defense

A Cluster-Randomized Trial in Nairobi, Kenya

- Project emerged out of a **cluster-randomized trial** seeking to reduce sexual violence among adolescents in Nairobi, Kenya
  - 4,100 sixth grade girls
  - Randomized treatment among 94 schools
  - Two years of follow-up (2016 to 2018)
- **Treatment:** ImPower
  - Locally designed 12-hour training program.
  - Empowerment, situational awareness, verbal skills, self-defense
- **Outcome**: incidence of sexual violence in prior 12 months

The Setting

Measurement Tool

- Due to size of study, outcome and covariate data was **self-reported** via survey

- Several efforts made to improve data collection quality
  - Multiple languages
  - Colloquial phrasing
  - Privacy maintenance

Motivating Questions

Results from the study showed **no protective effect** of treatment vs. standard of care

Repeated critique: misreporting error attenuated causal estimates toward zero

- Survey responses may be erroneous due to fear, shame, etc.
- Old canard: measurement error biases results toward the null

Motivating Questions

Results from the study showed **no protective effect** of treatment vs. standard of care

Repeated critique: misreporting error attenuated causal estimates toward zero

- Survey responses may be erroneous due to fear, shame, etc.
- Old canard: measurement error biases results toward the null

Research questions

1. With binary outcomes, how does non-differential* misreporting error affect the difference-in-means causal estimator?

2. How can a researcher can power a study with misreporting?

_____

*Meaning: misreporting behavior not influenced by the treatment itself

Outline

Running Example

Randomized controlled trial involving $i = 1, \ldots, 2n$ individuals.

Students randomized to receive either a **violence prevention program** ("intervention") or an unrelated training ("control").

Goal is to reduce students' experience of violence.
Outcome is binary:

- "yes, I experienced violence in the prior 12 months" $(= 1)$
- "no, I didn't experience violence in the prior 12 months" $(= 0)$

Outcome Definitions

- Each individual i has two potential outcomes (Rubin, 1974)
    - $Y_i(1) \in \{0, 1\}$, the outcome if treated
    - $Y_i(0) \in \{0, 1\}$, the outcome if given the control

Outcome Definitions

- Each individual i has two potential outcomes (Rubin, 1974)
  - $Y_i(1) \in \{0, 1\}$, the outcome if treated
  - $Y_i(0) \in \{0, 1\}$, the outcome if given the control
- Denote as $Y_i^{(t)} \in \{0, 1\}$ the realized, **true outcome** experienced by individual i. Follows definition

$$Y_i^{(t)} = W_i Y_i(1) + (1 - W_i) Y_i(0).$$

where $W_i \in \{0, 1\}$ is the treatment indicator.

Outcome Definitions

- Each individual i has two potential outcomes (Rubin, 1974)
    - $Y_i(1) \in \{0, 1\}$, the outcome if treated
    - $Y_i(0) \in \{0, 1\}$, the outcome if given the control
- Denote as $Y_i^{(t)} \in \{0, 1\}$ the realized, **true outcome** experienced by individual i. Follows definition

$$Y_i^{(t)} = W_i Y_i(1) + (1 - W_i) Y_i(0).$$

where $W_i \in \{0, 1\}$ is the treatment indicator.

- Denote as $Y_i^{(r)} \in \{0, 1\}$, as the **reported outcome** for individual i. In our setting, possible that $Y_i^{(r)} \neq Y_i^{(t)}$.

Reporting Classes

Reporting class	$Y_i^{(r)}$ when $Y_i^{(t)} = 0$	$Y_i^{(r)}$ when $Y_i^{(t)} = 1$
True (T_i)	0	1
~~False (F_i)~~	~~1~~	~~0~~
Never (N_i)	0	0
Always (A_i)	1	1

Table 1: Reporting behavior for each of the four reporting classes.

Indicators $T_i, N_i, A_i \in \{0, 1\}$ satisfy

$$T_i + N_i + A_i = 1$$

for all $i \in 1, \ldots, 2n$.

Response Classes

Response class	$Y_i(0)$	$Y_i(1)$
Decrease (D_i)	1	0
Increase (I_i)	0	1
Unsusceptible (U_i)	0	0
Predisposed (P_i)	1	1

Table 2: Potential outcomes for each of the four response classes.

See discussion in Hernán and Robins (2010).
Indicators $D_i, I_i, U_i, P_i \in \{0, 1\}$ satisfy

$$D_i + I_i + U_i + P_i = 1$$

for all $i \in 1, \ldots, 2n$.

Background and Motivation
ooooo

Results
ooooo●oooo

Power and Sample Size Analysis
ooooo

Methods in Context
ooooooo

Joint Distribution of Class Types

	Y(0)	**Y(1)**	**True-rep.**	**Always-rep.**	**Never-rep.**	**False-rep.**	
Decrease	1	0	$\overline{TD}$	$\overline{AD}$	$\overline{ND}$	x	$\overline{D}$
Increase	0	1	$\overline{TI}$	$\overline{AI}$	$\overline{NI}$	x	$\overline{I}$
Unsusceptible	0	0	$\overline{TU}$	$\overline{AU}$	$\overline{NU}$	x	$\overline{U}$
Predisposed	1	1	$\overline{TP}$	$\overline{AP}$	$\overline{NP}$	x	$\overline{P}$
			$\overline{T}$	$\overline{A}$	$\overline{N}$		

Table 3: Population proportions across response and reporting classes.

Bias Results (I)

- The $2n$ RCT units are assumed sampled from a (large) super-population of N_{sp} units. n units treated, n units control
- We consider estimating the super-population treatment effect

$$\tau = \frac{1}{N_{\text{sp}}} \sum_{i=j}^{N_{\text{sp}}} Y_j(1) - Y_j(0)$$

using the difference-in-means estimator,

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{2n} Y_i^{(r)} W_i - \frac{1}{n} \sum_{i=1}^{2n} Y_i^{(r)} (1 - W_i) \,.$$

Background and Motivation
ooooo

Results
ooooooo●●oo

Power and Sample Size Analysis
ooooo

Methods in Context
ooooooo

Bias Results (II)

Theorem (Bias of Difference-in-Means Estimator)

Define $\tau_i = Y_i(1) - Y_i(0)$ for $i = 1, 2, \ldots, N_{sp}$. The bias of the difference-in-means estimator in estimating τ is given by

$$\mathrm{Bias}(\hat{\tau}) = -\left(\bar{A} + \bar{N}\right)\tau - \mathrm{cov}\left(A, \tau\right) - \mathrm{cov}\left(N, \tau\right)$$
$$= -\overline{NI} + \overline{ND} - \overline{AI} + \overline{AD}$$

where $\mathrm{cov}\left(A, \tau\right)$ is the super-population covariance between A_i and τ_i, and $\mathrm{cov}\left(N, \tau\right)$ defined analogously.

Bias and Power Under Independence

As a direct corollary of Theorem 1, we see that if $A_i, N_i \perp\!\!\!\perp \tau_i$, then our estimate will be shrunk multiplicatively toward 0:

$$\frac{\mathbb{E}\left(\hat{\tau}\right)}{\tau} = 1 - \bar{A} - \bar{N} \qquad \text{under independence.}$$

Theorem

Suppose $\tau \neq 0$ and, in the super-population, $A_i, N_i \perp\!\!\!\perp \tau_i$. Then, the detection power is a strictly decreasing function of $\overline{A}$ and $\overline{N}$.

Interaction Between Class Types

Key result: <u>joint</u> distribution of response and reporting classes characterizes the bias of causal estimate.

Interaction Between Class Types

Key result: <u>joint</u> distribution of response and reporting classes characterizes the bias of causal estimate.

Why does the **joint distribution** matter? Consider our violence prevention example:

- Suppose $\overline{D} > 0, \overline{I} = 0 \Rightarrow$ some helped, none harmed

Interaction Between Class Types

Key result: <u>joint</u> distribution of response and reporting classes characterizes the bias of causal estimate.

Why does the **joint distribution** matter? Consider our violence prevention example:

- Suppose $\overline{D} > 0, \overline{I} = 0 \Rightarrow$ some helped, none harmed
- Among those in Decrease $(Y_i(0) = 1, Y_i(1) = 0)$ response class, some may feel violence is avoidable $\Rightarrow$ shame at experiencing violence yields Never-reporting

Interaction Between Class Types

Key result: joint distribution of response and reporting classes characterizes the bias of causal estimate.

Why does the **joint distribution** matter? Consider our violence prevention example:

- Suppose $\overline{D} > 0, \overline{I} = 0 \Rightarrow$ some helped, none harmed
- Among those in Decrease ($Y_i(0) = 1, Y_i(1) = 0$) response class, some may feel violence is avoidable $\Rightarrow$ shame at experiencing violence yields Never-reporting
- No misreporting among Predisposed or Immune.

Interaction Between Class Types

Key result: <u>joint</u> distribution of response and reporting classes characterizes the bias of causal estimate.

Why does the **joint distribution** matter? Consider our violence prevention example:

- Suppose $\overline{D} > 0, \overline{I} = 0 \Rightarrow$ some helped, none harmed
- Among those in Decrease $(Y_i(0) = 1, Y_i(1) = 0)$ response class, some may feel violence is avoidable $\Rightarrow$ shame at experiencing violence yields Never-reporting
- No misreporting among Predisposed or Immune.
- In this case, Never-reporters would be more prevalent among those responsive to treatment $\Rightarrow$ bigger bias!

Outline

1. **Background and Motivation**

2. **Results**

3. **Power and Sample Size Analysis**

4. **Methods in Context**

Sensitivity Model

- Practically it is often unrealistic to expect exact independence between reporting and response classes

- But deviations from independence may be small. Can bound them using a **sensitivity model**.

Sensitivity Model

- Practically it is often unrealistic to expect exact independence between reporting and response classes
- But deviations from independence may be small. Can bound them using a **sensitivity model**.

Definition

Under sensitivity model indexed by $\Gamma \geq 1$, every subgroup proportion in the joint distribution table differs by a multiplicative factor no greater than Γ from the proportion under row-column independence. For example,

$$\frac{1}{\Gamma} \leq \frac{\overline{TD}}{\overline{T} \cdot \overline{D}} \leq \Gamma$$

with analogous bounds for other entries in joint incidence table.

Joint Distribution of Class Types

	Y(0)	Y(1)	True-rep.	Always-rep.	Never-rep.	False-rep.	
Decrease	1	0	$\overline{TD}$	$\overline{AD}$	$\overline{ND}$	x	$\overline{D}$
Increase	0	1	$\overline{TI}$	$\overline{AI}$	$\overline{NI}$	x	$\overline{I}$
Unsusceptible	0	0	$\overline{TU}$	$\overline{AU}$	$\overline{NU}$	x	$\overline{U}$
Predisposed	1	1	$\overline{TP}$	$\overline{AP}$	$\overline{NP}$	x	$\overline{P}$
			$\overline{T}$	$\overline{A}$	$\overline{N}$		

Table 4: Population proportions across response and reporting classes.

Sensitivity model bounds the **level of deviation** from row/column independence in this table.

Worst-Case Sample-Size Calculations (I)

- **Proposed method:** compute sample size under *worst-case* power, constrained by the sensitivity model

Worst-Case Sample-Size Calculations (I)

- **Proposed method:** compute sample size under *worst-case* power, constrained by the sensitivity model
- User inputs (ideally derived from high-quality pilot study!)
  - Type I error bound α and Type II error bound β
    - $\Rightarrow$ power is $1 - \beta$
  - Estimates of $\boldsymbol{\pi}_{\mathsf{rep}} = \left(\overline{T}, \overline{A}, \overline{N}\right)$ and $\boldsymbol{\pi}_{\mathsf{res}} = \left(\overline{D}, \overline{I}, \overline{U}, \overline{P}\right)$
  - Worst-case independence deviation $\Gamma \geq 1$

Worst-Case Sample-Size Calculations (I)

- **Proposed method:** compute sample size under *worst-case* power, constrained by the sensitivity model
- User inputs (ideally derived from high-quality pilot study!)
  - Type I error bound α and Type II error bound β
    $\Rightarrow$ power is $1 - \beta$
  - Estimates of $\boldsymbol{\pi}_{\mathsf{rep}} = \left(\overline{T}, \overline{A}, \overline{N} \right)$ and $\boldsymbol{\pi}_{\mathsf{res}} = \left(\overline{D}, \overline{I}, \overline{U}, \overline{P} \right)$
  - Worst-case independence deviation $\Gamma \geq 1$
- Optimization variable $\boldsymbol{\delta}$ is population proportion table. Expected test statistic can be written as

$$t \left(\boldsymbol{\pi}_{\mathsf{rep}}, \boldsymbol{\pi}_{\mathsf{res}}, \boldsymbol{\delta}, n \right)$$

Worst-Case Sample-Size Calculations (III)

Obtain sample size by solving Optimization Problem 1:

$$
\begin{aligned}
\operatorname*{argmin}_{n} \max_{\boldsymbol{\delta}} \quad & t\left(\boldsymbol{\pi}_{\text{rep}}, \boldsymbol{\pi}_{\text{res}}, \boldsymbol{\delta}, n\right) \\
\text{subject to} \quad & \Phi\left(\Phi^{-1}(\alpha) - t\left(\boldsymbol{\pi}_{\text{rep}}, \boldsymbol{\pi}_{\text{res}}, \boldsymbol{\delta}, n\right)\right) \geq 1 - \beta, \\
& \boldsymbol{\delta}\mathbb{1}_3 = \boldsymbol{\pi}_{\text{rep}}^{\top}, \\
& \boldsymbol{\delta}^{\top}\mathbb{1}_4 = \boldsymbol{\pi}_{\text{rep}}^{\top}, \\
& \frac{1}{\Gamma} \leq \boldsymbol{\delta}\left/\left(\boldsymbol{\pi}_{\text{res}}^{\top}\boldsymbol{\pi}_{\text{rep}}\right)\right. \leq \Gamma,
\end{aligned}
\tag{1}
$$

where $\Phi(\cdot)$ is the CDF of a standard normal, and $\mathbb{1}_c$ is the length-c vector containing all ones.

This is a quadratic fractional programming problem, and can be solved by **Dinkelbach's Method**.

Outline

Returning to our Kenya example

- Baiocchi et al. (2019) estimated the annual baseline rate of sexual violence in this population to be approximately 7%

Returning to our Kenya example

- Baiocchi et al. (2019) estimated the annual baseline rate of sexual violence in this population to be approximately 7%

- Expectation from a pilot study: 50% reduction in sexual violence (we now know effect is null)

Returning to our Kenya example

- Baiocchi et al. (2019) estimated the annual baseline rate of sexual violence in this population to be approximately 7%

- Expectation from a pilot study: 50% reduction in sexual violence (we now know effect is null)

- Under $\alpha = 0.05$ and $\beta = 0.2$, a standard power analysis would estimate we would need to recruit 998 girls to be in our study

Posited parameter values

- Never-reporters expected in study population
  - Girls may experience shame or confusion
  - Robust literature (Cook et al., 2011; Fisher and Cullen, 2000).

  Always-reporters and False-reporters are assumed absent.

Posited parameter values

- Never-reporters expected in study population
    - Girls may experience shame or confusion
    - Robust literature (Cook et al., 2011; Fisher and Cullen, 2000).

    Always-reporters and False-reporters are assumed absent.

- We consider a grid of possible values

$$0 \leq \overline{N} \leq 0.20 \quad \text{and} \quad 1 \leq \Gamma \leq 2$$

Posited parameter values

- Never-reporters expected in study population
  - Girls may experience shame or confusion
  - Robust literature (Cook et al., 2011; Fisher and Cullen, 2000).

  Always-reporters and False-reporters are assumed absent.

- We consider a grid of possible values

$$0 \leq \overline{N} \leq 0.20 \quad \text{and} \quad 1 \leq \Gamma \leq 2$$

- Under the assumption that $\overline{I} = 0$ (i.e., no one is harmed by the treatment), we have:

$$\overline{D} = 0.035, \quad \overline{P} = 0.035, \quad \text{and} \quad \overline{U} = 0.930$$
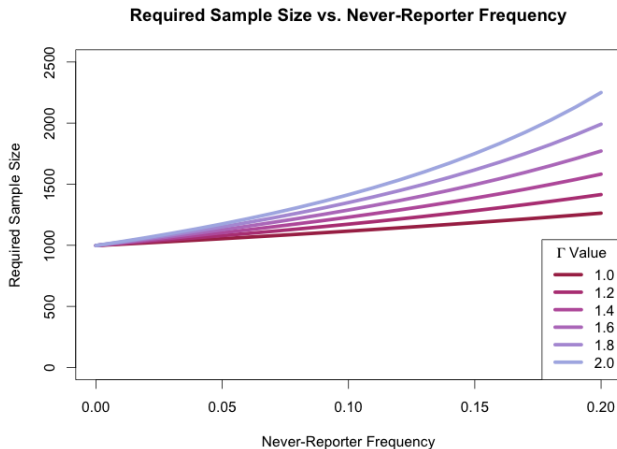
Results

Figure 1: Required sample size by Never-reporter frequency.

What's Going On?

- $\Gamma = 1$: sample size grows slowly with Never-reporter frequency
- As Γ gets larger, the algorithm allows for more **adversarial** allocations of the population proportions
    - Allocates as much of the population as possible to the $\overline{ND}$ subgroup, subject to Γ constraint
    - Yields the smallest possible causal estimate
- Hence, when $\overline{N} = 20\%$ and $\Gamma = 2$, the required sample size is nearly twice as large as the case when $\Gamma = 1$.

Wrap-Up

- Consider the effect of **non-differential misreporting error** on causal inference with a binary outcome

Wrap-Up

- Consider the effect of **non-differential misreporting error** on causal inference with a binary outcome
- Key contributions
  - Demonstrate that **joint distribution** of reporting and response classes characterizes the bias (and variance) of the difference-in-means estimator
  - Proposed a method for practitioners to identify adequate **sample sizes** in presence of misreporting

Wrap-Up

- Consider the effect of **non-differential misreporting error** on causal inference with a binary outcome
- Key contributions
  - Demonstrate that **joint distribution** of reporting and response classes characterizes the bias (and variance) of the difference-in-means estimator
  - Proposed a method for practitioners to identify adequate **sample sizes** in presence of misreporting
- **Takeaways**
  - The time to think about misreporting error is the **design phase** of the experiment!
  - Some error can be mitigated through **careful design choices**.
  - Some may only be addressable by recruiting more participants

Thanks!

Thanks to Mike & Rina!

arXiv: 2108.08944
Forthcoming in *AJE*.

References (I)

Baiocchi, M., Friedberg, R., Rosenman, E., Amuyunzu-Nyamongo, M., Oguda, G., Otieno, D., and Sarnquist, C. (2019). Prevalence and risk factors for sexual assault among class 6 female students in unplanned settlements of nairobi, kenya: Baseline analysis from the impower & sources of strength cluster randomized controlled trial. *PLoS one*, 14(6):e0213359.

Cook, S. L., Gidycz, C. A., Koss, M. P., and Murphy, M. (2011). Emerging issues in the measurement of rape victimization. *Violence against women*, 17(2):201–218.

Fisher, B. S. and Cullen, F. T. (2000). Measuring the sexual victimization of women: Evolution, current controversies, and future research. *Criminal justice*, 4:317–390.

Hernán, M. A. and Robins, J. M. (2010). Causal inference.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Proof of Theorem 1

Proof.

Define
$$Y_i^{(r)}(1) = (1 - N_i)(1 - A_i)Y_i(1) + A_i,$$
$$Y_i^{(r)}(0) = (1 - N_i)(1 - A_i)Y_i(0) + A_i$$

s.t. $Y_i^{(r)} = W_i Y_i^{(r)}(1) + (1 - W_i)Y_i^{(r)}(0)$. Now, proceed as normal!

$$\mathbb{E}(\hat{\tau}) = \mathbb{E}_R\left(\mathbb{E}_W\left(\frac{1}{n}\sum_{i=1}^{N_{\text{sp}}} Y_i^{(r)} W_i R_i - \frac{1}{n}\sum_{i=1}^{N_{\text{sp}}} Y_i^{(r)}(1 - W_i)R_i \mid \{R_i\}_{i=1}^{N_{\text{sp}}}\right)\right)$$

$$= \mathbb{E}_R\left(\frac{1}{2n}\left(\sum_{i=1}^{N_{\text{sp}}} R_i(Y_i^{(r)}(1) - Y_i^{(r)}(0))\right)\right)$$

$$= \frac{1}{N_{\text{sp}}}\sum_{i=1}^{N_{\text{sp}}}(1 - N_i)(1 - A_i)\tau_i = \frac{1}{N_{\text{sp}}}\sum_{i=1}^{N_{\text{sp}}}(1 - N_i - A_i)\tau_i$$

$$= \left(1 - \overline{N} - \overline{A}\right)\tau - \text{cov}(N, \tau) - \text{cov}(A, \tau).$$

The Intervention

- **ImPower**: a locally designed gender-based violence prevention program

- Goal is to target four causal pathways:
    - Empowerment
    - Situational awareness
    - Verbal skills
    - Physical self-defense

Our Approach (I)

- Model each respondent as a member of a **reporting class**, defining how the individual reports realized outcomes

  - True-reporter
  - False-reporter

  - Never-reporter
  - Always-reporter

- Also invoke notion of a **response class** (Hernán and Robins, 2010), defining how each individual responds to the treatment

  - Decrease
  - Increase

  - Unsusceptible
  - Presdisposed

- Show the **joint distribution** of reporting classes and response classes exactly characterizes error terms for a causal estimate (but not just the marginal distribution of reporting classes).

- Propose a novel minimax procedure to obtain adequately powered experiments in presence of misreporting.

Explanations for Misreporting Behavior

Why might someone be a Never-reporter?

- Fear of negative consequences if they report a violent event
- Feelings of shame for having been a victim of violence
- Confusion about what constitutes violence

How can researchers mitigate misreporting in the design phase?

- Use a technique like "randomized response," which offers a higher degree of anonymity
- Word questions to avoid triggers of shame
- Use detailed descriptive scenarios; use specific physical and verbal acts; or use common slang terms more familiar to students to clarify what constitutes "violence"

Sampling Mechanism

- Suppose our $2n$ units for the RCT are sampled from a very large super-population of N_{sp} units

- Once sampled, n units selected via simple random sample to receive the intervention

- Target of estimation is the super-population treatment effect,

$$\tau = \frac{1}{N_{\mathrm{sp}}} \sum_{i=1}^{N_{\mathrm{sp}}} Y_i(1) - Y_i(0) \, .$$

- $R_i \in \{0, 1\}$: indicator of being sampled into RCT.
  $W_i \in \{0, 1\}$: treatment indicator. Assume $R_i \perp\!\!\!\perp W_j$.

- Expectation of any estimator ϕ is defined as

$$\mathbb{E}(\phi) = E_R \left(E_W \left(\phi \mid \{R_i\}_{i=1}^{N_{\mathrm{sp}}} \right) \right) \, .$$

Response Classes

Define binary indicators $D_i, I_i, U_i, P_i \in \{0, 1\}$ reflecting whether each individual i falls into the Decrease, Increase, Unsusceptible, or Predisposed response classes.

Every individual belongs to exactly one class, so

$$D_i + I_i + U_i + P_i = 1 \text{ for all } i.$$

Reporting Classes

- Assume there are no False-reporters
- Define two fixed, binary constants,

$$N_i = \begin{cases} 1 & \text{if } i \text{ is a Never-reporter} \\ 0 & \text{otherwise} \end{cases}$$

$$A_i = \begin{cases} 1 & \text{if } i \text{ is an Always-reporter} \\ 0 & \text{otherwise} \end{cases}$$

where $N_i + A_i \leq 1$ and $N_i = A_i = 0$ signifies that someone is a True-reporter

- We can express the reported outcome as

$$Y_i^{(r)} = (1 - N_i)(1 - A_i)\left(W_i Y_i(1) + (1 - W_i)Y_i(0)\right) + A_i \, .$$

Worst-Case Sample-Size Calculations (II)

Define optimization variables:

$$\boldsymbol{\delta} = \left(\begin{array}{ccc} \overline{TD} & \overline{AD} & \overline{ND} \\ \overline{TI} & \overline{AI} & \overline{NI} \\ \overline{TU} & \overline{AU} & \overline{NU} \\ \overline{TP} & \overline{AP} & \overline{NP} \end{array} \right).$$

Expected test statistic can be written as a function of $\boldsymbol{\delta}$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}_{\mathsf{rep}}, \boldsymbol{\pi}_{\mathsf{res}})$:

$$t\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right) = n \cdot \frac{\mu_1\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right) - \mu_0\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right)}{\mu_1\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right) \cdot \left(1 - \mu_1\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right)\right) + \mu_0\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right) \cdot \left(1 - \mu_0\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right)\right)}.$$

where

$$\mu_1\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right) = \overline{I} + \overline{P} - \overline{NI} - \overline{NP} + \overline{AD} + \overline{AU},$$
$$\mu_0\left(\boldsymbol{\pi}, \boldsymbol{\delta}\right) = \overline{D} + \overline{P} - \overline{ND} - \overline{NP} + \overline{AI} + \overline{AU}.$$