Testing Attrition Bias in Field Experiments

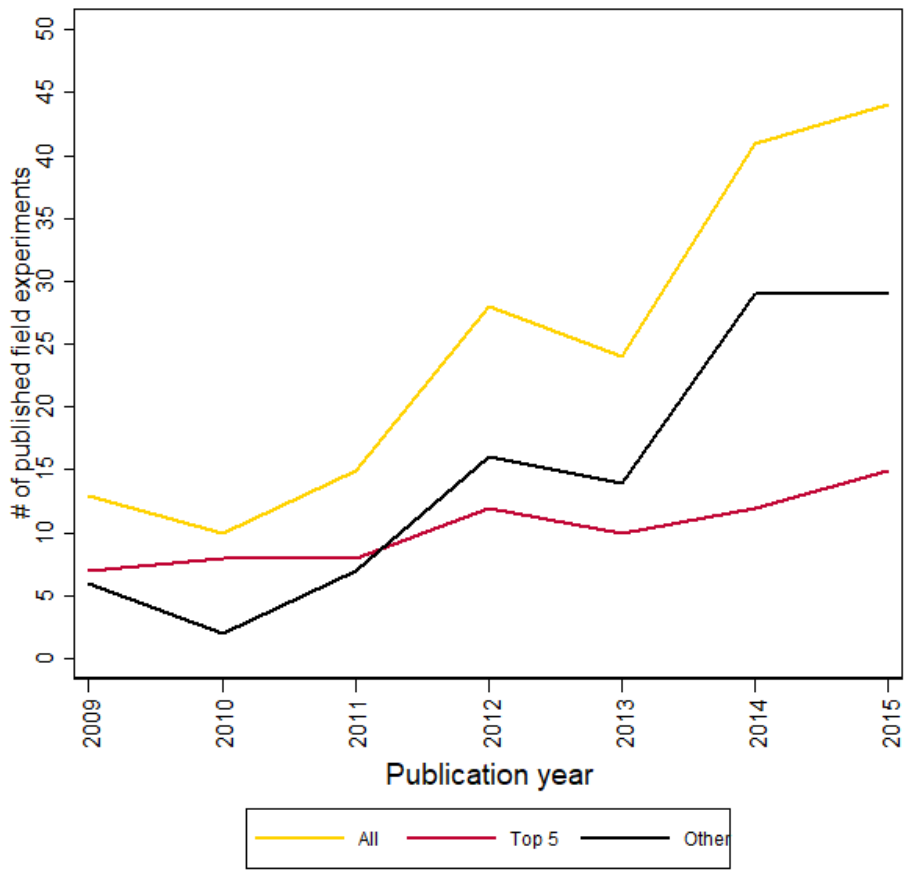
Dalia Ghanem, *UC Davis*
Sarojini R. Hirshleifer, *UC Riverside*
Karen Ortiz-Becerra, *University of San Diego*

2022 Stata Virtual Symposium

Motivation

Randomized control trials (RCTs) are increasingly important in economics

Published Field Experiments (2009-2015)



- 69% increase in top 5 + AEJ: Applied, EJ, JDE, JHR, REStat

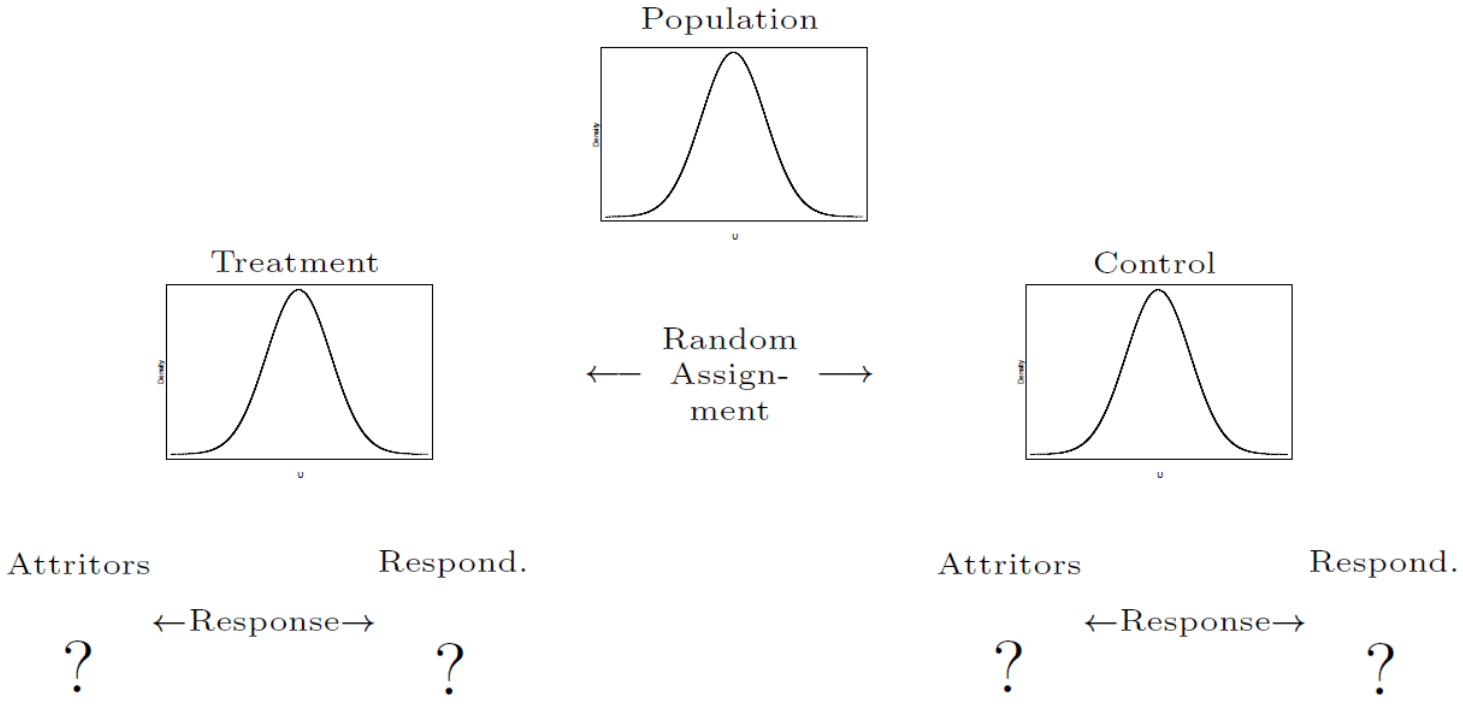
Motivation

- The appeal of field experiments is that they can generate internally valid estimates (when properly designed and implemented)
- Non-response on outcome measures (i.e. *attrition*), however, is a threat to that internal validity

- Researchers do their best to find everyone at follow-up, but:
  1. People migrate to locations that are out of reach for data collection
  2. Conflict, natural disasters and intimidation can make conditions unsafe for enumerators
  3. Ethical considerations associated with multiple attempts to ask for sensitive information (i.e. health outcomes)

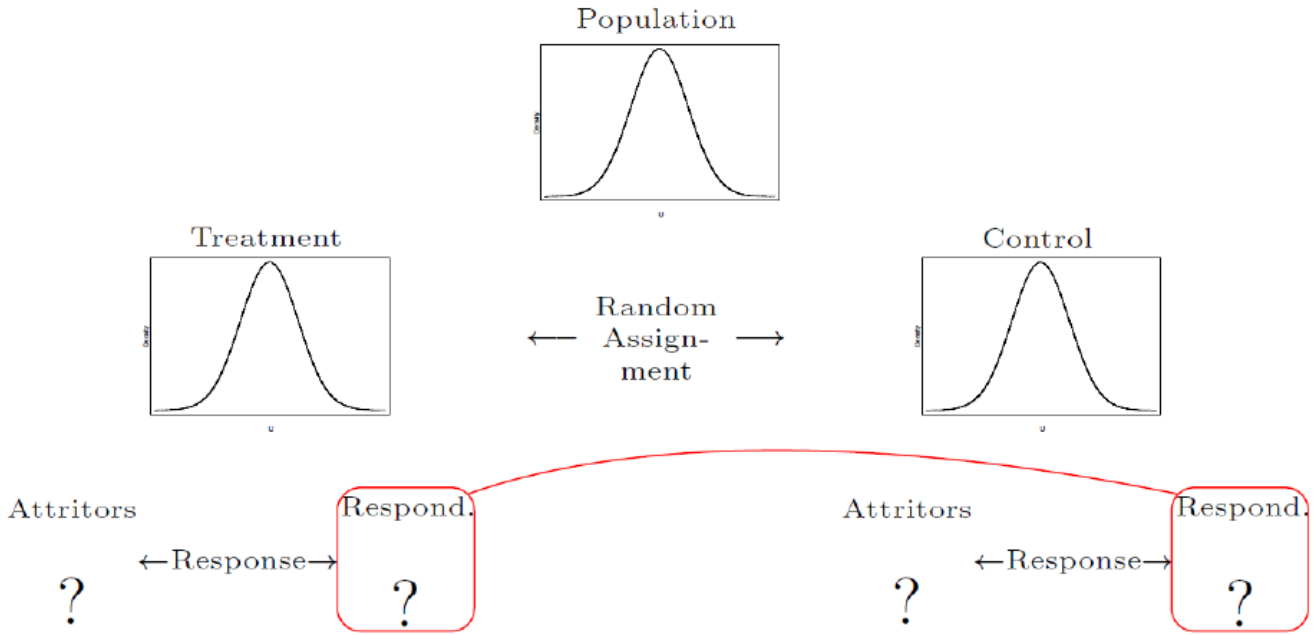
Big Picture: Internal Validity and RCTs

Suppose the outcome of interest $Y = \mu(D, U)$, where D denotes treatment status and U are determinants of the outcome/unobservables.



Big Picture: Internal Validity and RCTs

Suppose the outcome of interest $Y = \mu(D, U)$, where D denotes treatment status and U are determinants of the outcome/unobservables.



- Non-response raises the following question: are comparisons between treatment and control respondents internally valid?
- If yes, what (sub)population are they internally valid for?

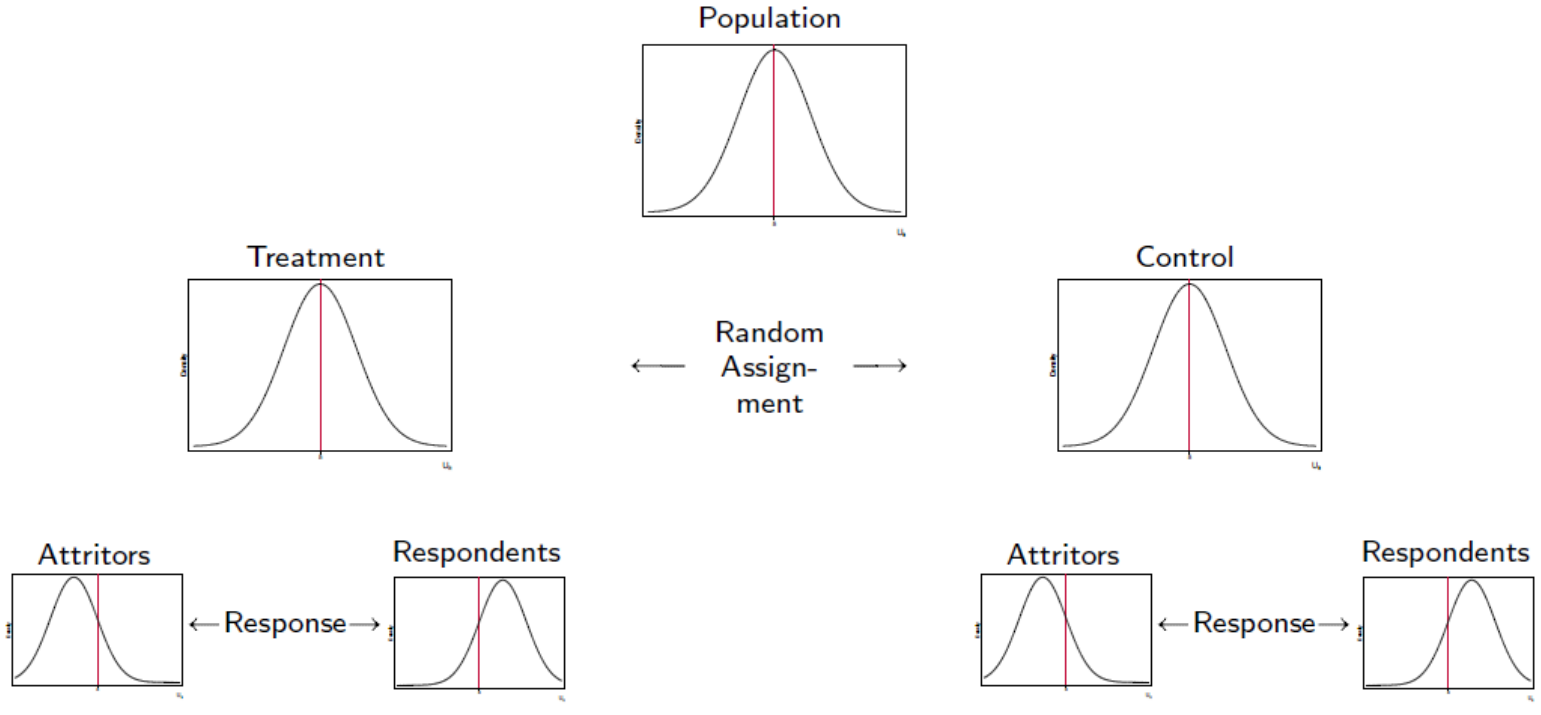
Two Main Types of Internal Validity

- **Internal validity for the respondents (IV-R):** allows to identify treatment effects for subpopulation of individuals with outcome data at endline

- **Internal validity for the study population (IV-P):** allows to estimate treatment effects for all individuals in the study
  - Particularly relevant when study population is representative of a larger population of interest

Big Picture: Key identification question I

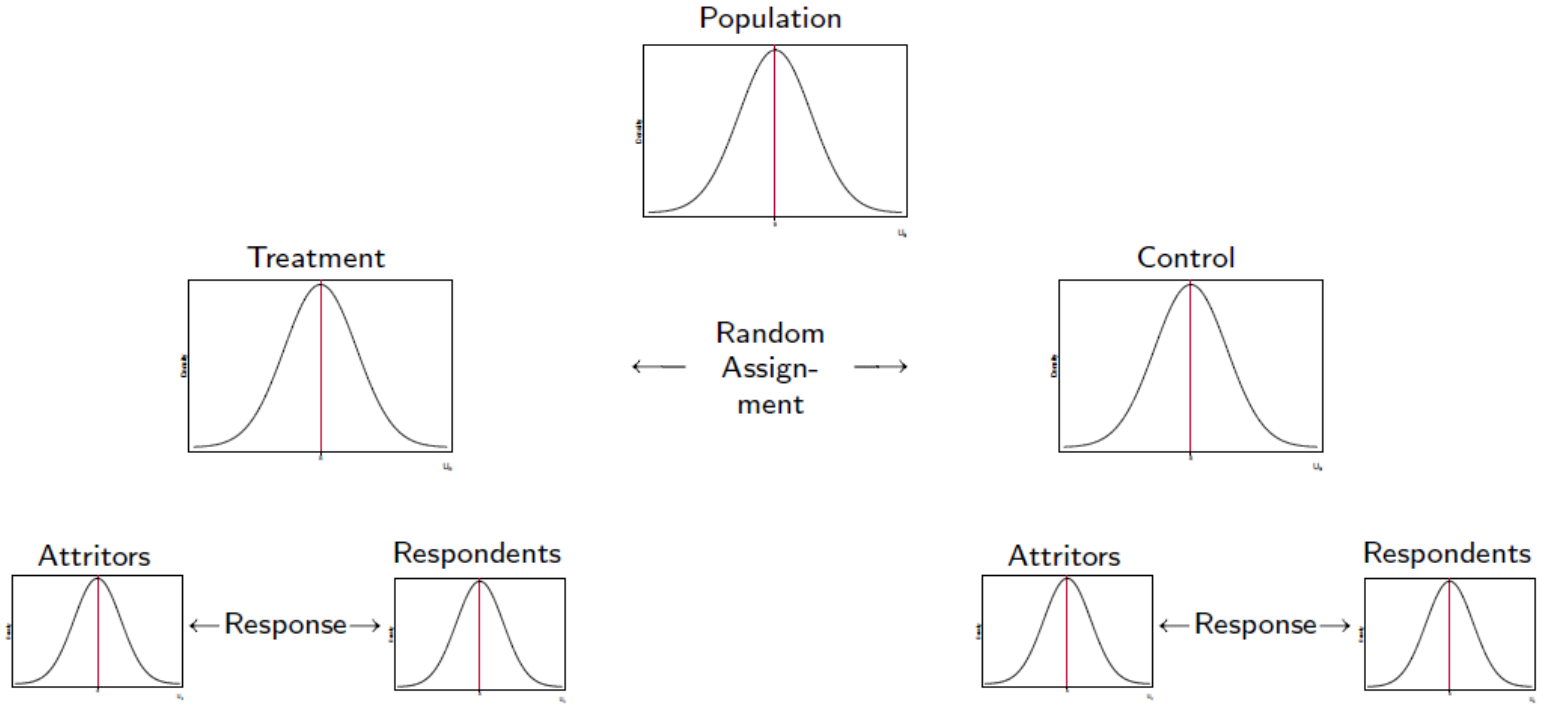
Do we still have *internal validity for the respondents* (IV-R)?



- If yes, we can identify the ATE-R (i.e. the average treatment effect for respondents)

Big Picture: Key identification question II

Do we still have *internal validity for the study population* (IV-P) ?



- If yes, we can identify the ATE (i.e. the average treatment effect for the study population)

Contributions

Document how authors test for attrition bias in field experiments:

- Systematic review shows that attrition rates and tests are common
- Many tests use baseline data, but there is no consensus on how to test

Contributions

Document how authors test for attrition bias in field experiments:

- Systematic review shows that attrition rates and tests are common
- Many tests use baseline data, but there is no consensus on how to test

Use panel identification framework to propose attrition tests:

- Establish identifying assumptions that ensure IV-R and IV-P
- Derive sharp testable restrictions on the *baseline outcome* distribution
- Propose testing procedures and provide *Stata* code to conduct tests

Contributions

Document how authors test for attrition bias in field experiments:

- Systematic review shows that attrition rates and tests are common
- Many tests use baseline data, but there is no consensus on how to test

Use panel identification framework to propose attrition tests:

- Establish identifying assumptions that ensure IV-R and IV-P
- Derive sharp testable restrictions on the *baseline outcome* distribution
- Propose testing procedures and provide *Stata* code to conduct tests

Provide insights on empirical practice using framework, simulations, and applications:

- Most commonly used test in the literature is not an appropriate test of internal validity in general
- Appropriate test depends on the object of interest

Today's Talk

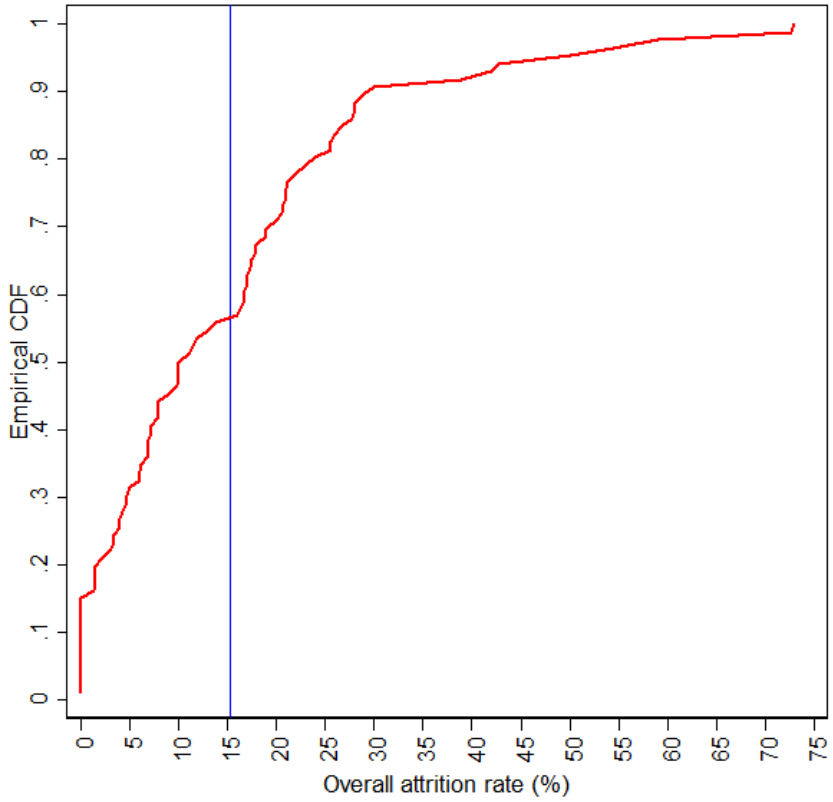
- Review of field experiment literature
- Testing internal validity using baseline outcome data
- Illustrate implementation of tests using Stata command *attregtest*
- Implications and recommendations for empirical practice

Review of field experiment literature

The sample: 93 papers published in ten respected general interest and applied journals during 2009-2015

- Field experiments with baseline data on at least one main outcome

Attrition is common in published field experiments



Review of field experiment literature

Most field experiments conduct at least one test (92%)

Review of field experiment literature

Most field experiments conduct at least one test (92%)

Two most common attrition tests:

- *Differential attrition rates:* determines if attrition rates are different across treatment and control groups
- *Selective attrition:* determines if the mean of baseline outcomes differs across treatment and control groups conditional on response status.

Review of field experiment literature

Most field experiments conduct at least one test (92%)

Two most common attrition tests:

- *Differential attrition rates:* determines if attrition rates are different across treatment and control groups
- *Selective attrition:* determines if the mean of baseline outcomes differs across treatment and control groups conditional on response status.

There is no consensus on which test to conduct:

Table: Distribution of field experiments by attrition test

Proportion of field experiments that conduct:		Selective attrition test		
		No	*Yes*	*Total*
Differential attrition rate test	*No*	10%	10%	21%
	Yes	29%	50%	79%
	Total	39%	61%	100%

Framework for RCTs with attrition

Panel identification framework with baseline outcome data

Outcome equation: $Y_{it} = \mu_t(D_{it}, U_{it})$

- D_{it} is treatment status of individual i in period t
- $t = 0$ at baseline and $t = 1$ at follow-up
- $T_i = 1$ if individual i is in treatment group, 0 otherwise

Response equation: $R_i = \xi(T_i, V_i)$

By random assignment $(U_{i0}, U_{i1}, V_i) \perp T_i$

Two Main Identification Questions

1. Do we have internal validity for the respondents (IV-R)?

$$\underbrace{Y_{i1} | T_i = 0, R_i = 1}_{\text{CR Observed Outcome}} \quad \stackrel{d}{=} \quad \underbrace{Y_{i1}(0) | T_i = 1, R_i = 1}_{\text{TR Counterfactual}}$$

2. Do we have internal validity for the study population (IV-P)?

   For $\tau = 0, 1$

$$\underbrace{Y_{i1} | T_i = \tau, R_i = 1}_{\text{TR/CR Observed Outcome}} \quad \stackrel{d}{=} \quad \underbrace{Y_{i1}(\tau)}_{\text{Study's Population Potential Outcome}}$$

$\Rightarrow$ **Approach:** start with assumptions that ensure IV-R and IV-P and derive the sharp testable restrictions on the baseline outcome distribution

Testable restriction of IV-R

Assume $(U_{i0}, U_{i1}) \perp T_i | R_i$ (IV-R Assumption)

1. (Identification) $Y_{i1} | T_i = 0, R_i = 1 \stackrel{d}{=} Y_{i1}(0) | T_i = 1, R_i = 1$
2. (Sharp Testable Restriction) $Y_{i0} | T_i = 0, R_i = r \stackrel{d}{=} Y_{i0} | T_i = 1, R_i = r$ for $r = 0, 1$.

Intuition:

- IV-R assumption: random assignment conditional on response
- Identification: control respondents are a good counterfactual for treatment respondents at follow-up
- Restriction: baseline outcome distribution is independent of treatment conditional on response status

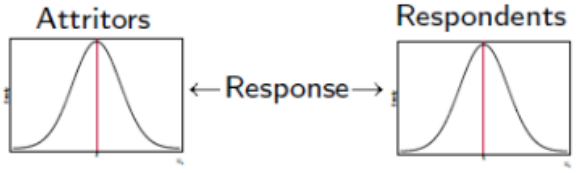
Testable restriction of IV-P

Assume $(U_{i0}, U_{i1}, V_i) \perp T_i + (U_{i0}, U_{i1}) \perp R_i | T_i$ (IV-P Assumption)

1. (Identification) $Y_{i1} | T_i = \tau, R_i = 1 \stackrel{d}{=} Y_{i1}(\tau)$ for $\tau = 0, 1$
2. (Sharp Testable Restriction) $Y_{i0} | T_i = \tau, R_i = r \stackrel{d}{=} Y_{i0}$ for $r = 0, 1$, $\tau = 0, 1$

Intuition:

- IV-P assumption: initial random assign. + independence of U and R conditional on T
- Identification: treatment and control respondents at follow-up identify the potential outcome distribution for the study population
- Restriction: baseline outcome distribution is independent of both treatment and response status



Empirical illustration of our tests: *Progresa*

Program: cash to eligible households conditional on children's school attendance

Study population: census of eligible households in 506 localities: 320 T, 186 C

Table: Difference in Mean Outcomes at First Follow-Up & Attrition Rates

Outcome	Baseline N	Attrition Rate	Diff. in Mean Outcomes for Respondents	
			T - C	P-val
School attendance	24,353	14.2%	0.043	0.00
Employment last week	31,237	9.6%	0.016	0.02

Empirical illustration of our tests: *Progresa*

Program: cash to eligible households conditional on children's school attendance

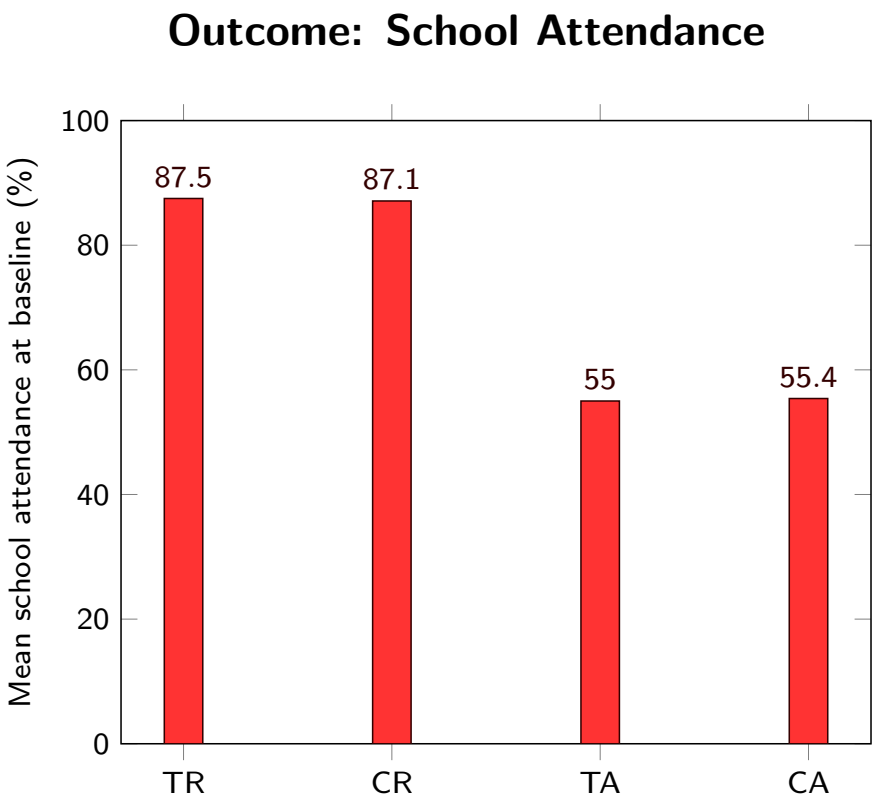
Study population: census of eligible households in 506 localities: 320 T, 186 C

Table: Difference in Mean Outcomes at First Follow-Up & Attrition Rates

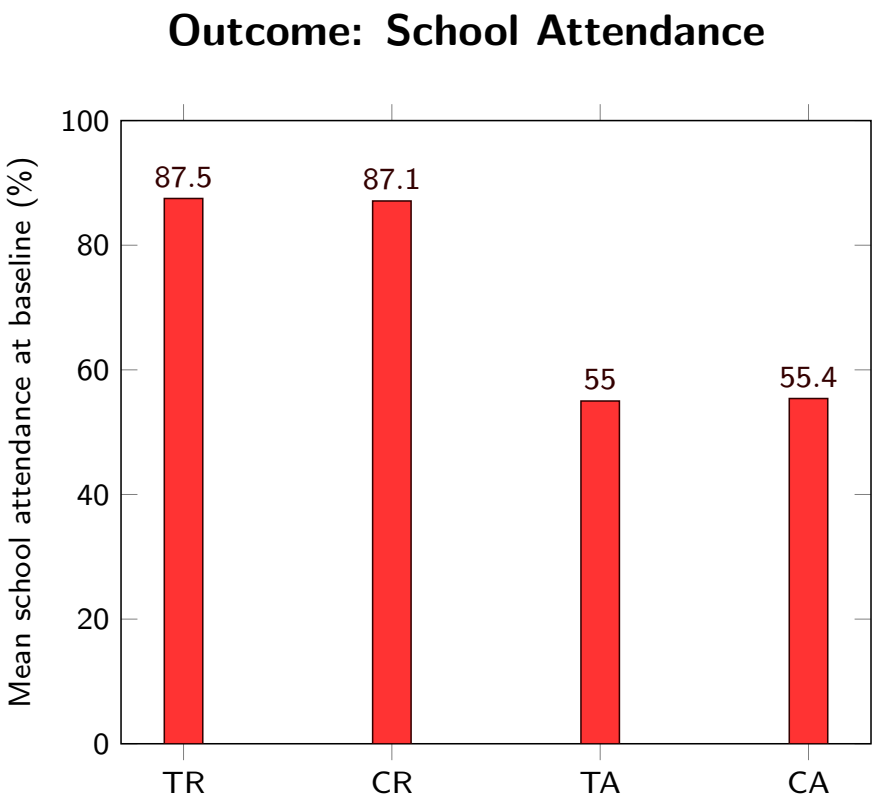
Outcome	Baseline N	Attrition Rate	Diff. in Mean Outcomes for Respondents	
			T - C	P-val
School attendance	24,353	14.2%	0.043	0.00
Employment last week	31,237	9.6%	0.016	0.02

- Do these estimated differences identify the causal impact of *Progresa*?
- Are they estimating the average treatment effect for the whole study population?
- Or, are they only estimating the average treatment effect for the respondents?

Empirical illustration of our tests: *Progresa*

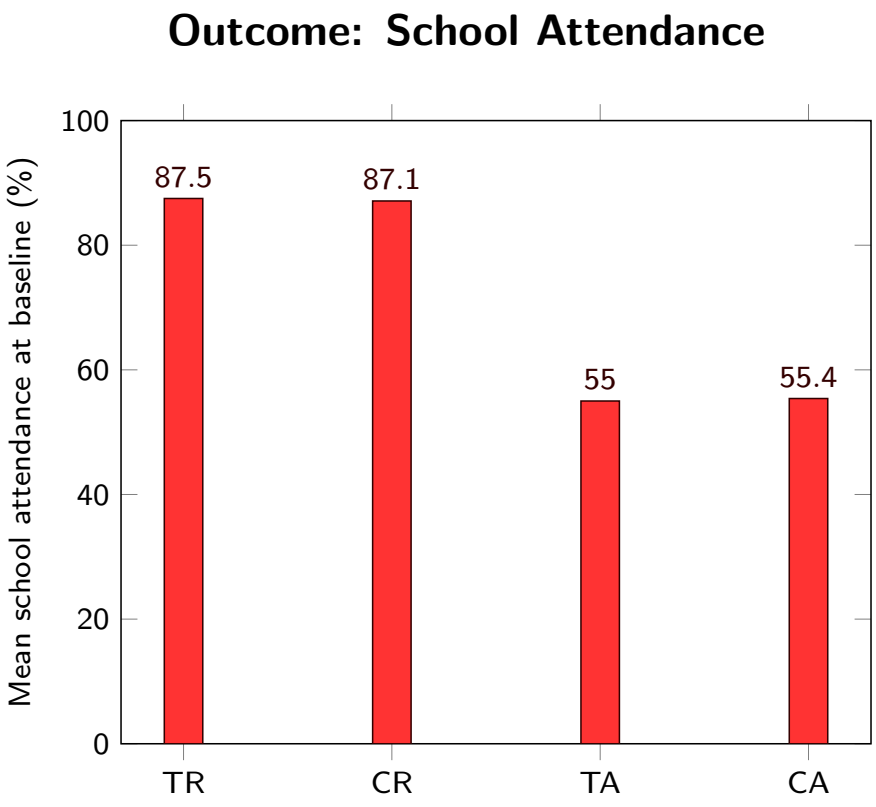


Empirical illustration of our tests: *Progresa*



Null of IV-P: reject (p-value=0.00)

Empirical illustration of our tests: *Progresa*



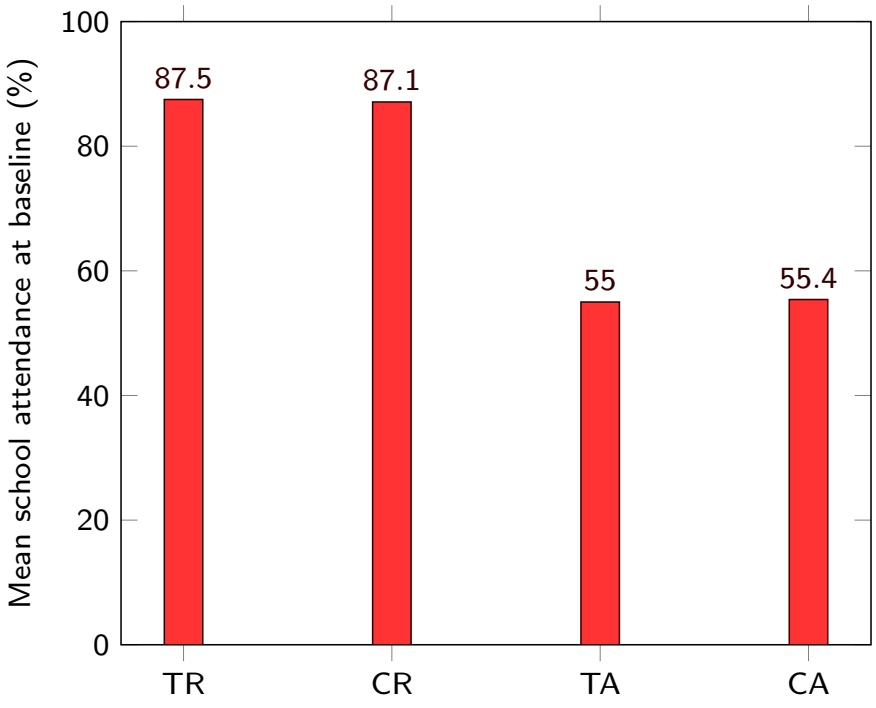
Null of IV-P: reject (p-value=0.00)

Null of IV-R: can't reject (p-value=0.81)

Can't reject identification of ATE-R

Empirical illustration of our tests: *Progresa*

Outcome: School Attendance

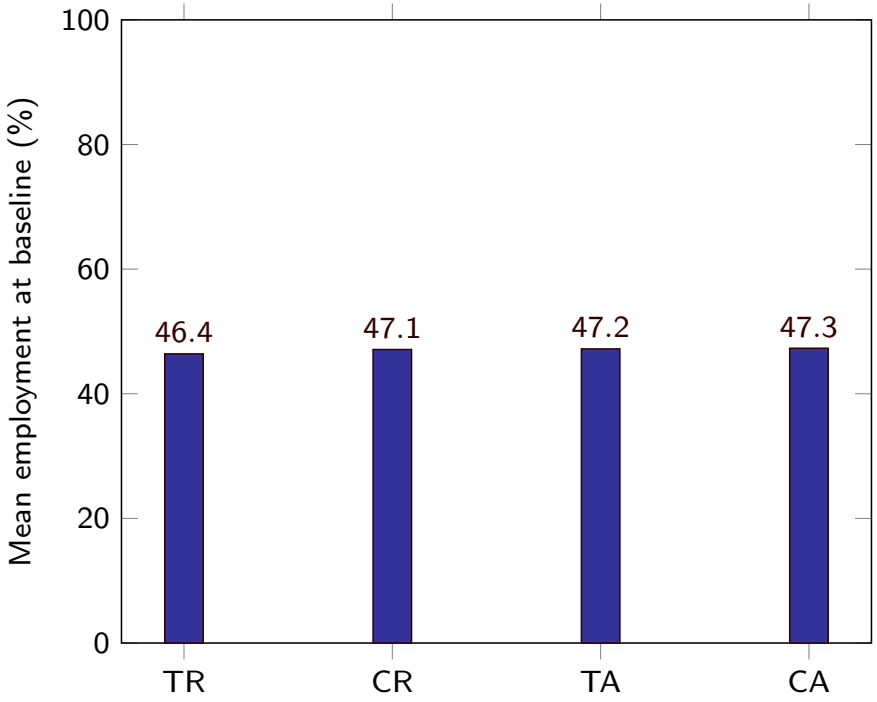


Null of IV-P: reject (p-value=0.00)

Null of IV-R: can't reject (p-value=0.81)

Can't reject identification of ATE-R

Outcome: Adult Employment



Null of IV-P: can't reject (p-value=0.86)

Null of IV-R: can't reject (p-value=0.83)

Can't reject identification of ATE

Test Implementation

- The distributional restrictions of IV-R and IV-P can be tested via randomization procedures ▸ Randomization Tests

Test Implementation

- The distributional restrictions of IV-R and IV-P can be tested via randomization procedures ▸ Randomization Tests
- The mean versions of the IV-R and IV-P tests are linear restrictions on regression coefficients

For completely randomized experiments:

$$Y_{i0} = \gamma_{11} T_i R_i + \gamma_{01}(1 - T_i) R_i + \gamma_{10} T_i (1 - R_i) + \gamma_{00}(1 - T_i)(1 - R_i) + \epsilon_i$$

$$H^1_{0,\mathcal{M}} : \gamma_{11} = \gamma_{01} \ \& \ \gamma_{10} = \gamma_{00},$$
$$H^2_{0,\mathcal{M}} : \gamma_{11} = \gamma_{01} = \gamma_{10} = \gamma_{00}.$$

Test Implementation

- The distributional restrictions of IV-R and IV-P can be tested via randomization procedures ▸ Randomization Tests
- The mean versions of the IV-R and IV-P tests are linear restrictions on regression coefficients

  For completely randomized experiments:

$$Y_{i0} = \gamma_{11} T_i R_i + \gamma_{01}(1 - T_i) R_i + \gamma_{10} T_i (1 - R_i) + \gamma_{00}(1 - T_i)(1 - R_i) + \epsilon_i$$

$$H^1_{0,\mathcal{M}} : \gamma_{11} = \gamma_{01} \ \& \ \gamma_{10} = \gamma_{00},$$
$$H^2_{0,\mathcal{M}} : \gamma_{11} = \gamma_{01} = \gamma_{10} = \gamma_{00}.$$

- Extension to stratified randomization/heterogeneous treatment effects in the paper

ATTREGTEST

Stata package that implements the regression-based attrition tests:

attregtest *baseline_y$_1$ baseline_y$_2$ ... [if] [in] , treatvar(varname) respvars(varlist) [options]*

ATTREGTEST

Stata package that implements the regression-based attrition tests:

attregtest *baseline_y_1 baseline_y_2 ... [if] [in] , treatvar(varname) respvars(varlist) [options]*

Main features:

- Field experiments with clustered or stratified designs
- Accommodates multiple treatment case
- Allows to pool waves in multiple follow-up case
- Can conduct test on multiple baseline outcomes at once
- Can export results as excel or tex file
- Help file with several examples

attregtest

Stata package that implements the regression-based attrition tests:

attregtest *baseline_y_1 baseline_y_2 ... [if] [in] , treatvar(varname) respvars(varlist) [options]*

Main features:

- Field experiments with clustered or stratified designs
- Accommodates multiple treatment case
- Allows to pool waves in multiple follow-up case
- Can conduct test on multiple baseline outcomes at once
- Can export results as excel or tex file
- Help file with several examples

Installation

ssc install attregtest, replace

Example - Progresa

- **Program:** cash conditional on children's school attendance
- **Focus:** first follow-up after 5 months

Table: Difference in Mean Outcomes at First Follow-Up & Attrition Rates

Outcome	Baseline N	Attrition Rate	Diff. in Mean Outcomes for Respondents	
			T - C	P-val
School attendance	24,353	14.2%	0.043	0.00
Employment last week	31,237	9.6%	0.016	0.02

Let:

- y_1 be children's school attendance
- y_2 be adult's employment last week

Example

```
. attregtest y1_base y2_base if fwup == 1 , treatvar(treat)
> respvars(resp_y1 resp_y2) vce(cluster id_locality2)
```

	(1) y1_base b/se	(2) y2_base b/se
treat1Xresponse	0.008	-0.005
	(0.021)	(0.019)
treat1	-0.004	-0.002
	(0.023)	(0.018)
response	0.317	-0.002
	(0.014)	(0.014)
Control Attritors(+)	0.554	0.473
Test of IV-R (p-val)	0.810	0.825
Test of IV-P (p-val)	0.000	0.860
N	24353	31237

(+) Mean baseline outcome control attritors
Standard errors are vce(cluster id_locality2)

- Report p-values of both tests and coefficients to calculate mean baseline outcome by treatment/response subgroup

Additional results in the paper

- Considerations when testing identifying assumptions by implication
  - Conditions under which baseline data can and cannot detect attrition bias in the follow-up ▸ Attrition tests as identification tests

Additional results in the paper

- Considerations when testing identifying assumptions by implication
  - Conditions under which baseline data can and cannot detect attrition bias in the follow-up ▸ Attrition tests as identification tests
- Role of baseline covariates in attrition tests ▸ Covariates
  - When, which, and how covariates should be included

Additional results in the paper

- Considerations when testing identifying assumptions by implication
  - Conditions under which baseline data can and cannot detect attrition bias in the follow-up ▸ Attrition tests as identification tests
- Role of baseline covariates in attrition tests ▸ Covariates
  - When, which, and how covariates should be included
- Extensive simulation study analyzing performance of proposed tests ▸ Simulations

Additional results in the paper

- Considerations when testing identifying assumptions by implication
  - Conditions under which baseline data can and cannot detect attrition bias in the follow-up ▸ Attrition tests as identification tests
- Role of baseline covariates in attrition tests ▸ Covariates
  - When, which, and how covariates should be included
- Extensive simulation study analyzing performance of proposed tests ▸ Simulations
- Empirical applications demonstrating empirical relevance of our results (26 outcomes across 4 papers) ▸ Applications

Implications for current empirical practice

Differential attrition rate test is not a valid test of internal validity:

- Provide theoretical conditions under which this test does not control size as test of IV-R ▸ Example 1
- Find many empirical examples that are consistent with this result (8/26 outcomes in applications)

Implications for current empirical practice

Differential attrition rate test is not a valid test of internal validity:

- Provide theoretical conditions under which this test does not control size as test of IV-R ▸ Example 1
- Find many empirical examples that are consistent with this result (8/26 outcomes in applications)

Most implementations of the selective attrition test constitute IV-R tests:

- A majority test an implication of the tests that we propose and do not use all available information in the baseline sample ▸ Sharp Test

Implications for current empirical practice

Differential attrition rate test is not a valid test of internal validity:

- Provide theoretical conditions under which this test does not control size as test of IV-R ▸ Example 1
- Find many empirical examples that are consistent with this result (8/26 outcomes in applications)

Most implementations of the selective attrition test constitute IV-R tests:

- A majority test an implication of the tests that we propose and do not use all available information in the baseline sample ▸ Sharp Test

Our applications provide promising results for field experiments where the study population is of interest:

- We can't reject IV-P for a surprisingly large proportion of outcomes (21/26)

Recommendations for empirical practice

1. When conducting attrition tests, authors should be explicit about the population of interest (IV-R/IV-P)

Recommendations for empirical practice

1. When conducting attrition tests, authors should be explicit about the population of interest (IV-R/IV-P)
2. Use outcome-specific approach to conduct attrition tests

Recommendations for empirical practice

1. When conducting attrition tests, authors should be explicit about the population of interest (IV-R/IV-P)
2. Use outcome-specific approach to conduct attrition tests
3. Differential attrition rate test should not be used as an IV-R test, as it may over-reject IV-R in practice

Recommendations for empirical practice

1. When conducting attrition tests, authors should be explicit about the population of interest (IV-R/IV-P)
2. Use outcome-specific approach to conduct attrition tests
3. Differential attrition rate test should not be used as an IV-R test, as it may over-reject IV-R in practice
4. Interpreting non-rejections – important to consider if relationship b/w outcome and determinants may have changed b/w baseline and follow-up

Recommendations for empirical practice

1. When conducting attrition tests, authors should be explicit about the population of interest (IV-R/IV-P)
2. Use outcome-specific approach to conduct attrition tests
3. Differential attrition rate test should not be used as an IV-R test, as it may over-reject IV-R in practice
4. Interpreting non-rejections – important to consider if relationship b/w outcome and determinants may have changed b/w baseline and follow-up
5. Proposed tests can aid researchers in interpreting their treatment effect estimates, but they should not be used as pre-tests to decide whether a correction is warranted or not

Summary and directions for future work

- This paper provides a formal treatment of the question of how to test for attrition bias in field experiments using baseline outcome data
- This formal analysis sheds light on current empirical practice
- Empirical applications support the empirical relevance of our results

Work in progress:

- Attrition corrections using baseline data: "Correcting Attrition Bias using Changes-in-Changes"

Thank you!

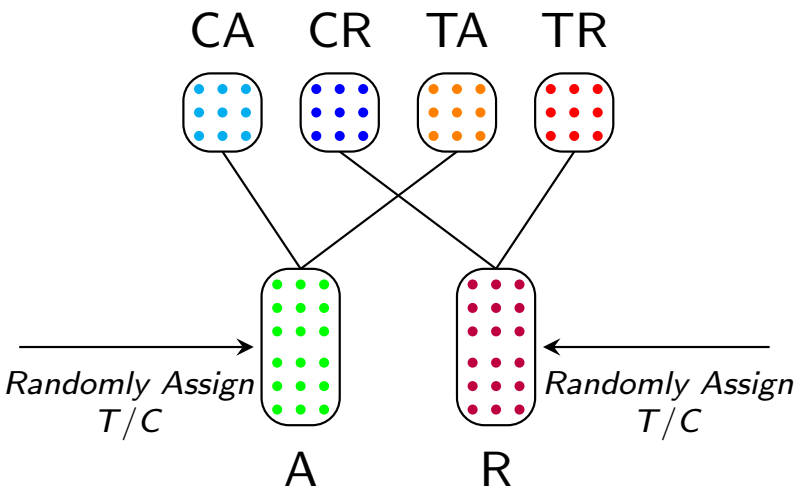
Questions or suggestions?

kortizbecerra@sandiego.edu

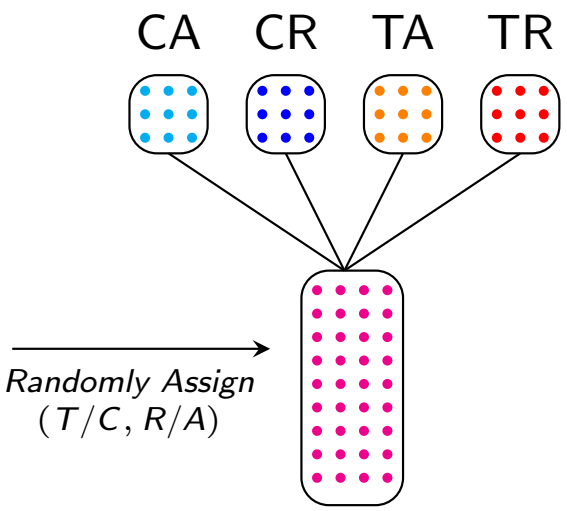
Scan for paper:

Sub-group randomization procedure

Panel A. Permutation for test of IV-R



Panel B. Permutation for test of IV-P



Specific Procedure:

1. For each permutation, compute the *joint* (distributional) statistic
2. Do this B times
3. Compute the p-value

back

Regression tests

Completely Randomized Experiments

$$Y_{i0} = \gamma_{11} T_i R_i + \gamma_{01}(1 - T_i) R_i + \gamma_{10} T_i (1 - R_i) + \gamma_{00}(1 - T_i)(1 - R_i) + \epsilon_i$$

$$H^1_{0,\mathcal{M}} : \gamma_{11} = \gamma_{01} \ \& \ \gamma_{10} = \gamma_{00},$$
$$H^2_{0,\mathcal{M}} : \gamma_{11} = \gamma_{01} = \gamma_{10} = \gamma_{00}.$$

Stratified Randomized Experiments

$$Y_{i0} = \sum_{s \in \mathcal{S}} \left[\gamma^s_{11} T_i R_i + \gamma^s_{10} T_i (1 - R_i) + \gamma^s_{01}(1 - T_i) R_i + \gamma^s_{00}(1 - T_i)(1 - R_i) \right] 1\{S_i = s\} + \epsilon_i$$

$$H^1_{0,\mathcal{M}} : \gamma^s_{11} = \gamma^s_{01} \ \& \ \gamma^s_{10} = \gamma^s_{00}, \ \text{for } s \in \mathcal{S},$$
$$H^2_{0,\mathcal{M}} : \gamma^s_{11} = \gamma^s_{01} = \gamma^s_{10} = \gamma^s_{00}, \ \text{for } s \in \mathcal{S}.$$

back

Role of covariates in attrition tests

Baseline covariates are only option when:

- Baseline outcome is not observed
- Baseline outcome is degenerate by design (e.g., job training program for unemployed people)

They can also help detect violations of internal validity when relationship b/w outcome and its determinants changes b/w baseline and follow-up

- Long-term follow-ups
- Short-term studies where population is at different points of life-cycle
- Short-term studies where aggregate shocks affect relationship b/w outcome and its determinants

back

Appropriate covariates in attrition tests

Given the outcome of interest $Y_{it} = \mu_t(D_{it}, U_{it})$, suppose that there is *a priori* information that a set of covariates satisfy the following

$$W_{it} = \nu_t(U_{it}) \text{ for } t = 0, 1, \tag{1}$$

then the testable restrictions of the IV-R and IV-P assumptions would be on the distribution of $(Y_{i0}, W_{i0}')'$.

Two types of covariates

1. Covariates that are themselves determinants of the outcome
2. "Proxy" variables determined by the same factors as the outcome

Remarks

- Testable restrictions are conditions on the joint distribution of the baseline outcome and covariates $Z_{i0} = (Y_{i0}, W_{i0}')'$
- Our outcome-specific approach implies that including other variables that do not satisfy (1) could lead to false rejection of IV-R/IV-P

back

Differential attrition rate test

- Most commonly used test (79% of field experiments)

$$H_0^{Diff} : P(R_i = 0 | T_i = 0) = P(R_i = 0 | T_i = 1)$$

- To understand relationship between differential attrition rates and the IV-R, we apply the LATE framework of potential compliance to potential response:

	Never-responders	*Control-only responders*	*Treatment-only responders*	*Always-responders*
$R_i(0), R_i(1)$	(0,0)	(1,0)	(0,1)	(1,1)
$P((R_i(0), R_i(1)))$	p_{00}	p_{10}	p_{01}	p_{11}

Under random assigment, $\underbrace{P(R_i = 0 | T_i = 0)}_{p_{00}+p_{01}} - \underbrace{P(R_i = 0 | T_i = 1)}_{p_{00}+p_{10}} = p_{01} - p_{10}$.

back

Results on differential attrition rate test

Summary:

- Equal attrition rates alone do not imply IV-R
  - The differential attrition rate test requires an additional assumption: *monotonicity* (treatment-only responders or control-only responders but not both)
- We demonstrate this through formal examples:
  - **Example 1**: Differential attrition rates, *and* internal validity
  - **Example 2**: Equal attrition rates, but *no* internal validity
- Even with the assumption of monotonicity: The differential attrition rate test has *no implication for the IV-P* (only the IV-R)
  - Specifically, the implication is that all respondents are always responders

back

Example 1: Internal validity and differential attrition rates

Assume

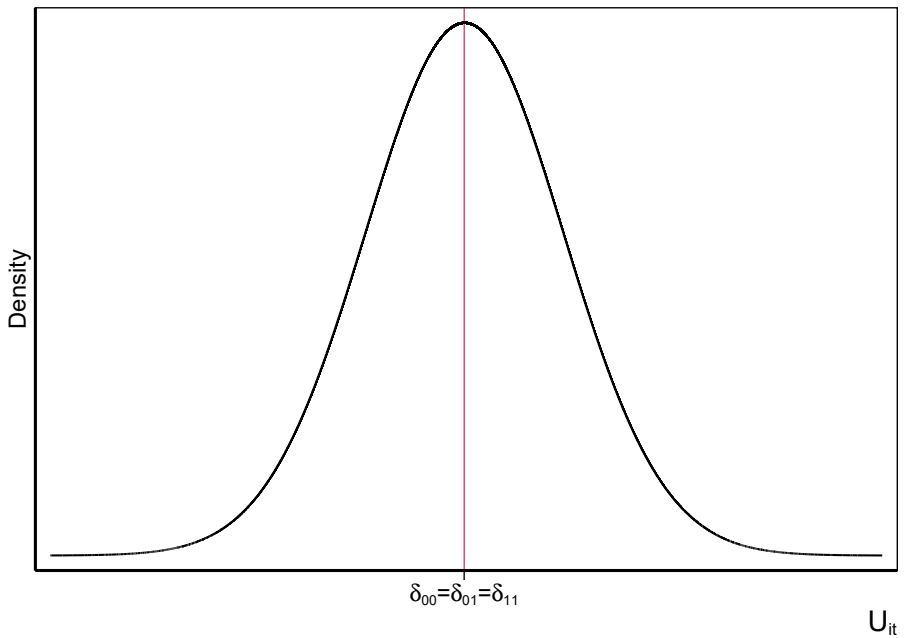
- "Missing-at-random" holds: $(U_{i0}, U_{i1}) \perp (R_i(0), R_i(1))$
- Treatment-only responders: $p_{01} > 0$
- Monotonicity in response holds: $p_{10} = 0$

Then, we have different attrition rates

$P(R_i = 0|T_i = 1) = p_{00} + p_{10}$ ($p_{10} \to 0$)
$P(R_i = 0|T_i = 0) = p_{00} + p_{01}$

Even though, under random assignment of treatment, IV-R and IV-P hold:
$(U_{i0}, U_{i1}) \perp (R_i(0), R_i(1)) \Rightarrow$
$(U_{i0}, U_{i1})|T_i, R_i \stackrel{d}{=} (U_{i0}, U_{i1})$

FIGURE: Distribution of U_{it}



Notes: $\delta_{r_0 r_1}$ for $(r_0, r_1) \in \{0, 1\}^2$ is the mean in each potential response group.

back

Example 2: Equal attrition rates and violation of internal validity

Assume

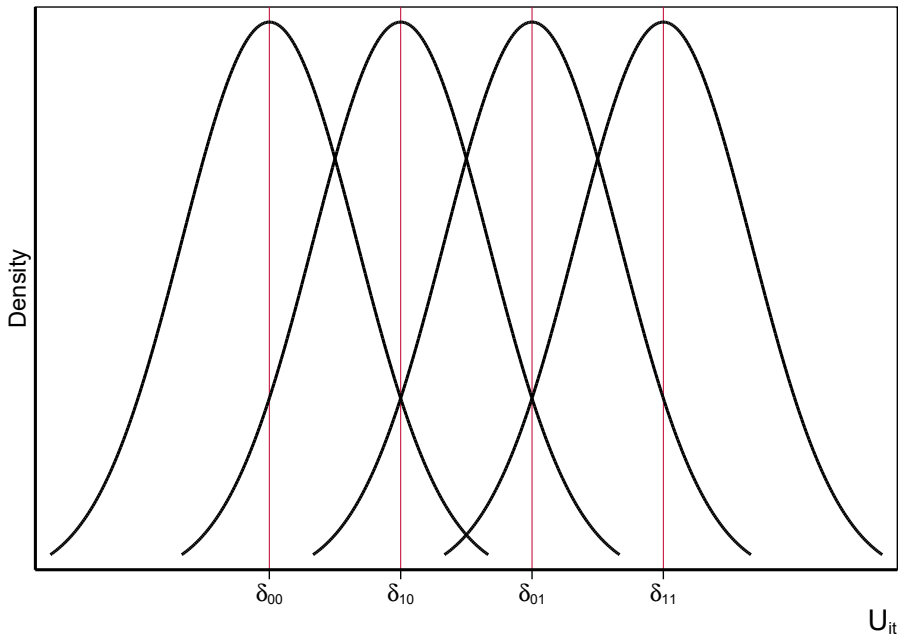
- No "missing-at-random": $(U_{i0}, U_{i1}) \not\perp (R_i(0), R_i(1))$
- Potential response violates monotonicity: $p_{01} > 0$ & $p_{10} > 0$
- Equal proportion of treatment-only and control-only responders: $p_{01} = p_{10}$

Then, we can have equal attrition rates

$P(R_i = 0 | T_i = 1) = p_{00} + p_{10}$
$P(R_i = 0 | T_i = 0) = p_{00} + p_{01}$

Even when IV-R does not hold.

Figure: Distribution of U_{it}

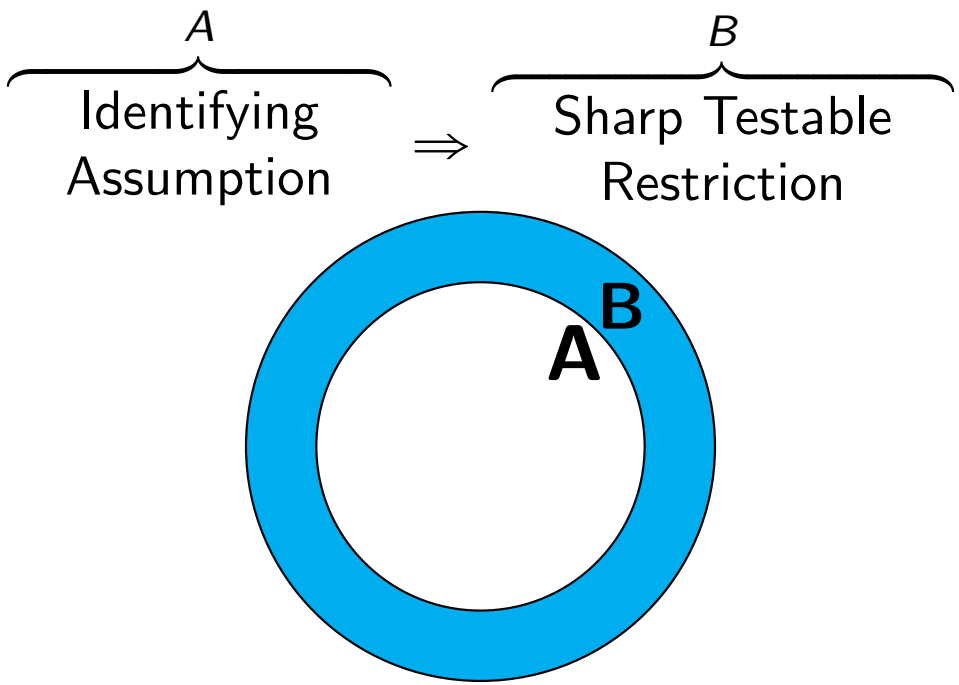


Notes: $\delta_{r_0 r_1}$ for $(r_0, r_1) \in \{0, 1\}^2$ is the mean in each potential response group.

back

Attrition Tests as Identification Tests

Testing identifying assumptions by implication

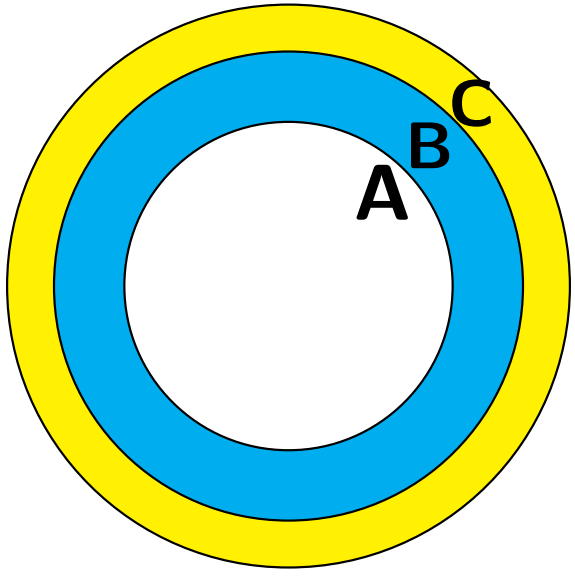


- Blue area: cases where sharp restriction holds, but identifying assump. is violated*
- Time homogeneity of structural function and unobservable distribution (Chernozhukov et al (2010)) rules out cases in the blue area
- Interpreting non-rejections: important to consider if relationship between outcome and its determinants may have changed between baseline and follow-up ▸ back

*The theoretical case where opposite holds is not empirically relevant.

Sharp Testable Restriction

$\overbrace{\text{Identifying Assumption}}^{A} \Rightarrow \overbrace{\text{Sharp Testable Restriction}}^{B} \Rightarrow \overbrace{\text{Implication of B}}^{C}$



- Blue area: cases where sharp restriction holds, but identifying assump. is violated*
- Using sharp testable restrictions avoids the yellow area ($B^c \cap C$) back

*The theoretical case where opposite holds is not empirically relevant.

Simulation design

Treatment Assignment

- Individual observations are randomly assigned to treatment

Outcome Equation

- Treatment effect heterogeneity which allows true ATE and ATE-R to differ
- Time-invariant unobservable *may* depend on potential response

Response Equation

- Assign individuals to one of the four types of compliance behavior according to certain proportions

Sample size and simulation replications

- N=2000, S=2000

back

Variants of the Simulation Design

Variants of DGP

DESIGN	I	II	III
IV-R/IV-P Assumption	Neither	IV-R Only	IV-P
Monotonicity in the Response Equation	Yes ($p_{10} = 0$)	Yes ($p_{10} = 0$)	Yes ($p_{10} = 0$)
Equal Attrition Rates	No	Yes ($p_{01} = 0$)	No
$(U_{i0}, U_{i1}) \perp (R_i(0), R_i(1))$	No	No	Yes

back

Simulation Results

Attrition Rate		Diff Att R Test	Tests of the IV-R Assumption ($\alpha = 0.05$)				Tests of the IV-P Assumption ($\alpha = 0.05$)		Diff. in Mean Y
			Mean Tests			KS Test	Mean Test	KS Test	
C	T	$\hat{p}_{0.05}$	CR-TR	CA-TA	Joint	Joint	Joint	Joint	Mean
Design I									
0.05	0.025	0.866	0.049	0.446	0.353	0.324	0.452	0.476	0.265
0.10	0.05	0.995	0.076	0.719	0.635	0.582	0.792	0.787	0.282
0.20	0.15	0.867	0.072	0.532	0.442	0.412	1.000	1.000	0.296
0.30	0.20	1.000	0.141	0.894	0.851	0.801	1.000	1.000	0.334
Design II†									
0.05	0.05	0.049	0.046	0.044	0.053	0.062	0.981	0.902	0.255
0.10	0.10	0.053	0.043	0.045	0.045	0.056	1.000	0.999	0.262
0.20	0.20	0.049	0.045	0.047	0.050	0.050	1.000	1.000	0.280
0.30	0.30	0.048	0.053	0.044	0.046	0.043	1.000	1.000	0.303
Design III⋆									
0.05	0.025	0.866	0.055	0.051	0.056	0.052	0.065	0.050	0.248
0.10	0.05	0.995	0.055	0.050	0.055	0.046	0.053	0.055	0.248
0.20	0.15	0.867	0.058	0.047	0.053	0.046	0.048	0.048	0.247
0.30	0.20	1.000	0.057	0.053	0.052	0.043	0.049	0.048	0.248

† IV-R only, ⋆ IV-R & IV-P ($n = 2000$, $S = 2000$)

back

Simulation Results

Attrition Rate		Diff Att R Test	Tests of the IV-R Assumption ($\alpha = 0.05$)				Tests of the IV-P Assumption ($\alpha = 0.05$)		Diff. in Mean Y
			Mean Tests			KS Test	Mean Test	KS Test	
C	T	$\hat{p}_{0.05}$	CR-TR	CA-TA	Joint	Joint	Joint	Joint	Mean
Design I									
0.05	0.025	0.866	0.049	0.446	0.353	0.324	0.452	0.476	0.265
0.10	0.05	0.995	0.076	0.719	0.635	0.582	0.792	0.787	0.282
0.20	0.15	0.867	0.072	0.532	0.442	0.412	1.000	1.000	0.296
0.30	0.20	1.000	0.141	0.894	0.851	0.801	1.000	1.000	0.334
Design II†									
0.05	0.05	0.049	0.046	0.044	0.053	0.062	0.981	0.902	0.255
0.10	0.10	0.053	0.043	0.045	0.045	0.056	1.000	0.999	0.262
0.20	0.20	0.049	0.045	0.047	0.050	0.050	1.000	1.000	0.280
0.30	0.30	0.048	0.053	0.044	0.046	0.043	1.000	1.000	0.303
Design III⋆									
0.05	0.025	0.866	0.055	0.051	0.056	0.052	0.065	0.050	0.248
0.10	0.05	0.995	0.055	0.050	0.055	0.046	0.053	0.055	0.248
0.20	0.15	0.867	0.058	0.047	0.053	0.046	0.048	0.048	0.247
0.30	0.20	1.000	0.057	0.053	0.052	0.043	0.049	0.048	0.248

† IV-R only, ⋆ IV-R & IV-P ($n = 2000$, $S = 2000$)

back

Simulation Results

Attrition Rate		Diff Att R Test	Tests of the IV-R Assumption ($\alpha = 0.05$)				Tests of the IV-P Assumption ($\alpha = 0.05$)		Diff. in Mean Y
			Mean Tests			KS Test	Mean Test	KS Test	
C	T	$\hat{p}_{0.05}$	CR-TR	CA-TA	Joint	Joint	Joint	Joint	Mean
Design I									
0.05	0.025	0.866	0.049	0.446	0.353	0.324	0.452	0.476	0.265
0.10	0.05	0.995	0.076	0.719	0.635	0.582	0.792	0.787	0.282
0.20	0.15	0.867	0.072	0.532	0.442	0.412	1.000	1.000	0.296
0.30	0.20	1.000	0.141	0.894	0.851	0.801	1.000	1.000	0.334
Design II†									
0.05	0.05	0.049	0.046	0.044	0.053	0.062	0.981	0.902	0.255
0.10	0.10	0.053	0.043	0.045	0.045	0.056	1.000	0.999	0.262
0.20	0.20	0.049	0.045	0.047	0.050	0.050	1.000	1.000	0.280
0.30	0.30	0.048	0.053	0.044	0.046	0.043	1.000	1.000	0.303
Design III⋆									
0.05	0.025	0.866	0.055	0.051	0.056	0.052	0.065	0.050	0.248
0.10	0.05	0.995	0.055	0.050	0.055	0.046	0.053	0.055	0.248
0.20	0.15	0.867	0.058	0.047	0.053	0.046	0.048	0.048	0.247
0.30	0.20	1.000	0.057	0.053	0.052	0.043	0.049	0.048	0.248

† IV-R only, ⋆ IV-R & IV-P ($n = 2000$, $S = 2000$)

back

Empirical Applications

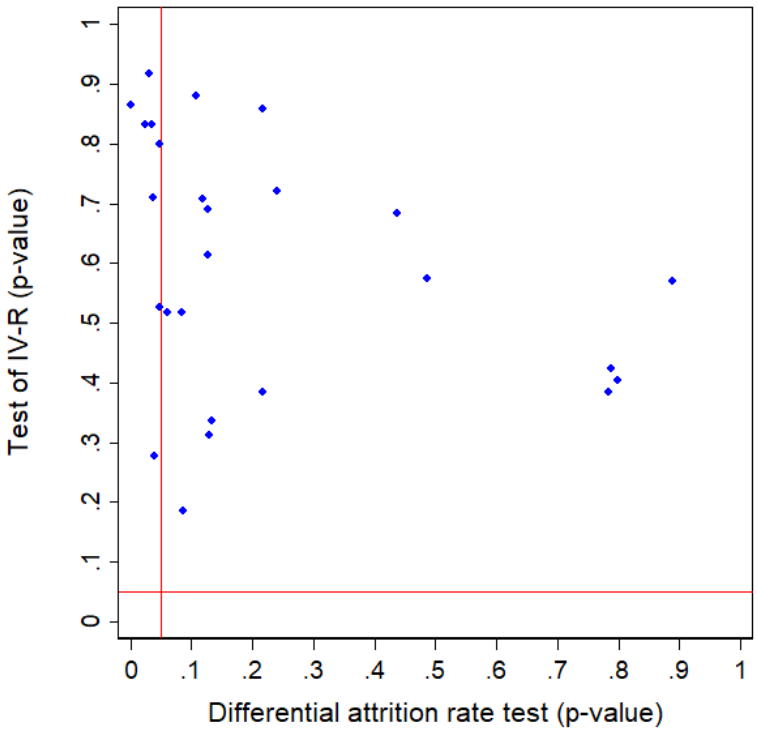
- 26 Outcomes of 4 articles from our review
  - highest attrition rates
  - publicly available data including attritors' baseline data
- Test Statistics
  - differential attrition rate test (outcome-level)
  - IV-R Tests (respondents only, attritors only and joint)
  - IV-P Test (joint)
- Authors' reported tests
  - differential attrition rate test (survey-level)
  - selective attrition test
- No consensus in the tests used or their implementation

back

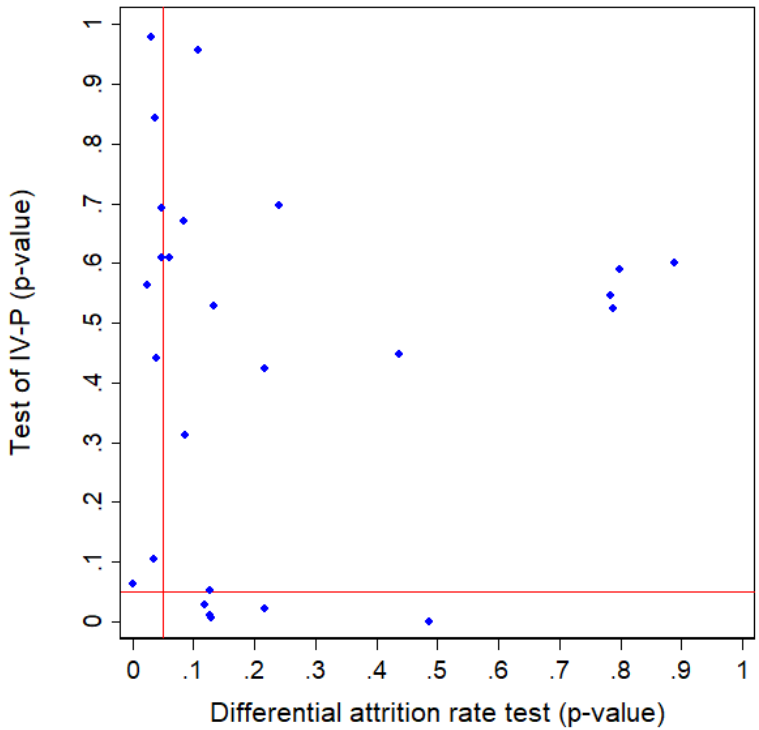
Summary of results $\alpha = 0.05$

Figure: P-values of Attrition Tests in Empirical Applications

Panel A. Tests of IV-R



Panel B.Test of IV-P



- Cannot reject the IV-R assumption for any of the 26 outcomes
- Cannot reject IV-P for 21/26 outcomes
- 8/22 outcomes consistent with the theoretical conditions of *Example 1*

back

Implications

Results underscore the empirical relevance of

- Concerns raised regarding the differential attrition rate test
- Outcome-specific approach to conducting attrition tests
- Internal validity for the study population

back