photo: Daniella Van Leggelo-Padilla / World Bank

Williams College ECON 523:

Program Evaluation for International Development

Lecture 4: Diff-in-Diff in a Panel Data Framework

Professor: Pamela Jakiela

2×2 Diff-in-Diff Specifications

Difference-in-Differences Estimation

To implement diff-in-diff in a regression framework, we estimate:

$$Y_{i,t} = \alpha + \beta D_i + \theta Post_t + \delta \left(D_i * Post_t \right) + \varepsilon_{i,t}$$

Where:

- D_i = treatment dummy
- $Post_i$ = dummy for post-treatment period
- $D_i * Post_t$ = interaction term

Panel data: multiple units, over time

- At least two time periods
- Two treatment groups, possible more units

	treatment	comparison
pre	$\bar{Y}^{treatment}_{pre}$	$\bar{Y}^{comparison}_{pre}$
post	$\bar{Y}^{treatment}_{post}$	$\bar{Y}^{comparison}_{post}$

Difference-in-Differences Estimation: Standard Errors

Only one correct method for calculating cell means and associated standard errors

- Multiple ways to handle the standard errors of the differences in means

Heteroskedastic SEs $\leftrightarrow$ SEs are independent

- Correct but conservative

Homoskedastic SEs $\leftrightarrow$ common variance

- Economists would never!

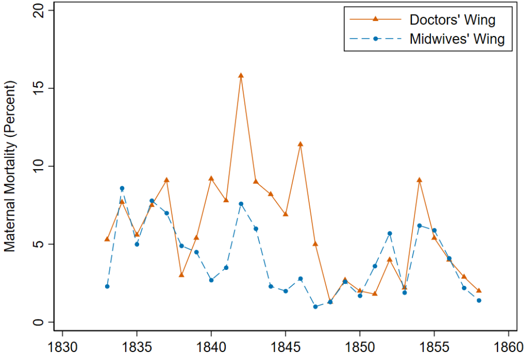
Calculate SE of within-year difference

$\Rightarrow$ Lower variance

	Doctors	Midwives
pre-1847	9.85 (1.34)	4.03 (0.92)
post-1847	3.53 (0.63)	3.13 (0.55)

Difference-in-Differences Estimation: Standard Errors

	Doctors	Midwives
pre-1847	9.85 (1.34)	4.03 (0.92)
post-1847	3.53 (0.63)	3.13 (0.55)



Difference-in-Differences Estimation: Standard Errors

	Doctors	Midwives	Difference
pre-1847	9.85 (1.34)	4.03 (0.92)	5.82 (0.90)
post-1847	3.53 (0.63)	3.13 (0.55)	0.4 (0.48)

$\Rightarrow SE = \sqrt{0.90^2 + 0.48^2} \approx 1.02$

Heteroskedasticity-robust SE:

$$\Rightarrow \sqrt{SE^2_{T,pre} + SE^2_{C,pre} + SE^2_{T,post} + SE^2_{C,post}} = \sqrt{1.34^2 + 0.92^2 + 0.63^2 + 0.55^2} \approx 1.83$$

Using $Y_{T,t=\tau} - Y_{C,t=\tau}$ as the Outcome Variable

Interacted 2×2 diff-in-diff specification equivalent* to regression of $Y_T - Y_C$ on $Post_t$:

$$Y_{T,t=\tau} - Y_{C,t=\tau} = \zeta + \lambda Post_t + \epsilon_{it}$$

where:

- $Y_{T,t=\tau} - Y_{C,t=\tau}$ = treatment vs. comparison difference in outcome
- λ = coefficient of interest (the treatment effect)
- ζ = selection bias (pre-treatment difference between T and C)

* Identical point estimates, different standard errors

Using ΔY_i as the Outcome Variable

Interacted 2×2 diff-in-diff specification also equivalent to first differences (in short panels):

$$Y_{i,t=2} - Y_{i,t=1} = \eta + \gamma D_i + \epsilon_{it}$$

where:

- $Y_{i,t=2} - Y_{i,t=1} =$ change (pre vs. post) in outcome of interest
- $\gamma =$ coefficient of interest (the treatment effect)
- $\eta =$ time trend (average change in comparison group)

Example: Minimum Wages and Employment in the Fast-Food Industry

Interacted 2×2 diff-in-diff specification also equivalent to first differences (in short panels):

$$\Delta \text{FTE}_i = \eta + \gamma \text{NJ}_i + \epsilon_i$$

where:

- ΔFTE_i = change in full-time employment in restaurant i
- γ = difference in mean change in NJ stores (vs. PA stores)
- η = constant (mean change in FTE in PA)

Example: Minimum Wages and Employment in the Fast-Food Industry

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

	Model	
Independent variable	(i)	(ii)
New Jersey dummy	2.33 (1.19)	2.30 (1.20)
Controls for chain and ownership[b]	no	yes
Controls for region[c]	no	no
Standard error of regression	8.79	8.78
Probability value for controls[d]	—	0.34

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are −0.237 and 8.825, respectively. All models include an unrestricted constant (not reported).

[b]Three dummy variables for chain type and whether or not the store is company-owned are included.

[c]Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

[d]Probability value of joint F test for exclusion of all control variables.

source: Card and Krueger (1994)

Diff-in-Diff with Fixed Effects

2×2 Diff-in-Diff in Panel Data

2×2 panel data diff-in-diff specification:

$$Y_{i,t} = \alpha + \beta EverTreated_i + \theta Post_t + \delta D_{i,t} + \varepsilon_{i,t}$$

where:

- $EverTreated_i$ = dummy for ever-treated unit(s)
- $Post_t$ = dummy for post-treatment period(s)
- $D_{i,t}$ = treatment dummy, equal to one if unit i is treated in period t
- δ = diff-in-diff estimate of treatment effect

2×2 Diff-in-Diff in Panel Data

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0	0	0	0	0
Unit 2	0	0	0	0	0
Unit 3	0	0	0	1	1
Unit 4	0	0	0	1	1
Unit 5	0	0	0	1	1

OLS specification:

$$Y = \alpha + \beta EverTreated_i + \theta Post_t + \delta D_{i,t}$$

2×2 Diff-in-Diff in Panel Data

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0	0	0	0	0
Unit 2	0	0	0	0	0
Unit 3	0	0	0	1	1
Unit 4	0	0	0	1	1
Unit 5	0	0	0	1	1
$\bar{D}_t$		0		0.6	

OLS specification:

$Y = \alpha + \beta EverTreated_i + \theta Post_t + \delta D_{i,t}$

By Frisch-Waugh-Lovell:

equivalent to regression on normalized D_{it}

$\rightarrow$ Subtract off mean $D_{i,t}$ in pre, post periods

2×2 Diff-in-Diff in Panel Data

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0	0	0	-0.6	-0.6
Unit 2	0	0	0	-0.6	-0.6
Unit 3	0	0	0	0.4	0.4
Unit 4	0	0	0	0.4	0.4
Unit 5	0	0	0	0.4	0.4
$\bar{D}_t$		0		0.6	

OLS specification:

$Y = \alpha + \beta EverTreated_i + \theta Post_t + \delta D_{i,t}$

By Frisch-Waugh-Lovell:

equivalent to regression on normalized D_{it}

$\rightarrow$ Subtract off mean $D_{i,t}$ in pre, post periods

2×2 Diff-in-Diff in Panel Data

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0.24	0.24	0.24	-0.36	-0.36
Unit 2	0.24	0.24	0.24	-0.36	-0.36
Unit 3	-0.16	-0.16	-0.16	0.24	0.24
Unit 4	-0.16	-0.16	-0.16	0.24	0.24
Unit 5	-0.16	-0.16	-0.16	0.24	0.24
$\bar{D}_t$		0		0.6	

OLS specification:

$Y = \alpha + \beta EverTreated_i + \theta Post_t + \delta D_{i,t}$

By Frisch-Waugh-Lovell:

equivalent to regression on normalized D_{it}

$\rightarrow$ Subtract off mean $D_{i,t}$ in pre, post periods

$\rightarrow$ Subtract off mean of de-meaned $D_{i,t}$ in T, C

Diff-in-Diff with Time Fixed Effects

Panel data diff-in-diff specification including time fixed effects:

$$Y_{i,t} = \alpha + \gamma EverTreated_i + \delta D_{i,t} + \nu_t + \varepsilon_{i,t}$$

where:

- $EverTreated_i$ = dummy for ever-treated unit(s)
- $D_{i,t}$ = treatment dummy, equal to one if unit i is treated in period t
- δ = diff-in-diff estimate of treatment effect
- ν_t = time-period fixed effects

Diff-in-Diff with Time Fixed Effects

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0	0	0	0	0
Unit 2	0	0	0	0	0
Unit 3	0	0	0	1	1
Unit 4	0	0	0	1	1
Unit 5	0	0	0	1	1
$\bar{D}_t$	0	0	0	0.6	0.6

OLS with fixed effects equivalent to a regression of: normalized $\tilde{Y}_{i,t}$ on normalized $\tilde{D}_{i,t}$

To normalize $\tilde{D}_{i,t}$, we

$\rightarrow$ Subtract off period-specific means, $\bar{D}_t$

Diff-in-Diff with Time Fixed Effects

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0	0	0	-0.6	-0.6
Unit 2	0	0	0	-0.6	-0.6
Unit 3	0	0	0	0.4	0.4
Unit 4	0	0	0	0.4	0.4
Unit 5	0	0	0	0.4	0.4
$\bar{D}_t$	0	0	0	0.6	0.6

OLS with fixed effects equivalent to a regression of: normalized $\tilde{Y}_{i,t}$ on normalized $\tilde{D}_{i,t}$

To normalize $\tilde{D}_{i,t}$, we

$\rightarrow$ Subtract off period-specific means, $\bar{D}_t$

Diff-in-Diff with Time Fixed Effects

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	0.24	0.24	0.24	-0.36	-0.36
Unit 2	0.24	0.24	0.24	-0.36	-0.36
Unit 3	-0.16	-0.16	-0.16	0.24	0.24
Unit 4	-0.16	-0.16	-0.16	0.24	0.24
Unit 5	-0.16	-0.16	-0.16	0.24	0.24
$\bar{D}_t$	0	0	0	0.6	0.6

OLS with fixed effects equivalent to a regression of: normalized $\tilde{Y}_{i,t}$ on normalized $\tilde{D}_{i,t}$

To normalize $\tilde{D}_{i,t}$, we

$\rightarrow$ Subtract off period-specific means, $\bar{D}_t$

$\rightarrow$ Subtract off mean of de-meaned $D_{i,t}$ in T, C

Diff-in-Diff with Time Fixed Effects

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
Unit 1	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$
Unit 2	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$
Unit 3	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$
Unit 4	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$
Unit 5	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$	$\tilde{Y}_{i,t}$

OLS with fixed effects equivalent to a regression of: normalized $\tilde{Y}_{i,t}$ on normalized $\tilde{D}_{i,t}$

To normalize $\tilde{D}_{i,t}$, we

$\rightarrow$ Subtract off period-specific means, $\bar{D}_t$

$\rightarrow$ Subtract off mean of de-meaned $D_{i,t}$ in T, C

Fixed effects absorb additional variation in Y

- Standard errors depend on the residuals

Diff-in-Diff with Fixed Effects

Why used fixed effects instead of dummies for post-treatment period and ever-treated group?

- Fixed effects "soak up" period-specific shocks, unit-specific variation better
  - Smaller residuals $\Rightarrow$ smaller standard errors $\Rightarrow$ statistical power
- Inclusion of time fixed effects yield should not lead to substantial changes in coefficients
  - Coefficients mechanically identical in balanced panels

Two-way fixed effects specification:

$$Y_{i,t} = \alpha + \eta_i + \nu_t + \delta D_{i,t} + \varepsilon_{i,t}$$

where η_i is an individual FE, ν_t is a time FE, and δ is DD estimator

Use two-way fixed effects with caution when treatment starts at different times in different units, treatment is continuous, or variance of treatment differs across treated units for other reasons, as we discuss further in the next module.

Example: Malawi's Ban on Traditional Birth Attendants

Does a ban on informal health providers save lives? Evidence from Malawi☆

Susan Godlonton [a,b], Edward N. Okeke [c,*]

[a] Department of Economics, Williams College, United States
[b] IFPRI, United States
[c] Department of Economics, Sociology and Statistics, RAND, United States

ARTICLE INFO

Article history:
Received 3 January 2015
Received in revised form 2 September 2015
Accepted 3 September 2015
Available online 11 September 2015

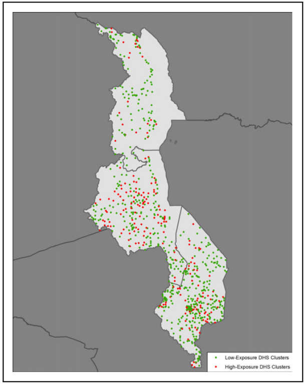
Keywords:
Informal health providers
Government bans
Child mortality

ABSTRACT

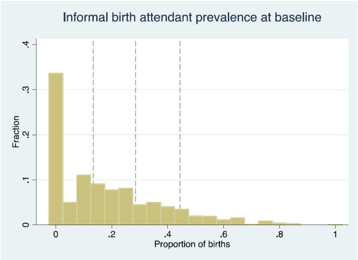
Informal health providers ranging from drug vendors to traditional healers account for a large fraction of health care provision in developing countries. They are, however, largely unlicensed and unregulated leading to concern that they provide ineffective and, in some cases, even harmful care. A new and controversial policy tool that has been proposed to alter household health seeking behavior is an outright ban on these informal providers. The theoretical effects of such a ban are ambiguous. In this paper, we study the effect of a ban on informal (traditional) birth attendants imposed by the Malawi government in 2007. To measure the effect of the ban, we use a difference-in-difference strategy exploiting variation across time and space in the intensity of exposure to the ban. Our most conservative estimates suggest that the ban decreased use of traditional attendants by about 15 percentage points. Approximately three quarters of this decline can be attributed to an increase in use of the formal sector and the remainder is accounted for by an increase in relative/friend-attended births. Despite the rather large shift from the informal to the formal sector, we do not find any evidence of a statistically significant reduction in newborn mortality on average. The results are robust to a triple difference specification using young children as a control group. We examine several explanations for this result and find evidence consistent with quality of formal care acting as a constraint on improvements in newborn health.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants



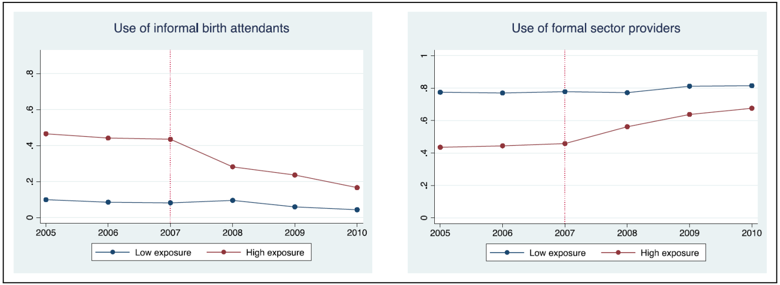
source: Godlonton and Okeke (2015)



Notes: This figure plots the distribution of DHS clusters by baseline prevalence of informal birth attendant use. Dotted lines from left to right indicate the 50th, 75th, and 90th percentiles respectively.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants



source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants

Godlonton and Okeke (2015) estimate regression specification:

$$Y_{ict} = \alpha + \theta HighExposure_c + \delta \left(HighExposure_c \times Post_t \right) + X_{ict}\beta + \tau_t + \varepsilon_{ict}$$

where:

- $HighExposure_c$ = indicator for (more) treated clusters (pre-ban use of TBAs above 75th percentile)
- $HighExposure_c \times Post_t$ = indicator for treated cluster-months
- δ = diff-in-diff estimate of treatment effect
- X_{ict} = set of control variables (e.g. household size, etc.)
- τ_t = fixed effect for month of birth (e.g. January 2007)
- ε_{ict} = mean-zero error term

Example: Malawi's Ban on Traditional Birth Attendants

Table 5
What was the effect of the ban on the use of formal and informal sector providers?

Variables	(1)	(2)	(3)	(4)	(5)	(6)
A. Birth attendant is informal attendant						
High exposure × *Post*	−0.189***	−0.190***	−0.184***	−0.187***	−0.154***	−0.188***
	(0.0146)	(0.0130)	(0.0141)	(0.0144)	(0.0126)	(0.0146)
High exposure	0.344***	0.321***	0.318***	0.320***	0.267***	
	(0.0143)	(0.0131)	(0.0123)	(0.0127)	(0.0110)	
Post				0.0134	−0.0655	−0.000915
				(0.0667)	(0.0908)	(0.0679)
Constant	0.0411***	0.0537	0.0512	1.848***	3.525***	0.265***
	(0.00204)	(0.0415)	(0.0410)	(0.284)	(0.440)	(0.0637)
Observations	19,607	18,673	18,673	18,673	12,491	18,673
R-squared	0.138	0.149	0.150	0.148	0.113	0.209
B. Birth attendant is formal sector provider						
High exposure × *Post*	0.145***	0.144***	0.143***	0.146***	0.109***	0.150***
	(0.0157)	(0.0136)	(0.0153)	(0.0152)	(0.0152)	(0.0165)
High exposure	−0.317***	−0.270***	−0.269***	−0.271***	−0.206***	
	(0.0177)	(0.0150)	(0.0152)	(0.0149)	(0.0155)	
Post				0.0660	0.132	0.00746
				(0.0794)	(0.0889)	(0.0974)
Constant	0.808***	0.726***	0.730***	−1.668***	−2.433***	0.446***
	(0.00257)	(0.0431)	(0.0429)	(0.391)	(0.479)	(0.0995)
Controls	No	Yes	Yes	Yes	Yes	Yes
Controls × *Post*	No	No	Yes	Yes	Yes	Yes
District-specific trend	No	No	No	Yes	Yes	No
Trimmed data	No	No	No	No	Yes	No
Cluster fixed effects	No	No	No	No	No	Yes
Observations	19,607	18,673	18,673	18,673	12,491	18,673
R-squared	0.088	0.132	0.134	0.131	0.104	0.218

Notes: for Panel A the dependent variable is an indicator for a birth attended by an informal birth attendant. For Panel B the dependent variable is an indicator for a birth attended by a formal-sector provider. Controls include an indicator for male births, an indicator for a multiple birth, birth order, dummies for mother's level of schooling, dummies for mother's age at birth, an indicator for women who are married or living with a partner, dummies for ethnicity and religion, dummies for the partner's educational attainment, distance to the nearest health facility, wealth quintile dummies, and a rural–urban indicator. Each column includes district and year × month fixed effects. Full set of coefficients is not shown to conserve space (see Table A.1). In Column 5, we exclude villages with baseline prevalence of 0 or 1 to account for 'floor' and 'ceiling' effects. Column 6 is equivalent to Column 3 except that district fixed effects have been replaced with cluster fixed effects. *Post* = 1 if birth occurs after December 2007. Standard errors in parentheses are clustered at the district level (there are 27 districts).
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants

Table 6
What was the effect of the ban on the use of other substitutes?

Variables	(1)	(2)	(3)	(4)	(5)	(6)
A. Birth attendant is a relative or friend						
High exposure × *Post*	0.0414***	0.0417***	0.0364***	0.0366***	0.0389***	0.0351***
	(0.00694)	(0.00725)	(0.00863)	(0.00918)	(0.0110)	(0.00962)
High exposure	−0.0256***	−0.0424***	−0.0396***	−0.0399***	−0.0496***	
	(0.00836)	(0.00933)	(0.00982)	(0.0101)	(0.0123)	
Post				−0.0476	−0.0367	0.121
				(0.0543)	(0.0812)	(0.0812)
Constant	0.105***	0.186***	0.184***	0.750***	0.251	0.202***
	(0.00151)	(0.0542)	(0.0536)	(0.236)	(0.329)	(0.0642)
Observations	19,607	18,673	18,673	18,673	12,491	18,673
R-squared	0.022	0.041	0.042	0.039	0.042	0.133
B. Birth was unattended						
High exposure × *Post*	0.00281	0.00322	0.00334	0.00247	0.00541	0.00116
	(0.00512)	(0.00491)	(0.00493)	(0.00557)	(0.00543)	(0.00518)
High exposure	0.000257	−0.00614*	−0.00622	−0.00629	−0.00931*	
	(0.00338)	(0.00339)	(0.00369)	(0.00393)	(0.00493)	
Post				0.0110	0.00680	−0.0164
				(0.0474)	(0.0572)	(0.0513)
Constant	0.0306***	0.0184	0.0173	−0.0440	−0.234	0.0319
	(0.000623)	(0.0267)	(0.0265)	(0.158)	(0.200)	(0.0346)
Controls	No	Yes	Yes	Yes	Yes	Yes
Controls × *Post*	No	No	Yes	Yes	Yes	Yes
District-specific trend	No	No	No	Yes	Yes	No
Trimmed data	No	No	No	No	Yes	No
Cluster fixed effects	No	No	No	No	No	Yes
Observations	19,607	18,673	18,673	18,673	12,491	18,673
R-squared	0.009	0.033	0.034	0.033	0.038	0.097

Notes: in Panel A (top) the dependent variable is an indicator for a birth attended by a relative or friend while in Panel B (bottom) the dependent variable is an indicator for an unattended birth. Controls include an indicator for male births, an indicator for a multiple birth, birth order, dummies for mother's level of schooling, dummies for mother's age at birth, an indicator for women who are married or living with a partner, dummies for ethnicity and religion, dummies for the partner's educational attainment, distance to the nearest health facility, wealth quintile dummies, and a rural–urban indicator. Each column includes district and year × month fixed effects. Full set of coefficients not shown to conserve space (see Table A.1). In Column 5, we exclude villages with baseline prevalence of 0 or 1 to account for 'floor' and 'ceiling' effects. Column 6 is equivalent to Column 3 except that district fixed effects have been replaced with cluster fixed effects. *Post* = 1 if birth occurs after December 2007. Standard errors in parentheses are clustered at the district level (there are 27 districts).
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants

Table 7
What was the effect of the ban on newbom deaths?

Variables	(1)	(2)	(3)	(4)	(5)	(6)
A. Child death within the first week						
High exposure × *Post*	−2.939	−2.742	0.319	−0.645	1.712	−0.344
	(3.500)	(3.530)	(3.669)	(3.706)	(4.121)	(3.508)
High exposure	5.311**	5.383*	4.528	4.661	2.850	
	(2.465)	(2.812)	(2.854)	(2.912)	(3.380)	
Post				−25.62	−33.96	−13.10
				(15.37)	(20.75)	(27.30)
Constant	21.39***	8.271	8.075	133.3***	139.7***	5.029
	(0.414)	(9.116)	(9.053)	(34.58)	(35.83)	(14.15)
Observations	35,246	33,748	33,748	33,748	22,317	33,748
R-squared	0.005	0.010	0.010	0.008	0.009	0.037
B. Child death within the first month						
High exposure × *Post*	−4.150	−4.414	−1.316	−1.908	−0.211	−2.760
	(4.242)	(4.274)	(4.369)	(4.515)	(4.603)	(4.337)
High exposure	6.659**	6.292*	5.465	5.428	3.899	
	(3.142)	(3.395)	(3.472)	(3.531)	(3.994)	
Post				−35.98	−54.18	−7.293
				(26.75)	(35.46)	(44.58)
Constant	31.90***	22.13*	21.72*	208.7***	231.5***	21.07
	(0.543)	(11.61)	(11.49)	(41.97)	(54.47)	(15.41)
Controls	No	Yes	Yes	Yes	Yes	Yes
Controls × *Post*	No	No	Yes	Yes	Yes	Yes
District-specific trend	No	No	No	Yes	Yes	No
Trimmed data	No	No	No	No	Yes	No
Cluster fixed effects	No	No	No	No	No	Yes
Observations	35,246	33,748	33,748	33,748	22,317	33,748
R-squared	0.005	0.012	0.012	0.010	0.012	0.038

Notes: in Panel A the dependent variable is an indicator for a newborn death within a week of being born. In Panel B the dependent variable is an indicator for a newborn death within a month of being born. Both variables have been scaled to allow coefficients to be interpretable as X per 1000 live births. Controls include indicators for male births, first births and young mothers (age < 18), dummies for mother's level of schooling, an indicator for women who are married or living with a partner, dummies for ethnicity and religion, dummies for the partner's educational attainment, distance to the nearest health facility, wealth quintile dummies, and a rural–urban indicator. Each column includes district and year × month fixed effects. Full set of coefficients is not shown to conserve space (see Table A.1). In Column 5, we exclude villages with baseline prevalence of 0 or 1 to account for 'floor' and 'ceiling' effects. Column 6 is equivalent to Column 3 except that district fixed effects have been replaced with cluster fixed effects. *Post* = 1 if birth occurs after December 2007. Standard errors in parentheses are clustered at the district level (there are 27 districts).

*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants

Table A.2
Was there an offsetting effect of relative/friend-attended births?

Variables	(1) High travel costs	(2) Low travel costs
Relative/friend-attended births	0.0430***	0.0267
	(0.0140)	(0.0190)
Child death within one week	−0.00244	0.00776
	(0.00499)	(0.00840)
Child death within one month	−0.00403	0.00467
	(0.00659)	(0.00836)
N	(0.00659)	(0.00836)

Notes: the sample in Column 1 is women who answered 'yes' when asked whether distance was a 'big problem' in accessing medical services. The sample in Column 2 is women who answered 'no'. The estimates are from the regression specification in Column 5 of Tables 6 and 7. Standard errors in parentheses are clustered at the district level (there are 27 districts).
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants

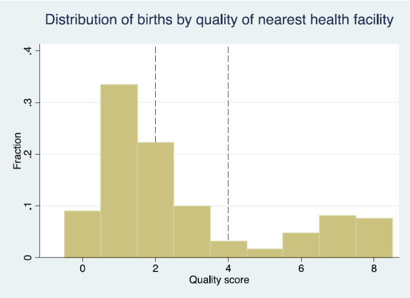
Table 9
Complier characteristics.

	$E(X \mid D_{1i} > D_{0i})$	$E(X)$	Relative likelihood
Male birth	0.509	0.500	0.9862
First birth	0.392	0.195	1.3172
Multiple pregnancy	0.067	0.043	1.3241
Young mother	0.493	0.116	1.4276
No maternal education	0.132	0.169	0.9862
No paternal education	0.052	0.108	0.6248
Has partner	0.812	0.869	0.9862
Poorest quintile	0.178	0.228	0.8345
Lives far from health facility	0.226	0.249	1.0483
Rural location	0.859	0.905	0.9448

Notes: the mean of each characteristic for compliers are in Column 1; population means are in Column 2, and the relative likelihood that compliers have a given characteristic are in Column 3. Details for how these are computed are in Section 6.5.2.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants



Notes: This figure shows the distribution of births by the quality score of the nearest health facility. The dotted lines indicate the 50th and 75th percentiles respectively. The set of indicators used in calculating the quality score are shown in Table A.7. Each indicator is assigned a score of one and we sum the scores for each facility.

source: Godlonton and Okeke (2015)

Example: Malawi's Ban on Traditional Birth Attendants

Table 10
Is the quality of formal care a constraint?

Variables	Nearest health facility is in the top quartile of quality distribution						Nearest health facility is in the bottom three quartiles of quality distribution					
	Child death within the first week						Child death within the first week					
High exposure × *Post*	−14.70**	−14.32*	−12.59	−14.42	−13.12	−12.84	−0.679	−0.333	2.491	2.377	5.432	1.653
	(6.725)	(7.100)	(8.439)	(8.651)	(9.448)	(8.425)	(4.418)	(4.586)	(4.633)	(4.656)	(5.253)	(4.420)
High exposure	8.537*	10.96*	10.44*	10.61*	8.673		4.950	4.942	4.188	4.094	2.445	
	(4.738)	(5.745)	(6.006)	(5.969)	(6.034)		(3.185)	(3.470)	(3.489)	(3.549)	(4.092)	
Post				−26.32	−47.46*	24.64				−29.31	−29.31	−44.60
				(15.79)	(24.46)	(46.70)				(22.08)	(26.48)	(27.03)
Constant	21.16***	−13.68	−17.15	66.70	103.3**	−4.514	30.46***	19.42	19.71	237.3***	296.6***	7.268
	(0.976)	(10.40)	(10.19)	(41.39)	(48.35)	(13.77)	(0.566)	(12.07)	(11.97)	(35.30)	(34.60)	(16.85)
Observations	8735	8570	8570	8570	4764	8570	25,666	25,178	25,178	25,178	17,553	25,178
R-squared	0.015	0.023	0.027	0.017	0.023	0.053	0.007	0.012	0.013	0.009	0.010	0.039
	Child death within the first month						Child death within the first month					
High exposure × *Post*	−17.33**	−17.02**	−16.00	−17.73*	−12.76	−17.69*	−1.482	−1.899	1.340	1.459	2.604	−0.0690
	(7.479)	(8.067)	(10.04)	(10.19)	(10.22)	(10.14)	(5.376)	(5.577)	(5.666)	(5.668)	(6.127)	(5.658)
High exposure	9.091	11.65*	11.25	11.64	8.405		6.391	5.658	4.839	4.584	3.835	
	(5.670)	(6.633)	(7.109)	(7.070)	(6.658)		(3.934)	(4.172)	(4.254)	(4.288)	(4.869)	
Post				−33.89	−45.00*	60.43				−41.40	−62.67	−60.54
				(34.30)	(25.51)	(89.20)				(34.54)	(45.67)	(35.73)
Constant	31.91***	13.77	11.46	122.1**	125.2	29.74	36.17***	23.61	23.38	227.9***	351.2***	19.69
	(1.009)	(25.26)	(26.43)	(51.73)	(77.49)	(36.08)	(0.599)	(13.94)	(13.69)	(37.19)	(41.04)	(18.61)
Controls	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Controls × *Post*	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
District-specific trend	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes	No
Trimmed data	No	No	No	No	Yes	No	No	No	No	No	Yes	No
Cluster fixed effects	No	No	No	No	No	Yes	No	No	No	No	No	Yes
Observations	8735	8570	8570	8570	4764	8570	25,666	25,178	25,178	25,178	17,553	25,178
R-squared	0.019	0.028	0.030	0.019	0.024	0.054	0.007	0.013	0.014	0.011	0.013	0.039

Notes: coefficients have been scaled to allow them to be interpreted as *X* per 1000 live births. The sample in Panel A (left) consists of births to households with access to high-quality formal care (equal to one if the nearest health facility is within the top quartile of the quality distribution) while the sample in Panel B (right) consists of births to households where the nearest health facility is in the bottom three quartiles of the quality distribution. The dependent variables are shown at the top of each set of results. Controls include indicators for male births, first births and young mothers (age < 18), dummies for mother's level of schooling, an indicator for women who are married or living with a partner, dummies for ethnicity and religion, dummies for the partner's educational attainment, distance to the nearest health facility, wealth quintile dummies, and a rural–urban indicator. Each column includes district and year × month fixed effects. In Column 5, we exclude villages with baseline prevalence of 0 or 1 to account for 'floor' and 'ceiling' effects. Column 6 is equivalent to Column 3 except that district fixed effects have been replaced with cluster fixed effects. *Post* = 1 if birth occurs after December 2007. Standard errors in parentheses are clustered at the district level (there are 27 districts).

*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)

Continuous Variation in Treatment Intensity

Example: Malawi's Ban on Traditional Birth Attendants

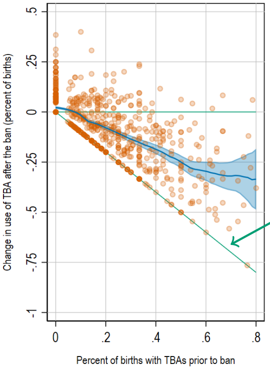
Alternative regression specification (that Godlonton and Okeke don't use):

$$Y_{i,t} = \alpha + \gamma \left(PreMeanTBA_c \times Post_t \right) + \beta X_{ict} + \eta_c + \tau_t + \varepsilon_{ict}$$

where:

- $PreMeanTBA_c$ = level of TBA use in cluster c before TBA ban
- γ = diff-in-diff estimate of treatment effect
- X_{ict} = set of control variables (eg household size, etc.)
- η_c = fixed effect for DHS cluster c
- τ_t = fixed effect for month of birth (eg January 2007)
- ε_{ict} = mean-zero error term

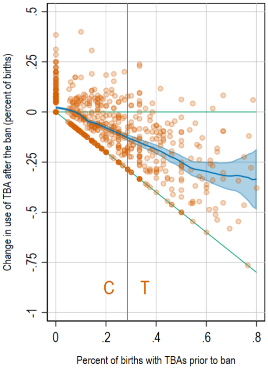
Example: Traditional Birth Attendants in Malawi



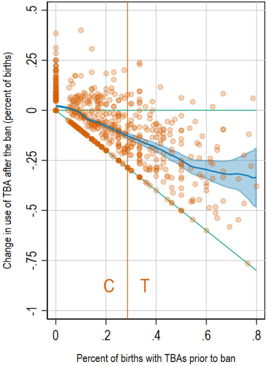
Post-ban mean must be positive

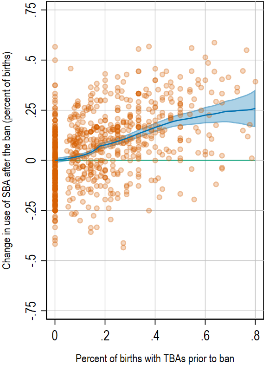
$\Rightarrow$ Decline $\uparrow$ with pre-ban mean

Example: Traditional Birth Attendants in Malawi

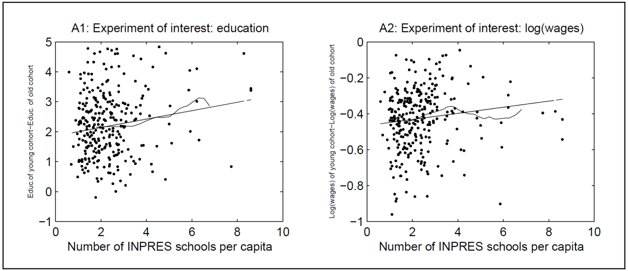


Example: Traditional Birth Attendants in Malawi





Example: School Construction in Indonesia



source: Duflo (2000)

Example: School Construction in Indonesia

Main empirical specification in Duflo (2001):

$$S_{ijk} = \alpha + \eta_j + \beta_k + \gamma \left(Intensity_j * Young_i \right) + \boldsymbol{C}_j \delta + \varepsilon_{ijk}$$

where:

- S_{ijk} = education of individual i born in region j in year k
- η_j = region of birth fixed effect
- β_k = year of birth fixed effect
- $Young_i$ = dummy for being 6 or younger in 1974 (treatment group)
- $Intensity_j$ = INPRES schools per thousand school-aged children
- $\boldsymbol{C}_j$ = a vector of region-specific controls (that change over time)

Example: School Construction in Indonesia

Dependent Variable: Years of Education

	Obs.	OLS (1)	OLS (2)	OLS (3)
Panel A: Entire Sample				
$Intensity_j * Young_i$	78,470	0.124	0.150	0.188
		(0.025)	(0.026)	(0.029)
Panel B: Sample of Wage Earners				
$Intensity_j * Young_i$	31,061	0.196	0.199	0.259
		(0.042)	(0.043)	(0.050)
Controls Included:				
YOB∗enrollment rate in 1971		No	Yes	Yes
YOB∗other INPRES programs		No	No	Yes

Sample includes individuals aged 2 to 6 or 12 to 17 in 1974. All Specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummis and the number of children in the region of birth (in 1971). Standard errors are in parentheses.

Example: School Construction in Indonesia

Dependent Variable: Log Hourly Wages (as Adults)

	Obs.	OLS (1)	OLS (2)	OLS (3)
Panel A: Sample of Wage Earners				
$Intensity_j * Young_i$	31,061	0.0147	0.0172	0.027
		(0.007)	(0.007)	(0.008)
Controls Included:				
YOB$*$enrollment rate in 1971		No	Yes	Yes
YOB$*$other INPRES programs		No	No	Yes

Sample includes individuals aged 2 to 6 or 12 to 17 in 1974. All Specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummis and the number of children in the region of birth (in 1971). Standard errors are in parentheses.

Testing Common Trends

How Can We Test the Common Trends Assumption?

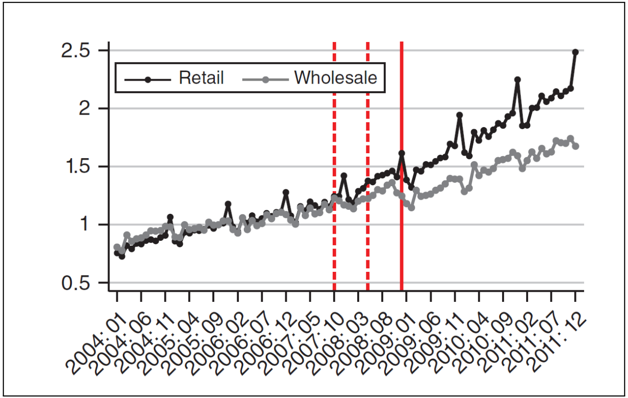
The **common trends** assumption: in the absence of treatment, outcomes in the treatment (i.e. ever-treated) group and the comparison group would have evolved along similar trajectories

- Common trends relates to potential outcomes without treatment
  - We can never observe the (treatment group) counterfactual
- It is fundamentally impossible to test the common trends assumption

Approaches to defending (or perhaps evaluating) the common trends assumption:

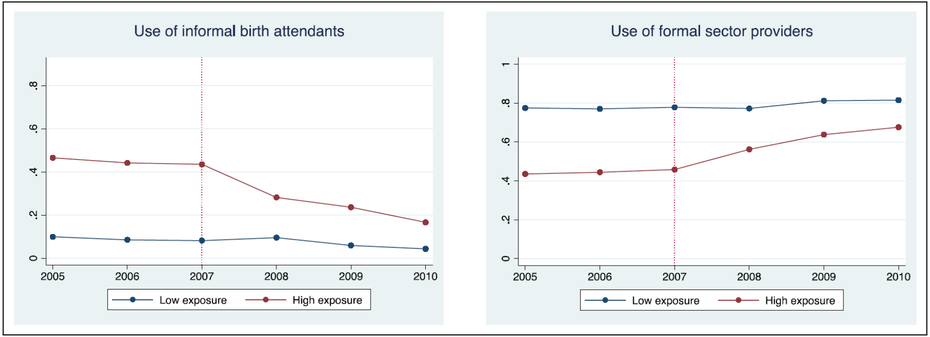
1. Comparing pre-treatment trends in the treatment and comparison groups
2. Conducting a **falsification test** (sometimes called a **placebo test**)
3. Triple differences: identifying an additional comparison group within the treatment group

Pre-Trends: A Picture Is Worth a Thousand Words



source: Naritomi (2019)

Pre-Trends: A Picture Is Worth a Thousand Words



source: Godlonton and Okeke (2015)

Testing Pre-Trends in a Regression

Godlonton and Okeke (2015) test for differences in pre-treatment trends:

$$Y_{ict} = \alpha + \beta HighExposure_c + \lambda Time_t + \gamma \left(HighExposure_c \times Time_t \right) + \varepsilon_{ict}$$

where:

- Y_{it} = outcome variable in cluster i at time t
- $HighExposure_c$ = indicator for (eventually) treated clusters
- $Time_t$ = (linear) measure of months from start of data set
- γ = measures equality of time trends between treatment, control
- ε_{it} = mean-zero error term

Sample is restricted to observations from before the ban on traditional birth attendants

Testing Pre-Trends in a Regression

Table 2
Test of parallel time trends.

Variables	(1) Birth attended by informal attendant	(2) Birth attended by formal-sector provider	(3) Child death within the first week	(4) Child death within the first month
High exposure	0.566	−0.419	0.0402	0.0340
	(0.505)	(0.572)	(0.0439)	(0.0537)
Time trend	−0.000558	0.000560	−5.75e-05	−0.000110*
	(0.000349)	(0.000442)	(4.30e-05)	(5.58e-05)
High exposure × time trend	−0.000388	0.000175	−6.52e-05	−5.12e-05
	(0.000902)	(0.00102)	(8.21e-05)	(9.83e-05)
Constant	0.401*	0.459	0.0499**	0.0860***
	(0.195)	−0.247	(0.0230)	(0.0301)
Observations	9277	9277	25,696	25,696
R-squared	0.171	0.100	0.002	0.002

source: Godlonton and Okeke (2015)

Testing Pre-Trends: Implications for Practice

- A compelling test of the equality of pre-trends requires lots of pre-treatment data
  - At a minimum, you need two pre-treatment periods
  - Statistical power can be a serious issue with limited pre-treatment data
- Often makes sense to disaggregate data as much as possible (e.g. months instead of years)
  - Treatment and comparison groups should be impacted by the same period-specific shocks
- Whenever possible, graph your data **and** conduct a statistical (i.e. regression) test

Falsification (or Placebo) Tests

A placebo or **falsification test** looks for effects that shouldn't be there using:

- A different outcome (that should not be impacted)
- A different (i.e. not real) definition of treatment
- A different sample (i.e. one not impacted by treatment)

Unlike tests of pre-trends, falsification tests typically use the same diff-in-diff regression specifications as the main analysis (except for the one placebo element being tested)

Falsification Test Example 1: Alternative Outcomes

Table 4
Is treatment correlated with observables?

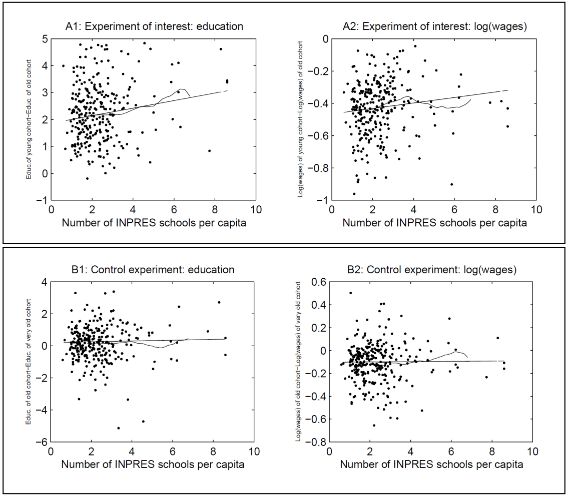
Variables	(1) Twin birth	(2) Male birth	(3) First birth	(4) Young mother	(5) Number of children ever born	(6) Mother's age at first birth	(7) Mother is Christian	(8) No education (mother)	(9) No education (spouse)
High exposure	0.0089	0.00306	−0.0095	−0.000565	0.237***	−0.00518	0.0172	0.0362***	0.00916
	−0.00578	−0.0122	−0.00854	−0.00758	−0.07	−0.113	−0.015	−0.0113	−0.00889
Post	0.00852	0.00697	−0.0209*	−0.00985	−0.00716	−0.143	−0.00901	0.0228*	0.00111
	−0.00858	−0.0134	−0.0112	−0.00737	−0.0616	−0.0949	−0.00953	−0.0123	−0.00908
High exposure × *Post*	−0.00693	−0.00823	0.00431	0.00415	−0.0992	−0.0479	−0.00602	−0.0164	0.00414
	−0.00744	−0.0156	−0.0122	−0.00865	−0.0713	−0.0641	−0.00817	−0.00994	−0.00847
Constant	0.549***	0.935***	−0.187	0.477***	17.16***	15.01***	0.868***	0.479***	−0.0109
	−0.12	−0.17	−0.157	−0.104	−0.795	−1.294	−0.121	−0.163	−0.121
Observations	19,680	19,680	19,680	19,680	19,680	19,680	19,680	19,680	19,192
R-squared	0.009	0.003	0.005	0.008	0.03	0.021	0.355	0.068	0.039

Variables	(10) Unemployed	(11) Household head is female	(12) Household size	(13) Household has bicycle	(14) Household has electricity	(15) Household has radio	(16) Poorest wealth quintile	(17) Rural location	(18) Distance to nearest facility
High exposure	−0.0155	−0.00487	0.00358	0.0149	−0.0385***	−0.0479***	0.0627***	0.0580***	2.067***
	−0.0173	−0.014	−0.0679	−0.0246	−0.00998	−0.0113	−0.0143	−0.0159	−0.296
Post	−0.0227**	−0.0122	−0.0784	0.0122	−0.000171	−0.0200***	−0.0155	0.00227	−0.0321
	−0.00914	−0.0139	−0.0578	−0.0111	−0.00514	−0.00683	−0.0102	−0.00766	−0.1
High exposure × *Post*	−0.00356	0.00421	−0.0965	−0.00849	−0.00296	0.00277	0.00149	−0.00327	−0.014
	−0.0116	−0.0119	−0.0629	−0.0151	−0.00567	−0.00755	−0.0148	−0.00607	−0.067
Constant	−0.929***	−0.259	10.56***	1.737***	−0.183***	−0.313***	−0.431***	0.971***	5.341***
	−0.137	−0.177	−0.765	−0.152	−0.0642	−0.0923	−0.134	−0.102	−1.293
Observations	19,680	19,680	19,680	19,680	19,680	19,680	19,680	19,680	19,211
R-squared	0.086	0.022	0.028	0.043	0.062	0.046	0.053	0.192	0.242

Notes: all columns include district and month × year fixed effects. Standard errors in parentheses are clustered at the district level (there are 27 districts).
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)

Falsification Test Example 2: Alternative Sample / Placebo Treatment



source: Duflo (2000)

Falsification Test Example 2: Alternative Sample / Placebo Treatment

Dependent Variable: Years of Education

	Obs.	OLS (1)	OLS (2)	OLS (3)
Panel A: Entire Sample				
$Intensity_j * Younger_i$	78,488	0.009	0.018	0.008
		(0.026)	(0.027)	(0.030)
Panel B: Sample of Wage Earners				
$Intensity_j * Younger_i$	30,255	0.012	0.024	0.079
		(0.048)	(0.048)	(0.056)
Controls Included:				
YOB∗enrollment rate in 1971		No	Yes	Yes
YOB∗other INPRES programs		No	No	Yes

Sample includes individuals aged 12 to 24 in 1974. All Specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummis and the number of children in the region of birth (in 1971). Standard errors are in parentheses.

Falsification Tests: Implications for Practice

Falsification tests are fundamentally context-specific:

- Which outcomes and/or groups should not be impacted?
  - Could there be spillovers onto groups that weren't directly treated?
  - Could treatment have unintended consequences?
    - Example: impacts of maternity leave on attitudes toward LGBTQ issues

There is not a one-size-fits-all approach to coming up with a good falsification test

- You need to know your setting and your data, and framing matters

Triple-Differences as a Test of Common Trends

Table 8
Effect of the ban on child mortality — triple difference specification.

Variables	Child death (1)	(2)	(3)
High exposure × *Post* × *Treated*	−1.320	−1.332	3.008
	(3.518)	(3.474)	(3.468)
High exposure × *Treated*	4.093*	4.051*	2.060
	(2.122)	(2.132)	(2.457)
Treated × *Post*	1.209	1.639	−2.541
	(2.113)	(2.148)	(2.379)
High exposure × *Post*	−0.00819	0.0151	−0.502
	(1.552)	(1.538)	(1.672)
High exposure	−0.207	−0.252	−0.431
	(1.062)	(1.061)	(1.102)
Post		−9.929	−15.05*
		(7.668)	(8.550)
Treated	0.548	0.261	−0.115
	(8.593)	(8.499)	(8.792)
Constant	6.118	59.99***	62.21***
	(3.688)	(10.73)	(11.59)
District-specific trend	No	Yes	Yes
Trimmed data	No	No	Yes
Observations	122,301	122,301	79,596
R-squared	0.008	0.007	0.008

Notes: dependent variable is a child death in year *t*. It has been scaled to allow coefficients to be interpretable as *X* per 1000 live births. *Treated* is equal to 1 for newborns and equal to 0 for children aged 2–5 years in year *t*. Each column includes district and year × month fixed effects, the full set of controls and their interactions with *Post* and *Treated*. In Column 3, we exclude villages with baseline prevalence of 0 or 1 to account for 'floor' and 'ceiling' effects. *Post* = 1 if birth occurs after December 2007. Standard errors in parentheses are clustered at the district level (there are 27 districts).
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

source: Godlonton and Okeke (2015)