Webinar Series in Applied Quantitative Analysis

Date	Topic
	Session One
February 29	Potential Outcomes and Omitted Variable Bias I (Theory) ✓
March 7	Potential Outcomes and Omitted Variable Bias II (Application) ✓
	Session Two
March 21	Difference-in-differences I (Theory)
March 28	Difference-in-differences II (Application)
	Session Three
April 25	Power analysis, clustering and sample size calculations I (Theory)
May 2	Power analysis, clustering and sample size calculations II (Application)
	Session Four
May 23	Fixed-effects I (Theory)
May 30	Fixed-effects II (Application)
	Session Five
June 20	Instrumental variables I (Theory)
June 27	Instrumental Variables II (Application)
	Session Six
July 25	Lagged dependent variables and the Arellano-Bond Estimator I (Theory)
August 1	Lagged dependent variables and the Arellano-Bond Estimator II (Application)

AIR®
Advancing Evidence.
Improving Lives.

https://www.airp3-africa.org/applied-quantitative-analysis-webinar-series

AERC
CREA

Difference-in-Differences I

Ashu Handa
Institute Fellow – AIR
Kenan Eminent Professor of Public Policy – UNC-CH
March 21, 2024

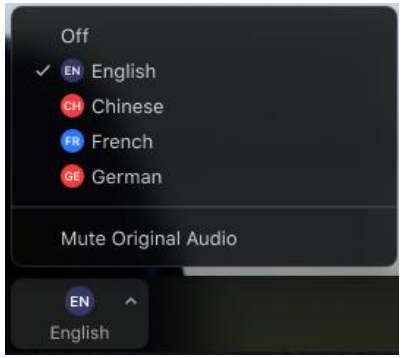
[Some slides are courtesy of Prof. Gustavo Angeles]

AIR®
Advancing Evidence.
Improving Lives.

AERC
CREA

Écoute de l'interprétation d'une langue
Windows | macOS

1. Dans les contrôles de votre réunion/webinaire, cliquez sur Interprétation .
2. Cliquez sur la langue que vous souhaitez entendre. (Nous aurons le français) Pas besoin de choisir l'anglais, c'est la langue de la salle Zoom principale



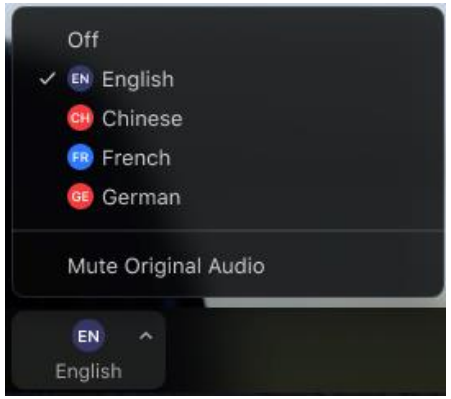
3. (Facultatif) Pour entendre uniquement la langue interprétée, cliquez sur Couper le son original.

Remarques:

- Vous devez rejoindre l'audio de la réunion via l'audio/VoIP de votre ordinateur. Vous ne pouvez pas écouter l'interprétation linguistique si vous utilisez les fonctions audio de connexion ou d'appel téléphonique.
- En tant que participant rejoignant une chaîne linguistique, vous pouvez retransmettre sur le canal audio principal canal si vous réactivez votre audio et parlez.

Listening to language interpretation
Windows | macOS

1. In your meeting/webinar controls, click Interpretation .
2. Click the language that you would like to hear. (We will have French) No need to choose English, that is the language in the main Zoom room



3. (Optional) To hear the interpreted language only, click Mute Original Audio.

Notes:

- You must join the meeting audio through your computer audio/VoIP. You cannot listen to language interpretation if you use the dial-in or call me phone audio features.
- As a participant joining a language channel, you can broadcast back into the main audio channel if you unmute your audio and speak.

What is your best estimate of the ‘effect’ of job training on wages?

Wages ($ per hour) for Person A and B according to job training status					
	Person A		Person B (no training)		
Period 1 (before)	Before training	$12	Before training	$8	
Period 2	After training	$16	No training	$10	
Period 2 – Period 1		$4		$2	DD=$4-$2 = $2

DiD – focus on the simple 2x2 case (most common)

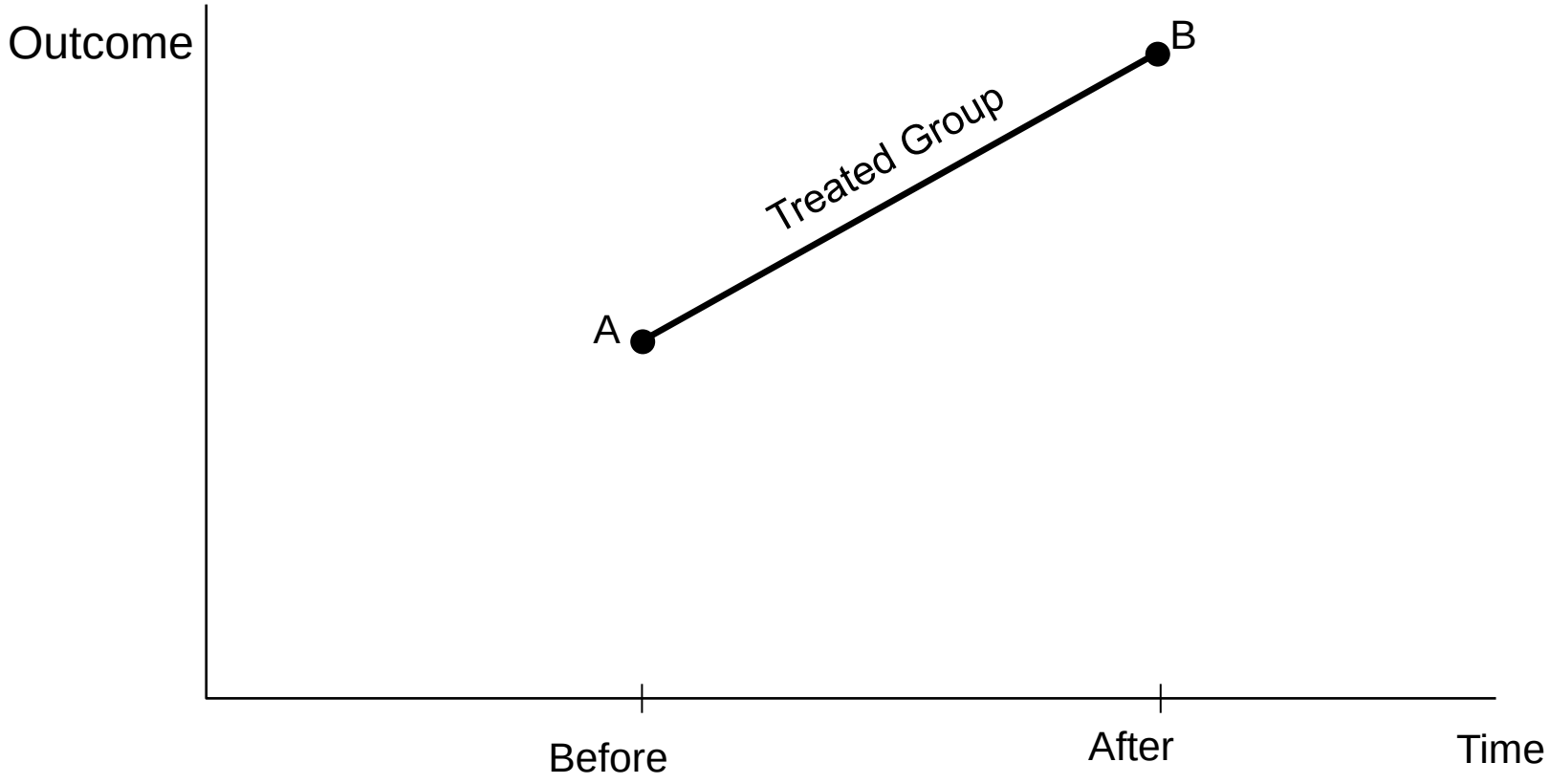
- Typical goal is to identify causal effect of some ‘treatment’ on the outcome
  - Can be a specific intervention, or a policy change that affects one group and not another
- Two groups: one exposed to the treatment (‘treated’), one not exposed (‘not-treated’ or ‘comparison’)
- Two time periods: Before the intervention and after the intervention

- Extensions with staggered treatment – will not cover that today

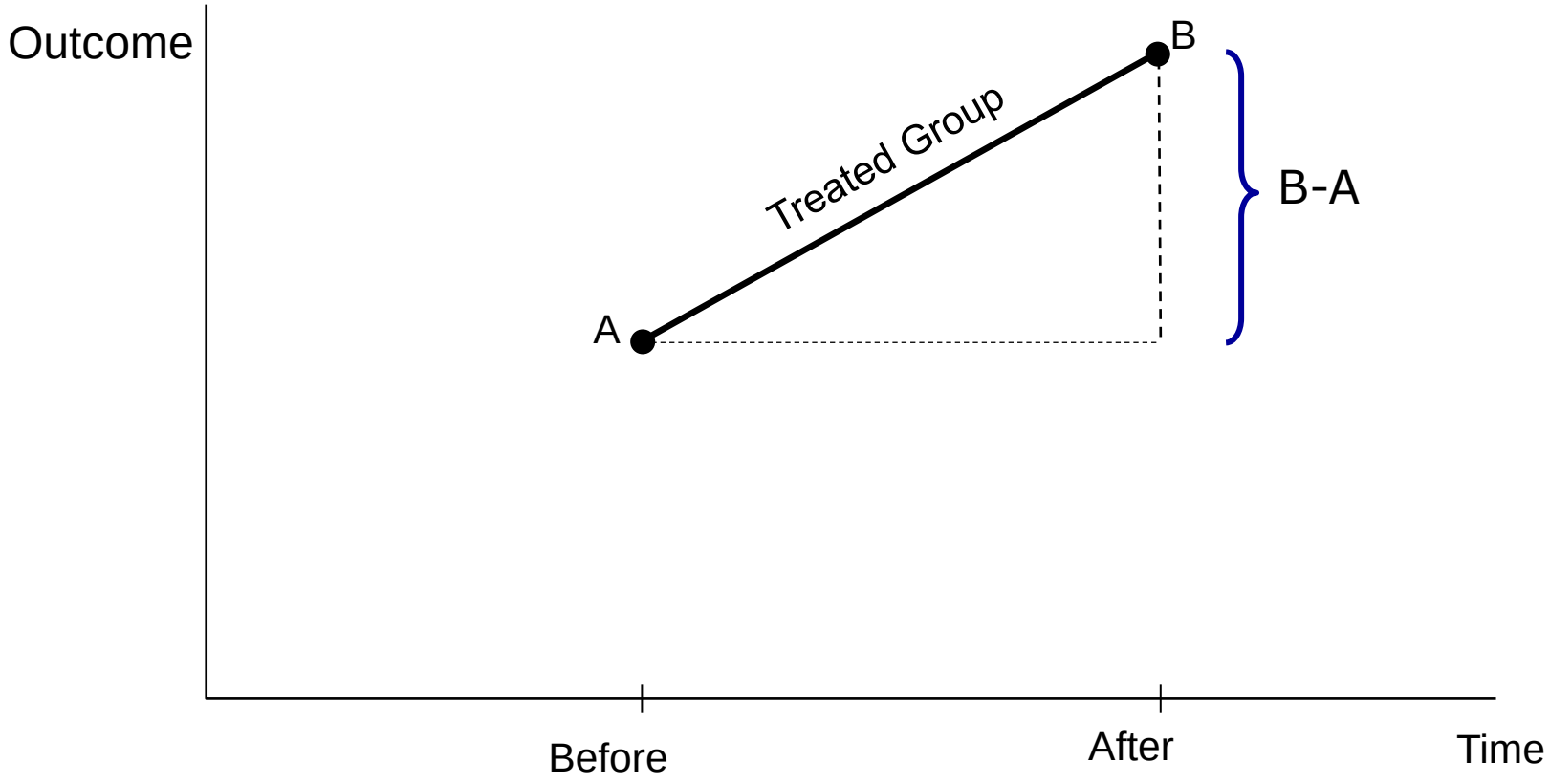
The classic 2x2 matrix (we will use this later)

	Before	After
Non-treated (comparison)		
Treated		

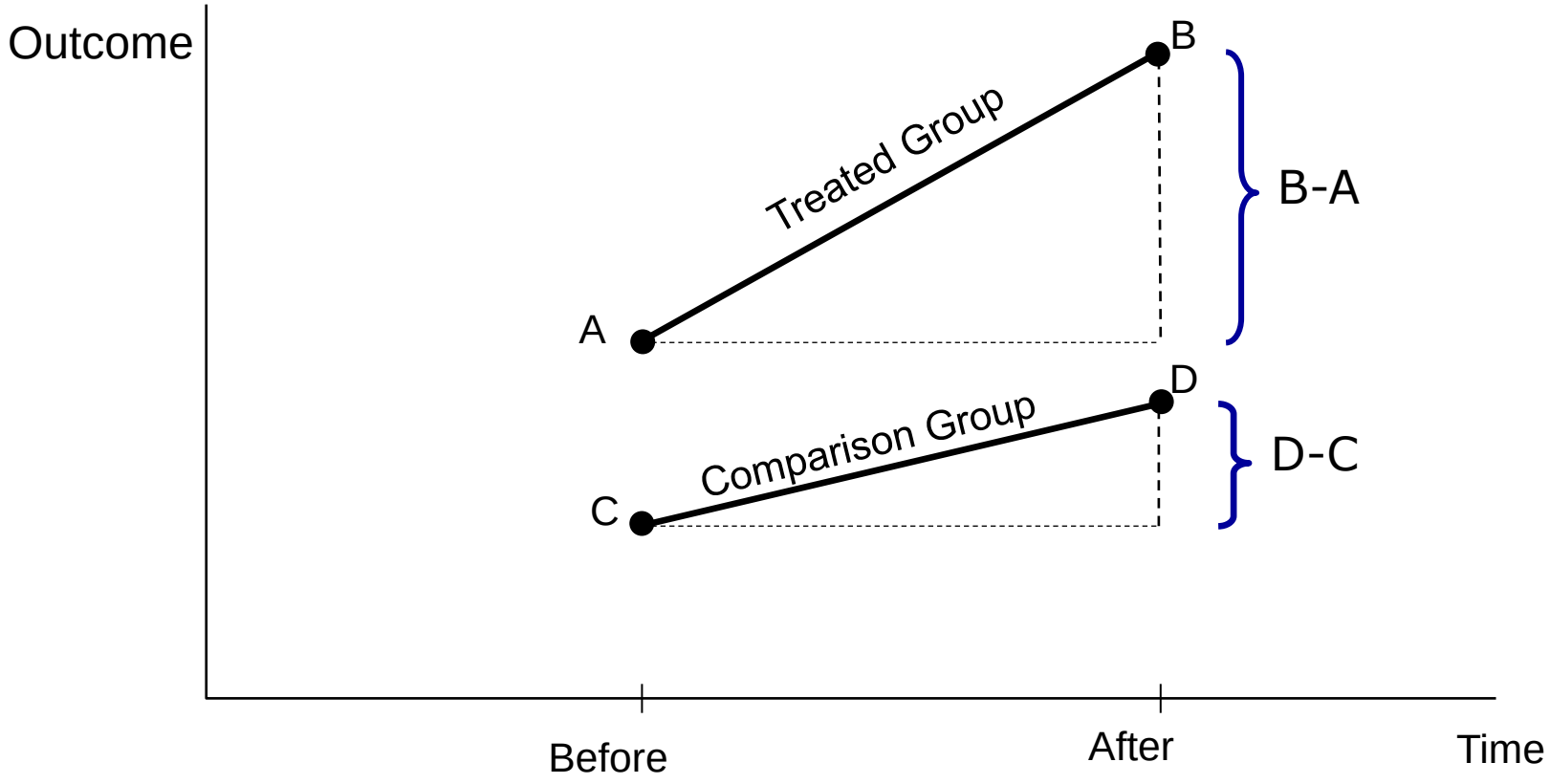
Difference-in-differences (DID) in graphs



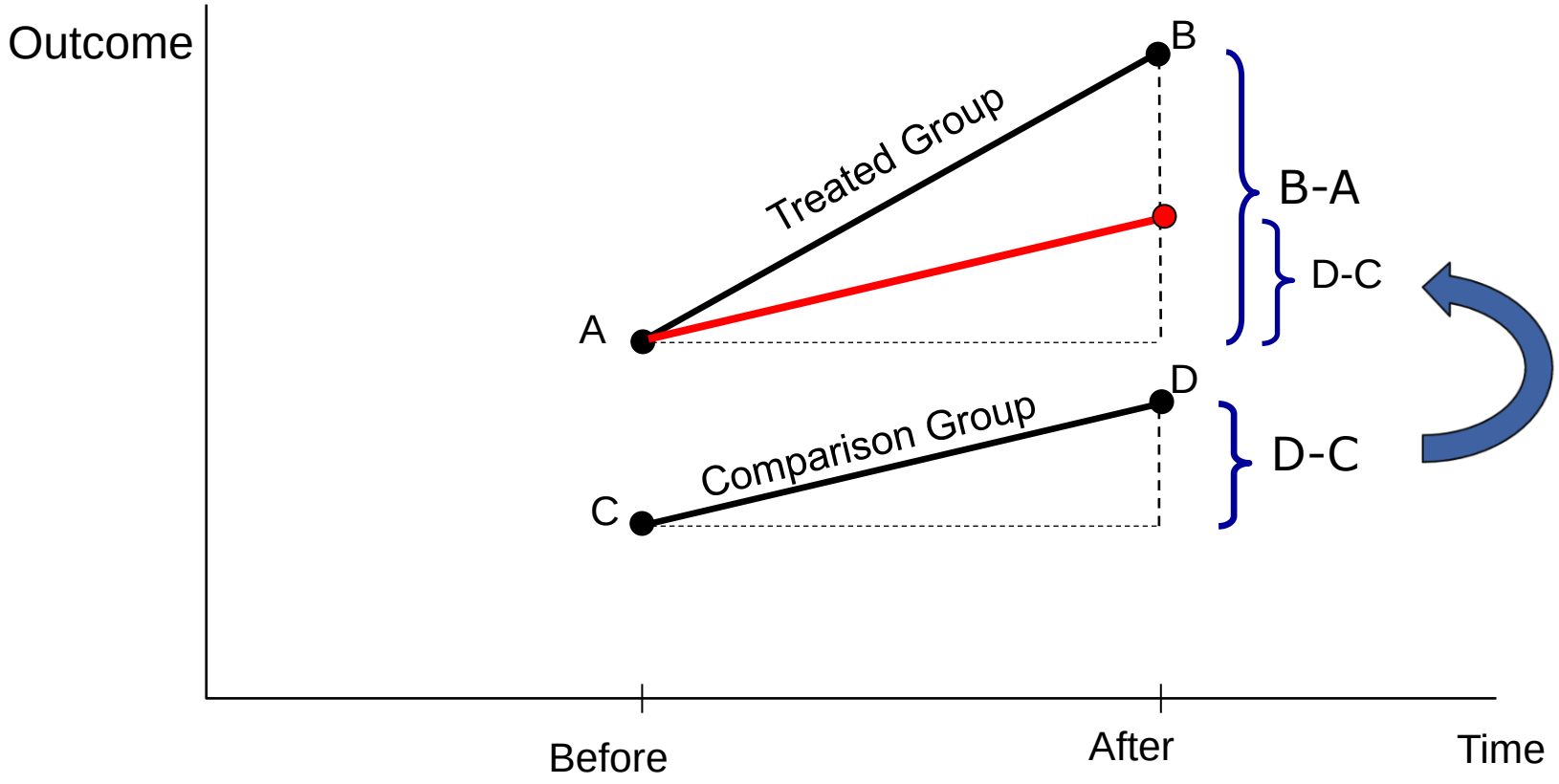
Difference-in-differences (DID) in graphs



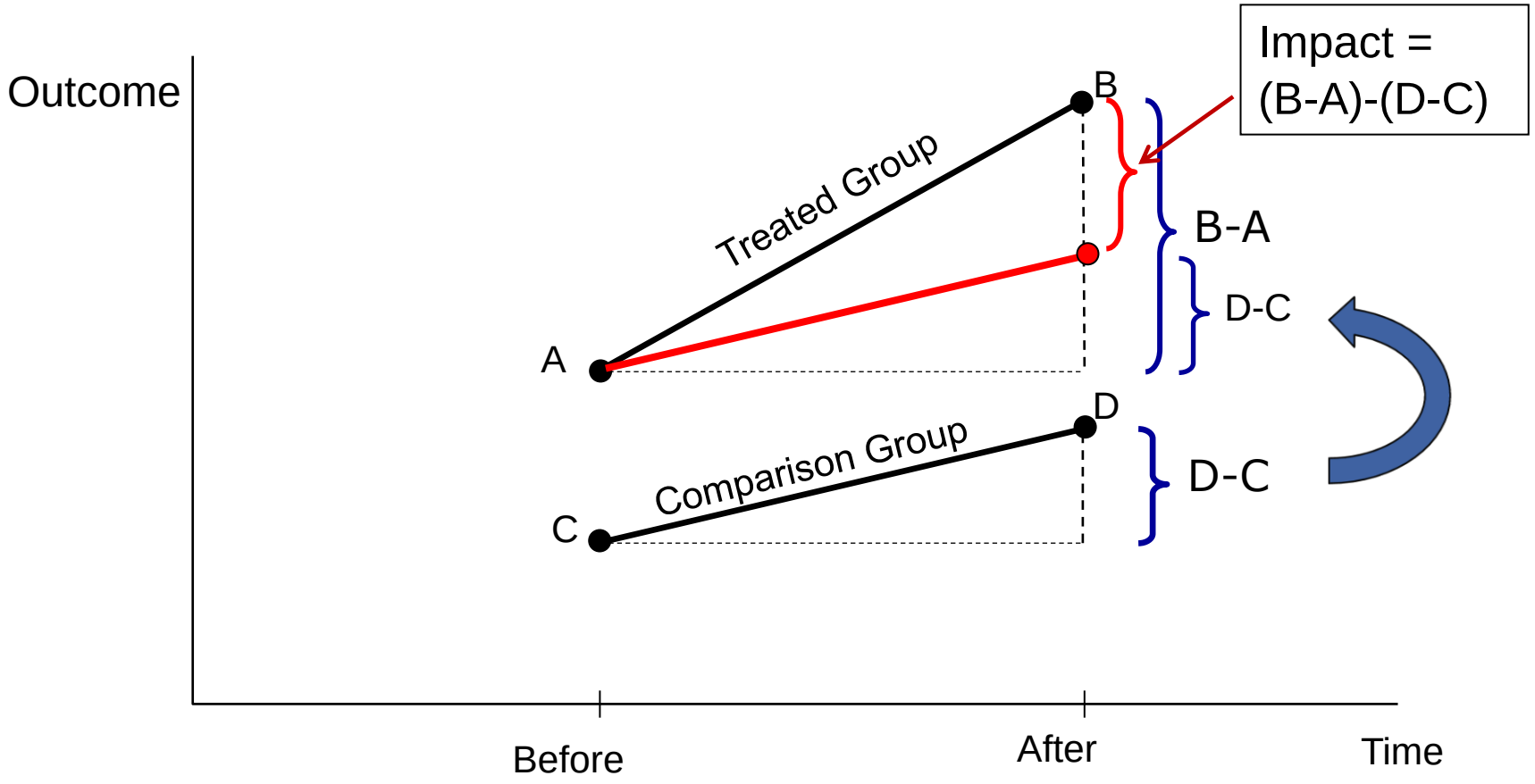
Difference-in-differences (DID) in graphs



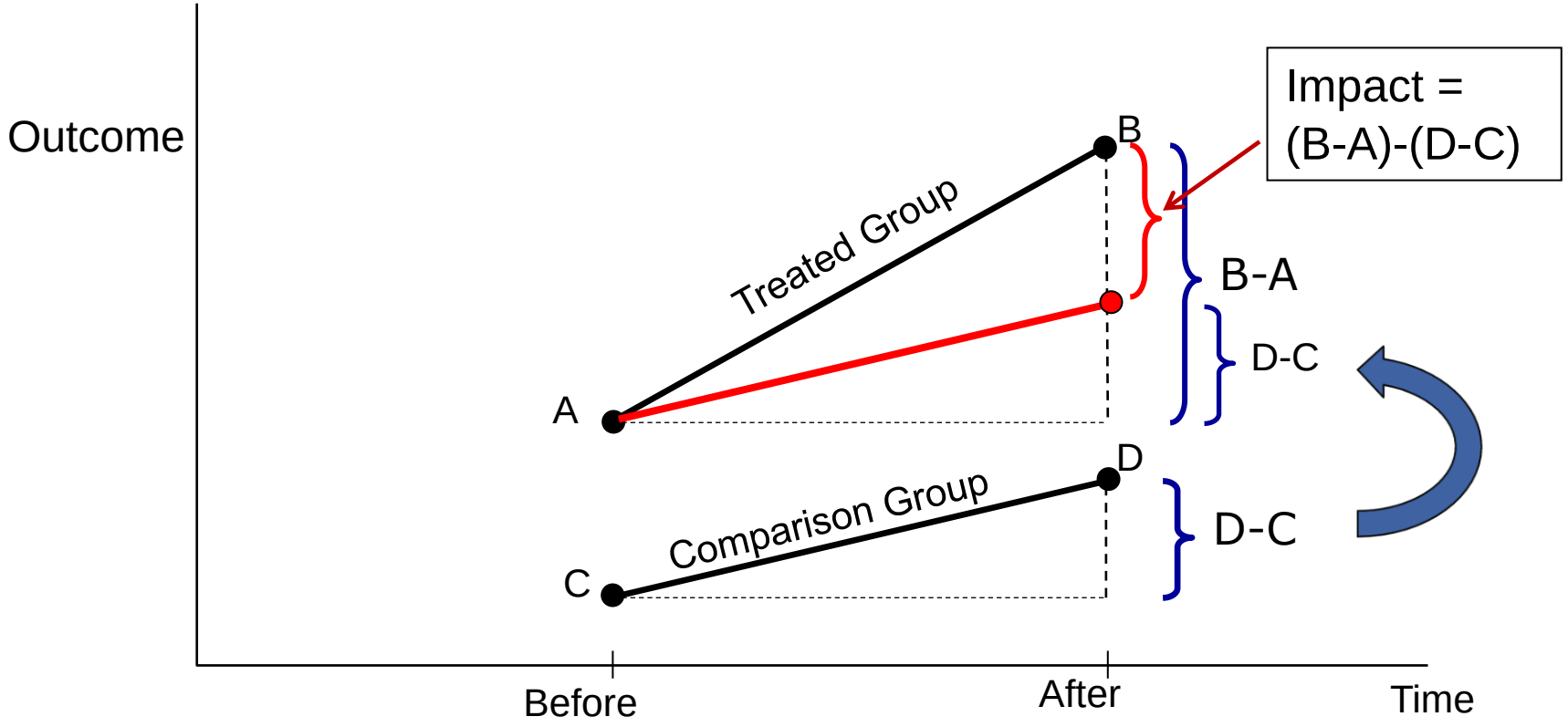
Difference-in-differences (DID) in graphs



Difference-in-differences (DID) in graphs



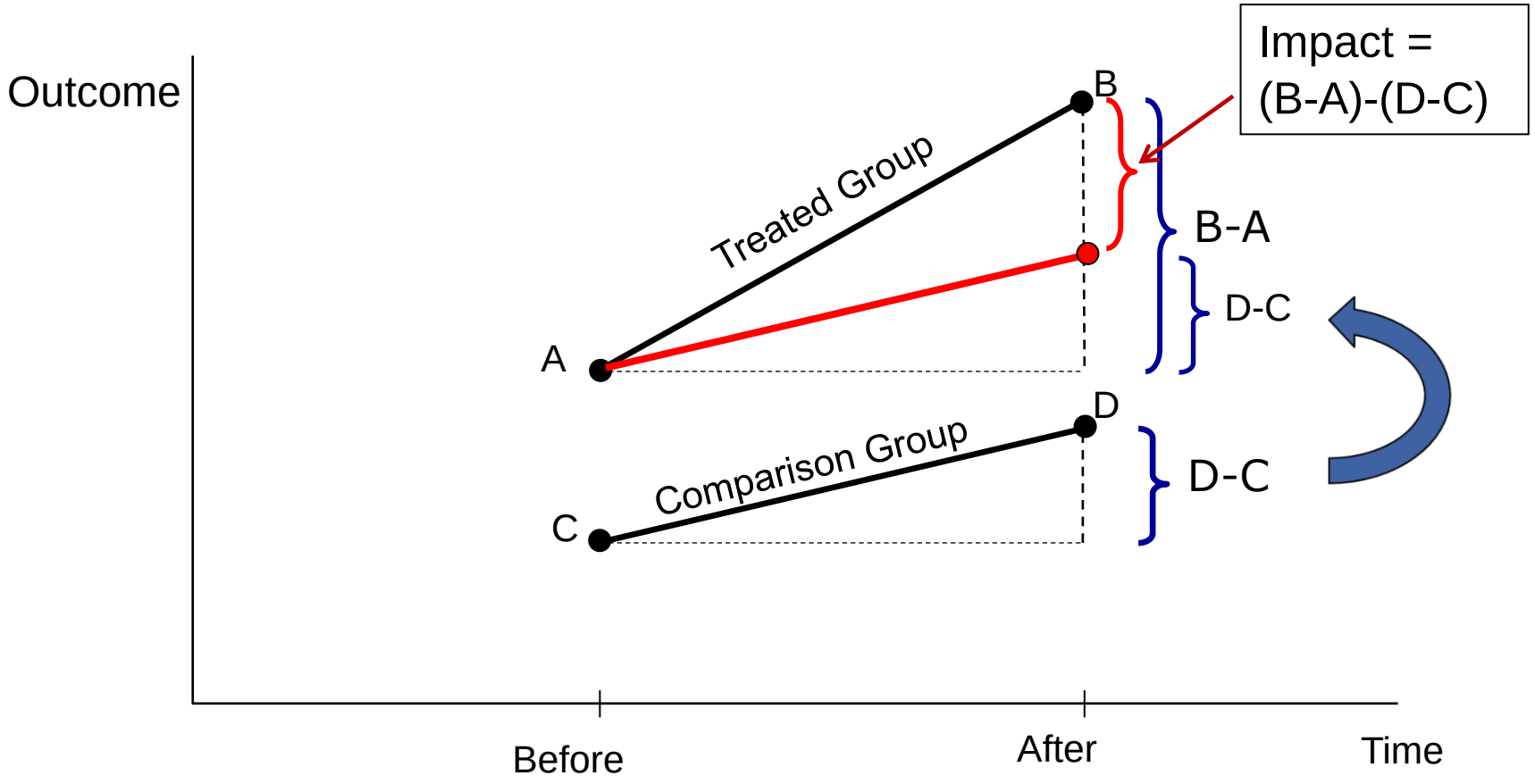
The Difference-in-differences 2x2 table



	BEFORE	AFTER	DIFFERENCE
Comparison	**C**	**D**	**D – C**
Treated	**A**	**B**	**B – A**
Diff-in-diff			**(B-A) – (D-C)**

DD Impact

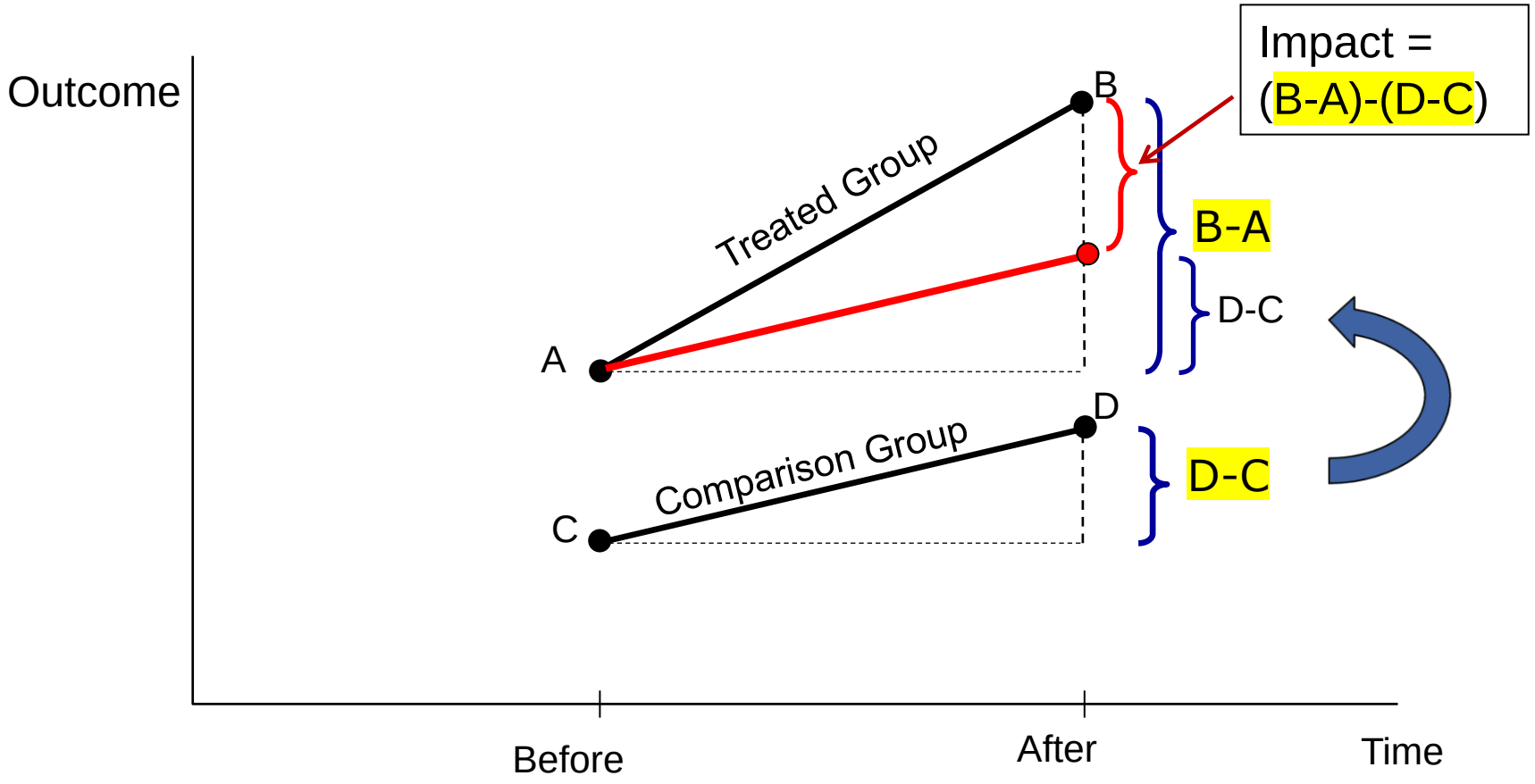
Key assumption behind DID is that of 'parallel trends'



Assumption: "Parallel Trends"

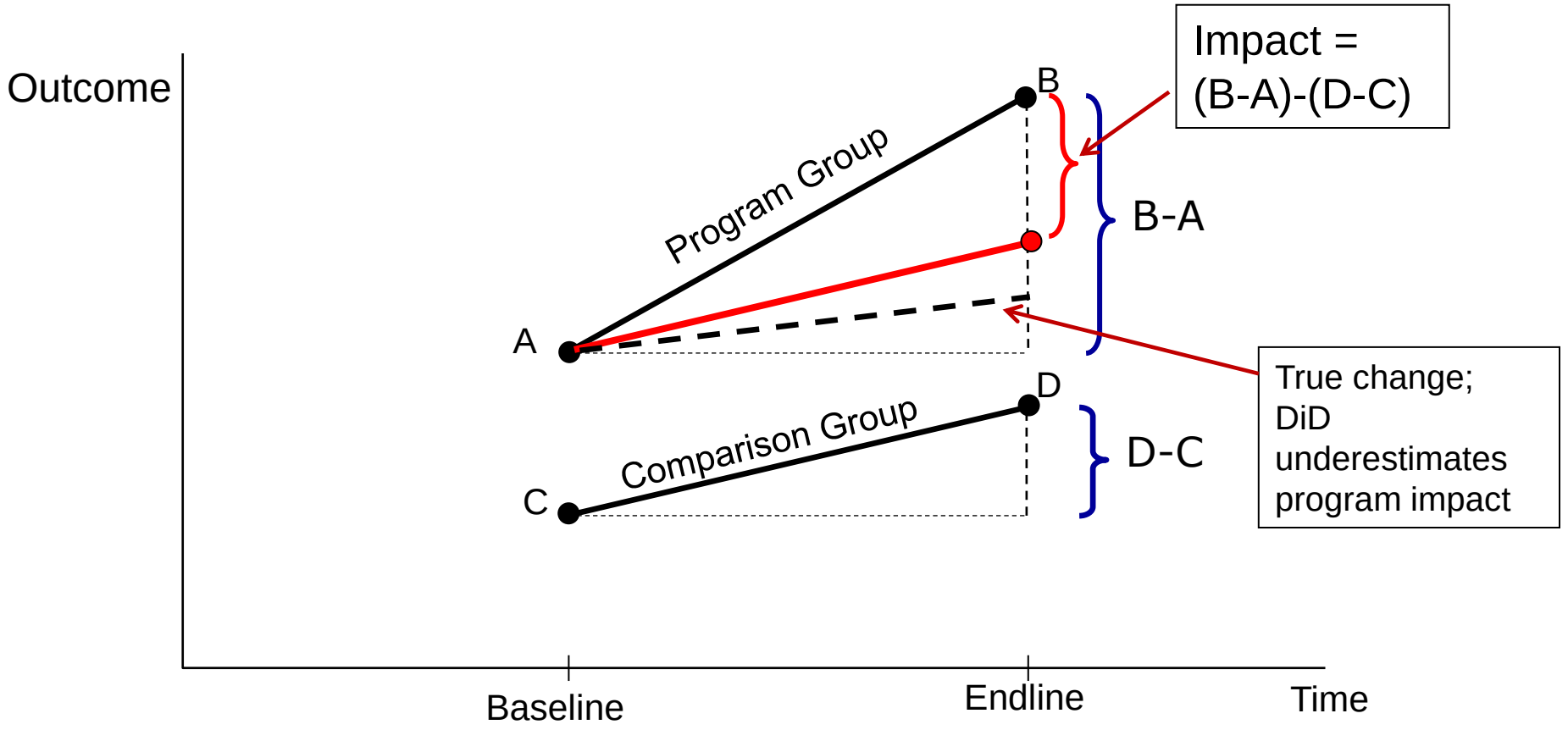
In absence of the treatment, the treated group would have had the same change in the outcome as the change in the Comparison group.

Assumption behind DID--it is the change, not the level!!



Important: Impact estimate does not depend on similar "levels" before the intervention. It depends on the two groups having similar behavioral responses to the intervention, that is, on having the "same change", the "same response".

DID: Parallel Trends could be a strong assumption



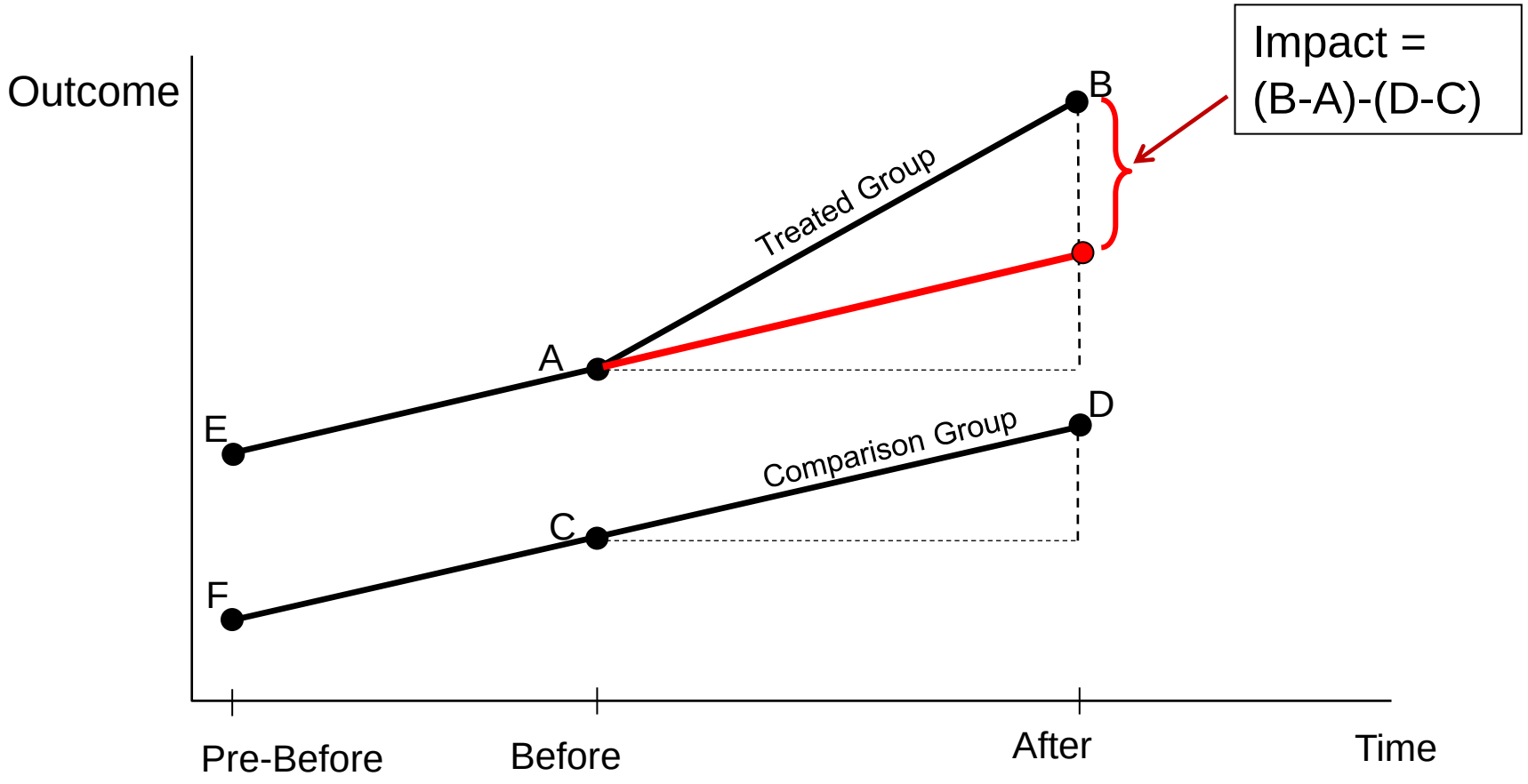
Assumption: "Parallel Trends"

Limitations: - Strong assumption; true change could've been different

- Might be more credible for "short" time interval, but that could reduce the magnitude of impact to estimate.

Important: The parallel trends assumption is about the time period after the program was implemented, it is not about the past.

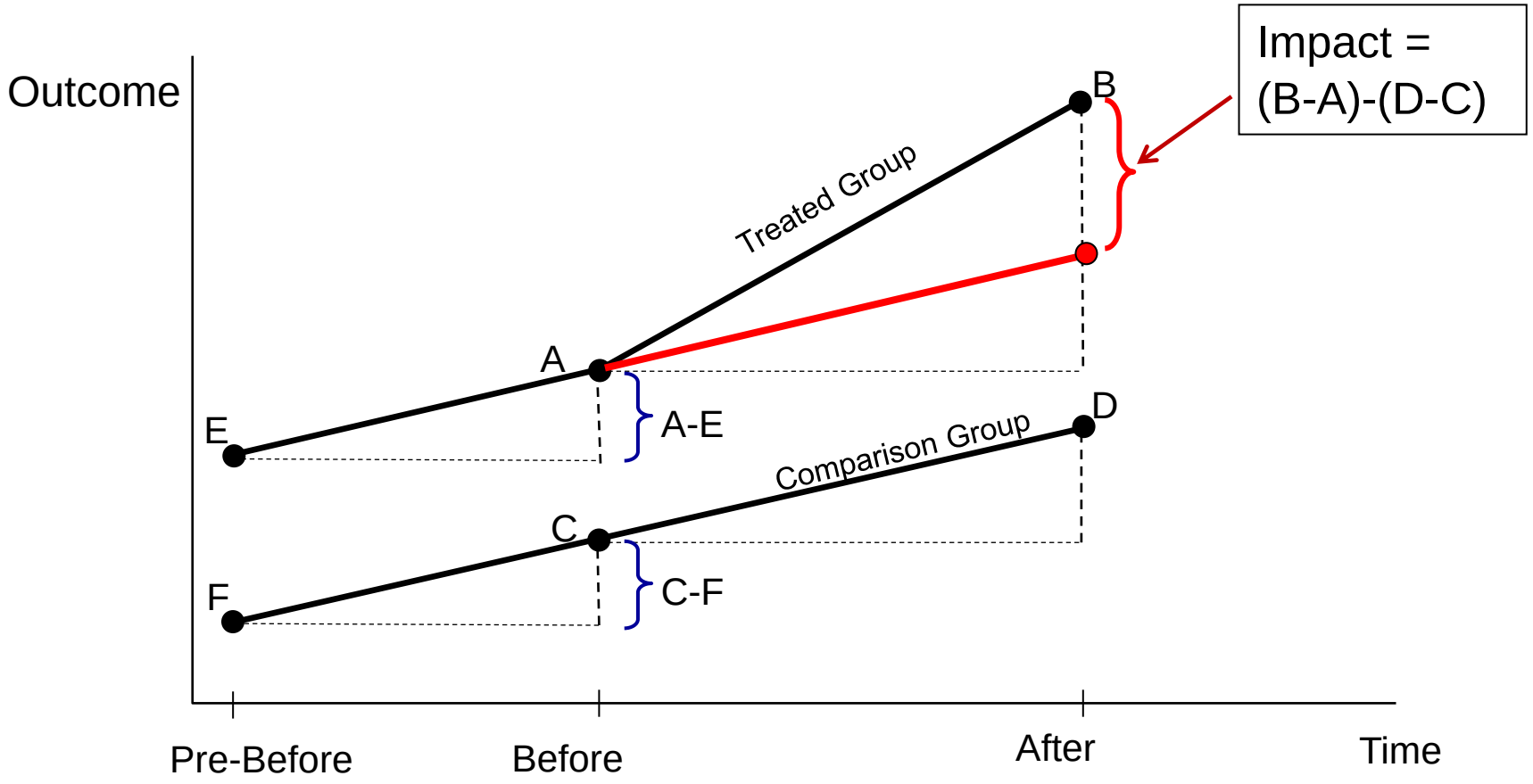
DiD: Testing the “Parallel trends” assumption



Because the parallel trends assumption is about the time the intervention was implemented, it cannot be tested. But we can look at the past as a proxy.

You need data prior to your analysis period! This can be difficult.

DiD: Testing the “Parallel trends” assumption



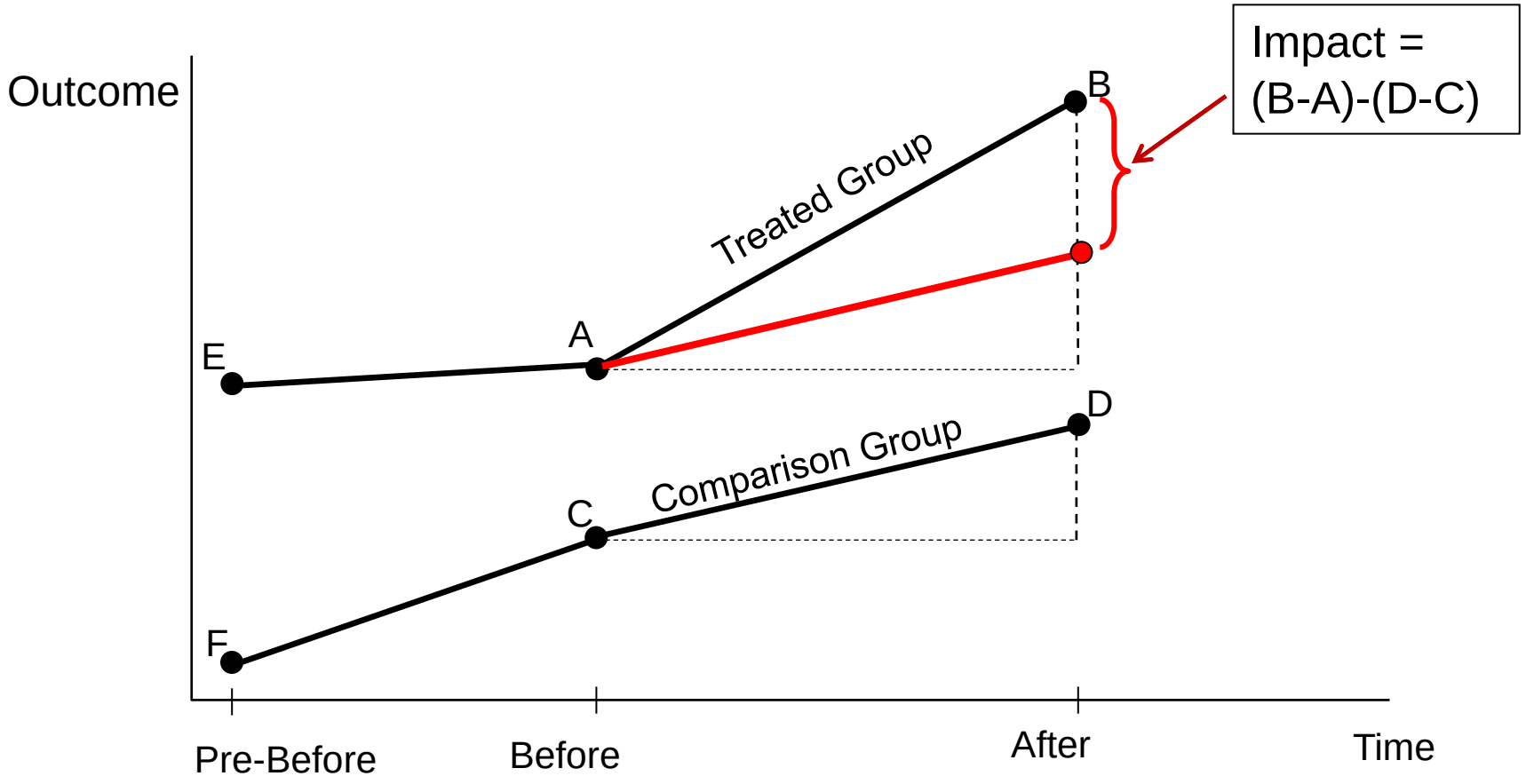
You need prior data!

In this example, the “Parallel trends assumption” holds if: (A-E)=(C-F)

Problems: - Prior data may not be available

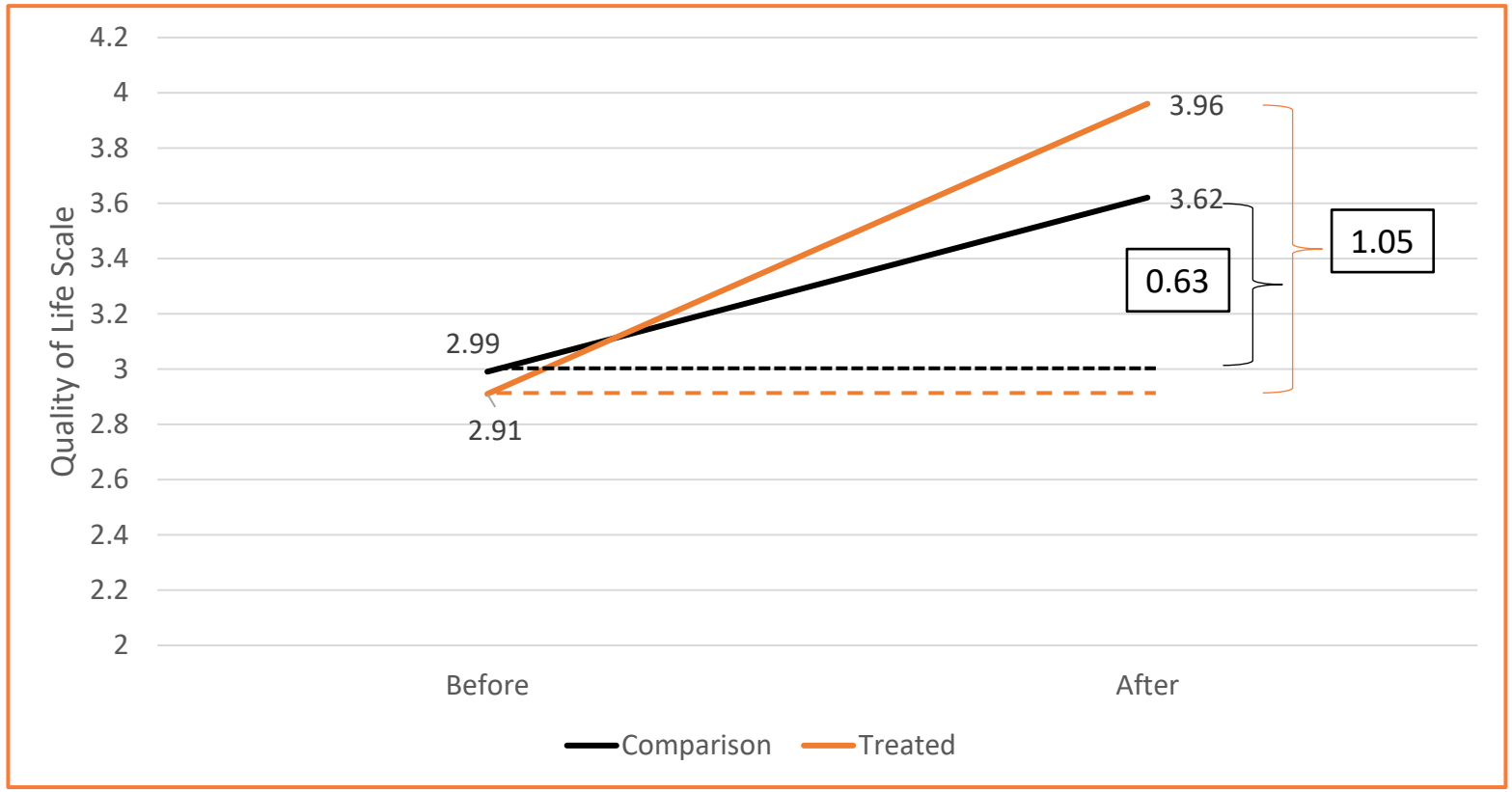
- Prior behavior is only an indication of future behavior.

DiD is less credible if past trends were different



In this case parallel trends did not occur in the recent past.

Impact of a cash transfer program on quality of life



	Before	After	Difference
Comparison	2.99	3.62	0.63
Treated	2.91	3.96	1.05
			1.05-0.63 = 0.42

DID Impact

How do we estimate DD in the regression framework?

- $Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$
  - Treat =1 if observation is in the treated group, 0 otherwise
  - Post=1 if observation is from post-intervention period (after), otherwise=0
  - Treat*Post is a dummy interaction!!! The two variables multiplied together
  - We can add covariates to this basic specification, focus on the essentials for now

$$Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$$

Treat=1 if treated group, 0 if comparison group
Post=1 if post-intervention period (after), 0 if pre-intervention

	Before (Post=0)	After (Post=1)	Difference
Comparison (Treat=0)	Treat=0, Post=0 B_0	Treat=0, Post=1 $B_0 + B_2$	
Treated (Treat=1)	Treat=1, Post=0 $B_0 + B_1$	Treat=1, Post=1 $B_0 + B_1 + B_2 + B_3$	

$$Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$$

Treat=1 if treated group, 0 if control group
Post=1 if post-intervention period (after), 0 if pre-intervention

	Before (Post=0)	After (Post=1)	Difference
Comparison (Treat=0)	Y = B0	Y = B0 + B2	B2
Treated (Treat=1)	Y = B0 + B1	Y = B0 + B1 + B2 + B3	B2+B3
Difference	B1	B1 + B3	B3

DID Impact

Questions

1. What coefficient(s) give(s) the time trend?
2. What coefficient(s) give(s) the difference between the two groups before the intervention?
3. How do you interpret B_0?

B_2

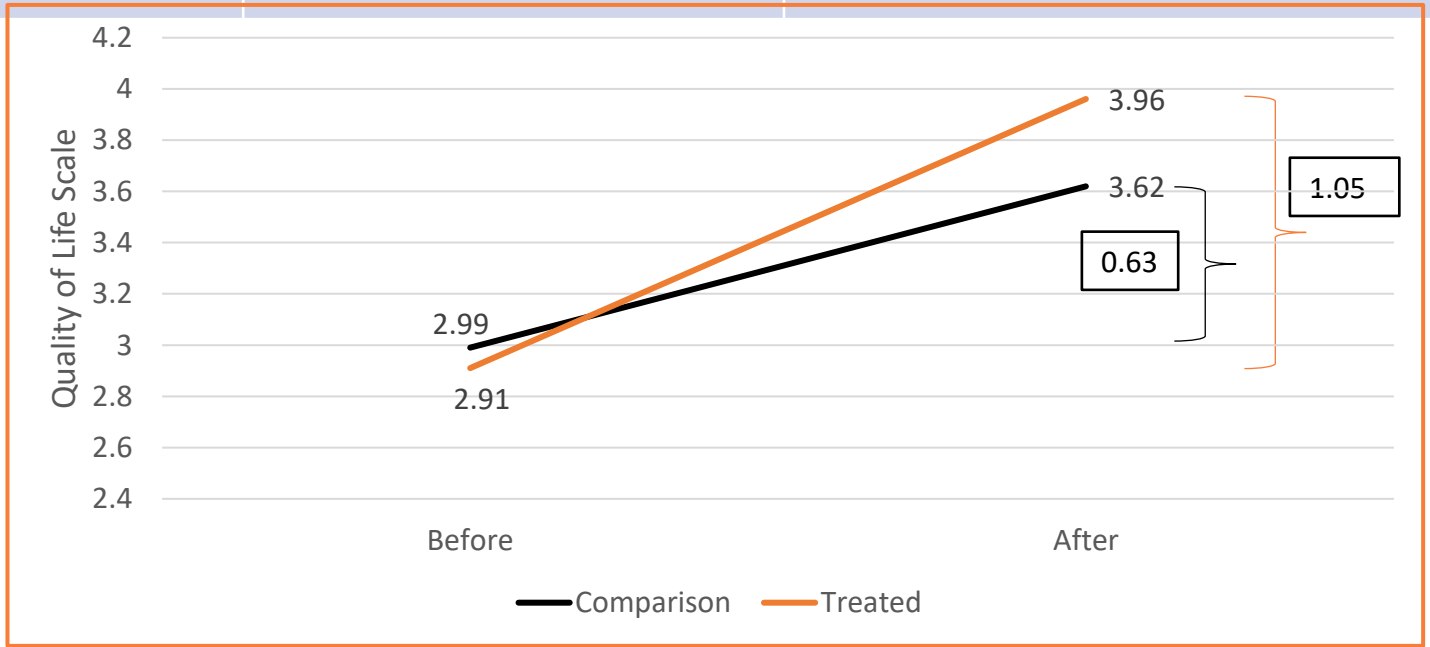
B_1

$$Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$$

$$Y = 2.99 \quad -0.08*(Treat) + 0.63*(Post) + 0.42*(Treat*Post)$$

$B_0 = 2.99$ $B_1 = -0.08$ $B_2 = 0.63$ $B_3 = 0.42$

	Before (Post=0)	After (Post=1)	Difference
Comparison (Treat=0)	**Y=2.99 (B_0)**	**Y=2.99 + 0.63 ($B_0 + B_2$)**	0.63
Treated (Treat=1)	**Y = 2.99 + -0.08 ($B_0 + B_1$)**	**Y = 2.99+ -0.08 + .63+.42 ($B_0 + B_1 + B_2 + B_3$)**	0.63+ 0.42 = 1.05
	-0.08		0.42



$$Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$$

$$Y = 2.99 \quad -0.08*(Treat) + 0.63*(Post) + 0.42*(Treat*Post)$$

```
. regress swb_avg treat post DD
```

Source	SS	df	MS
Model	1130.773	3	376.924334
Residual	6209.03454	5,123	1.21199191
Total	7339.80754	5,126	1.43187818

Number of obs	=	5,127	
F(3, 5123)	=	311.00	
Prob > F	=	0.0000	
R-squared	=	0.1541	
Adj R-squared	=	0.1536	
Root MSE	=	1.1009	

Is the impact statistically significant?

swb_avg	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
treat	-.0842445	.0462952	-1.82	0.069	-.175003	.0065139
post	.6262741	.0536708	11.67	0.000	.5210565	.7314918
DD	.4227154	.0654844	6.46	0.000	.294338	.5510928
_cons	2.996556	.0379397	78.98	0.000	2.922178	3.070934

B1
B2
B3
B0

Impact of cash transfer program in Zimbabwe on number of key assets owned
Can you write down the regression equation?

	Before (Post=0)	After (Post=1)	Difference
Comparison (Treat=0)	2.97 B_0	3.21	0.24 B_2
Treated (Treat=1)	2.66	3.33	0.67
	$B_1 = -0.31$		$B_3 = 0.43$

$$Y = B_0 + B_1*(Treat) + B_2*(Post) + B_3*(Treat*Post) + e_i$$

$$Y = 2.97 - 0.31*(Treat) + 0.24*(Post) + 0.43*(Treat*Post)$$

B_0 is the mean of the comparison group in the before period;
B_1 is the difference between the two groups in the before period;
B_2 is the time trend;

$$Y = 2.97 - 0.31*(Treat) + 0.24*(Post) + 0.43*(Treat*Post)$$

Source	SS	df	MS	Number of obs	=	5,133
				F(3, 5129)	=	39.86
Model	421.534695	3	140.511565	Prob > F	=	0.0000
Residual	18080.3126	5,129	3.52511456	R-squared	=	0.0228
				Adj R-squared	=	0.0222
Total	18501.8473	5,132	3.60519237	Root MSE	=	1.8775

ownkey	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
treat	-.3118427	.0789312	-3.95	0.000	-.4665817	-.1571038
post	.2375297	.0915052	2.60	0.009	.0581405	.4169188
DD	.4328079	.1116309	3.88	0.000	.2139636	.6516522
_cons	2.973872	.0647039	45.96	0.000	2.847024	3.100719