MBA 753 : Causal Inference Methods for Business Analytics

Dr. Nivedita Bhaktha

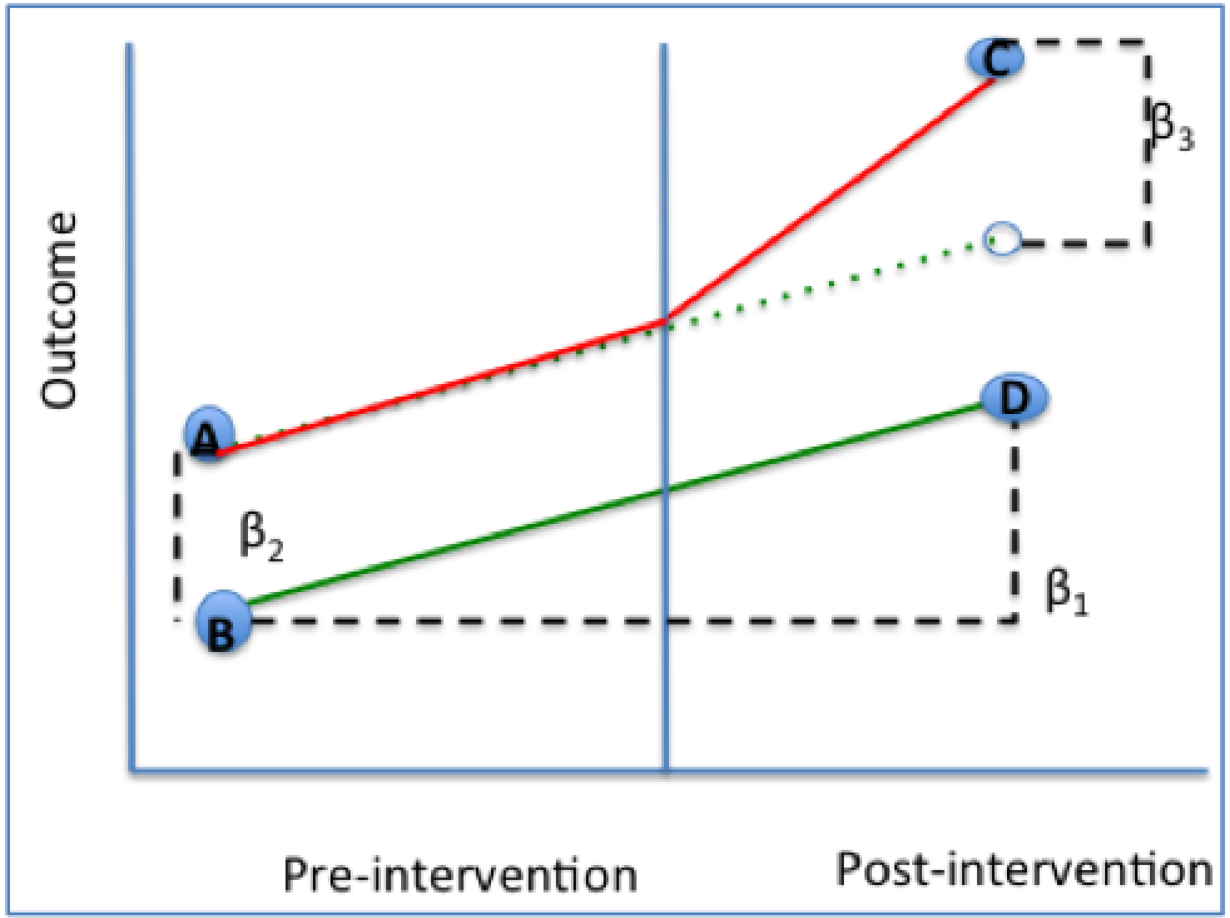
02.09.2024

Agenda

- Difference-in-Differences

Difference-in-Differences

DID Model



- 2 group – T & C
- parallel trends
- No spill over effect
- Time series data
- controls temporal effects
- Need not have comparable trt & C groups

Source: Jose Fernandez, 2024

DID - Tabular

	Before Change	After Change	Difference
Group 1 (Treat)	Y_{t1}	Y_{t2}	ΔY_t $= Y_{t2}-Y_{t1}$
Group 2 (Control)	Y_{c1}	Y_{c2}	ΔY_c $=Y_{c2}-Y_{c1}$
Difference			$\Delta\Delta Y$ $\Delta Y_t - \Delta Y_c$

DID Estimate

- Basic idea of DID estimate from the tabular form

Diff-in-Diff estimate = (Treatment_post - Treatment_pre) - (Control_post - Control_pre)

Exam scores

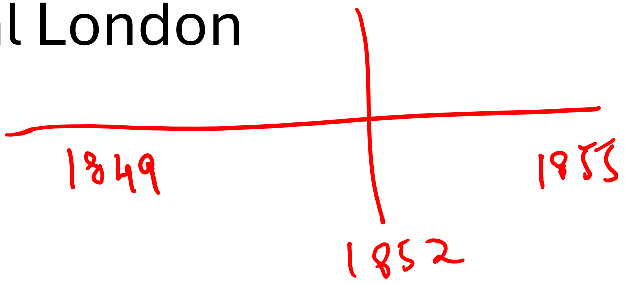
	Pre	Post	Difference
Treatment	50	85	35
Control	35	55	20
Difference	15	30	**15**

Increase in scores due to the intervention for the treated group

Example – John Snow’s Experiment

- 1849: London’s worst cholera epidemic claims 14,137 lives
  - Two companies supplied water to much of London: the Lambeth Waterworks Co. and the Southwark and Vauxhall Water Co.
    - Both got their water from the Thames
  - John Snow believed cholera was spread by contaminated water
- 1852: Lambeth Waterworks moved their intake upriver
  - Everyone knew that the Thames was dirty below central London
- 1853: London has another cholera outbreak

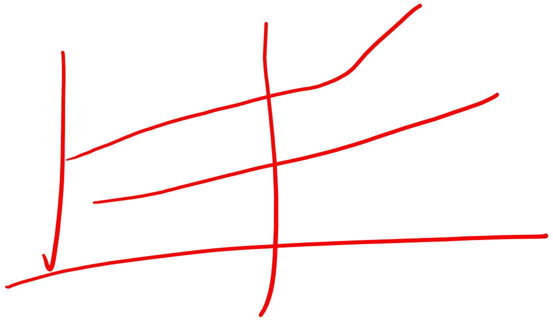
L → U



- Are Lambeth Waterworks customers less likely to get sick?

Example – John Snow's Experiment

- Mortality data showed that very few cholera deaths were reported in areas of London that were only supplied by the Lambeth Waterworks
- Snow hired John Whiting to visit the homes of the deceased to determine which company (if any) supplied their drinking water
- Using Whiting's data, Snow calculated the death rate
  - C Southwark and Vauxhall: 71 cholera deaths/10,000 homes
  - T Lambeth: 5 cholera deaths/10,000 homes
- Southwark and Vauxhall responsible for 286 of 334 deaths
  - Southwark and Vauxhall moved their intake upriver in 1855



DID Regression

$$Y = \beta_0 + \beta_1 * T + \beta_2 * P + \beta_3 * T * P + \epsilon$$

DID

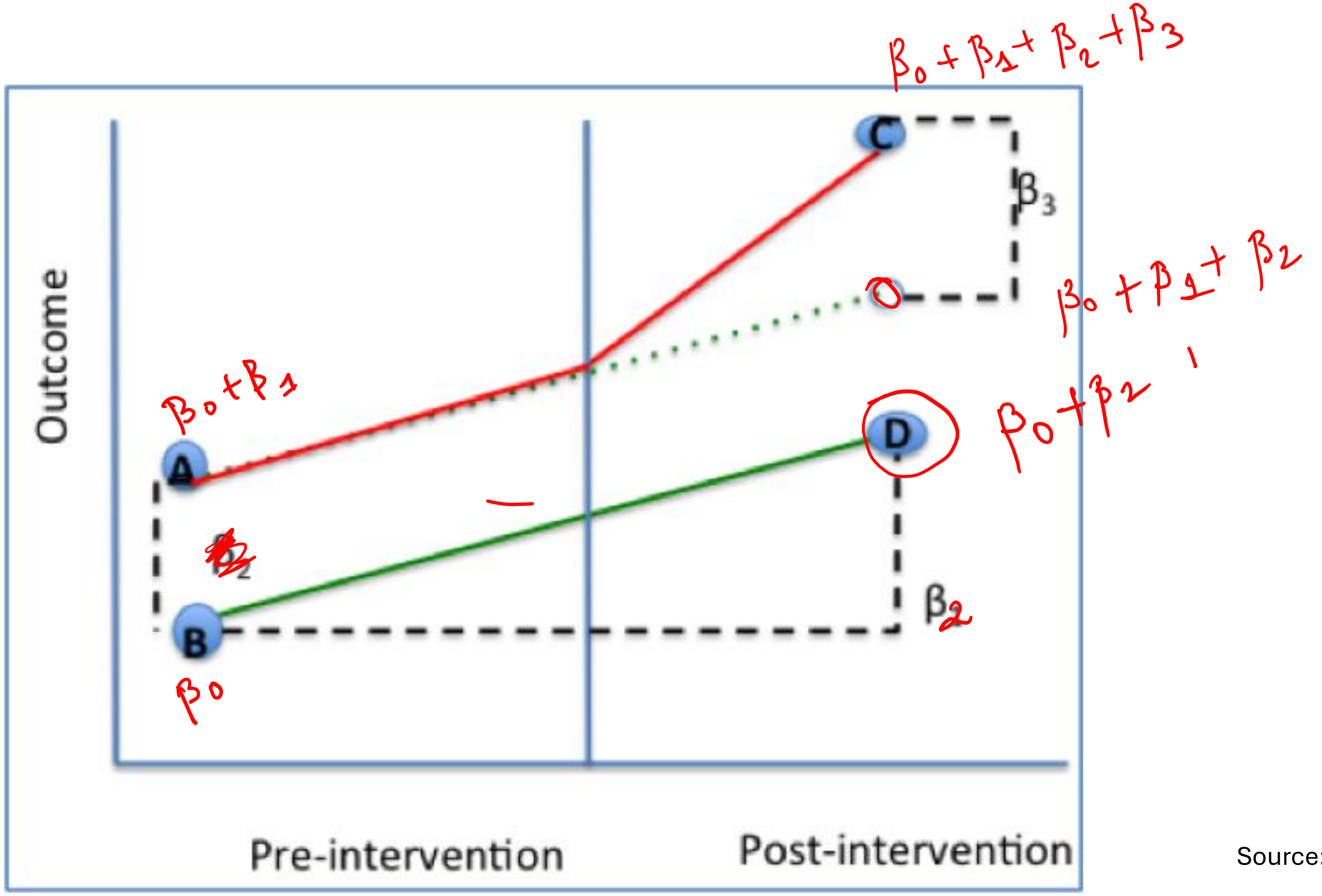
where

- Y is the outcome variable
- T is the dummy variable indicating the treatment (=1) and control (=0) group;
- P is a dummy variable indicating pre (=0) and post (=1) treatment;
- T*P is a dummy variable indicating whether the outcome was observed in the treatment group AND it was observed after the intervention (=1)

DID Regression Coefficients in Tabular Form

	Before Change	After Change	Difference
Group 1 (Treat)	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\Delta Y_t = \beta_2 + \beta_3$
Group 2 (Control)	β_0	$\beta_0 + \beta_2$	$\Delta Y_c = \beta_2$
Difference			$\Delta\Delta Y = \beta_3$

DID Regression Coefficients



Source: Columbia Public Health

Y — T & P

$Y = 35 + 15\,T + 20\,P + 15\,TP + \varepsilon$

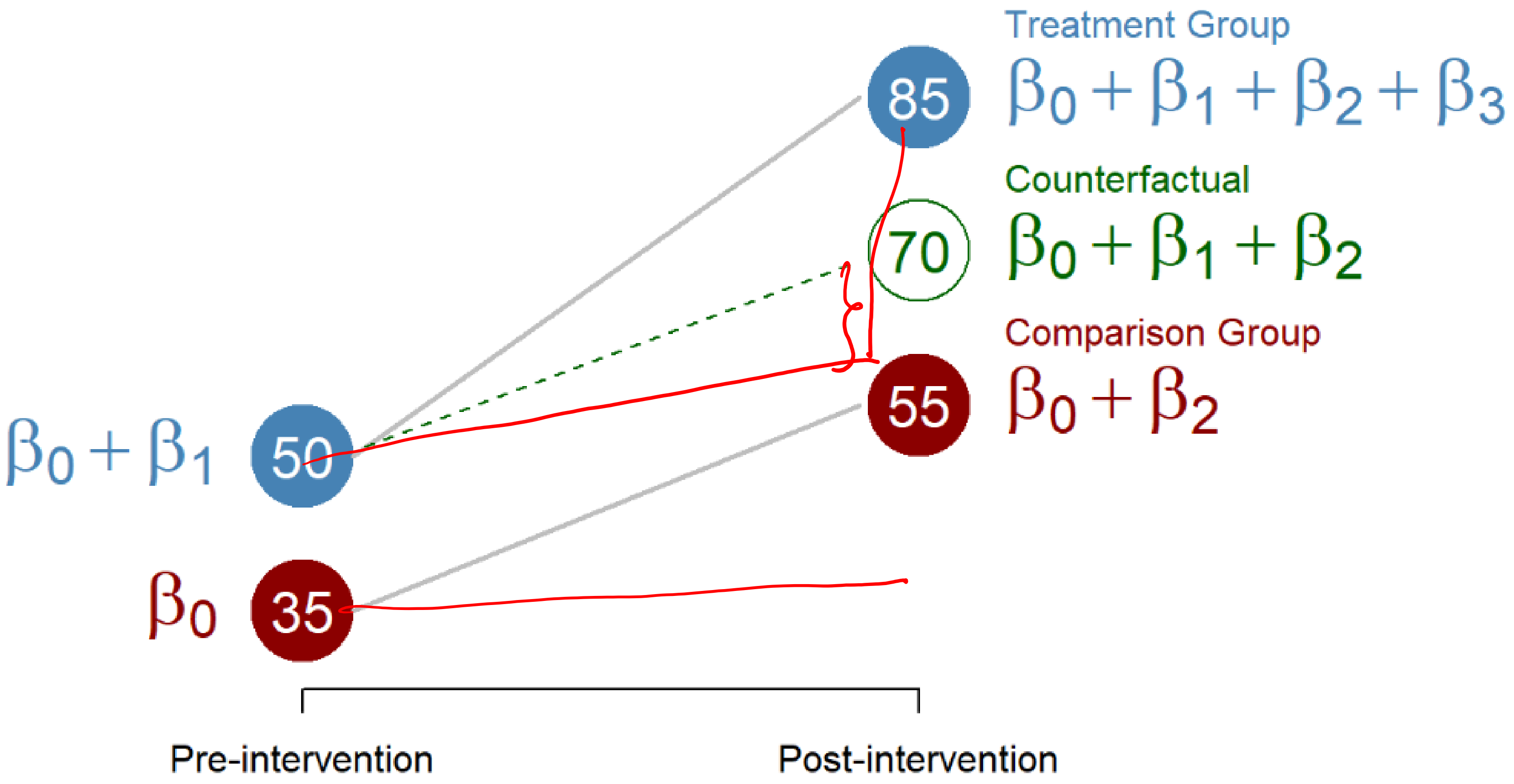
DID Regression Interpretation

Subject	Outcome	Treatment	Post	Treatment * Post
1	74	1	1	1
1	46	1	0	0
2	96	1	1	1
2	54	1	0	0
3	50	0	1	0
3	30	0	0	0
4	60	0	1	0
4	40	0	0	0

	Dependent variable:
	Outcome
Intercept (B0) C grp pre Int	35.00^{***}
	(0.00)
Treatment (B1)	15.00^{***}
	(0.00)
Post-treatment (B2)	20.00^{***}
	(0.00)
Diff in Diff (B3)	15.00^{***}
	(0.00)

$y = 35 + 15T + 20P + 15TP + \varepsilon$

DID Regression Interpretation



$H_0: b_i = 0$ a

$H_0: \beta_1 = 0$

$H_0: \beta_3 = 0$

DID Hypothesis Testing

β	HYPOTHESES
b_0	Is the average outcome of the control group before the treatment $\neq 0$?
b_1	Is the difference between the control and treatment group before the treatment $\neq 0$?
b_2	Is the difference between the average outcome ~~of the control group~~ before and after the treatment $\neq 0$?
b_3	Difference in difference estimator. Does the treatment have an impact?

DID Example

- The U.S. Dept. of Housing and Urban Development, along with state and local governments, subsidize several housing projects to create affordable living spaces. This policy has become increasingly important in the past yeas, as rent and housing costs increase in several US cities. However, there are several critics to housing programs, including the belief that the development of affordable housing might cause a decline in housing value in nearby neighborhoods.
- Research question: Do affordable housing projects reduce prices of nearby houses?

T
AHP present

C
AHP not present

House value before & after AHP

Results Discussion

1. Is the comparison group in Time = 1 different from zero?
2. Do houses in the treatment group cost more than houses in the control group?
3. Do the prices of houses in the control group increase over time?
4. Does the construction of an affordable housing site change the price of the nearby houses?

	Dependent variable:
	Housing prices
Intercept (B0)	216,673***
	(3,742)
Treatment group (B1)	16,461***
	(5,292)
Post-treatment (B2)	35,231***
	(5,292)
Diff in Diff (B3)	34,066***
	(7,484)

Example – Card & Krueger, 1994

- Investigated the relationship between a rise in minimum wage and employment.
- Theory: Increases in the minimum wage lead to a reduction in the employment
- They applied a difference-in-difference the design to look at two groups of fast-food restaurants:
  - fast-food restaurants in New Jersey where the minimum wage increased from 4.25$ to 5.05$ per hour (treatment group) in 1992 AND
  - fast-food restaurants in Pennsylvania where the minimum wage did not increase (control group).

Example – Card & Krueger, 1994

- They collected employment data before and after the minimum wage was approved.
- Research question: Do increases in the minimum wage affect employment?
- Theory: An increase in the minimum wage is negatively correlated with employment.

Card & Krueger, 1994 - Codebook

Variable name	Description
ID	Unique identifier for fast food
Treatment	Pre-treatment (=0) and post-treatment (=1)
Group	1 if NJ (treatment); 0 if PA (Control)
Empl	# of full time employees
C.Owned	If owned by a company (=1) or not (=0)
Hours.Opening	Number hours open per day
PA2	Easton and other PA areas
Shore	New Jersey Shore

Soda	Price of medium soda, including tax
Fries	price of small fries, including tax
Chain	1 = BK, 2 = KFC, 3 = Roys, 4 = Wendys
SouthJ	South New Jersey
CentralJ	Central New Jersey
NorthJ	North New Jersey
PA1	Northeast suburbs of Philadelphia

DID Issues

- Standard Errors in DID
  - Practical applications of DID strategies use data from many years: not just 1 pre and 1 post period
  - The variables of interest only vary at a group level and outcome variables are often serially correlated

- Solution:
  - Clustering standard errors at the group level and time-level
  - Bootstrapped SE

DID Issues

- Threats to validity
  - Non-parallel trends
    - Ashenfelter dip: earnings often fall just prior to entering a training program, which complicates measurement of treatment effect
  - Other simultaneous treatment/intervention
  - Functional form dependence
    - Assumption: DID regression equation is linear
    - Matching DID
  - Multiple treatment times
    - Treatment occurs at different times for treated

- Robustness check for DID design: Falsification test using placebo outcome

Recap

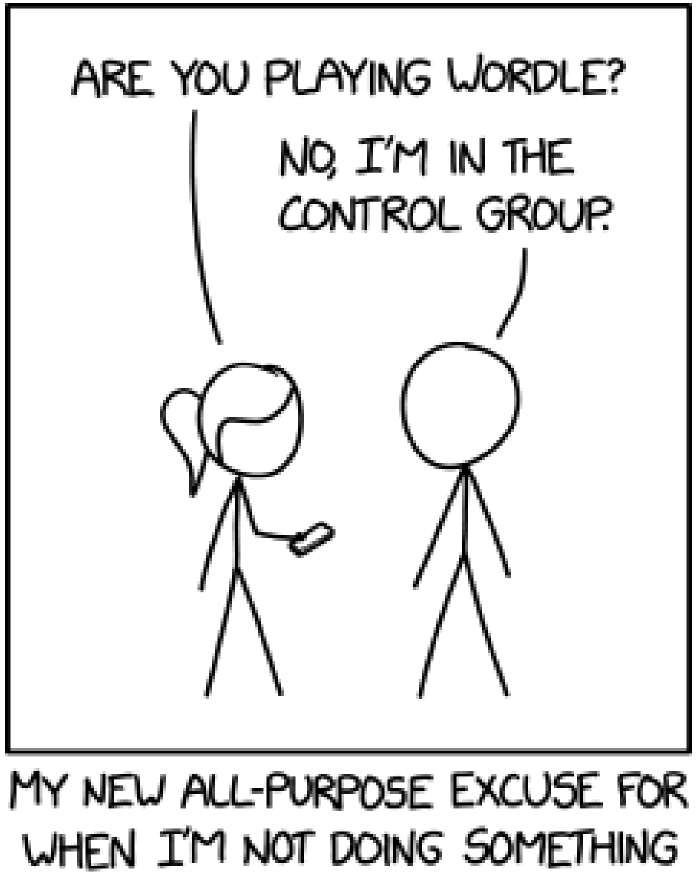
Summary

- DID is a widely used alternative to RCT
  - DID Setup: two or more groups, with units observed in two or more periods.
- Often need more than 2 periods to test:
  - Pre-treatment trends for treatment and control to see if parallel trends assumption is plausible or not
- Use robustness checks
- Objectives achieved:
  - Can understand the advantages of DID design
  - Can interpret the regression coefficients of DID design
  - Can understand the issues of DID design

References

- Card, David, and Alan B Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *Am. Econ. Rev.* 84 (4): 772–93.
- Scott Cunningham, Causal Inference: The Mix Tape, Yale University Press.
- Cook, T. D., & Campbell, D. T. (2007). *Experimental and quasi-experimental designs for generalized causal inference*.
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Thank You ☺


ARE YOU PLAYING WORDLE?
NO, I'M IN THE CONTROL GROUP.
MY NEW ALL-PURPOSE EXCUSE FOR WHEN I'M NOT DOING SOMETHING