Difference-in-Differences

JAN 27, 2020

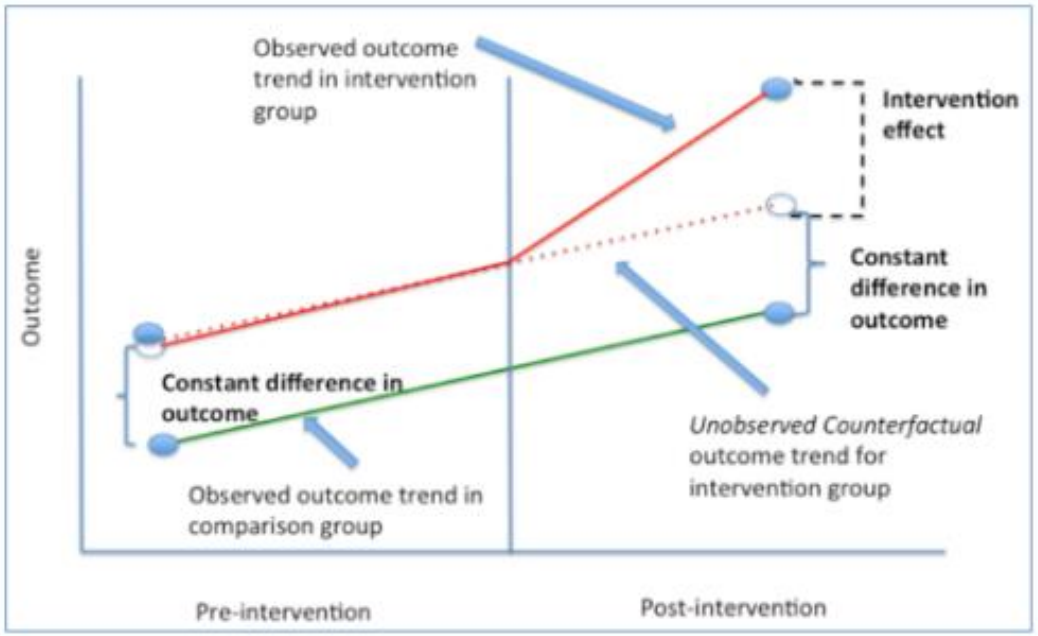
Difference-in-Differences (DID)

What is difference-in-differences?

- Quasi-experimental analysis
- Compares changes over time between intervention and control groups
- Also known as pre-post with control
- Need at least two groups and at least two time points

Why use difference-in-differences?

- More rigorous than a pre-post analysis
  - Accounts for previous trends in the outcome (e.g., if outcome is improving over time from factors not related to the intervention, this improvement may be incorrectly attributed to the intervention)
- Accounts for differences between intervention and control group that are constant over time



https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

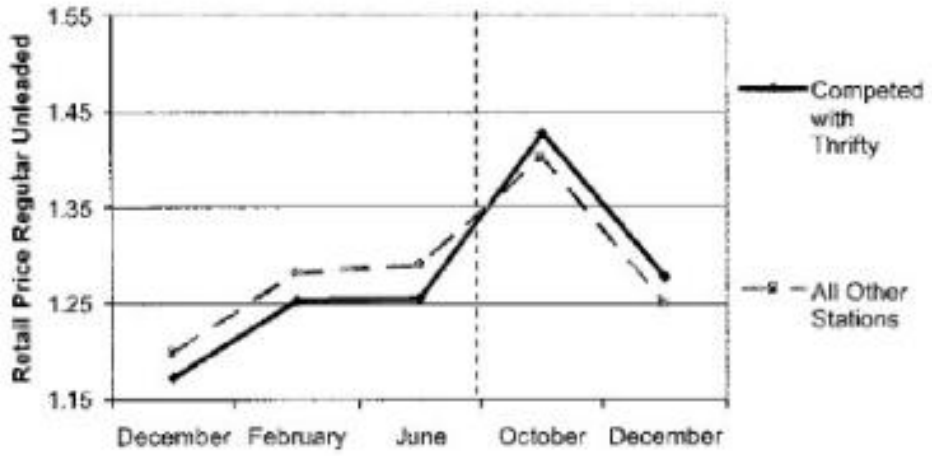
Difference-in-Differences (DID)

Parallel trends assumption

- In the absence of the intervention, the difference in the outcome between treated and controls would be constant over time (i.e, outcomes would move in parallel between groups).
- This can be tested by comparing trends in the pre-period. A difference-in-differences analysis using two pre-intervention time periods would not be significant. Though some have argued against this approach, it is probably sufficient here.

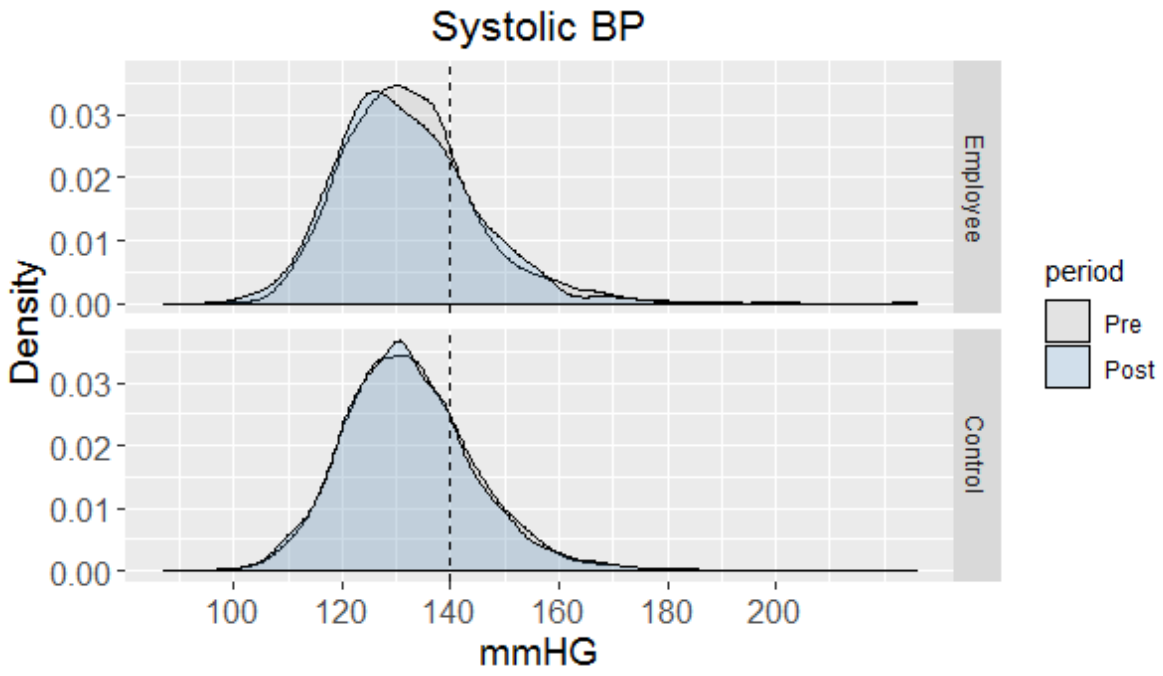
Stable Unit Treatment Value Assumption (SUTVA)

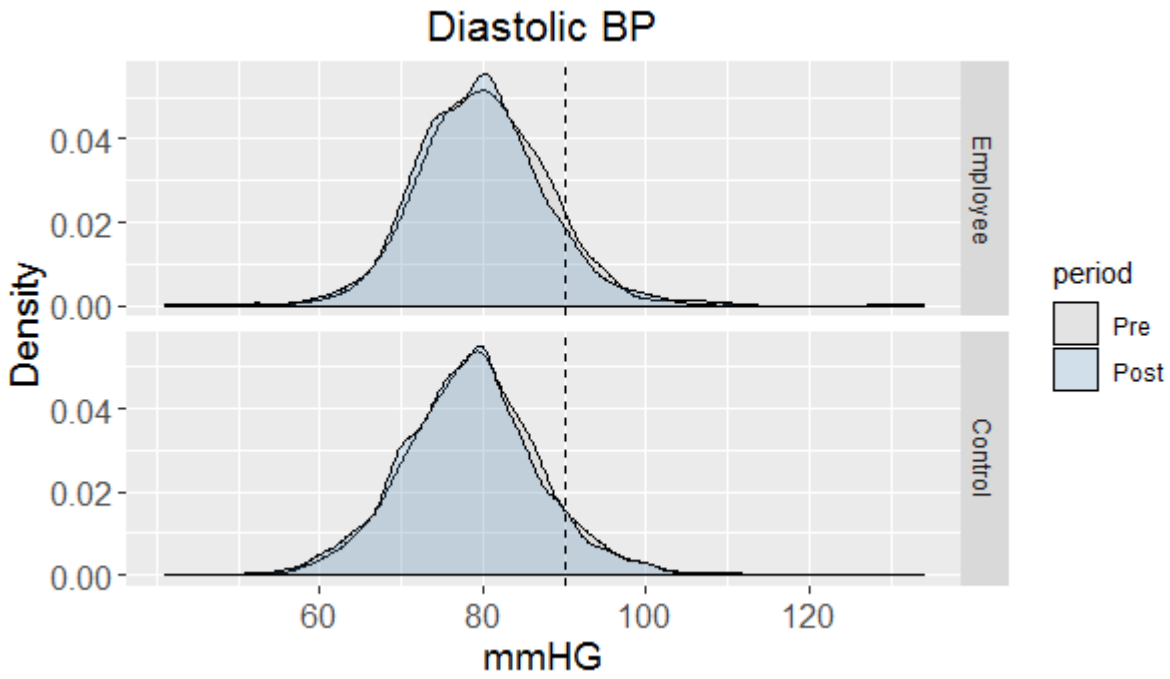
- The composition between groups is the same over time (i.e., no changes in populations before and after the intervention). Note you could adjust for this using multivariable regression.
- There are no spillover effects between groups



https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Hypertension (Results)

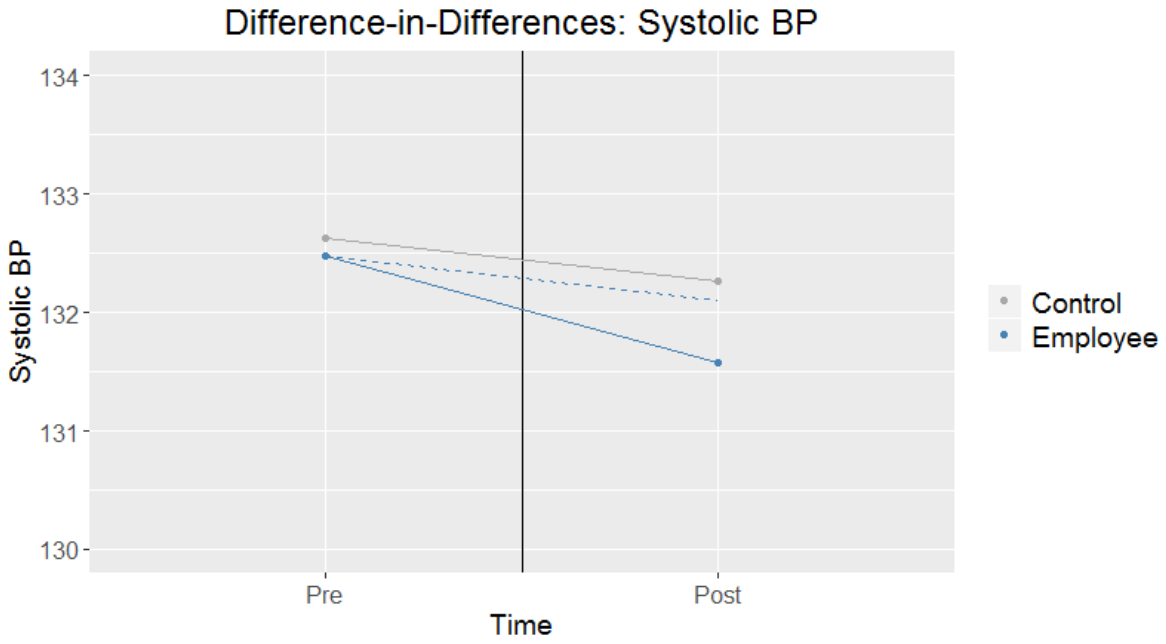




Hypertension (Difference-in-Differences)

Blood Pressure Difference-in-Differences

	Coefficient	SE	95% CI	p-value
Systolic BP				
Employee	-0.158	0.410	-0.963 to 0.646	0.700
Post-Period	-0.365	0.238	-0.832 to 0.102	0.126
Emp*Post (DID)	**-0.522**	**0.585**	**-1.669 to 0.624**	**0.372**
Diastolic BP				
Employee	1.374	0.266	0.853 to 1.896	< 0.001
Post-Period	-0.606	0.160	-0.921 to -0.291	< 0.001
Emp*Post (DID)	**-0.014**	**0.375**	**-0.749 to 0.722**	**0.971**



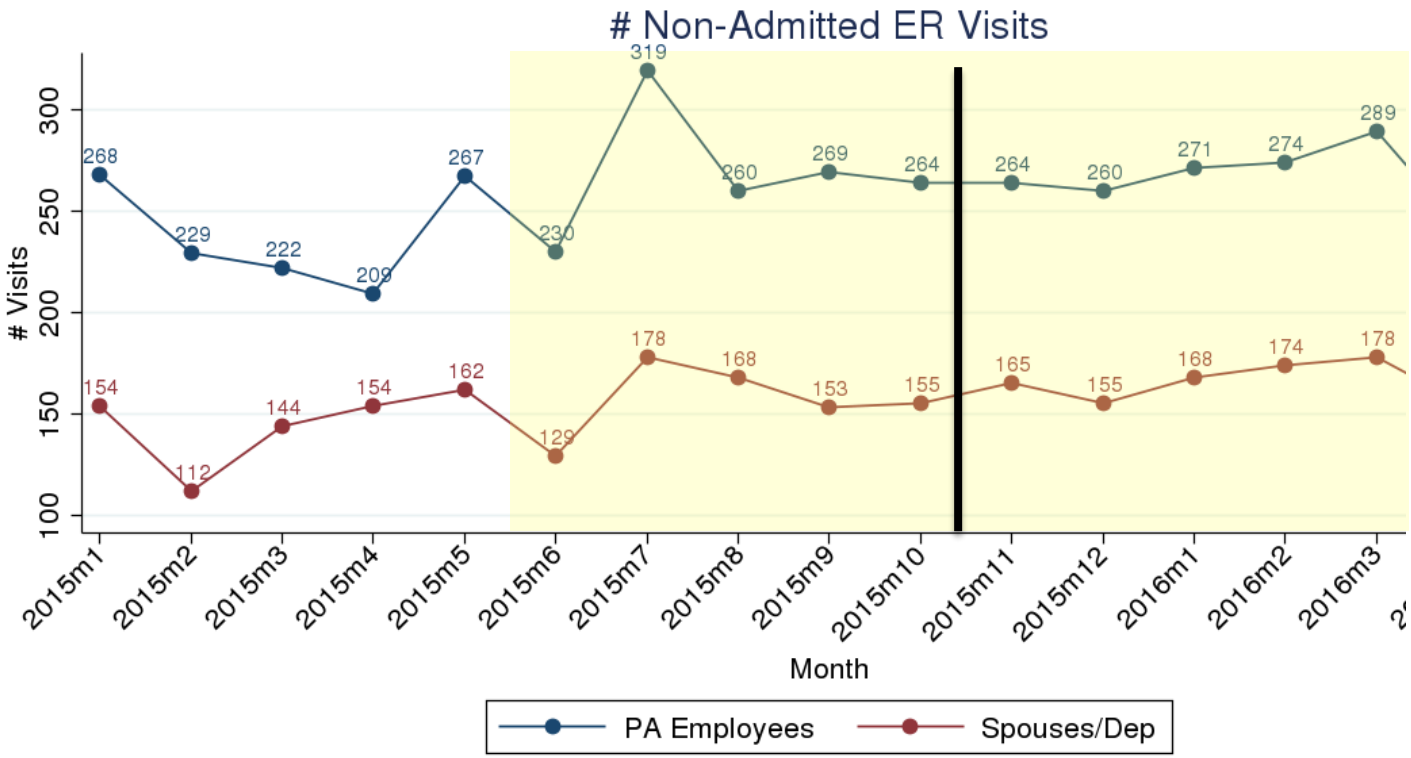
Example R Code

```
bp <- c(132.47,131.58,132.63,132.26)
time <- c(0,1,0,1)
group <- c("Employee","Employee","Control","Control")

data <- data.frame(bp=bp,time=time,group=group) %>% mutate(time =
factor(time,levels=c("0","1"),labels=c("Pre","Post")))

ggplot(data,aes(time,bp,color=group)) +
  geom_point() +
  ylim(130,134) +
  scale_colour_manual(values=c("darkgrey","steelblue")) +
  geom_vline(xintercept = 1.5) +
  xlab("Time") +
  ylab("Systolic BP") +
  ggtitle("Difference-in-Differences: Systolic BP") +
  theme(legend.title=element_blank(),plot.title=element_text(hjust=0.5,size=18),axis.title.x=element_text(size=1
  4),axis.title.y=element_text(size=14),
        axis.text.x=element_text(size=12),axis.text.y=element_text(size=12),legend.text=element_text(size=14)) +
  geom_segment(aes(1,132.63,xend=2,yend=132.26),col="darkgrey",size=0.5) +
  geom_segment(aes(1,132.47,xend=2,yend=131.58),col="steelblue",size=0.5) +
  geom_segment(aes(1,132.47,xend=2,yend=132.10),col="steelblue",lty=2,size=0.5)
```

Difference-in-Differences

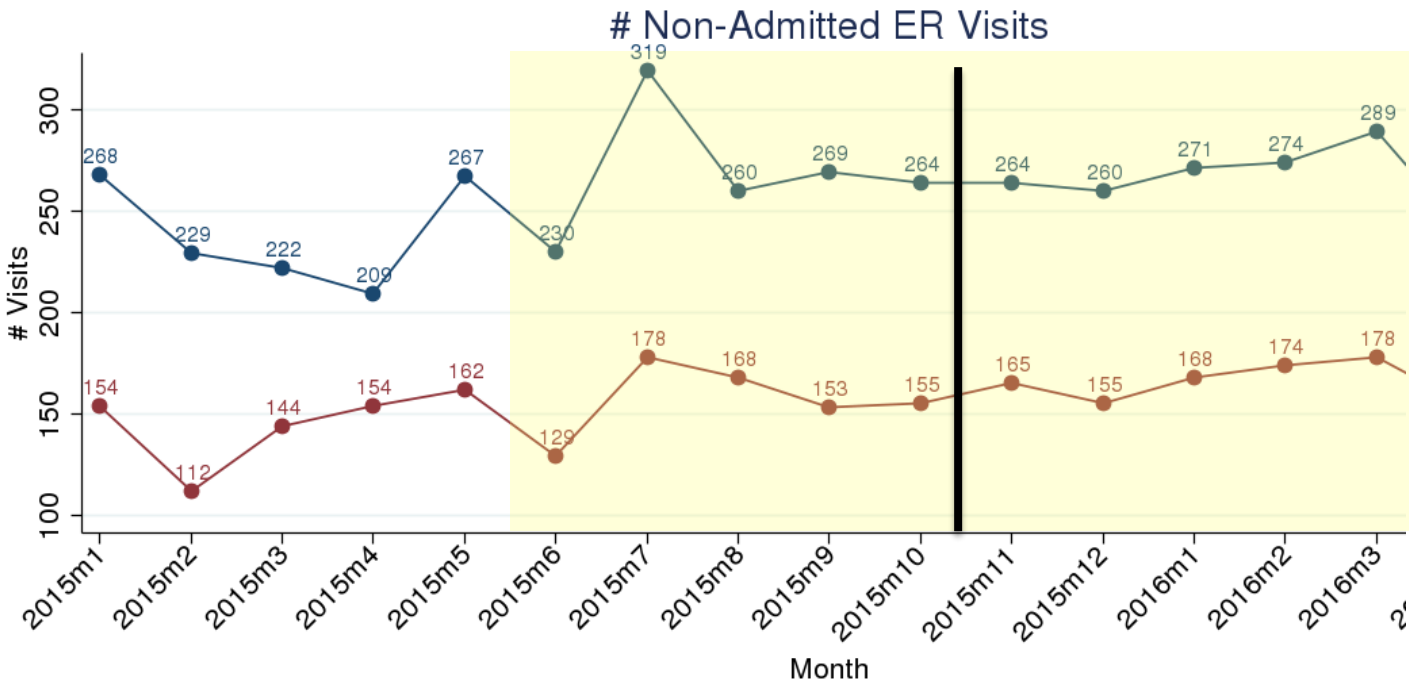


Assumes similar pre-trends

Small window = Less population variation

Raymond Lim (RJLIM@mail.med.upenn.edu)

Difference-in-Differences



	5 month Pre	5 month Post	Change
PA Employees	1342	1358	+16 (+1.2%)
Spouses/Dep>18	783	840	+57 (+7.3%)
			DiD = -6.1%

Raymond Lim (RJLIM@mail.med.upenn.edu)

Example R Code

```
# Estimating the DID estimator (using the multiplication method, no need to generate the interaction)

didreg1 = lm(y ~ treated*time, data = mydata)
summary(didreg1)
```

```
Call:
lm(formula = y ~ treated * time, data = mydata)

Residuals:
       Min          1Q      Median          3Q         Max
-9.768e+09 -1.623e+09  1.167e+08  1.393e+09  6.807e+09

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.581e+08  7.382e+08   0.485   0.6292
treated       1.776e+09  1.128e+09   1.575   0.1200
time          2.289e+09  9.530e+08   2.402   0.0191 *
treated:time -2.520e+09  1.456e+09  -1.731   0.0882 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.953e+09 on 66 degrees of freedom
Multiple R-squared:  0.08273, Adjusted R-squared:  0.04104
F-statistic: 1.984 on 3 and 66 DF,  p-value: 0.1249
```

https://www.princeton.edu/~otorres/DID101R.pdf