Photo credit: Dave DiCello

17-803 Empirical Methods

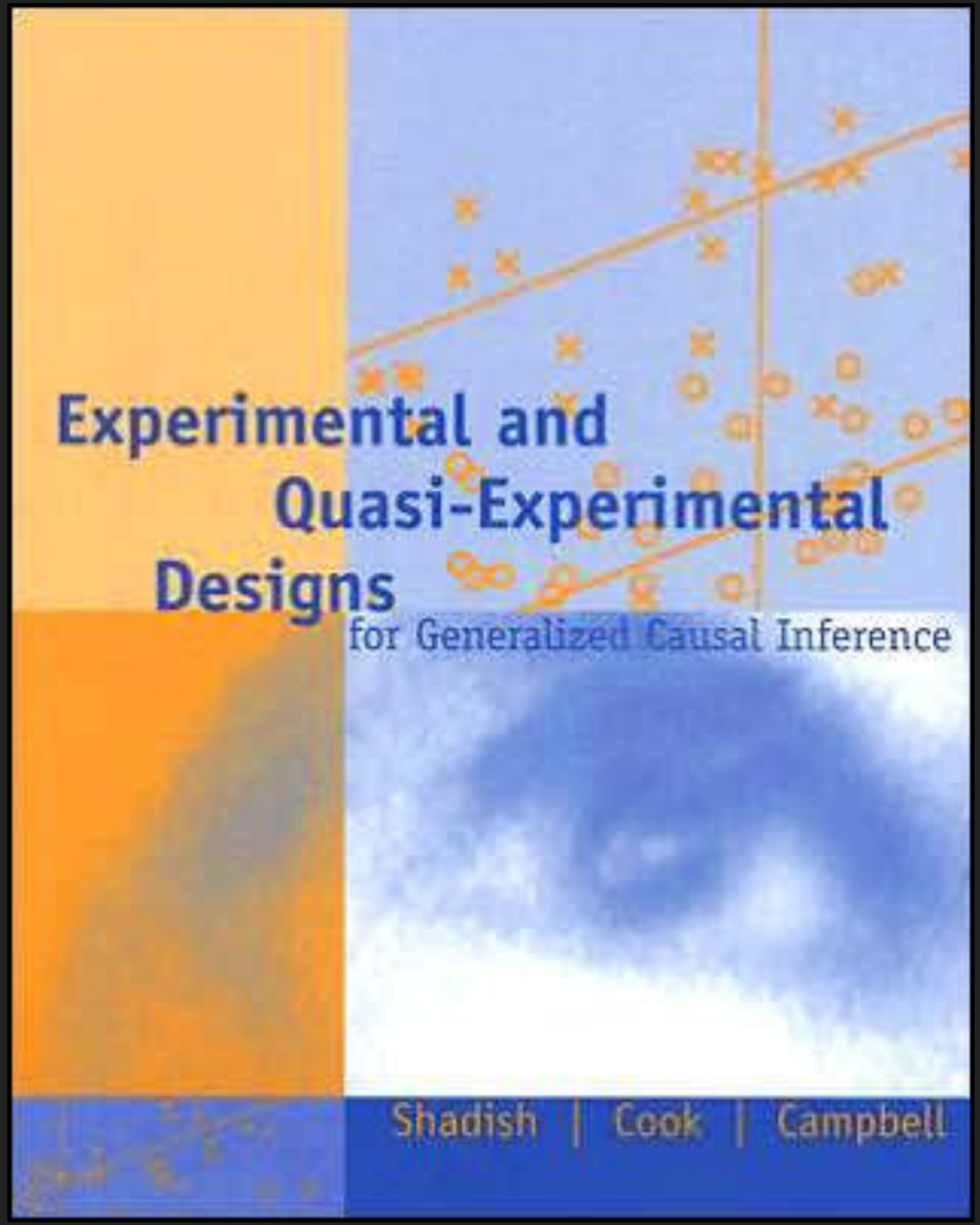
Bogdan Vasilescu, Institute for Software Research

Interrupted Time Series

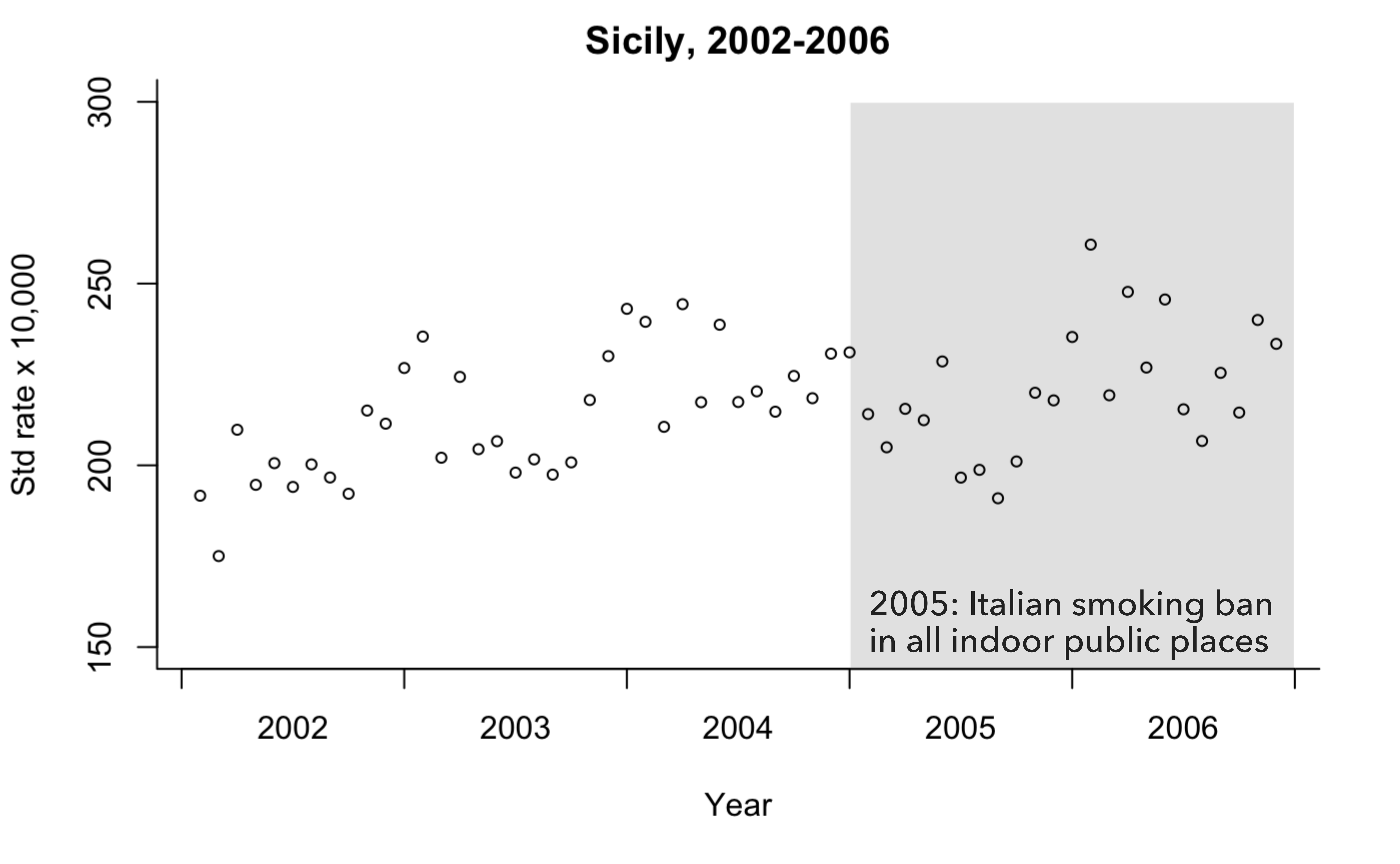
Thursday, April 8, 2021

Plan for Today

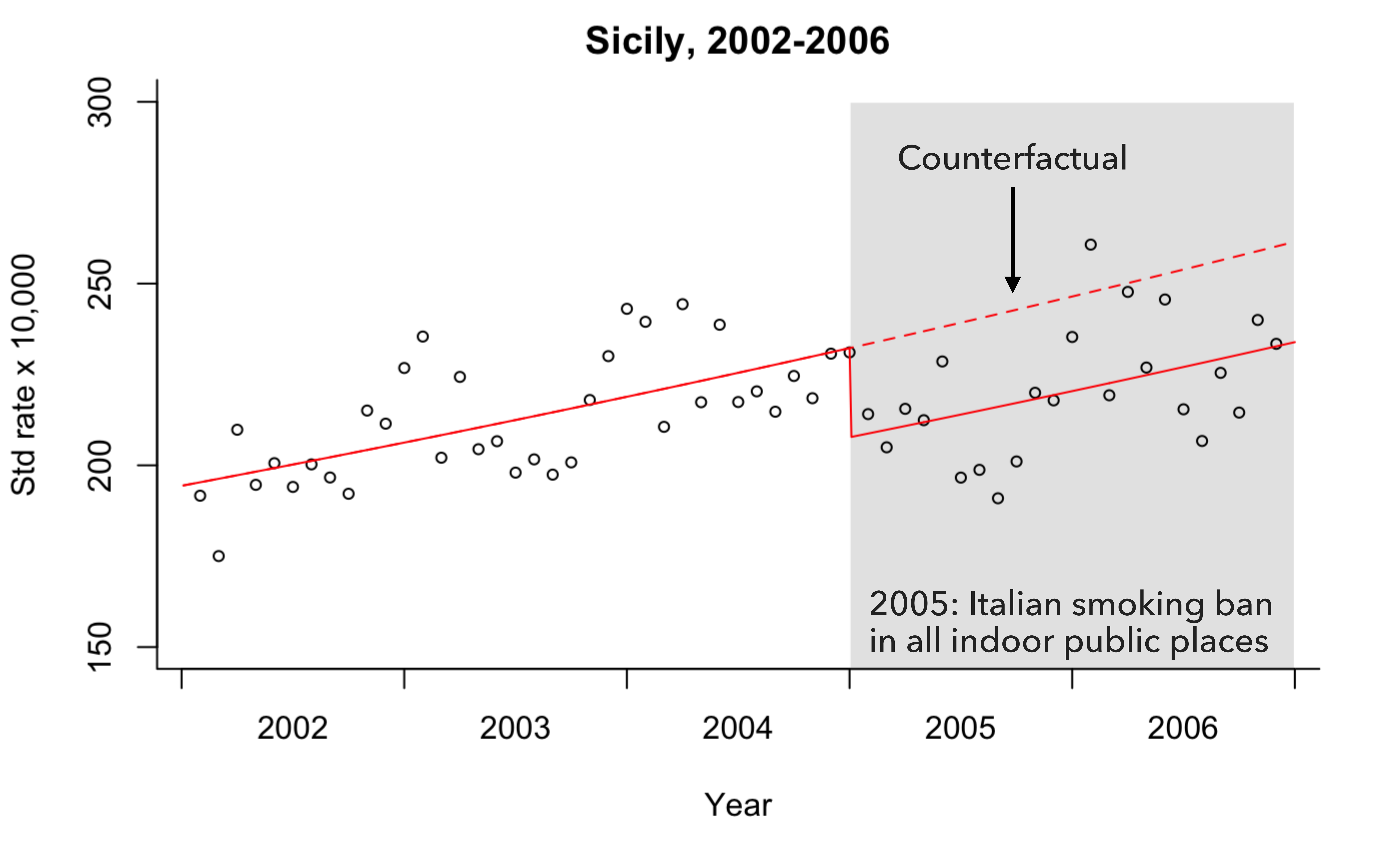
- Segmented regression of interrupted time series data
  - Mini lecture
  - A few examples



Hospital Admissions for Acute Coronary Events



Hospital Admissions for Acute Coronary Events

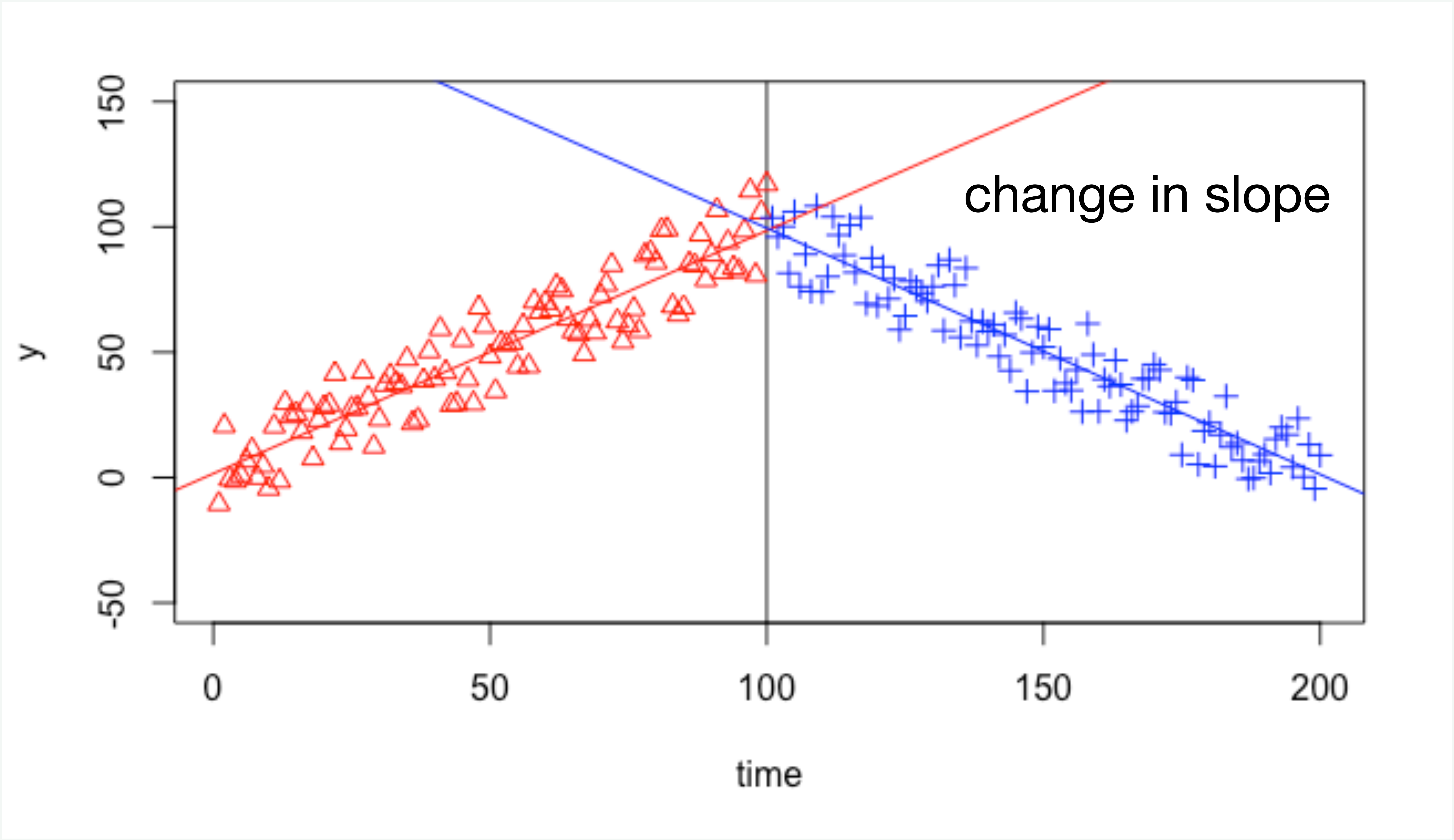


Interrupted Time Series Design

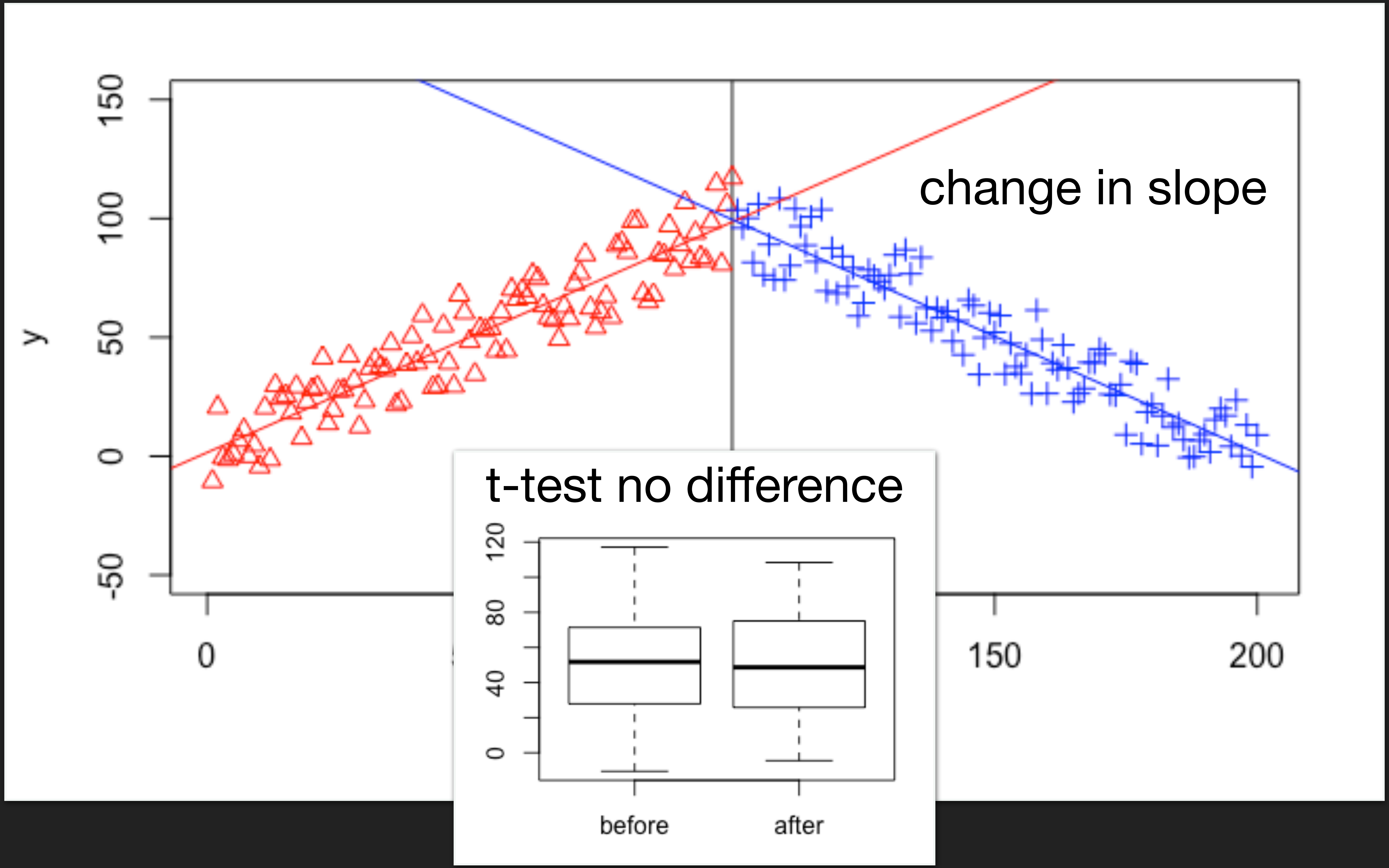
- One of the strongest quasi-experimental design to evaluate longitudinal effects of time-delimited interventions.
- How much did an intervention change an outcome of interest?
  - immediately and over time;
  - instantly or with delay;
  - transiently or long-term;
- Could factors other than the intervention explain the change?

Modeling 101

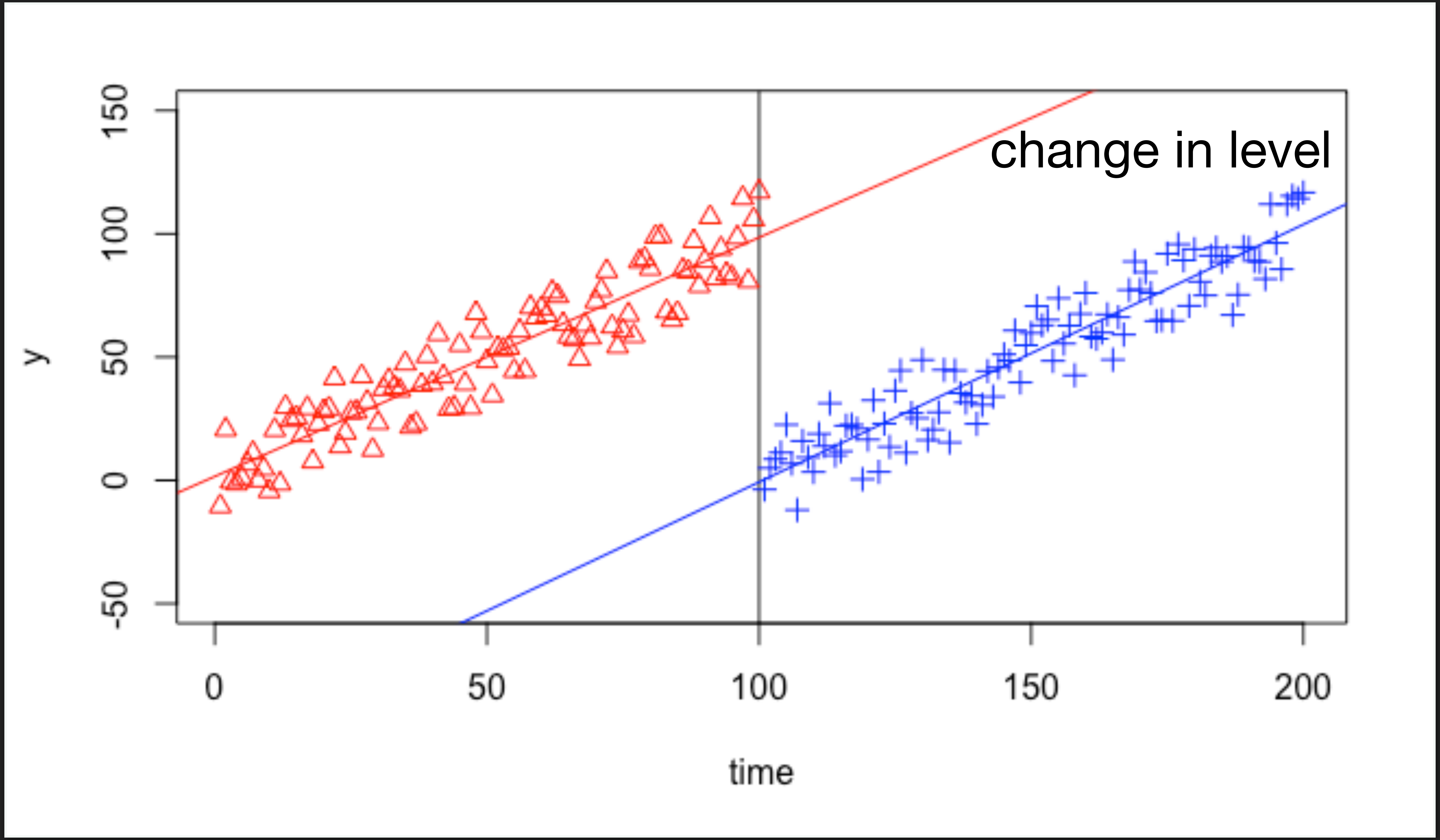
Evaluating the Effects of an Intervention



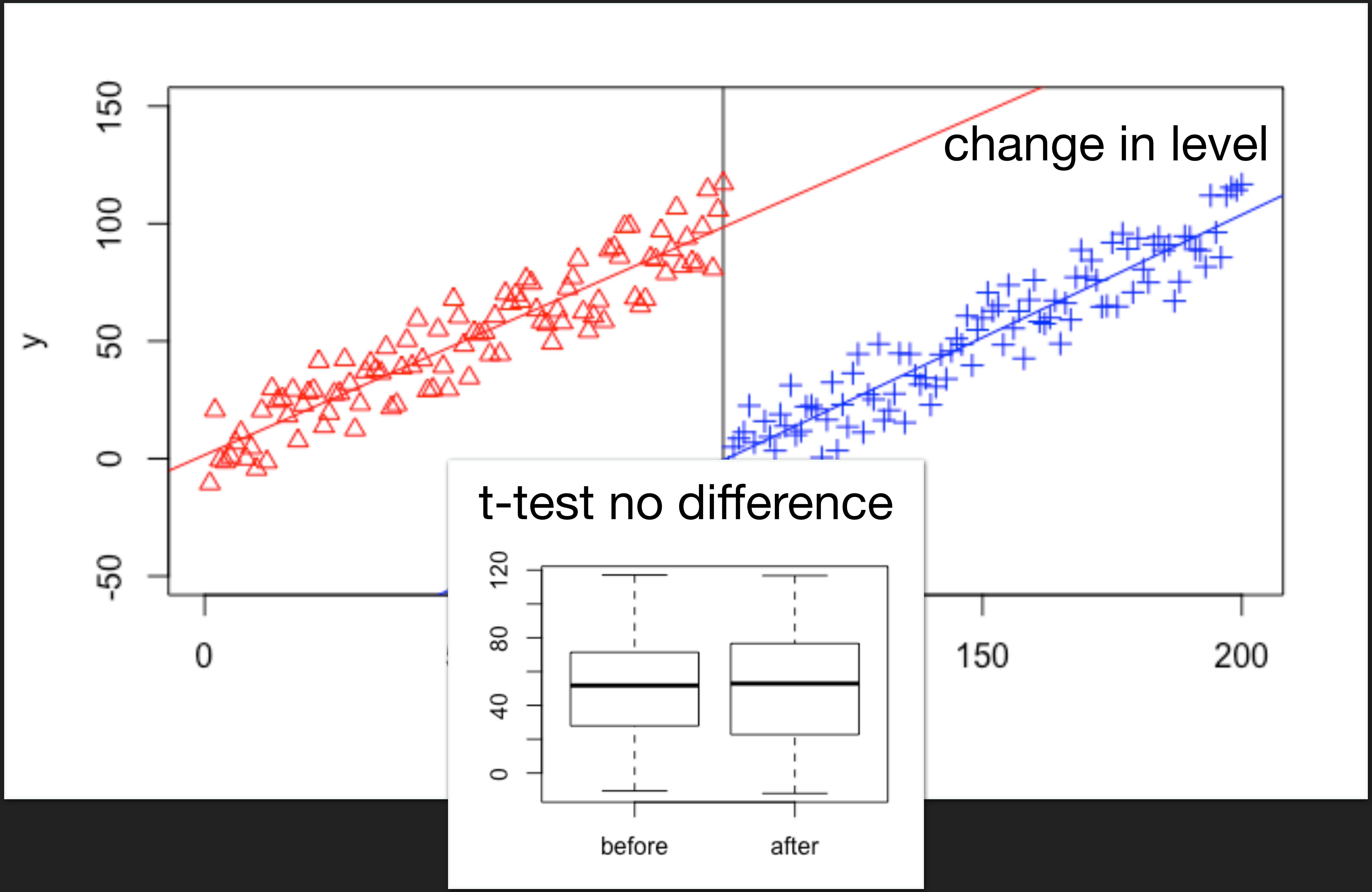
Evaluating the Effects of an Intervention



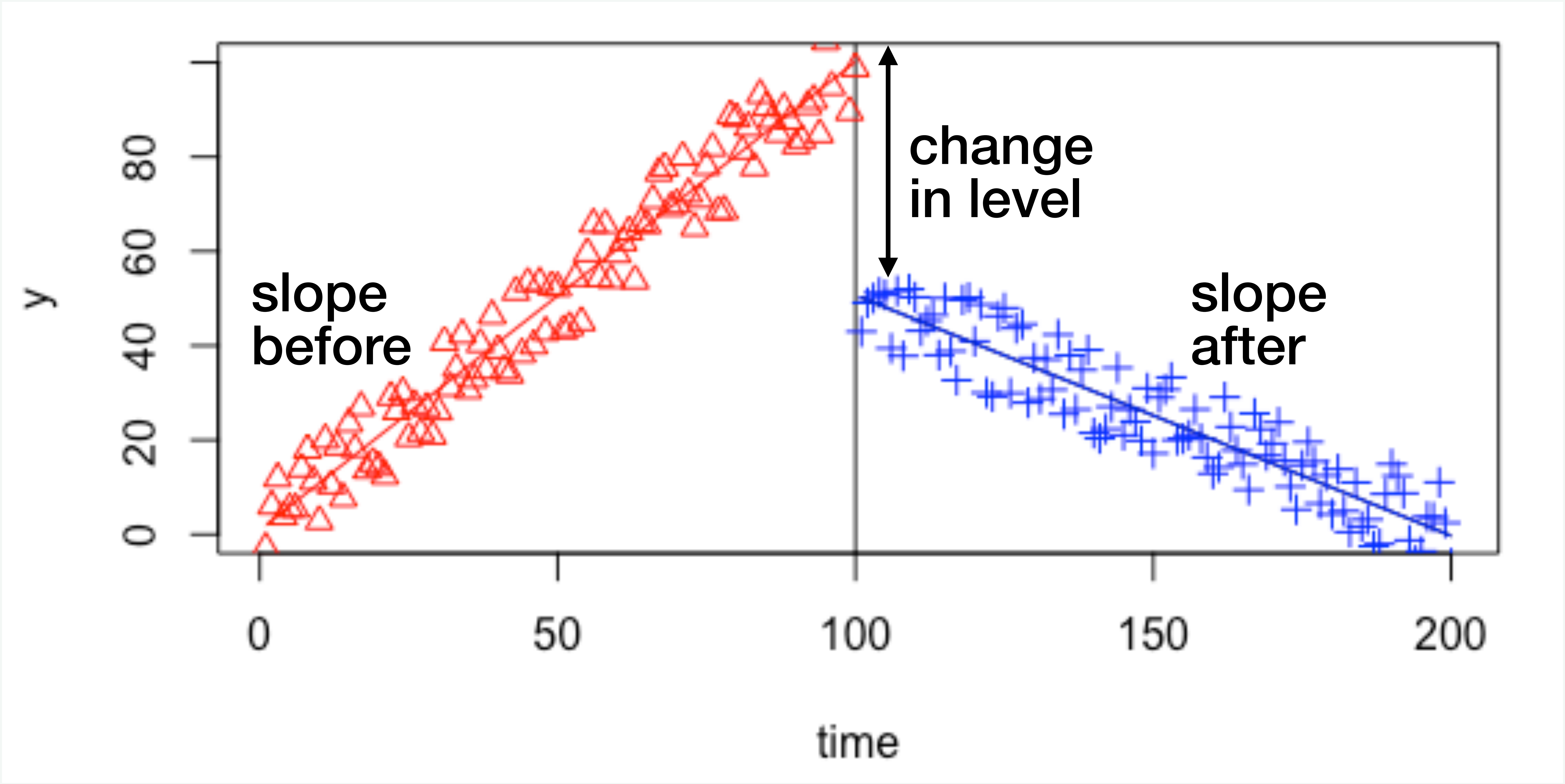
Evaluating the Effects of an Intervention



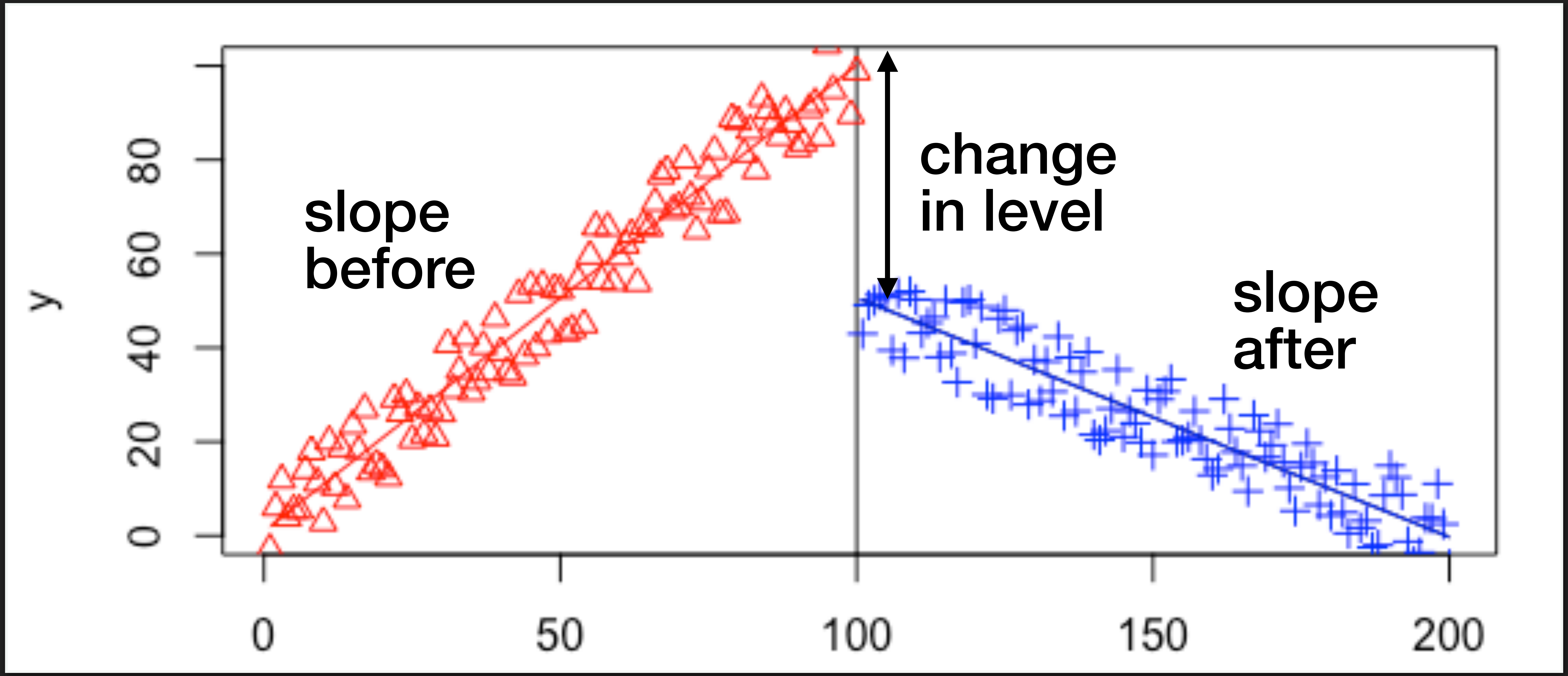
Evaluating the Effects of an Intervention



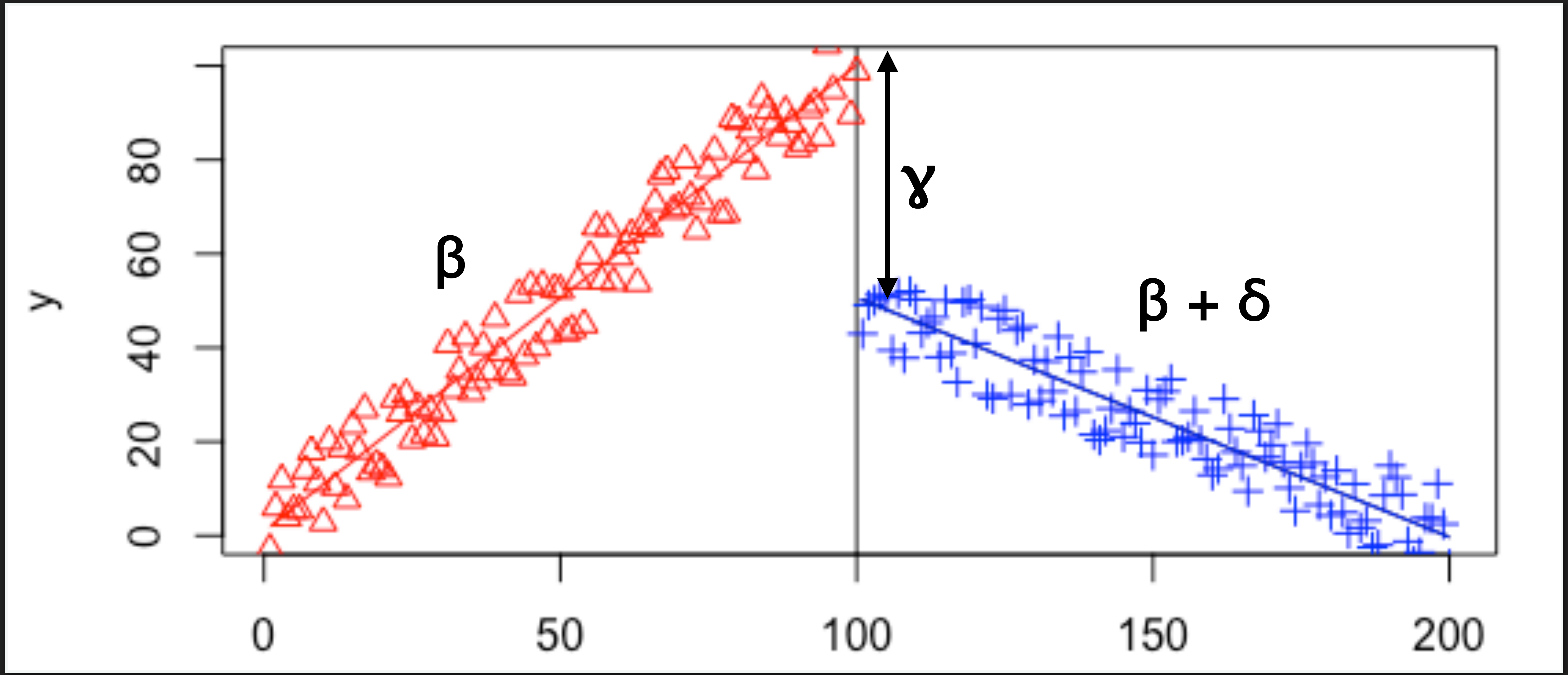
Segmented Regression Analysis of Interrupted Time Series Data



slope
before
change
in level
slope
after
y
0
20
40
60
80
0
50
100
150
200
time



time:	1	2	3	… … …	100	101	102 … … …	200
time after intervention:	0	0	0	… … …	1	2	3 … … …	100
intervention:	F	F	F	… … …	T	T	T … … …	T



time:	1	2	3 … … …	100	101	102	… … …	200
time after intervention:	0	0	0 … … …	1	2	3	… … …	100
intervention:	F	F	F … … …	T	T	T	… … …	T

$$
\begin{aligned}
y_i = \alpha + \beta \cdot \mathrm{time}_i &+ \\
\gamma \cdot \mathrm{intervention}_i &+ \\
\delta \cdot \mathrm{time\_after\_intervention}_i &+ \varepsilon_i
\end{aligned}
$$

Two examples, presented by Jenna and Simon

We Discussed:

- Trockman, A., Zhou, S., Kästner, C., & Vasilescu, B. (2018). Adding sparkle to social coding: an empirical study of repository badges in the npm ecosystem. In Proceedings of the 40th International Conference on Software Engineering (pp. 511-522).
  - See Jenna’s slides at the end of this deck.
- Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross−Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics, 27(4), 299-309.
  - See Simon’s slides at the end of this deck.

One more example:
The Florida "Stand your ground" paper

Debate Around "Stand Your Ground" Laws

- Self-defense laws, removing the duty to retreat and allowing the use of lethal force in situations (inside and outside the home) where an individual perceives a threat of harm.
- Advocates:
  - the increased threat of retaliatory violence deters would-be burglars.
- Critics:
  - weakening the punitive consequences of using force may serve to escalate aggressive encounters.

Box. States That Have Enacted "Stand Your Ground" Laws[a]

State Name (Year Original Law Signed)

Utah (1994)[b]
Florida (2005)
Alabama (2006)
Alaska (2006)
Arizona (2006)
Georgia (2006)
Indiana (2006)
Kansas (2006)
Kentucky (2006)
Louisiana (2006)
Michigan (2006)
Mississippi (2006)
Oklahoma (2006)
South Carolina (2006)
South Dakota (2006)
Tennessee (2007)
Texas (2007)
West Virginia (2008)
Montana (2009)
Nevada (2011)
New Hampshire (2011)
North Carolina (2011)
Pennsylvania (2011)

Florida Natural Experiment

- Florida was the first state to implement a stand your ground law, removing the duty to retreat principle.
- Idea: Use the years that have elapsed since the enactment of the Florida law to assess its impact on rates of homicide and homicide by firearm.

Box. States That Have Enacted "Stand Your Ground" Laws[a]

State Name (Year Original Law Signed)
Utah (1994)[b]
Florida (2005)
Alabama (2006)
Alaska (2006)
Arizona (2006)
Georgia (2006)
Indiana (2006)
Kansas (2006)
Kentucky (2006)
Louisiana (2006)
Michigan (2006)
Mississippi (2006)
Oklahoma (2006)
South Carolina (2006)
South Dakota (2006)
Tennessee (2007)
Texas (2007)
West Virginia (2008)
Montana (2009)
Nevada (2011)
New Hampshire (2011)
North Carolina (2011)
Pennsylvania (2011)

Potential Limitations of Interrupted Time Series Designs

- The possibility that other factors that occur simultaneously may distort estimates of intervention effects, e.g.,
  - national changes in social or economic variables (e.g., a recession)
  - events that have a profound and lasting impact on society (e.g., natural disasters).

- Study design features to address limitations:
  - analysis of homicide rates in 4 comparison states (New York, New Jersey, Ohio, and Virginia),
  - analysis of control outcomes (suicide and suicide by firearm).

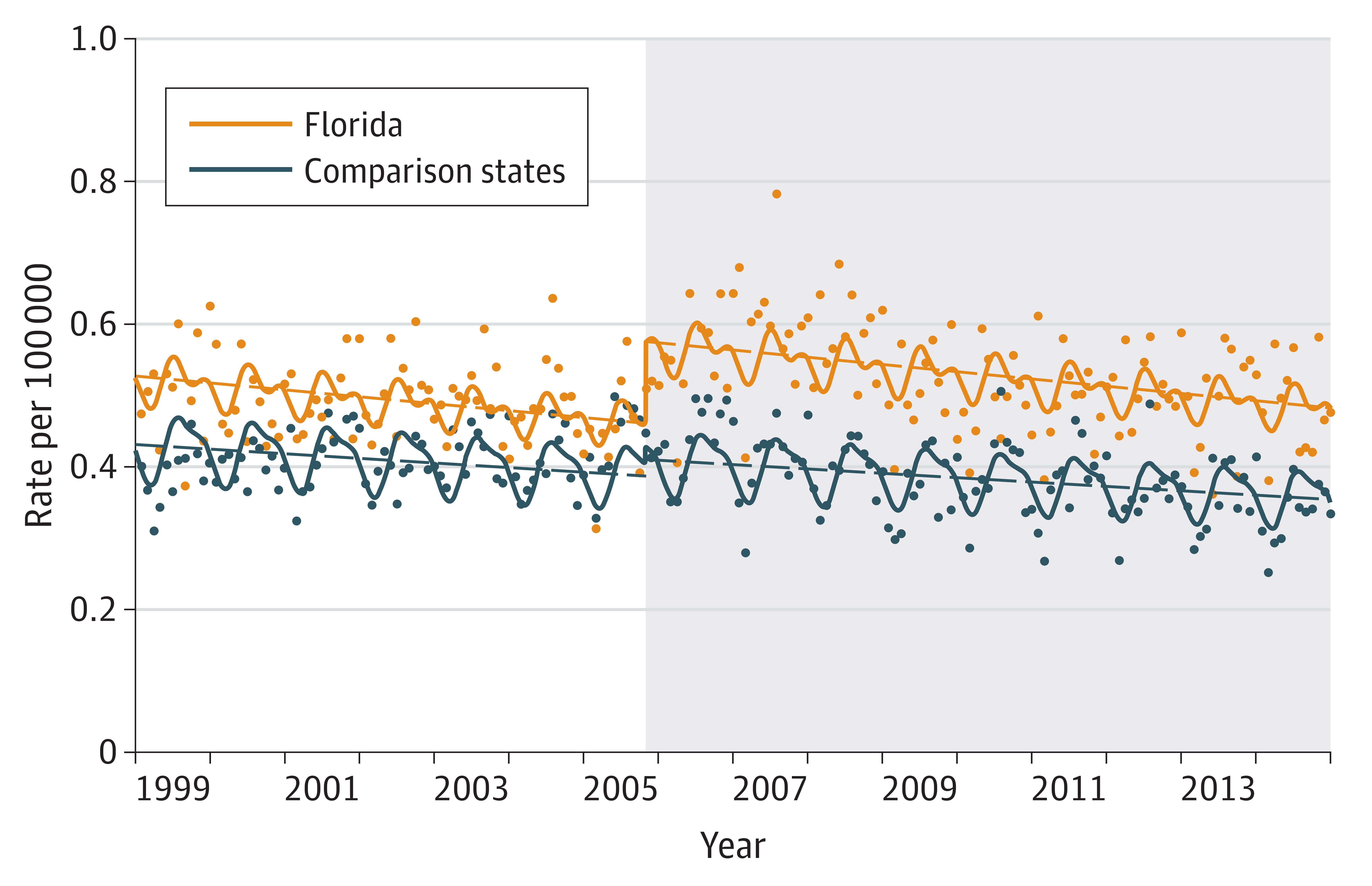
Data Sources

- Monthly death totals for Florida between Jan 1999 and Dec 2014, from CDC.
- Classified cases by:
  - place of occurrence (within or outside the State of Florida),
  - cause of death (homicide or suicide),
  - mechanism (firearms or other means), and
  - month of occurrence.

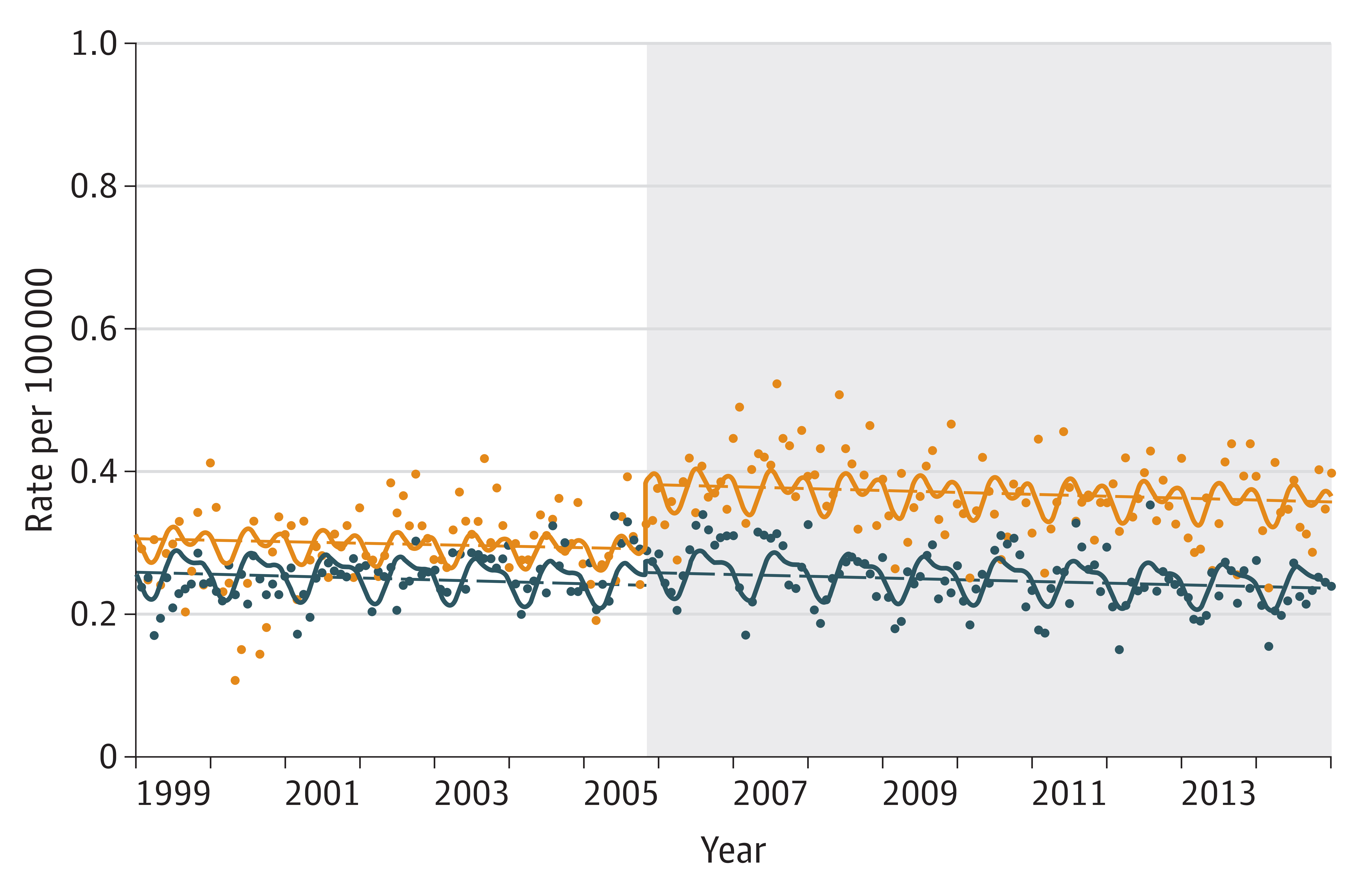
Data Analysis

- Evaluate whether post-intervention trends in homicide and homicide by firearm in Florida differed significantly from pre-intervention trends.
- Segmented quasi-Poisson regression analysis to analyze trends in both periods and estimate an effect size, taking underlying trends into account.
- Because of time sequencing of data points used in time series analysis, residual autocorrelation can lead to the violation of regression assumptions.
  - Generate robust standard errors (using a sandwich estimator) to produce more conservative estimates of uncertainty.

Homicide Rates in Florida and Comparison States



Homicide by Firearm Rates in Florida and Comparison States



Discussion

- Since Florida's stand your ground law took effect in October 2005, rates of homicide (+24.4% through 2014) and homicide by firearm (+31.6%) in the state have significantly increased.
- These increases appear to have occurred despite a general decline in homicide in the United States since the early 1990s.
- In contrast, rates of homicide and homicide by firearm did not increase in states without a stand your ground law (New York, New Jersey, Ohio, and Virginia), or for either suicide or suicide by firearm.
- Findings support the hypothesis that increases in the homicide and homicide by firearm rates in Florida are related to the stand your ground law.

Credits

- Graphics: Dave DiCello photography (cover)
- Shadish, William R., Thomas D. Cook, and Donald Thomas Campbell. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin, 2002.
  - Chapter 6: Interrupted time series
  - Chapter 7: Regression discontinuity design
- Morgan, S. L., & Winship, C. (2015). Counterfactuals and causal inference. Cambridge University Press.
  - Chapter 11: Repeated Observations and the Estimation of Causal Effects
- Humphreys, D. K., Gasparrini, A., & Wiebe, D. J. (2017). Evaluating the impact of Florida's "stand your ground" self-defense law on homicide and suicide by firearm: an interrupted time series study. JAMA Internal Medicine, 177(1), 44-50.
- Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. International Journal of Epidemiology, 46(1), 348-355.
- Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., & Armstrong, B. (2013). Time series regression studies in environmental epidemiology. International Journal of Epidemiology, 42(4), 1187-1195.
- Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross−Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. Journal of Clinical Pharmacy and Therapeutics, 27(4), 299-309.
- Trockman, A., Zhou, S., Kästner, C., & Vasilescu, B. (2018). Adding sparkle to social coding: an empirical study of repository badges in the npm ecosystem. In Proceedings of the 40th International Conference on Software Engineering (pp. 511-522).

Jenna's slides

Adding Sparkle to Social Coding:
An Empirical Study of Repository Badges in the *npm* Ecosystem

ASHER TROCKMAN, SHURUI ZHOU, CHRISTIAN KÄSTNER,
BOGDAN VASILESCU

Service	Master	Develop
CI Status	PASSED	PASSED
Test coverage	codecov 24%	codecov 46%
Test report	Test Report master	Test Report develop
Documentation	Doxygen master	Doxygen develop

Problem

Developers infer the quality of [open-source software] projects using visible cues, known as signals, collected from personal profile and repository pages.

GitHub **repository badges** can be seen as …

- easily observable signals used by maintainers to convey underlying qualities of their projects
- a game-like incentive designed to engage participants

Badges are a potentially impactful feature in transparent, social coding environments. **However, the value and effects of badges are not well understood.**

Main Research Questions

[RQ1] What are the most common badges and what does displaying them intend to signal?

[RQ2] To what degree do badges correlate with qualities that developers expect?

Overview of Methods: Mixed Methods

[RQ1] What are the most common badges and what does displaying them intend to signal?

- Conducted two online surveys targeting *npm* package maintainers and corresponding GitHub contributors
- Observed the frequency and historical adoption of badges among 294,941 *npm* packages through repository mining (collected a multidimensional longitudinal data set of *npm* packages)

[RQ2] To what degree do badges correlate with qualities that developers expect?

- Built regression models to test hypotheses regarding developer perceptions (collected when exploring RQ1)

Hypotheses

[H1] The adoption of quality-assurance badges correlates with other indicators of code quality (metric: test suite size).

[H2] The adoption of quality-assurance badges correlates with increased user confidence and attractiveness (metric: downloads).

[H3] The adoption of a quality-assurance badge, and even more so of a coverage badge, correlates with more external contributors including tests (metric: percentage of PRs with tests).

[H4] The adoption of dependency-management badges correlates with fresher dependencies (metric: freshness, see below).

Hypotheses

[H5] The adoption of a link-related badge does not correlate with either popularity or code quality.

[H6] The adoption of popularity-related badges in popular packages correlates with more future downloads (metric: monthly downloads).

[H7] The adoption of a support-related badges correlates with more responsive support (metric: issue closing time).

[H8] The number of badges correlates non-linearly with popularity.

Regression Analysis

Proceed in 3 complementary steps per hypothesis

1. **Correlation** – look for correlations between presence of badges and difference in the quality they are signaling; independent of causual relationships, confounds, or historic trends
   - Use the **non-parametric WMW test to compare distributions and report Cliff's delta**

2. **Additional Information** – explore whether badges add info to explain the qualities beyond readily available signals (stars, issues, downloads, dependent packages, etc.)
   - Use **hierarchical linear regression** comparing the fit of a base model including only readily available signals and control variables to a fully model with badge predictors
   - Follows a model fit and diagnostics process like the one we learned in class and did in the homework

1. **Longitudinal Analysis** – reveals whether introducing a first badge has an observable effect on the package's quality as the package evolves
   - Use **time series regression discontinuity design (RDD)** and **multiple regression**

RDD

Estimates the magnitude of a function's discontinuity between its values at points just before and just after an intervention

Based on the assumption that in the absence of an effect, the function's trend after the intervention would be continuous in the same way as prior to the intervention

[In This Domain] The earliest display of a badge is the intervention and by aligning the history on the intervention date the authors can compare 9 month trends before and after an intervention across many package

Multiple regression is then used to estimate the trend in the response before the badge adoption (*time*) and the changes in level (*intervention*) and trend (*time_after_intervention*) after the badge adoption. The authors also control for confounds in the multiple regression to evaluate whether the change could be attributed to other factors than the intervention.

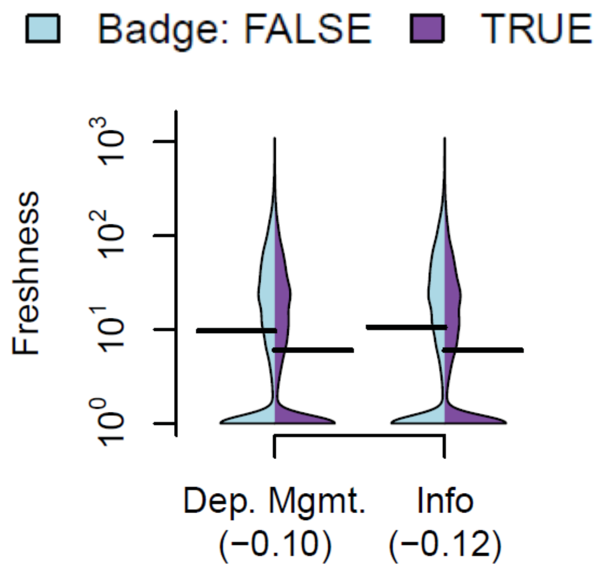
Example: Dependency Management

Response Variable: dependency freshness – metric score based on how many dependencies declared in a package have a newer version that existed on *npm* at the time

[H4] dependency-manager badges correlate with more up-to-date secure dependencies operationalized with freshness metric

[H5] expect a marginal effect from information-related badges

Correlation: Dependency Management



(a) Dependency freshness

Supporting H4 and, surprisingly, contradicting H5,
Fig. 2a reveals a small, but statistically significant difference:

Packages with a dependency-manager badge or an information badge tend to have overall fresher dependencies than packages without.

Additional Information: Dependency Management

Base model: explain freshness given **stars, dependents, dependencies, contributors** and **a control for time** since package was last updated

- Explains 17.3 % of the deviance

Full model: additionally models the presence of **dependency-manager badges and information badges** and their interaction

- explains 17.4 % of the deviance.

The odds of having fresh dependencies increase by 27% for packages **with dependency-manager badges (H4)**.

The **effect of information badges** is a 17% increase in odds **(H5)**.

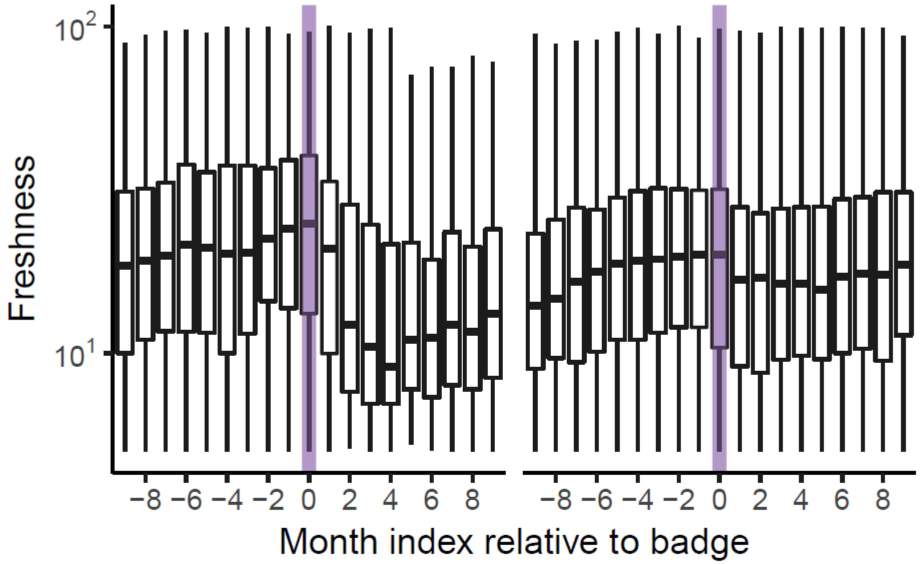
	Basic Model response: *freshness = 0* 17.3% deviance explained		**Full Model** response: *freshness = 0* 17.4% deviance explained		**RDD** response: log(*freshness*) $R^2_m = 0.04$, $R^2_c = 0.35$	
	Coeffs (Err.)	LR Chisq	Coeffs (Err.)	LR Chisq	Coeffs (Err.)	Sum sq.
(Interc.)	$3.54\ (0.03)^{***}$		$3.50\ (0.03)^{***}$		$1.45\ (0.09)^{***}$	
Dep.	$-1.78\ (0.01)^{***}$	32077.8^{***}	$-1.79\ (0.01)^{***}$	32292.8^{***}	$-0.04\ (0.02)$	3.01
RDep.	$0.22\ (0.01)^{***}$	610.3^{***}	$0.21\ (0.01)^{***}$	560.6^{***}	$-0.01\ (0.02)$	0.11
Stars	$-0.08\ (0.00)^{***}$	301.4^{***}	$-0.09\ (0.00)^{***}$	311.2^{***}	$0.00\ (0.01)$	0.00
Contr.	$-0.24\ (0.01)^{***}$	500.5^{***}	$-0.25\ (0.01)^{***}$	548.7^{***}	$-0.04\ (0.02)^{*}$	4.39^{*}
lastU	$-0.65\ (0.01)^{***}$	12080.9^{***}	$-0.64\ (0.01)^{***}$	11537.9^{***}	$0.01\ (0.02)$	0.37
hasDM			$0.24\ (0.03)^{***}$	116.1^{***}	$0.45\ (0.08)^{***}$	2.43
hasInf			$0.11\ (0.02)^{***}$	48.3^{***}	$0.04\ (0.05)$	0.45
hasDM:hasInf			$-0.05\ (0.04)$	1.9	$-0.32\ (0.10)^{**}$	
hasOther			$0.01\ (0.01)$			
time					$0.03\ (0.00)^{***}$	82.99^{***}
intervention					$-0.93\ (0.03)^{***}$	1373.22^{***}
time_after_intervention					$0.11\ (0.00)^{***}$	455.56^{***}
time_after_intervention:hasDM					$-0.10\ (0.01)^{***}$	230.36^{***}
time_after_intervention:hasInf					$-0.00\ (0.01)$	1.14
time_after_intervention:hasDM:hasInf					$0.03\ (0.01)^{**}$	10.62^{**}

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$;
Dep: dependencies; RDep: dependents; Contr.: contributors; lastU: time since last update; hasDM: has dependency-manager badge; hasInf: has information badge; hasOther: adopts additional badges within 15 days

Longitudinal Analysis: Dependency Management

A trend is already visible from the longitudinal freshness data plotted for those packages in Fig. 3a.

The adoption of any badges correlates to an improvement in freshness, especially for dependency-manager badges.



(a) Monthly freshness scores, rel. to dependency-manager (left) and information badges (right).

Longitudinal Analysis: Dependency Management

The RDD model confirms the trend:

The adoption of (any) badges correlates to a strong improvement in freshness, by about a factor 2.5 on average.

Interpretation Derivation:

- Coefficient for *intervention*
- e^0.93 factor decrease in freshness score

After adoption freshness slightly decays again over time.

Interpretation Derivation:

Sum of the coefficients for *time* and *time_after_intervention* in the model, which expresses the slope of the post-intervention trend

	Basic Model response: *freshness = 0* 17.3% deviance explained		**Full Model** response: *freshness = 0* 17.4% deviance explained		**RDD** response: log(*freshness*) $R^2_m = 0.04$, $R^2_c = 0.35$	
	Coeffs (Err.)	LR Chisq	Coeffs (Err.)	LR Chisq	Coeffs (Err.)	Sum sq.
(Interc.)	3.54 (0.03)***		3.50 (0.03)***		1.45 (0.09)***	
Dep.	−1.78 (0.01)***	32077.8***	−1.79 (0.01)***	32292.8***	−0.04 (0.02)	3.01
RDep.	0.22 (0.01)***	610.3***	0.21 (0.01)***	560.6***	−0.01 (0.02)	0.11
Stars	−0.08 (0.00)***	301.4***	−0.09 (0.00)***	311.2***	0.00 (0.01)	0.00
Contr.	−0.24 (0.01)***	500.5***	−0.25 (0.01)***	548.7***	−0.04 (0.02)*	4.39*
lastU	−0.65 (0.01)***	12080.9***	−0.64 (0.01)***	11537.9***	0.01 (0.02)	0.37
hasDM			0.24 (0.03)***	116.1***	0.45 (0.08)***	2.43
hasInf			0.11 (0.02)***	48.3***	0.04 (0.05)	0.45
hasDM:hasInf			−0.05 (0.04)	1.9	−0.32 (0.10)**	
hasOther			0.01 (0.01)			
time					0.03 (0.00)***	82.99***
intervention					−0.93 (0.03)***	1373.22***
time_after_intervention					0.11 (0.00)***	455.56***
time_after_intervention:hasDM					−0.10 (0.01)***	230.36***
time_after_intervention:hasInf					−0.00 (0.01)	1.14
time_after_intervention:hasDM:hasInf					0.03 (0.01)**	10.62**

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$;
Dep: dependencies; RDep: dependents; Contr.: contributors; lastU: time since last update; hasDM: has dependency-manager badge; hasInf: has information badge; hasOther: adopts additional badges within 15 days

Regression Analysis: Threats to Validity

1. Imperfect measures
   - Operational measures don’t capture all aspects of a software quality
   - E.g. Large test suites as an indicator of good testing practices

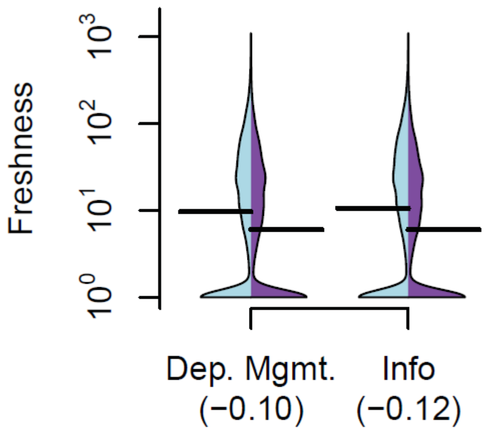
2. Badges vs practices
   - Cannot distinguish between effects of practice adoption from effects of badge adoption; results can only be interpreted as exploring the reliability of the signal that a badge provides

3. Beyond correlations
   - None of the three analysis steps can establish a causal relationship between badges and the studied qualities

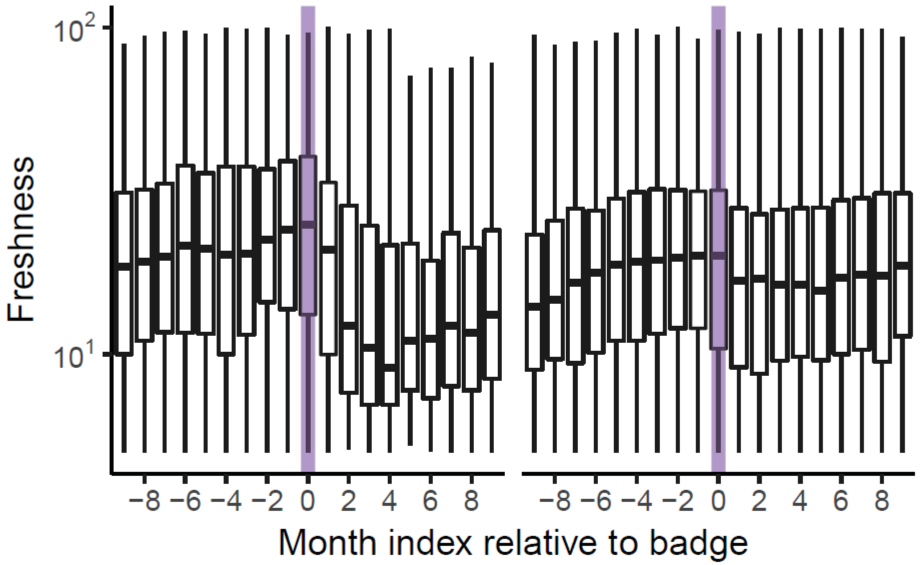
Extra Slides

	Basic Model response: *freshness = 0* 17.3% deviance explained		Full Model response: *freshness = 0* 17.4% deviance explained		RDD response: log(*freshness*) $R^2_m = 0.04$, $R^2_c = 0.35$	
	Coeffs (Err.)	LR Chisq	Coeffs (Err.)	LR Chisq	Coeffs (Err.)	Sum sq.
(Interc.)	$3.54\ (0.03)^{***}$		$3.50\ (0.03)^{***}$		$1.45\ (0.09)^{***}$	
Dep.	$-1.78\ (0.01)^{***}$	32077.8^{***}	$-1.79\ (0.01)^{***}$	32292.8^{***}	$-0.04\ (0.02)$	3.01
RDep.	$0.22\ (0.01)^{***}$	610.3^{***}	$0.21\ (0.01)^{***}$	560.6^{***}	$-0.01\ (0.02)$	0.11
Stars	$-0.08\ (0.00)^{***}$	301.4^{***}	$-0.09\ (0.00)^{***}$	311.2^{***}	$0.00\ (0.01)$	0.00
Contr.	$-0.24\ (0.01)^{***}$	500.5^{***}	$-0.25\ (0.01)^{***}$	548.7^{***}	$-0.04\ (0.02)^{*}$	4.39^{*}
lastU	$-0.65\ (0.01)^{***}$	12080.9^{***}	$-0.64\ (0.01)^{***}$	11537.9^{***}	$0.01\ (0.02)$	0.37
hasDM			$0.24\ (0.03)^{***}$	116.1^{***}	$0.45\ (0.08)^{***}$	2.43
hasInf			$0.11\ (0.02)^{***}$	48.3^{***}	$0.04\ (0.05)$	0.45
hasDM:hasInf			$-0.05\ (0.04)$	1.9	$-0.32\ (0.10)^{**}$	
hasOther			$0.01\ (0.01)$			
time					$0.03\ (0.00)^{***}$	82.99^{***}
intervention					$-0.93\ (0.03)^{***}$	1373.22^{***}
time_after_intervention					$0.11\ (0.00)^{***}$	455.56^{***}
time_after_intervention:hasDM					$-0.10\ (0.01)^{***}$	230.36^{***}
time_after_intervention:hasInf					$-0.00\ (0.01)$	1.14
time_after_intervention:hasDM:hasInf					$0.03\ (0.01)^{**}$	10.62^{**}

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$;
Dep: dependencies; RDep: dependents; Contr.: contributors; lastU: time since last update; hasDM: has dependency-manager badge; hasInf: has information badge; hasOther: adopts additional badges within 15 days



(a) Dependency freshness



(a) Monthly freshness scores, rel. to dependency-manager (left) and information badges (right).

Simon's slides

Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research

Wagner, Zhang, et al.

Journal of Clinical Pharmacy and Therapeutics

Simon Chu 4/8/2021
17803 Empirical Methods

Interrupted Time Series Analysis (ITS)

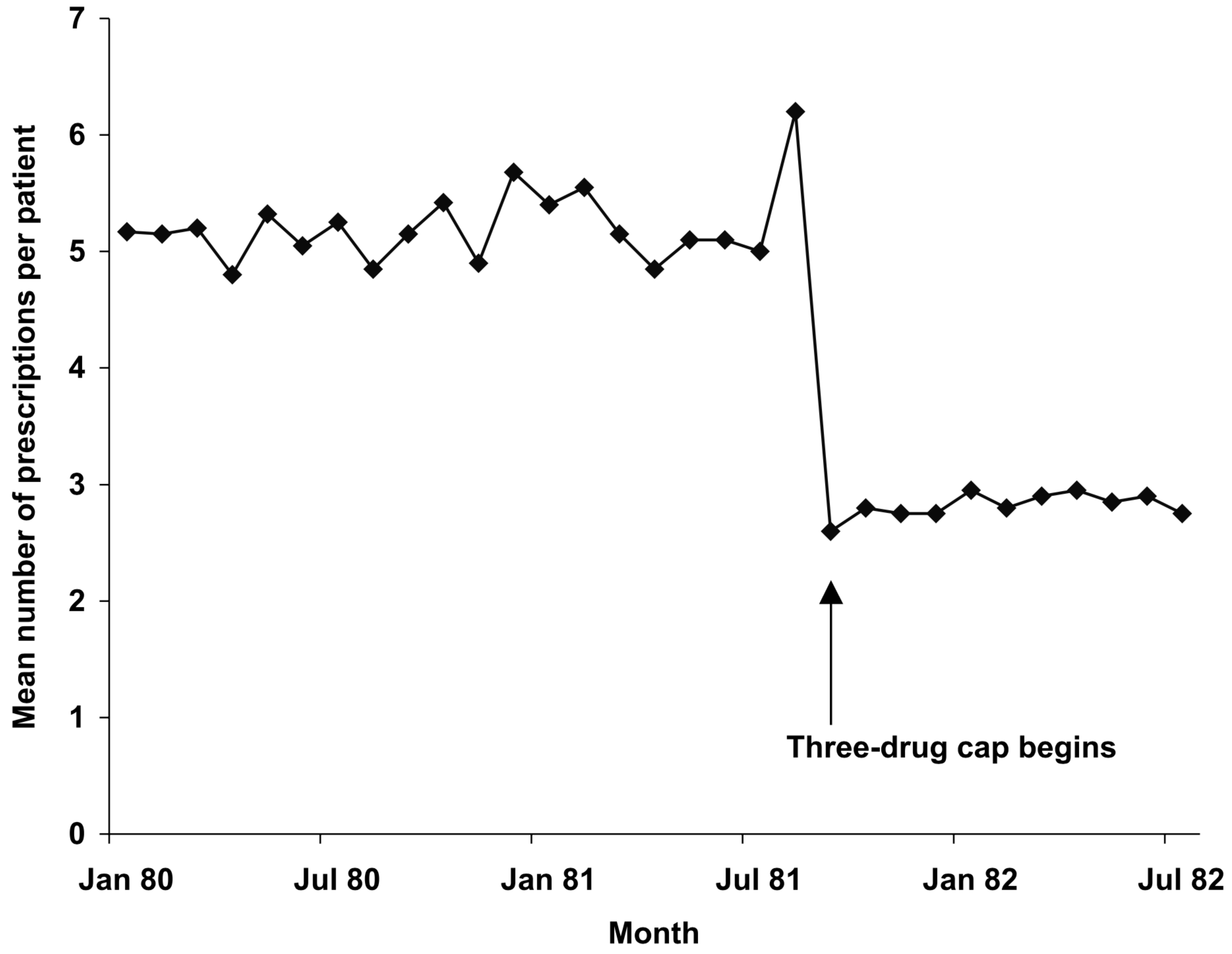
ITS in a Nutshell

- **Statistical analysis** method involving tracking a **long-term** period before and after a point of **intervention** to **assess** the intervention's **effects**
- AKA, quasi-experimental time series analysis
- Widely used in political science, economics, sociology, etc.
- Now in medication

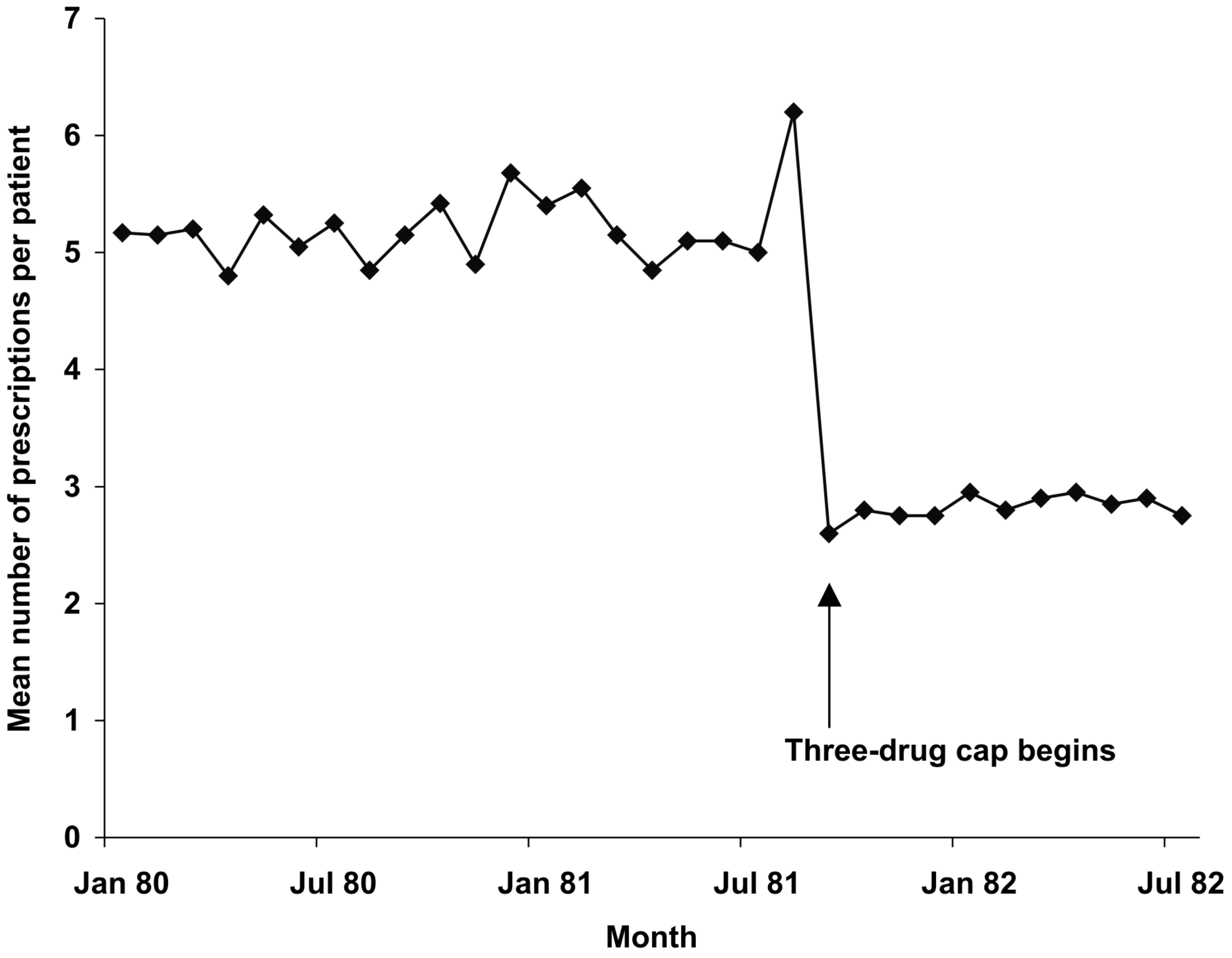
ITS in a Nutshell

- Time Series → Data over a period
- Interrupted → Intervention(s)
- Segmented Regression Analysis $\in$ ITS
  - Requires a **sufficient number** of time points **before** and **after** the intervention for segmented regression analysis
  - Requires data collected regularly over time, and organized at equally spaced intervals.

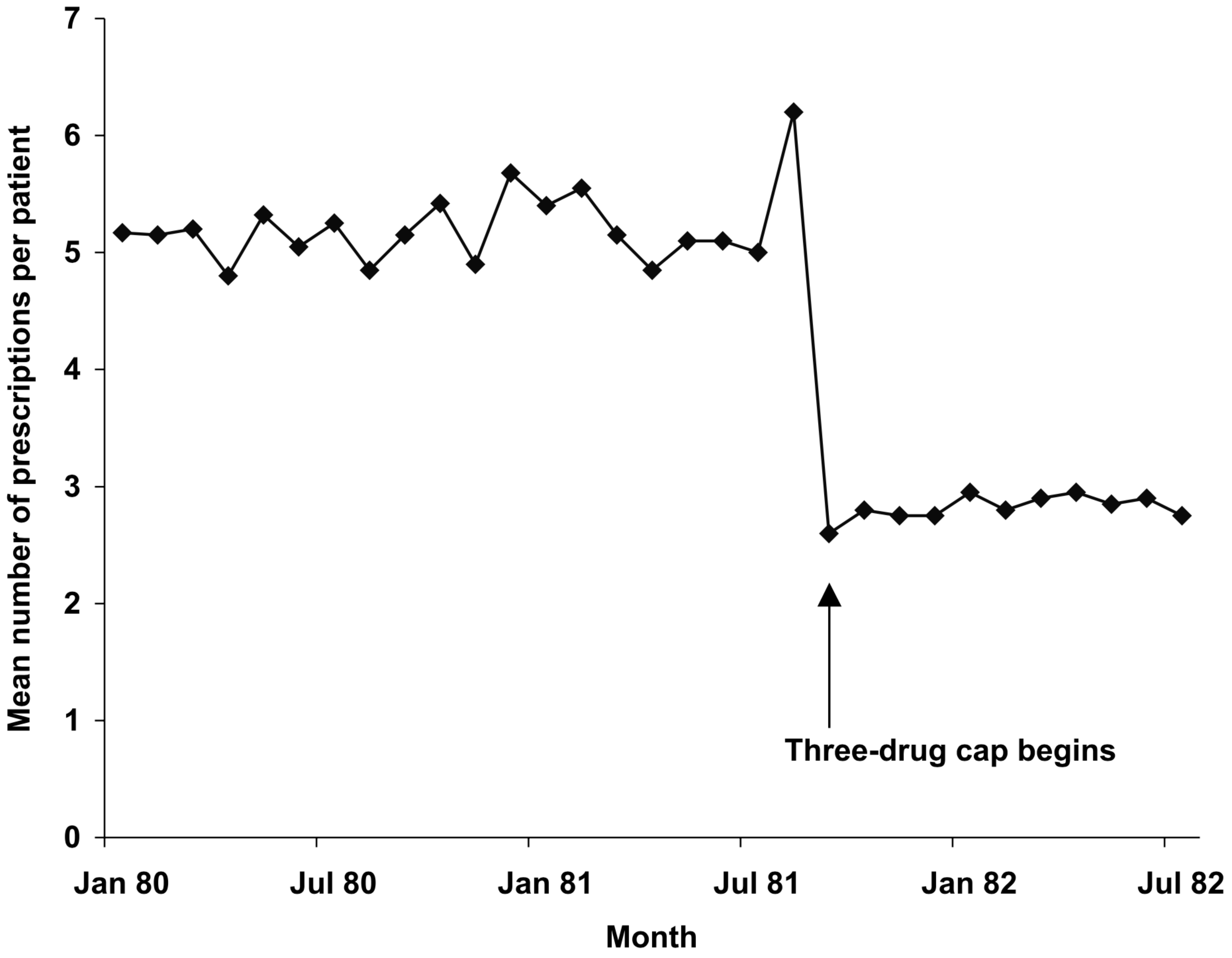
Running Example



Question: Is the change in level and trend the result of chance alone, or the factors other than intervention?



$$Y_t = \beta_0 + \beta_1{}^{*}\text{time}_t + \beta_2{}^{*}\text{intervention}_t + \beta_3{}^{*}\text{time after intervention}_t + e_t \quad (1)$$



$$Y_t = \beta_0 + \beta_1{}^{*}\text{time}_t + \beta_2{}^{*}\text{intervention}_t + \beta_3{}^{*}\text{time after intervention}_t + e_t \tag{1}$$

time_t	Time since Start
intervention_t	Intervention Indicator
Time after intervention_t	# of Months after Intervention
e_t	Error Term
Y_t	Mean # of Prescriptions/Patient/ Month

$$Y_t = \beta_0 + \beta_1{}^{*}\text{time}_t + \beta_2{}^{*}\text{intervention}_t + \beta_3{}^{*}\text{time after intervention}_t + e_t \quad (1)$$

β0	Baseline Level
β1	Baseline Trend
β2	Level Change
β3	Trend Change

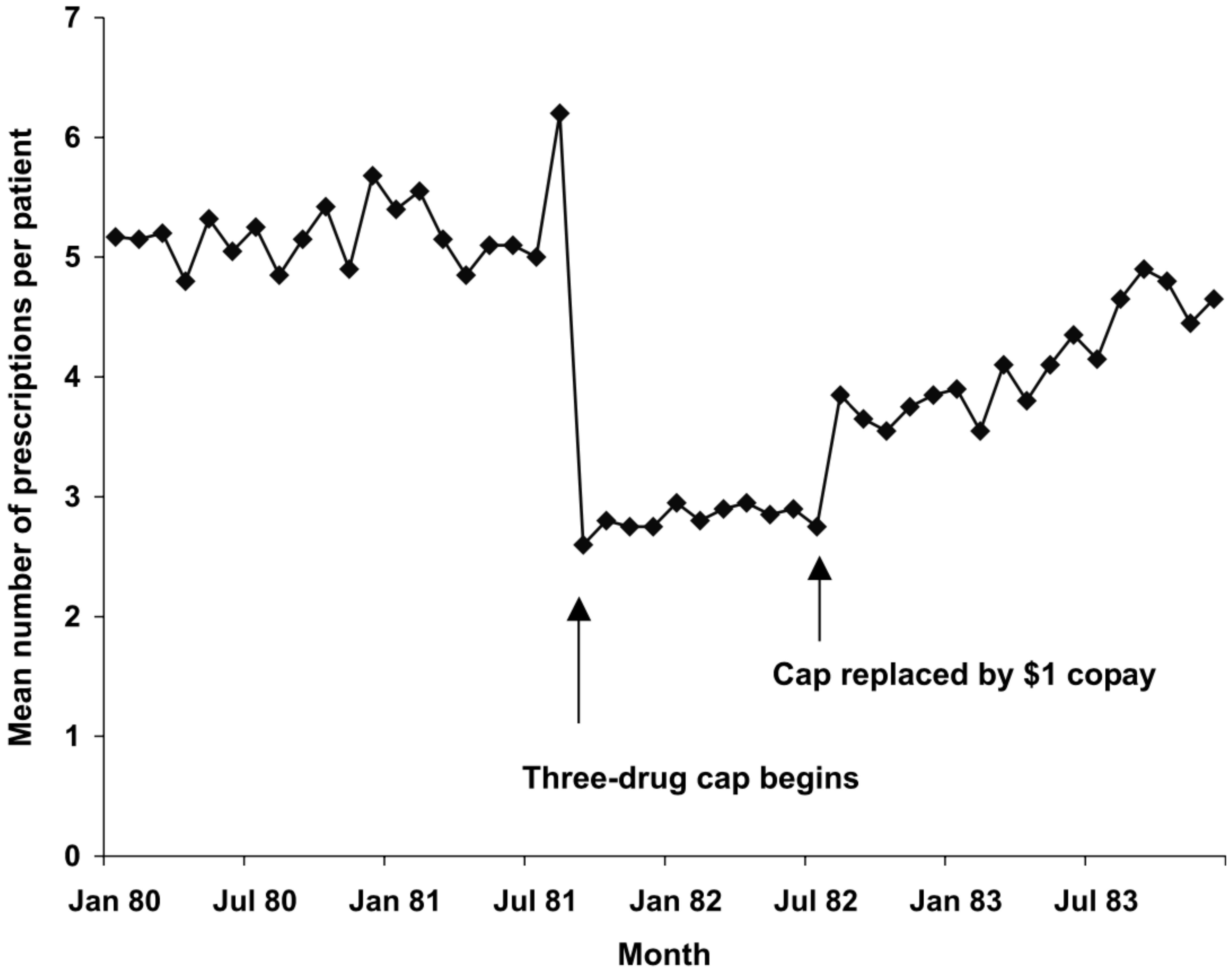
Data Fitting

	Coefficient	Standard error	t-statistic	P-value
a. Full segmented regression model				
Intercept β_0	5·1389	0·0748	68·69	<0·0001
Baseline trend β_1	0·003481	0·006791	0·51	0·6128
Level change after cap β_2	−2·5931	0·1572	−16·49	<0·0001
Trend change after cap β_3	0·0263	0·0193	1·36	0·1849
b. Most parsimonious segmented regression model				
Intercept β_0	5·1677	0·0311	166·38	<0·0001
Level change after cap β_2	−2·3736	0·0563	−42·14	<0·0001

Report the Intervention Effect

- (Absolute) Level/Trend Changes
  - Avg. # of prescriptions/patient/month dropped 2.6
- Percentage/Rate of Changes Based on the Baseline Trend/Level Changes
  - Avg. # of prescriptions/patient/month decreased by 46%

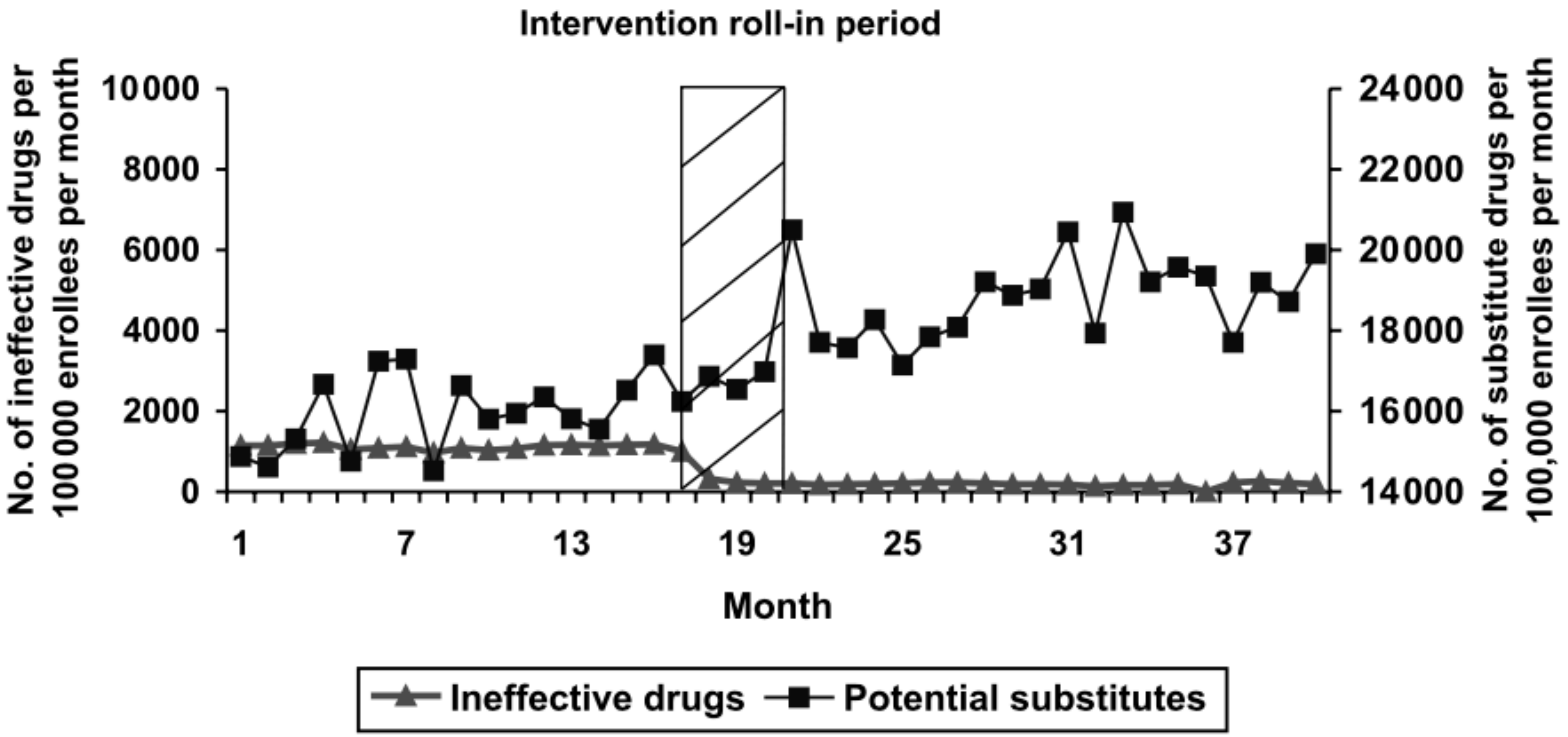
Multiple Interventions



$$
\begin{aligned}
Y_t = \beta_0 &+ \beta_1 \times time_t + \beta_2 \times intervention1_t \\
&+ \beta_3 \times time\ after\ intervention1_t \\
&+ \beta_4 \times intervention2_t \\
&+ \beta_5 \times time\ after\ intervention2_t + e_t
\end{aligned}
\tag{4}
$$

Lagged Effect

What is it?



Lagged Effect

Modeling the Lagged Effect

- How to Manage the Transition Period?
  - Exclude the data from the transition period in the time series analysis
  - Model the period as a separate segment (analyze separately)

Autocorrelation

Collinearity?

- (Seasonal/Cyclic) Patterns
- # of Prescription in Jan. 2021 ≈ # of Prescription in Jan. 2020 (than other months)

Autocorrelation

Detection

- Supported by Software (proc autoreg in SAS)
- Visual inspection (residuals vs. time plot)
  - No pattern → good, no autocorrelation
  - Pattern → bad, autocorrelation → mitigation

Autocorrelation

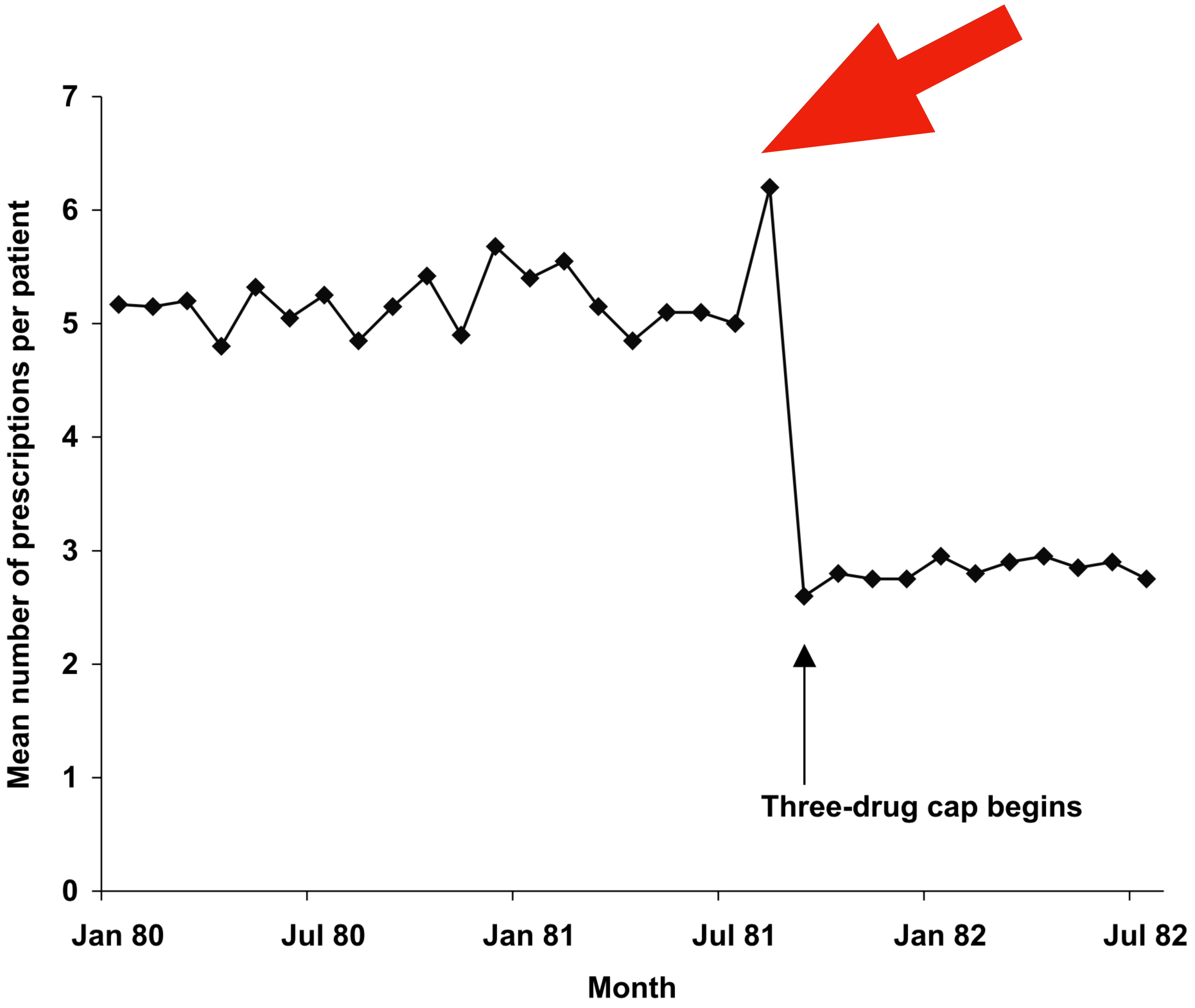
Consequences When Fail to Consider

- Underestimate Standard Errors
- Overestimated Significance of the Effect of an Intervention

Wild Data Points

Outliers

“Anticipatory Demand”



Wild Data Points

Causes & Mitigation

- some caused by measurement errors
- Some can actually be explained
    - “Anticipatory Demand”
- Some are caused by random variation
    - Carry out the analysis **with** and **without** the wild data point to evaluate its impact.

Bias Control

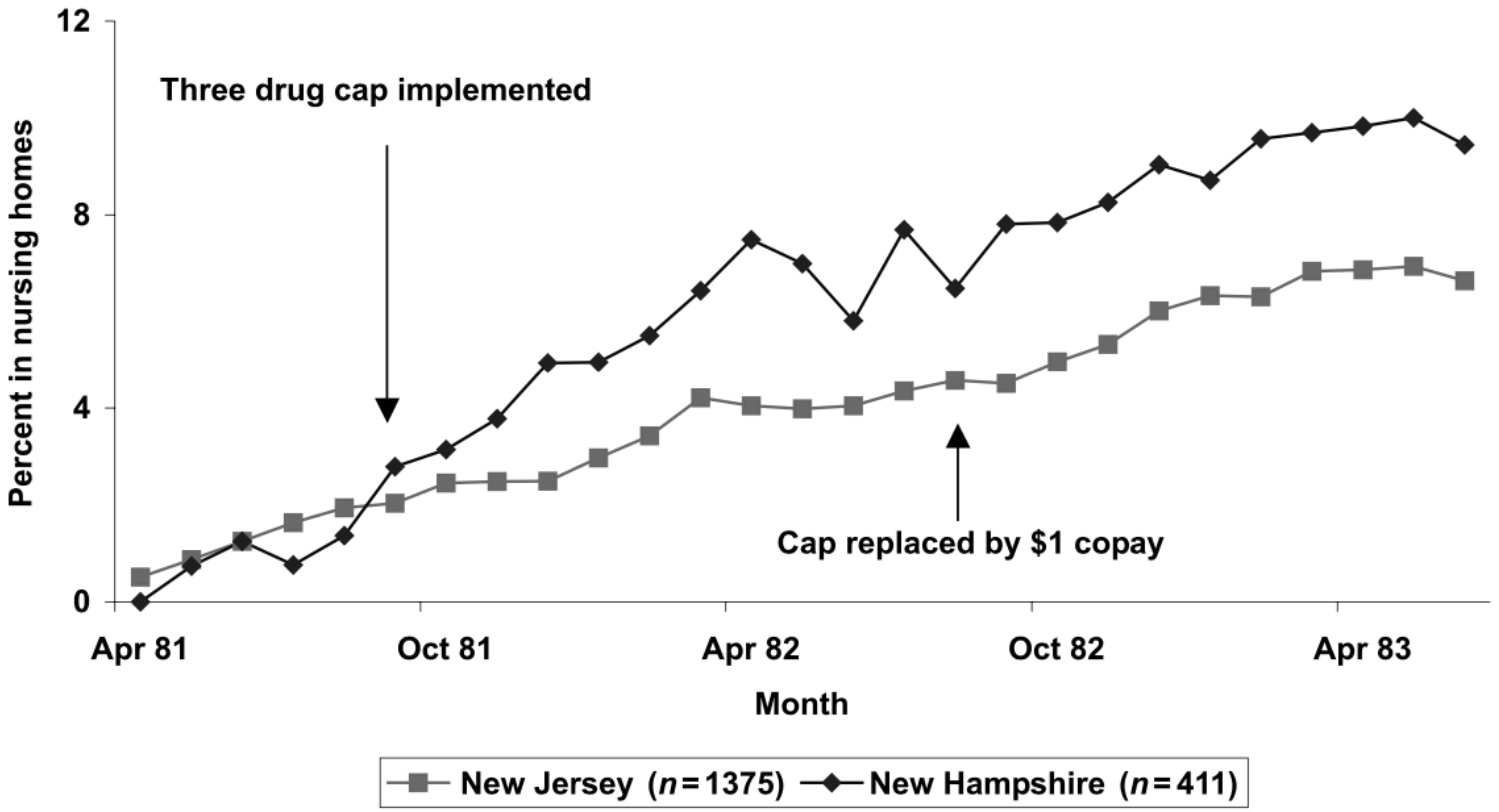
Sources of biases

- **Co-interventions (**simultaneously occurring interventions)
- **Seasonal changes** that occurs at the time of intervention
- Changes in **composition** of **study population**
- Changes in **measurement** at the time of intervention
- …

Bias Control

Mitigate Biases

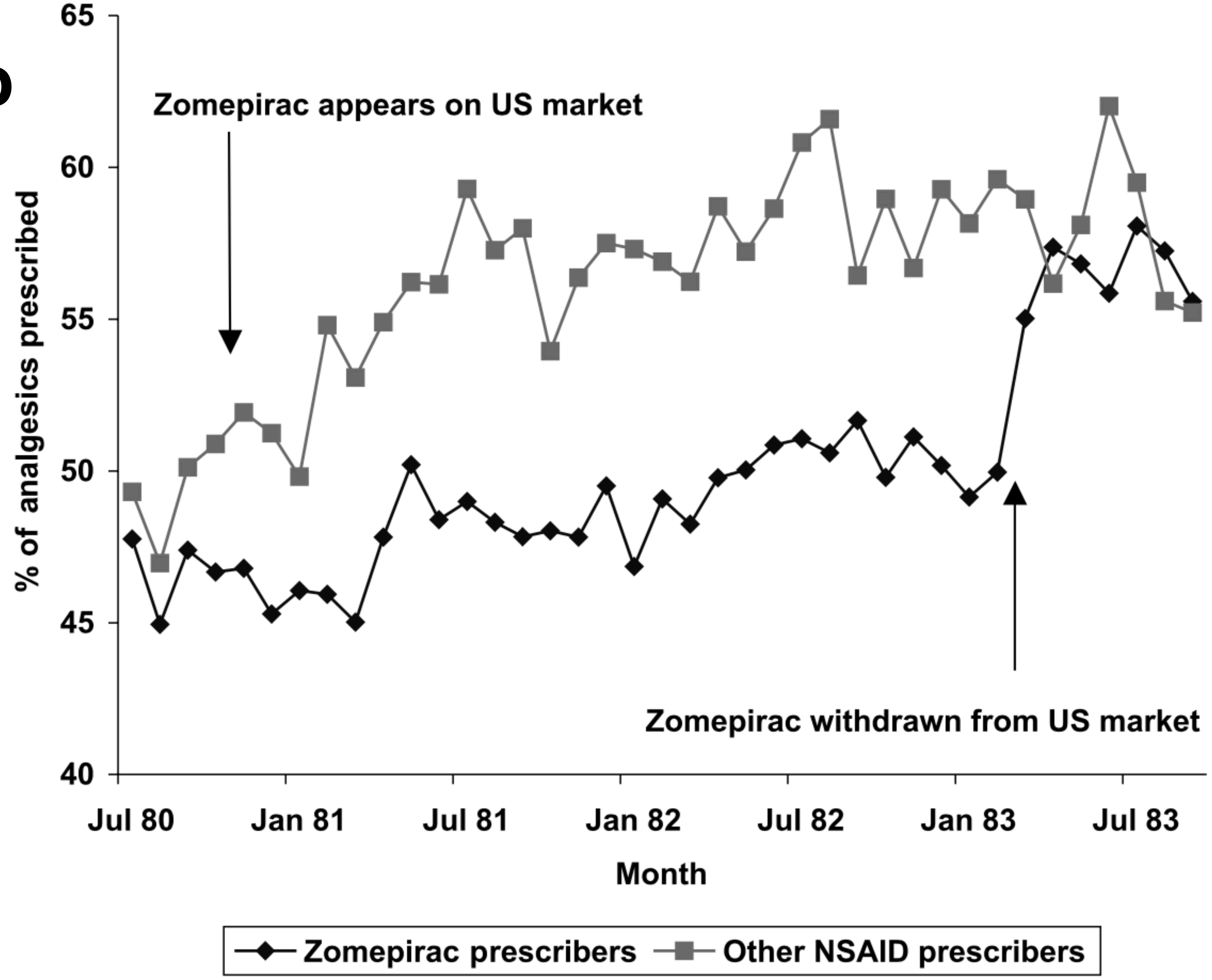
- Use control group
- Comparing the effect in the intervention group with that in the control group
- Separating the intervention effect from others that may have occurred at the same time



Bias Control

"Less Desirable" Control Group

- Use control group



Stratification

- Intervention effect can be studied separately in each group
  - “Staff model”
  - “Group model”

Final Regression Model

Full Model vs. Parsimonious Model?

- Both full and the most parsimonious models will not correctly estimate the effect of the intervention if confounders exist.
- Important measured confounders should be added to the model (regardless of statistical significance)
  - Such as baseline trend, an important control variable for secular trends.

Strengths

Summary

- Allow analysts to control for prior trends in the outcome and to study the dynamics of change in response to an intervention
- Address important threats to internal validity (history/maturation, even without a control group)
- Estimate changes in the trend of the effect over time.
- Visually display the dynamics of response to intervention
  - delayed, abrupt, or gradual
  - Effect persists/is temporary

Weaknesses

Summary

- Doesn’t support non-linear patterns
    - Can deploy Box-Jenkins Model, but it requires 50 time points, which medication use research lacks.
- Does not allow control for individual-level covariates (New Hampshire Medicaid Enrollees), less information <—> Cross-section analysis methods