PHS 2000B: Econometrics Interrupted Time Series 2023

Maggie McConnell
Department of Global Health and Population
Harvard T. H. Chan School of Public Health

Plan of presentation

1. Recap of RDD
2. Background – program evaluation – quasi experiments
3. **Interrupted time series model**
4. Practical examples
5. Timing of effects
6. Autocorrelation and standard errors
7. Controlled interrupted time series

When is RDD valid?

- Treatment is assigned on the basis of a continuous observable variable – running variable z
- There is a discontinuity in the probability of receiving “treatment” (T) at some cut-off value of the running variable – threshold c
- The only thing that differs for groups above and below the threshold is the likelihood of program participation/eligibility
  - Continuity of outcome at threshold if untreated by the program – no natural jump

Assessment of Validity of RDD

- We want the distribution of z to be smooth around c
  - No bunching (i.e. no signs of manipulation)

- We want to see a jump in the probability of receiving T at c
  - For sharp RD this jump goes from 0 to 1
  - For fuzzy RDD this jump is somewhere in the range $0<p<1$

- We want the characteristics of those just above and below the threshold to be similar

Practical Tips: Assessing the Validity of RDD

1) Show characteristics (SES, age...) are balanced around cutoff.
   - One way to do this is to estimate the same main regression as RD but your "outcomes" are the covariates. Look for jumps near the cutoff.
   - Similar to balance tables in RCTs

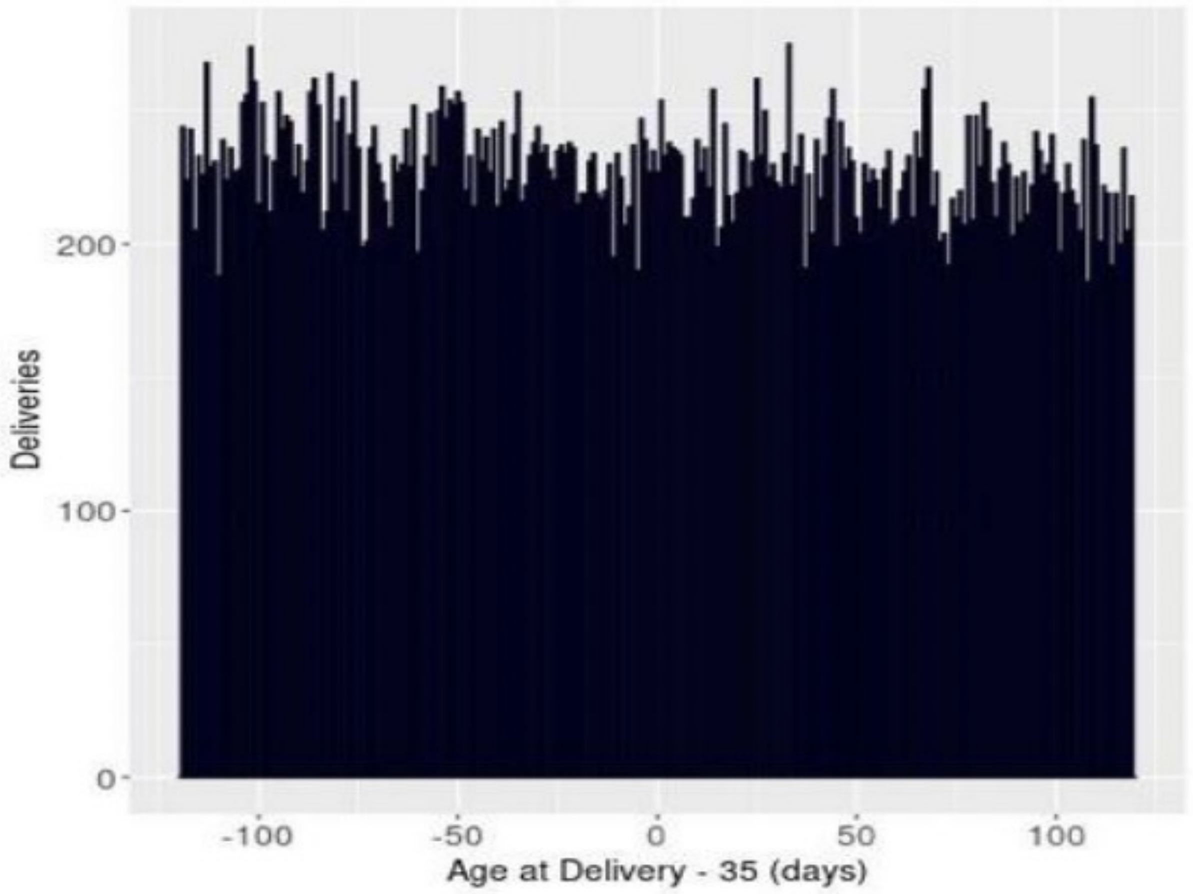
2) Look for "bunching" in the running variable around cutoff
   - Should not be extra mass in distribution near cutoff
   - Can use McCrary 2008 density test to test this formally

3) Estimate causal effect using different bandwidths and check stability of estimates — instability suggests wrong functional form
   - can use formal procedures to choose "optimal" bandwidth (Imbens and Kalyanaraman 2009)

4) Estimate "false cutoff"
   - Should not find significant effects at the wrong cutoff

Do you see bunching here?

eFigure 4. Histogram of All Deliveries to Individuals Within 120 Days of the AMA Cutoff



Legend: Shown is the number of deliveries by the running variable, i.e., the number of days between the expected date of delivery and the individual's 35th birthday. Sample includes all individuals with an expected date of delivery within 120 days of the 35th birthday.

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

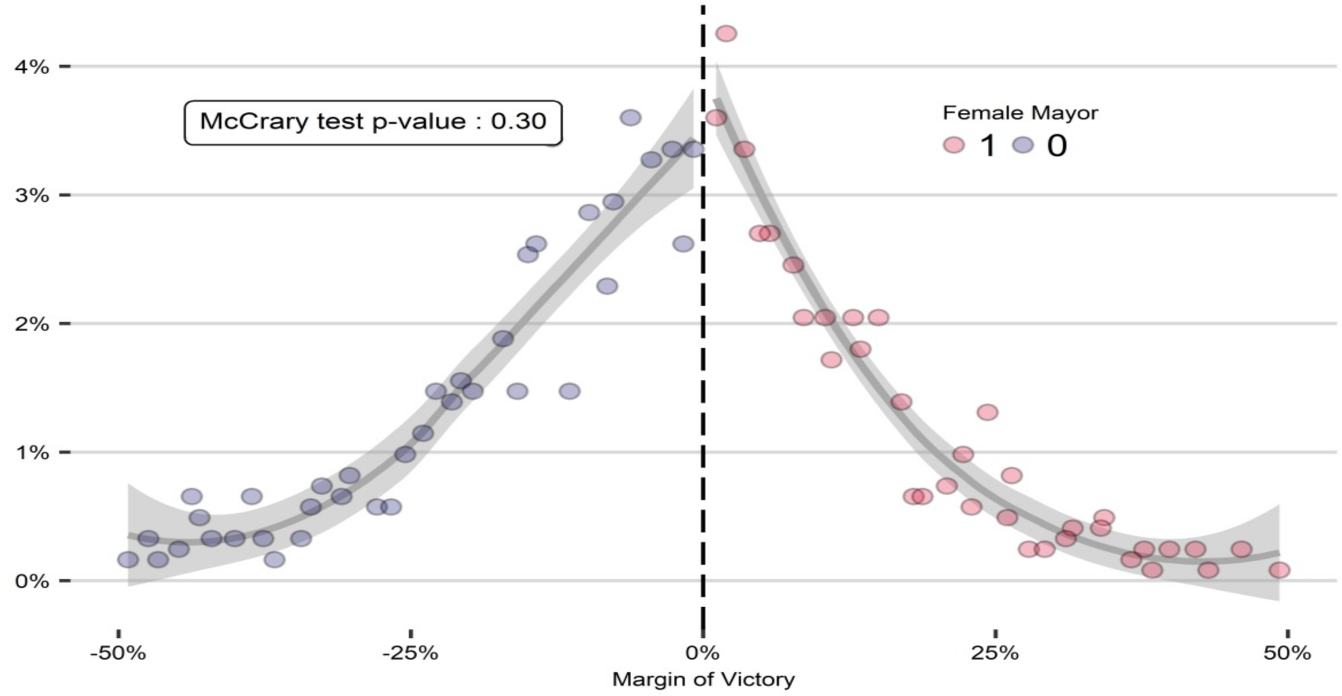
Under pressure: Women's leadership during the COVID-19 crisis

Raphael Bruce [a], Alexsandros Cavgias [b], Luis Meloni [c,*], Mário Remígio [c]

[a] Insper Institute of Education and Research, Brazil
[b] University of Barcelona, Spain
[c] University of São Paulo, Brazil

What about bunching here?

Figure A2: McCrary Test



Notes: This figures displays the McCrary density test for the running variable around the cutoff (McCrary, 2008).

What are we estimating?

If we get through all these checks, we can feel more confident assessing changes in outcome variables at the cutoff as:

- The average causal effect of the program at the threshold for sharp RDD
- The impact of crossing the threshold (kind of an analogy to ITT) for fuzzy RDD
    - With some assumptions we can scale this by the change in the probability of participating in treatment at the threshold to get the Complier Average Causal Effect of the program at the threshold
    - <u>Compliers</u>: Those who get the treatment only because they are above the threshold
        - Can we identify compliers in the data?

Formal RDD Assumptions

- Requires takings limits as we narrow the window around the threshold.
- Straightforward if we assume treatment effect is homogenous (ACE)
  - Continuity of outcome at threshold if untreated – no natural jump
  - Discontinuity in treatment at threshold
- If effect is heterogeneous, we need more assumptions
  - Assuming treatment effect continuous in z and probability of treatment is independent of size of treatment effect we can estimate complier average causal effect (CACE)
- See additional slides from Monday for formal specifications

Strength of RDDs

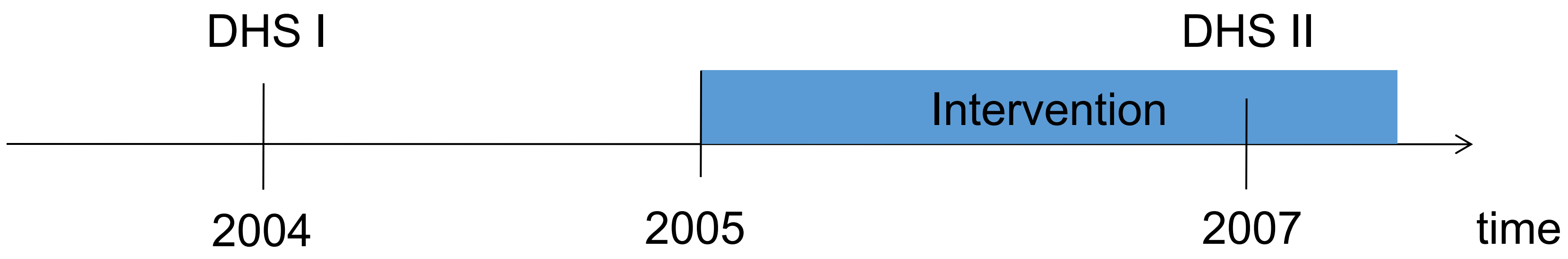
- RDD models have many potential applications
- The assumption that treatment assignment is discontinuous around a very specific threshold is plausible in many settings and can be tested
- The degree to which populations are the same to the “left” and “right” of the threshold can be tested empirically – test balance of observables
- The plausibility of RDD is possible to assess visually – can lend strength to conclusions
- RDD estimate the causal effect of treatments in practice, which may be very different from the controlled environments generated for RCTs

Moving on to ITS -- Impact evaluation

- Impact evaluation attempts to estimate the causal effect of an intervention
- The intervention is usually a policy change that is designed to have an impact, but is not designed to answer research questions
- Given the intervention is happening anyway, we can try to design an evaluation answering the research question: what was the impact of the program on our outcomes of interest?
- Often carried out ex-post using available data

Introduction to impact evaluation

- Example:
  - An NGO has provided vaccines to all newly born children since 2005 in District A in a region experiencing a high burden of diarrheal disease. We find out that there is a (baseline) Demographic and Health Survey (DHS) before and after the intervention.
  - The timeline looks as follows



  - Based on looking at the baseline and post-intervention data, the NGO states: *"In District A, the infant mortality has dropped from 122 per 1000 in 2004 to 97 per 1000 in 2007, highlighting the effectiveness of the program"*
  - Is this convincing?

Pre-Post estimation

- We can measure the outcome of interest at two time points:
  - At baseline, i.e., the pre-intervention measure
  - In the follow-up survey, i.e., the post-intervention measure
- The pre-post estimator of the effect of the intervention is the difference between the post-intervention measure and the pre-intervention measure
  - This is a **single difference**
- The counterfactual for this causal estimate is that the outcome would not have changed without the intervention
- We need a better counterfactual regarding what would have happened without the intervention

Time series data

- Examples of typical time series data would include:
  - Number of cases/deaths reported in each state, month, and year over some period of time (public records)
  - Number of cases/patients treated in a health facility in a each month or day over some period of time (administrative data)
  - Number of Google searches for things like “flu” on a given day (online data)

Structure of time series data

- Generally, a time series is a sequence of data across regular time intervals
- The main unit of analysis is a day or month (in a given area or unit of interest)
- Most typically, outcomes in time series models are count variables, such as the number of events occurring in a specific time window
- One observation per time period

Main logic of interrupted time series analysis

- Time series data allows us to carefully model the general time trends in our outcome of interest

- Most programs/policies are introduced at a very specific time point: if the policy worked, we should observe that the outcome shifts onto a different trajectory after the policy change happened

- The main assumption of this interrupted time series (ITS) approach is that the pre-intervention trend can be continued after the intervention as the counterfactual

What is the counterfactual in an ITS

- In a randomized trial the counterfactual is what happens in the control group – we think this is what would have happened without the treatment
- In ITS, we fit a trend to the data in the pre intervention period – the counterfactual is that the data would have follows this trend in the post intervention period without the trend
- Assumes we have adequately fit the trend in the data
- Assumes no other new interventions happening in the post intervention period that could affect the outcome

Ideal conditions for ITS

In order to implement ITS we need the following conditions

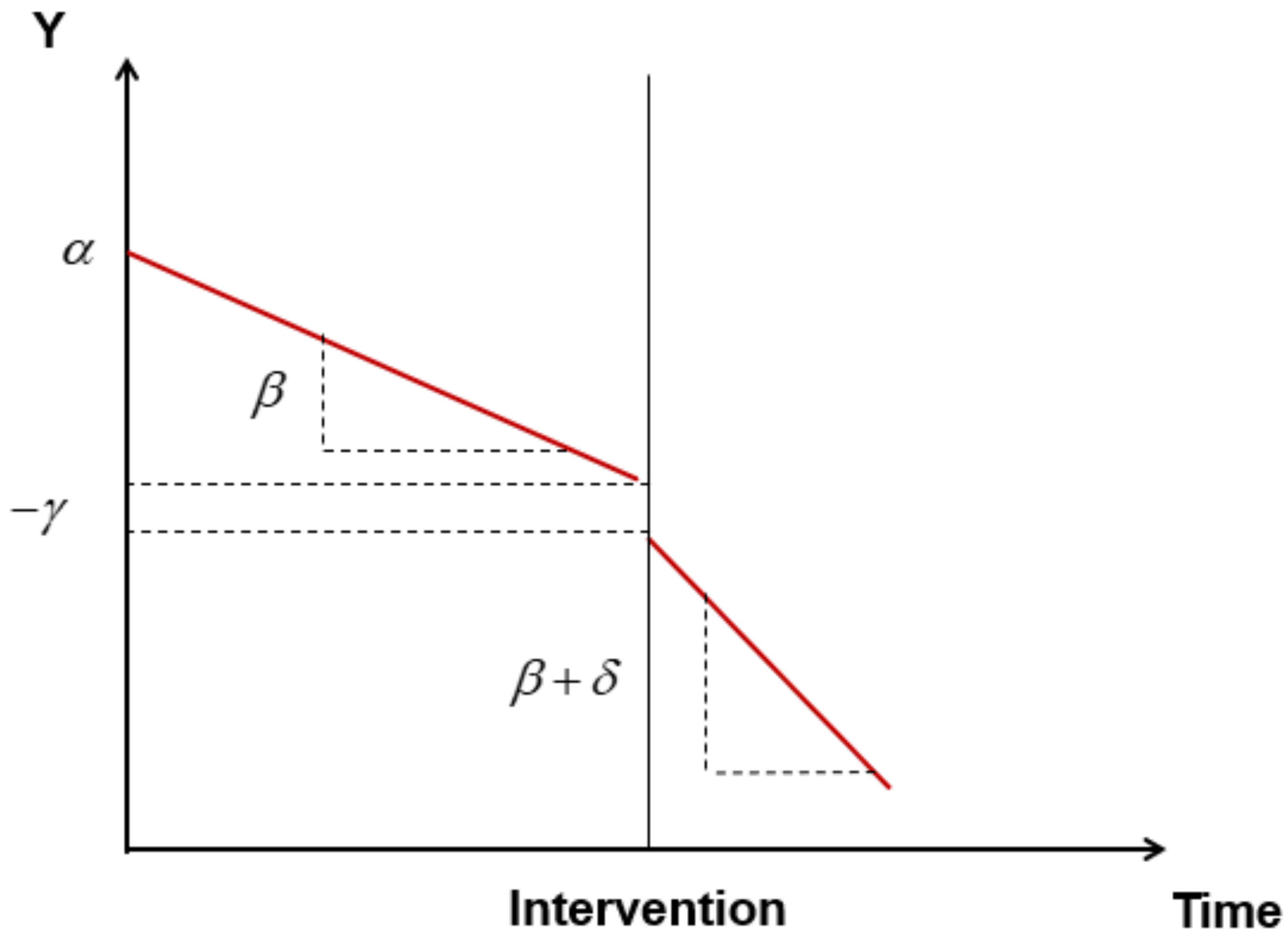
- A clearly defined pre- and post-treatment period
- High frequency, time series data on the outcome of interest
- (Ideally) an outcome where we anticipate that
    - The outcome will have a large response to the policy change
    - The response to the policy change will either be very rapid or with a clearly defined lag

A simple model for intercept and slope in interrupted time series

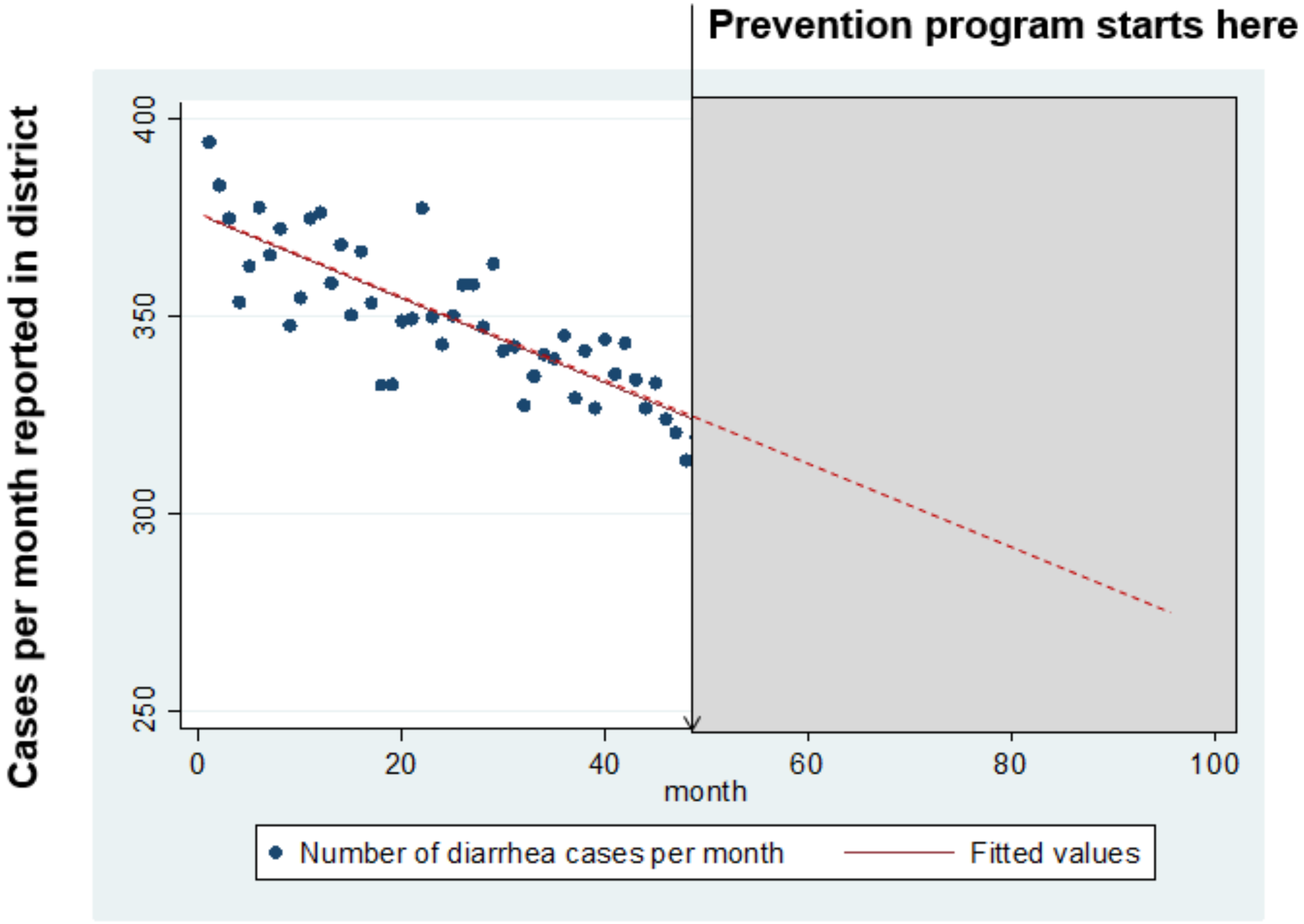
$$Y_t = \alpha + \beta time + \gamma Post + \delta Post * time + \epsilon_t$$

- Y_t: Outcome of interest at time t
- $Post$: an indicator for the "post-policy change" period
- α: level of the outcome Y at time $t = 0$
- β: linear time trend
- γ: level shift post-policy change
- δ: gradient shift post-policy change

Interpretation of coefficients

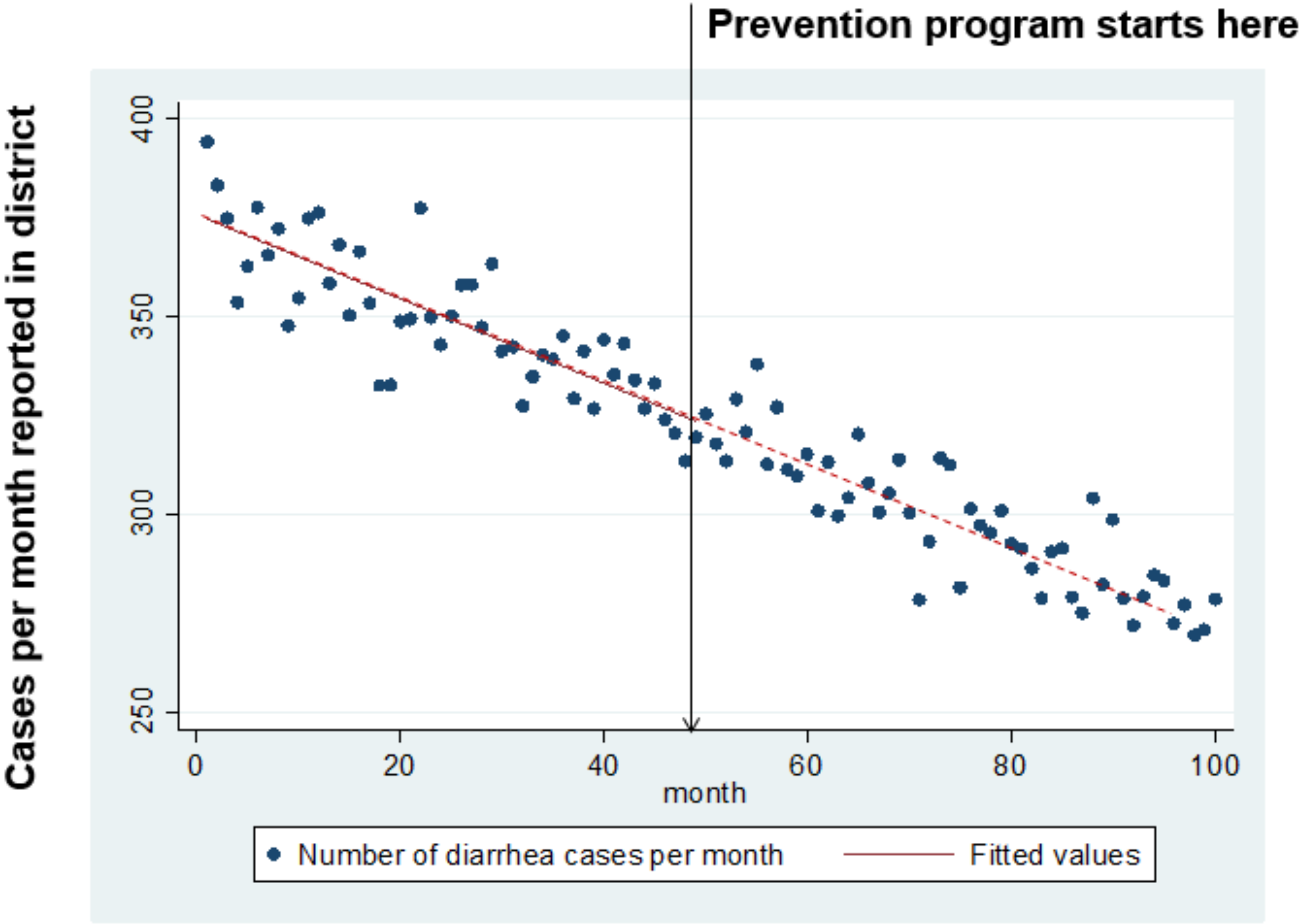


Example 1: Diarrhea incidence



Example 1: Diarrhea incidence

- How would α, β, γ, and δ look here?



What would you conclude in a pre-post evaluation design about the program's impact?

Example 2: Smoking ban and admissions for acute coronary events (ACE)

- **Law change:** End indoor smoking
- **Outcome:** Hospital admissions for acute coronary events

- How well does the policy change + outcome fit the ideal conditions for ITS?
  - Would we expect a large change in the outcome from this change in law?
  - Would we expect the change to be either immediate or with a predictable lag?

Example 2: Smoking ban and admissions for acute coronary events (ACE)

- **Law change:** End indoor smoking
- **Pre-intervention period:** Jan 2002 – Dec 2004
- **Post-intervention period:** Jan 2005 – Dec 2006
- **Source:** Bernal et al, 2017
- What is your visual assessment?

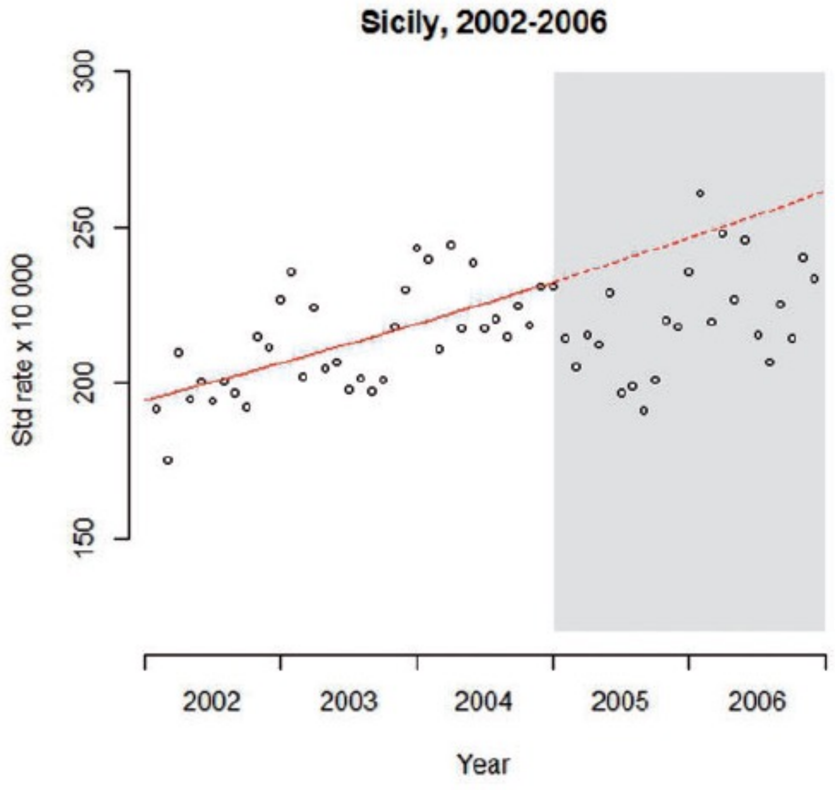


Figure 1 Scatter plot of example dataset. Standardized (Std) rate of ACE over time. White background, pre-intervention period; grey background, post-intervention period; continuous line, pre-intervention trend; dashed line, counterfactual scenario

Example 2: Smoking ban and admissions for acute coronary events (ACE)

- **Estimated impacts:**
  - Immediate reduction
  - No change in trend

- **Source:** Bernal et al, 2017

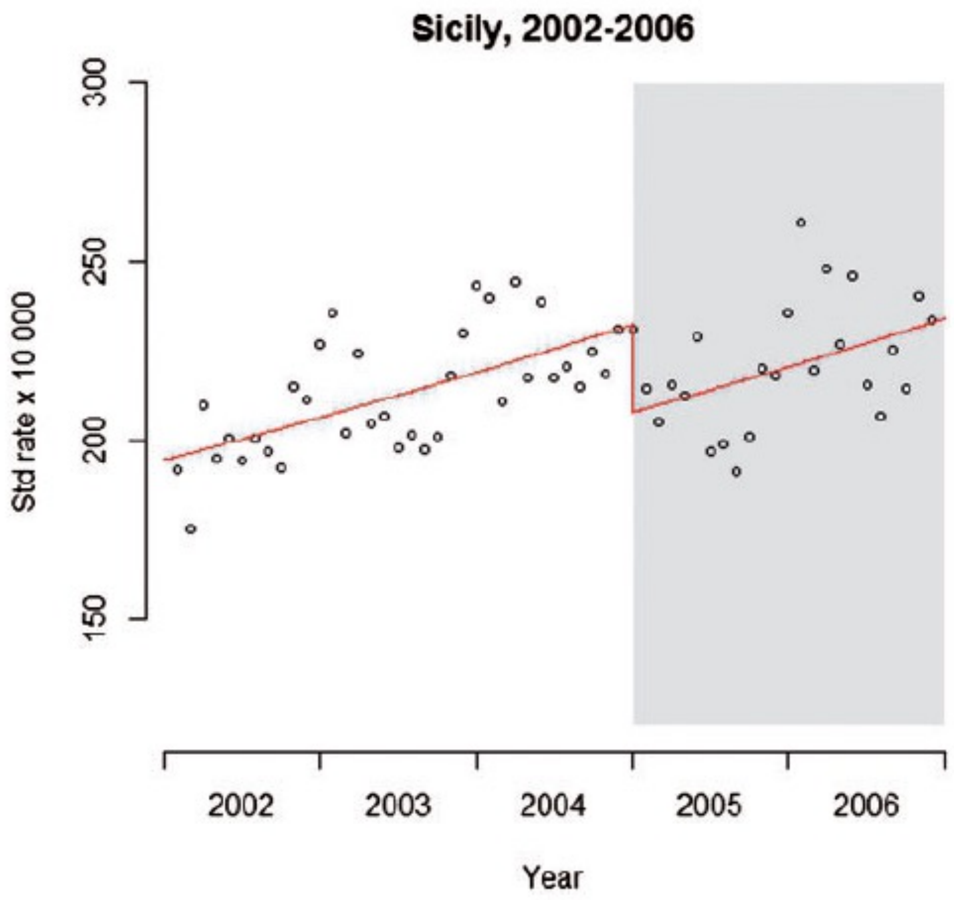
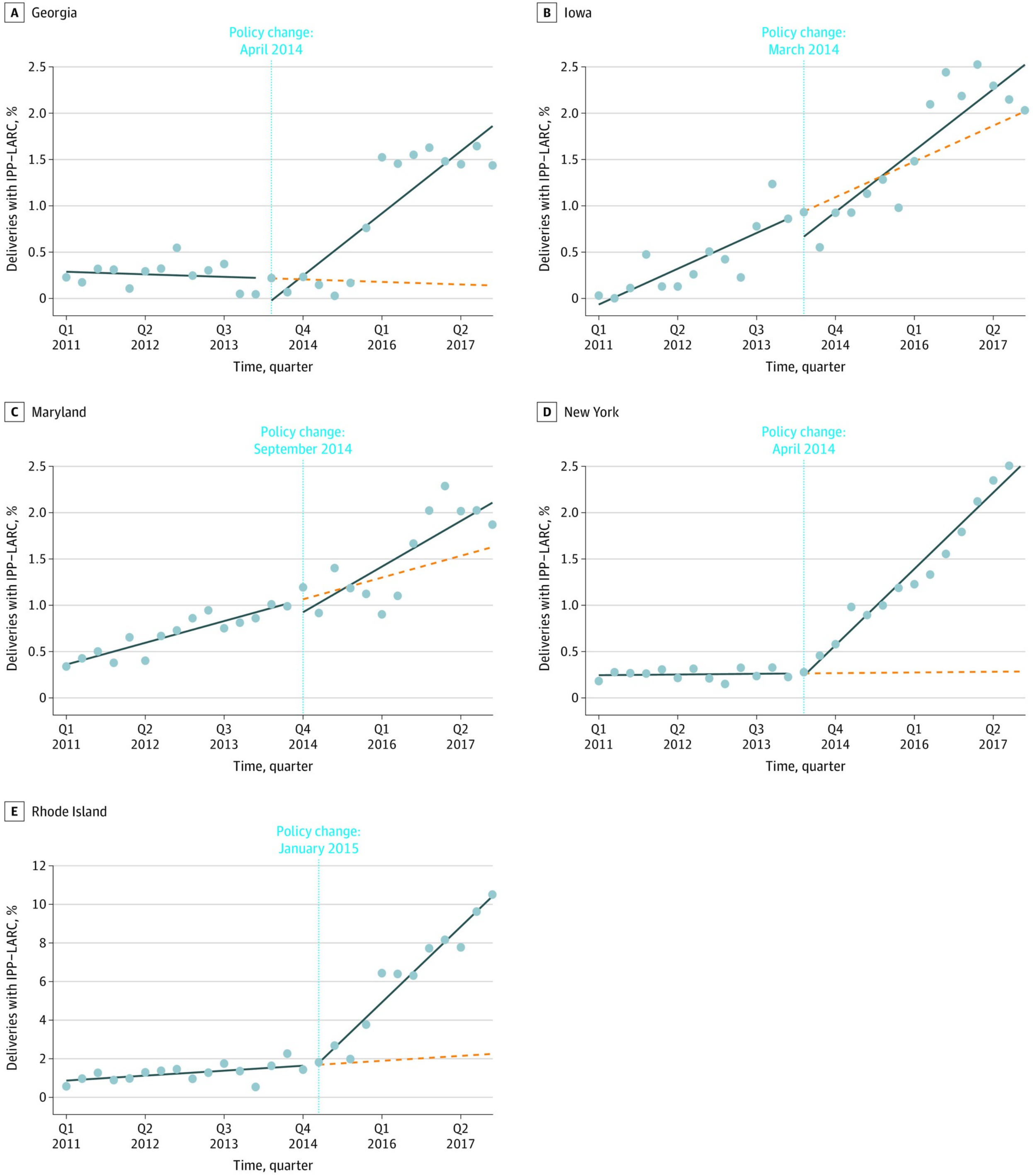


Figure 3 Interrupted time series with level change regression model. Line: predicted trend based on the unadjusted regression model

Example 3: Change in payment for immediate postpartum LARCs

- *Policy change*: change in Medicaid policy that provided reimbursement for immediate postpartum long-acting reversible contraception methods (IPP-LARC)
- Outcome: % of Medicaid-insured deliveries with IPP-LARC
- How well does the policy change + outcome fit the ideal conditions for ITS?
  - Would we expect a large change in the outcome from this change in law?
  - Would we expect the change to be either immediate or with a predictable lag?



- ITS can accommodate very different pre-period trends
- Why would there be different patterns across different states?

Challenge 1: Getting the timing right

- In some instances, the impact of the policy change should be reasonably immediate conditional on the policy being implemented as stated
  - For example: handwashing on diarrhea, or speeding laws on traffic accidents

- In many other cases, it is not clear how soon we should expect an impact on outcomes
  - Particularly challenging for outcomes like cancer or children's height which are the result of longer-term processes

- There may possibly be anticipation effects as well when policy is announced rather than implemented
  - Conversely it may take time for people to become aware of and respond to policy changes

- We need to have a clear conceptual framework for when we would expect impact relative to the timing of the reform

Specifying the impact model

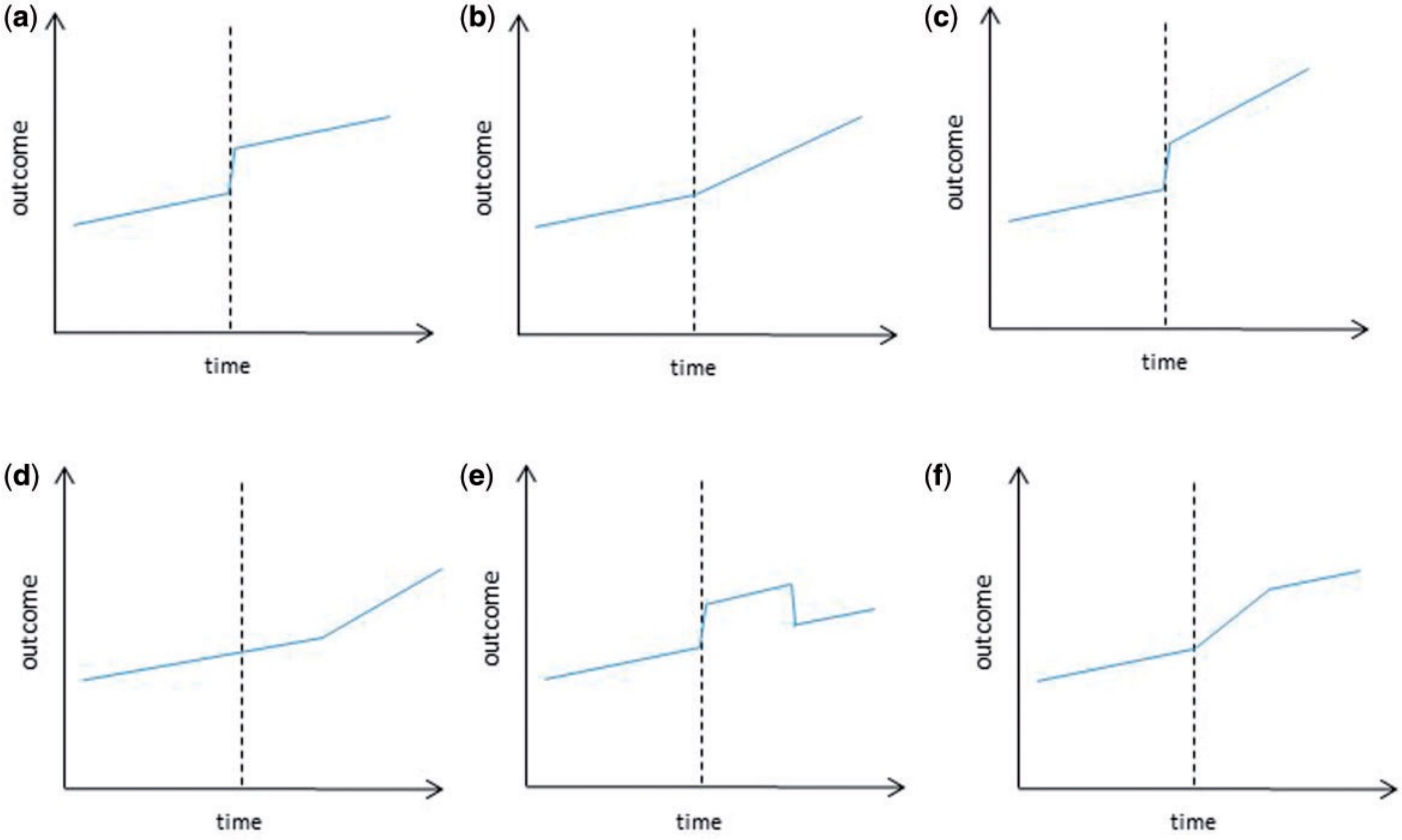
- Before running actual regressions, we need to be clear regarding our hypotheses on how the policy change affected the outcomes of interest

- **Intercept shift:** are we expecting an immediate shift in the outcome?

- **Slope shift:** are we expecting outcome trajectories to change?

- **Timing and duration change:** are we expecting changes to start immediately or to start with a lag? To last, or to fade out?

Potential changes in outcome trajectories



Source: Bernal et al, 2017

Challenge 2: Seasonality of outcomes

- Most time series data has daily, monthly, or quarterly records with rather pronounced seasonality in the outcome
- If seasonal effects are not controlled for, impact estimates are likely to be biased because programs are likely to start in either high or low incidence months (with biases in the opposite direction)
- In regression models, we can:
    - Model the seasonality directly by using “season dummies”: i.e. an indicator variable for each month
    - Create a “de-trended” time series by first regressing outcome on trend model and season effects, and then just analyzing deviations from trend

Challenge 3: Confounding concerns

- One of the main strengths of ITS is that we are only comparing units of interest to themselves, and thus we don't worry about cross-sectional confounding

- The main weakness is the concern that the policy change coincides with other things happening which may affect the outcome
  - For example
    - Some of the LARC policy changes occurred in 2014
    - This is also when the Affordable Care Act allowed expansion of Medicaid to more low-income adults
    - How might this affect the interpretation of the results?

- One solution is a comparator outcome
  - Outcome that should *not* be affected by the policy change but would be affected by confounding events

- Another solution – control or comparison region/group
  - Controlled ITS models
  - Synthetic control (we will get to this next week)

Challenges with inference with ITS – problems with standard errors

- Even when we de-trend the data, it seems somewhat likely that two adjacent observations (observed value today and observed value last month) are not independent

- Autocorrelated errors still give unbiased coefficient estimates, but will likely result in incorrect standard errors
  - Autocorrelation means less “information” in each observation than if the data are independent
  - Estimated standards errors tend to be too small if we incorrectly assume independent observations

- While autocorrelation may be less common in public health time series data, it is possible to test for it under some assumptions (Ex: Durban Watson test)

Controlled interrupted time series (CITS)

- Given the confounding concerns about confounding with other things happening at the same time → would be nice to have control regions
  - These areas did not get the treatment and should not see a treatment effect
  - If we do see treatment effect in control areas, suggests the "intervention" effect due to some common factor affecting both the treated & control areas
- Assume we have a second region/hospital that did not get the intervention. This is the control group.
- Counterfactual – jump and change in trend in the control group would have been the same as in the treatment group without treatment

Controlled Interrupted Time Series

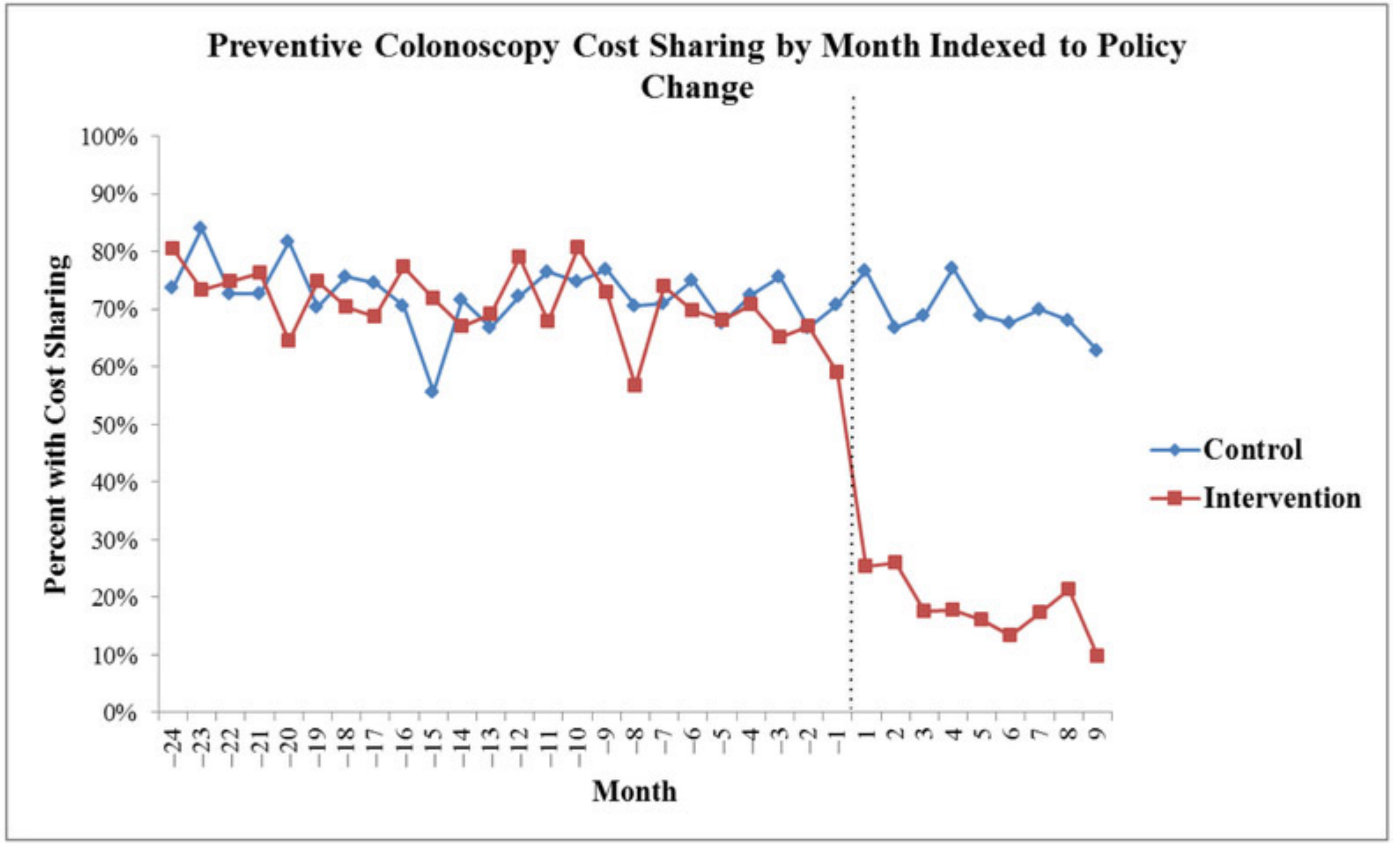
$$Y_{it} = \alpha + \beta_1 time + \beta_2 Post + \beta_3 Post * time + \beta_4 Treated + \beta_5 time * Treated + \beta_6 Treated * Post + \beta_7 time * Post * Treated + \epsilon_{it}$$

- Y_{it}: Outcome of interest at time t in region i
- $Post$: an indicator for the "post-policy change" period
- *Treated*: an indicator for the region "treated" by the policy change
- β_4: difference in levels at time $t = 0$
- β_5: difference in slope at baseline
- β_6: difference in intercept shift post policy change (relative to shift in control)
- β_7: difference in slope change (relative to change in slope in control)

Challenges with Controlled ITS

- Control group does not need to be identical to treatment group in the pre-intervention period
  - Different intercept
  - Different slope
- Counterfactual: changes from trend seen in the control group in the post intervention period would have occurred in the intervention group without the intervention
- Less convincing if the control group is dissimilar to the treatment group
- We have to <u>really</u> believe our modeling of trends in the control group

CITS Example



Policy change: elimination of cost sharing for cancer screening post ACA

Control group: ”grandfathered” plans not required to change cost-sharing

How plausible is the identifying assumption?

Control group trends would be expected to continue for treatment group without intervention

Am J Manag Care. Author manuscript; available in PMC 2018 Nov 7.
Published in final edited form as:
Am J Manag Care. 2015 Jul; 21(7): 511–517.
PMCID: PMC622041
NIHMSID: NIHMS98830
PMID: 2624774

ACA-Mandated Elimination of Cost Sharing for Preventive Screening Has Had Limited Early Impact

Shivan J. Mehta, MD, MBA,[1] Daniel Polsky, PhD,[1] Jingsan Zhu, MBA,[1] James D. Lewis, MD, MSCE,[1] Jonathan T. Kolstad, PhD,[2] George Loewenstein, PhD,[3] and Kevin G. Volpp, MD, PhD[1]

Summary: Interrupted time series

- ITS analysis offers an interesting empirical approach to assessing policy impact when we have routinely collected temporal data
- The main identification assumption is that the timing of the reform is not correlated with other things affecting the outcome and observed trends in the data would have continued without the intervention
- **ITS works best if the policy start is clearly defined and impact is immediate and large**

Summary: Interrupted time series

- CITS can offer a more robust alternative that controls for confounding changes that happened at the same time

- Requires a credible control group that experienced the same confounding changes but not the treatment we are interested in
  - This is not always easy to find
  - We will discuss alternative strategy for finding a control group (synthetic control) next week

Recommended readings

- Bernal, Cummins, and Gasparrini (2017). Interrupted time series regression for evaluation of public health interventions: a tutorial [Link]
- Bhaskaran, Gasparrini, Hajat, Smeeth, and Armstrong (2013). Time series regression studies in environmental epidemiology [Link]

Additional Slides

Over-dispersion in ITS

- Given that most time series outcomes are count variables, the use of Poisson models seems ideal

- In many cases, we may be worried about over-dispersion, i.e., the Poisson assumption that mean of the outcome = variance of the outcome
  - This is easy to handle in our regression models (see notes from PHS 2000A)

- For large samples, mean will be normally distributed (Law of Large Numbers)

- Over-dispersion is not a problem if linear regression models are used

Autoregressive integrated moving average (ARIMA) models in ITS

- In many instances, time series can be best predicted by their own lagged values as well as previous deviations from their usual trends:
  - Autoregressive: a function of past values
  - Moving average: a function of the average of several previous values
  - Integrated: a function of previous changes in values

- These models are very flexible, easy to implement in practice, and may help remove complex serial correlation structures