Evaluating impacts of population shocks using an interrupted time series approach

Alison Gemmill

June 12, 2023

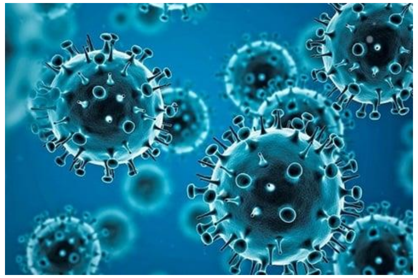
SPER

(Some slides borrowed from Tim Bruckner at UC Irvine

And Liz Stuart at Johns Hopkins)

Rationale

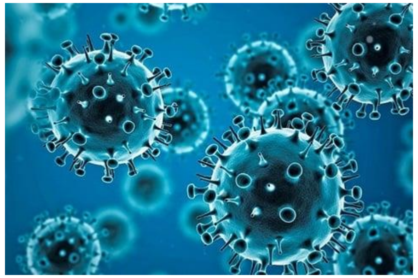
- Population is the unit of interest
- Use of time series data (monthly, weekly, etc.)
- Interruption has a well-defined time of onset
  - Policy evaluation
  - Acute population shocks (COVID-19, Great Recession, Terrorist attacks)
- Strong casual inference/natural experiments



ACOG
The American College of Obstetricians and Gynecologists

Distinct advantages

- Denominators are not available
- No comparison group within the population exists (although ITS can also be used with comparison groups)



ACOG
The American College of
Obstetricians and Gynecologists

General approach of an ITS design

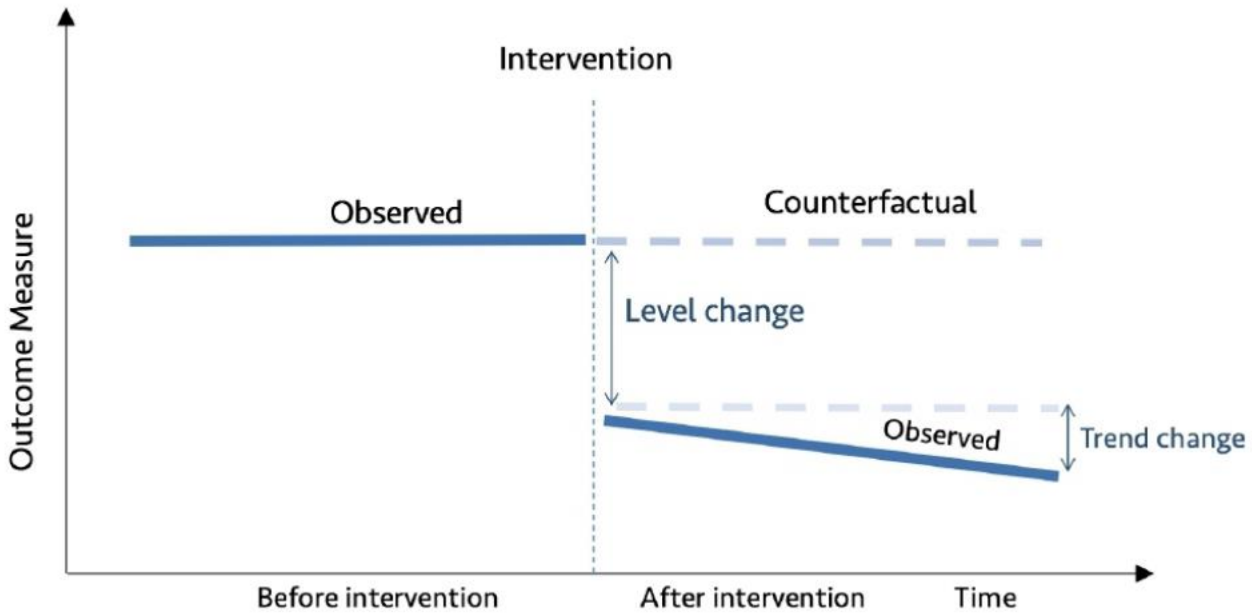


Figure 1 Diagrammatic representation of single interrupted time series.

- This method allows you to compare observed values of preterm birth with *counterfactual values* extrapolated from patterns in the pre-shock data

Source: Hategeka et al. (2020), BMJ Global Health

The nuts and bolts

Interrupted time series regression for the evaluation of public health interventions: a tutorial

James Lopez Bernal,[1,*] Steven Cummins,[1] Antonio Gasparrini[1]

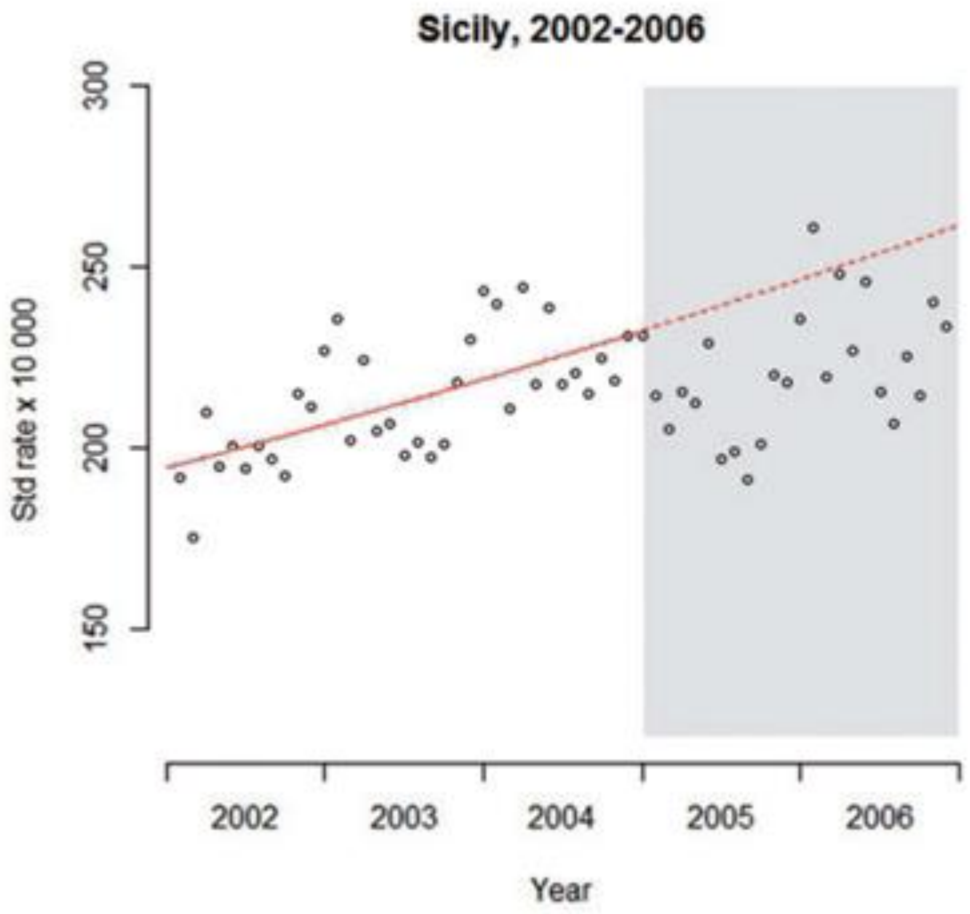
Bernal et al. 2017 *IJE*

Table 1. Excerpt from the example dataset

Year	Month	Time elapsed (T)	Smoking ban[a] (X)	ACEs (Y)	Std popn
2004	1	25	0	914	381656.3
2004	2	26	0	808	383680
2004	3	27	0	937	383504.2
2004	4	28	0	840	386462.9
2004	5	29	0	916	383783.1
2004	6	30	0	828	380836.8
2004	7	31	0	845	383483
2004	8	32	0	818	380906.2
2004	9	33	0	860	382926.8
2004	10	34	0	839	384052.4
2004	11	35	0	887	384449.6
2004	12	36	0	886	383428.4
2005	1	37	1	831	388153.2
2005	2	38	1	796	388373.2
2005	3	39	1	833	386470.1
2005	4	40	1	820	386033.2
2005	5	41	1	877	383686.4
2005	6	42	1	758	385509.3
2005	7	43	1	767	385901.9
2005	8	44	1	738	386516.6
2005	9	45	1	781	388436.5
2005	10	46	1	843	383255.2
2005	11	47	1	850	390148.7
2005	12	48	1	908	385874.9

ACEs, hospital admissions for acute coronary event; Std popn, age-standardized population in person-years.[16]

[a]Smoking ban: 0, smoking ban not in place; 1, smoking ban in place.



Bernal et al. 2017 *IJE*

Basic ITS approaches: Segmented regression

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 TX_t$$

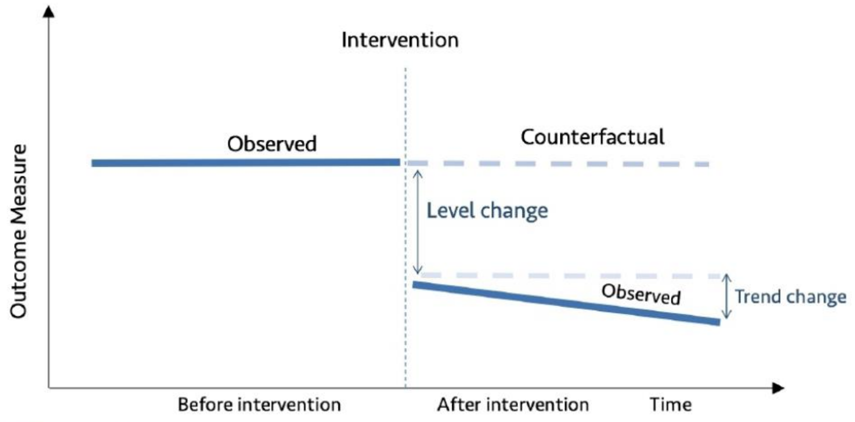
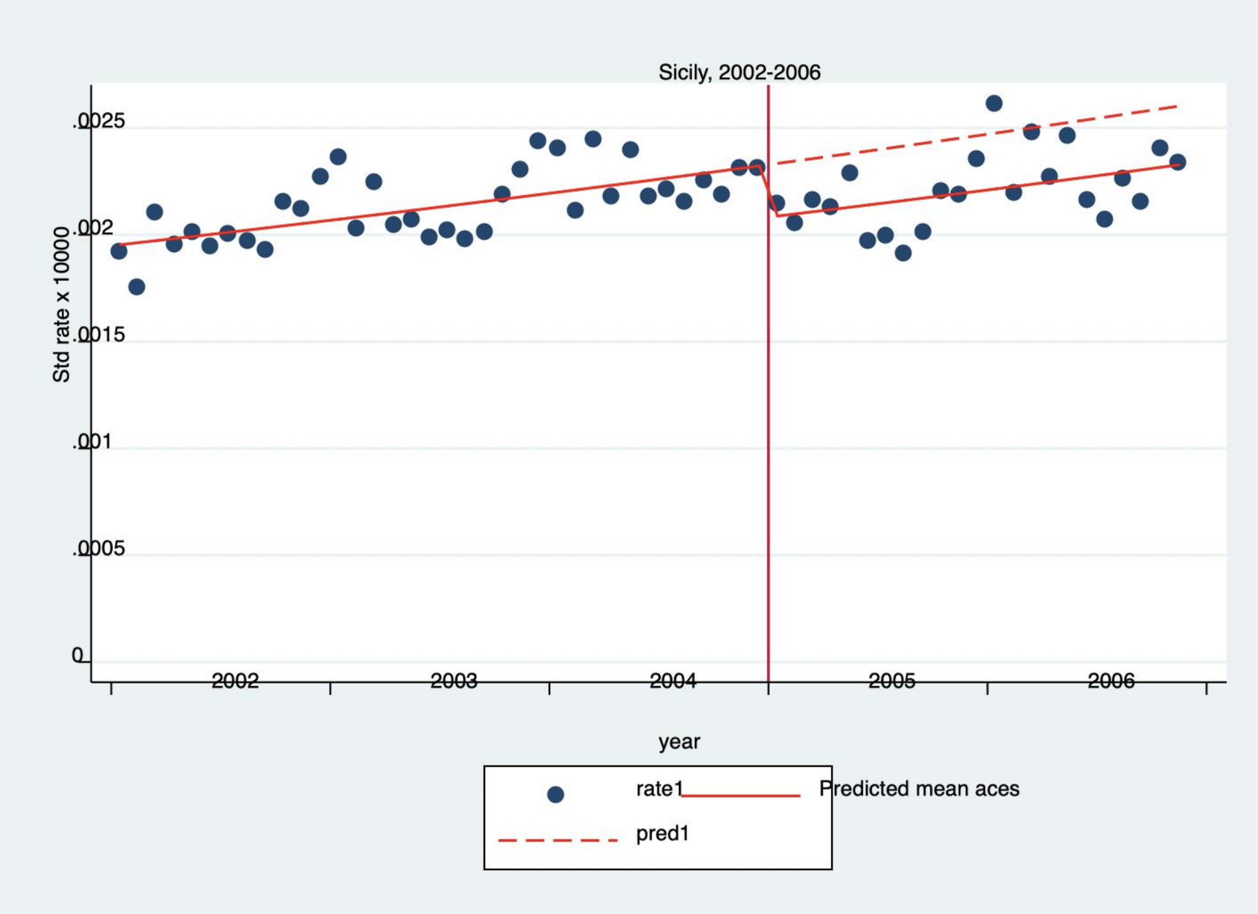


Figure 1 Diagrammatic representation of single interrupted time series.

- β_0 represents the baseline level at $T = 0$
- β_1 is interpreted as the change in outcome associated with a time unit increase (representing the underlying pre-intervention trend)
- β_2 is the level change following the intervention and β_3 indicates the slope change following the intervention (using the interaction between time and intervention: TX_t)

Bernal et al. 2017 *IJE*

Example: Segmented regression results



Pitfalls

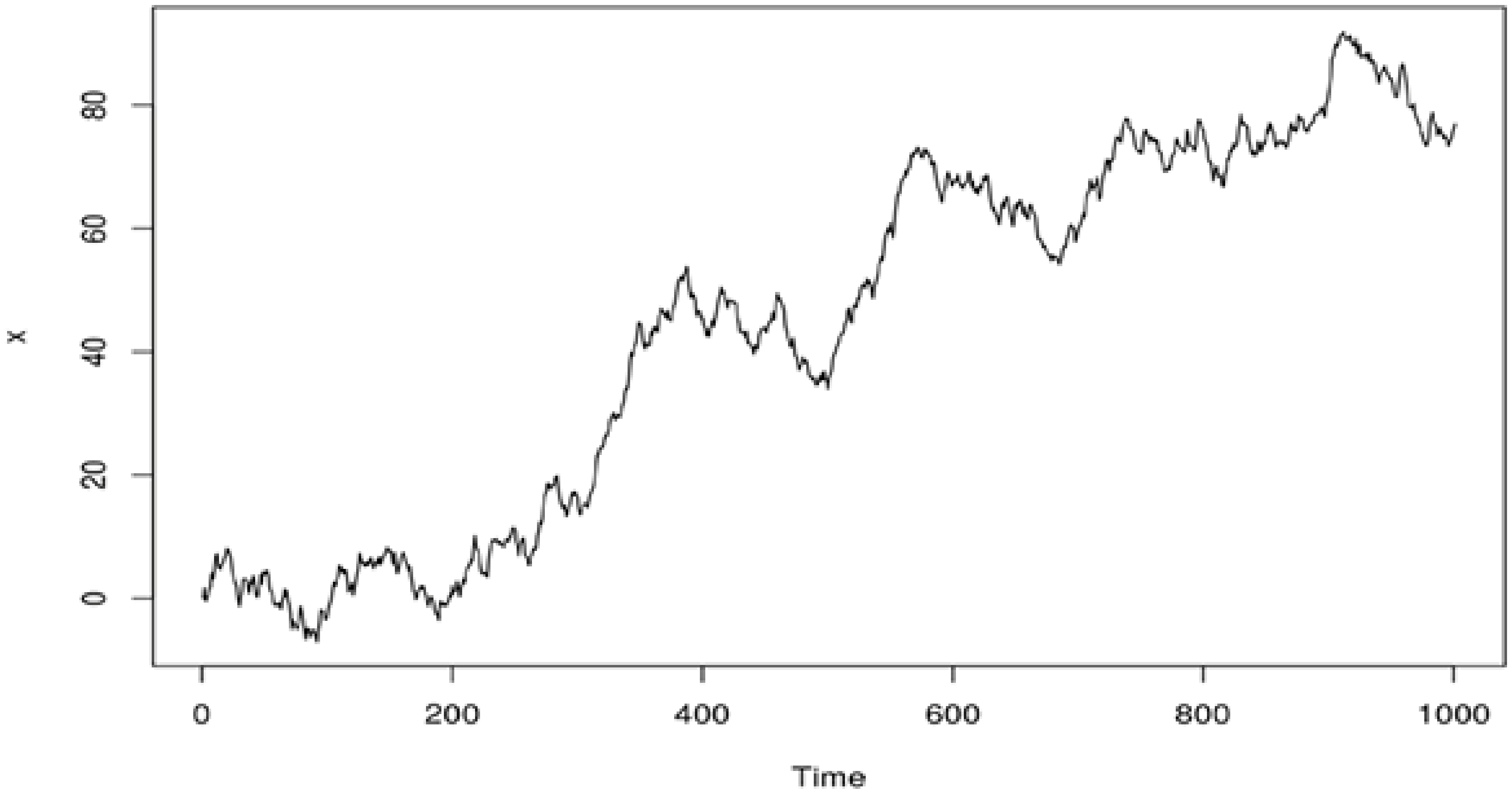
- Does not account for seasonality or other types of autocorrelation
- May not capture short-term changes or fluctuations after interruption

Bernal et al. 2017 *IJE*

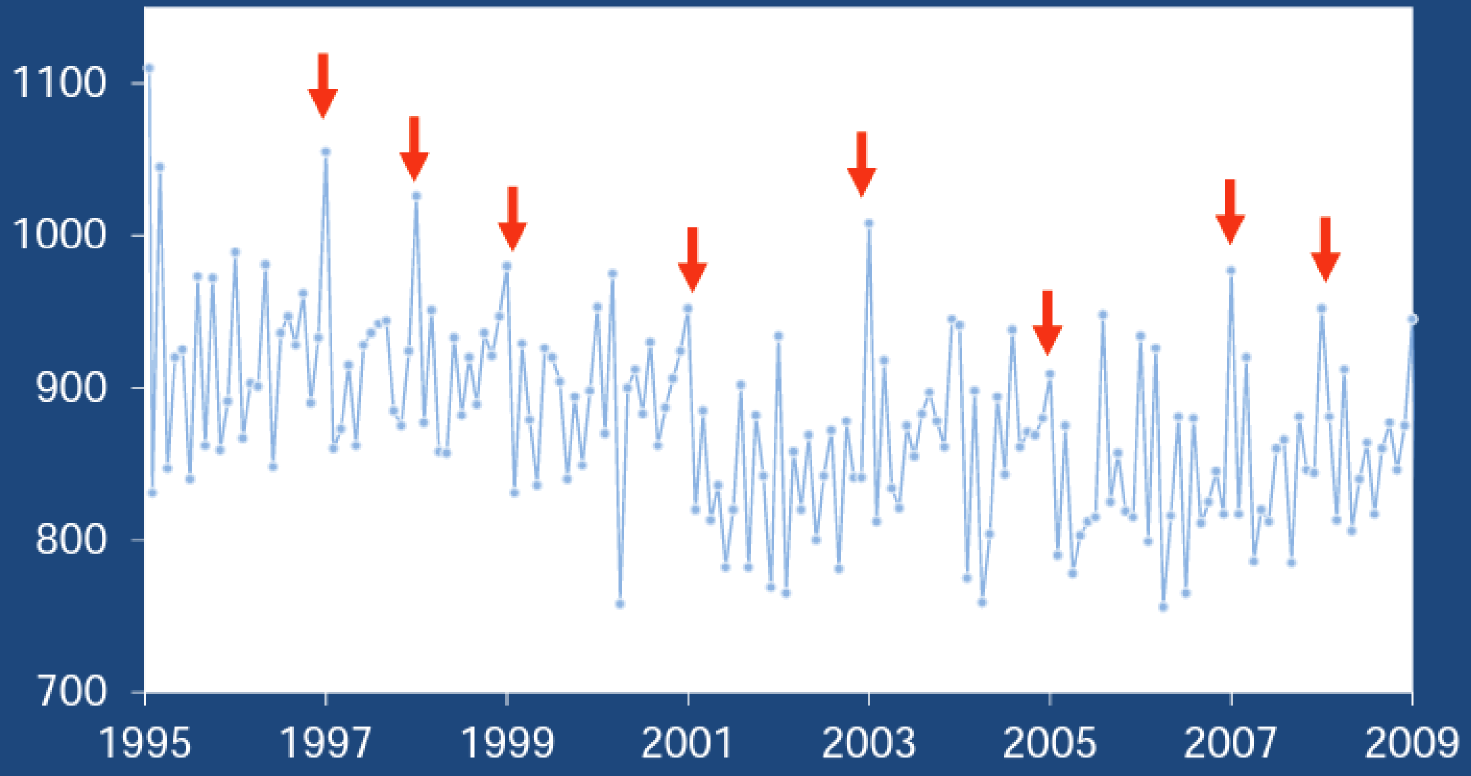
Importance of accounting for autocorrelation in time series data

- Patterns in outcome variable may include trend, seasonality, and other autocorrelation "signatures"
- Failure to identify and control for autocorrelation in the pre-intervention period often leads to falsely attributing an "effect" to the intervention itself
  - Or, leads to artificially precise standard errors
- Strengthens causal inference (especially when there's no comparison group)

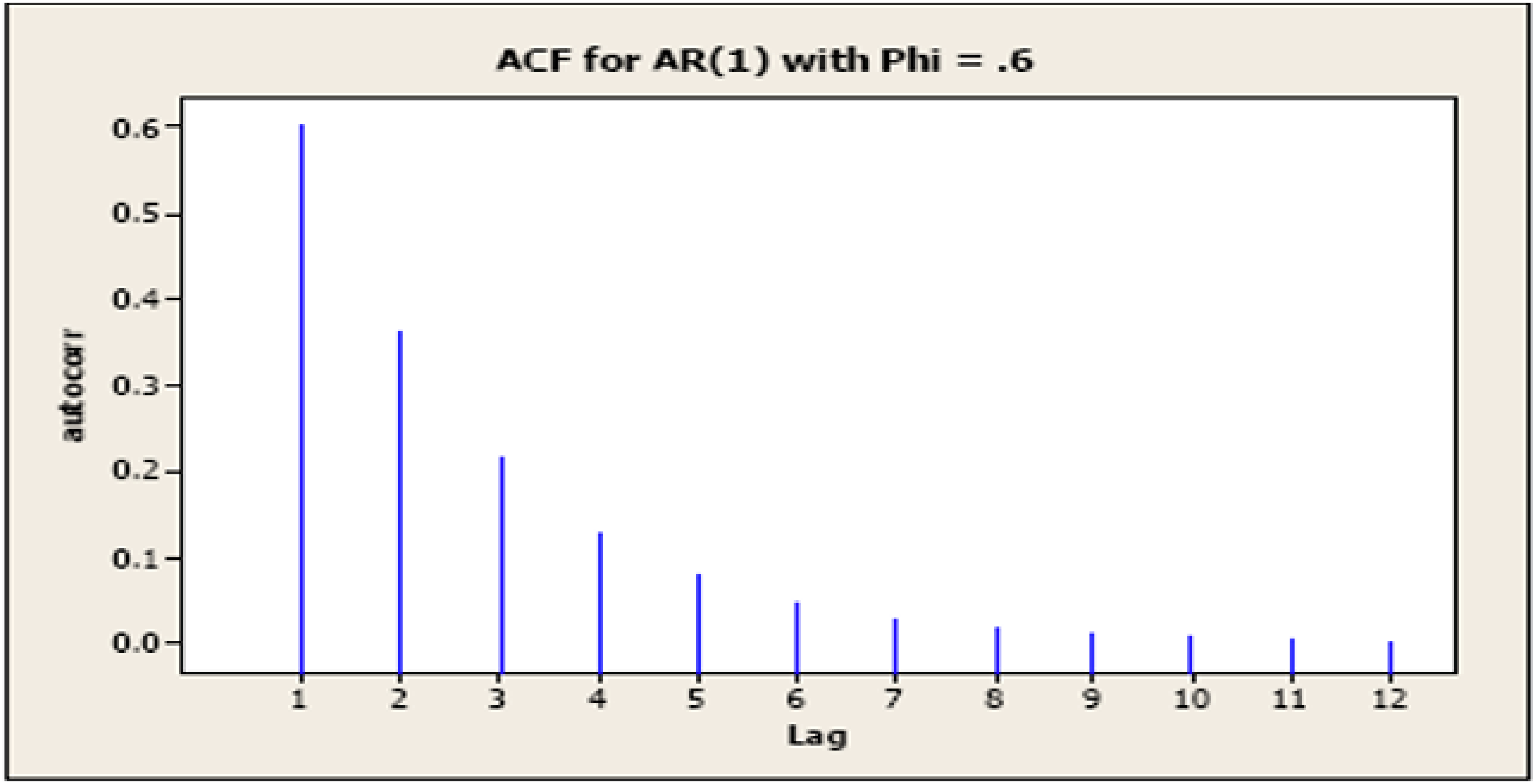
Trend



Seasonality



Memory


ACF for AR(1) with Phi = .6
autocorr
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
2
3
4
5
6
7
8
9
10
11
12
Lag

Logic of ITS that accounts for autocorrelation

- Identify autocorrelation of an outcome (Y) *before* intervention to derive statistically expected values of Y after intervention
  - Counterfactual is derived from the history of Y

- Earlier values of Y are used to remove patterns, so that expected values of residuals = 0

- Intervention (X) may cause Y only if it predicts Y better than the history of Y itself
  - Granger causality; conservative

Generating counterfactuals that account for autocorrelation

1. Yearly time trends + month fixed effects

Early assessment of the relationship between the COVID-19 pandemic and births in high-income countries

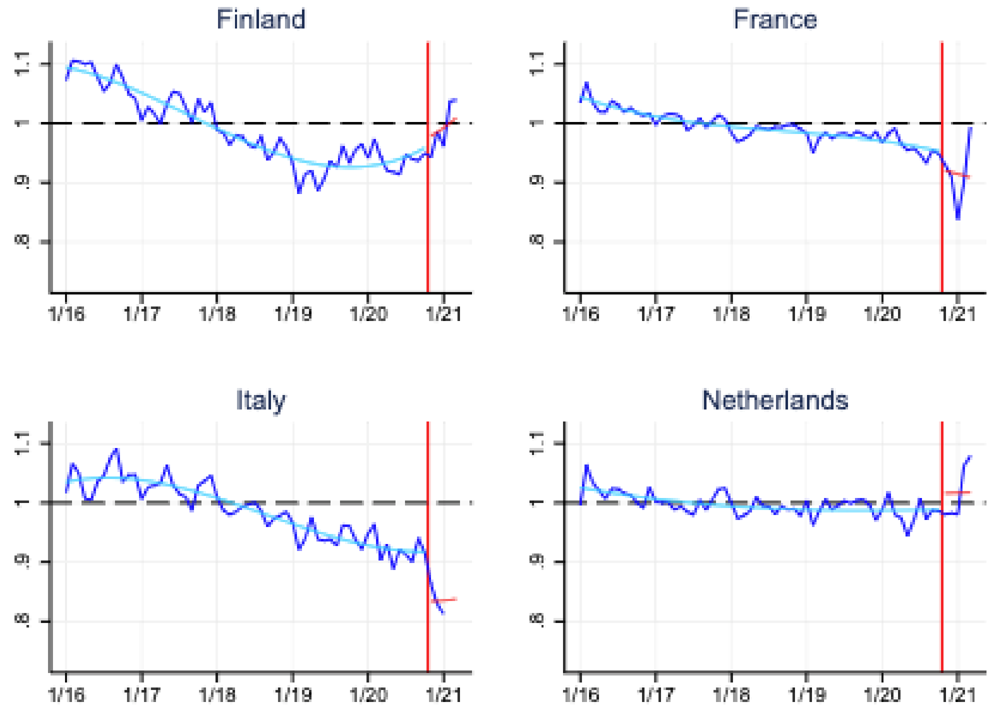
Arnstein Aassve[a,b,1], Nicolò Cavalli[b,c], Letizia Mencarini[a,d], Samuel Plach[a], and Seth Sanders[e]

Subnational variations in births and marriages during the COVID-19 pandemic in South Korea

Myunggu Jung

D. Susie Lee

$$GFR_t \ or \ GMRf_t = a_m + \beta COVID19_t + \sum_{i=1}^{3} \gamma Time_t^i + \epsilon_t$$



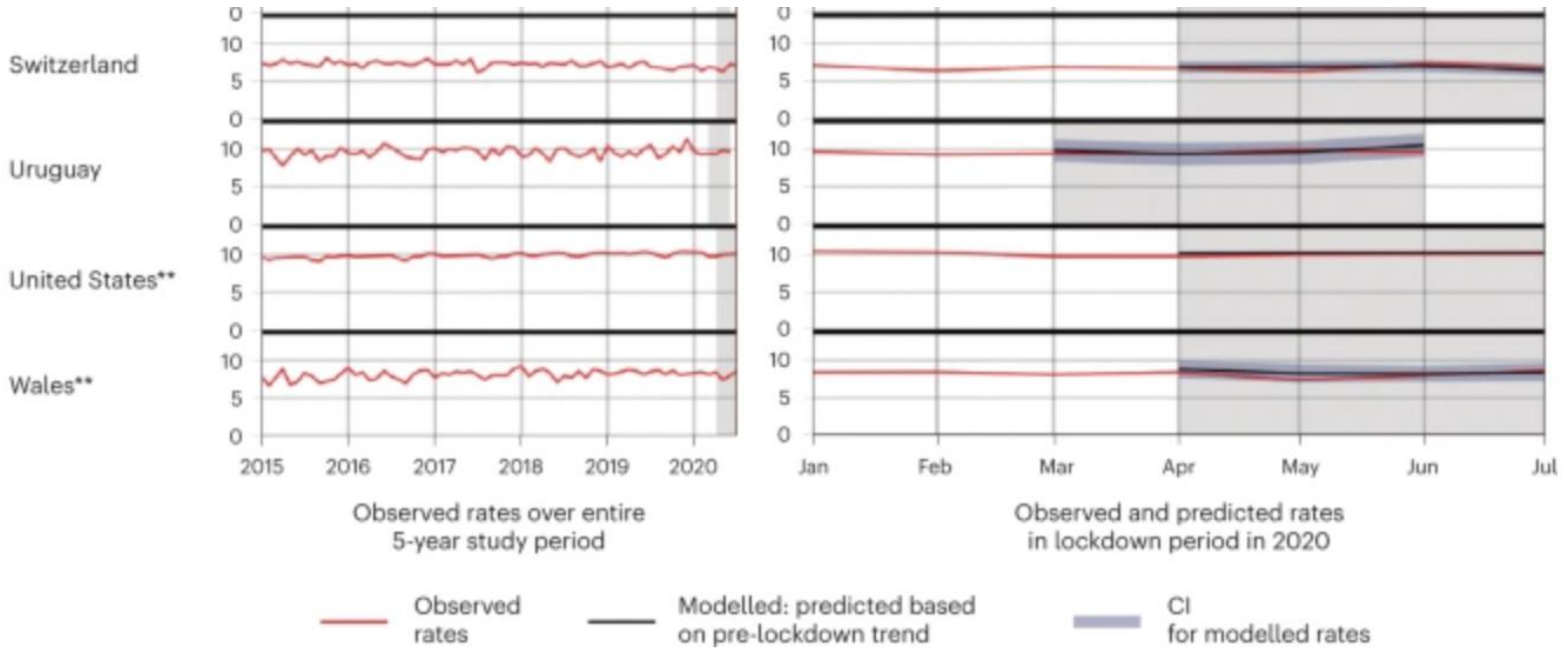
- a_m denotes fixed effects of month m
- COVID19 is a dummy variable equal to 1 from November 2020 and 0 otherwise
- $Time_t$ refers to the month and the year; γ^i are estimates of the linear, quadratic, and cubic time trends
- Coefficent β estimates the average changes in the GFRs or GMRfs between the pre-pandemic and during-pandemic periods after controlling for temporal trends and seasonality.

nature human behaviour

Article https://doi.org/10.1038/s41562-023-01522-y

Changes in preterm birth and stillbirth during COVID-19 lockdowns in 26 countries

- Accounted for seasonality with month fixed effects
- Trend evaluated using either a linear, square, quadratic, logarithmic, or second-order polynomial effect



Patterned Outcomes, Unpatterned Counterfactuals, and Spurious Results: Perinatal Health Outcomes Following COVID-19

Alison Gemmill*, Joan A. Casey, Claire E. Margerison, Jennifer Zeitlin, Ralph Catalano, and Tim A. Bruckner

- Be careful with month/year fixed effects!

2. Adding Fourier terms

GYNECOLOGY

The impact of the COVID-19 pandemic on abortion care utilization and disparities by age

Isabel R. Fulcher, PhD; Chiamaka Onwuzurike, MD; Alisa B. Goldberg, MD, MPH; Alischer A. Cottrill, BSc; Jennifer Fortin, MPH; Elizabeth Janiak, ScD

$$\log(E[Y_t|t]) = \beta_0 + \beta_1 t + \beta_2 days_t + \beta_3 \cos\left(\frac{2\pi t}{12}\right) + \beta_4 \sin\left(\frac{2\pi t}{12}\right) + \beta_5 \cos\left(\frac{4\pi t}{12}\right) + \beta_6 \sin\left(\frac{4\pi t}{12}\right)$$

- Modeled monthly counts of abortions (Y_t) using a GLM w/ log link and Poisson distribution
- Time t for each month captures trends
- $Days_t$ is a term for number of days that clinic was open in month t
- Rest of terms are Fourier terms used to capture seasonality (pairs of sine and cosine functions)

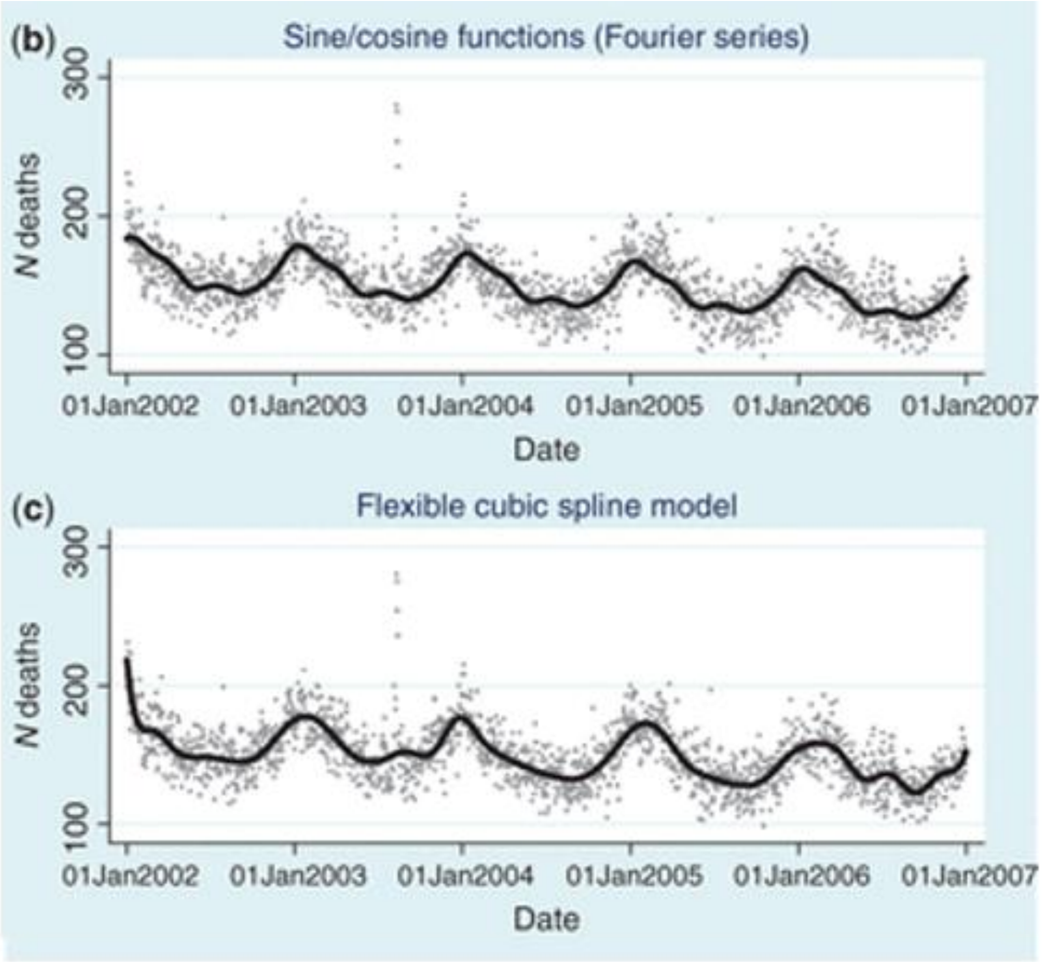
3. Adding flexible spline functions

- Uses a number of different polynomial (most commonly cubic) curves that are joined smoothly end-to-end to cover the full period
- In generating the spline basis, it is necessary to decide how many knots (join-points) there should be, which governs how many end-to-end cubic curves will be used and therefore how flexible the curve will be
  - Too few → fail to capture the main long-term patterns closely
  - Too many → can result in a very ‘wobbly’ function which may compete with the variable of interest to explain the short-term variation of interest, widening confidence intervals of relative risk estimates

Bhaskaran et al. 2013

More on Fourier and Spline functions

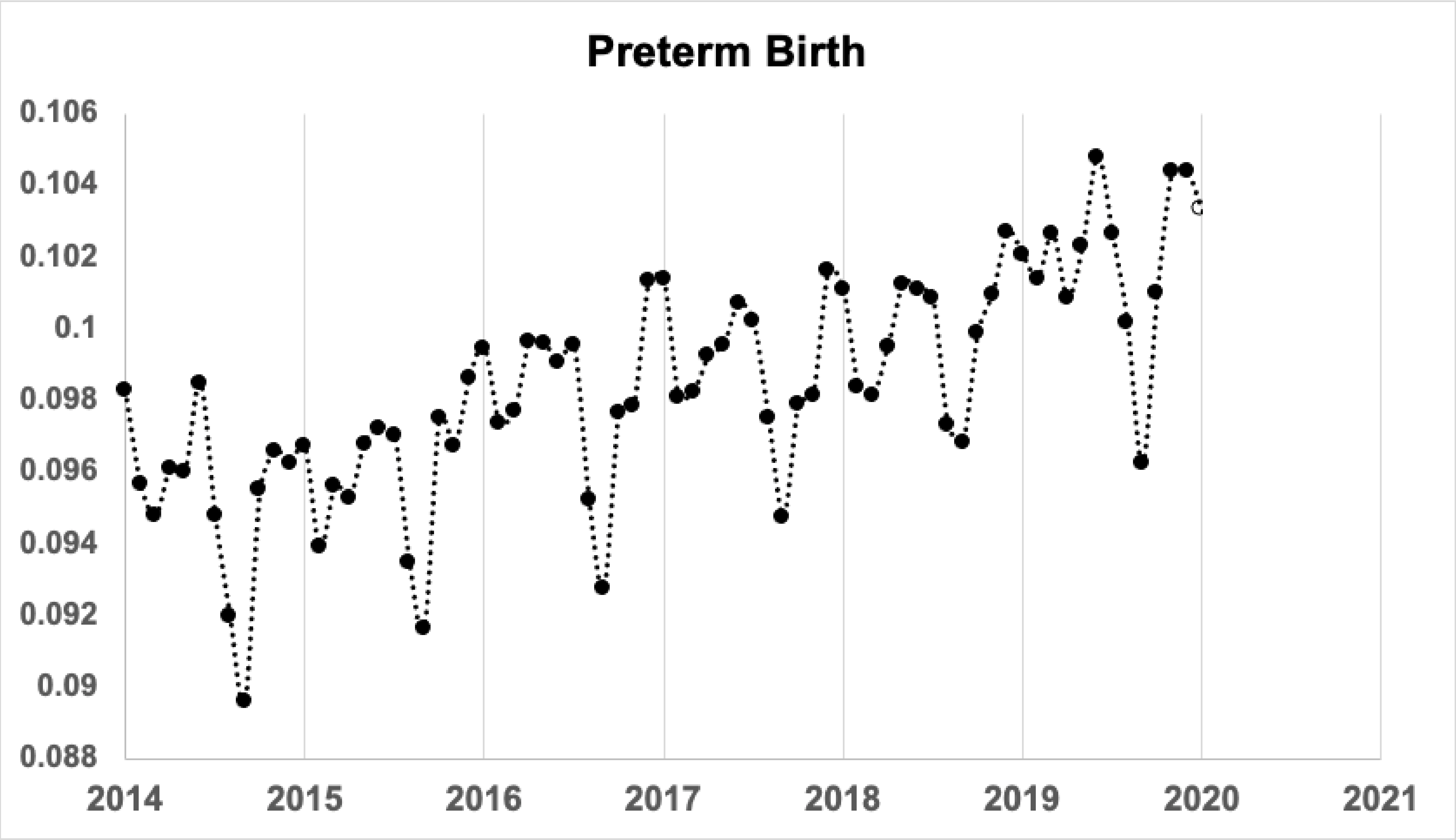
	Fourier	Splines
Pros	Uses few parameters; models long-term patterns smoothly	Can capture seasonal patterns in a way that is allowed to vary from one year to the next
Cons	The modeled seasonal pattern is always forced to be the same from one year to the next, which may not reflect the data well	Need to balance having too few or too many join-points (no consensus on optimal # of join-points); could miss short-term fluctuations or perturbations



Bhaskaran et al. 2013

4. ARIMA transfer functions (models using Box-Jenkins routines)

- Detect patterns in temporal data
  - Seasonality, trend, plateaus, spike and decay, etc.
- ARIMA
  - Autoregressive (AR): captures the tendency for high or low values to be remembered into the subsequent time periods
  - Integrated (I): characterizes non-stationarity (e.g., secular trend, strong seasonality)
  - Moving average (MA): similar to an AR term in that it captures “memory” of a high or low value but disappears much more quickly than an AR and is often characterized as an “echo.”


Preterm Birth
0.106
0.104
0.102
0.1
0.098
0.096
0.094
0.092
0.09
0.088
2014
2015
2016
2017
2018
2019
2020
2021

ARIMA models for outcomes

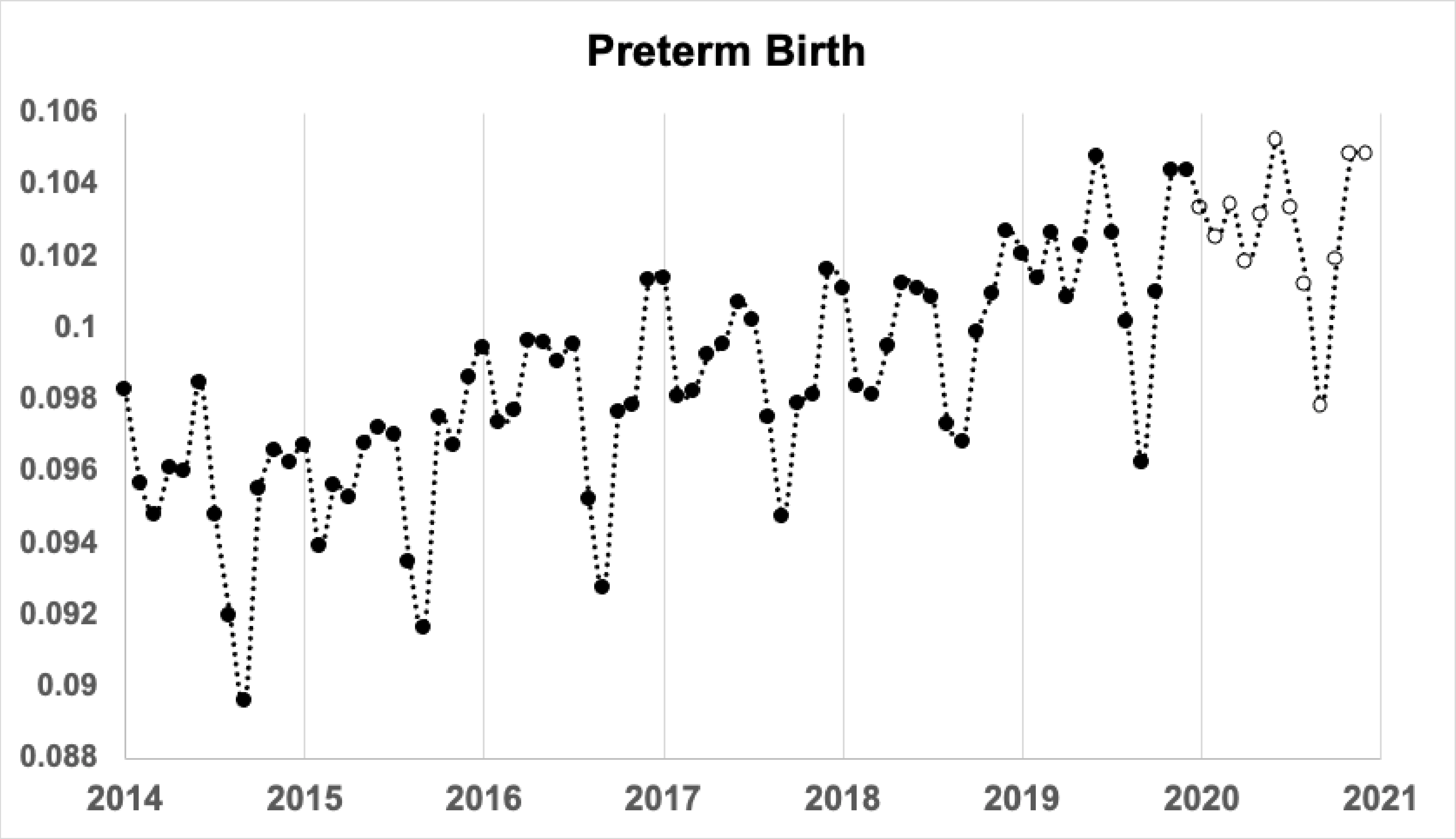
Outcome	Constant	AR	I	MA
Number of births				
Incidence of preterm birth				
Incidence of extreme preterm birth				
Incidence of cesarean delivery				
Sex ratio at birth				

ARIMA models for outcomes

Outcome	Constant	AR	I	MA
Number of births	none	AR(3)	I(12)	MA(2)
Incidence of preterm birth	yes	AR(1), AR(12)	none	none
Incidence of extreme preterm birth	yes	AR(12)	none	MA(9)
Incidence of cesarean delivery	yes	AR(3), AR(12)	none	MA(9)
Sex ratio at birth	yes	AR(12)	none	none

All outcomes, except sex ratio at birth, contained additional patterning beyond a seasonal/calendar month effect (i.e., AR at lag other than 12 or I term other than 12)

Gemmill et al. 2022


Preterm Birth
0.106
0.104
0.102
0.1
0.098
0.096
0.094
0.092
0.09
0.088
2014
2015
2016
2017
2018
2019
2020
2021

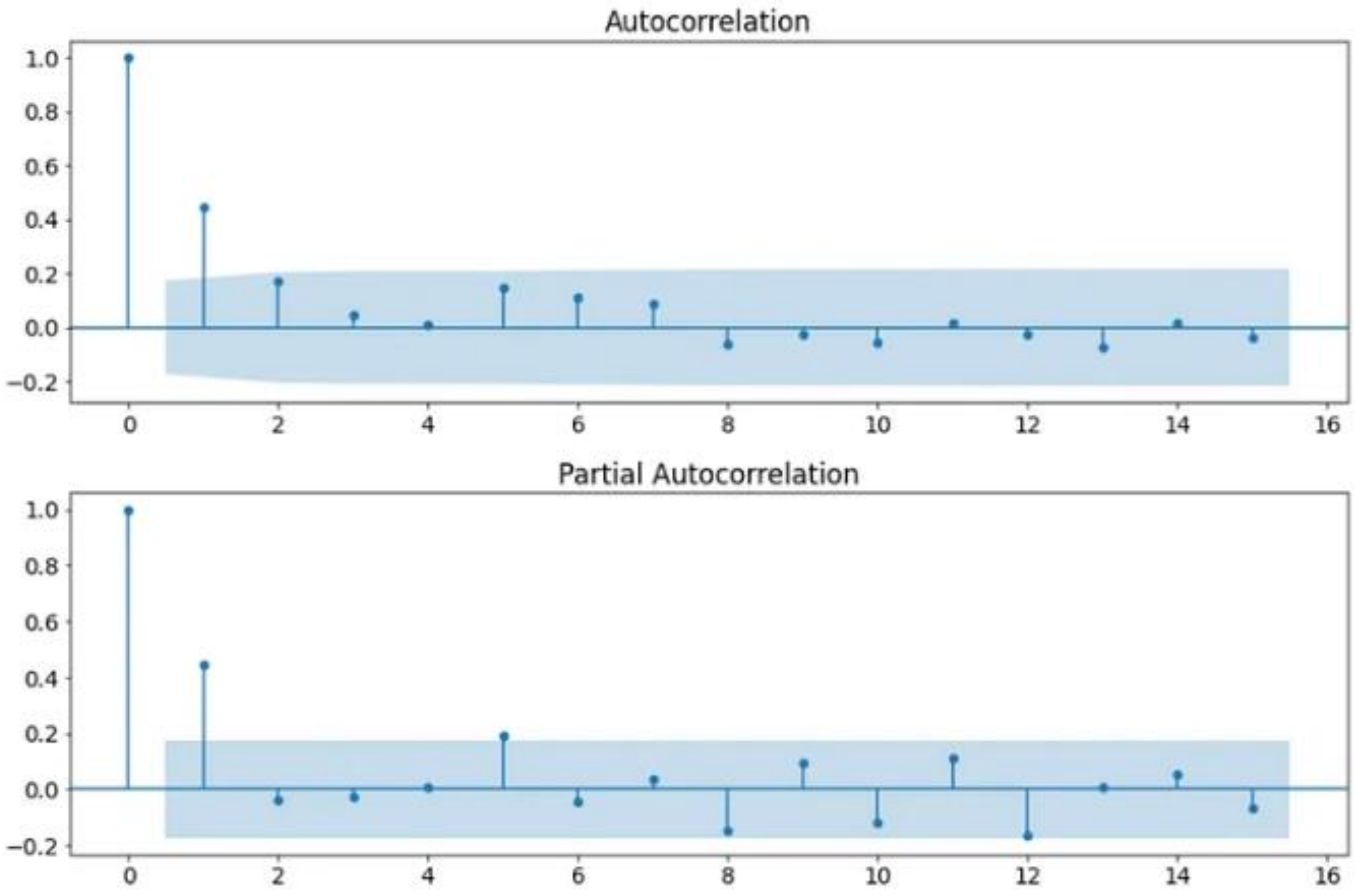
Considerations

Always check residuals for temporal autocorrelation

- Investigate Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

- ACF and PACF answer the following questions:
  - Is the observed time series white noise (random)?
  - Is an observation related to adjacent observations or other observations?

  - And for ARIMA approaches: Can time series be modeled with an AR, MA, or I term?



Example of an ACF and a PACF plot. (Image by the author via Kaggle)

AR(1) process:
Geometric decay in ACF plot
In PACF, cuts off at lag 1

Lots of decision rules and testing

Auto.arima() functions in R and Stata (algorithm based)

Number of observations

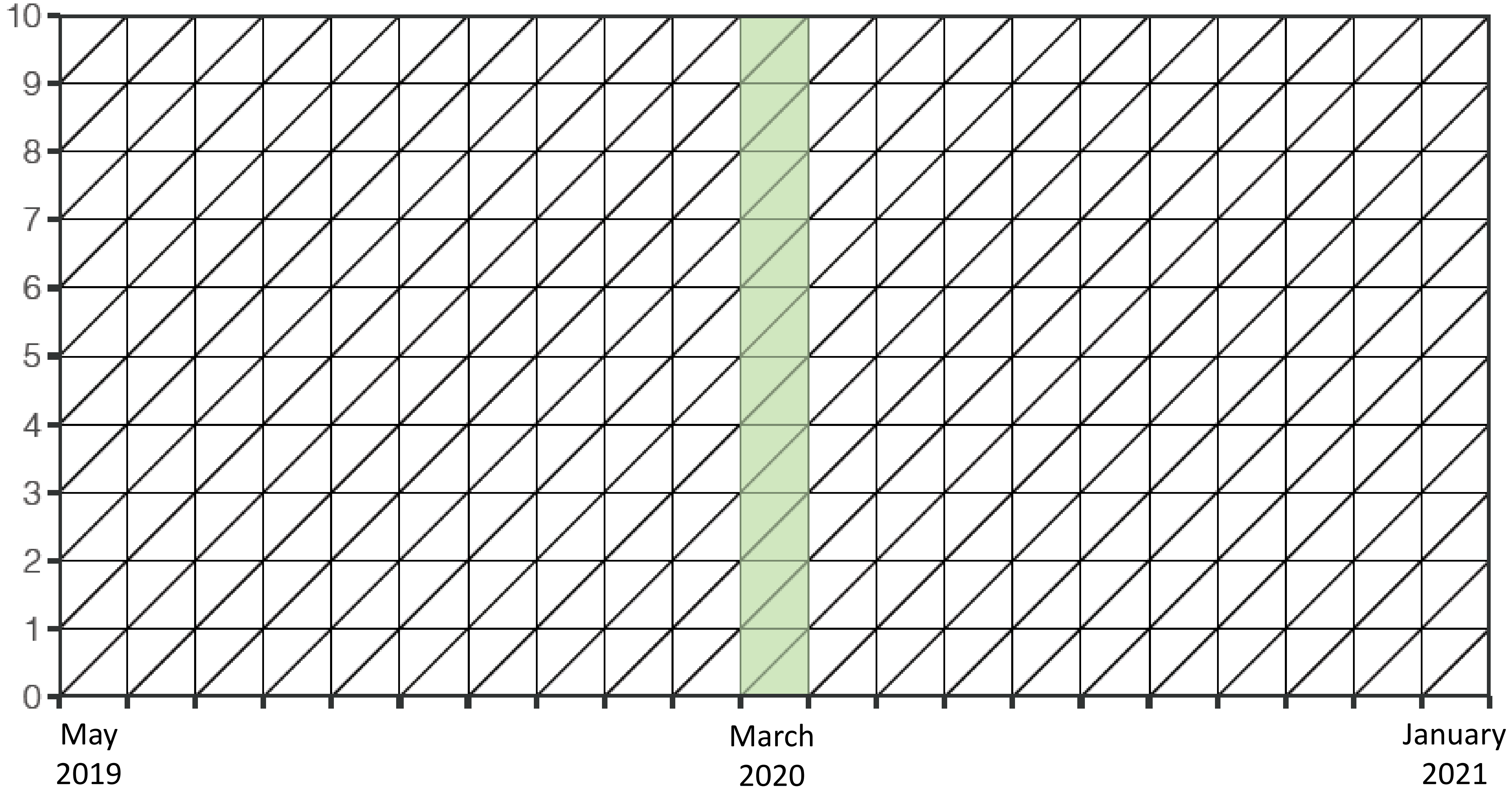
- Depends on method
- For ARIMA, at least 50 pre-intervention time points needed as a rule of thumb

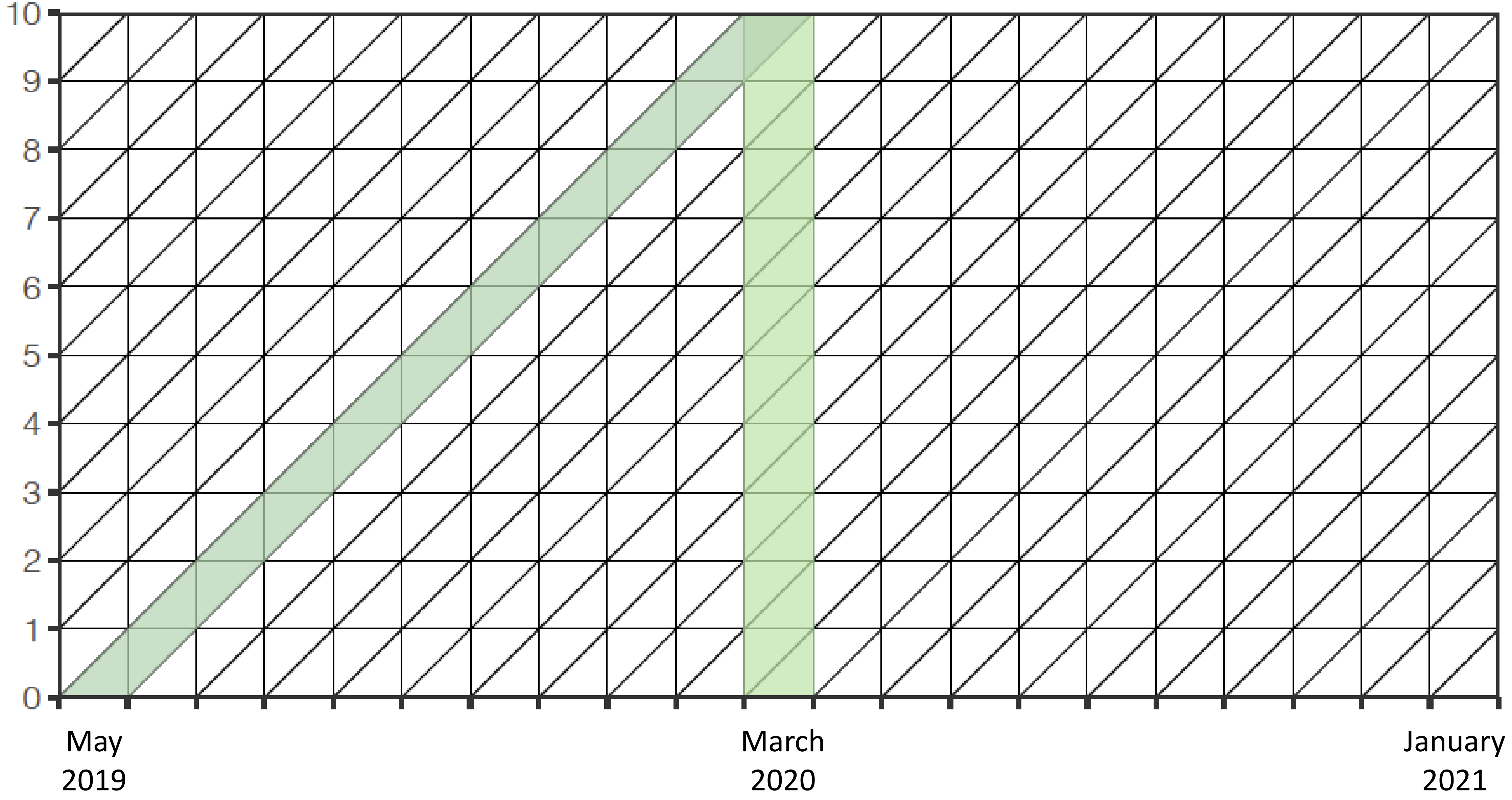
- Number of observations that contribute to each month also matters, because fewer little n observations means more random variation gets introduced
  - But, in our study of the Muslim ban, we detected an association with very small numbers in the population (an average of only 1600 births per month)

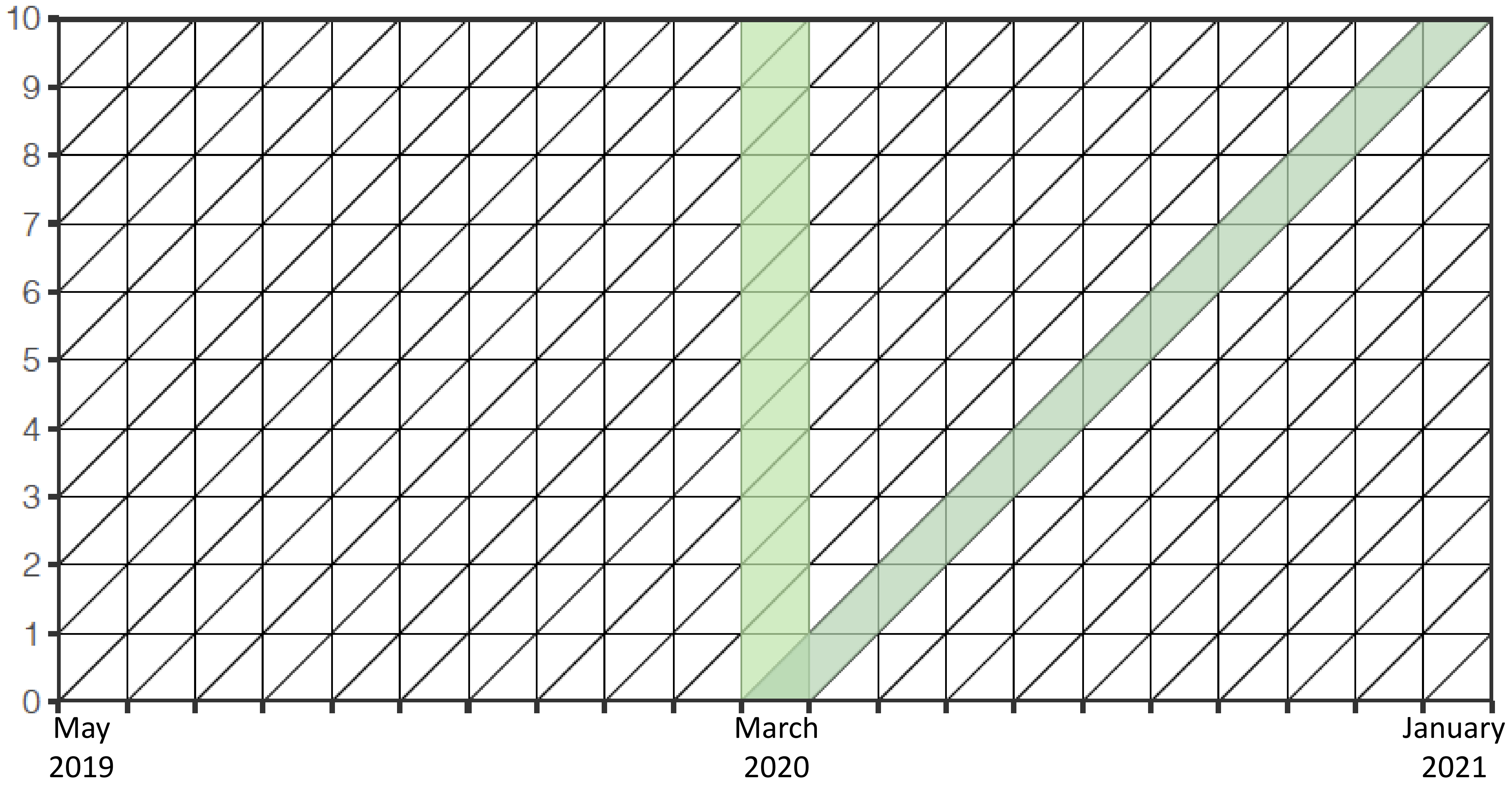
Temporal relationship between onset and outcomes

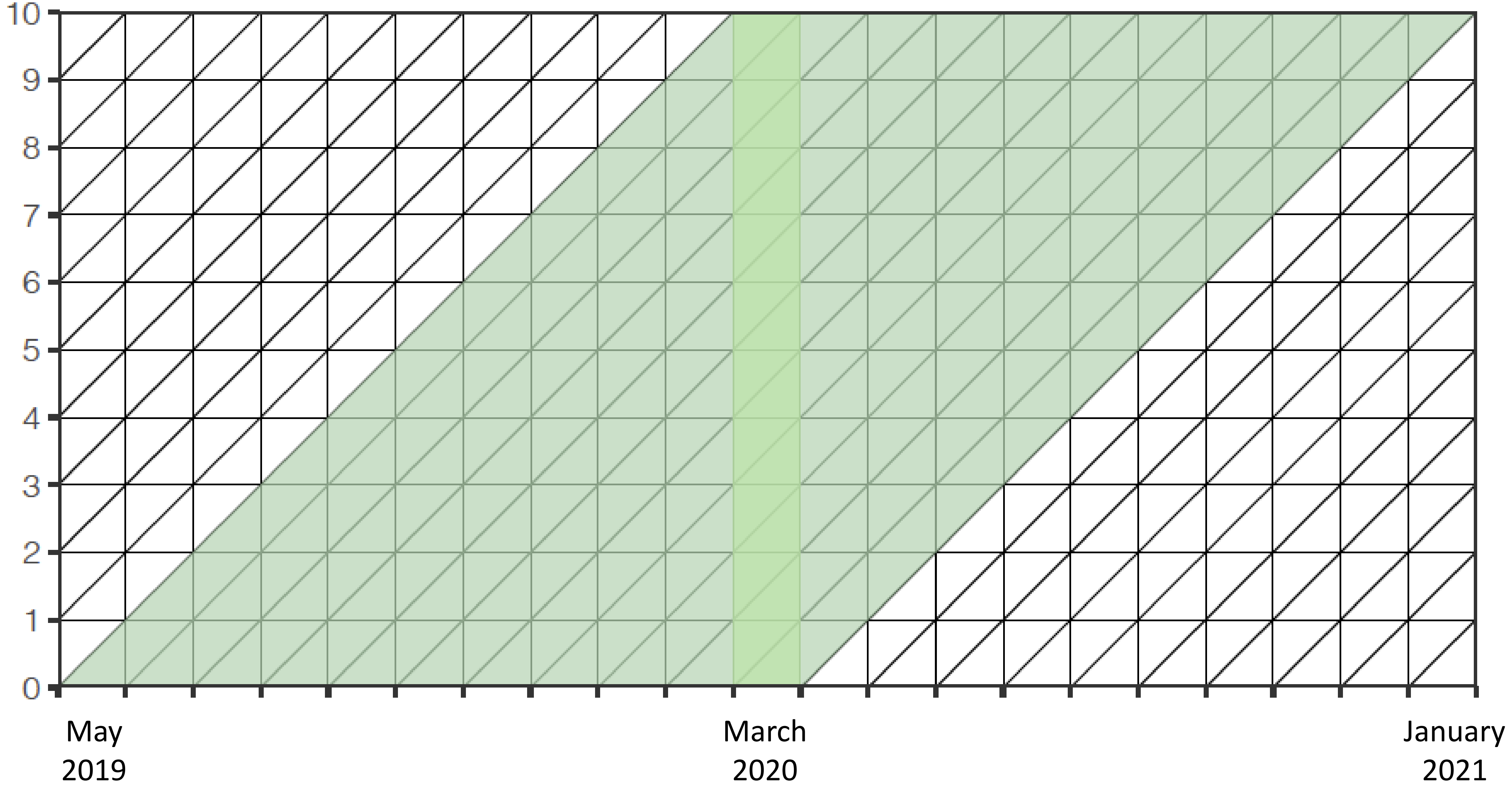
- Birth cohorts comprise members from different conception cohorts

- Lexis diagrams are our friends!


10
9
8
7
6
5
4
3
2
1
0
May 2019
March 2020
January 2021


10
9
8
7
6
5
4
3
2
1
0
May 2019
March 2020
January 2021


10
9
8
7
6
5
4
3
2
1
0
May 2019
March 2020
January 2021



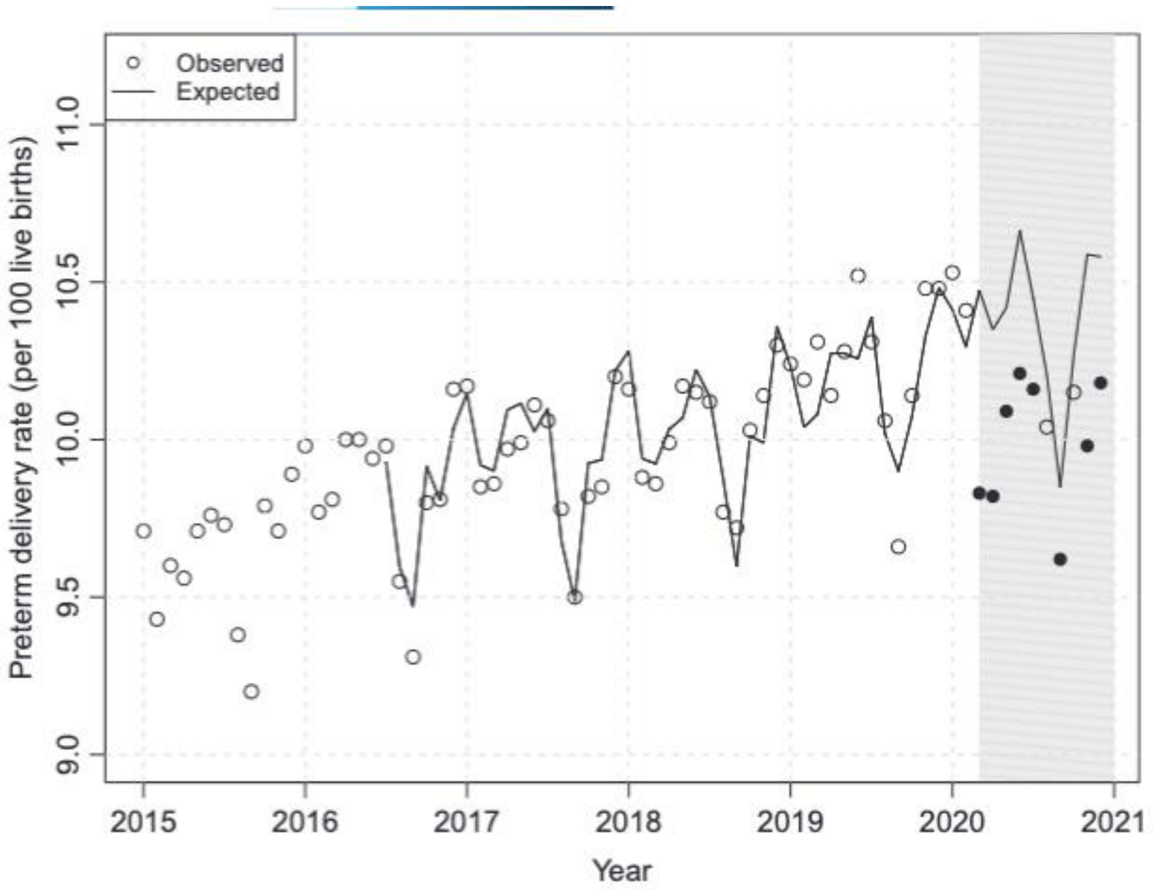
10
9
8
7
6
5
4
3
2
1
0
May 2019
March 2020
January 2021

Other considerations

- Defining policy/event onset (e.g., passage of a law vs. implementation)
- How to measure counterfactuals that incorporate several years of pandemic changes??

Covid and Preterm Birth

Unexpected reduction in preterm birth in 2020



Changes in preterm birth and caesarean deliveries in the United States during the SARS-CoV-2 pandemic

Alison Gemmill[1] | Joan A. Casey[2] | Ralph Catalano[3] | Deborah Karasek[4,5] | Claire E. Margerison[6] | Tim Bruckner[7]

Exposure to the early COVID-19 pandemic and early, moderate and overall preterm births in the United States: A conception cohort approach

Claire E. Margerison[1] | Tim A. Bruckner[2] | Colleen MacCallum-Bridges[1] | Ralph Catalano[3] | Joan A. Casey[4] | Alison Gemmill[5]

We detected a 5-6% reduction in preterm births.

Exposure to the early COVID-19 pandemic may have promoted longer gestation among close-to-term pregnancies and/or reduced risk of later preterm delivery among gestations exposed in the first trimester.

Covid and Fertility

Work with Jenna Nobles (lead), Florencia Torche, and Sungsik Hwang

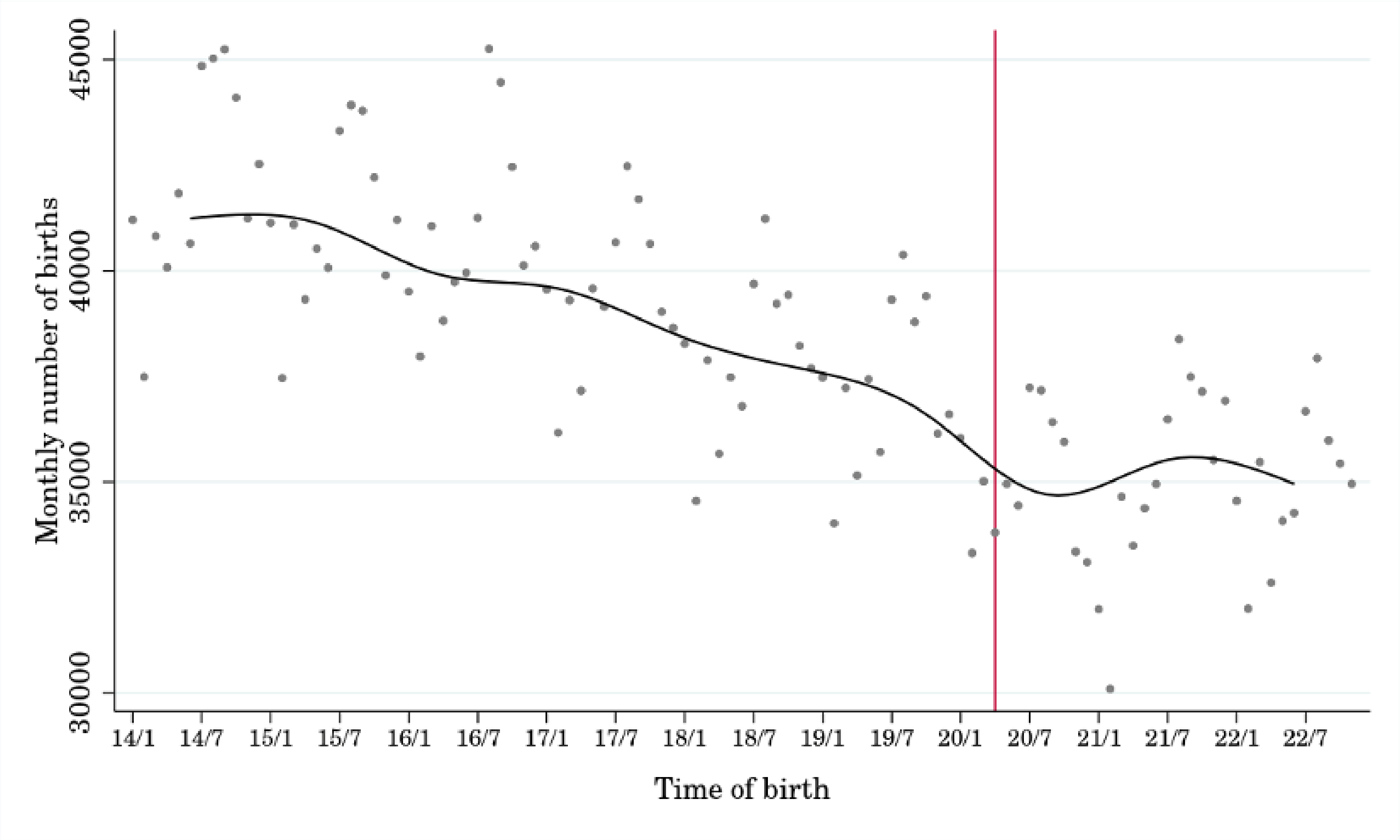
Data

- California birth certificate data from January 2014 through November 2022
- 12% of births in the US
- 39% Hispanic, 35% white, 15% Asian or Pacific Islander, 5% Black, 4% multiracial, <1% Native American/Alaskan Native)

Methods

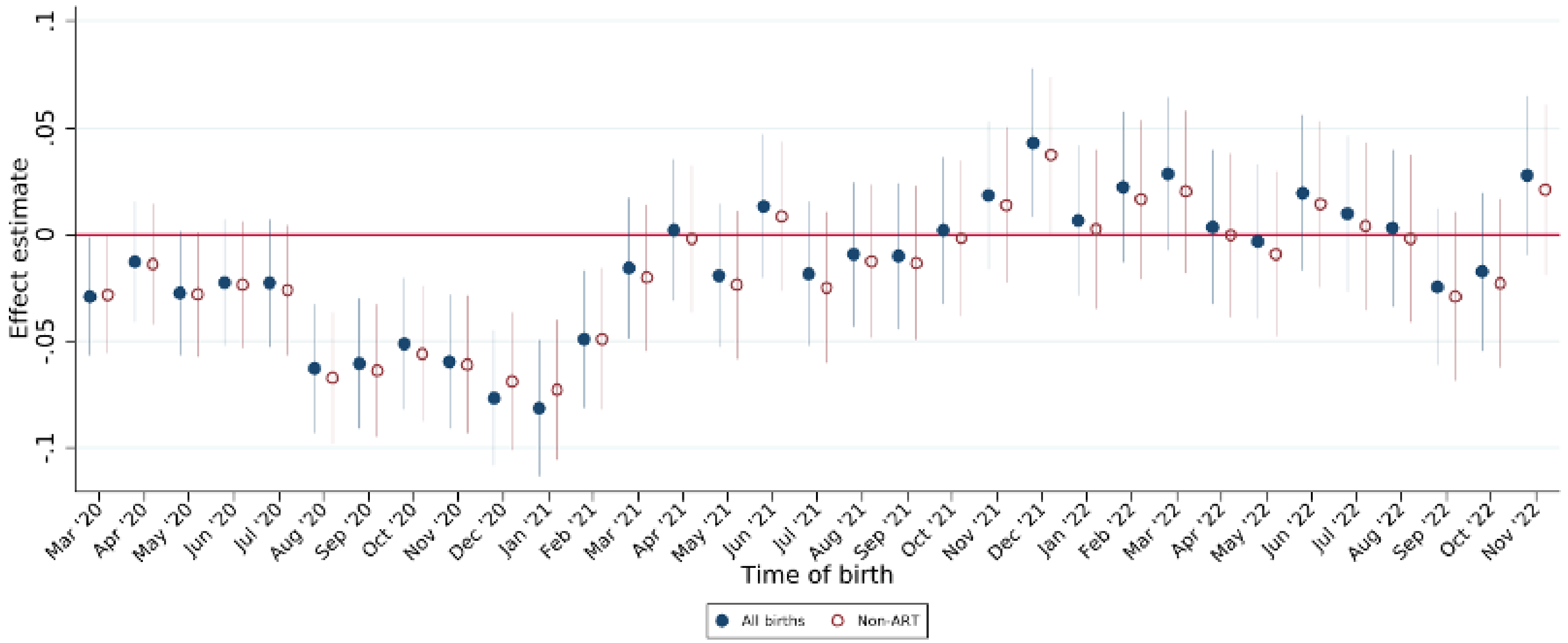
- Interrupted Time Series models w/ monthly birth data
- ARIMA/Box-Jenkins models used to estimate the counterfactual using data from Jan 2014 – Feb 2020
- "Effects" are estimated by examining the deviation between expected (from counterfactual) and observed in the COVID period
- Examine heterogeneity by maternal age, parity, race/ethnicity, nativity, education (stratified time-series models)

Births in California by month, 2014-November 2022



Source: California Department of Public Health. Red line indicates the onset of the pandemic in March 2020.

Proportional deviation in California birth counts by month: all births, and births without Assisted Reproductive Technology, March 2020 – Nov 2022



What if you have comparison groups?

- Comparative interrupted time series
- Bayesian structural time series (Causal Impact package in R)
- Synthetic control method (AugSynth package in R)

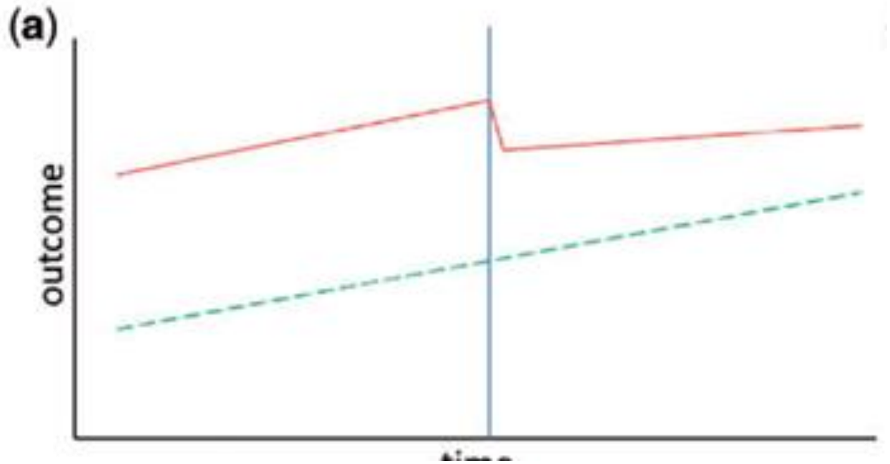
- Important that control is theorized to be unaffected!

Comparative interrupted time series

The use of controls in interrupted time series studies of public health interventions

James Lopez Bernal,[1]* Steven Cummins[1] and Antonio Gasparrini[1,2]

- Use a comparison population that is theorized to be unaffected by the event
- Similar to difference-in-difference designs
- Intervention and control groups may exhibit similar autocorrelation, but need to account for any further autocorrelation in intervention group that remains after adjustment for controls

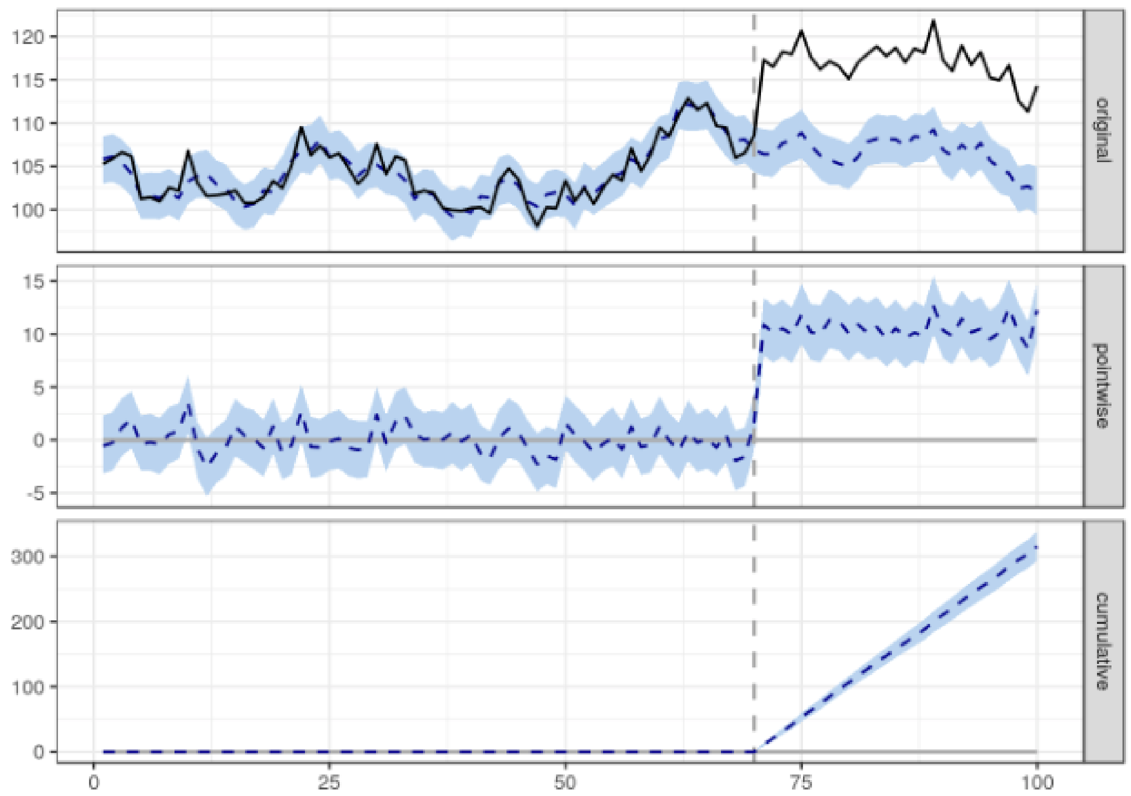


Bayesian Structural Time Series

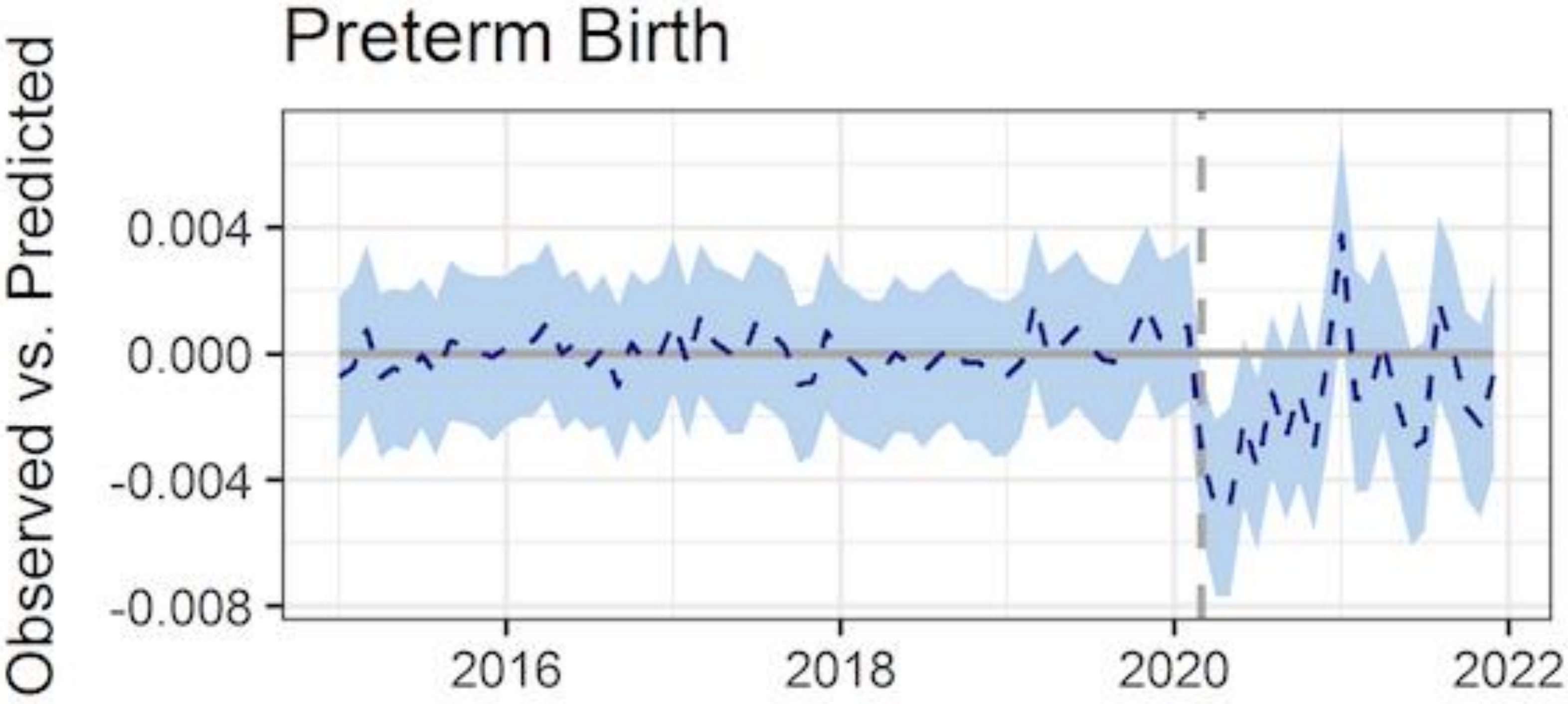
- See CausalImpact Package in R for a useful tutorial

```
impact <- CausalImpact(data, pre.period, post.period)
```

```
plot(impact)
```

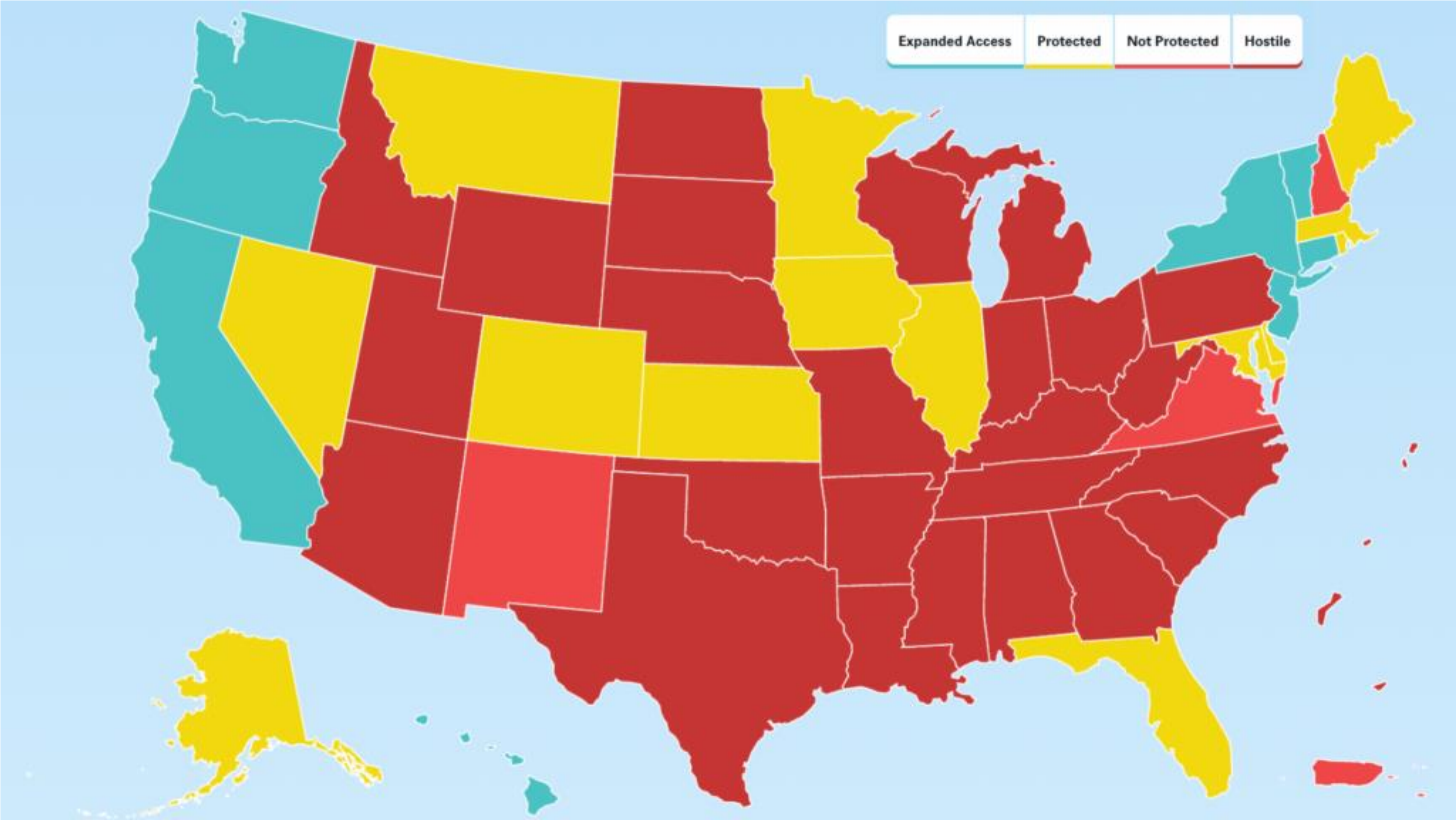


CausalImpact results for preterm birth in US



Synthetic Control methods

- Good to use when you have one intervention site with multiple comparison sites

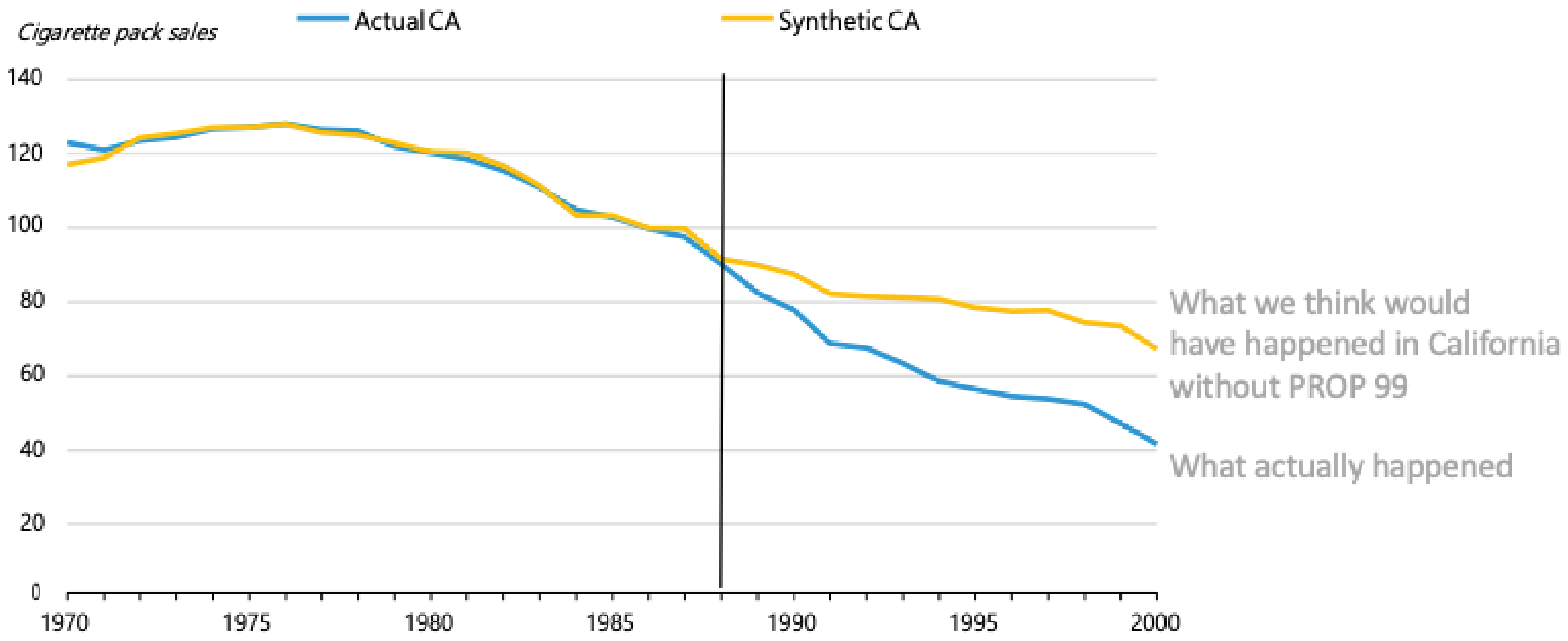


Synthetic control methods

- 1) Build a “synthetic comparison group” by taking a weighted average of other similar “donor” units so that our synthetic comparison group is “as much like” our “treatment” unit as possible before the policy change or event occurred. and then
- 2) Use the observed outcome trajectory of our synthetic comparison group to represent the counterfactual outcome trajectory for our treatment unit.

- See slides from Elizabeth Stuart: https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/ViewPresentation.cfm?file=306613.pdf

Figure 1: The Seminal Application: Annual Cigarette Sales for California and its Synthetic Comparison Group Before and After PROP 99



Graph from Robert McClelland and Sarah Gault, 2017, *The Synthetic Control Method as a Tool to Understand State Policy*, Washington, DC: Urban Institute.

Conclusions

- Lots of different ways to model counterfactuals using time series data!
- If using monthly or weekly data, make sure you check you always check residuals for temporal autocorrelation
  - Using ACF and PACF
- If autocorrelation is still present, iterate and/or adapt model

- Email: agemmill@jhu.edu

Another Challenge: Terminology!

- Difference in differences
- Event study methods
- Synthetic control
- Augmented synthetic control
- Marginal structural model
- Interrupted time series
- Comparative interrupted time series
- Two-way fixed effects
- Panel data methods
- ....

Are these even all different things???

See slides from Elizabeth Stuart: https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/ViewPresentation.cfm?file=306613.pdf