

*American Journal of Epidemiology* Submitted Manuscript

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

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

**Correspondence Address:** Correspondence to Alison Gemmill, Department of Population, Family and Reproductive Health, Johns Hopkins School of Public Health, 615 N. Wolfe Street, Room E4148, Baltimore, MD 21205 (e-mail: [agemmill@jhu.edu](mailto:agemmill@jhu.edu))

ORIGINAL UNEDITED MANUSCRIPT

© The Author(s) 2022. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

**Affiliations:** Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States (Alison Gemmill); Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York, United States (Joan A. Casey); Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, Michigan, United States (Claire E. Margerison); Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Center of Research in Epidemiology and Statistics Sorbonne Paris Cité, Université de Paris and Institut National de la Santé et de la Recherche Médicale, Institut National de la Recherche Agronomique, Paris, France (Jennifer Zeitlin); School of Public Health, University of California, Berkeley, Berkeley, California, United States (Ralph Catalano); Program in Public Health, University of California, Irvine, Irvine, California, United States (Tim A. Bruckner)

**Funding:** This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (1R01HD103736-01A1).

**Data Availability Statement:** All data are available from publicly available sources.

**Thanks:** N/A

**Conference presentation:** N/A

**Preprint Information:** N/A

**Disclaimer:** The views expressed in this article are those of the authors and do not reflect those of the National Institutes of Health.

**Conflict of Interest:** The authors report no conflicts of interest.

**Running Head:** Perinatal Covid Counterfactuals

**Key words:** Covid-19 pandemic; perinatal health; counterfactuals

**Abbreviations:** ITS, interrupted time series; ARIMA, autoregressive integrated moving average; AR, autoregressive term; I, integrated term; MA, moving average term; ePTB, extreme preterm birth; Coef, coefficient; SE, standard error.

## **Abstract**

The epidemiologic literature estimating the indirect or secondary effects of the COVID-19 pandemic on pregnant people and gestation continues to grow. Our assessment of this scholarship, however, leads us to suspect that the methods most commonly used may lead researchers to spurious inferences. This suspicion arises because the methods do not account for temporal patterning in perinatal outcomes when deriving counterfactuals, or estimates of the outcomes had the pandemic not occurred. We illustrate the problem in two ways. First, using monthly data from US birth certificates, we describe temporal patterning in five commonly used perinatal outcomes. Notably, for all but one outcome, temporal patterns appear more complex than much of the emerging literature assumes. Second, using data from France, we show that using counterfactuals that ignore this complexity produces spurious results. We recommend that subsequent investigations on COVID-19 and other perturbations use widely available time-series methods to derive counterfactuals that account for strong temporal patterning in perinatal outcomes.

The epidemiologic literature includes several reports counterintuitive declines in preterm birth during the early months of the COVID-19 pandemic (1–4). These short-term improvements in perinatal and maternal health offer an opportunity to better understand preterm birth and identify new prevention strategies. Sorting out how changes in clinical care, disruption of social ties (5), widened economic inequality (6), and changes in environmental, occupational, and infectious exposures (7) affected preterm birth would seem an important, albeit difficult, research program for epidemiology. We believe that such a research program should start with careful assessment of the current literature reporting changes in perinatal outcomes early in the pandemic. Based on our assessment of that scholarship (8, 9), we are concerned that the study designs most commonly used have a significant limitation. The limitation stems from what we view as a suboptimal choice of counterfactuals, or outcomes that would be expected had the COVID-19 pandemic not occurred. We elaborate on this study design challenge for estimating COVID-19 “effects” especially on perinatal outcomes, illustrate how use of naïve counterfactuals produces spurious results, and recommend alternative methods that, while not novel, appear largely unused by perinatal epidemiologists.

### **The Challenge**

Describing the pandemic’s effect on maternal and perinatal outcomes requires “counterfactuals” or estimates of the outcomes had the pandemic not occurred. Investigators often assume that the mean of an outcome’s pre-pandemic values is its expected value and, therefore, serves as its intra-pandemic counterfactual. However, this assumption does not apply if the outcome in pre-pandemic cohorts exhibits patterns over time. Such patterns, or

“autocorrelation,” imply that the expected value of measurements is not their mean, but instead an extrapolation from the history.

Autocorrelation in birth rates can arise from exogenous forces such as culture or climate that induce patterns (e.g., seasonality) in behavior and biology or by factors that are endogenous to the system. Autocorrelation can confound tests like ours because “programmed” high or low rates in the perinatal outcome can coincide with, for example, the onset of an epidemic.

Figure 1 plots the monthly rate of preterm birth in the US from January 2014 to December 2019 (represented by closed circles). Visual inspection of the time series suggests several key features. First, the overall series appears to be trending upward, which aligns with reports that preterm birth rates have risen 7% since 2014 (10). Second the series appears to be highly seasonal, meaning that peaks and troughs occur roughly at the same time each year (11–13). In addition to seasonality, there may be other types of autocorrelation not easily detected by visual inspection. These may include the tendency to remain elevated or depressed, or to oscillate, after high or low values.

### **The Temptation**

A naïve approach, used in many papers assessing the impact of the pandemic on perinatal outcomes and summarized in two meta-analyses (8, 9), ignores autocorrelation and makes questionable counterfactual approximations. This work often focuses on a clinically-defined population and compares outcomes observed immediately prior to the pandemic to those observed after (14–17). Such an approach does not consider well-documented patterns of seasonality and trend in perinatal outcomes described above (11–13).

Much more common in the pandemic-perinatal literature is the use of what we call the “stacked calendar” approach (2–4, 18–21). This approach adjusts for patterns in the data by comparing the outcomes of a given period to the outcomes of the same period but in a prior year or years (e.g., comparing March of 2020 to Marches of 2015-2019). For example, an analysis of preterm birth and stay-at-home orders in Tennessee compared births occurring between March 22 to April 30, 2020 to births occurring between the same time period but aggregated over 2015-2019 (18). A similar approach using data from California compared preterm birth rates occurring between April and July 2020 to the same period four years prior (e.g., 2016-2019) (19). These approaches adjust for seasonality by capturing month-of-year effects and often control for individual-level sociodemographic factors (e.g., race/ethnicity, maternal age) that are believed to induce seasonality in perinatal outcomes. The “stacked calendar” approach, however, does not account for other types of autocorrelation, including upward or downward longer-term trends and the tendency for high or low values in one month to persist into subsequent months. This omission is particularly important for an outcome like preterm birth, which as described above, has been increasing in the US since 2014.

### **The Alternative**

One way to devise counterfactuals that account for autocorrelation is to use time series methods widely applied in engineering and in the natural and social sciences to systematically detect and mathematically model temporal patterning (22, 23). Patterns detected by these methods include seasonality and other cycles as well as linear trends. Other detected shapes include “plateaus” or the tendency for a high or low value to persist for several or more cohorts and to then drop abruptly back to previous levels, or “spike and decay” patterns, which describe

a high or low value followed by similarly high or low, but declining, values (24, 25). After such estimation of a counterfactual value, the researcher can implement an Interrupted Time Series (ITS) study design in which the COVID-19 pandemic serves as the interruption of interest. ITS methods are well-described in the epidemiologic literature (26–28).

One of the many approaches to accounting for autocorrelation in time series data is to use an ARIMA model. These models, developed by Box and Jenkins, can have three components: the Autoregressive, or AR term, which captures the tendency for high or low values to be remembered into the subsequent time periods; the Integrated, or I term, which characterizes non-stationarity (e.g., secular trend, strong seasonality); and the Moving Average, or MA term, which is 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.” Use of the AR, I, and MA components to parsimoniously describe autocorrelation in the perinatal outcome removes the threat of confounding due to predictably “scheduled” patterns.

Using monthly data from US birth certificates from January 2014 to December 2019, we evaluated five perinatal outcomes of interest to assess presence of autocorrelation. These outcomes include monthly birth counts, rate of preterm birth (<37 weeks gestational age), rate of extreme preterm birth (<28 weeks gestational age), rate of cesarean delivery, and the sex ratio at birth. Box-Jenkins methods detected autocorrelation in all five series (Table 1). We detected strong seasonality in all five series, and all series except for sex ratio at birth contained additional patterning other than a seasonal component (e.g., other than an AR or I at lag 12 months). Note that our calculation of rates for this exercise uses births as the denominator, rather than conceptions, which yields time series that are sensitive to both seasonal changes in conceptions

as well as changes in the risk among conception cohorts. Alternatively, one could estimate autocorrelation in conception cohort patterns as well.

Using these patterns, the researcher can derive statistically expected, or “counterfactual”, values for the months during the COVID-19 pandemic. The open circles of Figure 1 show expected monthly preterm birth rates for all 12 months of 2020 based on monthly preterm birth rates from January 2014 to December 2019. The researcher could then determine whether the difference between the observed and expected values for the hypothesized months during the COVID-19 pandemic differ detectably from 0. Whereas we use software from Scientific Computing Associates Corp (River Forest, IL), the ARIMA routines are also available in other standard packages (e.g., R [Vienna, Austria], SAS [Cary, NC], Stata [College Station, TX]).

### **Real-world example of spurious results: COVID-19 lockdown and extreme preterm birth in France**

We provide one example which compares the stacked calendar approach to a time-series approach. We examine the continuous outcome of the weekly rate of extreme preterm birth (ePTB; <28 weeks gestational age) in France, a country which imposed a strict nationwide lockdown from the 17<sup>th</sup> of March and lifted on the 10<sup>th</sup> of May, in response to the COVID-19 pandemic (29). We focus on ePTB because early reports from Denmark and Ireland found that reductions in preterm birth were concentrated among this population (2, 3).

Using the stacked calendar approach, we compared the rate of ePTB during the lockdown period in 2020 to the same eight weeks in the previous year—that is, Spring 2019. Results indicate a reduced odds of ePTB during the lockdown period (adjusted Odds Ratio [OR]= 0.84; 95% CI: 0.71 - 0.99). By contrast, when we employ a time-series approach, we identified more



nuanced patterns in ePTB before the lockdown period. After controlling for these patterns, the lockdown results differ substantially from the stacked calendar approach in that we do not reject the null for ePTB (OR = 0.98; 0.82 - 1.17).

Additional exploration provides some insight into the discrepant findings. One large positive outlier in the week beginning April 9, 2019, which lies outside the 99% confidence interval for the weekly ePTB rate over the entire series (30), drives the divergence between the stacked calendar counterfactual and the observed pandemic values. By contrast, the time-series approach does not heavily weight this Spring 2019 outlier when arriving at forecasts of the counterfactuals for the pandemic period. We attribute this Spring 2019 outlier, as well as the inability of the stacked calendar approach to account for more nuanced patterning in ePTB, as inducing a spurious inference that ePTB fell during the first lockdown period in France.

## **Conclusion**

Perinatal outcomes show strong patterning over time. Evaluating an “interruption” like the COVID-19 pandemic requires generating counterfactuals that take these patterns into account. While the literature includes prior applications of ITS for public health research (e.g., Bernal et al. 2017) (26), nearly all the perinatal epidemiology literature on the secondary effects of COVID-19 (i.e., >90% of articles in a recent meta-analysis) fails to fully control for time-dependent autocorrelation and/or does not consider the possibility (7). The temptation to ignore such patterns may lead future researchers to use suboptimal study designs that yield spurious results. This circumstance would not move the field forward in that it would not be reproducible with the use of more appropriate study designs and could result in policies or clinical guidance that would be based on erroneous inferences.

Although the simple execution and interpretation of “stacked calendar” methods appeals to researchers, there are two reasons why derivation of counterfactuals using ITS seems more appropriate. First, ITS uses the full information of the dataset in providing a more stable reference group and minimizes random error if, as in the stacked calendar example from France, one outlier occurs during the time (i.e., weeks in March/April 2019) used as a “narrow” reference. Second, accounting for autocorrelation corrects for “confounding” by calendar time in the presence of temporal patterns—whether they arise due to biologic causes, changes in surveillance, or other non-stochastic reasons. Whereas ITS practitioners have considerable flexibility in their approach for deriving counterfactuals in the presence of autocorrelation, there are several requirements. For example, ITS studies that employ Box-Jenkins methods provide adequate study power with at least 50 evenly spaced time units before the interruption and require consistent data collection over these units (31). As it relates to the COVID-19 pandemic, monthly birth data starting in January 2015 (i.e., 62 monthly values from January 2015 to February 2020) would meet this requirement.

Identification of COVID-19 effects on population health outcomes has its own unique challenges (e.g., disentangling what is meant by “exposure” to the pandemic; see Dmitris and Platt 2021) (32) which remain salient regardless of study design used—ITS or otherwise. Nevertheless, we hope this discussion helps researchers address a common challenge related to the derivation of counterfactuals in autocorrelated time series data that are especially common in perinatal outcomes research.

## References

1. Gemmill A, Casey JA, Catalano R, et al. Changes in preterm birth and cesarean deliveries in the United States during the SARS-CoV-2 pandemic. *Paediatr Perinat Epidemiol.* 2021; 10.1111/ppe12811.
2. Hedermann G, Hedley PL, Bækvad-Hansen M, et al. Danish premature birth rates during the COVID-19 lockdown. *Arch Dis Child Fetal Neonatal Ed.* 2021;106(1):93–95.
3. Philip RK, Purtill H, Reidy E. Unprecedented reduction in births of very low birthweight (VLBW) and extremely low birthweight (ELBW) infants during the COVID-19 lockdown in Ireland: a ‘natural experiment’ allowing analysis of data from the prior two decades. *BMJ Glob Health.* 2020;5(9):e003075.
4. Khalil A, von Dadelszen P, Draycott T, et al. Change in the incidence of stillbirth and preterm delivery during the COVID-19 pandemic. *JAMA.* 2020;324(7):705–706.
5. Davenport MH, Meyer S, Meah VL, et al. Moms are not OK: COVID-19 and maternal mental health. *Front Glob Womens Health.* 2020;1:1.
6. Perry BL, Aronson B, Pescosolido BA. Pandemic precarity: COVID-19 is exposing and exacerbating inequalities in the American heartland. *Proc Natl Acad Sci U S A.* 2021;118(8): e2020685118.
7. Kotlar B, Gerson E, Petrillo S, et al. The impact of the COVID-19 pandemic on maternal and perinatal health: a scoping review. *Reprod Health.* 2021;18:10.
8. Chmielewska B, Barratt I, Townsend R, et al. Effects of the COVID-19 pandemic on maternal and perinatal outcomes: a systematic review and meta-analysis. *Lancet Glob Health.* 2021;9(6):e759–e772.

9. Yang J, D'Souza R, Kharrat A, et al. COVID-19 pandemic and population-level pregnancy and neonatal outcomes: a living systematic review and meta-analysis. *Acta Obstet Gynecol Scand*. 2021; in press.
10. Hamilton BE, Martin JA, Osterman MJK. Births: Provisional data for 2019. Vital Statistics Rapid Release; no 8. Hyattsville, MD: National Center for Health Statistics. May 2020.
11. Bodnar LM, Simhan HN. The prevalence of preterm birth and season of conception. *Paediatr Perinat Epidemiol*. 2008;22(6):538–545.
12. Darrow LA, Strickland MJ, Klein M, et al. Seasonality of birth and implications for temporal studies of preterm birth. *Epidemiology*. 2009;20(5):699–706.
13. Currie J, Schwandt H. Within-mother analysis of seasonal patterns in health at birth. *Proc Natl Acad Sci U S A*. 2013;110(30):12265–12270.
14. Greene NH, Kilpatrick SJ, Wong MS, et al. Impact of labor and delivery unit policy modifications on maternal and neonatal outcomes during the coronavirus disease 2019 pandemic. *Am J Obstet Gynecol MFM*. 2020;2(4):100234.
15. Kumari V, Mehta K, Choudhary R. COVID-19 outbreak and decreased hospitalisation of pregnant women in labour. *Lancet Glob Health*. 2020;8(9):e1116–e1117.
16. Kc A, Gurung R, Kinney MV, et al. Effect of the COVID-19 pandemic response on intrapartum care, stillbirth, and neonatal mortality outcomes in Nepal: a prospective observational study. *Lancet Glob Health*. 2020;8(10):e1273–e1281.
17. Lemon L, Edwards RP, Simhan HN. What is driving the decreased incidence of preterm birth during the coronavirus disease 2019 pandemic? *Am J Obstet Gynecol MFM*. 2021;3(3):100330.

18. Harvey EM, McNeer E, McDonald MF, et al. Association of Preterm Birth Rate With COVID-19 Statewide Stay-at-Home Orders in Tennessee. *JAMA Pediatr.* 2021;175(6):635–637.
19. Main EK, Chang SC, Carpenter A, et al. Singleton preterm birth rates for racial and ethnic groups during the coronavirus disease 2019 pandemic in California. *Am J Obstet Gynecol.* 2021;224(2):239–241.
20. Handley SC, Mullin AM, Elovitz MA, et al. Changes in preterm birth phenotypes and stillbirth at 2 Philadelphia hospitals during the SARS-CoV-2 pandemic, March-June 2020. *JAMA.* 325(1):87–89.
21. Janevic T, Glazer KB, Vieira L, et al. Racial/ethnic disparities in very preterm birth and preterm birth before and during the COVID-19 pandemic. *JAMA Netw Open.* 2021;4(3):e211816.
22. Box GEP, Jenkins GM, Reinsel GC, et al. *Time Series Analysis: Forecasting and Control.* 5th ed. Hoboken, NJ: Wiley; 2015.
23. McCleary R, Hay RA, Meindinger EE, et al. *Applied Time Series Analysis for the Social Sciences.* Beverly Hills, CA: Sage; 1980.
24. Chatfield C. *Time-series forecasting.* Boca Raton, FL: CRC Press; 2001.
25. Liu LM, Hudak GB, Box GEP, Muller ME, Tiao GC. Forecasting and Time Series Analysis Using the SCA Statistical System, Volume 1. Scientific Computing Associates Corporation, Chicago; 1994.
26. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol.* 2017;46(1):348–355.

27. Helfenstein U. The use of transfer function models, intervention analysis and related time series methods in epidemiology. *Int J Epidemiol.* 1991;20(3):808–815.
28. Turner SL, Karahalios A, Forbes AB, et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. *J Clin Epidemiol.* 2020;122:1–11.
29. Pullano G, Valdano E, Scarpa N, et al. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit Health.* 2020;2(12):e638–e649.
30. Chen C, Liu LM. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association.* 1993;88(421):284-297.
31. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. *J Clin Epidemiol.* 2011; 64(11):1252–1261.
32. Dimitris MC, Robert WP. Consider this before using the SARS-CoV-2 pandemic as an instrumental variable in an epidemiological study. *Am J Epidemiol.* 2021;in press.

**Table 1.** Time-series results predicting monthly values of selected birth outcomes in the US from January 2014 to December 2019 as a function of autocorrelation.

Parameter and Lag (in months)	Births		Preterm birth		Extreme preterm birth		Cesarean delivery		Sex ratio at birth	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Constant			0.1077	0.0055	0.0062	0.0001	0.3159	0.0025	1.0473	0.0011
Autoregressive term										
1			0.4278	0.1150						
3	0.5865	0.1110					0.3196	0.1312		
12			0.8595	0.0726	0.7105	0.0713	0.7343	0.0981	0.5499	0.1036
Integrated term										
12	(yes) <sup>a</sup>	(yes) <sup>a</sup>								
Moving average term										
2	-0.3308	0.1318								
9					-0.3105	0.1392	-0.4364	0.1454		

Abbreviations: Coef, coefficient; SE, standard error.

<sup>a</sup>The series exhibited strong autocorrelation at lag 12 months, which required removal of seasonal cycles by taking the 12th difference (i.e., values at month  $t - 12$  subtracted from values at month  $t$ ) to render the series mean stationary.

**Figure 1.** Monthly rate of preterm birth (per 100 live births) in the US. Solid circles plot observed rates from January 2014 to December 2019. Open circles plot forecasted rates from January 2020 to December 2020 based on time-series modeling of seasonality and other forms of autocorrelation. Januarys are demarcated with vertical lines.

ORIGINAL UNEDITED MANUSCRIPT



