brushing aside the additional and possibly untenable assumptions that accompany such a decision.

We therefore encourage readers to rely on other available and up-to-date tutorials[7,11,12] and text books[2,3,13] when designing a study that uses a natural experiment to evaluate the effects of a public health programme.

References

1. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. *Int J Epidemiol* 2018;**47**:2082–93.
2. Angrist JD, Pischke J-S. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press, 2008.
3. Angrist JD, Pischke J-S. *Mastering 'Metrics: The Path From Cause to Effect.* Princeton, NJ: Princeton University Press, 2014.
4. Cartwright N. Are RCTs the gold standard? *BioSocieties* 2007;**2**:11–20.
5. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;**210**:2–21.
6. Bor J. Capitalizing on natural experiments to improve our understanding of population health. *Am J Public Health* 2016;**106**:1388.
7. Abadie A, Cattaneo MD. Econometric methods for program evaluation. *Annu Rev Econ* 2018;**10**:465–503.
8. Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ* 2016;**25**:1514–28.
9. Doudchenko N, Imbens GW. *Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A synthesis.* Cambridge, MA: National Bureau of Economic Research, 2016.
10. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc* 2010;**105**:493–505.
11. Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Public Health* 2017;**38**:39–56.
12. Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. Synthetic control methodology as a tool for evaluating population-level health interventions. *J Epidemiol Community Health* 2018;**72**:673–78.
13. Khandker S, Koolwal GB, Samad H. *Handbook on Impact Evaluation: Quantitative Methods and Practices.* Washington, DC: World Bank, 2009.

Difference in difference, controlled interrupted time series and synthetic controls

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

[1]Department of Public Health Environments and Society, London School of Hygiene and Tropical Medicine, London, UK and [2]Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author. Department of Public Health Environments and Society, London School of Hygiene and Tropical Medicine, London, UK. E-mail: james.lopez-bernal@lshtm.ac.uk

We thank Benmarhnia and Rudolph[1] for their critical appraisal of our recent article on the use of controls in interrupted time series (ITS) studies.[2] This offers the opportunity to clarify some important issues related to ITS and controlled ITS (CITS) designs and their comparison with other methods applied for public health evaluation. In particular, we argue that Benmarhnia and Rudolph based their assessment on three incorrect premises: that ITS without control is not a valid design for assessing causal relationships; that CITS is just another name for the difference-in-difference (DID) design that they advocate; and that the synthetic control methodology represents an alternative to CITS.

A DID design typically refers to a controlled before-and-after study in which the outcome is measured at a single baseline (pre-intervention) time point and a single post-intervention time point, or where pre- and post-intervention means are compared but where 'time' is not incorporated into the model.[3–5] The counterfactual is estimated based on the control group alone and assumes that trends are parallel in the two groups. While approaches can be used to ensure that the two groups are as similar as possible, such as using synthetic controls, the assumption that trends are parallel is not verifiable using this design; therefore, it is highly susceptible to confounding due to between-group differences.

ITS, conversely, uses multiple consecutive pre- and post-intervention observations in a single population and incorporates time. The counterfactual is estimated by extrapolation of the pre-intervention trend and assumes that in the absence of an intervention the trend would remain constant. Because the observations are undertaken in the same population, between-group differences do not present a problem, and the strict temporal structure allows fine control for underlying trends and measured time-varying confounders. However, other events occurring around the time of the intervention can be a source of confounding.[6–8]

Both ITS and DID are generally regarded as intermediate designs in the hierarchy of quasi-experimental designs.[5] However, CITS combines the ITS design with one or more control series, allowing both within- and between-group comparisons and strengthening the control for potential confounders. As such, it provides a more flexible and structured inferential framework, and it is regarded as a more powerful design than DID.[5,8,9] Notably, as an extension of DID, it allows the assumption of parallel trends to be verified and for differences in trend between two groups to be adjusted for. As an extension to ITS, it allows time-varying confounders, including contemporaneous events, that affect both groups to be controlled for.[2]

Synthetic controls are not an alternative approach to CITS; rather, as we describe, they can be used to identify a suitable control series for use in a CITS, and they are thus complementary approaches.[2] The use of synthetic controls in CITS studies has been described in detail elsewhere.[10] Nevertheless, synthetic controls rely on the availability of multiple suitable controls with various measures on several characteristics; a scenario which is often not available in practice.

There are scenarios where no suitable unaffected control group exists—for example, evaluation of national or international policies or events, such as the impact of smoking cessation legislation or the financial crisis— therefore, a CITS design would not be possible.[11,12] Benmarhnia and Rudolph suggest that under such circumstances, uncontrolled ITS is not a suitable alternative and that the intervention should not be evaluated if an 'approach with identification assumptions that are more closely aligned with the data' is not possible. We disagree. ITS is the most powerful study design available where no control group exists.[5,6] Of course, the evidence emerging from an ITS study has to be interpreted in a broader context, taking into account biological plausibility, magnitude of effect

and consistency across settings.[13] Further, the inability to control for possible contemporaneous events should be explicitly acknowledged. Nevertheless, this is not an excuse not to evaluate using the best available approach. Lack of evaluation of such interventions would simply perpetuate the evaluative bias that exists with complex interventions that are challenging to study.[3]

References

1. Benmarhnia T, Rudolph KE. A rose by any other name still needs to be identified (with plausible assumptions). *Int J Epidemiol* 2019;**48**:2061–62.
2. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. *Int J Epidemiol* 2018;**47**:2082–93.
3. Craig P, Cooper C, Gunnell D *et al*. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *J Epidemiol Community Health* 2012;**66**:1182–86.
4. Rockers PC, Røttingen JA, Shemilt I, Tugwell P, Bärnighausen T. Inclusion of quasi-experimental studies in systematic reviews of health systems research. *Health Policy* 2015;**119**:511–21.
5. Soumerai SB, Starr D, Majumdar SR. How do you know which health care effectiveness research you can trust? A guide to study design for the perplexed. *Prev Chronic Dis* 2015;**12**:E101.
6. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002;**27**:299–309.
7. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol* 2017;**46**:348–55.
8. Shadish WR, Cook TD, Campbell DT, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2002.
9. Lopez Bernal JA, Andrews N, Amirthalingam G. The use of quasi-experimental designs for vaccine evaluation. *Clin Infect Dis* 2018; doi:10.1093/cid/ciy906.
10. Linden A. Combining synthetic controls and interrupted time series analysis to improve causal inference in program evaluation. *J Eval Clin Pract* 2018;**24**:447–53.
11. Barone-Adesi F, Gasparrini A, Vizzini L, Merletti F, Richiardi L. Effects of Italian smoking regulation on rates of hospital admission for acute coronary events: a country-wide study. *PLoS One* 2011;**6**:e17419.
12. Laliotis I, Ioannidis JPA, Stavropoulou C. Total and cause-specific mortality before and after the onset of the Greek economic crisis: an interrupted time-series analysis. *Lancet Public Health* 2016;**1**:e56–65.
13. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;**58**:295–300.