Received: 5 August 2016 | Accepted: 8 August 2016
DOI 10.1111/jep.12638

WILEY Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

ORIGINAL ARTICLE

Challenges to validity in single-group interrupted time series analysis

Ariel Linden DrPH,[1,2]

[1] President, Linden Consulting Group, LLC, Ann Arbor, MI, USA

[2] Research Scientist, Division of General Medicine, Medical School, University of Michigan, Ann Arbor, MI, USA

Correspondence
Ariel Linden, Linden Consulting Group, LLC, 1301 North Bay Drive, Ann Arbor, MI 48103, USA.
Email: alinden@lindenconsulting.org

Abstract

Rationale, aims and objectives Single-group interrupted time series analysis (ITSA) is a popular evaluation methodology in which a single unit of observation is studied; the outcome variable is serially ordered as a time series, and the intervention is expected to "interrupt" the level and/or trend of the time series, subsequent to its introduction. The most common threat to validity is *history*—the possibility that some other event caused the observed effect in the time series. Although history limits the ability to draw causal inferences from single ITSA models, it can be controlled for by using a comparable control group to serve as the counterfactual.

Method Time series data from 2 natural experiments (effect of Florida's 2000 repeal of its motorcycle helmet law on motorcycle fatalities and California's 1988 Proposition 99 to reduce cigarette sales) are used to illustrate how history biases results of single-group ITSA results—as opposed to when that group's results are contrasted to those of a comparable control group.

Results In the first example, an external event occurring at the same time as the helmet repeal appeared to be the cause of a rise in motorcycle deaths, but was only revealed when Florida was contrasted with comparable control states. Conversely, in the second example, a decreasing trend in cigarette sales prior to the intervention raised question about a treatment effect attributed to Proposition 99, but was reinforced when California was contrasted with comparable control states.

Conclusions Results of single-group ITSA should be considered preliminary, and interpreted with caution, until a more robust study design can be implemented.

KEYWORDS

causal inference, interrupted time series analysis, quasi-experimental

1 | INTRODUCTION

Single-group interrupted time series analysis (ITSA) is an increasingly popular evaluation methodology for observational data in which a single unit of observation (eg, an individual, a city, or a country) is studied; the dependent variable is a serially ordered time series, and multiple observations are captured in both the pre- and post-intervention periods. The study design is called an *interrupted time series* because the intervention is expected to "interrupt" the level and/or trend of the time series, subsequent to its introduction.[1,2] ITSA has been argued to generally have strong internal validity, primarily through its control over *regression to the mean*,[1–4] and good external validity, particularly when the unit of measure is at the population level, or when the results can be generalized to other units, treatments, or settings.[2,5]

ITSA has been used in many areas of study, such as assessing the effects of community interventions,[6,7] public policy,[8] regulatory actions,[9] and health technology assessment.[10] ITSA has also been proposed as a more flexible and rapid design to be considered in health research before defaulting to the traditional 2-arm randomized controlled trial.[11] In addition, systematic reviews of the literature increasingly include studies using ITSA as the primary research design.[12]

Despite its widespread use, the single-group ITSA design remains a vastly inferior evaluation approach to those utilizing a comparable control group to serve as the counterfactual—a fundamental element of the potential outcomes framework.[13,14] With a comparable control group, factors other than the intervention that are responsible for shifting the time series will likely be observed in both groups and thus not mistaken for a treatment effect. Moreover, events that affect the time series in the treatment group prior to initiation of the intervention can be used in the matching process to ensure that the shift in the time series does not confound the results. Conversely, without a comparable control group, the impact on the time series by an event outside the intervention may be mistaken for a treatment effect.

Other literature has provided both a comprehensive description of the ITSA design and methodological guidance in its implementation (see Box and Tiao,[15] Glass *et al*,[16] and McDowall *et al*[17] for using autoregressive integrated moving-average models; and Crosbie,[18] Gottman,[19] Linden and Adams,[20] Linden,[21] McKnight *et al*,[22] Simonton,[23] and Velicer and McDonald[24] for using ordinary least-squares regression-based models). The purpose of the current paper, however, is to offer a nontechnical discussion of how factors that impact the time series outside of the intervention may be mistaken for a treatment effect when using the single-group ITSA model, but captured when using a comparable control group to serve as the counterfactual. This problem is illustrated using data from 2 natural experiments; the effect of Florida's 2000 motorcycle helmet law repeal on motorcycle fatality rates, and the effect of California's 1988 Proposition 99 antismoking initiative on cigarette sales.

2 | USUAL THREATS TO VALIDITY IN SINGLE-GROUP ITSA DESIGNS

While the single-group ITSA design can control for many threats to validity, the remaining threats that the design does not control for are crucial (see Campbell & Stanley[1] and Shaddish *et al*[2] for a comprehensive description of the threats to validity in ITSA and many other evaluation designs).

History is the principal threat to validity—the possibility that some event other than the intervention caused the observed effect in the time series.[2] There are at least 2 scenarios where the effect of history may be misconstrued. First, when the change in the time series is immediate and drastic, it is easy to ignore the possibility that some other factor may be the cause. And even if there is an alternative explanation for the effect, information may not always be available to identify those factors. Thus, the investigator is likely to argue that the effect is causally related to the intervention without further study. In the second scenario, some factor may cause a directionally correct change in the time series prior to the intervention. Thus, any additional change in the time series subsequent to the introduction of the intervention may be argued to be a continuation or magnified effect of that prior factor and not a treatment effect.[21,25] In either of these scenarios, the inclusion of a comparable control group will clarify these issues.

Instrumentation, or a change in how the time series is measured, is another threat to validity that may erroneously appear as a treatment effect in a single-group ITSA.[2] While documentation should be obtained indicating how and when the instrumentation changed, it may nevertheless be impossible to control for this bias in a single-group ITSA. However, with the inclusion of a comparable control group, the change in instrumentation should impact both time series equally, thereby nullifying its effect.

Selection may bias the single-group ITSA if the serial observations are cross-sectional and the characteristics (or composition) of the group under study are different before and after the introduction of the intervention (selection is not a factor in a single-group ITSA where the same group, or individual, undergoes surveillance over the duration of the study). Selection may be controlled for by finding a control group that is comparable to the treatment group on pre-intervention characteristics (at the very least, the groups should be comparable on the pre-intervention level and trend of the outcome under study).[20,21]

Threats to *statistical conclusion validity* apply as much to ITSA as to any other design, such as low power, violated test assumptions, and unreliability of measurement.[2] While these issues are important, their discussion is beyond the scope of this paper (the reader is referred to references[15–24] for a comprehensive discussion of the relevant statistical issues in ITSA models).

3 | EXAMPLE 1: THE REPEAL OF FLORIDA'S MOTORCYCLE HELMET LAW

On July 1, 2000, the State of Florida partially repealed its motorcycle helmet law by exempting adult motorcyclists (aged 21 years and older) and moped riders from wearing a helmet—provided that they carry motorcycle insurance coverage with a minimum of $10 000 in medical benefits for injuries sustained in a motorcycle accident. The law continued to require helmets for riders younger than 21 years of age. Several studies have examined the effect of the Florida helmet repeal on motorcycle fatalities and have collectively concluded that weakening of the helmet law led to increased motorcycle fatalities.[26–29] A major shortcoming common to all these studies is that no contrasts were made with other comparable states.

For the current analysis, all motor vehicle fatality data for all states were retrieved from the Fatal Accident Reporting System database for the years 1975 to 2014 (which is all the data available in the system).[30] Annual issues of *Highway Statistics* provided motorcycle registration data for the periods of 1996 to 2001, and years between 1975 and 1996 were retrieved from the 1995 summary volume.[31] Statistical analyses were conducted using `ITSA`, a program written for STATA to conduct single-group and multiple-group interrupted time series analyses.[21]

Figure 1a presents the raw motorcycle fatality counts in Florida annually from 1975 to 2014. As shown, motorcycle deaths were decreasing annually from 1975 until the repeal in 2000, followed immediately by a sharp jump in deaths that continued to rise annually thereafter until 2014. Figure 1b presents annual motorcycle fatalities as a percent of total motor vehicle deaths. The overall behavior of this time series is nearly identical to that of raw motorcycle deaths (Figure 1a). The percentage of motorcycle deaths relative to all motor vehicle fatalities decreased annually between 1975 and 2000, followed by an immediate (and thereafter increasing) rise.

On the face of it, these 2 figures (Figure 1a,b) lend compelling support for the hypothesis that Florida's helmet repeal led to increased motorcycle fatalities—both in raw counts and relative to all other motor vehicle deaths. Additionally, based on these figures alone, most relevant threats to validity[2] could be ruled out. For example, *regression to the mean* can be ruled out as a rival explanation because the lengthy pre-intervention time series shows a consistent decrease in

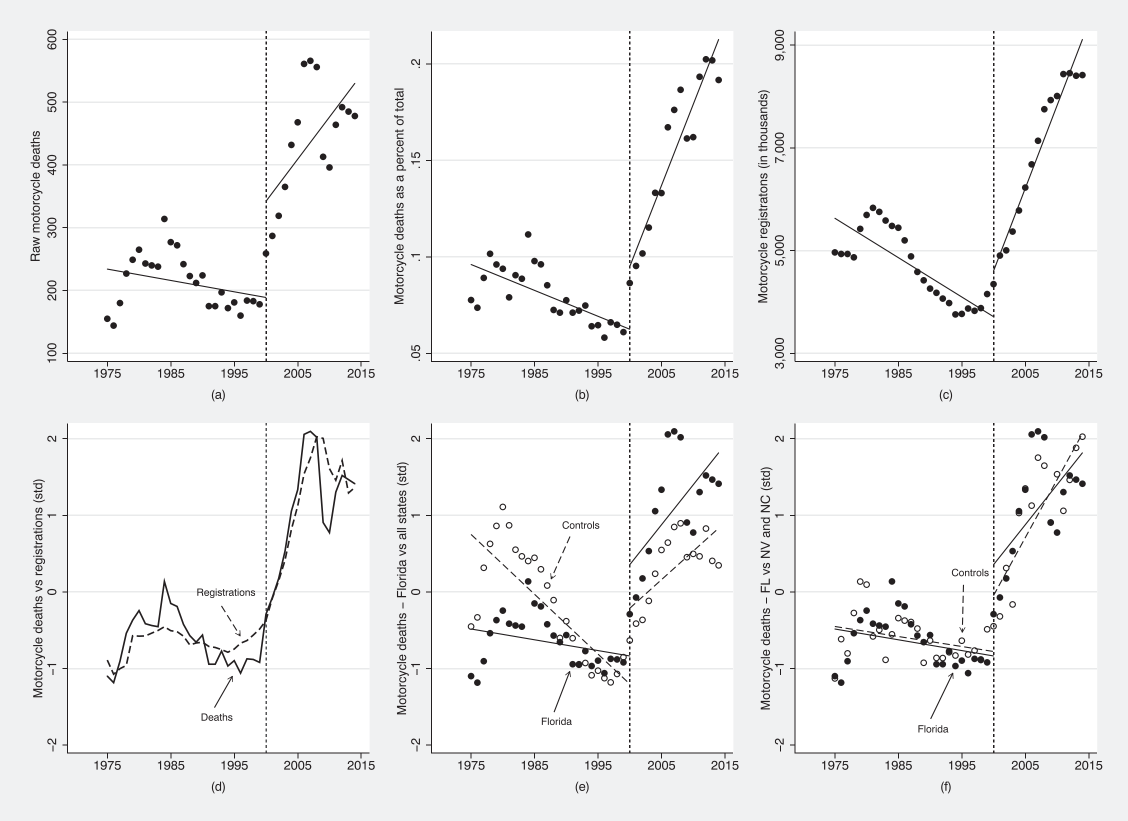

FIGURE 1 Florida motorcycle deaths and registrations from 1975 to 2014, using single-group and multiple-group ITSA designs

motorcycle deaths over the entire period. Thus, the jump in the level of the time series immediately following the repeal cannot be viewed as a response to an outlier observation occurring immediately prior to the repeal. *Selection bias* may pose a threat to validity if the characteristics of those who died after the repeal differed systematically from those who died prior to the repeal, with the most likely case being made for more deaths for motorcyclists older than 21 years of age. However, neither Muller[27] nor Kyrychenko and McCartt[29] found differential fatality rates based on the age cutoff. History is a plausible threat to validity only if another event or action had occurred simultaneously with the repeal, given that the trend in fatalities was decreasing annually prior to the repeal. However, given such an immediate and dramatic effect on the time series concomitant with the repeal, it may appear rather unlikely that any other factor could have caused the effect outside the intervention.

However, Figure 1c and 1d cast doubt on the assertion that the helmet law repeal caused the increase in motorcycle deaths. As illustrated, motorcycle registrations followed a nearly identical historic pattern as motorcycle deaths (with a very high correlation between them of 0.95). Most notable in this time series is the sharp increase in motorcycle registrations commencing in 2000—after many years of declining rates. In light of these data, one may revise the prior hypothesis to now consider that the helmet law repeal is associated with more people registering motorcycles, which in turn is associated with more deaths.

Figure 1e and 1f offer a complete rebuttal for any causal association between Florida's helmet law repeal and the rise in motorcycle fatalities. In Figure 1e, motorcycle fatalities in Florida are compared to those of all other States (excluding Arkansas, Kentucky, Michigan, Pennsylvania and Texas—states that repealed their helmet laws during some point in the same timeframe under study). The time series were ipsatively standardized[32] so that they could be compared on the same scale. As shown, nationally there was an even sharper downward trend in motorcycle deaths prior to 2000 than in Florida. However, similar to Florida, there was both an immediate and prolonged increase in motorcycle deaths after 2000. As one can see from the intermingled observations between the 2 time series, the trends are not statistically different from each other. Although not shown, national motorcycle registrations followed a similar annual trajectory to that in Florida. Thus, one may now further conclude that there was some event that caused people to register motorcycles in large numbers throughout the country starting in 2000, and this in turn was associated with increasing annual motorcycle fatalities.

Finally, one may argue that the comparison between Florida and all other states is biased because the 2 are not comparable on either baseline level or trend of the outcome variable.[20,21] To address this concern, an optimal matching algorithm was implemented to identify states that matched Florida on both baseline level and trend of standardized motorcycle fatalities. As illustrated in Figure 1f, Nevada and North Carolina were virtually indistinguishable from Florida across the entire time series from 1975 to 2014, including, and most importantly, the year 2000, in which the Florida helmet law was repealed.

In summary, this example demonstrates that a seemingly irrefutable treatment effect detected on reviewing data from a single time series can be disproven when that time series is contrasted with that of a comparable control group.

4 | EXAMPLE 2: CALIFORNIA'S PROPOSITION 99 ANTI-SMOKING INITIATIVE

In 1988, California passed the voter-initiative Proposition 99, which was a widespread effort to reduce smoking rates by raising the cigarette excise tax by 25 cents per pack and to fund anti-smoking campaigns and other related activities throughout the state (see Breslow and Johnson[33] and Siegel[34] for a comprehensive discussion of this initiative). Several studies have shown that cigarette consumption in California after the passage of Proposition 99 in 1988 was lower than the average national trend and lower than the linearly extrapolated pre-intervention trend in California (see Breslow and Johnson,[33] Glantz,[35] Fichtenberg and Glantz,[36] and among others).

Per capita cigarette sales (in packs) is the most widely used indicator of smoking prevalence found in the tobacco research literature[37] and serves here as the aggregate outcome variable under study, measured annually at the state level from 1970 until 2000 (with 1989 representing the first year of the intervention). The current data were obtained from Abadie *et al*,[38] who obtained the data from Orzechowski and Walker.[39] Eleven states were discarded from the dataset because of their adoption of some other large-scale tobacco control program at some point during California's intervention period under study between 1989 and 2000, leaving 38 states as potential controls (Abadie *et al*[37]).

Figure 2a illustrates the annual time series of cigarette sales per capita in California from 1970 to 2000. As shown, per capita cigarette sales began to decrease in 1976 and continued its downward trajectory until 2000. There does appear to have been an "interruption" in the time series coinciding with the initiation of Proposition 99, after which the annual trend decreased more so than prior to 1999.

Given that the internal validity of the ITSA design rests on the premise that the interruption in the time series is associated with the introduction of the treatment, treatment effects may seem less plausible if a shift in the time series appears prior to the actual intervention. Such a shift would indicate that an external factor was already influencing the time series and imply that any additional shifts may simply be a continuation of that factor's impact. Using these same cigarette sales data, Linden and Yarnold[25] found that numerous structural breaks occurred prior to the actual initiation of Proposition 99 in 1989, including perfect structural breaks in 1983 and 1985. Figure 2b illustrates that the linear trend between 1983 and 1989 is nearly identical to the linear trend following the introduction of Proposition 99, casting doubt on whether there was an intervention effect associated with Proposition 99, or simply an additional structural break due to some factor outside of the intervention.

Figure 2c illustrates the comparison of California to all other states that had not yet implemented any anti-smoking campaign. As shown, the annual linear trend in cigarette sales after 1989 is decreasing much more so in California than in the other states, pointing to an intervention effect associated with Proposition 99. However, as in the previous example, one could argue that the comparison between California and all other states is biased because the 2 are not comparable on either baseline level or trend of the outcome variable.[20,21] To address this concern, an optimal matching algorithm was implemented to identify states that matched California on both baseline level and trend of per capita cigarette sales. As illustrated in Figure 2d, Colorado, Idaho, and Montana were very comparable to California in both level and trend of cigarette sales across the entire pre-intervention period spanning from 1970 to 1989. However, California's cigarette sales declined much more so than these control states after the initiation of Proposition 99, indicating a treatment effect.

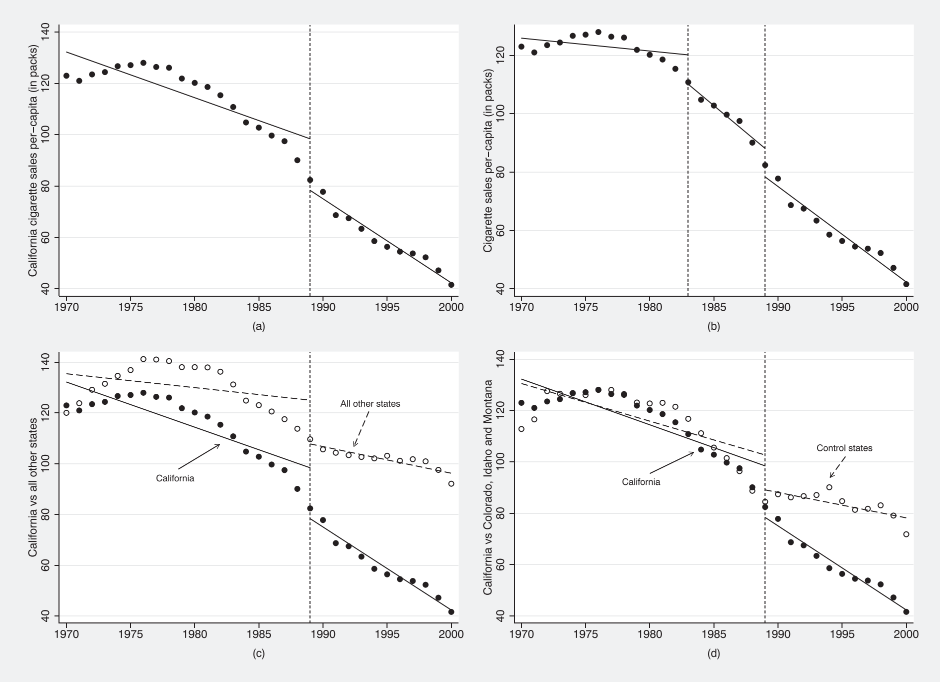

FIGURE 2 Cigarette sales in California from 1970 to 2000, using single-group and multiple-group ITSA designs

In summary, this example demonstrates that when some factor causes a shift in the time series prior to the actual introduction of the intervention, it raises the concern that any shift subsequent to the introduction of the intervention, may be related to this prior factor, rather than the intervention. To control for this confounder, the treated group's pre-intervention time series is matched to that of a comparable control group. The result here was that Proposition 99 appeared to be effective when contrasted to a comparable control group.

5 | DISCUSSION

The 2 examples presented in this paper illustrate how the single-group ITSA model can easily provide misleading results about the effects of an intervention, because the effects of other competing factors cannot be identified, or controlled for. In the first example, a seemingly unquestionable treatment effect was reversed when contrasted with a comparable control group. Conversely, in the second example, a debatable treatment effect (due to a preexisting directionally correct trend in the time series) was reinforced when the treatment group was contrasted with a comparable control group. In short, even with an extensive number of pre- and post-intervention observations to control for regression to the mean and other biases, the single-group ITSA design may be no better than the simple single-group pretest-posttest design for causal inference. Thus, a more robust ITSA design must be employed if inferences about the intervention are to be considered valid and casual.

As demonstrated in the present examples, using a comparable control group to serve as the counterfactual provides a robust approach for assessing treatment effects. Only when contrasted with a comparable control group can the effect of the intervention (or lack thereof) be isolated from other rival factors. Moreover, other anomalies observed in the time series (such as changes in instrumentation, selection bias, etc) can alert the investigator to other potential sources of confounding.

When multiple nontreated units are available, investigators can choose from at least 3 different matching methods suitable for time series data. This includes the matching process implemented in the present examples (ie, finding those nontreated units that are nonstatistically different from the treated unit on pre-intervention levels and trend of the outcome variable),[21] a synthetic controls approach[37] or propensity score-based weighting[20] (which can also be extended to longitudinal data with multiple treated units[40] and for censored data.[41,42] The ITSA framework with a comparable control group can be further strengthened by implementing a cross-over design, wherein the groups switch their treatment assignment at a given time-point (ie, the treatment group switches to control and the control switches to treatment), and the outcomes change in accordance with the exposure to the intervention.

When a control group is simply not available, a version of the cross-over design can be implemented with a single group as well. Here the intervention is administered and withdrawn, repeatedly. The results may be considered a causal effect of the intervention if the treatment effect changes in a similar fashion after each successive administration. A limitation of any cross-over design, however, is that it requires the ability to control the treatment assignment, thereby restricting its application from most natural experiments (see Barlow *et al*[43] for many other ITSA design alternatives to improve causal inference over the basic single-group design).

In summary, this paper illustrated 2 cases in which erroneous conclusions may be drawn about the effectiveness of an intervention when using the single-group ITSA design for evaluation. Absent a comparable control group as a contrast, there is no assurance that the effect of external factors have been identified and controlled for. Thus, the results should be considered preliminary—and interpreted with caution—until a more robust study design can be implemented. Given the popularity and widespread use of the single-group ITSA design, it is important for investigators to be cognizant of its limitations and to strive to add features that maximize its validity and improve causal inference.

ACKNOWLEDGEMENT

I wish to thank Julia Adler-Milstein for reviewing the manuscript and providing many helpful comments.

REFERENCES

1. Campbell DT, Stanley JC. *Experimental and Quasi-experimental Designs for Research*. Chicago, IL: Rand McNally; 1966.
2. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin; 2002.
3. Linden A. Estimating the effect of regression to the mean in health management programs. *Dis Manag Health Outcomes*. 2007;15:7–12.
4. Linden A. Assessing regression to the mean effects in health care initiatives. *BMC Med Res Methodol*. 2013;13:1–7.
5. Linden A, Adams J, Roberts N. The generalizability of disease management program results: getting from here to there. *Manag Care Interface*. 2004;17:38–45.
6. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci*. 2000;1:31–49.
7. Gillings D, Makuc D, Siegel E. Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *Am J Public Health*. 1981;71:38–46.
8. Muller A. Florida's motorcycle helmet law repeal and fatality rates. *Am J Public Health*. 2004;94:556–558.
9. Briesacher BA, Soumerai SB, Zhang F, *et al*. A critical review of methods to evaluate the impact of FDA regulatory actions. *Pharmacoepidemiol Drug Saf*. 2013;22:986–994.
10. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *Int J Technol Assess Health Care*. 2003;19:613–623.
11. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. *Clin Transl Med*. 2013;2:1–6.
12. Effective Practice and Organisation of Care (EPOC). Interrupted time series (ITS) analyses. 2015. EPOC Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services. Available at: http://epoc.cochrane.org/epoc-specific-resources-review-authors
13. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66:688–701.
14. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc*. 2005;100:322–331.

15. Box GEP, Tiao GC. Intervention analysis with applications to economic and environmental problems. *J Am Stat Assoc*. 1975;70:70–79.
16. Glass GV, Willson VL, Gottman JM. *Design and Analysis of Time-series Experiments*. Boulder: University of Colorado Press; 1975.
17. McDowall D, McCleary R, Meidinger EE, Hay RA. *Interrupted Time Series Analysis*. Newbury Park, CA: Sage Publications, Inc.; 1980.
18. Crosbie J. Interrupted time-series analysis with brief single-subject data. *J Consult Clin Psychol*. 1993;61:966–974.
19. Gottman JM. *Time-series analysis. A Comprehensive Introduction for Social Scientists*. New York: Cambridge University Press; 1981.
20. Linden A, Adams JL. Applying a propensity-score based weighting model to interrupted time series data: Improving causal inference in program evaluation. *J Eval Clin Pract*. 2011;17:1231–1238.
21. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. *The Stata J*. 2015;15:480–500.
22. McKnight S, McKean JW, Huitema BE. A double bootstrap method to analyze linear models with autoregressive error terms. *Psychol Methods*. 2000;5:87–101.
23. Simonton DK. Cross-sectional time-series experiments: some suggested statistical analyses. *Psychol Bull*. 1977;84:489–502.
24. Velicer WF, McDonald RP. Cross-sectional time series designs: a general transformation approach. *Multivar Behav Res*. 1991;26:247–254.
25. Linden A, Yarnold PR. Using machine learning to identify structural breaks in single-group interrupted time series designs. *J Eval Clin Pract*. DOI:10.1111/jep.12544.
26. Hotz GA, Cohn SM, Popkin C, *et al.* The impact of a repealed motorcycle helmet law in Miami-Dade County. *J Trauma*. 2002;52:469–474.
27. Muller A. Florida's motorcycle helmet law repeal and fatality rates. *Am J Public Health*. 2004;94:556–558.
28. Turner P, Hagelin C, Chu X, Greenman M, Read J, West M. *Florida Motorcycle Helmet Use Observational Survey and Trend Analysis*. Center for Urban Transportation Research, University of South Florida, Tampa, Florida. 2004.
29. Kyrychenko SY, McCartt AT. Florida's weakened motorcycle helmet law: effects on death rates in motorcycle crashes. *Traffic Inj Prev*. 2006;7:55–60.
30. U.S. Department of Transportation, National Highway Traffic Safety Administration, National Center for Statistics and Analysis. Fatality analysis reporting system (FARS). Available at: ftp://ftp.nhtsa.dot.gov/fars/ Accessed July 5, 2016.
31. U.S. Department of Transportation, Federal Highway Administration. Highway statistics (multiple years). Available at http://www.fhwa.dot.gov/policyinformation/statistics.cfm Accessed July 5, 2016.
32. Yarnold PR, Linden A. Using machine learning to model dose–response relationships via ODA: eliminating response variable baseline variation by ipsative standardization. *Optim Data Anal*. 2016;5:41–52.
33. Breslow L, Johnson M. California's Proposition 99 on tobacco, and its impact. *Annu Rev Public Health*. 1993;14:585–604.
34. Siegel M. The effectiveness of state-level tobacco control interventions: a review of program implementation and behavioral outcomes. *Annu Rev Public Health*. 2002;23:45–71.
35. Glantz S. Changes in cigarette consumption, prices, and tobacco industry revenues associated with California's Proposition 99. *Tob Control*. 1993;2:311–314.
36. Fichtenberg C, Glantz S. Association of the California tobacco control program with declines in cigarette consumption and mortality from heart disease. *N Engl J Med*. 2000;343:1772–1777.
37. 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.
38. Abadie A, Diamond A, Hainmueller J. SYNTH: STATA module to implement synthetic control methods for comparative case studies. Statistical Software Components S457334, Department of Economics, Boston College. 2014. Available from: https://ideas.repec.org/c/boc/bocode/s457334.html
39. Orzechowski W, Walker RC. *The Tax Burden on Tobacco. Historical Compilation*, vol. 40. Arlington, VA: Orzechowski & Walker; 2005.
40. Linden A, Adams JL. Evaluating health management programmes over time. Application of propensity score-based weighting to longitudinal data. *J Eval Clin Pract*. 2010;16:180–185.
41. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
42. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: an introduction to survival analysis. *Dis Manag*. 2004;7:180–190.
43. Barlow DH, Hayes SC, Nelson RO. *The Scientist Practitioner: Research and Accountability in Clinical and Educational Settings*. New York: Pergamon Press; 1984.

How to cite this article: Linden A. Challenges to validity in single-group interrupted time series analysis. *J. Eval. Clin. Pract*. 2017;23:413–418. https://doi.org/10.1111/jep.12638