Challenges to validity in single-group interrupted time series analysis -- even with a crossover design

Ariel Linden, DrPH[1,2]

[1] President, Linden Consulting Group, LLC - Ann Arbor, MI alinden@lindenconsulting.org

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

Corresponding Author Information:
Ariel Linden, DrPH
Linden Consulting Group, LLC
1301 North Bay Drive
Ann Arbor, MI USA 48103
Phone: (971) 409-3505
Email: alinden@lindenconsulting.org

Key Words: interrupted time series analysis, quasi-experimental, causal inference, crossover design

Running Header: validity of interrupted time series analysis

Acknowledgement: I wish to thank Grant Miller for providing the data used in the second example of this paper. I also wish to thank Julia Adler-Milstein for reviewing the manuscript and providing many helpful comments.

Author Manuscript

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jep.12668

This article is protected by copyright. All rights reserved.

ABSTRACT

Rationale, aims and objectives: The basic single-group ITSA design has been shown to be susceptible to the most common threat to validity -- *history* -- the possibility that some other event caused the observed effect in the time series. A single-group ITSA with a crossover design (in which the intervention is introduced and withdrawn one or more times) should be more robust. In this paper, we describe and empirically assess the susceptibility of this design to bias from *history*.

Method: Time series data from two natural experiments (the effect of multiple repeals and reinstatements of Louisiana's motorcycle helmet law on motorcycle fatalities, and the association between the implementation and withdrawal of Gorbachev's anti-alcohol campaign with Russia's Mortality Crisis) are used to illustrate that *history* remains a threat to ITSA validity, even in a crossover design.

Results: Both empirical examples reveal that the single-group ITSA with a cross-over design may be biased due to *history*. In the case of motorcycle fatalities, helmet laws appeared effective in reducing mortality (while repealing the law increased mortality), but when a control group was added, it was shown that this trend was similar in both groups. In the case of Gorbachev's anti-alcohol campaign, only when contrasting the results against those of a control group, was the withdrawal of the campaign found to be the more likely culprit in explaining the Russian Mortality Crisis than the collapse of the Soviet Union.

Author Manuscript

This article is protected by copyright. All rights reserved.

Conclusions: Even with a robust crossover design, single-group ITSA models remain susceptible to bias from *history*. Therefore, a comparable control group design should be included, whenever possible.

Author Manuscript

This article is protected by copyright. All rights reserved.

1. INTRODUCTION

Single-group interrupted time-series analysis (ITSA) is a popular evaluation strategy for observational data in which a single unit is studied (e.g., an individual, a city, or a country), 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]. It has been maintained that ITSA generally has strong internal validity, primarily through its control over *regression to the mean* [1,2,3,4] and good external validity, particularly when the analysis occurs at the population level, or when the results can be generalized to other units, treatments or settings [2,5].

Recently, the validity of the basic single-group ITSA design (consisting of a pre-intervention phase and an intervention phase) has been scrutinized. Linden [6] illustrated that this design can either fail to identify the effects of external factors on the time series, resulting in a false causal attribution, or conversely, confuse the causal interpretation when a directionally-correct change in the time series also occurs prior to the intervention. In both cases, the inclusion of a comparable control group clarifies causal effects.

Shaddish et al [2] suggest adding design features such as the removal of the treatment at a known time, or by extension, adding multiple replications, to improve the validity of the single-group ITSA design. In essence, this is a single-group version of the cross-over design, in which the intervention is introduced and withdrawn, one or more times. The underlying premise is that it would be increasingly unlikely that external events will impact the time series coincidentally

This article is protected by copyright. All rights reserved.

with each successive crossover, and thus, the results can be considered a causal effect of the intervention if the time series changes accordingly [2].

The purpose of the current paper is to offer a non-technical discussion of how, even with the addition of a crossover design, factors other than the intervention may be mistaken for a treatment effect (or withdrawal of treatment) when only a single group is being evaluated. By way of example, it will be shown that the effects of external events can only be identified and controlled for by utilizing a comparable control group to serve as the counterfactual -- a fundamental element of the potential outcomes framework [7,8]. With the inclusion of a comparable control group, factors other than the intervention that are responsible for shifting the time series in each crossover phase will likely be observed in both groups, and thus not mistaken for an effect of treatment or withdrawal. Likewise, directionally-correct changes that do occur in the intervention group but not in the control group may be interpreted as causal [9,10]. This problem is illustrated using data from two natural experiments; the effect of multiple repeals and reinstatements of Louisiana's motorcycle helmet law on motorcycle fatalities, and the association between the implementation and withdrawal of Gorbachev's anti-alcohol campaign and Russia's Mortality Crisis in the early 1990s.

2. THREATS TO VALIDITY IN THE BASIC AND CROSSOVER SINGLE-GROUP ITSA DESIGNS

While the basic single-group ITSA design (consisting of a pre-intervention and intervention phase) can control for many threats to validity, the remaining threats that the design does not

Author Manuscript

This article is protected by copyright. All rights reserved.

control for, are critical [1,2]. Consequently, investigators typically add features to the single-group design, such as one or more treatment crossovers, with the intent of mitigating the influence of these remaining biases.

History is the possibility that some event other than the intervention caused the observed effect in the time series [2], and it is the principal threat to validity of any single-group ITSA design. There are at least two scenarios where the effect of *history* may be overlooked or misinterpreted. First, some factor may cause a directionally-correct change in the time series prior to the intervention. As such, any additional change in the time series subsequent to the introduction of the intervention may be considered a continued or magnified effect of that prior factor rather than a treatment effect [9]. Recently, sensitivity tests adapted from the regression-discontinuity literature [11] have been applied to the ITSA design to identify these false treatment effects [12]. In the second scenario, the change in the time series after initiation of the intervention is immediate and drastic, and as such, it is easy to ignore the possibility that some other factor may be the cause. 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 [6].

A crossover design is considered a more robust approach to control for *history* than the basic single-group ITSA, by virtue of the fact that there are more comparisons involved, making it less likely that some external event will impact the time series repeatedly at each crossover phase. For example, adding a specific time-point when the treatment is deliberately withdrawn,

Author Manuscript

This article is protected by copyright. All rights reserved.

allows the investigator to evaluate change in the time series between three distinct phases, (pre-intervention versus intervention phase, intervention versus post-intervention, pre-intervention versus post-intervention). Conversely, in a simple ITSA design, the intervention is considered on-going, thereby limiting the comparison to only the pre-intervention and intervention phase. If the time series changes in the expected direction after each crossover, then it may be harder to argue that the effect was caused by some external event. In theory, as more crossovers are added to the study, the threat of *history* should diminish accordingly. However, other external events may in fact be causing changes in the time series coincidentally with the initiation or withdrawal of the intervention, regardless of the number of crossovers. Only with the inclusion of a control group for comparison, will the effect of external events on the time series be identified.

While *history* is the most common threat to validity, there are three other threats to which ITSA is susceptible. *Instrumentation,* or a change in how the time series is measured, is a threat to validity that may erroneously appear as a treatment effect in both the basic single-group ITSA and cross-over design [2]. As an example, in health management interventions, patients' health behaviors are sometimes measured on different scales over time or a particular scale may be altered [13,14]. As a result, the measurements will be both inconsistent and unreliable. In general, while documentation should be obtained indicating how and when the instrumentation changed, it may nevertheless be impossible to control for this bias in either a single-group ITSA or crossover design. However, with the inclusion of a comparable control group, the change in instrumentation should impact both time series equally, thereby nullifying its effect.

Author Manuscript

This article is protected by copyright. All rights reserved.

Selection may bias both the single-group ITSA and crossover design if the serial observations are cross-sectional and the characteristics (or composition) of the group under study are different in any two (or more) crossover phases of the study (selection is not a factor in either a single-group ITSA or crossover design 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) [9,10].

Threats to *statistical conclusion validity* common to any study design also apply to ITSA. These include 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 Box and Tiao [15], Glass et al. [16], McDowall et al. [17], Crosbie [18]; Gottman [19]; Linden [9]; Linden & Adams [10]; McKnight et al. [20]; Simonton [21]; and Velicer and McDonald [22].

In the following two empirical examples, we demonstrate the susceptibility of the single-group ITSA crossover design to bias from *history*. In the first example, the data is analyzed as a single-group ITSA where the estimated effects substantiate the hypothesized effects. Next, the time series for the intervention group is contrasted with that of a control group to see if the treatment effects still stand. If the time series for both groups are similar across all phases, then we have demonstrated that the single-group ITSA results are biased due to *history* (because

Author Manuscript

This article is protected by copyright. All rights reserved.

events outside of the intervention are impacting the control group as well). In the second example the data is analyzed as a single-group ITSA where the estimated effects substantiate the hypothesized effects. However, given that our focus in this example is on the cause of the spike in the time series at the time of the intervention's withdrawal, the time series for the intervention group is contrasted with that of several different control groups as sensitivity analyses. *History* may be the source of bias if the time series spikes in control groups where we hypothesize they should remain stable (because an external event was causing a spike in both groups similarly).

3. EXAMPLE 1: THE EFFECT OF MULTIPLE REPEALS AND REINSTATEMENTS OF LOUISIANA'S MOTORCYCLE HELMET LAW ON MOTORCYCLE DEATHS

Louisiana first enacted a universal motorcycle helmet law in 1968. Then, in 1976 the law was partially repealed to require helmet use only by riders under the age of 18. In 1982, the universal helmet law was reinstated. In 1999, the law was amended to require helmet use only by motorcyclists under the age of 18 and riders over 18 who did not have a minimum of $10,000 in medical insurance coverage. In 2004, the universal helmet law was again reinstated [23]. This unusual policy situation of multiple repeals/reinstatements (4 distinct phases) of the law provides an excellent opportunity for testing the validity of the single-group ITSA cross-over design to assess the impact of the helmet law (and its repeal) on motorcycle fatalities.

For the current analysis, all motor vehicle fatality data for all states were retrieved from the Fatal Accident Reporting System (FARS) database for the years 1975 to 2014 (which is all the data available in the system) [24]. Annual issues of *Highway Statistics* provided motorcycle

Author Manuscript

This article is protected by copyright. All rights reserved.

registration data for the periods of 1996 to 2001, and years between 1975 and 1996 were retrieved from the 1995 summary volume [25]. Statistical analyses were conducted using `ITSA`, a program written for Stata to conduct single-group and multiple-group interrupted time series analyses [9,10]. When comparisons are made between times series on different scales or at different levels of the same scale, the time series were ipsatively standardized [26] to allow for comparisons on the same scale.

Figure 1A presents the raw motorcycle annual fatality counts in Louisiana from 1975 to 2014. As shown, motorcycle deaths increased from 1975 to 1982, the period in which the helmet law applied only to riders under the age of 18. Motorcycle deaths decreased annually from 1982 until 1999, the period in which the universal helmet law was reinstated. Motorcycle deaths increased annually from 1999 to 2004, the period in which the helmet law was once again applied only to riders under the age of 18, and annual motorcycle deaths appeared to stabilize (that is, the linear trend was flat) between 2004 and 2014, the period in which the universal helmet law was once again reinstated.

As clearly illustrated in Figure 1A, the annual motorcycle fatalities were directionally consistent with the expected effect. That is, Louisiana's multiple helmet law repeals were associated with increased motorcycle fatalities, and the universal helmet law reinstatements were associated with reduced or flat annual fatalities. In addition, one could argue that most likely threats to validity [2] could be ruled out. For example there were a sufficient number of annual observations in each treatment phase (i.e. repeal and reinstatement periods) to rule out *regression*

Author Manuscript

This article is protected by copyright. All rights reserved.

to the mean as a rival explanation [3,4]. *Selection bias* may pose a threat to validity if the characteristics of those who died in each treatment phase differed systematically from those who died in the alternate treatment phase, with the most likely case being made for differential fatality rates based on the age cutoff of 18 years. However, an analysis by age group did not bear that out (data not shown). *History* is a plausible threat to validity only if another event or action had occurred concomitantly with the initiation of each treatment phase. However, it is easy to dismiss the possibility that any other factor could have caused the effect outside of the policy changes, given such an immediate and dramatic effect that occurs simultaneously with the initiation of each and every treatment phase.

However, Figure 1B casts doubt on the hypothesis that the repeated repeals of the helmet law in Louisiana *caused* the increase in motorcycle deaths. As shown, motorcycle registrations followed a nearly identical historic pattern as motorcycle deaths, with the exception of the last few years of the dataset, where a change in methodology for estimating registered vehicles may have distorted the reporting of registrations (a good example of *instrumentation* [1,2]). In light of these data, one may revise the prior hypothesis to now consider that motorcycle registrations hold the primary association with motorcycle deaths in Louisiana, rather than the repeal and reinstatement of the helmet law.

Figure 1C offers a complete rebuttal for any causal association between Louisiana's multiple helmet law repeals and reinstatements and the change in motorcycle fatalities. Here, standardized motorcycle fatalities in Louisiana are compared to those of all other States

Author Manuscript

This article is protected by copyright. All rights reserved.

(excluding Arkansas, Florida, Kentucky, Michigan, Pennsylvania and Texas – states that repealed their helmet laws during some point in the same timeframe under study). As shown, the general behavior of the time series in Louisiana is similar to that of all other states, irrespective of the multiple repeals and reinstatements of the universal helmet law in Louisiana.

Figure 1D illustrates standardized motorcycle registrations in Louisiana and the control states over the entire observation period. When paired with the corresponding motorcycle fatalities presented in Figure 1C, the close relationship between registrations and fatalities remains evident (until the mid-2000s, when the change in methodology for estimating registrations began to distort the measurement of motorcycle registrations [see http://www-fars.nhtsa.dot.gov/common/FARS%20Encyclopedia%20VMT%20Changes.pdf for the details of the methodologic changes].

In summary, despite improving the robustness of the single-group ITSA design by including multiple treatment cross-overs (i.e. repeal to reinstatement, etc.), the effects of *history* still biased the results of the evaluation. Moreover, this bias was only revealed when Louisiana's time series was contrasted with that of the control states. The results of this analysis suggest that motorcycle fatalities are *not* causally related to the helmet law (i.e. they do not decrease as a result of the law being enacted/reinstated, and they do not increase as a result of the law being repealed). However, given that motorcycle fatalities are so closely associated to motorcycle registrations, an alternate hypothesis may simply be that with more motorcycles on the road, there will be a likewise increase in the number of fatalities.

Author Manuscript

This article is protected by copyright. All rights reserved.

4. EXAMPLE 2: THE ASSOCIATION BETWEEN GORBACHEV'S ANTI-ALCOHOL CAMPAIGN AND RUSSIA'S MORTALITY CRISIS

In 1985, Mikhail Gorbachev initiated an anti-alcohol campaign throughout the Soviet Union. The campaign was unprecedented in scale and scope, and it operated through both supply-side and demand-side channels, simultaneously raising the effective price of drinking and subsidizing substitutes for alcohol consumption. At the height of the campaign, official alcohol sales had fallen by as much as two-thirds. In practice, the campaign lasted beyond its official end in 1988, as restarting state alcohol production took time, and alcohol prices remained elevated [27].

All-cause mortality in the Soviet Union decreased during the campaign years but rose precipitously between 1990 and 1994 (a period which has been referred to as the "Russian Mortality Crisis"). Because this episode also overlapped with Russia's political and economic transition to capitalism and democracy, the underlying cause of the mortality crisis has been subject to considerable debate. However, Bhattacharya et al [27] provide a compelling argument that the mortality crisis was mostly attributable to the coincident termination of the Gorbachev anti-alcohol campaign rather than the political and economic transition.

For the present example, we reanalyze data from Bhattacharya et al [27] in two ways. First we conduct a single-group ITSA with a cross-over design to evaluate the effect of introducing and then withdrawing the anti-alcohol campaign on mortality in Russia between the years 1960 and 2005. Next, we conduct separate multiple-group ITSAs to compare the mortality rates in Russia to those of three groups of countries that would be expected to have a change in

Author Manuscript

This article is protected by copyright. All rights reserved.

mortality rates proportionate with their exposure to the anti-alcohol campaign and ethnic/religious composition. The first group consists of former Baltic/Western Soviet states that were exposed to the campaign and have a small share of Muslims (Latvia, Lithuania, Estonia, Ukraine, Belarus, and Moldova). The second group consists of former Soviet States that were exposed to the campaign and have a large share of Muslims (Armenia, Azerbaijan, Georgia, Uzbekistan, Kazakhstan, Krygyzstan, and Turkmenistan), and the third group consists of non-Soviet Eastern European countries that were not exposed to the anti-alcohol campaign at all, but did undergo a similar political and economic transition (the Czech Republic, the Slovak Republic, Hungary, and Poland) [27]. Considering the ethnic composition is important in this analysis because Islam prohibits the use of intoxicants, and thus the anti-alcohol campaign would not be expected to significantly reduce mortality in countries with large Muslim concentrations [27].

In all analyses, the pre-intervention phase (i.e. pre anti-alcohol campaign) spans from 1960 to 1984; the intervention phase spans from 1985 to 1990; and the post-intervention phase spans from 1991 to 2005 (3 distinct phases). As in the previous example, all statistical analyses were conducted using `ITSA` [9,10], and for comparisons between groups, the time series were ipsatively standardized to allow for comparisons between groups on the same scale.

As illustrated in Figure 2, deaths per thousand increased linearly in Russia from 1960 to 1985. There was a sharp drop in deaths immediately following the introduction of the anti-alcohol campaign in 1986, and the level appears attenuated until 1991, at which point annual

Author Manuscript

This article is protected by copyright. All rights reserved.

deaths returned to the same level and trend of the time series prior to the anti-alcohol campaign. The results from this single-group ITSA indicate that overall mortality rates in Russia decreased during the period of the anti-alcohol campaign. However, this design does not provide an unambiguous answer as to whether the decrease in mortality rates was causally associated with the campaign, and whether the sharp rise in mortality rates starting in 1991 was causally associated with the withdrawal of the campaign.

Figure 2B shows the comparison of annual mortality rates in Russia versus other former Soviet states that were also exposed to the campaign, and have low Muslim concentrations. As expected, these states exhibited a trend in overall mortality similar to that of Russia, thereby confirming that the anti-alcohol campaign was equally effective in other states exposed to intervention with a similar demographic composition.

Figure 2C shows the comparison of annual mortality rates in Russia versus other non-Soviet Eastern European countries that were not exposed to the anti-alcohol campaign but did undergo a political and economic transition as well. As illustrated, mortality rates in non-Soviet countries rose over time prior to the campaign (between 1960 and 1985), then flattened during the period of the campaign, but in contrast to Russia, declined after 1991. While this may call into question the causal relationship between the campaign and mortality (given that both Russia and non-Soviet Eastern bloc countries experienced a flat trend in mortality during the campaign years), it does suggest that withdrawal of the campaign was more causally associated with the

Author Manuscript

This article is protected by copyright. All rights reserved.

Russian Mortality Crisis than the political and economic transitions (given that the transition occurred across all these countries contemporaneously).

Figure 2D shows the comparison of annual mortality rates in Russia versus those of former Soviet states with higher concentrations of Muslims. As hypothesized, the trend in mortality rates of the Central Asia/Caucasus countries followed a completely different pattern than that of Russia. In these former Soviet states, mortality rates declined from 1960 to 1990 and then appeared to stabilize over the remaining duration of the study. If the Russian Mortality Crisis was entirely due to the political and economic transition and not withdrawal of the anti-alcohol campaign, we would expect to see a comparable increase in mortality in this set of former Soviet states that also experienced the transition but were not sensitive to anti-alcohol policy.

In summary, this example again highlights the effect of *history* in complicating the interpretation of ITSA results -- even when utilizing a cross-over design. Only when contrasting the results against those of a control group (or in the case, against different sets of control groups) was the treatment effect of Gorbachev's anti-alcohol campaign substantiated, and relatedly, the withdrawal of the campaign found to be the more likely culprit in explaining the Russian Mortality Crisis than Russia's transition to capitalism and democracy.

5. DISCUSSION

The two examples presented in this paper suggest that the single-group ITSA crossover design is just as vulnerable to the effects of *history* as the basic single-group ITSA design [Linden itsa

Author Manuscript

This article is protected by copyright. All rights reserved.

bias]. In the first example, a seemingly unquestionable treatment effect (and withdrawal effect) across multiple crossovers was reversed when contrasted with a comparable control group. In the second example, a negative change in the time series after withdrawal of the intervention was attributed to an external event, and only when compared to a control group did the withdrawal of the intervention receive correct attribution for that effect. In short, even with the addition of one or more crossovers, a single-group ITSA remains susceptible to threats to validity that limit the ability to draw causal inferences about the effects of the intervention.

As demonstrated in the present examples (in addition to those presented in Linden [6]), using a control group to serve as the counterfactual is the most robust approach for assessing treatment effects. Only when contrasted with a comparable control group can the effect of the intervention (or withdrawal of the intervention) (i.e., the *history* threat) be isolated from other rival factors. Moreover, other anomalies observed in the time series (e.g., changes in instrumentation, selection bias) can alert the investigator to other potential sources of confounding.

When multiple non-treated units are available, investigators can choose from at least three different matching methods suitable for time-series data. This includes the matching process implemented in the present examples (i.e., finding non-treated units that are non-statistically different from the treated unit on pre-intervention levels and trend of the outcome variable), a synthetic controls approach [28] or propensity score-based weighting [10] (which can also be extended to situations in which multiple treated units are available [29] and for

Author Manuscript

This article is protected by copyright. All rights reserved.

censored data [30,31]. Investigators should also consider the use of an instrumental variables strategy in cases where some of the right-hand-side covariates are endogenous [32,33,34]. Most statistical software packages have commands designed to implement these approaches (such as `XTIVREG` in Stata).

Finally, while this paper has illustrated that the crossover design does not ensure improved validity when implemented in a single-group study, the crossover design can further enhance the robustness of an ITSA study that includes a comparable control group to serve as the counterfactual. In such a study, the groups switch their treatment assignment at a given time-point (i.e. the treatment group switches to control and the control switches to treatment), and the outcomes change in accordance with the exposure to the intervention. Although clearly difficult to implement in practice, the design when properly executed, may possibly be considered as good as randomized (see Barlow et al [35] and Biglan et al [36] for other ITSA design alternatives to improve causal inference over the basic single-group design).

In summary, this paper illustrated that *history* -- the foremost threat to validity in the basic single-group ITSA design – persists even when adding one or more treatment crossovers to the study. Absent a comparable control group as a contrast, there is simply no assurance that the effect of external factors have been identified and controlled for, regardless of whether the time series follow the expected pattern after each crossover. Thus, even when using a single-group ITSA crossover design, the results should be considered preliminary -- and interpreted with caution -- until a more robust study design can be implemented. Given the popularity and

Author Manuscript

This article is protected by copyright. All rights reserved.

widespread use of single-group ITSA designs, it is important for investigators to be cognizant of their limitations, and to strive to add a comparable control group to maximize validity and improve causal inference.

Author Manuscript

This article is protected by copyright. All rights reserved.

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. *Disease Management and Health Outcomes* 2007;15:7-12.
4. Linden A. Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology* 2013;13:1-7.
5. Linden A, Adams J, Roberts N. The generalizability of disease management program results: getting from here to there. *Managed Care Interface* 2004;17:38-45.
6. Linden A. Challenges to validity in single-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice DOI: 10.1111/jep.12638*
7. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974;66:688–701.
8. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association* 2005;100:322–331.
9. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. *The Stata Journal* 2015;15:480–500.

Author Manuscript

This article is protected by copyright. All rights reserved.

10. Linden A, Adams JL. Applying a propensity-score based weighting model to interrupted time series data: Improving causal inference in program evaluation. *Journal of Evaluation in Clinical Practice* 2011;17:1231–1238.

11. Linden A, Adams JL. Combining the regression-discontinuity design and propensity-score based weighting to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice* 2012;18(2):317-325.

12. Linden A, Yarnold PR. Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice Epub April 2016*

13. Linden A, Roberts N. Disease management interventions: What's in the black box? *Disease Management* 2004;7(4):275-291.

14. Linden A, Butterworth S, Roberts N. Disease management interventions II: What else is in the black box? *Disease Management* 2006;9(2):73-85.

15. Box GEP, Tiao GC. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* 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. (1993) Interrupted time-series analysis with brief single-subject data. *Journal of Consulting and Clinical Psychology* 1993;61:966-974.

Author Manuscript

This article is protected by copyright. All rights reserved.

19. Gottman JM. *Time-series analysis. A Comprehensive Introduction for Social Scientists*. New York: Cambridge University Press; 1981.
20. McKnight S, McKean JW, Huitema BE. A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, 2000;5:87-101.
21. Simonton DK. Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychological Bulletin* 1977;84:489-502.
22. Velicer WF, McDonald RP. (1991) Cross-sectional time series designs: A general transformation approach. *Multivariate Behavioral Research* 1991;26:247-254.
23. Chaudhary GH, Neil K, Solomon MG, Preusser DF, Cosgrove LA. *Evaluation of the Reinstatement of the Helmet Law in Louisiana*. National Highway Traffic Safety Administration. 2008. Available at: www.nhtsa.gov/DOT/NHTSA/Traffic%20Injury%20Control/Articles/Associated%20Files/810956.pdf Accessed September 7, 2016
24. 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 September 7, 2016.
25. U.S. Department of Transportation, Federal Highway Administration. Highway statistics (multiple years). Available at: http://www.fhwa.dot.gov/policyinformation/statistics.cfm Accessed on September 7, 2016.

Author Manuscript

This article is protected by copyright. All rights reserved.

26. Yarnold PR, Linden A. Using machine learning to model dose-response relationships via ODA: eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis* 2016;5:41-52.
27. Bhattacharya J, Gathmann C, Miller G. The Gorbachev anti-alcohol campaign and Russia's mortality crisis. *American Economic Journal: Applied Economics* 2013;5:232-60.
28. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 2010;105:493–505.
29. Linden A, Adams, JL. Evaluating health management programmes over time. Application of propensity score-based weighting to longitudinal data. *Journal of Evaluation in Clinical Practice 2010;16:*180-185.
30. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11: 550–60.
31. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: An introduction to survival analysis. *Disease Management* 2004;7:180-190.
32. Linden A, Adams J. Evaluating disease management program effectiveness: an introduction to instrumental variables. *Journal of Evaluation in Clinical Practice* 2006;12(2):148-154.
33. Baltagi BH. *Econometric Analysis of Panel Data*. 4th ed. New York: Wiley; 2008.
34. Becketti S. *Introduction to Time Series Using Stata*. College Station, TX: Stata Press; 2013.

Author Manuscript

This article is protected by copyright. All rights reserved.

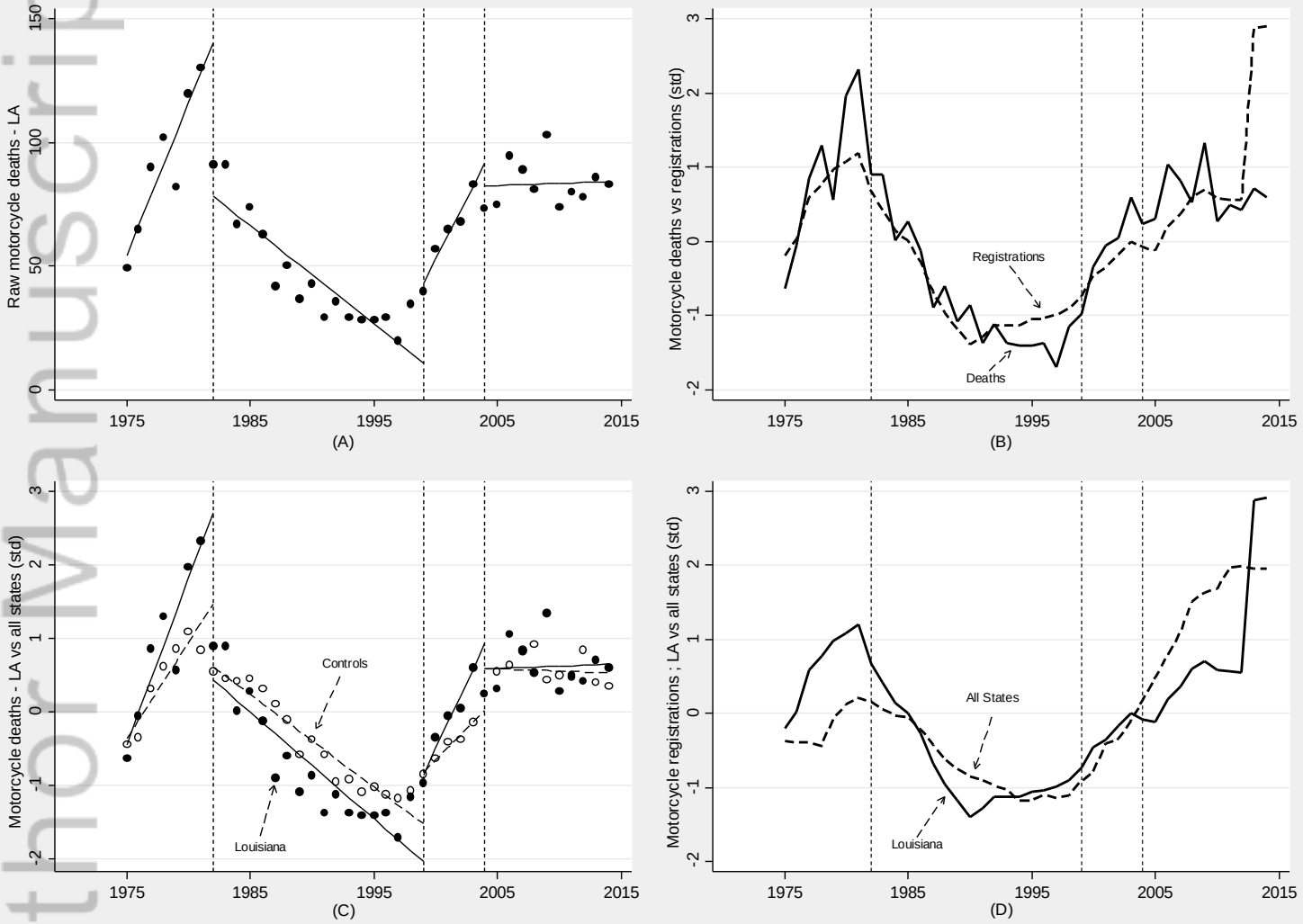
35. Barlow DH, Hayes SC, Nelson RO. *The Scientist Practitioner: Research and Accountability in Clinical and Educational Settings*. New York: Pergamon Press; 1984.

36. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prevention Science* 2000;1: 31–49.

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript



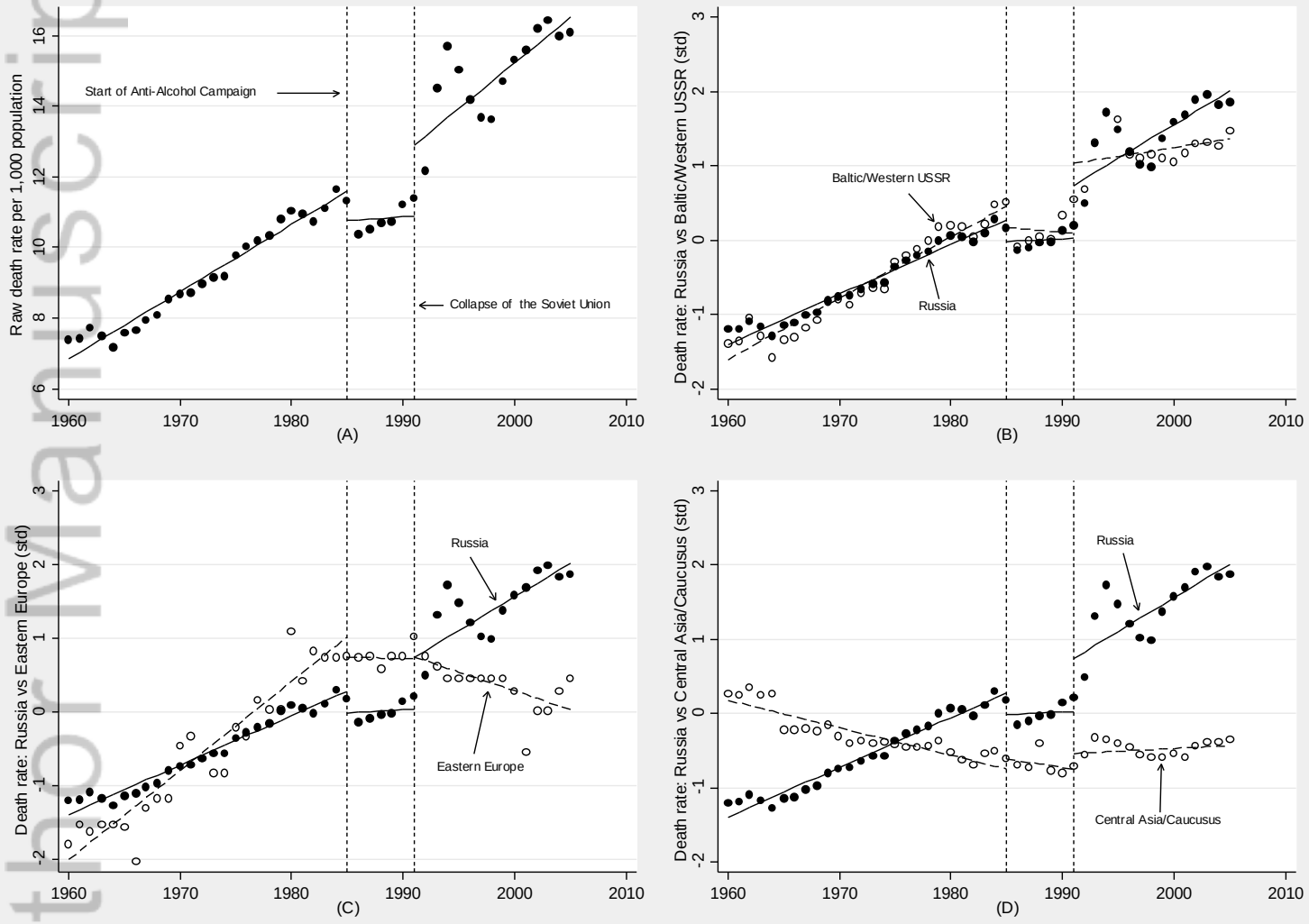
This article is protected by copyright. All rights reserved.

Figure 1. Louisiana motorcycle deaths and registrations from 1975 to 2014, using single-group and multiple-group ITSA designs

Author Manuscript

This article is protected by copyright. All rights reserved.

Author Manuscript



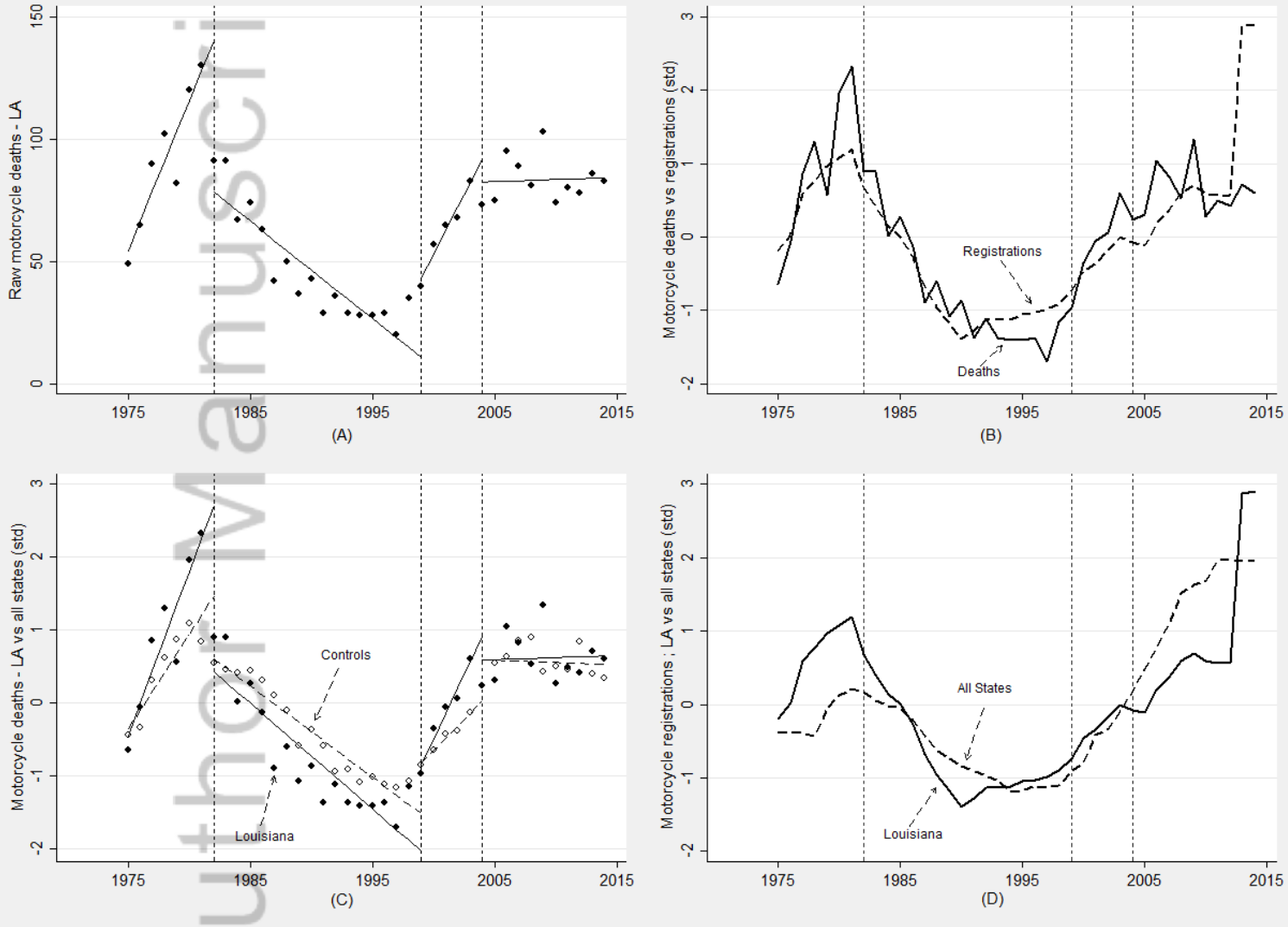
This article is protected by copyright. All rights reserved.

Figure 2. Raw death rates per 1,000 population in Russia and other former Soviet and non-Soviet states from 1960-2005, using single-group and multiple-group ITSA designs

Author Manuscript

This article is protected by copyright. All rights reserved.

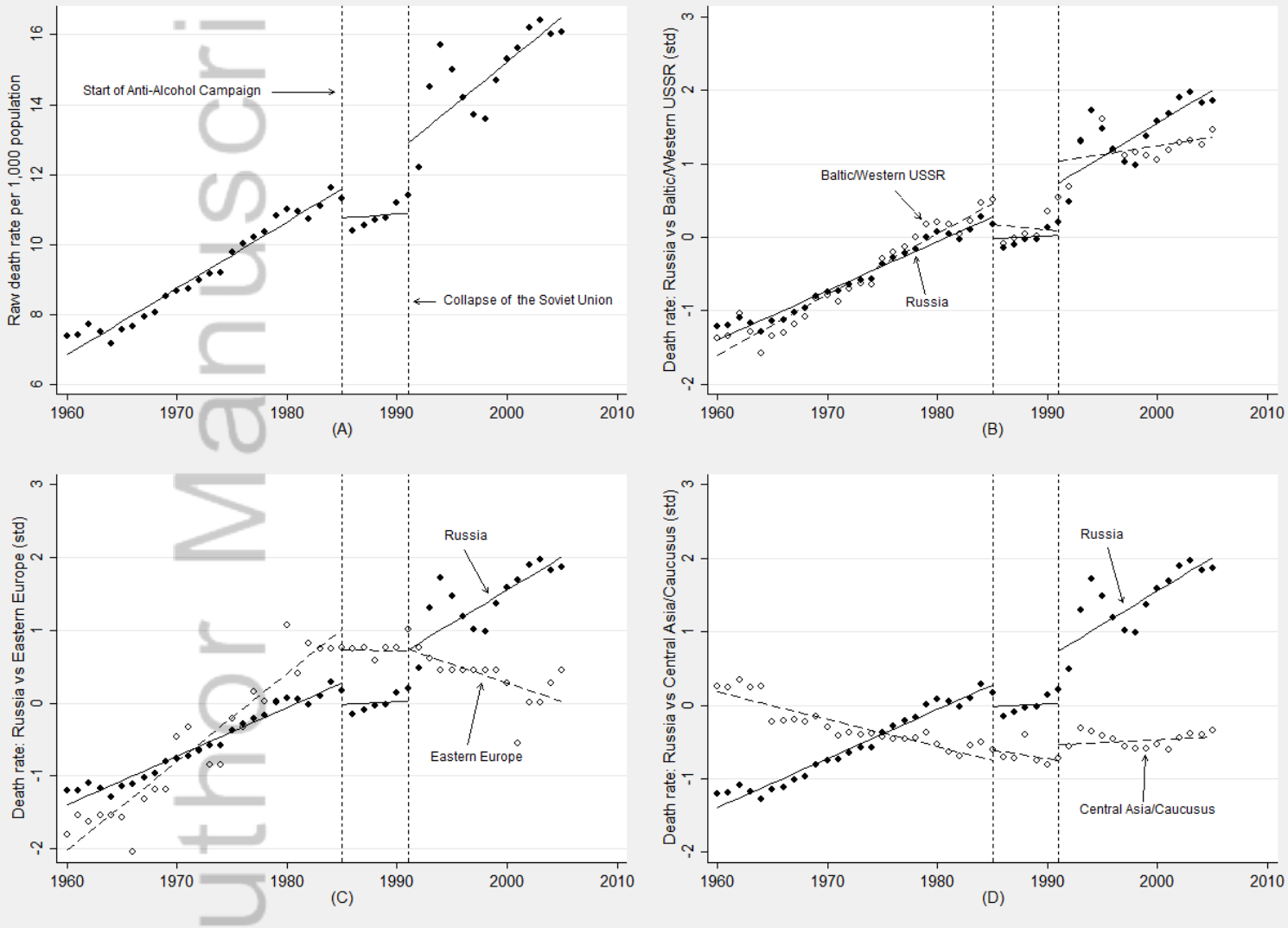
Author Manuscript



JEP_12668_F1.tif

This article is protected by copyright. All rights reserved.

Author Manuscript



JEP_12668_F2.tif

This article is protected by copyright. All rights reserved.