# Conducting interrupted time series analysis for single and multiple group comparisons

**Article** *in* Stata Journal · June 2015

CITATIONS
17

READS
2,568

**1 author:**

Ariel Linden
University of Michigan
**104** PUBLICATIONS **1,299** CITATIONS

SEE PROFILE

The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–20

# Conducting interrupted time series analysis for single and multiple group comparisons

Ariel Linden
Linden Consulting Group, LLC
Ann Arbor, MI
alinden@lindenconsulting.org

**Abstract.** This article introduces the `itsa` command, which performs interrupted time series analysis for single and multiple group comparisons. In an interrupted time series analysis, an outcome variable is observed over multiple, equally spaced time periods before and after the introduction of an intervention which is expected to *interrupt* its level and/or trend. The `itsa` command estimates the effect of an intervention on an outcome variable for either a single treatment group or when compared with one or more control groups. Additionally, its options allow the user to control for autocorrelated disturbances and to estimate treatment effects over multiple periods.

**Keywords:** st0001, itsa, interrupted time series, quasi-experimental designs, casual inference

## 1 Introduction

In considering the impact of large scale interventions (e.g. population-based health interventions, media campaigns, dissemination of professional guidelines) or public policy changes (e.g. new laws or taxes), researchers are often faced with an effective sample size of N=1, where the treated group may be the local community, state, or an even larger unit. It is also fairly common in these situations that the only data available are reported at an aggregate level (e.g. morbidity or mortality rates, average costs, median incomes, etc.). If multiple observations on an outcome variable of interest in the pre- and post-intervention periods can be obtained, an interrupted time series analysis (ITSA) offers a quasi-experimental research design with a potentially high degree of internal validity (Campbell and Stanley [1966], Shaddish et al. [2002]). Naturally, when the treated groups outcomes can also be contrasted with those of one or more comparison groups, the internal validity is further enhanced by allowing the researcher to potentially control for confounding omitted variables.

Interrupted time series analysis has been used in many areas of study, such as assessing the effects of community interventions (Biglan et al. [2000], Gillings et al. [1981]), public policy (Muller [2004]), regulatory actions (Briesacher et al. [2013]), and health technology assessment (Ramsay et al. [2003]), to name but a few. ITSA has also been proposed as a more flexible and rapid design to be considered in health research before defaulting to the traditional two-arm randomized controlled trial (Riley et al. [2013]), and systematic reviews of the literature are increasingly including studies that have used

 

ITSA as their primary research design (Cochrane Effective Practice and Organisation of Care (EPOC) [2013]).

This article introduces the new `itsa` command, which performs interrupted time series analysis using two OLS regression-based approaches available in the official Stata package, `newey` and `prais`. Additionally, `itsa` can estimate treatment effects for multiple treatment periods.

# 2 Method and Formulas

Statistical analyses used for ITSA must account for autocorrelated data. The two general approaches historically used in ITSA are autoregressive integrated moving-average (ARIMA) models (see Box and Tiao [1975], Glass et al. [1975], and McDowall et al. [1980]) and ordinary least-squares (OLS) regression models designed to adjust for autocorrelation (see, among others, Crosbie [1993];Gottman [1981]; McKnight et al. [2000] Simonton [1977a] and Velicer and McDonald [1991]). `itsa` relies on OLS rather than ARIMA-based regression methods because the former is often more flexible and broadly applicable in an interrupted time series context (Box and Jenkins [1976], Velicer and Harrop [1983]).

## 2.1 The single-group analysis

When there is only one group under study (no comparison groups) the standard ITSA regression model assumes the following form (Huitema and McKean [2000a], Linden and Adams [2011], Simonton [1977a], Simonton [1977b]):

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 X_t T_t + \epsilon_t \tag{1}$$

$Y_t$ is the aggregated outcome variable measured at each equally-spaced time-point $_t$, $T_t$ is the time since the start of the study, $X_t$ is a dummy (indicator) variable representing the intervention (pre-intervention periods 0, otherwise 1), and $X_t T_t$ is an interaction term. These terms are displayed in the lower half of Figure 1. In the case of a single group study, $\beta_0$ represents the intercept, or starting level of the outcome variable. $\beta_1$ is the slope, or trajectory of the outcome variable until the introduction of the intervention. $\beta_2$ represents the change in the level of the outcome that occurs in the period immediately following the introduction of the intervention (compared to the counterfactual). $\beta_3$ represents the difference between pre- and post-intervention slopes of the outcome. Thus, we look for significant $P$-values in $\beta_2$ to indicate an immediate treatment effect, or in $\beta_3$ to indicate a treatment effect over time (Linden and Adams [2011]).

When the random error terms follow a first-order autoregressive (AR1) process:

$$\epsilon_t = \rho \epsilon_{t-1} + u_t \tag{2}$$

where, the autocorrelation parameter $\rho$ is the correlation coefficient between adjacent

error terms, such that $|\rho| < 1$, and the disturbances $u_t$ are independent $N(0,\sigma^2)$ (see Kutner et al. [2005] for a detailed discussion of autocorrelation in time-series models).

Identification in both the single and multiple group models is driven by the functional form assumptions of the ITSA model. By design, a single-group ITSA has no comparable control group; rather, the pre-intervention trend projected into the treatment period serves as the counterfactual. We assume that any time varying unmeasured confounder is relatively slowly changing so that it would be distinguishable from the sharp jump of the intervention indicator. This underscores the need for caution with these methods if there are multiple policy shifts in the time window around the implementation of the intervention.

The assumptions necessary for causal inference in the single-group ITSA may seem plausible when the pre-intervention trend is flat followed by a significant change in the outcome variable immediately following the introduction of the intervention and then sustained over time. However, these assumptions may seem less plausible if there already is an existing trend in the time series prior to the intervention. While the ITSA literature does not address the topic of testing for interruptions in the level and trend of the outcome variable ($\beta_2$ and $\beta_3$ of Equation 1) prior to the actual period in which the intervention started, we can look to the regression-discontinuity literature to provide guidance for applicable robustness tests. In practice, this would simply entail testing for interruptions after replacing the true intervention start period with other pseudo-start periods along the pre-intervention continuum. In an adaptation of Imbens and Lemieux (2008) for ITSA, an investigator could use the median time point of the pre-intervention period to test for an interruption. In a sufficiently long time series, the median time-point of the pre-intervention period is a good choice of a pseudo-start period to maximize power to detect a significant jump (as the subsample will be evenly split on both sides). For shorter time series, a simple iterative process of testing each pre-intervention time period as the pseudo-start period may be a good approach. In using such robustness tests, the underlying assumptions of the single-group ITSA may be challenged if interruptions in the level and/or trend of the outcome variable are found to exist at other time-points prior to the true initiation of the intervention.

## 2.2 The multiple-group analysis

When one or more control groups are available for comparison, the regression model in Equation (1) is expanded to include four additional terms ($\beta_4$ to $\beta_7$) ( Linden and Adams [2011], Simonton [1977a], Simonton [1977b]):

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 X_t T_t + \beta_4 Z + \beta_5 Z T_t + \beta_6 Z X_t + \beta_7 Z X_t T_t + \epsilon_t \qquad (3)$$

Here $Z$ is a dummy variable to denote the cohort assignment (treatment or control) and $ZT_t$, $ZX_t$ and $ZX_tT_t$ are all interaction terms among previously described variables. Now, when examining Figure 1, the coefficients of the lower line $\beta_0$ to $\beta_3$ represent the control group and the coefficients of the upper line $\beta_4$ to $\beta_7$ represent values of the treatment group. More specifically, $\beta_4$ represents the difference in the level (intercept) of the outcome variable between treatment and controls prior to the intervention, $\beta_5$

represents the difference in the slope (trend) of the outcome variable between treatment and controls prior to the intervention, $\beta_6$ indicates the difference between treatment and control groups in the level of the outcome variable immediately following introduction of the intervention, and $\beta_7$ represents the difference between treatment and control groups in the slope (trend) of the outcome variable after initiation of the intervention compared to pre-intervention (akin to a difference-in-differences of slopes). If the multiple-group model follows a first-order autoregressive process, the random error term is defined as in Equation (2).

A multiple-group ITSA may be particularly valuable when there is an exogenous policy shift that affects all of the groups. The key assumption is that the change in the level and/or trend in the outcome variable is presumed to be the same for both the control group and, counterfactually, for the treatment group had it not received the intervention. In other words, we assume that confounding omitted variables affect both treatment and control groups similarly. A major strength of the multiple-group ITSA is the ability to test for comparability between groups on observed covariates and in particular, the two parameters $\beta_4$ and $\beta_5$, which play a particularly important role in establishing whether the treatment and control groups are balanced on both the level and the trajectory of the outcome variable in the pre-intervention period. If these data were from a randomized controlled trial, we would expect similar levels and slopes prior to the intervention. However, in an observational study where equivalence between groups cannot be assured, any observed differences will likely raise concerns about the ability to draw causal inferences about the relationship between the intervention and the outcomes (Linden and Adams [2011]).

To reduce the threat of confounding, investigators may attempt to emulate the randomization process with observational data by finding control groups that are comparable to the treatment group on observed pre-intervention covariates. One approach to finding comparable controls out of a pool of potential candidates is via an iterative process in which each non-treated group is compared separately to the treatment group using the model defined in Equation 3. Those groups who have *P*-values greater than 0.05 (or a higher threshold) on both $\beta_4$ and $\beta_5$ can be selected as controls for inclusion in the final model. This method can be easily expanded to other available covariates, however there is a diminished likelihood of finding good controls as the number of covariates is increased. If achieving balance on many covariates is an important factor, two alternative approaches to `itsa` should be considered: the synthetic controls approach described by Abadie et al. (2010) and implemented in Stata using the `synth` package (Abadie et al. [2014]), or the propensity-score weighting technique described by Linden and Adams (2011).

## 2.3 Data variables corresponding to model parameters

Table 1 displays the variables used in both regression models (1) and (3), using an artificial example with one intervention period. There are two individuals (or groups) in these data (`ID` = 1, 2) with 6 observations each (`T`). `X` indicates that there are 2 pre-intervention observations, followed by 4 observations in the intervention period (the

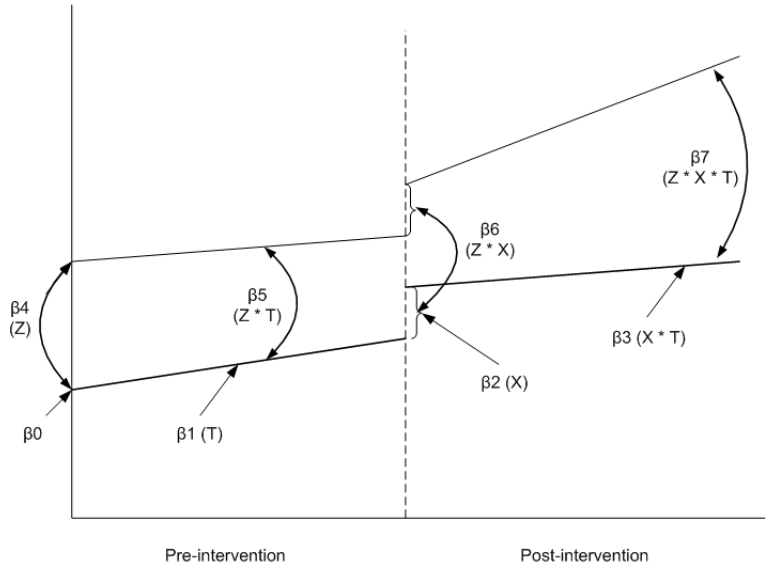


Figure 1: Visual depiction of a single-group (lower line) and multiple-group (upper and lower lines) interrupted time series design, from Linden and Adams (2011)

Legend:
*Single-group* - $\beta_0$: intercept; $\beta_1$: slope prior to intervention; $\beta_2$: change in level in the period immediately following intervention initiation (compared to counterfactual); $\beta_3$: difference between pre- and post-intervention slopes

*Multiple-group* - $\beta_0$ to $\beta_3$ represent the control group; $\beta_4$ to $\beta_7$ represent the treatment group; $\beta_4$: difference in the level between treatment and control prior to intervention; $\beta_5$: difference in the slope between treatment and control prior to intervention; $\beta_6$: difference in the level between treatment and control in the period immediately following intervention initiation; $\beta_7$: difference between treatment and control in the slope after initiation of the intervention compared to pre-intervention.

intervention commences when `T`=3). `XT` is an interaction term of `X`*`T`, which starts in the observation period immediately following the start of the intervention (`T`=4) and runs sequentially until the last observation when `T`=6 (see Huitema and McKean [2000a] for an exposition on the appropriateness of commencing the sequence in the observation period *after* the start of the intervention). Here we transform `XT` = (`T` - 3) * `X`, so that it runs sequentially starting at 1. Additional variables are required for a multiple-group analysis. `Z` indicates the treatment status, where `Z`=1 for the treatment group and `Z`=0 for the control group. `ZT`, `ZX`, and `ZXT` are additional interaction terms used in multiple-group comparisons, as described above in `Section 2.2`. When multiple treatment periods are specified, additional variables are added to the dataset, corresponding to

each treatment period respectively (see `Section 4.3` for an example).

Upon running the `itsa` command, all variables required for the corresponding single or multiple-group model are automatically generated and added to the dataset.

Table 1: Covariates used in a single-group ITSA (T, X, XT) and multiple-group ITSA (T, X, XT, Z, ZX, ZXT) corresponding to regression models (1) and (3), respectively.

| ID | T | X | XT | Z | ZT | ZX | ZXT |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 2 | 0 | 0 | 1 | 2 | 0 | 0 |
| 1 | 3 | 1 | 0 | 1 | 3 | 1 | 0 |
| 1 | 4 | 1 | 1 | 1 | 4 | 1 | 1 |
| 1 | 5 | 1 | 2 | 1 | 5 | 1 | 2 |
| 1 | 6 | 1 | 3 | 1 | 6 | 1 | 3 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 4 | 1 | 1 | 0 | 0 | 0 | 0 |
| 2 | 5 | 1 | 2 | 0 | 0 | 0 | 0 |
| 2 | 6 | 1 | 3 | 0 | 0 | 0 | 0 |

## 2.4 Models

`itsa` allows the user to choose between two OLS regression-based models specifically designed for time series data. The first, `newey`, estimates the coefficients by OLS regression but produces Newey-West standard errors to handle autocorrelation in addition to possible heteroskedasticity (see [TS] **newey**). The second model, `prais`, uses the generalized least-squares method to estimate the parameters in a linear regression model in which the errors are assumed to follow a first-order autoregressive (AR1) process. More specifically, `prais` offers several methods to transform the original observations based on the pooled autocorrelation estimate $p$ in order to remove the correlation between first-order errors (i.e., the correlation between the errors of each observation period and those of the preceding observation period) (see [TS] **prais**).

The type of model that an investigator will choose for conducting time series analysis will likely depend on a combination of factors, with primary attention on the number of lags in the data for which autocorrelation is present. In general, the investigator first fits an OLS model using either `regress` or `newey` (with lag(0) specified) and then tests

for autocorrelation in the error distribution. It is important to test for the presence of autocorrelated errors when using regression-based time series methods, as such tests provide critical diagnostic information regarding the adequacy of the time series model (i.e., whether tests and confidence intervals on the regression coefficients are satisfactory, whether important variables have been left out of the time series regression model and because autocorrelated errors are produced when the functional form of the variables included in the model is incorrect; Huitema and McKean [2000b]). The official Stata package offers several post-estimation commands for this purpose (see [R] **regress postestimation time series**). In addition, there is a comprehensive and versatile user-written program `actest` (Baum and Schaffer [2013]), which is downloadable from SSC, with the default being the Cumby-Huizinga general test for autocorrelation (Cumby and Huizinga [1992]).

# 3 The `itsa` command

## 3.1 Syntax

`itsa` *depvar* [ *indepvars* ] [ *if* ] [ *in* ] [ *weight* ] , *trperiod(numlist)* [ `single` `treatid(#)` `contid(`*numlist*`)` `prais` `lag(#)` `figure` `posttrend` `replace` `prefix(`*string*`)` *model_options* ]

A dataset for a single panel must be declared to be time-series data using `tsset` *timevar*. When the dataset contains multiple panels, a strongly balanced panel dataset using `tsset` *panelvar timevar* must be declared; see [TS] **tsset**. *indepvars* may contain factor variables; see [U] **11.4.3 Factor variables**. *depvar* and *indepvars* may contain time-series operators; see [U] **11.4.4 Time-series varlists**. `aweights` are allowed; see [U] **11.1.6 weight**. See [TS] **newey postestimation** and [TS] **prais postestimation** for features available after estimation.

## 3.2 Options

`trperiod(`*numlist*`)` specifies the time period when the intervention begins. The values entered for time period must be in the same units as the panel time variable specified in `tsset` *timevar*; see [TS] **tsset**. More than one period may be specified. `trperiod()` is required.

`single` indicates that `itsa` will be used for a single group analysis. Conversely, omitting `single` indicates that `itsa` is for a multiple group comparison.

`treatid(#)` specifies the identifier of the single treated unit under study when the dataset contails multiple panels. The value entered must be in the same units as the panel variable specified in `tsset` *panelvar timevar*; see [TS] **tsset**. When the dataset contains data for only a single panel, `treatid()` must be omitted.

`contid(`*numlist*`)` specifies a list of identifiers to be used as control units in the multiple group analysis. The values entered must be in the same units as the panel variable specified in `tsset` *panelvar timevar*; see [TS] **tsset**. If `contid()` is not specified, all non-treated units in the data will be used as controls.

`prais` specifies that a `prais` model should be estimated. If `prais` is not specified, `itsa` will use `newey` as the default model.

`lag(#)` specifies the maximum lag to be considered in the autocorrelation structure when a `newey` model is chosen. If the user specifies `lag(0)`, the output is the same as `regress, vce(robust)`; Default is `lag(0)`. An error message will appear if both `prais` and `lag()` are specified, as `prais` implements an AR(1) model, by design.

`figure` produces a line plot of the predicted *depvar* variable combined with a scatter plot of the actual values of *depvar* over time. In a multiple group analysis, `figure` plots the average values of all controls used in the analysis (more specifically, data for specified controls are collapsed and the monthly observations averaged).

`posttrend` produces post-treatment trend estimates using `lincom`, for the specified model. In the case of a single-group ITSA, one estimate is produced. In the case of a multiple-group ITSA, an estimate is produced for the treatment group, the control group, and the difference. In the case of multiple treatment periods, a separate table is produced for each treatment period.

`replace` replaces variables created by `itsa` if they already exist. If `prefix()` is specified, only variables created by `itsa` with the same prefix will be replaced.

`prefix(`*string*`)` adds a prefix to the names of variables created by `itsa`. Short prefixes are recommended.

*model_options* specifies all available options for `prais` when the `prais` option is chosen; otherwise all available options of `newey` other than `lag()`.

## 3.3 Saved results

As `itsa` passes all user-entered information to `prais` and `newey`, all results saved by those commands are available. Additionally, `itsa` generates several key time series variables and adds them to the current dataset, as described in `Section 2.3`. These additional variables allow the user to further estimate treatment effects using `arima` or other time-series models.

Below is a cross reference to default names for those variables that appear in the regression output tables (and used when `posttrend` is specified). Variables starting with `_z` are added to the dataset only when a multiple-group comparison is specified. (`trperiod`) is a suffix added to certain variables indicating the start of the intervention period. This is particularly helpful for differentiating between added variables when multiple interventions are specified (see the example presented in `Section 4.3`). If the user specifies a `prefix()`, it will be applied to all variables generated by `itsa`:

_t is the time since start of study

_x(trperiod) is a dummy variable representing the intervention periods (pre-intervention periods 0, otherwise 1)

_x_t(trperiod) is an interaction of _x and _t

_z is a dummy variable to denote the cohort assignment (treatment or control)

_z_x(trperiod) is an interaction of _z and _x

_z_x_t(trperiod) is an interaction of _z, _x, and _t

_s_*depvar*_pred is the predicted value generated after running itsa for a single group

_m_*depvar*_pred is the predicted value generated after running itsa for a multiple-group comparison

# 4 Examples

In 1988, California passed the voter-initiative Proposition 99 which was a wide-spread effort to reduce smoking rates by raising the cigarette excise tax by 25 cents per pack and fund anti-smoking campaigns and other related activities throughout the state (for a comprehensive discussion of this initiative see Abadie et al. [2010]). Per-capita cigarette sales (in packs) is the most widely used indicator of smoking prevalence found in the tobacco research literature (Abadie et al. [2010]) and serves here as the aggregate outcome variable under study, measured at the state level from 1970 until 2000 (with 1989 representing the first year of the intervention). The current data file was obtained from the synth package (Abadie et al. [2014]) which originally obtained the cigarette sales data and average retail price of cigarettes from Orzechowski and Walker (2005). 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. [2010]).

## 4.1 Single-group interrupted time-series analysis

In this example, we use itsa to assess the impact of Proposition 99 in reducing California's per-capita cigarette sales (in packs), using a single-group design. More specifically, we assess whether the introduction of Proposition 99 resulted in a shift in the level and trend of per-capita cigarette sales compared with those of the pre-intervention period (as described in Section 2.1).

First, we load the data and declare the dataset as panel:

(*Continued on next page*)

```
. use cigsales, clear
. tsset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 1970 to 2000
                delta:  1 unit
```

Next, we specify a single-group ITSA with State 3 (California) as the treatment group, and start of the intervention is 1989, request post-intervention trend estimates and plot the results. The model is estimated using `newey` with 1 lag:

```
. itsa cigsale, single treat(3) trperiod(1989) lag(1) posttrend fig
       panel variable:  state (strongly balanced)
        time variable:  year, 1970 to 2000
                delta:  1 unit
Regression with Newey-West standard errors          Number of obs  =        31
maximum lag: 1                                      F(  3,    27)  =    331.45
                                                    Prob > F       =    0.0000
```

| cigsale | Coef. | Newey-West Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -1.779474 | .3834188 | -4.64 | 0.000 | -2.566184 | -.9927632 |
| _x1989 | -20.0581 | 4.724395 | -4.25 | 0.000 | -29.75175 | -10.36444 |
| _x_t1989 | -1.494652 | .4368201 | -3.42 | 0.002 | -2.390933 | -.5983715 |
| _cons | 134.0053 | 4.600271 | 29.13 | 0.000 | 124.5663 | 143.4442 |

Post-Intervention Linear Trend: 1989

Treated: _b[_t]+_b[_x_t1989]

| Linear Trend | Coeff | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Treated | -3.2741 | 0.2688 | -12.1803 | 0.0000 | -3.8257 | -2.7226 |

As shown in the regression table, the starting level of the per-capita cigarette sales was estimated at 134 packs, and sales appeared to decrease significantly every year prior to 1989 by 1.78 packs ($P$ <0.0001, CI = -2.57, 0.99). In the first year of the intervention (1989) there appeared to be a significant decrease in per-capita cigarette sales of 20.06 packs ($P$ <0.0001, CI = -29.75, -10.36), followed by a significant decrease in the annual trend of sales (relative to the pre-intervention trend) of 1.49 packs per-capita per year ($P$ = 0.002, CI = -2.39, -0.60). We also see from the `lincom` estimate produced by specifying `posttrend`, that after the introduction of Proposition 99, per-capita cigarette sales deceased annually at a rate of 3.27 packs (95% CI -3.83, -2.72). Figure 2 provides a visual display of these results.

To ensure that we estimated a model that accounts for the correct autocorrelation structure, we use `actest` (Baum and Schaffer [2013]), to test for autocorrelation.

```
. actest , lags(6)
Cumby-Huizinga test for autocorrelation
  H0: variable is MA process up to order q
  HA: serial correlation present at specified lags >q
```

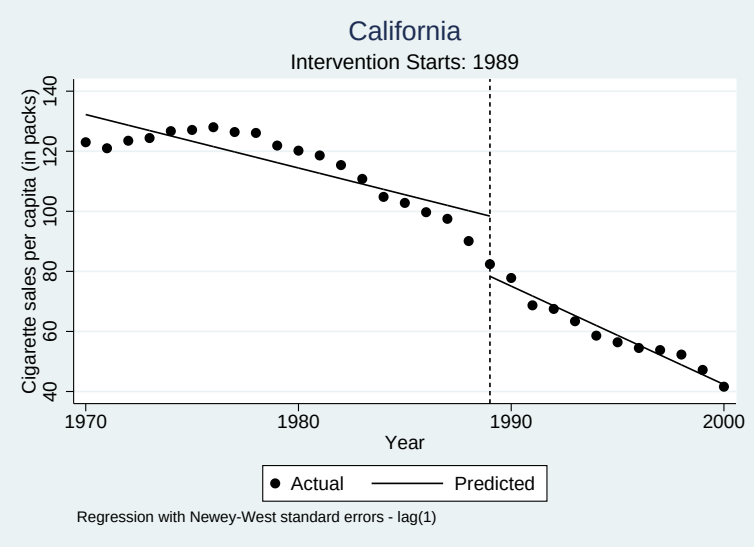


Figure 2: Single-group interrupted time series analysis with Newey-West standard errors and one lag

| H0: q=0 (serially uncorrelated) HA: s.c. present at range specified | | | | H0: q=specified lag-1 HA: s.c. present at lag specified | | | |
|---|---|---|---|---|---|---|---|
| lags | chi2 | df | p-val | lag | chi2 | df | p-val |
| 1 - 1 | 15.242 | 1 | 0.0001 | 1 | 15.242 | 1 | 0.0001 |
| 1 - 2 | 15.255 | 2 | 0.0005 | 2 | 3.300 | 1 | 0.0693 |
| 1 - 3 | 15.325 | 3 | 0.0016 | 3 | 1.192 | 1 | 0.2749 |
| 1 - 4 | 15.896 | 4 | 0.0032 | 4 | 0.000 | 1 | 0.9880 |
| 1 - 5 | 16.057 | 5 | 0.0067 | 5 | 1.113 | 1 | 0.2914 |
| 1 - 6 | 16.078 | 6 | 0.0133 | 6 | 2.051 | 1 | 0.1521 |

```
Test allows predetermined regressors/instruments
Test requires conditional homoskedasticity
```

As shown in the right-side panel of the output table, autocorrelation is present at lag 1, but at no higher lag orders (up to the 6 lags tested). Thus, our initial model specifying lag(1) should correctly account for this autocorrelation.

An alternative approach is to rerun `itsa` specifying the `prais` option, which is inherently designed to estimate an AR1 model. Here we specify the option `rhotype(tscorr)`, which bases $p$ on the autocorrelation of the residuals, and add robust standard errors.

(*Continued on next page*)

```
. itsa cigsale, single treat(3) trperiod(1989) replace prais rhotype(tscorr) //
  vce(robust)
      panel variable:  state (strongly balanced)
       time variable:  year, 1970 to 2000
               delta:  1 unit
```

(*output omitted*)

```
Prais-Winsten AR(1) regression -- iterated estimates
Linear regression                                 Number of obs =      31
                                                  F(  3,    27) =  609.24
                                                  Prob > F      =  0.0000
                                                  R-squared     =  0.9011
                                                  Root MSE      =  2.5964
```

| cigsale | Coef. | Semirobust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -1.843139 | .4538631 | -4.06 | 0.000 | -2.77439 | -.9118892 |
| _x1989 | -6.094491 | .8840197 | -6.89 | 0.000 | -7.90835 | -4.280633 |
| _x_t1989 | -1.998494 | .9191 | -2.17 | 0.039 | -3.884332 | -.1126568 |
| _cons | 128.1931 | 3.958813 | 32.38 | 0.000 | 120.0703 | 136.316 |
| rho | .9424635 | | | | | |

```
Durbin-Watson statistic (original)    0.535242
Durbin-Watson statistic (transformed) 1.342728
```

As the estimates produced using `prais` are transformed, they are not directly comparable to those of `newey` which are produced using an OLS model. However, these results confirm a significant decrease in the annual trend of sales (relative to the pre-intervention trend) of 2 packs per-capita per year ($P = 0.039$, CI = -3.88, -0.11). `prais` provides the Durbin-Watson $d$ statistic as an indicator of how well the model corrects for first-order autocorrelation. $d$ can take on values between 0 and 4 and under the null hypothesis, $d$ is equal to 2. Values of $d$ less than 2 suggest positive autocorrelation ($p>0$), whereas values of $d$ greater than 2 suggest negative autocorrelation ($p<0$); see [R] **regress postestimation time series**. As discussed previously, there are several more intuitive and flexible tests of autocorrelation, however none of them can currently be used in conjunction with `prais`.

## 4.2 Multiple-group interrupted time-series analysis

In this example, we use `itsa` to assess the impact of Proposition 99 in reducing California's per-capita cigarette sales (in packs), using a multiple-group design. More specifically, we now compare California's experience to that of the other 38 States in the data file.

(*Continued on next page*)

```
. itsa cigsale, treat(3) trperiod(1989) lag(1) replace fig
       panel variable:  state (strongly balanced)
        time variable:  year, 1970 to 2000
                delta:  1 unit
Regression with Newey-West standard errors          Number of obs  =      1209
maximum lag: 1                                      F(  7,  1201)  =    364.04
                                                    Prob > F       =    0.0000
```

| cigsale | Coef. | Newey-West Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -.5477701 | .2941289 | -1.86 | 0.063 | -1.124834 | .0292935 |
| _z | -2.041967 | 5.75639 | -0.35 | 0.723 | -13.33567 | 9.251731 |
| _z_t | -1.231704 | .4641182 | -2.65 | 0.008 | -2.142276 | -.321131 |
| _x1989 | -17.25168 | 3.815452 | -4.52 | 0.000 | -24.73737 | -9.765987 |
| _x_t1989 | -.5035089 | .5252893 | -0.96 | 0.338 | -1.534096 | .5270778 |
| _z_x1989 | -2.806417 | 5.841839 | -0.48 | 0.631 | -14.26776 | 8.654929 |
| _z_x_t1989 | -.9911435 | .6657528 | -1.49 | 0.137 | -2.297311 | .3150244 |
| _cons | 136.0472 | 3.818559 | 35.63 | 0.000 | 128.5554 | 143.539 |

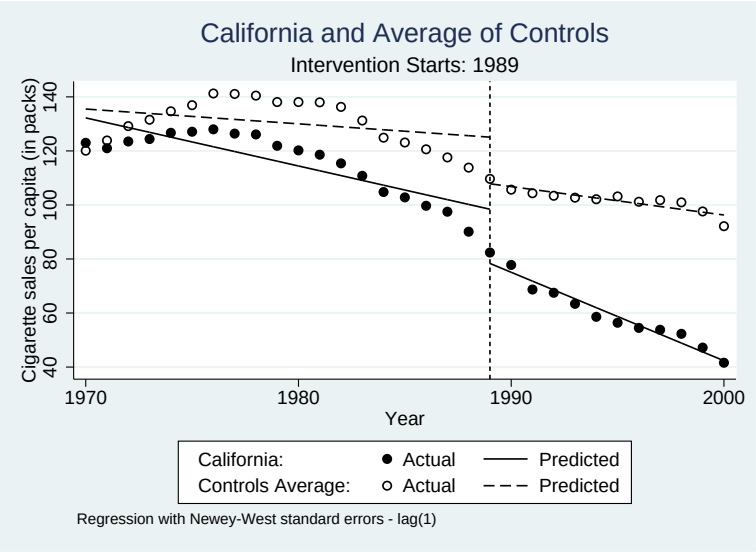


Figure 3: Multiple-group interrupted time series analysis with Newey-West standard errors and one lag. All 38 "non-treated" states are used for comparison.

As shown in the regression table, the initial mean level difference between California and the remaining states (_z) was not significant ($P = 0.723$, CI = -13.33, 9.25), but the difference in the mean baseline slope (_z_t) was significant ($P = 0.008$, CI = -2.14, -0.32). This is verified upon visual inspection of Figure 3, as the trajectory of mean cigarette sales for the 38 states appears to rise higher than in California and that level remains elevated throughout the duration of the observation period. Given this

differential pattern of change in the baseline, one could argue that the 38 other states were not comparable to California and thus treatment effect estimates for _z_x1989 and _z_x_t1989 may be biased (in the present case, both estimates are not-statistically significant). Therefore, this model could be improved by limiting the choice of control groups to only those with similar values on these two variables.

In the following example, we limit the analysis to only those states that are comparable to California on baseline level and trend of the outcome variable, as described in Section 2.2. Comparability in the current context is defined as a having a $P$-value greater than 0.10 on both _z and _z_t. Three comparison states meet this criteria: Colorado, Idaho and Montana.

```
. itsa cigsale, treat(3) trperiod(1989) contid(4 8 19) lag(1) replace posttrend fig
       panel variable:  state (strongly balanced)
        time variable:  year, 1970 to 2000
                delta:  1 unit
Regression with Newey-West standard errors          Number of obs  =       124
maximum lag: 1                                      F(  7,   116)  =    251.48
                                                    Prob > F       =    0.0000
```

| cigsale | Coef. | Newey-West Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -1.464503 | .3837773 | -3.82 | 0.000 | -2.224622 | -.7043836 |
| _z | 2.046198 | 6.218666 | 0.33 | 0.743 | -10.27065 | 14.36305 |
| _z_t | -.3149707 | .5330632 | -0.59 | 0.556 | -1.37077 | .7408282 |
| _x1989 | -13.58866 | 4.180499 | -3.25 | 0.002 | -21.86867 | -5.308658 |
| _x_t1989 | .4746428 | .4992501 | 0.95 | 0.344 | -.514185 | 1.463471 |
| _z_x1989 | -6.469433 | 6.185239 | -1.05 | 0.298 | -18.72008 | 5.781212 |
| _z_x_t1989 | -1.969295 | .6533782 | -3.01 | 0.003 | -3.263393 | -.6751973 |
| _cons | 131.9591 | 4.355318 | 30.30 | 0.000 | 123.3328 | 140.5853 |

```
                Comparison of Linear Post-Intervention Trends: 1989
Treated    : _b[_t] + _b[_z_t] + _b[_x_t1989] + _b[_z_x_t1989]
Controls   : _b[_t] + _b[_x_t1989]
Difference : _b[_z_t] +  _b[_z_x_t1989]
```

| Linear Trend | Coeff | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Treated | -3.2741 | 0.2594 | -12.6234 | 0.0000 | -3.7878 | -2.7604 |
| Controls | -0.9899 | 0.2883 | -3.4336 | 0.0008 | -1.5608 | -0.4189 |
| Difference | -2.2843 | 0.3878 | -5.8905 | 0.0000 | -3.0523 | -1.5162 |

As shown in both the regression table and verified upon visual inspection of Figure 4, the treatment group is comparable to controls on both baseline level and trend. While there is no statistically significant treatment effect during the first year of the intervention (_z_x1989), there is a statistically significant annual reduction in the pre-post trend compared to that of controls of 1.97 per-capita cigarette sales per year ($P$ = 0.003, CI = -3.26, -0.68). Additionally, we see from the posttrend output that the treatment group decreased annual cigarette sales in the post-intervention period by 3.27

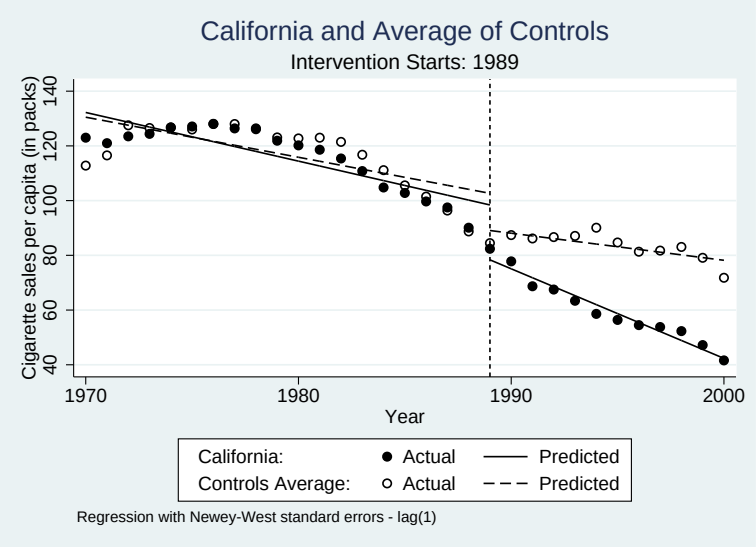


Figure 4: Multiple-group interrupted time series analysis with Newey-West standard errors and one lag. Three states, comparable on the baseline level and trend of the outcome, are used for comparison.

packs, the control group decreased sales over the same period by only 1 pack, and the difference between them is 2.28 packs per-capita per year.

These results highlight the importance of ensuring that treatment and control units are comparable on the pre-intervention level and trend of the outcome variable when conducting a multiple-group ITSA. As described in `Section 2.2`, an iterative process can be used in which each non-treated group is compared separately to the treatment group. Those groups who have $P$-values greater than a specified threshold on both $\beta_4$ and $\beta_5$ of Equation 3 can be retained as controls for inclusion in the final model. This approach can be easily extended to other covariates as well, however if achieving balance on many covariates is an important factor, two alternative approaches to `itsa` should be considered: the synthetic controls approach described by Abadie et al. (2010) and implemented in Stata using the `synth` package (Abadie et al. [2014]), or the propensity-score weighting technique described by Linden and Adams (2011).

## 4.3 Multiple treatment periods

`itsa` can accomodate design variations in which the effect of multiple treatment periods are of interest. For example, the researcher may be interested in studying the effects of an intervention which is introduced, withdrawn, and reintroduced. Or an intervention which is followed by a separate intervention at a later point in time (see Barlow et al.

[1984] for many other design alternatives).

For exposition, in the following example we add a fictitious intervention to the cigarette sales data, starting in 1982. Thus, we re-estimate the single-group ITSA from `Section 4.1`, now with one additional intervention period.

```
. itsa cigsale, single treat(3) trperiod(1982 1989) lag(1) replace posttr fig
       panel variable:  state (strongly balanced)
        time variable:  year, 1970 to 2000
                delta:  1 unit
Regression with Newey-West standard errors          Number of obs  =        31
maximum lag: 1                                      F(  5,    25)  =    657.58
                                                    Prob > F       =    0.0000
```

| cigsale | Coef. | Newey-West Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _t | -.2038463 | .3343824 | -0.61 | 0.548 | -.8925198 | .4848273 |
| _x1982 | -8.040472 | 2.875704 | -2.80 | 0.010 | -13.96309 | -2.117849 |
| _x_t1982 | -3.639012 | .378855 | -9.61 | 0.000 | -4.419278 | -2.858745 |
| _x1989 | -9.285162 | 2.561198 | -3.63 | 0.001 | -14.56005 | -4.010275 |
| _x_t1989 | .5687319 | .3153918 | 1.80 | 0.083 | -.0808296 | 1.218293 |
| _cons | 125.2333 | 2.281397 | 54.89 | 0.000 | 120.5347 | 129.932 |

Post-Intervention Linear Trend: 1982

Treated: _b[_t]+_b[_x_t1982]

| Linear Trend | Coeff | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Treated | -3.8429 | 0.2237 | -17.1757 | 0.0000 | -4.3037 | -3.3821 |

Post-Intervention Linear Trend: 1989

Treated: _b[_t]+_b[_x_t1982]+_b[_x_t1989]

| Linear Trend | Coeff | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Treated | -3.2741 | 0.2793 | -11.7205 | 0.0000 | -3.8495 | -2.6988 |

The interpretation of all coefficients up to the second intervention is as before. That is, the first intervention period is compared to the pre-intervention period. However, the additional coefficients for the second intervention period `_x1989` and `_x_t1989` are now compared to those of the prior (first) intervention period.

As shown in both the regression table and verified upon visual inspection of Figure 5, there is evidence of a "treatment effect" beginning in 1982, and no additional decrease in annual sales after the implementation of the second intervention (which in reality was the true intervention period).

We can demonstrate this effect further via the `posttrend` option, which estimates the post-intervention trends separately after the first and second intervention periods. As shown in the `posttrend` output, the annual decrease in cigarette sales after 1982 was

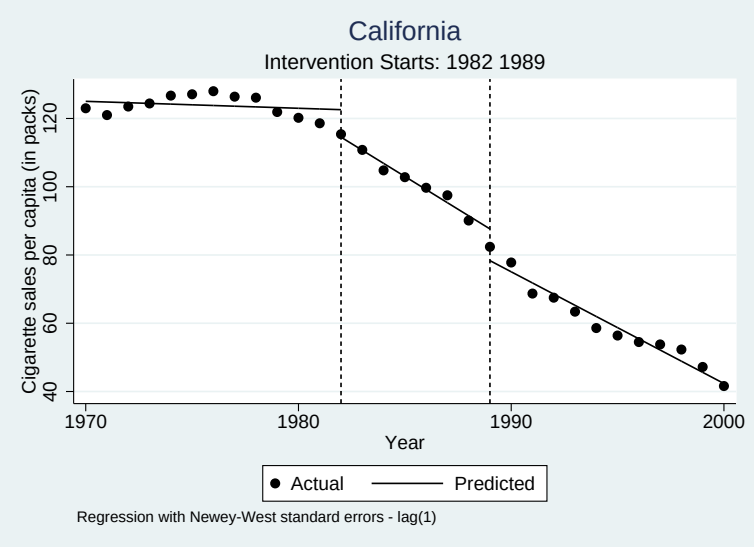


Figure 5: Single-group interrupted time series analysis with Newey-West standard errors and two intervention periods.

3.84 packs per year, while the decrease in annual sales after 1989 was slightly less, at 3.27 packs per year (the difference is 0.569, which appears in the original regression table as `_x_t1989`). Thus, the results of this exercise reveal an additional utility of defining multiple treatment periods in `itsa` when analyzing single-group data; it allows for the testing of "interruptions" away from the true start of the intervention, as described in `Section 2.1`. As demonstrated in this example, a statistically significant reduction in the trend of cigarette sales in California began several years prior to implementation of Proposition 99. This result also further highlights the importance, even when using interrupted time-series analysis, of finding a comparable control group to represent the counterfactual.

# 5 Discussion

While the randomized controlled trial remains the gold standard research design, there are situations in which this design is not feasible or practical, such as when large scale interventions or policy changes target the entire population. When data are available for multiple time-points in both the pre- and post-intervention periods, interrupted time-series designs offer a robust quasi-experimental alternative for evaluating treatment effects (Campbell and Stanley [1966], Shaddish et al. [2002]).

In this paper we have demonstrated the basic implementation of `itsa` to estimate treatment effects for a single treatment group, a multiple-group comparison, and when

more than one intervention has been employed sequentially. Additional important issues were also addressed, such as: criteria for choosing and specifiying a model, testing for autocorrelation, robustness testing for interruptions prior to the true intervention start period, and choosing comparable controls. More complex models can easily be estimated with `itsa` by including additional covariates to control for confounding, seasonal effects and the impact of external events. Moreover, the addition of key time series variables to the data-set after running `itsa` allows for further estimation of treatment effects using more complex OLS models or `arima`, assuming that more sophisticated time series modelling is warranted and the availability of a sufficient number of observations.

## 6 Acknowledgement

I would like to thank Nicholas J. Cox for his support while developing `itsa` and Steven J. Samuels for creating the `posttrend` option and assistance with several post-review changes to `itsa`. I would also like to thank Michael J. Harvey for his assistance with LaTeX, and Steven J. Samuels, Roger B. Newson, John L. Adams, Andrew Ryan, and Julia Adler-Milstein for their review and helpful comments on the manuscript. I would also like to thank the anonymous reviewer and chief editor for their thoughtful reviews and recommendations for improving both the manuscript and the `itsa` command.

## 7 References

Abadie, A., A. Diamond, and J. Hainmueller. 2010. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 105: 493–505.

———. 2014. synth: Stata module to implement synthetic control methods for comparative case studies. Statistical Software Components s457334, Boston College Department of Economics. Downloadable from http://ideas.repec.org/c/boc/bocode/s457334.html.

Barlow, D. H., S. C. Hayes, and R. O. Nelson. 1984. *The scientist practitioner: Research and accountability in clinical and educational settings. Pergamon general psychology series*. New York: Pergamon Press.

Baum, C. F., and M. E. Schaffer. 2013. actest: Stata module to perform Cumby-Huizinga general test for autocorrelation in time series. Statistical Software Components s457668, Boston College Department of Economics. Downloadable from http://ideas.repec.org/c/boc/bocode/s457668.html.

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

Box, G. E. P., and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden Day.

Box, G. E. P., and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* 70: 70–79.

Briesacher, B. A., S. B. Soumerai, F. Zhang, S. Toh, S. E. Andrade, J. L. Wagner, A. Shoaibi, and J. H. Gurwitz. 2013. A critical review of methods to evaluate the impact of FDA regulatory actions. *Pharmacoepidemiology and drug safety* 22: 986–994.

Campbell, D. T., and J. C. Stanley. 1966. *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Cochrane Effective Practice and Organisation of Care (EPOC). 2013. EPOC Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services. Available at: http://epocoslo.cochrane.org/epoc-specific-resources-review-authors.

Crosbie, J. 1993. Interrupted time-series analysis with brief single-subject data. *Journal of Consulting and Clinical Psychology* 61: 966–974.

Cumby, R. E., and J. Huizinga. 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica* 60: 185–195.

Gillings, D., D. Makuc, and E. Siegel. 1981. Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *American Journal of Public Health* 71: 38–46.

Glass, G. V., V. L. Willson, and J. M. Gottman. 1975. *Design and analysis of time-series experiments*. Boulder: University of Colorado Press.

Gottman, J. M. 1981. *Time-series analysis. A comprehensive introduction for social scientists*. New York: Cambridge University Press.

Huitema, B. E., and J. W. McKean. 2000a. Design specification issues in time-series intervention models. *Educational and Psychological Measurement* 60: 38–58.

———. 2000b. A simple and powerful test for autocorrelated errors in OLS intervention models. *Psychological Reports* 87: 3–20.

Imbens, G. W., and T. Lemieux. 2008. Regression discontinuity designs: a guide to practice. *Journal of Econometrics* 142: 615–635.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied linear statistical models*. New York: McGraw-Hill Irwin.

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

McDowall, D., R. McCleary, E. E. Meidinger, and R. A. Hay. 1980. *Interuppted time series analysis*. Newbury Park, CA: Sage Publications, Inc.

McKnight, S., J. W. McKean, and B. E. Huitema. 2000. A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods* 5: 87–101.

Muller, A. 2004. Florida's motorcycle helmet law repeal and fatality rates. *American Journal of Public Health* 94: 556–558.

Orzechowski, W., and R. C. Walker. 2005. *The tax burden on tobacco. Historical compilation*, vol. 40. Arlington, VA: Orzechowski & Walker.

Ramsay, C. R., L. Matowe, R. Grilli, J. M. Grimshaw, and R. E. Thomas. 2003. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *International journal of technology assessment in health care* 19: 613–623.

Riley, W. T., R. Glasgow, L. Etheredge, and A. P. Abernethy. 2013. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. *Clinical and translational medicine* 2: 1–6.

Shaddish, S. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Simonton, D. K. 1977a. Cross-sectional time-series experiments: Some suggested statistical analyses. *Psychological Bulletin* 84: 489–502.

———. 1977b. Erratum to Simonton. *Psychological Bulletin* 84: 1097.

Velicer, W. F., and J. W. Harrop. 1983. The reliability and accuracy of time series model identification. *Evaluation Review* 7: 551–560.

Velicer, W. F., and R. P. McDonald. 1991. Cross-sectional time series designs: A general transformation approach. *Multivariate Behavioral Research* 26: 247–254.

**About the author**

Ariel Linden is a health services researcher specializing in the evaluation of health care interventions. He is both an independent consultant and adjunct associate professor at the University of Michigan in the department of Health Management and Policy, where he teaches program evaluation.