

Use of Interrupted Time Series Analysis in Evaluating Health Care Quality Improvements

Robert B. Penfold, PhD; Fang Zhang, PhD

From the Group Health Research Institute and the Department of Health Services Research, University of Washington, Seattle, Wash (Dr Penfold); and Harvard Pilgrim Healthcare Research Institute and Department of Population Medicine, Harvard University, Boston, Mass (Dr Zhang)

The views expressed in this report are those of the authors and do not necessarily represent those of the US Department of Health and Human Services, the Agency for Healthcare Research and Quality or the American Board of Pediatrics Foundation.

The authors declare that they have no conflict of interest.

Publication of this article was supported by the Agency for Healthcare Research and Quality and the American Board of Pediatrics Foundation.

Address correspondence to Robert B. Penfold, PhD, Group Health Research Institute and Department of Health Services Research, University of Washington, 1730 Minor Ave, Suite 1600, Seattle, WA 98115 (e-mail: Penfold.r@ghc.org).

Received for publication March 14, 2013; accepted August 2, 2013.

ABSTRACT

Interrupted time series (ITS) analysis is arguably the strongest quasi-experimental research design. ITS is particularly useful when a randomized trial is infeasible or unethical. The approach usually involves constructing a time series of population-level rates for a particular quality improvement focus (eg, rates of attention-deficit/hyperactivity disorder [ADHD] medication initiation) and testing statistically for a change in the outcome rate in the time periods before and time periods after implementation of a policy/program designed to change the outcome. In parallel, investigators often analyze rates of negative outcomes that might be (unintentionally) affected by the policy/program. We discuss why ITS is a useful tool for quality improvement. Strengths of ITS include the ability to control for secular trends in the data (unlike a 2-period before-and-after *t* test), ability to evaluate outcomes using population-level data, clear graphical presentation of results, ease of conducting stratified analyses, and ability to evaluate both intended and unintended consequences of interventions. Limitations of ITS include the need for a minimum of 8 time periods before and 8 after an interven-

tion to evaluate changes statistically, difficulty in analyzing the independent impact of separate components of a program that are implemented close together in time, and existence of a suitable control population. Investigators must also be careful not to make individual-level inferences when population-level rates are used to evaluate interventions (though ITS can be used with individual-level data). A brief description of ITS is provided, including a fully implemented (but hypothetical) study of the impact of a program to reduce ADHD medication initiation in children younger than 5 years old and insured by Medicaid in Washington State. An example of the database needed to conduct an ITS is provided, as well as SAS code to implement a difference-in-differences model using preschool-age children in California as a comparison group.

KEYWORDS: interrupted time series; quality improvement; quasi-experimental; research design

ACADEMIC PEDIATRICS 2013;13:S38–S44

THIS IS PERHAPS an unprecedented time for health care in terms of policy, information technology, and organizational change. Simultaneous efforts to improve the quality of care in this chaotic environment are ongoing. Rigorous methods for evaluating both positive, intended and negative, unintended consequences of interventions and policies are needed in order to determine what policies and interventions are effective and which are not. Randomized trials are often not the best approach.^{1,2} In addition to the enormous expense of funding such trials, the answers often cannot be provided on a timeline consistent with the need to make decisions. Further, the number of possible options to compare often makes infeasible the number of trial arms and the number of participants to be recruited. Most trials also use strict inclusion and exclusion criteria, which limit the generalizability of the results. Although randomized trials may be considered the gold standard of causal evidence (because randomization theoretically balances the intervention and control groups with respect to confounders and thereby reduces the potential for

unmeasured confounding), quasi-experimental designs, informed by extensive qualitative work about decision making, are likely the best way to move the discipline of quality improvement and implementation science forward.

Interrupted time series (ITS) is arguably the strongest quasi-experimental research design^{3–5}—particularly when the investigator does not have control over the implementation of an intervention, such as the inability to randomize clinicians or clinics or conduct a sequential rollout of the intervention. Here we give a brief description of ITS analysis, including how to construct the analytic database and perform the regression analysis. We also discuss the pros and cons of using ITS and compare the approach to a randomized trial. We provide readers with the basic tools to conduct their own ITS analyses.

A BRIEF DESCRIPTION OF ITS ANALYSIS

In the context of quality improvement, ITS is best understood as a simple but powerful tool used for evaluating the impact of a policy change or quality improvement program

on the rate of an outcome in a defined population of individuals. A time series—repeated observations of a particular event collected over time—is divided into 2 segments in the simplest case. The first segment comprises rates of the event before the intervention or policy, and the second segment is the rates after the intervention. “Segmented regression” is used to measure statistically the changes in level and slope in the postintervention period compared to the preintervention period. In other words, segmented regression is used to measure immediate (level) changes in the rate of the outcome as well as changes in the trend (slope). “Segmented” simply refers to a model with different intercept and slope coefficients for the pre- and postintervention time periods. An investigator may use a single time series describing only the intervention/policy site or (more strongly) compare the changes at the intervention/policy site to changes at another site where no intervention/policy occurred.

STRENGTHS OF ITS ANALYSIS

A notable strength of ITS with respect to evaluating the impact of quality improvement efforts using observational data is that the approach controls for the effect of secular trends in a time series of outcome measures. For example, suppose that an intervention is introduced at a hospital to reduce medication errors. Researchers find that the medication error rate in the year after the intervention is significantly lower than in the year preceding (ie, a *t* test comparing the postintervention rate to the preintervention rate is significant). However, the trend in the medication error rate was sloping downward for several years in this same hospital. Using a pre–post design, the researchers incorrectly attribute the annual reduction in the medication error rate to the intervention when in fact the decrease was likely due to other factors.

Figure 1 shows 2 scenarios: one in which the intervention was effective (white squares) and one in which the intervention was not (black diamonds). A simple comparison of the before-and-after mean rates in the black series will be statistically significant; however, the comparison would not be significant after controlling for the trend. In contrast, the white series has an identical trend before the intervention but decreases at a faster rate after the intervention.

In addition, there was an immediate drop in the medication error rate at the time of the intervention. The ITS design and use of segmented regression allow an investigator to test the change in level (ie, a change in the intercept) and change in slope associated with the intervention or change in policy while controlling for the overall trend in the outcome rate of interest.

Another powerful characteristic of ITS is that analyses can be conducted with respect to population rates rather than at the individual level. It is advisable to model the data using population rates when there is a clear linear trend in the population rates rather than in the log odds.^{6–10} Because the ITS approach evaluates changes in rates of an outcome at the population level, confounding by

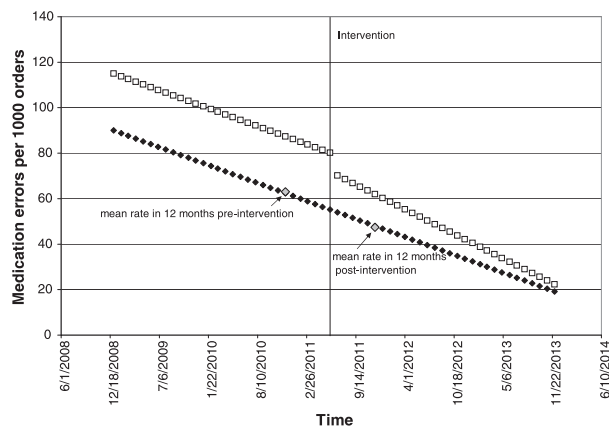


Figure 1. Hypothetical example of an intervention to reduce medication error rates. White squares indicate an example in which the intervention was effective; black diamonds, intervention not effective.

individual-level variables will not introduce serious bias unless it occurred simultaneously with the intervention. Standardization¹¹ is typically used to adjust for population shifts over time (ie, changes in the composition of the population with respect to individuals’ characteristics/traits).

A third characteristic of ITS is that the method readily lends itself to the analysis of unintended consequences of interventions and policy changes. Just as in the analysis of the outcomes of interest, investigators can construct time series of the rates of other potentially negative population-level events. Soumerai and colleagues have shown, for example, that medication authorization policies focusing on a psychotropic medication class or capping the number of allowable prescriptions decrease medication adherence rates, increase emergency department utilization rates, and increase hospital admission rates.^{12–15} Comparable studies have not been done in pediatric populations.

A fourth strength of ITS is that the investigator can easily conduct stratified analyses in order to evaluate the differential impact of an intervention or policy change on subpopulations of individuals (eg, by age, sex, race). For example, Du et al¹⁶ recently reported on the effect of the US Food and Drug Administration’s (FDA) black box warning on the increased risk of suicidal ideation in people receiving atomoxetine. The authors’ objective was to evaluate the impact of the warning on rates of medication initiation (including stimulants) for the treatment of attention-deficit/hyperactivity disorder (ADHD). Overall, adults were 3 times more likely to use atomoxetine. The authors therefore analyzed 3 time series of data separately for those aged 12 years or younger, those aged 13 to 18 years, and those aged over 18 years. The results of that study (Figs. 1 and 2 in Du et al¹⁶ in particular) clearly show that the impact of the black box warning differed across the 3 age groups.

Fifth, ITS provides extremely clear and easy-to-interpret graphical results. Even in the absence of the statistical output from a corresponding segmented regression model, presenting administrators and policy makers with graphs such as that shown in Figure 1 make a potent message. The reader can easily identify when the change occurred,

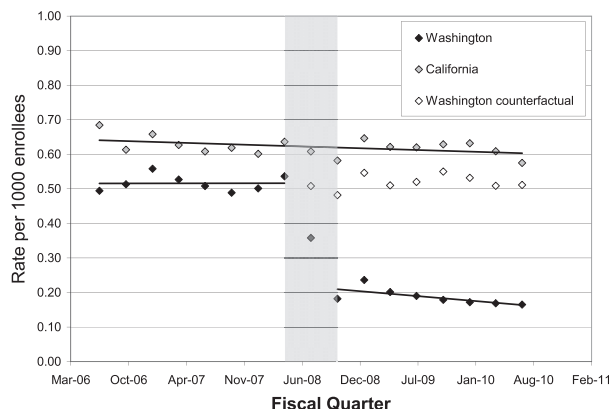


Figure 2. Example of an intervention to decrease ADHD medication initiation in children younger than 5 years and insured by Medicaid.

what was happening before the change, and what happened immediately after the change as well as in the follow-up period. One particularly stark example is presented by Rodgers and Topping¹⁷ in their study of the effects of drawstring requirements for upper outerwear on child death due to drawstring entanglement between 1985 and 2009; the risk was cut in half. Few statistical approaches can illustrate the effect of a policy with such clarity and impact.

ITS EXAMPLE—HYPOTHETICAL IMPACT OF A PROGRAM TO REDUCE ADHD MEDICATION USE IN PRESCHOOLERS

The American Academy of Pediatrics guideline on ADHD treatment recommends behavior therapy as the first-line treatment for ADHD in children.¹⁸ Although the 2011 guideline supports use of medications in children as young as 4 years of age, use in this population remains controversial.¹⁹ Wolraich et al²⁰ reported that more than 90% of pediatricians start medication at a sometimes or greater frequency in children with ADHD. The Preschool-Age Treatment Study reported that preschool-age children experienced a high rate of adverse events (30%), with up to 11% of the children discontinuing treatment as a result of adverse events.²¹

Figure 2 depicts the impact of a program^{22,23} that includes efforts to reduce initiation rates of ADHD medications in preschool children insured by Medicaid in the state of Washington (quarterly rates shown). Community clinicians that write ADHD medication prescriptions for Medicaid-enrolled children aged younger than 5 years in Washington State are required to consult with a pediatric psychiatrist before the prescription will be approved.^{24,25} Although this program is operational in Washington State, the data presented in this example are hypothetical.

The numerator for rates in this figure is all children younger than 5 years continuously enrolled in Medicaid for at least 10 months in a year, with a first prescription for an ADHD medication, including methylphenidate, dextromethylphenidate, amphetamine, dextroamphetamine, lis dexamphetamine, clonidine, guanfacine, and modafinil.

The denominator is all children younger than 5 years, insured by Medicaid, and enrolled in the calendar year for at least 10 months. Crude rates were calculated separately for Washington (where the program was implemented) and California (where no program was implemented). In this hypothetical example, California is the control state.

Visualization of the impact of the hypothetical program is clear. Before the program beginning in the third quarter of 2008, rates of initiation of ADHD medications were a little more than 0.50 per 1000 enrollees younger than 5 years. After the program, rates quickly dropped to about 0.20 per 1000 enrollees. It also appears from the graph that rates continued to decline slowly over the next 7 fiscal quarters. In contrast, rates of ADHD medication initiation in California averaged about 0.62 per 1000 enrollees in this population, and the rate was mostly stable. No contemporaneous change in rates occurs in the third quarter of 2008. There is perhaps a modest secular decline in the rate of initiation over the entire 17 fiscal quarters. Note that it is often desirable to include the 95% confidence intervals for the point estimates of rates in each time period. The power to detect a change in rates associated with an intervention is greatly influenced by the variability in the rates over time (ie, when the outcome rate of interest is going up and down), and this will naturally tend to occur in time series that are based on small populations. Graphing the 95% confidence intervals quickly conveys that the changes in rates are indeed significant (or not).

The next step in the analysis is to quantify the magnitude of the program's impact and test the statistical significance of the changes in level (ie, the rapid drop in rates immediately after the launch of the program) and slope (ie, the gradual decline in rates over the remainder of the follow-up period). This is accomplished by segmented regression.

Before proceeding to the statistical analysis, a brief discussion of assumptions is warranted. First, the validity of comparisons between the intervention and control groups depends on the assumption of exchangeability. In our example, the comparison between initiation rates in Washington and California is valid to the extent that the population of Medicaid enrollees younger than 5 years in Washington and California are more or less the same. In this example, we might question the assumption of exchangeability because the proportion of enrollees in California who are of Hispanic race/ethnicity is much greater than in Washington. Other factors, such as access to child psychiatrists or rural/urban distribution of the population, could also affect the validity of the comparison. The difference in size of the populations is controlled by using population rates. This example highlights the importance of choosing an appropriate control group for ITS. In cases where there is some imbalance (there is likely no perfect control group), investigators should consider matching individuals in the intervention group to individuals in the control group on relevant characteristics (eg, age, sex, race, urban/rural) or otherwise standardizing the groups in order to minimize any bias due to differences in the composition of the populations.

A second critical assumption in ITS is that the outcome of interest would remain unchanged in the absence of the policy, program, or intervention. Figure 2 shows the counterfactual outcome for Washington in gray diamonds. In a randomized trial, we know the counterfactual outcome (ie, what would have happened in the absence of the intervention) with certainty because the investigator purposefully withheld the intervention from the control group (which by randomization or block randomization is more or less the same as the intervention group). In ITS with observational data, the counterfactual outcome is supplied by the comparison population, thus emphasizing the need for this comparison population to be as similar as possible to the population where the program was implemented.

A third feature of Figure 2 is the shaded area centered around the implementation date (third quarter of 2008). Most policies or programs are implemented with a pilot phase (eg, a subset of sites within a system) or otherwise implemented over a period of time before they are more widely implemented. An investigator may therefore wish to evaluate the impact of the program once it is fully implemented. Figure 2 illustrates how this is done within the ITS approach. Rates are censored over the implementation period. That is, the segmented regression model is fit using only the data after the implementation period. In this example, the value for the third quarter of 2008 is censored.

SEGMENTED REGRESSION

In technical terms, the goal of the regression analysis is to estimate the interaction terms between implementation of a policy/program and time. We also wish to estimate the effects relative to the control population. Fortunately, the somewhat complicated presentation offered in most texts is easily simplified. Table 1 shows each of the data elements needed to fit the regression model. The “Quarter” column is simply the label for each time period. The next 2 rows are the crude (or adjusted when necessary) rates of the outcome variable for the intervention and control populations.

The fourth column is the difference in rates (ie, the Washington rate minus the California rate for every time period). Taking the difference allows the investigator to collapse the 2 time series into 1 in order to estimate a difference-in-differences effect—that is, to estimate how the change in the intervention population differed from the change in the control population over the same time period.

The fifth column is a binary variable indicating the time periods in which the policy/program was in effect. In our example, all fiscal quarters before July 2008 are coded 0 and all quarters after July 2008 are coded 1. This binary variable captures the interaction between the policy/program implementation and time. The regression coefficient on this variable is interpreted as the immediate impact on the level of the outcome (ie, an intercept change). The sixth column is simply an indicator of time and in this example covers the 17 time periods (fiscal quarters). The coefficient of “Time” captures the overall secular trend in rates over the entire time period (eg, if ADHD medication

Table 1. Data Table to Estimate the Regression Model

Quarter	WA	CA	WA_CA	Program	Time	Time_After
Jul-06	0.494	0.684	-0.190	0	1	0
Oct-06	0.513	0.613	-0.100	0	2	0
Jan-07	0.558	0.658	-0.100	0	3	0
Apr-07	0.527	0.627	-0.100	0	4	0
Jul-07	0.508	0.608	-0.100	0	5	0
Oct-07	0.489	0.619	-0.130	0	6	0
Jan-08	0.501	0.601	-0.100	0	7	0
Apr-08	0.536	0.636	-0.100	0	8	0
Jul-08	0.358	0.608	-0.250	0	9	0
Oct-08	0.182	0.582	-0.400	1	10	1
Jan-09	0.236	0.646	-0.410	1	11	2
Apr-09	0.201	0.621	-0.420	1	12	3
Jul-09	0.190	0.620	-0.430	1	13	4
Oct-09	0.179	0.629	-0.450	1	14	5
Jan-10	0.172	0.632	-0.460	1	15	6
Apr-10	0.169	0.609	-0.440	1	16	7
Jul-10	0.165	0.575	-0.410	1	17	8

initiation rates were generally declining over time in preschool-age children). Finally, the “Time After” variable is coded 0 before the policy/program is implemented, then sequentially numbers time periods after implementation. The regression coefficient on this variable captures the continuing effect of the policy/program—that is, the slope of the change in successive time periods (if any).

The regression model used to fit these data is straightforward:

$$\text{Rate}_t = \beta_0 + \beta_1 \text{time}_t + \beta_2 \text{program}_t + \beta_3 \text{time after program}_t + e_t$$

In this specific example (using the variable names from Table 1), the model is:

$$\text{WA_CA}_t = \beta_0 + \beta_1 \text{time}_t + \beta_2 \text{program}_t + \beta_3 \text{time_after}_t + e_t$$

Although the equation above has a linear specification, polynomial and nonlinear regression models can be used if the data exhibit nonlinear patterns. Indeed, careful examination of the time series for nonlinear patterns is critical because fitting a linear model to a nonlinear time series will lead to incorrect attribution of the change to the policy/program when in fact the change was simply due to the underlying nature of the trend in the data.

Another important caveat: the regression model for an ITS must be estimated using an autoregressive form. The technical details of autoregression are beyond the scope of this article; however, the core of the issue is that observations taken over time are correlated. Observations nearer together in time are often more strongly correlated. Stated another way, an observation at time t is linearly related to observations that precede it. Sometimes this occurs in the time periods immediately surrounding a particular observation, and other times it occurs at regular lags. A good example of the latter pattern are data with seasonality. Seasonality might occur, for example, in rates of ADHD medication initiation among school-age children, where

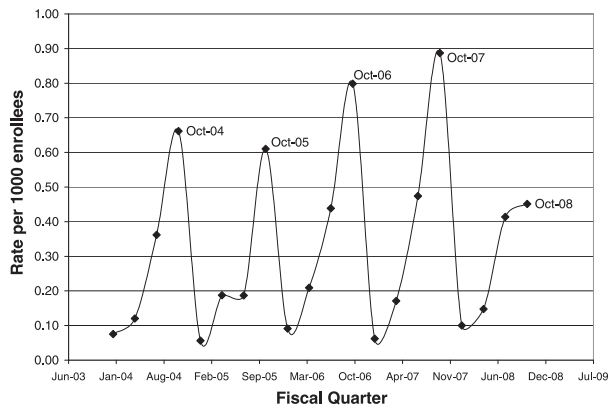


Figure 3. Example of seasonality in the initiation of ADHD medication in children aged 5 years and insured by Medicaid in Washington State.

initiation rates may be highest in October (after the start of school and potential identification of inattentive behaviors). Failing to account for the correlated nature of time series data will often lead to spurious conclusions regarding the effect of the policy/program under evaluation. An example of such data is shown in Figure 3, which illustrates rates of ADHD medication in kindergarten children insured by Medicaid in Washington State.

Fortunately, the practical investigator can be spared the complicated treatment of autoregressive integrated moving average (ARIMA) modeling. Estimating an autoregressive model can be painlessly accomplished in SAS software using PROC AUTOREG.²⁶ The AUTOREG function will automatically test for correlations in the data, estimate autoregressive parameters to be included in the model, and estimate the final parameters with the autoregressive parameters assumed given. The SAS code to implement the model in our example is shown in Figure 4.

The first line of this program specifies the data set to use as well as an output SAS data set of the estimated regression parameters. Line 2 is the model statement and specifies the dependent (CA_WA) and independent variables (Time, Program, Time_After). The fourth line specifies the modeling options. This example uses maximum likelihood to fit the model (method = ml). The model also specifies that 6 lags (t-1, t-2, t-3, t-4, t-5, t-6) should be tested and parameters included in the final model if statistically significant. For data with seasonal trends, investigators will often specify an nlag of 12 (ie, one for each month). Other ARIMA-related seasonal adjustment tools can also be used to reduce noise in the time series, including the US Bureau of the Census X-11 Seasonal Adjustment program.^{27,28} Line 4 also specifies that these lags should be entered into the model using backward elimination in order to fit the most parsimonious model. The DWPROB

```
PROC AUTOREG DATA=ITS_EG.ADHD OUTEST=ITS_EG.ADHD_param_estmts;
MODEL WA_CA = time program time after/
METHOD=ml NLAG=6 BACKSTEP DWPROB LOGLIKL;
OUTPUT OUT=ITS_EG.out p=pvar r=rvar;
RUN;
```

Figure 4. SAS code to implement an autoregressive ITS model.

Table 2. Parameter Estimates With AR Parameters Assumed Given

Variable	df	Parameter Estimate	Standard Error	t Value	Approx Pr > t
Time	1	-0.0045	0.006	-0.80	0.437
Program	1	-0.2602	0.043	-6.03	<0.0001
Time after	1	0.0002	0.009	-0.02	0.981

option specifies that a Durbin-Watson test is to be used to test for the presence of autocorrelation. Finally, the LOGLIKL option specifies that the log likelihood for the overall model be produced in order to assess the overall quality of the model.

The fifth line of the program in Figure 4 contains an output statement. The OUT = statement produces an SAS data set to store the predicted values and residuals from the model. The p = option specifies the full model predicted values; p = is a variable for the predicted mean; and the r = option is a variable for the residuals.²⁶

An abbreviated example of the output from running the model in Figure 4 is presented in Table 2. The output from the AUTOREG procedure includes a section for Ordinary Least Squares Estimates, Autoregressive Error Analysis, and Final Model Estimation. The final model estimation section includes the Fit Summary, Durbin-Watson Statistics, Parameter Estimates, and Parameter Estimates with AR Parameters Assumed Given.

The parameter estimates for “Program” and “Time After” are the main coefficients of interest. Recall that the parameter for “Time” controls for the overall secular trend in rates, which is generally treated as a nuisance variable, the effect of which should be removed in order to estimate the true impact of the policy/program. In our example, rates of initiation over the entire time period were not trending downward to a statistically significant degree. The “Time After” coefficient is also not significant, indicating that the apparent downward trend in Washington rates after the program was not significant. Finally, the coefficient for “Program” was highly significant, as is obvious from the graph. The hypothetical program was associated with a 0.26 decrease in the rate per 1000 enrollees of ADHD medication among children younger than 5 years and insured by Medicaid.

DISCUSSION

As we have demonstrated above, the ITS approach to policy/program evaluation has several advantages. The approach is easy to do and provides powerful, easy-to-understand results. ITS controls for secular trends in the data and therefore reduces bias that might be present in a simple 2-time-period model (ie, simple pre-post measurement and analysis). ITS does not require adjustment for individual-level characteristics.⁴

There are 3 important threats to validity in ITS analyses. The most serious of these is history.³ ITS analysis is only valid to the extent that the policy/program of interest was the only thing that changed at the demarcated point in time. The example above controls for one of these: the

overall secular trend. Other changes that could have affected the outcome variable are commonly referred to as competing interventions. In our hypothetical example, some competing interventions might be a FDA warning on prescribing to preschoolers, a Washington State policy on prior authorization for prescribing ADHD medications to preschoolers, or some other exogenous event such as a widely publicized stimulant overdose in a preschool-age child. Rigorous ITS analysis requires thoughtful consideration of competing interventions, and any study reporting the results of an ITS should include a discussion of possible competing interventions.

It should be noted, however, that a competing intervention must occur contemporaneously with the policy/program of interest in order to be a competing intervention. When other factors are hypothesized to change the outcome rate of interest, the investigator may add multiple “interruptions” to the time series and test the independent effect of each change. Such an approach may be particularly useful if the investigator desires to parse the impact of several, sequentially introduced components of a multi-component policy/program.

Another threat to validity in ITS is changes in instrumentation or the ability to measure the outcome rate of interest. For example, missing prescription data might falsely give the impression that ADHD medication initiation rates had fallen when in fact the data were simply missing. Hacker et al²⁹ reported on the rate of behavioral health screening in a large pediatric practice after the introduction of electronic medical records (EMR). After the EMR was introduced, paper records were still required in order to measure the true rate at which behavioral health screening was performed. Much of the paper data was not transferred to the EMR, giving the false impression that behavioral health screening decreased much more than occurred in reality.

A third important threat is selection bias, particularly if the composition of the intervention group changes at the same time as the policy/program.³ In our example, the validity of the study would be undermined if the composition of the preschool-age Medicaid population changed at same time the program to decrease prescribing occurred—for example, if preschool-age children were preferentially enrolled in managed care plans and prescription record data were not captured. Similarly, changes in eligibility categories such as a large decrease in preschool-age children in foster care would threaten the validity of the study. Again, investigators must carefully examine the time series data in order to verify that the composition of the population under study did not change across the pre- and post-policy/program periods.

LIMITATIONS

Although ITS has many strengths, there are important limitations to be aware of. First, estimating the level and slope parameters requires a minimum of 8 observations before and after the policy/program implementation in order to have sufficient power to estimate the regression

coefficients. In our example, ITS could not reasonably be used to evaluate the impact of the program on prescribing until 8 quarters (2 years!) after the program began. Estimating monthly or even weekly data may be feasible; however, the unit of time used in the analysis will generally depend on the lag between implementation of the policy/program the hypothesized impact, particularly when the impact occurs gradually.

A second related limitation is that there needs to be a sufficient number of time periods between interventions in order to estimate their impact independently. In general, this means that there should be at least 8 observations between components of a multicomponent program. This is often infeasible and/or undesirable. Although ITS can be implemented with fewer observations in the postimplementation period, this approach involves bootstrapping confidence intervals around values in the postimplementation period and requires stronger assumptions about the possible counterfactual outcomes.³⁰ Other methods, such as group sequential conditional probability ratio testing, offer robust alternatives to ITS that are capable of detecting changes in time series data with as little as 1 postpolicy/program data point.^{31–34}

A third limitation is the nonexistence of a suitable control population. In some cases, it is possible to use a nonequivalent control group³; however, even this may not exist. Consider the case of the FDA black box warning on antidepressants and suicidal ideation in children and adolescents. Although the warning was only intended for clinicians in the United States, rates of prescribing antidepressants to youth worldwide were likely affected by the FDA warning, thereby eliminating any contemporaneous control group. In such cases, ITS may still be conducted on the intervention group; however, the strength of inference is weaker in the absence of the counterfactual outcome.

Finally, ITS cannot be used to make inferences about individual-level outcomes when the series is a set of population rates. In our example, it would not be appropriate to conclude that the likelihood of any individual preschool-age child being prescribed an ADHD medication was lower after implementation of the program. Although it is tempting to make such inferences, it is an ideal example of the ecological fallacy. In order to make person-level inferences, an investigator would need to construct a time series of within-person measurements (eg, an individual's medication adherence rate measured over time with an interruption demarcating an intervention intended to increase adherence).

CONCLUSION

ITS is a simple but powerful approach to policy/program evaluation. Although the approach has limitations, few statistical approaches are as elegant in design and powerful in audience impact. ITS is particularly useful when a randomized trial is infeasible or unethical. Because ITS is the strongest quasi-experimental design, its value in quality improvement and program evaluation cannot be

understated. Paired with comprehensive qualitative data regarding the implementation of policies/programs, ITS may be more useful and certainly less expensive than comparable randomized trials designed to answer similar questions.

REFERENCES

- Gruenewald PJ. Analysis approaches to community evaluation. *Eval Rev*. 1997;21:209–230.
- Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci*. 2000;1:31–49.
- Shadish W, Cook T, Campbell D. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Mass: Houghton Mifflin; 2002.
- Wagner AK, Soumerai SB, Zhang F, et al. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002;27:299–309.
- Fan E, Laupacis A, Pronovost PJ, et al. How to use an article about quality improvement. *JAMA*. 2010;304:2279–2287.
- Soumerai SB, McLaughlin TJ, Ross-Degnan D, et al. Effects of a limit on Medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia. *N Engl J Med*. 1994;331:650–655.
- Soumerai SB, Ross-Degnan D, Gortmaker S, et al. Withdrawing payment for nonscientific drug therapy. Intended and unexpected effects of a large-scale natural experiment. *JAMA*. 1990;263:831–839.
- Soumerai SB, Avorn J, Ross-Degnan D, et al. Payment restrictions for prescription drugs under Medicaid. Effects on therapy, cost, and equity. *N Engl J Med*. 1987;317:550–556.
- 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.
- McDowall D, McCleary R, Meindinger EE, et al. *Interrupted Time Series Analysis*. Beverly Hills, Calif: Sage; 1980.
- Briesacher BA, Zhao Y, Madden JM, et al. Medicare part D and changes in prescription drug use and cost burden: national estimates for the Medicare population, 2000 to 2007. *Med Care*. 2011;49:834–841.
- Adams AS, Zhang F, LeCates RF, et al. Prior authorization for antidepressants in Medicaid: effects among disabled dual enrollees. *Arch Intern Med*. 2009;169:750–756.
- Soumerai SB, Zhang F, Ross-Degnan D, et al. Use of atypical antipsychotic drugs for schizophrenia in Maine Medicaid following a policy change. *Health Aff (Millwood)*. 2008;27:w185–w195.
- Zhang Y, Adams AS, Ross-Degnan D, et al. Effects of prior authorization on medication discontinuation among Medicaid beneficiaries with bipolar disorder. *Psychiatr Serv*. 2009;60:520–527.
- Soumerai SB, Ross-Degnan D, Avorn J, et al. Effects of Medicaid drug-payment limits on admission to hospitals and nursing homes. *N Engl J Med*. 1991;325:1072–1077.
- Du DT, Zhou EH, Goldsmith J, et al. Atomoxetine use during a period of FDA actions. *Med Care*. 2012;50:987–992.
- Rodgers GB, Topping JC. Safety effects of drawstring requirements for children's upper outerwear garments. *Arch Pediatr Adolesc Med*. 2012;166:651–655.
- Subcommittee on Attention-deficit/Hyperactivity Disorder Steering Committee on Quality Improvement and Management. ADHD: clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics*. 2011;128:2011–2654.
- Ghuman JK, Ghuman HS. Pharmacologic intervention for attention-deficit hyperactivity disorder in preschoolers: is it justified? *Paediatr Drugs*. 2013;15:1–8.
- Wolraich ML, Bard DE, Stein MT, et al. Pediatricians' attitudes and practices on ADHD before and after the development of ADHD pediatric practice guidelines. *J Atten Disord*. 2010;13:563–572.
- Greenhill L, Kollins S, Abikoff H, et al. Efficacy and safety of immediate-release methylphenidate treatment for preschoolers with ADHD. *J Am Acad Child Adolesc Psychiatry*. 2006;45:1284–1293.
- Hilt RJ, Romaine MA, McDonnell MG, et al. The Partnership Access Line: evaluating a child psychiatry consult program in Washington State. *JAMA Pediatr*. 2013;167:162–168.
- Thompson JN, Varley CK, McClellan J, et al. Second opinions improve ADHD prescribing in a Medicaid-insured community population. *J Am Acad Child Adolesc Psychiatry*. 2009;48:740–748.
- Hilt R. Primary care principles for child mental health, version 4.0. Available at: http://www.palforkids.org/docs/Care_Guide/Care_Guide_4.0_WA_Online_Version.pdf. Accessed July 25, 2013.
- Hilt R. Partnership access line Washington. Available at: <http://www.palforkids.org/resources/>. Accessed July 26, 2013.
- SAS. The AUTOREG procedure. In: *SAS/ETS 9.2 User's Guide*. Cary, NC: SAS Institute; 2008. Available at: <http://support.sas.com/documentation/cdl/en/etsug/60372/PDF/default/etsug.pdf>. Accessed July 26, 2013. Pages 313–428.
- Bobbit LG, Otto MC. Effects of forecasts on the revisions of seasonally adjusted data using the X-11 adjustment procedure. In: *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*. 1990:449–453. Washington, DC; American Statistical Association.
- Buszuwski JA. Alternative ARIMA forecasting horizons when seasonally adjusting producer price data with X-11-ARIMA. In: *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*. Washington, DC; American Statistical Association. 1987:488–493.
- Hacker K, Penfold R, Zhang F, et al. Impact of electronic health record transition on behavioral health screening in a large pediatric practice. *Psychiatr Serv*. 2012;63:256–261.
- Zhang F, Wagner AK, Soumerai SB, et al. Methods for estimating confidence intervals in interrupted time series analyses of health interventions. *J Clin Epidemiol*. 2009;62:143–148.
- Brown JS, Kulldorff M, Chan KA, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf*. 2007;16:1275–1284.
- Cook AJ, Tiwari RC, Wellman RD, et al. Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the Mini-Sentinel pilot. *Pharmacoepidemiol Drug Saf*. 2012;21(suppl 1):72–81.
- Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Stat Med*. 2010;29:284–295.
- Lieu TA, Kulldorff M, Davis RL, et al. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care*. 2007;45(10 suppl 2):S89–S95.