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 intervention 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

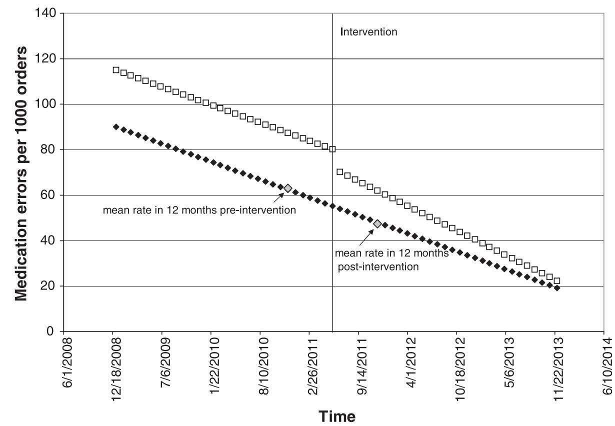

Figure 1. *Hypothetical example of an intervention to reduce medication error rates. White squares indicate an example in which the intervention was effective; black diamonds, intervention not effective.*

individual-level variables will not introduce serious bias unless it occurred simultaneously with the intervention. Standardization[11] 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 al[16] 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 times 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 al[16] 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,