Did the Protocol Change Work?
Interrupted Time Series Evaluation for Health Care Organizations
Carol Conell. Division of Research. Kaiser Permanente Northern California, Oakland, CA
Alexander C. Flint. Department of Neuroscience. Kaiser Permanente, Redwood City, CA

ABSTRACT

BACKGROUND: This paper introduces interrupted time series (ITS) analysis to the novice user. The primary objective of the paper is to enable intermediate level SAS® programmers to apply this well-established econometric technique. As an illustration of ITS, we evaluate a one-time protocol change at Kaiser Permanente Northern California. Specifically we consider the impact on statin administration of requiring physicians admitting ischemic stroke patients to opt-out of ordering statins instead of requiring them to opt-in. .

METHODS: We use PROC AUTOREG to investigate the structure of the outcome time series and evaluate whether the protocol change resulted in an increase in early statin administration. We also investigate the sensitivity of the estimates to the number of time periods available. Finally we show how the approach can be used to explore possible heterogeneity of treatment effects (HTE).

CONCLUSIONS: We show that the protocol change substantially increased early statin administration. The conclusion was invariant to analyzing one or two years before and after the protocol change and to analyzing a quarterly rather than a monthly time series.

Using the quarterly time period enabled us to create separate time series for minor, moderate and moderate to severe strokes and to show that the protocol change resulted in substantial increases in early statin administration both in moderate and in moderate to severe strokes but not in minor strokes.

Aggregating data and applying time series methods is a good way to evaluate the impact of protocol changes and similar interventions. When the intervention timing is well-defined, this approach can eliminate the need for data on individual-level confounders as well as problems associated with selection bias. If the effect is immediate, the approach requires only a moderate number of time points.

INTRODUCTION

Over the last decade the fraction of healthcare that is provided by organizations using an electronic healthcare record (EHR) has increased—partly in response to federal incentives and requirements—and the extent of penetration by EHR will continue to grow in coming years. More recently, organizations have begun transforming the EHR from a convenient way of documenting care to an active tool for improving the quality of care by incorporating active protocols. Moreover, the EHR often makes it feasible to observe exactly what decisions were made before or after each protocol change was implemented. As a result, analysts are increasingly asked to use routinely collected data to evaluate the impact of policy and protocol changes in healthcare.

Ideally, analysts would help plan for changes and could ensure that stepped wedge designs are used, but frequently the request for evaluation occurs after the changes are made. However, both in healthcare and in other contexts where workflow is routinely documented in an electronic record and protocols are used to control work, requests for rapid evaluation are frequently made after a change. A simple interrupted time series (ITS) approach can work well as a response, but this approach, although well-established in economics, is much less used in healthcare.

Time series analysis needs to allow for the fact that observations that are temporally close to each other often have correlated errors due to the underlying processes generating the data. Fortunately, PROC AUTOREG and other SAS® procedures in the ETS module provide a systematic way to identify and fit an appropriate model making it unnecessary to master econometric theory to evaluate the impact of a specific protocol change. However, the numerous procedural options and the differences between set up for ETS and STATISTICS modules can also be daunting. This paper is designed to enable intermediate level SAS® programmers to use time series analysis to evaluate a simple one-time protocol or policy change when electronic information from an EHR or other routinely gathered organizational data source is available. For this purpose, it provides a very simplified introduction to time series and the use of PROC AUTOREG in this situation.

A CASE STUDY: EVALUATING THE IMPACT OF PROTOCOL CHANGE

The specific case considered here involved a very simple change in the standard Kaiser Permanente Northern California (KPNC) admission protocol for ischemic stroke patients. Behavioral economic theories and empirical evidence across varied settings suggest that decision makers often accept whatever default an organization sets
Interrupted Time Series for Health Care, continued

(Johnson and Goldstein 2003). In order to maximize the practice of evidence based medicine, after determining that administering statins during hospitalization is beneficial, KPNC changed the stroke admission protocols. Before this change (henceforth, the Opt-In period) the protocol included an optional section that made it easy to order statins but did not require completion of this section. After the change (the Opt-Out period) admitting physicians who did not want to order a statin had to supply information on why they regarded a statin order as unsuitable.

The expected result was an abrupt and permanent rise in the likelihood of ordering statins right after admission, increasing the probability of “early statin,” defined as statin administered within 8 hours of arriving at the hospital. The expectation was that admitting physicians would be more likely to accept the default statin order, producing a shift to early statin. This change would eliminate delays due to physicians too busy managing the other pressing aspects of the case to think through whether statins were currently indicated, especially for patients with difficulty swallowing—common after a stroke—who would require non-oral administration. Although statins on admission are not always appropriate—either because of counter-indications or recent doses, there was agreement that they were appropriate for more patients than received them under the old protocol.

After the protocol change, we were asked to evaluate how much switching to an Opt-Out protocol accelerated statin orders and improved outcomes. For the sake of simplicity, this paper only considers the impact on “early statin.”

The original evaluation proposal required collecting detailed information on individual patients and analyzing the impact of the change at the individual level. Patients' medical histories and initial clinical presentation would be used to predict the propensity to order statins given their individual characteristics. The time-series approach minimizes the need for individual-level information by aggregating individual data over periods encompassing enough events to smooth out the impact of individual level variation.

For ease and clarity of presentation, we use a simplified analysis here. More complete results on the actual evaluation appear in Flint et al (Forthcoming).

THE INTERRUPTED TIME SERIES APPROACH

Individual level models for evaluating this protocol change have several disadvantages:

- With a binary outcome the best approach to modeling at the individual level involves logistic regression analysis, but the real objective was estimating the change in the fraction of patients who received “early statins.” This meant translating the change in the odds into changes in the probability, which requires making and justifying specific assumptions.

- At best, incorporating individual level factors requires substantial data collection and cleaning, slowing down the evaluation process and raising its cost. Since physicians’ decisions depend on individual clinical presentation, never completely captured by the EHR, there will always be unmeasured variation not reflected in the individual-level model.

- On the other hand, if a model includes irrelevant individual patient, provider, or practice measures, estimated confidence intervals will be inflated inappropriately.

- Individual level models also require careful consideration of how to incorporate higher level processes, including temporal trends. One approach that is sometimes used is simply to include a dummy variable for the period. However, this can result in conflating temporal trends with the impact of the intervention.

ITS models avoid these problems by fitting aggregate information for periods including sufficient individual observations to produce a series largely unaffected by random individual level variation. These models offer several advantages:

- They can model the probability of an outcome directly.

- They avoid the need for incorporating individual level factors.

- They eliminate the problem of selection bias reflecting unmeasured clinical characteristics.

- They avoid inflating confidence interval estimates by incorporating irrelevant individual level predictors.

- They facilitate identifying two possible changes in an outcome in response to a protocol or policy change: an immediate shift at the time of the change and a change in slope following the intervention.

- They make it easy to control for the overall trend in the outcome.

In our case, we calculated the proportion receiving “early statins” by month, with months defined relative to the date when the protocol change occurred—“interrupting” the time series. We used PROC AUTOREG to examine the
structure of this monthly time series and to determine whether the protocol change worked: as indicated by a significant increase in the likelihood of receiving “early statins” when the protocol was changed.

Using this approach also allowed us to check for heterogeneous treatment effects (HTE) in the impact of the protocol. In particular, we examined how the change affected patients with more and less severe clinical presentations and determined that all benefited from the change and there was evidence of HTE. In particular, the absolute change in the fraction receiving early statins increased most in those with less serious strokes, put the the change led to a greater increase in the relative chance of receiving early statins among patients presenting with “moderate to serious” strokes. Since the number of patients per month with severe strokes was not high enough to result in a smooth time series for the highest risk class, in the subgroup analyses we aggregated observations using quarters.

BASICS IN EVALUATING THE IMPACT OF THE OPT-OUT PROTOCOL

STEP ONE: CREATE A TIME SERIES DATABASE TO EVALUATE OPT-OUT.

The basic analysis database requires only 3 variables. The database for evaluating the change to Opt-Out on September 2011 includes 48 months from September 2009-August 2011, equally divided into the 2-year pre-intervention period, (September 2009-August 2011), and the post-intervention period (September 2011-August 2013):

- The outcome measure, **Outcome**: For this study, **Outcome** was P_Stat8h: defined as the probability of “early statin administration” and operationalized as the fraction of subjects in each observation period who received “early statins.”

- The dichotomous **Intervention** to assess: For this study this was Opt-Out, coded 0 for months 0-23,1 for months 24-47

- The measure of time **Trend**. Numbered from 0 for September 2009 to 47 for August 2013.

In order to test whether the slope of the trend in **Outcome** changed when the intervention occurred, we included two additional trend measures: **TrendPre**, which captures the trend during the pre-intervention period and is 0 after the intervention, and **TrendPst**, which captures the trend during the post-intervention period.

Selecting Time Periods

The time series periods should be selected to ensure the following:

- **Sufficient events in each period** for the law of large numbers to ensure reasonable approximations of the period-specific probability of a positive outcome. In most projects, 80-100 events should achieve this. The simple trend graph in Figure 1 illustrates a series with sufficient observations to create a fairly stable measure. On the average, there were 120 subjects per month, with a range of 45-166.

- **Sufficient observations both pre- and post-intervention**. Efficiency is highest if there is the same number of time points before and after the intervention. A minimum of 50 is sometimes recommended to ensure that one can identify the time series structure of the database (Zhang et al 2009). However, in the absence of seasonality or higher than second order autoregression, as few as 8 or 9 periods before and after can be used to estimate a simple one-time change in level (Janos 2015) and even fewer if there is no autoregression and the intervention is expected to produce an immediate effect. Obviously there is a trade-off between the number of individual observations aggregated in each time point and the number of time points.

- **Evenly spaced periods**.

- **Each period (observation) is either before or after the intervention**.

For a simple one-time change with an immediate effect, the most efficient design includes the same number of months before and after the intervention. If it is possible to include a whole number of years in both the pre- and post-intervention periods, the possible impact of seasonality can be largely ignored. Otherwise, if there is reason to expect much seasonal variation in the outcome measure, one should test for seasonality. In this case, the intervention: Opt-Out, occurred in September 2011 and data on the outcome was consistently available from September 2009 – August 2013 so the basic analysis examines this 48-month period.

**From Individual Observations to Time Series Observations.**

An easy approach, illustrated below, is to use a Data Step to partition the individual study data into an appropriate set of time periods and then use the SUMMARY procedure to aggregate the information for each time period observation.
The Data Step: Associate Time Period with Each Individual Observation and Create Basic Predictors.

Choose the appropriate time period and determine how many months before and after the intervention will be in the study:

- We chose single months as the basic observation periods in this evaluation, because there were about 100 observations per month. As discussed above, this number is sufficient to produce a well-behaved estimate of the proportion with a positive outcome (Early Statins).

- Given the timing of the change, there was consistent information on when individuals presented at the hospital, how severe their strokes were based on the Modified NIH stroke scale (mNIHss), and when they received statins. To make it easier to check for seasonality and any other time series structure, we wanted to start with a period that was symmetric around the time of the protocol change and included at least 2 years before and after, enabling us to include approximately 50 observations in the basic time series. If individual observations are much rarer one might want to use quarters or years as the basic period; if they are much more common one could use days or weeks.

- The code below shows how we associated a time period (Trend) to each individual observation using the Intck function.

- Adding the number of months in the study period before the protocol change (NStudModPreChange) to the number of months relative to the protocol change ensures that Trend is coded 0 for the first time period observation. As a result the coefficient associated with Trend provides a straightforward estimate of the linear temporal trend and the intercept term estimates the value at the beginning of the study period.

- The first argument in the Intck function is the type of basic time period, here months. This must be in quotes. The second argument (Intervention_date) is the Protocol Change Date. The third argument is the individual observation date (IndivDate). Intck calculates the number of months between the second and third arguments.

- The final argument, "c", indicates that the periods should be exactly one month long with a new period beginning on the protocol change date. If this is omitted, months will be based on calendar months, so that if the change occurred mid-month the period including the change would include both days before and after the change.

- Intervention is coded as 0 before the Protocol Change and 1 after.

- TrendPre and TrendPst are identical with Trend before and after the Protocol Change respectively.

```plaintext
Data IndivDa;
Set IndivDa (keep=IndivDate Outcome);
Trend = sum (Intck ("months", Intervention_date, IndivDate, "c"), NStudMoPreIntervention);
 Intervention = max (0, Trend > NStudMoPreIntervention);
TrendPre = max (0, Trend * (Not (Intervention)));
TrendPst = max (0, Trend * Intervention));
```

The Summary Step: Aggregate Information into the Time Periods.

Once the individual level database includes all of the above variables, creating the time series database with PROC SUMMARY is extremely straightforward.

The code below creates the basic time series database: Monthly_Time_Series_Data.

The summary statement uses the option nway to indicate that an observation should be created for each value of the class variable, Trend, without also creating an observation for the whole study population.

Each time series observation includes the measure Outcome, which in this evaluation is the fraction of individual strokes during the observation period who received Early Statins, i.e., statins within 8 hours of arrival at the hospital. This is calculated as the mean of Outcome. The data set also includes an estimate of the standard error of this outcome, std_Outcome.

Each observation in the time series database created by the Summary procedure also includes the value from the first observation in time period for each of the variables listed in the Id statement: Intervention, TrendPre, and
**TrendPst.** Since these variables depend only on Trend, they are coded identically for all observations during each period when Trend defines the period.

```sas
PROC SUMMARY data= IndivData (where= (0 <= Trend <= 47)) nway;
  Var Outcome ;
  Class Trend ;
  Id Intervention TrendPre TrendPst ;
  Output out= Monthly Time Series Data
    (rename=(_freq=nstroke) drop= _type_ )
    mean( Outcome )= Outcome std( Outcome )= std_Outcome ;
run;
```

Figure 1 plots the values of the *Outcome (P_Stath)* by month. This plot suggests that there was an increase in early statins following the intervention. However this plot also suggests that there might have been an upward trend reflecting more gradual changes due to factors other than the policy change, such as physician education about statins.

**Figure 1. Outcome (Early Statin) as a function of Time, for the Study Period. Intervention Sept 2011.**

---

**STEP TWO: FIT THE INTERRUPTED TIME SERIES MODEL.**

Once the time series database has been created, we use PROC AUTOREG to determine whether the Opt-Out protocol increased the likelihood of early statins (P_Stath), by fitting a series of time series models.

For our purposes, two features distinguish the AUTOREG procedure from the REG procedures:

- The input data set must be a Time Series Data Set containing observations corresponding to equally spaced time periods;
- The procedure calculates time-series diagnostics including Durbin-Watson statistics based on an autoregressive process of order k if one specifies dw= k dwprob.

The key question for this evaluation was whether there was a causal impact of the intervention on early statin use. Most evaluations of interrupted time series will include Trend, as we did. The reason is that making quasi-causal inferences using AUTOREG requires that the error terms in the model fit are independently and identically distributed (i.i.d.) once one allows for the time series structure of the outcome. Figure 1 shows an upward trend in the fraction receiving early statins, consistent with the fact that the decision to make the protocol change reflected increasingly
clear evidence that statins would benefit these patients. Advisories to this effect had circulated within the organization as well as appearing in the literature during the period preceding the change.

We estimate the impact of the protocol change (Intervention) on the probability of early statin (Outcome) controlling for Trend with the simple code shown below. The AUTOREG procedure uses the standard SAS conventions in specifying the input and the basic model. I.e., the dependent variable (Outcome) is on the left hand side of the equation in the Model statement and the predictors on the right hand side and options follow the “/”.

We specify 3 sets of options on the model statement:

• The first set of options (DW=k dwprob) calculates and displays the Durbin-Watson statistics for orders up to k. This enables us to check for autoregressive processes, as discussed below.

• The second set of options (Method= ml loglikl indicates that the model estimation should use maximum likelihood estimation and include the log likelihood value for the model in the output.

• The third option produces diagnostic plot outcome. AUTOREG can produce six diagnostic plots designed to make it easy to check for the fit of the model and the existence and order of autoregressive processes in the error terms. Here we ask for only two: the “standardresidual” plot is shown in Figure 2 and the “fitplot” plot in Figure 3, below.

\[\text{Proc AUTOREG data=Time-Series Data Set;}
\text{Model Outcome = Intervention Trend/ DW=k dwprob}
\text{Method=ml loglikl}
\text{plots (UNPACKpanel only) = (standardresidual fitplot)}\]

Check for Autoregressive Processes and Evaluate Model Fit

It is critical to evaluate model fit before attempting to interpret the model substantively. AUTOREG provides both tabular and graphic output to check the adequacy of each model. Table 1 displays tabular data on the Durbin-Watson statistics for autoregressive processes of order up to 12. We started by checking for an autoregressive process in the error terms. Since each observation simply aggregated information on all the eligible stroke admissions during a given time period, there was no obvious autoregressive process or seasonality. Moreover, neither Table 1 nor any of the figures (1-3) display any obvious evidence of autoregression. Nevertheless, with time series data, it is always necessary to rule out autoregression and this question will always be raised by reviewers. We specified DW=12 dwprob to allow for an autoregressive process AR(12) and calculate the Durbin-Watson statistics necessary to check for significant autoregressive of order 12 or smaller. Output 1 displays the Durbin-Watson statistics for autoregression of Orders 1-12, along with the probability that the Durbin-Watson statistic (DW) is small enough to suggest positive autocorrelation (Pr < DW) or large enough to indicate negative autocorrelation (Pr > DW). Note that if one detects evidence of AR(n) then there is no additional information to be gained from the reported DW of order n+1 or higher.

Since we found no evidence of autoregression based on the Durbin-Watson statistics, we only show two of the six diagnostic plots that PROC AUTOREG can produce. Figure 2 plots the residual for each time period based on the Basic Model. In contrast to the long runs of high or low residuals we see a basic pattern consistent with white noise. Figure 3 plots the observations along with the predictions and a 95 per cent confidence interval corresponding to the Basic Model. This confirms that the model is fitting the time series fairly well. We see that many of the observations are very close to the predicted values and all but 1 falls within the 95% confidence interval. Since we did not expect and do not find evidence of any significant autoregression, we do not need to include any lagged values of Outcome as predictors in fitting this model.
Table 1. AUTOREG Diagnostics for Basic Model

<table>
<thead>
<tr>
<th>Order</th>
<th>DW</th>
<th>Pr &lt; DW</th>
<th>Pr &gt; DW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7172</td>
<td>0.1010</td>
<td>0.8990</td>
</tr>
<tr>
<td>2</td>
<td>2.3214</td>
<td>0.8440</td>
<td>0.1560</td>
</tr>
<tr>
<td>3</td>
<td>2.0372</td>
<td>0.5818</td>
<td>0.4182</td>
</tr>
<tr>
<td>4</td>
<td>1.6234</td>
<td>0.1492</td>
<td>0.8508</td>
</tr>
<tr>
<td>5</td>
<td>1.5712</td>
<td>0.1496</td>
<td>0.8504</td>
</tr>
<tr>
<td>6</td>
<td>1.9920</td>
<td>0.7250</td>
<td>0.2750</td>
</tr>
<tr>
<td>7</td>
<td>1.8765</td>
<td>0.6491</td>
<td>0.3509</td>
</tr>
<tr>
<td>8</td>
<td>1.8813</td>
<td>0.7194</td>
<td>0.2806</td>
</tr>
<tr>
<td>9</td>
<td>1.5599</td>
<td>0.3686</td>
<td>0.6394</td>
</tr>
<tr>
<td>10</td>
<td>1.7227</td>
<td>0.6536</td>
<td>0.3464</td>
</tr>
<tr>
<td>11</td>
<td>1.5590</td>
<td>0.4984</td>
<td>0.5016</td>
</tr>
<tr>
<td>12</td>
<td>1.4221</td>
<td>0.3726</td>
<td>0.6274</td>
</tr>
</tbody>
</table>

Figure 2. Residual Plot for Outcome (Early Statin) as a function of Time Period for the Basic Model
Select A Model

Just as with regression models, we can select the best model using the Schwarz’s Bayesian information criterion (SBC), which is one of several statistics used to judge overall model fit. In particular, we checked two alternatives to the basic model. Table 2 shows the SBC for each of the competing models considered along with the estimated impact of intervention, i.e., the coefficient associated with Intervention. In the standard SAS output, the fit statistics are found in a table labeled Ordinary Least Squares Estimates and the parameter estimates in a table labeled Parameter Estimates. Alternatively, one can combine the results from several models into a single table using ODS.

In order to facilitate comparing models, we used ODS to store the output from the analysis systematically in a set of databases for each Model. In this evaluation, we gave each model a name ModName: Basic, No Trend, and Intervention Dependent Trend:

- **Basic** is the basic model discussed earlier that allows changes in the Outcome to depend on both an immediate permanent Intervention effect and a linear temporal trend. In any situation where there might be forces other than the Intervention operating that result in increases in an Outcome, it is important to allow for a possible trend to avoid attributing all changes between the pre- and post-intervention periods to the Intervention. We refer to this as the “Basic” model, because it is a logical default model, since one of the most important reasons for using time-series models is to avoid attributing temporal trends to an intervention.

  \[ \text{OUTCOME}_t = \beta_0 + \beta_1 \text{ Intervention}_t + \beta_2 \text{ Trend}_t + \mu_t, \]

- **No Trend** allows for only an immediate permanent Intervention effect and no temporal trend.

  \[ \text{OUTCOME}_t = \beta_0 + \beta_1 \text{ Intervention}_t + \mu_t \]

- **Intervention Dependent Trend** relaxes the assumption that the same temporal trend operates in the pre- and post-Intervention periods: i.e., it allows for the possibility of distinct trends before and after the intervention. This is appropriate when the impact of the Intervention changes over time. This can occur when the impact of the Intervention depends on decision-makers recognizing and adapting to changes in the environment. This type of mechanism will tend to produce a gradual change after the intervention, with little or none of its long-term effect appearing as an immediate effect. The effect of the basic Intervention term might not be significant, even if the Intervention worked. Instead the value of the TrendPst coefficient would
be significantly greater than that of TrendPre. Separate trend terms might also improve the model fit if a process that increases the Outcome is already operating before the intervention and the Intervention merely accelerates the process, e.g., if physicians are gradually switching to a more effective procedure and the intervention just results in some of these switches being accelerated. In this type of situation the coefficient associated with that TrendPst might well be not only significantly smaller than that of TrendPre but actually 0. Similarly, an intervention might render whatever was causing the trend inoperative. For instance, a protocol could be implemented requiring that some procedure that had been becoming more common be used universally.

\[ OUTCOME_t = \beta_0 + \beta_1 \cdot Intervention_t + \beta_2 \cdot TrendPre_t + \beta_3 \cdot TrendPst_t + \mu_t \]

In all models: \( OUTCOME_t \) is the Outcome, i.e., \( P_{\text{Stat8H}} \) for observation \( t \) in the time series, \( \cdot Intervention_t \) is the value of Opt-In for observation \( t \), \( \cdot Trend_t \) indicates the study month for observation \( t \) in the time series (or in the analyses of the quarterly time series, the middle month of the quarter), and as discussed previously, \( TrendPre_t \) fits a possible trend during the Opt-Out period but is 0 during the Opt-In period, while \( TrendPst_t \) fits a trend operating only during the Opt-In periods and is 0 during the Opt-Out period. Since \( TrendPre_t + TrendPst_t = Trend_t \) if there is no change in trend will lead to estimating two nearly identical trend terms and preferring the more parsimonious Basic model, which will have a smaller SBC.

AUTOREG can allow for an error term \( \mu_t \) that is the sum of a normal i.i.d. error term \( \epsilon_t \) and an autoregressive error term of order.

\[ \mu_t = \sum_{n=1}^{k} \phi_n \mu_{t-n} + \epsilon_t \]

However, AUTOREG starts by fitting the OLS model where the error term does not have any autoregressive terms, and only fits the model incorporating autoregression if the analyst includes autolag = k in the option list to indicate the need to allow for an autoregressive process of order AR(k).

Then we placed the following code block before the AUTOREG procedure call.

```plaintext
ODS graphics on;
ODS output FitSummary = FitModName Dwtest = dwFitModName ParameterEstimates = ParOLSModName Diagnosticspanel = diagpanlModName;
```

The resulting ODS Output makes it easy to produce the model comparison Tables 2-5 using the output detailed in the table notes in Table 2.

Since smaller values of SBC indicate a better fit, adjusted for differences in the number of degrees of freedom in different models, Table 2 indicates clearly that the Basic model is the most appropriate. Omitting the trend (No Trend) results in overestimating the impact of the intervention by incorrectly attributing all of the change observed to the intervention. As is typical when a model includes irrelevant factors, allowing for a change in the trend once the intervention was in place (Intervention Dependent Trend) does not alter the point estimate for the intervention effect. However, the partial multicollinearity present when one includes an intervention dependent trend, a basic trend and an indicator of the intervention, this model produces excessively conservative confidence intervals. Fitting this model could result in a false negative: i.e., incorrectly identifying the impact of the intervention as nonsignificant.
Table 2. Comparing Alternative Models for the Full Population  
Based on Monthly Data for 2 Years before and 2 Years after the Intervention

<table>
<thead>
<tr>
<th>Model</th>
<th>Temporal Trend</th>
<th>SBC¹</th>
<th>Impact of Intervention²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Simple Linear Trend</td>
<td>-188</td>
<td>.04* (.02)</td>
</tr>
<tr>
<td>No Trend</td>
<td>None</td>
<td>-176</td>
<td>.10*** (.01)</td>
</tr>
<tr>
<td>Intervention Dependent Trend</td>
<td>Distinct Trend before and after</td>
<td>-181</td>
<td>.04 (.04)</td>
</tr>
</tbody>
</table>

1. The Schwarz’s Bayesian information criterion (SBC) = 2ln(L) + ln(N)k, where ln(L) stands for the log likelihood associated with the model, N is the number of observations, and k is the number of degrees of freedom used up in fitting the model. All else being equal, a model with a smaller SBC value is preferred. The ln(N)k term is a penalty that helps to ensure parsimony in model selection. This statistic is labelled cvalue1 in the line where label1=’SBC’ in the FitSummary ODS Output Table.

2. This is the coefficient associated with the dichotomous Intervention term. It measures the immediate, permanent impact of the intervention. The point estimate is labeled Estimate, the standard error is labelled StdErr, and the significance level is labelled Probt in the observation in the ParameterEstimates ODS Output table where Variable = Intervention (Opt-Out in this study). *** indicates the probability of a higher value of the t-statistic is <.001, ** indicates between .001 and .01, * indicates between .01 and .5.

SENSITIVITY ANALYSES: LENGTH OF TIME SERIES, ALTERNATIVE PERIODS.

Having concluded that the protocol change did increase the likelihood of early statins, we want to explore how robust our findings are to different lengths of time series and alternative periods. In practice, the interest in a speedy evaluation often discourages the use of ITS to evaluate the impact of protocol changes or similar interventions. Decision-makers often pressure for information too quickly to allow a 24 month post-intervention period like that used in this study. Moreover, using months as the time period worked well for the entire population of strokes but was not appropriate for critical sub-groups of strokes, particularly for the patients presenting with moderate to severe symptoms, who made up about one tenth of the population and were particularly likely not to receive early statins before the protocol change.

We performed two alternative analyses: the first used monthly data but only 24 months (12 before and 12 after the intervention), the second used quarterly data for all 48 months.

Since we had created a 48-month time series analysis database, the only modification to the code required to check the impact of a shorter time series was to include an appropriate “where statement” on the data option in the basic call to PROC AUTOREG. Since we had set the database up so that 0<=Trend < 24 corresponded to the pre-intervention period and 24<=Trend <=47 corresponded to the post-intervention period, we modified the input data reference as shown below:

Proc AUTOREG data=Monthly_Time_Series_Data(where=(12 <= Trend <=36));

Shifting to a quarterly time series required repeating the two-step process used to create the initial monthly time series database. Rather than simply substituting “quarters” for months in the Intck function, we actually changed the name of the time series variable associated with each observation from Trend to Quarter so that we could continue to code Trend in months. This made it easier to compare the results. In the interest of simplicity we assigned Trend the value for each quarter corresponding to the middle month of the 3-month quarter, i.e., the average value for the quarter assuming an equal distribution of the individual events throughout the quarter. Subsequently, we performed the same analyses as previously, using the quarterly time series.

Tables 3 and 4 below report results for the first and second alternatives, respectively. Details on the table source are in Table 2 above. Substantively, both of these analyses lead to the same conclusions as the initial modeling: introducing an opt-out protocol improved patients’ chances of getting early statins substantially and significantly.

The fact that using a quarterly series rather than a monthly one (Table 4 versus Table 2) does not change the estimate of the intervention effect is extremely encouraging, as using the quarterly observations makes it possible to apply the time series technique to relatively small subgroups, in particular to patients with “moderate to severe”
Interrupted Time Series for Health Care, continued

strokes as we see below.

Examining a shorter period (Table 3 versus Table 2) leads to identifying a "No Trend" model as best. This results in a somewhat higher estimate of the impact of intervention, although the one-year and two-year estimates are not inconsistent given the estimated confidence limits.

**Table 3. Comparing Alternative Models for the Full Population:**
Based on Monthly Data for Only 1 Year Before and 1 Year After the Intervention

<table>
<thead>
<tr>
<th>Model</th>
<th>Temporal Trend</th>
<th>SBC</th>
<th>Impact of Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Simple Linear Trend</td>
<td>-90</td>
<td>.05 (.03)</td>
</tr>
<tr>
<td>No Trend</td>
<td>None</td>
<td>-92</td>
<td>.07*** (.01)</td>
</tr>
</tbody>
</table>

**Table 4. Comparing Alternative Models for the Full Population**
Based on Quarterly Data for 2 Years Before and 2 Years After the Intervention

<table>
<thead>
<tr>
<th>Model</th>
<th>Temporal Trend</th>
<th>SBC</th>
<th>Impact of Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Simple Linear Trend</td>
<td>-78</td>
<td>.04** (.02)</td>
</tr>
<tr>
<td>No Trend</td>
<td>None</td>
<td>-69</td>
<td>.10*** (.01)</td>
</tr>
<tr>
<td>Intervention Dependent Trend</td>
<td>Distinct Trend before and after Intervention</td>
<td>-75</td>
<td>.04 (.04)</td>
</tr>
</tbody>
</table>

**HETEROGENEITY OF TREATMENT EFFECTS IN TIME SERIES ANALYSES.**

The more information we amass, the clearer it becomes that there is substantial diversity within almost all diagnostic categories with consequent differences in the relative risks and benefits within patient groups who have a single diagnosis. Consequently, a major question for evaluations of this, and many other protocol changes, is who will benefit, and who if anyone will be harmed? As a result many evaluations call for exploration of possible heterogeneity of treatment effects (HTE).

An important methodological problem is posed by the fact that more severe presentations are typically relatively rare, so that even when the whole study population is fairly substantial the more severe subgroup is relatively small. This is particularly problematic, because improving medical care for those with severe disease is especially important, given their high acuity and typically greater risks of major negative consequences.

The ischemic strokes in this study followed this typical pattern. Over half (56%) presented with relatively mild stroke symptoms, scores 1-4 on the modified MNIH Stroke Scale (mNIHss); they had relatively short stays, very low risk of death in the short run and symptoms that often resolved even in the absence of any treatment. One third (33%) presented with moderate symptoms, mNIHss scores 5-15. Just over one tenth (11%) presented with "moderate to severe" mNIHss scores of 16 or higher. Figure 4 graphs the probability of early statin administration by severity for each of these three subgroups.

As Figure 4 and Table 5 (below) show, the fraction receiving early statins both before and after the protocol change was much higher for those with less severe symptoms. All the issues that might lead physicians to opt out of prescribing statins on admission were more prominent for those with more severe presentations: difficulty swallowing, multiple and complicated needs for management, less available information about the best approaches and less experience with similar cases. The absolute increase in the fraction of patients who received early statins based on a simple before and after comparison was greater the lower the severity. However the relative increase was actually greatest for the moderate to severe group.

Obviously there were not enough subjects per month in the Moderate to Severe group for a monthly time series specific to this group. However, once we shifted to quarters, the number of observations average about 45, allowing us to conduct this analysis even though this sample size is not as high as optimal. We created separate quarterly
time series for each of the 3 severity groups just described and ran the same sequence of analyses on each. Figure 4 (below) shows the observed probability of early statins by severity for the 3 subgroups. One indication of the impact of the smaller frequencies, is the amount of random variation visible in the series for the more severe (and smaller) subgroups. It also appears when we disaggregate that the upward temporal trend in early statins is probably confined to the mild strokes.

**Figure 4. Probability of Early Statins by Stroke Severity**

**Based on Quarterly Data for 2 Years Before and 2 Years After the Intervention**

Table 5 reports the impact of the intervention for the best model, selected based on SBC, as previously. The impact was significant for all subgroups, and only the Mild subgroup appeared to have a trend that needed to be fit. This may reflect the fact that many of those in this group were hospitalized for 24 hours or less, primarily for observation purposes, and actually had discharge orders written very early in their stay. During the study period this may have led to early statin administration for many with very mild strokes.
Table 5. Impact of Intervention as a function of Stroke Severity
Based on Quarterly Data for 2 Years Before and 2 Years After the Intervention

<table>
<thead>
<tr>
<th>Stroke Severity</th>
<th>Mild</th>
<th>Moderate</th>
<th>Moderate to Severe</th>
<th>Any</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact Intervention</td>
<td>.07**(.02)</td>
<td>.06**(.02)</td>
<td>.04**(.02)</td>
<td>.04 **(.02)</td>
</tr>
<tr>
<td>Best Model</td>
<td>Basic</td>
<td>No Trend</td>
<td>No Trend</td>
<td>Basic</td>
</tr>
<tr>
<td>Number of Strokes</td>
<td>3407</td>
<td>2028</td>
<td>696</td>
<td>6131</td>
</tr>
<tr>
<td>Number of Strokes per Period</td>
<td>104-242</td>
<td>85-153</td>
<td>21-47</td>
<td>210-444</td>
</tr>
<tr>
<td>Early Statins Before (%)</td>
<td>20</td>
<td>14</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>Early Statins After (%)</td>
<td>34</td>
<td>20</td>
<td>10</td>
<td>26</td>
</tr>
</tbody>
</table>

DISCUSSION
Interrupted time series analysis worked well for evaluating the impact of the change in the stroke protocol: from one allowing admitting physicians to opt-in and order statins to one requiring them to either order statins immediately or actively opting-out from doing so. ITS enabled us to abstract away from the complex circumstances of individual stroke admissions by aggregating strokes into either monthly or quarterly time series. The findings were fundamentally invariant when we substituted quarterly for monthly series and were also substantively the same when we shortened the pre- and post-intervention periods used in the analysis from two years to only one year before and after the protocol change.

Moreover, using the time series approach, we were able to investigate whether the impact of the protocol change depended on the severity of the stroke presentation. We found that the absolute change in the chance of receiving early statins was greatest for those with mild strokes and smallest for those with moderate to severe strokes, but the change in the relative risk was actually greatest for the moderate to severe group.

When applying this approach, it is important to consider carefully the nature of the mechanisms that one expects to account for the impact of any protocol change or other intervention. In this case, the impact of the protocol change was direct and immediate and the timing of the change was well-defined. Every ischemic stroke admission involved using a stroke protocol embedded in the KPNC EHR. Consequently, physicians faced one set of choices during the pre-intervention period and a different set in the post-intervention period. Moreover, no other changes to the protocols occurred during the study period, which could have confused the picture. In addition, the decisions about ordering statins were made by a large number of physicians, so that even if physicians are influenced by their own immediately preceding cases, the series as a whole is unlikely to exhibit autoregression. As a result, although, as in any time series study, it was important to check for possible autoregressive processes, the apparent absence of such processes was unsurprising.

However, in many situations autoregressive processes are important. We have discussed how to identify these processes here but not provided extensive detail on interpreting models of outcomes affected by significant autoregression, due to their limited relevance to this case.

Fortunately, the AUTOREG procedure makes it easy to identify autoregression and allow for it, as well as to check how well alternative models fit any series.

CONCLUSION
Aggregating data and applying time series methods provides a way to evaluate the impact of protocol changes and similar interventions. When the timing of these interventions is well-defined, this approach avoids the need to collect
substantial data on individual level confounders and problems associated with selection bias. If the effect of the intervention is immediate, the approach requires a very moderate number of time points to obtain a stable solution.

REFERENCES
Johnson EJ, Goldstein D. 2003. Do defaults save lives?

ACKNOWLEDGMENTS
The research was approved by the KFRI IRB board and supported by a Kaiser Permanente Northern California Community Benefit Grant to Jeffrey Klingman and Alexander C. Flint. This manuscript benefitted from helpful comments by Amy Alabaster, Lynn Conell-Price and Christopher Price.

RECOMMENDED READING
Interrupted Time Series for Evaluating Health Care

CONTACT INFORMATION
Your comments and questions are valued and encouraged.

Your comments and questions are valued and encouraged. Contact the author at:
Carol Conell
Division of Research, Kaiser Permanente
2000 Broadway Oakland, CA 94612
510 891-3574
FAX: 510-891-3606
Carol.Conell@kp.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.