Paper 4674-2020

Time After Time: Difference-in-Differences and Interrupted Time Series Models in SAS®

E Margaret Warton, Kaiser Permanente Northern California Division of Research

ABSTRACT

Healthcare and other epidemiological researchers are increasingly turning to difference-in-differences(D-I-D) and interrupted time series models (ITS) to analyze pre- and post-changes in outcomes around an intervention or exposure. These models are often used in quasi-experimental studies with non-randomized exposures using retrospective observational data, and they allow a causal interpretation and adjustment for secular trends in the outcome of interest. D-I-D models compare the rate of change an outcome measure before and after an exposure in exposed and control groups based on a single measure in each period. D-I-D data can be modeled with repeated-measures generalized linear models with an interaction term between time period and the exposure variable. The ITS – an extension of the D-I-D design - compares trends in an outcome over multiple pre- and post-time period measures, allowing for a discontinuity (an "interruption") in both control and exposure average rates during the study period. Models of this type account for situations where rates of an outcome of interest shift in both exposed and control groups at a specific time point, such as the change from ICD9 to ICD10 diagnosis codes during 2015. We will discuss the statistical basis of these methods and illustrate data structure, modeling methods, power calculations and interpretation of model estimates for both models with data from studies performed in a large integrated healthcare system.

INTRODUCTION

Difference-in-Differences (D-I-D) and Interrupted Time Series (ITS) models have long been used by economists to study changes in outcomes after an economic "shock". Epidemiology and healthcare researchers have lagged behind in their use, but this is changing as the knowledge of these methods and tools for using them have become more widespread. These longitudinal designs allow researchers to answer important questions about the causal impact of interventions or "interruptions" in long-term processes. Analysis of time series data using these methods can reduce costs and resources required to analyze the impact of interventions compared with clinical trials, especially in settings with large repositories of historical administrative data such as integrated healthcare systems. The ability to study the impact of changes in care provision protocols on healthcare-related outcomes makes these study designs particularly useful in healthcare research. This paper will discuss the study designs and their statistical framework, fundamental modeling and data requirements, and will illustrate some methods for analysis using data from a prior research study on the impact of a change in healthcare insurance benefits on medication costs.

OVERVIEW OF D-I-D AND ITS PRE-POST STUDY DESIGNS

D-I-D and ITS models are both study designs based on longitudinal or time-series data that is collected over multiple timepoints. In the case of D-I-D and ITS models in their simplest form, the outcome measure is collected at multiple timepoints other measures are collected at baseline. These study designs are intended to test the effect of an "interruption" that occurs at a known timepoint within the data collection period. The D-I-D design measures the outcome at only two timepoints, before and after the interruption, with the ITS design

requires multiple outcome measures of the outcome both before and after the interruption. In addition, the interruption should be "instantaneous"; the impact occurs immediately instead of gradually over time. These designs have been used by economists to study such interruptions as the adoption of new minimum wage, smoking restrictions, and motorcycle helmet laws. In the area of healthcare delivery research, these design can be used to study instantaneous interruptions with known starting dates such as bringing new laboratory testing equipment on-line, changes to healthcare benefits effective on a specific date, or the beginning of a pilot intervention to improve handwashing at a hospital.

D-I-D OVERVIEW

The D-I-D design is conceptually simple: measure the change in an outcome between the pre- and post-periods for an exposed group and an unexposed group, then subtract one from the other to see the "difference in the differences." In order to use the D-I-D analytic approach, a longitudinal cohort is divided into at least two groups with subjects exposed and unexposed to the condition or treatment of interest. Outcome measures must be available for members of both groups before and after the exposure time point for the exposed group. While the time points do not have to be specified calendar dates or even the same for every subject, this is the simplest way to set up a D-I-D study. The study design is very useful for measuring the results of programs, policies or protocols that are implemented at a specific time and are applied to a subgroup within a population. As with any study design requiring an unexposed comparison group, the identification of an appropriate unexposed group is key. The unexposed group should be as similar as possible to the exposed group, preferably observed over the same time period, and differing only in the exposure.

The only data required for a basic D-I-D analysis are an exposure flag, outcome measures, identified as pre or post, and a study ID for each individual. If the exposed and unexposed groups do not have equal distributions of outcome predictors, propensity scores can be used to improve the comparability of the two groups. The simplest D-I-D models are used with continuous outcomes, as changes in continuous outcomes are more easily interpreted, but binary outcomes can also be examined with the D-I-D study design (Warton, Parker 2018).

ITS OVERVIEW

The ITS study design estimates the impact of an interruption in an outcome based on multiple measures taken before and after the interruption. Unlike the D-I-D design, ITS models compare slopes of best-fit lines through the pre- and post-period outcomes separately to detect differences. The impact of the interruption is modeled with a step function variable, an indicator that is zero in all periods before the interruption and one afterwards. To calculate the pre and post-slopes, data for an ITS model must include multiple measures of a variable at time points separated by even intervals around the interruption.

Single Group ITS

While the strongest ITS design includes both an exposed and unexposed group, ITS models can be used in situations where there is no unexposed group If an adequate number of measures are available before and after the protocol change, no other interruptions affecting the outcome occur within the study period, the protocol change is essentially instantaneous, and the pre- and post-periods can by synchronized for any cyclic patterns endemic in the outcome, then ITS with no unexposed group is a legitimate causal analytic method.

Robust ITS

When both an exposed and unexposed group are available, the stronger Robust ITS analysis is possible. These types of situations arise when an interruption is only partial, when, for example, certain hospitals within a system implement a change, while others do not, or if

an intervention is rolled out over a long timeframe. If the unexposed hospitals are similar to the intervention hospitals with a long enough time period to make multiple evenly-spaced measures both before and after the interruption, the unexposed group will capture temporal changes that would (counterfactually) affect the intervention group in the same fashion had the interruption not occurred.

D-I-D AND ITS: SIMILAR MODELS THAT ANSWER DIFFERENT QUESTIONS

The D-I-D and ITS designs both compare an outcome (continuous, binary or ordinal) before and after a change, and both are well-suited to the analysis of planned and natural experiments. The D-I-D focuses on the difference in means between groups or the rate ratios for only one measure each in the pre and post periods and answers the question: was the amount of change different between the unexposed and exposed groups between these two points in time? It is estimated by comparing at the slopes of the lines between the pre- and post- measures using an interaction term between time and the exposure in a model that allows the slopes to differ.

The single group ITS, on the other hand, compares the slopes of least squares lines fit to multiple measures of the outcome on either side of a known interruption and estimates the value of the "jump" between the two periods caused by the interruption as well as comparing the slopes before and after the interruption. Instead of a single measure pre and post, ITS requires multiple evenly spaced measures, and preferably more than 8 or 9 on each side of the interruption. In addition, the ITS is more likely to be affected by cyclical trends, so if the outcome changes in a regular way over time, it is important that both the pre and post periods are matched on the cycle. Comparing monthly rates of influenza from April through December to rates in January through August for a January handwashing intervention with an ITS design would be a bad idea, as pre-post changes in slope would likely be due to the seasonality of influenza, not the intervention.

The robust ITS design compares pre- and post-interruption slopes between the exposure groups, but, like the D-I-D, it captures temporal trends with an unexposed group. The pre/post change in the unexposed group is subtracted from that of the exposed group, and the remaining change is due to the interruption, so long as the pre- and post-periods are comparable in regard to seasonality or other cyclical trends, and the unexposed group is an appropriate comparison group.

The D-I-D analysis requires only two outcome measurement timepoints for each group and answers the question "Did the two different groups change the same amount and in the same direction around the interruption?" The simple ITS requires multiple outcome measures both pre- and post-interruption for only one group, and answers the questions: (1) "Did the interruption change the outcome rate?", and (2) "How big was the jump (discontinuity) at the interruption point?" The robust ITS design requires multiple outcome measures pre- and post-interruption and answers multiple questions: (1)"Did the interruption cause a change in the outcome rate for either group?", (2)" Was the pre- to post-period change in the outcome rates between the groups significant?", and (3) "Was the immediate outcome change at the interruption point the same for both groups?"

CAUSALITY AND THE D-I-D AND ITS STUDY DESIGNS

Results from both D-I-D and ITS designs can be interpreted causally. Both designs maintain the temporal order of events and are based around a known intervention or interruption timepoint. It is always possible that some other unmeasured interruption is affecting the pre- and post-period outcome measures or that a trend in another confounding variable is responsible for the changes seen between periods or exposure groups. To check for trends over time and suggestions that other factors are at play, graphical inspections are useful. It is especially helpful to review trends in outcome measures before the beginning of the study period to ensure that the outcome rates appear stable prior to the study period.

In some cases, D-I-D and ITS study designs are used with static cohorts for which all measures are available for all members. In this case, all individuals serve as their own controls, and so confounding by measured and unmeasured factors is removed within each exposure group. Thus, in a single group ITS analysis, a static cohort strengthens the argument for a causal model result. However, this does not remove potential confounding due to differences in the distributions of these factors between the exposure groups in the D-I-D and Robust ITS with designs. Comparing distributions of potential confounders between groups in D-I-D and two-group ITS designs may suggest the use of propensity score weighting as part of the modeling process.

D-I-D and ITS designs are also used with dynamic cohorts, where the units measured at each timepoint fluctuate over time. When using dynamic cohorts in this way, it is important to choose an unexposed group that represents what would have happened to the exposed group had they not experienced the interruption. For the single ITS design where no unexposed group is available, using a short enough timespan to limit the effects of other secular trends that might vary between the pre- and post-periods is key.

BACKGROUND AND DATA FOR EXAMPLES

Healthcare costs have been rising rapidly in the United States for many years. In the past, nearly all health care plans provided by Kaiser Permanente of Northern California (KPNC) had no deductible and very low co-pays for healthcare visits and prescription medications. However, during the past ten to fifteen years, employers and individuals have begun to purchase health insurance with higher deductibles and co-pays to reduce premium costs. In order to offset the increasing out-of-pocket costs for employees, some employers have opted for new benefit programs that reduce costs on effective treatments that have been shown to improve patient health. Encouraging use of these preventive treatments, the theory goes, decreases the cost of providing care over time. Such plans, when part of a health insurance policy, are known as Value-Based Insurance Design (VBID) benefits.

Prescription medications have experienced many years of increasing costs, and with higher deductibles and co-pays, patient out-of-pocket costs have increased as well. Higher costs are hypothesized to lead to potentially detrimental changes in patient behavior, including taking lower doses than recommended or stopping a beneficial medication altogether. In 2013, KPNC began offering a VBID pharmacy benefit option to provide certain prescription medications for free, including drugs to treat high cholesterol, diabetes, and hypertension. We wanted to take advantage of this natural experiment to determine if the VBID medication benefit reduced adherence to certain medications among patients with a deductible plan (Reed, Mary. 2016). However, we did not have a large enough sample of patients who had a VBID medication plan added to an existing high deductible plan perform the analysis. Instead, we identified a cohort of patients on a non-deductible plan in 2013 whose employers switched to a deductible plan with a VBID benefit starting at the beginning of 2014. Our comparison cohort consisted of patients on a non-deductible plan in 2013 who switched to a deductible plan with no VBID benefit in 2014. All members of both cohorts had continuous healthcare coverage and full drug benefits throughout the study period, to ensure that they were likely to fill their prescriptions at KPNC pharmacies. The study timing, cohort composition and sample sizes are illustrated in Figure 1.

If the hypothesis that cost is a barrier to adherence which later affects health is correct, then demonstrating that the intervention had the intended effect of lowering out-of-pocket costs to the members with the plan is essential; if costs are not significantly reduced, adherence won't be affected, nor will downstream health. This retrospective analytic cohort built around a "natural experiment" is used to demonstrate both D-I-D and ITS study designs. Monthly medication costs in the years before and after the intervention provide the outcome measures of interest for these analyses: Annual and monthly out-of-pocket medication costs for all prescriptions filled at KPNC pharmacies from 2012-2014. KPNC

pharmacies regularly provide three months' worth of medication (90-100 days) per fill for chronic disease medications which can lead to high volatility in monthly out-of-pocket costs, so the monthly outcome is a three-month rolling average cost.

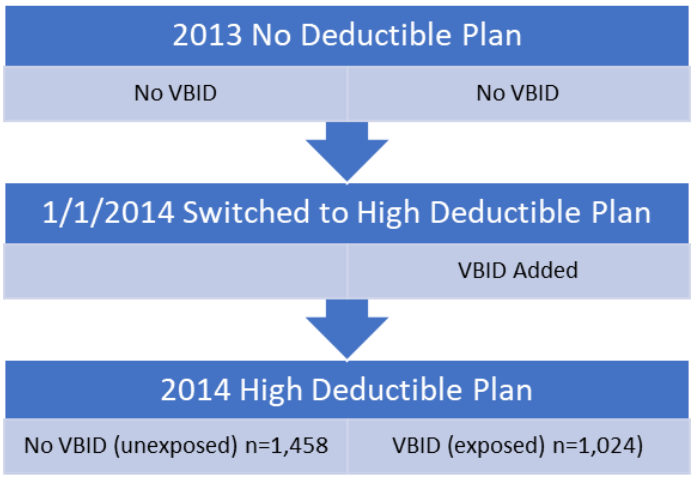

Figure 1. Cohort Timeline and Exposure Status

DIFFERENCE-IN-DIFFERENCES DESIGN AND ANALYSIS

D-I-D IN PICTURES AND EQUATIONS

A graphical illustration helps elucidate the D-I-D study design. In Figure 2, the left end of the solid line is the mean outcome value of the unexposed group for the pre-period. The bracket labeled $\boldsymbol{\beta}_0$ indicates how far from zero this starting point is in the unexposed group, which translates to the intercept value in the D-I-D model. The left end of the dashed line is the mean outcome value in the pre-period among the exposed group, and the bracket labeled $\boldsymbol{\beta}_{exp}$ indicates the different between the exposed and unexposed groups in the pre-period. The height of the right end of the solid line is the mean outcome value measured at the post time point for the unexposed, and this change in height is captured by $\boldsymbol{\beta}_{post}$, the slope coefficient in the model. This slope estimate captures any secular trend in the outcome over time in the unexposed group. The slope of the dashed line indicates the change in the outcome between the pre- and post-periods among those who experienced the exposure. The difference in slopes between the two groups is represented by $\boldsymbol{\beta}_{interaction}$, the coefficient of an interaction term between exposure group and time in the model. This value is the difference in the final value of the exposed group had it experience the same rate of change (counterfactually) as the unexposed group, in which case lines for both groups would have been parallel. In other words, this coefficient is equivalent to subtracting the change in the unexposed group from the change in the exposed group, and it provides an estimate of the effect of the exposure adjusted for background trends.

The basic equation for the D-I-D model is

$$\boldsymbol{\mu}_{it} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_{post}*\text{Post} + \boldsymbol{\beta}_{exp}*\text{Exposed} + \boldsymbol{\beta}_{interaction}*\text{Post}*\text{Exposed} + \boldsymbol{\varepsilon}_{it}$$

where $\boldsymbol{\mu}_{it}$ is the expected mean value for subject *i* at time *t*, Post is a binary indicator that the outcome measurement was made in the post period, Exposed is a binary indicator that the subject is in the exposure group during the post period. The Exposed indicator does not change between pre and post periods – it simply identifies the two different groups in the cohort. Finally, $\boldsymbol{\varepsilon}_{it}$ is the error term for the outcome measure of subject *i* at time *t*. As usual, errors are assumed to be normally distributed with a mean of zero. The model equation includes only outcome, time, and exposure measures – it includes no other

subject-level measures. The coefficient on the interaction term answers the main D-I-D question: Was the impact of the interruption the same in the two groups?

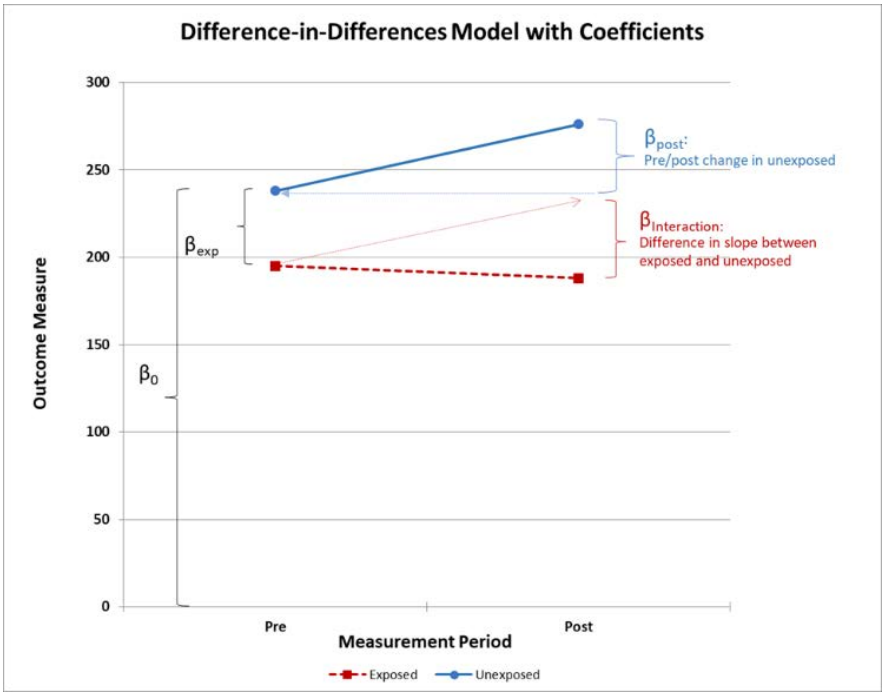

Figure 2. Difference-in-Differences Study Design with Model Coefficients

SAS® PROCEDURES FOR MODELING D-I-D DESIGNS

The D-I-D model is a repeated measures design; outcome values for a given subject are assumed to be correlated while outcomes between subjects are assumed independent. Methods used to model D-I-D studies need to account for the correlation within subjects as well as providing results for comparisons between the exposure groups.

Repeated Measures ANOVA (the GLM Procedure)

Repeated measures ANOVA using PROC GLM in SAS is an adequate method for a quick D-I-D model analysis. The procedure accounts for correlation within subjects, provides mean outcome values in each exposure group at each time period and statistical significance of the interaction term, and can compare the differences in the mean outcome values among the four time*exposure groups. It does not generate model coefficients or allow for the generation of individual predicted values; it simply tests for differences in mean values between groups.

Repeated Measures Linear Regression model (the MIXED Procedure)

A more flexible analytic approach when the outcome is continuous is the repeated measures linear model using PROC MIXED(Wolfinger, Russ and Chang, Ming, 1995). Using a mixed model with random intercepts means that individuals in given group have intercepts that can vary around the group mean. However, the slopes of individual members' trajectories in an exposure group are constrained to match the slope of all other group members, which is equal to the slope between the pre- and post-period group means. PROC MIXED tests the significance of the interaction term while accounting for correlation between measures. Unlike the ANOVA model, this method provides estimates of model parameters and can be used to create predicted outcome values. Propensity score weighting is allowed when required, and the LSMEANS and ESTIMATE statements generate summary predicted values for the different groups in the cohort. In addition, model coefficients can be used to predict

mean outcome values in the full cohort under different counterfactual scenarios. For example, to estimate the mean adherence in the full cohort during the pre-period assuming no exposure in the post-period, keep only the POST=0 records, set EXPOSED=0 for all cohort members, and generate predicted values using the model coefficients. Repeating this for all time periods and exposures can provide more realistic estimates of the outcome than the default methodology used by the LSMEANS statement.

Repeated Measures Non-Linear Models (the GENMOD Procedure)

D-I-D methods are commonly used in studies with binary outcomes. When the outcome measure is binary, the pre- and post-period group means represent the proportion of individuals in each group experiencing the outcome at each of the study time points. These proportions can be estimated as odds or probabilities on the log scale (relative risks) or as proportions (absolute risks). Note that in the binary case, the values of the outcome measure are constrained to 0 and 1 for all individuals. Therefore, a random-intercept mixed model for a binary outcome with only two time points is equivalent to a population-level marginal model and can be modeled using PROC GENMOD.

There are two main questions to consider when deciding on a model for a binary outcome D-I-D analysis: (1) how common is the outcome and (2) which is more interpretable - absolute or relative risks? For very rare outcomes, the odds ratio from a logistic regression model is an unbiased estimate of the relative risk. For more common outcomes, however, relative risks should be directly estimated using log-binomial or log-Poisson models (Zou 2004; Spiegelman and Hertzmark 2005). In these models, the log link is specified, and so coefficient estimates are on the log scale, and the results are referred to as the "Relative" D-I-D, the ratio of rate ratios, or the RRR. If the magnitudes of the underlying outcome rates are of peripheral interest with the main analytic focus being on the relative change between groups, then calculating relative risks and RRR is a reasonable analytic method.

Usually, however, the magnitudes of the pre to post changes in risk of the outcome between the exposure groups is of interest, and so estimates of absolute risk are required. In these cases, a binomial or Poisson model with the identity link yields estimates of absolute risk differences and prevalence in the exposure groups in the pre- and post-periods.

Propensity Score Modeling (the PSMATCH procedure)

Exposed and unexposed groups in a randomized control trial generally have similar distributions of measured characteristics due to the magic of randomization. Starting with exposed and unexposed groups with very similar characteristics in observational trials is desirable but not always achievable. The chief aim of a propensity score model is to generate weights, which, when applied to a cohort, remove any associations between baseline covariates and exposure status. Propensity scores are generated by modeling the probability of being in the exposed group using the available measured characteristics for the cohort. The "propensity" to be exposed is used to weight the cohort, hopefully resulting in similar distributions of potential confounders in the exposed and unexposed groups. PROC PSMATCH performs many of the steps required for generating and analyzing propensity score models(Warton and Parker, 2018).

Power Calculations for D-I-D Modeling (the GLMPOWER procedure)

PROC GLMPOWER can be used to estimate power or required sample size for a D-I-D model with a continuous outcome, as described by John Castelloe (Castelloe, 2014) and illustrated in detail in a previous paper (Warton and Parker, 2018). Power calculations require estimates of pre and post outcome values and standard deviations for both groups, the ratio of exposed to unexposed members, and the expected correlation between pre and post measures. Based on these estimates, the user can create an exemplary dataset to generate power and sample size results.

D-I-D ANALYSIS: INSURANCE BENEFIT CHANGES AND ANNUAL MEDICATION COSTS

Graphical Analysis

The switch from a no deductible to a high-deductible healthcare plan increases patient's out-of-pocket costs for services and medications immediately and, often, substantially. All members of the study cohort experienced this change on 1/1/2014. An essential first step to any proposed D-I-D analysis is to graphically review the data to determine if the study design is appropriate and whether propensity scoring may be called for. Figure 3 is a plot of 15 observations each from the exposed and unexposed groups. The plot suggests that no VBID group members had substantially higher annual medication costs than VBID group members in 2013. The no VBID group members had higher variability in costs in 2013 in the change in costs between 2013 and 2014. The randomly selected VBID group members are clustered much more closely together in 2013 and 2014, in general. In addition, it appears that the general trend for the no VBID cohort is roughly flat: the number of members with increased annual costs appears to be roughly offset by the number of members with decreased annual costs. The VBID cohort, however, appears to have more members with an overall decline in annual medication costs which may suggests that, despite moving to higher deductible plan, out-of-pocket medication costs may be decreasing due to the VBID plan.

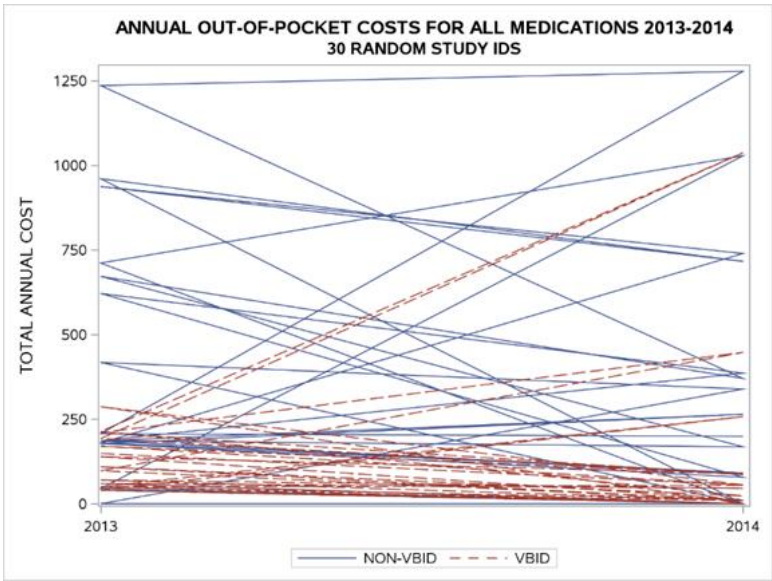

Figure 3. Annual Medical Costs for Thirty Random Cohort Members

Another important plot is of the mean out-of-pocket medication costs for the two groups from 2012 through 2014(Figure 4). This plot highlights differences in outcome trends before the study period that might bias D-I-D results. This plot matches the intuition gained from Figure 3; the mean of the VBID group members starts at a lower annual cost than the no VBID group mean, and, while the no VBID group mean cost appears to increase between 2013 and 2014, the VBID group does not. The key area of this plot, however, is the 2012-2013 section. While the VBID group costs are clearly much lower in 2012 and 2013, both groups show the same gradual rise in the outcome before the beginning of the study period in 2013. This suggests that the unexposed group is a reasonable comparison group, as it likely captures secular trends that would impact the VBID group had they not received VBID benefits during 2014. Had these slopes been more divergent, results from a D-I-D analysis would be questionable.

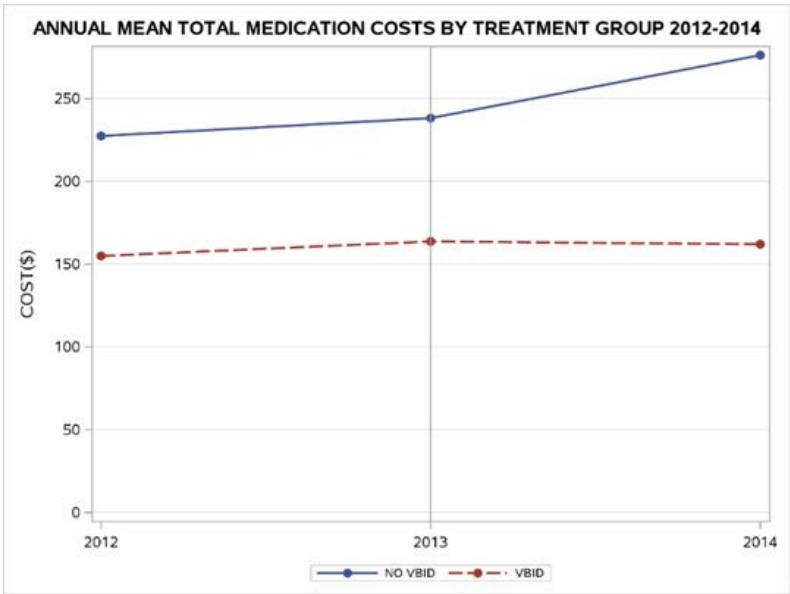

Figure 4. Mean Annual Medication Costs by Exposure Group 2012-2014

Estimating Propensity Weights

Figures 3 and 4 show that the exposed VBID group is paying less per year than the no VBID group at all time points, suggesting that propensity score weighting maybe be in order. Propensity scores can be used in different ways: (1) Matching--use the propensity score variable to match individual reference subjects to exposed subjects, (2) Stratification—run stratified analyses, usually by quintiles of the propensity score, and then aggregate the results for a weighted treatment effect, (3) Weighting-- use the inverse of the propensity scores as weights to estimate the average treatment effect (ATE) or average effect in the treated (ATT) or (4) Covariate adjustment-- include the propensity score as an individual covariate in the D-I-D model. Empirical observations and analyses of Monte Carlo simulations indicate that matching and weighting are more effective at balancing differences than stratification or covariate adjustment (Austin 2011b). It is very important to respect the time ordering of data when generating propensity scores for a D-I-D or ITS analysis: no measurement made after baseline should be included in the propensity score model, including pre- or post-period outcomes of interest.

The process for propensity scoring in this cohort was demonstrated in a previous paper (Warton and Parker, 2018) using the code:

```
TITLE "VBID D-I-D: PROC PSMATCH FOR PROPENSITY SCORES";
PROC PSMATCH DATA=DID_DATA;
   CLASS EXPOSED;
   PSMODEL EXPOSED(TREATED=LAST)=AGE_AT_INDEX FEMALE LOW_SES HISPANIC BLACK
      ASIAN OTHER RACE_MISS ENGLISH COMORBID1-COMORBID3;
   ASSESS PS LPS VAR=(FEMALE WHITE AGE_AT_INDEX LOW_SES HISPANIC BLACK
      ASIAN OTHER ENGLISH COMORBID0-COMORBID3) / VARINFO NLARGESTWGT=6
  PLOTS(NODETAILS)=(CDFPLOT BOXPLOT(DISPLAY=(PS AGE_AT_INDEX))
      STDDIFF(REF=0.10) ) WEIGHT=ATEWGT(STABILIZE=NO) ;
 OUTPUT OUT(OBS=ALL)=GRID.PS_OUT LPS=LPS PS=PS ATEWGT(STABILIZE=NO)=ATE_WT;
RUN;
```

The ATT weights can be calculated from the PROC PSMATCH output with the code:

```
DATA GRID.DID_DATA_PS;
    SET GRID.PS_MODEL_OUT;
    ATT_WT = PS/(1-PS);
RUN;
```

Figure 5 illustrates that propensity score matching process has done a good job removing differences in measured values between the exposed and unexposed groups, with all weighted observations falling within ±0.10 standardized differences of zero and very little difference in the cumulative distribution of propensity scores after weighting.

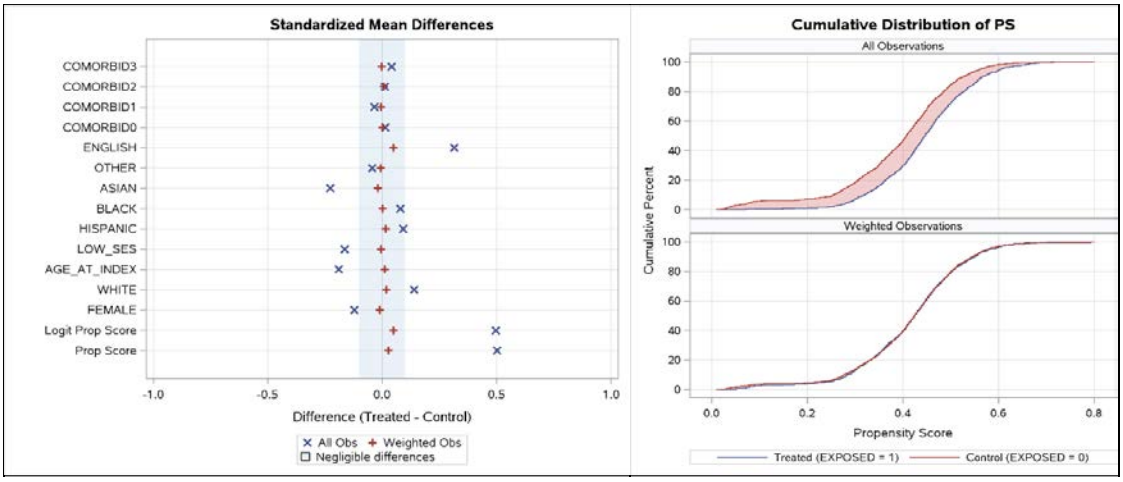

Figure 5. PROC PSMATCH Propensity Score Plots

D-I-D MODEL FOR ANNUAL MEDICATION COSTS

This example D-I-D analysis is intended to answer the questions "How much did the annual out-of-pocket costs for medications change with the switch to a high-deductible plan in 2014?" and "Did having the VBID benefit help offset the increase in out-of-pocket costs for medications after the switch to a high-deductible plan?"

The macro code below runs the D-I-D model for annual medication cost (RXALL_COST) by exposure group using the dataset whose first 10 rows are shown in Table 1. The MODEL statement includes all combination of the POST and EXPOSED variables, and the correlation is allowed to be unstructured. The WEIGHT statement specifies which weights from the propensity score model we wish to use, if any. The ESTIMATE statement requests the D-I-D estimate of the interaction term ($\boldsymbol{\beta}_{interaction}$), while the LSMEANS statement requests the weighted means of the various combinations of EXPOSED and POST for the cohort. The model statement is wrapped in macro code to allow different model specifications to be run quickly, as illustrated below:

```
%MACRO DID_MODELS(WTSTRING =, MODELTYPE=, OUTCOMEVAR= );
  PROC MIXED DATA = DID_DATA_PS order=formatted;
   CLASS POST EXPOSED;
   MODEL &OUTCOMEVAR. =POST|EXPOSED / SOLUTION;
   RANDOM INT/SUBJECT=STUDY_ID TYPE=UN ;
   &WEIGHTSTRING;
   LSMEANS POST|EXPOSED / DIFF;
   ESTIMATE 'D-I-D' EXPOSED*POST 1 -1 -1 1;
   FORMAT EXPOSED EXPOSED_MIXED. POST PREPOST_MIXED.;
   TITLE2 "RANDOM INTERCEPT: PRE/POST &OUTCOMEVAR. BY EXPOSURE GROUP";
   TITLE3 "&MODELTYPE: INCLUDING LEAST-SQUARES MEANS ESTIMATES";
  RUN;
%MEND DID_MODELS;
%DID_MODELS(WSTRING=,MODELTYPE=UNADJUSTED,OUTCOMEVAR=RXALL_COST);
%DID_MODELS(WTSTRING=WEIGHT ATE_WT,MODELTYPE=ATE WT,OUTCOMEVAR=RXALL_COST);
```

Note that if the outcome were binary and relative ratios were of interest, then a random intercept model with repeated measures in the GENMOD procedure with a binomial or Poisson distribution and logit link would be appropriate as discussed above. For this

outcome, absolute differences are more meaningful than relative risk or odds ratios and so PROC MIXED is used for simplicity.

Obs	STUDY_ID	EXPOSED	POST	ATE_WT	RXALL_COST
1	1001	0	0	1.0819	58.30
2	1001	0	1	1.0819	0.00
3	1002	0	0	1.03485	60.10
4	1002	0	1	1.03485	43.30
5	1003	0	0	2.04064	191.80
6	1003	0	1	2.04064	50.05
7	1004	1	0	1.95129	40.00
8	1004	1	1	1.95129	0.00
9	1005	0	0	1.42721	206.60
10	1005	0	1	1.42721	260.28

Table 1. Sample Data for D-I-D Analysis from the DID_DATA_PS Dataset

The *Solution for Fixed Effects* table in Output 1 includes the coefficients of interest from the model. The ATE weighted model estimates are β_0=238.39, β_{post}=38.76, β_{exp}=-73.13, and $\beta_{interaction}$ =-44.34. These coefficients indicate that the no VBID group started 2013 with a (weighted) mean annual medication cost of $238.38, and the VBID group started $73.13 lower, a significant difference. The mean change between pre- and post-periods in the no VBID group was an increase of $38.76, again a significant change. The coefficient on the interaction term is likewise highly significant, indicating that annual costs decreased by

Solution for Fixed Effects

Effect	post	EXPOSED	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			238.39	7.3466	2480	32.45	<.0001
post	(DIFF) 2014		38.7631	7.4065	2480	5.23	<.0001
post	(REF) 2013		0	.	.	.	.
EXPOSED		(DIFF) VBID	-73.1331	10.9406	2480	-6.68	<.0001
EXPOSED		(REF) NO VBID	0	.	.	.	.
post*EXPOSED	(DIFF) 2014	(DIFF) VBID	-44.3422	10.4943	2480	-4.23	<.0001
post*EXPOSED	(DIFF) 2014	(REF) NO VBID	0	.	.	.	.
post*EXPOSED	(REF) 2013	(DIFF) VBID	0	.	.	.	.
post*EXPOSED	(REF) 2013	(REF) NO VBID	0	.	.	.	.

Least Squares Means

Effect	post	EXPOSED	Estimate	Standard Error	DF	t Value	Pr > |t|
post	(DIFF) 2014		218.41	5.4703	2480	39.93	<.0001
post	(REF) 2013		201.82	5.4703	2480	36.89	<.0001
EXPOSED		(DIFF) VBID	162.47	7.2046	2480	22.55	<.0001
EXPOSED		(REF) NO VBID	257.77	6.3450	2480	40.63	<.0001
post*EXPOSED	(DIFF) 2014	(DIFF) VBID	159.68	8.1070	2480	19.70	<.0001
post*EXPOSED	(DIFF) 2014	(REF) NO VBID	277.15	7.3466	2480	37.72	<.0001
post*EXPOSED	(REF) 2013	(DIFF) VBID	165.25	8.1070	2480	20.38	<.0001
post*EXPOSED	(REF) 2013	(REF) NO VBID	238.39	7.3466	2480	32.45	<.0001

Output 1. Fixed Effects Solutions and Least Squares Means from D-I-D Model

$44.34 more in the VBID group than in the no VBID group. Based on these results, we have evidence that the VBID plan more than offset the increased medication costs resulting from the switch to a high deductible plan in the first year.

The *Least Squares Means* results for the ATE weighted model show that mean costs for the no VBID group increased from $238.39 in 2013 to $277.15 in 2014, while the VBID group mean costs actually decreased slightly from $165.25 to $159.68 following the switch to a high deductible plan. Overall, the switch to a higher deductible plan led to an increase in mean medication costs from $201.82 to $218.41 in 2013 to 2014 across both groups, while the difference between VBID and non-VBID groups was $162.47 versus $257.77, respectively, when averaged over the entire study period.

Thus, the answers to our questions are: (1) Annual medication costs increased by $38.67 when switching to a high-deductible plan and (2) the VIBD benefit completely offset the cost increase and resulted in even lower annual medication costs after the switch.

INTERRUPTED TIME SERIES DESIGN AND ANALYSIS

ITS IN PICTURES AND EQUATIONS

As with D-I-D, a (slightly more complex) graphical illustration can be useful in understanding the interrupted time series study design. In their very helpful 2011 paper, Linden and Adams present a basic illustration of the ITS design which is adapted for our particular analytic example in Figure 6 (Linden, Ariel and Adams, John, 2011).The coefficients shown at the top of the figure (β_0 through β_3) represent the pre- and post-interruption trajectories of the unexposed group, while those at the bottom of the figure(β_4 through β_7) represent differences between the exposed and unexposed group trajectories.

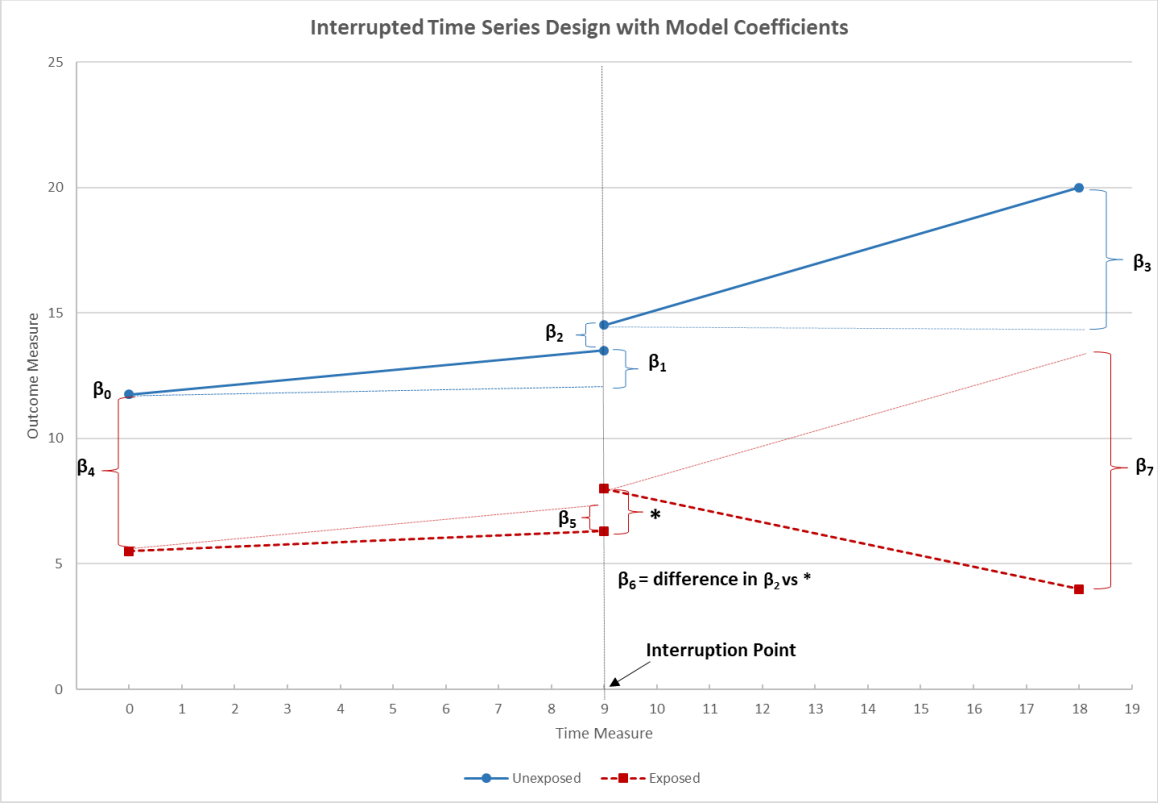

Figure 6. Interrupted Time Series Design with Model Coefficients

The coefficients in Figure 6 represent the following:

β_0 : The predicted mean outcome of the unexposed group at time $t=0$.

β_1 : The difference in outcomes from time t to time $t+1$ before the interruption. In other words, the slope of the outcome in the unexposed group in the pre-interruption period.

β_2 : The amount the outcome "jumps" at the interruption point among the unexposed group, measured as the first post-interruption outcome value minus the last pre-interruption outcome value.

β_3 : The difference in outcomes from time t to time $t+1$ after the interruption. The slope of the outcome in the unexposed group in the post-interruption period.

β_4 : The difference in outcome levels at time $t=0$ in the exposed group compared to the unexposed group.

β_5 : The difference in slope in the exposed group compared to the unexposed group in the pre-interruption period.

β_6 : The difference the outcome "jump" at the interruption between exposed and unexposed groups.

β_7 : The difference in slopes after the interruption between exposed and unexposed groups.

The Single ITS Model

The basic equation for the single group ITS model can modeled as the top portion of Figure 7, where the coefficients are only related to the unexposed group:

$$\mu_{it} = \beta_0 + \beta_1 * \text{Time} + \beta_2 * \text{Interruption} + \beta_3 * \text{Interruption} * \text{Time} + \varepsilon_{it}$$

In the single group ITS model, the only variables required are a time variable marking evenly spaced outcome measures (Time), a dummy variable indicating that a specific time point is before (=0) or after (=1) the interruption (Interruption), and the outcome measure values at each time point during the study period. The Interruption*Time interaction will have values of zero at all time points before the interruption, and the same values as the Time variable following the interruption. As can be surmised by this description, the data for a single group ITS model consists of a long dataset with one observation per time point with the Time, Interruption, and outcome values for that time point. Based on this model, we can answer two main questions about the outcome in the single group: (1) How much did the outcome value change from right before to right after the interruption change? and (2) How much did the outcome trajectory (slope) change after the interruption? The first question is answered by the coefficient on the Interruption variable (β_1), while the second is answered using coefficient on the interaction term Interruption*Time (β_3).

The Robust ITS Model

The equation for the Robust ITS model includes all the coefficients from Figure 7:

$$\mu_{it} = \beta_0 + \beta_1 * \text{Time} + \beta_2 * \text{Interruption} + \beta_3 * \text{Interruption} * \text{Time} + \beta_4 * \text{Exposed} + \beta_5 * \text{Exposed} * \text{Time} + \beta_6 * \text{Exposed} * \text{Interruption} + \beta_7 * \text{Exposed} * \text{Interruption} * \text{Time} + \varepsilon_{it}$$

This version of the ITS Model requires the addition of only one variable to those required by the single ITS model: a binary variable indicating that the cohort member is exposed or not at the end of the study period. Note that Exposed=1 for those who are exposed even before the exposure occurs (i.e.. during the pre-period). As was the case for the single group ITS, the data for this model consist of a long dataset with one observation per time point that includes the Time, Exposed, Interruption, and outcome values at that specific outcome measure timepoint. With the coefficient estimates from this model, we can answer several important questions: (1) How much did the outcome value change from right before to right

after the interruption change for both groups, (2) is that change different between groups, (3) did the outcome trajectory (slope) change after the interruption compared to before the interruption, and (4) is that slope change different between the groups? We can also estimate the outcome at other specific timepoints during the study period.

SAS® MODELING METHODS FOR INTERRUPTED TIME SERIES

Repeated Measures Linear Regression Model (The MIXED Procedure)

As with the D-I-D model for continuous outcomes, ITS models can be run using PROC MIXED in SAS. The model is built in a similar manner, with the outcome measure on the left of the MODEL statement and the variables names for the Time, Exposure, and Interruption values along with their various interaction terms on the right. PROC MIXED provides estimated model parameters, allows for propensity score weighting, flexible correlation structures, and estimates from the LSMEANS and ESTIMATE statements.

Repeated Measures Non-Linear Models (The GENMOD Procedure)

When an outcome of interest is binary, the measurement timepoint group means represent the proportion of individuals in each group experiencing the outcome at each of the study time points. These proportions can be estimated using PROC GENMOD as odds or probabilities on the log scale (relative risks) or as proportions (absolute risks) for each group at each time point. Factors that influence the choice of a continuous or binary outcome for an ITS analysis are similar to those outlined for the D-I-D model.

Repeated Measures Autoregressive Models (The AUTROEG and ARIMA Procedures)

When autoregressive correlation is a suspected issue in an analysis − when current outcome measures can predict future outcome measures − the SAS AUTOREG procedure can be used for Single ITS designs, and the ARIMA procedure can be used for Robust ITS modeling. Both procedures account for autocorrelation between repeated measures and can be used to model ITS under certain data constraints. This analytic option will not be discussed in this paper, but suggested references for these procedures include Conell and Flint, 2016 and Dickey, 2019.

ITS ANALYSIS: INSURANCE BENEFIT CHANGES AND ANNUAL MEDICATION COSTS

ITS Data Considerations

When approaching an ITS analysis, it is important to consider factors that might affect the underlying data in ways that could bias results. First the timing of the interruption must be known, and the impact of the interruption must be close to instantaneous. In our example, high-deductible plans became active on 1/1/2014, and so a patient filling a prescription on 12/31/2013 might pay a lower cost than if filling the prescription on 1/2/2014. If, however, there was a month or two delay in charging the new costs after the first of the year, the interruption would not be instantaneous. In this situation, removing the period of the implementation phase from the analysis may be justified.

Another consideration when planning an ITS analysis is the number of outcome measures in the pre- and post-periods and how they are arranged in time. The ITS design requires that outcome measures be evenly spaced in time and a sufficient number of observations before and after the interruption. One rule of thumb is 8-9 minimum number of measures before and after the interruption, although some earlier authors recommend no fewer than 48 total measures. Also, having an equal number of measures before and after the interruption maximizes the power of the basic ITS design(Conell and Flint, 2016).

Another important factor to consider is whether the outcome displays any cyclic patterns when measured over time. Patterns such as seasonality will need to be addressed when choosing which time periods to include in the analysis. For annual seasonality, for example,

the pre- and post-period measures could be matched by month to remove any spurious findings generated by differing outcome trajectories over the course of the year. Autoregression, the situation where knowing the value of the outcome at time *t* provides information about the likely outcome at time *t+1*, can also lead to bias in ITS analyses, and testing for the presence of autoregression should be part of any ITS analysis (see Connel and Flint, 2016).

Outcome Measure and Graphical Analysis for ITS

The outcome measure for the ITS analysis examples is the three-month rolling average of monthly medication costs (RXCOST_MAS3MOS). Monthly costs fluctuated in the first few months of the VBID program as pharmacy systems were updated, so reduce the chances of spurious results due to the implementation phase, and to address potential seasonality in our outcome measure, the study period for this analysis includes the nine months prior to the interruption (4/1-12/31/2013) and the nine months starting in the fourth month after the interruption (4/1-12/31/2014).

Plots of actual monthly medication costs for 30 random VBID and no VBID cohort members in Figure 7 show the expected spikes corresponding to a 90-100 day fill pattern. The smoothed data include fewer spikes and indicate that the VBID cohort members are paying less for medications before the switch to the non-deductible plan. This plot also suggests that the general trend in medication costs was a slight rise over time for both groups, and that costs jumped for no VBID members at the deductible switch point (month 13), and then remained roughly flat. For the VBID group, the plot suggests that costs jumped for some members, but in general, costs decreased during the post-period.

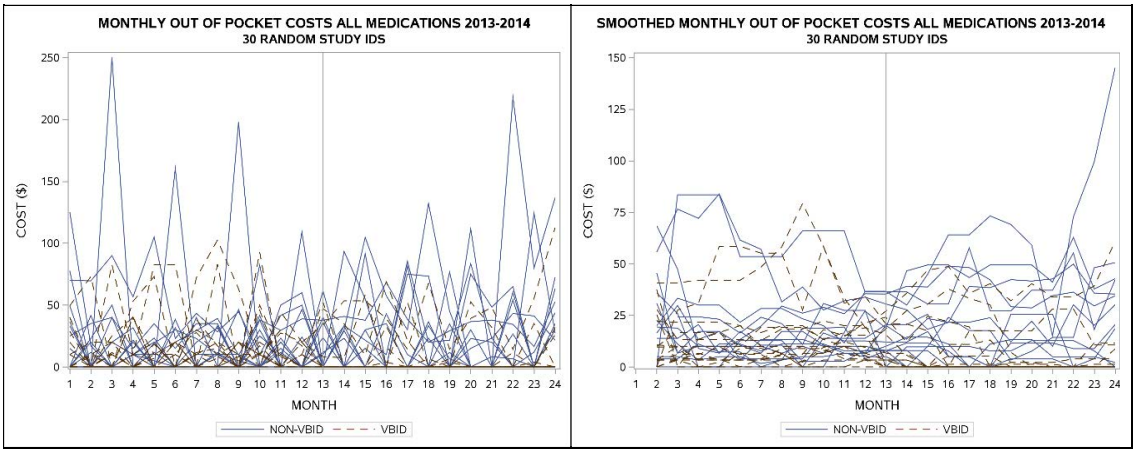

Figure 7. Monthly and Smoothed Monthly Medication Costs in 2013-2014

In Figure 8, a plot of mean out-of-pocket medication costs by group for the pre- and post-periods, these trends are confirmed. Removing the instability in costs in the first three months of the implementation phase and including nine equally spaced measures in the both pre- and post-interruption periods meets the basic requirements for an ITS analysis. More measures in both periods would be preferable, of course, and based on Figure 8, it is possible that the implementation phase lasted longer than 3 months; there is some evidence that the decline in monthly medication costs slows after month 19 for the VBID group, for example.

The ITS model fits separate regression lines through the outcome measure values before and after the interruption point. Because the model is a random intercept model, the slope for each individual is constrained to match those within that individual's treatment group. A single ITS model compares the pre- and post-interruption slope values along with the change in intercepts at the interruption point, while ITS model with an unexposed group compares the slopes and intercepts of all four trajectory regression lines.

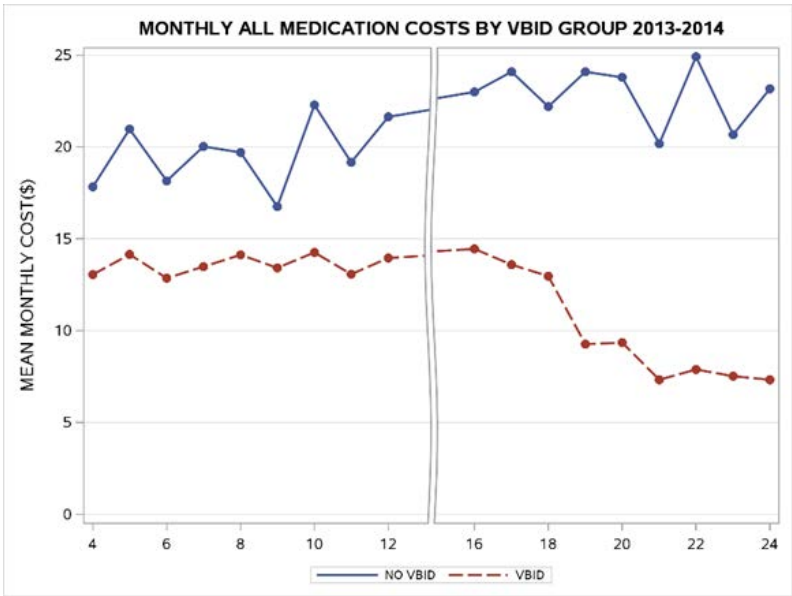

Figure 8. Mean Monthly Costs for All Medications for Pre- and Post-Interruption Study Periods

ITS Single Group Analysis and Monthly Out-of-Pocket Medication Costs

The basic single ITS model will use data from the VBID group with the RXALL_MA3MOSS outcome. Sample data for this analysis are displayed in Table 2.

Obs	STUDY_ID	MONTHNUM	GAP_MONTHNUM	INTERVENTION	RXALL_MA3MOS
1	1010	4	1	0	0.00
2	1010	5	2	0	23.33
3	1010	6	3	0	26.67
4	1010	7	4	0	33.33
5	1010	8	5	0	23.33
6	1010	9	6	0	23.33
7	1010	10	7	0	23.33
8	1010	11	8	0	13.33
9	1010	12	9	0	20.00
10	1010	16	10	1	24.37
11	1010	17	11	1	20.40
12	1010	18	12	1	20.40
13	1010	19	13	1	20.40
14	1010	20	14	1	20.40
15	1010	21	15	1	33.08
16	1010	22	16	1	29.21
17	1010	23	17	1	25.84
18	1010	24	18	1	6.35
19	1011	4	1	0	26.85
20	1011	5	2	0	16.67
21	1011	6	3	0	3.33
22	1011	7	4	0	6.67
23	1011	8	5	0	6.67
24	1011	9	6	0	16.35

Table 2. Sample Data for Single ITS Analysis from the SINGLE_ITS_DATA Dataset

The key variables are STUDY_ID, which identifies the units for repeated measures, GAP_MONTHNUM, which is the time variable for our specific study period,

INTERVENTION, the Interruption variable which flags time periods after the interruption, and the outcome variable RXALL_MA3MOS. This model will answer two primary questions: (1) How much did the switch to a high-deductible plan with a VBID change the monthly cost of medications, and (2) did the monthly rate of change in medications costs change after the switch?

The hypothesis is switching to a high-deductible plan increases monthly medication costs overall, but the VBID plan would help to offset that increase because patients receive certain medications at no cost.

We used PROC MIXED for this analysis, as it allows for repeated measures with random intercepts and differing correlation structures with continuous outcomes. Below is the SAS code to generate the single ITS model coefficient estimates:

```
PROC MIXED DATA = SINGLE_ITS_DATA METHOD=ML
  PLOTS(MAXPOINTS=20000)=(RESIDUALPANEL(UNPACK) VCIRYPANEL(UNPACK) );
  CLASS STUDY_ID;
  MODEL RXALL_MA3MOS=GAP_MONTHNUM INTERVENTION GAP_MONTHNUM*INTERVENTION/S;
  REPEATED / SUBJECT = STUDY_ID  TYPE =  UN;
  ESTIMATE 'PRE-PERIOD SLOPE' GAP_MONTHNUM 1 / CL E;
  ESTIMATE 'POST-PERIOD SLOPE' GAP_MONTHNUM 1 GAP_MONTHNUM*INTERVENTION 1
  / CL E;
  ESTIMATE 'PRE/POST GAP' GAP_MONTHNUM 1 INTERVENTION 1
  GAP_MONTHNUM*INTERVENTION 10 /CL E;
RUN;
```

This code requests various residual fit plots, specifies an unstructured correlation structure, and generates estimates of the slopes in both periods, as well as an estimate of the "jump" due to the intervention.

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.6732	0.4722	1023	26.84	<.0001
GAP_MONTHNUM	0.07958	0.04536	1023	1.75	0.0797
INTERVENTION	10.4079	2.4161	1023	4.31	<.0001
GAP_MONTH*INTERVENTI	-0.9856	0.1545	1023	-6.38	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
PRE-INTERVENTION SLOPE	0.07958	0.04536	1023	1.75	0.0797	0.05	-0.00943	0.1686
POST-INTERVENTION SLOPE	-0.9060	0.1460	1023	-6.20	<.0001	0.05	-1.1926	-0.6195
PRE/POST GAP	0.6314	1.0235	1023	0.62	0.5374	0.05	-1.3770	2.6399

Output 2. Single Group ITS Model Coefficients and Linear Combination Estimates

The coefficient estimates from this model in the "Solution for Fixed Effects" table in Output 2 estimate a starting monthly cost at time t=0 of $12.67 ($\beta_0$), and a monthly cost increase of $0.08 ($\beta_1$), although this rate may not be different from zero (p-value = 0.08). The interruption moved the monthly cost up by $10.41 ($\beta_2$), while the cost per month in the post-period decreased substantially, by $0.99 ($\beta_3$), with both estimates highly significant.

Of note in this model specification, the time variable GAP_MONTHNUM in our dataset is a sequential count of months from 1 to 18, with the interruption occurring between months 9 and 10. To calculate the change due to the interruption alone, the final estimate statement calculates the linear combination for the net change between months 9 and 10, which is β_1*(10-9) + β_1*1+ β_1*10*1 or $0.08 + 10.41-9.86 = $0.63.

The answers to the two primary questions are: (1) The switch to a high-deductible plan increased average monthly costs by $0.63, but the change is likely not different from zero (p-value 0.54), and (2) the monthly rate of change in medication costs moved from essentially flat before the switch to a *decrease* of nearly $1.00 per month after the VBID was introduced with the high-deductible plan.

A commonly used alternative data specification for ITS models uses a variation on the time measure for the period after the interruption, re-starting the count of time units at the first time unit after the interruption as "post-interruption time"=1. In this case, the model specification is the same, but the ESTIMATE statements would need to change for the pre- to post-interruption gap. While the coefficients change, the final results do not.

Robust ITS Analysis and Monthly Out-of-Pocket Medication Costs

The data structure for the two-group ITS analysis is identical to that presented in Table 2, except that a binary variable is added to identify the exposure group. If propensity score or other weighting is desired, the weight variable(s) would also be added. For the example analysis, the VBID group is designated in the variable EXPOSED (VBID = 1), and both ATE and ATT weights are included in the modeling dataset. The ATE-weighted model had slightly better fit characteristics, so these estimates are used to report results.

This model will answer several questions: (1) How much did the monthly cost of medications change from right before to right after the switch to a high-deductible plan? (2) Is this jump amount different between the VBID and no VBID groups? (3) Did the rate of change in monthly medication costs change after the high-deductible plan switch? And (4) is the rate of change in costs after the switch different for VBID vs no VBID groups?

```
PROC MIXED DATA = ROBUST_ITS_DATA METHOD=ML
  PLOTS(MAXPOINTS=60000)=(RESIDUALPANEL(UNPACK) VCIRYPANEL(UNPACK));
  CLASS STUDY_ID EXPOSED INTERVENTION;
  MODEL RXALL_MA3MOS = GAP_MONTHNUM INTERVENTION INTERVENTION*GAP_MONTHNUM
                       EXPOSED EXPOSED*GAP_MONTHNUM EXPOSED*INTERVENTION
                       EXPOSED*INTERVENTION*GAP_MONTHNUM/ S ;
  WEIGHT ATE_WT;
  REPEATED / SUBJECT = STUDY_ID  TYPE = UN  R;
  FORMAT EXPOSED EXPOSED_MIXED. INTERVENTION PREPOST_MIXED.;
RUN;
```

The model results from SAS for the ATE weighted analysis are shown in Output 3. The *Solution for Fixed Effects* section of the SAS output displays coefficient estimates and significance levels. Based on these estimates, the VBID group began the study with a mean monthly cost $5.77 lower than the $18.65 starting point for the no VBID group, and the pre-interruption period increase in monthly costs was higher in the no VBID group ($0.13) than the VBID group($0.13-0.09=$0.04), although the difference is not significant. The decrease in monthly costs after the switch to the high-deductible plan was quite small for the no VBID group ($0.13-0.15=-$0.02), and not significant, an indication that monthly medication costs held steady after the switch in this group. The switch to the high deductible plan had a much larger impact on the VBID group compared to the no VBID group ($8.36, highly significant). However, the monthly medications costs decreased significantly (-$0.94) per month for the VBID group, and this, combined with a lower starting cost, led to lower costs for medications in the VBID group over the entire study period.

A plot of the estimated VBID and no VBID mean monthly costs based on these coefficients at each monthly time point during the study shows the monthly costs for VBID group members rising from $12.88 to $13.23 in the pre period, followed by a jump to $14.69 at

Solution for Fixed Effects							
Effect	EXPOSED	INTERVENTION	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			18.6467	0.5208	2480	35.80	<.0001
GAP_MONTHNUM			0.1272	0.04952	2480	2.57	0.0103
INTERVENTION		(DIFF) 2014	3.9587	1.9948	2480	1.98	0.0473
INTERVENTION		(REF) 2013	0	.	.	.	.
GAP_MONTH*INTERVENTI		(DIFF) 2014	-0.1482	0.1335	2480	-1.11	0.2671
GAP_MONTH*INTERVENTI		(REF) 2013	0	.	.	.	.
EXPOSED	(DIFF) VBID		-5.7711	0.7379	2480	-7.82	<.0001
EXPOSED	(REF) NO VBID		0	.	.	.	.
GAP_MONTHNUM*EXPOSED	(DIFF) VBID		-0.08749	0.07017	2480	-1.25	0.2126
GAP_MONTHNUM*EXPOSED	(REF) NO VBID		0	.	.	.	.
EXPOSED*INTERVENTION	(DIFF) VBID	(DIFF) 2014	8.3547	2.8264	2480	2.96	0.0031
EXPOSED*INTERVENTION	(DIFF) VBID	(REF) 2013	0	.	.	.	.
EXPOSED*INTERVENTION	(REF) NO VBID	(DIFF) 2014	0	.	.	.	.
EXPOSED*INTERVENTION	(REF) NO VBID	(REF) 2013	0	.	.	.	.
GAP_MO*EXPOSE*INTERV	(DIFF) VBID	(DIFF) 2014	-0.9414	0.1892	2480	-4.98	<.0001
GAP_MO*EXPOSE*INTERV	(DIFF) VBID	(REF) 2013	0	.	.	.	.
GAP_MO*EXPOSE*INTERV	(REF) NO VBID	(DIFF) 2014	0	.	.	.	.
GAP_MO*EXPOSE*INTERV	(REF) NO VBID	(REF) 2013	0	.	.	.	.

Output 3. Results of Robust ITS Model and ATE weights

the benefit switch, and then a sharp decrease to $6.29 per month by the end of the study period. In contrast, while the no VBID group experienced a similar climb from $18.65 to $19.79 in the pre period with a slightly larger jump to $22.40 at the beginning of the post period, the no VBID group experienced only a very small decrease in costs after the interruption, ending the study period at $22.23 per month for medications.

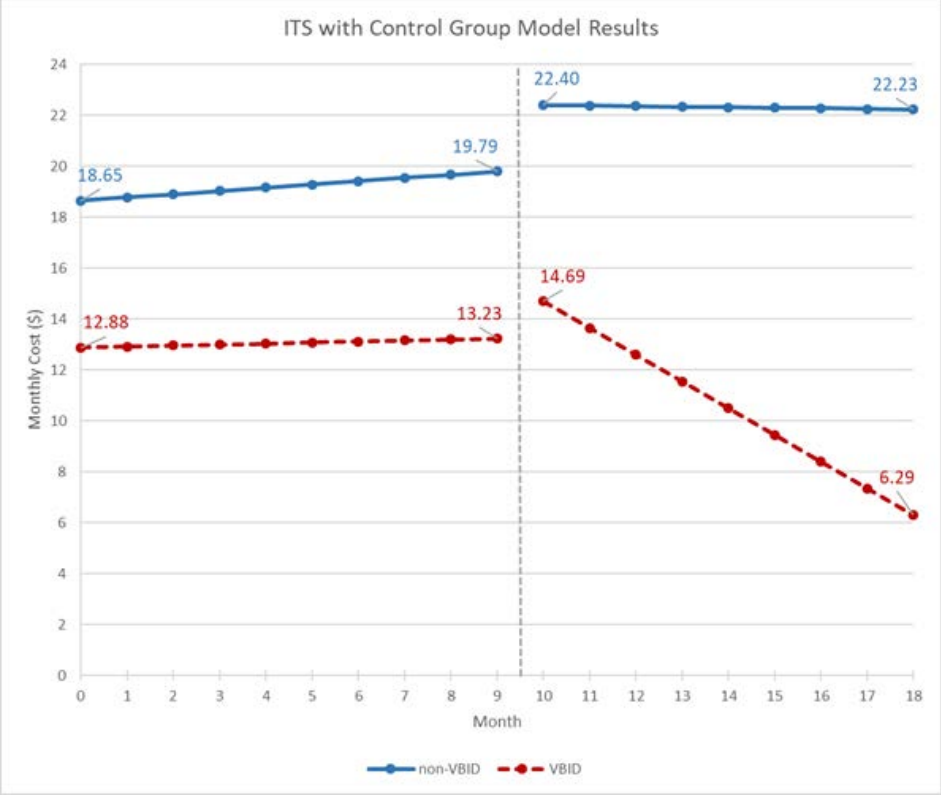

Figure 9. ITS Model Estimates Plot of Monthly Medication Costs

Non-Panel Data and ITS Models

The example data used in this paper consists of panel data - repeated measures over time on the same individuals. However, many ITS analyses are performed on summarized measures which do not include the same individuals over time. It is possible to analyze time series data on dynamic cohorts in a similar fashion, so long as sufficient numbers of evenly spaced measures are available in the pre- and post-interruption periods, and a single interruption point is clearly identifiable. For example, after creating a dataset containing the mean values for the smoothed monthly medication costs, a single ITS analyses on the VBID group could be performed with any of the analysis procedures for single group Robust ITS.

CONCLUSION

Difference-in-differences and Interrupted Time Series study designs are important tools in analyzing the impacts of operational changes made in healthcare delivery. The causal nature of these study designs, along with the ability to generate data retrospectively from large administrative databases, makes them a logical choice for analyzing "natural experiments" that result from operational interventions or changes not made in coordination with research partners. These types of studies have long been used in other research fields, especially economics, and this paper is intended to further their adoption by healthcare and epidemiologic researchers by outlining the basic concepts and providing examples that can be used by analysts using these methods for the first time.

REFERENCES

Austin, P. C. 2009. 'Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples', *Stat Med*, 28: 3083-107.

Castelloe, John. 2014. 'Power and Sample Size for MANOVA and Repeated Measures with the GLMPOWER Procedure, Paper SAS030-2014.' *In SAS Global Forum 2014 Conference*, Washington, D.C.: SAS Institute Inc.

Connel, Carol and Flint, David. 2016. 'Did the Protocol Change Work? Interrupted Time Series Evaluation for Health Care Organizations.' *SAS Conference Proceedings: Western Users of SAS Software 2016*, San Jose, CA : Western Users of SAS Software. Available at https://www.lexjansen.com/wuss/2016/14_Final_Paper_PDF.pdf

Dickey, David. 2019. "The Best of Both Worlds: Forecasting Using Time Series with Inputs, Paper 118-2019.' *In SAS Conference Proceedings: SouthEast SAS Users Group 2019*, Williamsburg, Virginia: SouthEast SAS Users Group. Available at https://www.lexjansen.com/sesug/2019/SESUG2019_Paper-118_Final_PDF.pdf

Linden, Ariel and Adams, John. 2011. 'Applying a Propensity Score-Based Weighting Model to Interrupted Time Series Data: Improving Causal Inference in Programme Evaluation', *Journal of Evaluation in Clinical Practice*,17: 1231-1238.

Reed, M. E., E. M. Warton, E. Kim, M. D. Solomon, and A. J. Karter. 2017. 'Value-Based Insurance Design Benefit Offsets Reductions In Medication Adherence Associated With Switch To Deductible Plan', *Health Affairs* (Millwood), 36: 516-23.

Spiegelman, D., and E. Hertzmark. 2005. 'Easy SAS calculations for risk or prevalence ratios and differences', Am J Epidemiol, 162: 199-200.

Warton, E Margaret and Parker, Melissa. 2018. "Oops, I D-I-D It Again! Advanced Difference-in-Differences Models in SAS®." *SAS Conference Proceedings: Western Users of SAS Software 2018*, Sacramento, CA : Western Users of SAS Software. Available at https://www.lexjansen.com/wuss/2018/25_Final_Paper_PDF.pdf

Wolfinger, Russ and Chang, Ming. 1995. "Comparing the SAS® GLM and MIXED Procedures for Repeated Measures." Proceedings of the SAS Global Users Group 1995 Conference 20. Cary, NC: SAS Institute, Inc. Available at https://support.sas.com/rnd/app/stat/papers/abstracts/mixedglm.html

Zou, G. 2004. 'A modified Poisson regression approach to prospective studies with binary data', Am J Epidemiol, 159: 702-6.

ACKNOWLEDGMENTS

Many thanks to Dr. Mary Reed for allowing me to use the data from the VBID Pharmacy Medication study in the paper and to Melissa Parker and Wendy Dyer for their help in preparing this manuscript. The VBID study was funded by the Kaiser Permanente Northern California Community Benefits Program.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Margaret Warton
Kaiser Permanente Division of Research
510-891-3129
Margaret.M.Warton@kp.org