

Abstract Title Page
Not included in page count.

Title: A Unified Framework for Estimating Minimum Detectable Effects for Comparative Short Interrupted Time Series Designs

Authors and Affiliations:

Cristofer Price, Abt Associates (e-mail: Cristofer_Price@Abtassoc.com)

Fatih Unlu, Abt Associates (e-mail: Fatih_Unlu@Abtassoc.com)

Abstract Body

Limit 4 pages single-spaced.

Problem / Background / Context:

Description of the problem addressed, prior research, and its intellectual context.

The Comparative Short Interrupted Time Series (C-SITS) design is a frequently employed quasi-experimental method, in which the pre- and post-intervention changes observed in the outcome levels of a treatment group is compared with those of a comparison group where the difference between the former and the latter is attributed to the treatment. The increase in the availability and quality of extant data (e.g., state test scores, graduation rates, and college application rates in primary and secondary education and cognitive, language, and socio-emotional assessments in pre-school settings) has made the use of C-SITS designs a more viable option for assessing the impacts of interventions. Despite the recent growth in its use, the existing resources on how to estimate minimum detectable effects for this design are still very limited. One such resource is Schochet (2008) which shows that the variance of the difference-in-difference estimator (which can be considered as a special application of C-SITS) critically depends on sample sizes and the cluster-level (if applicable) and individual-level correlations between the pre- and post-test outcome measures. Extending Bloom (1999 and 2003), Dong and Maynard (2013) consider a particular application of the C-SITS model which includes separate linear time trends for the treatment and comparison group and the treatment effect is estimated separately for each follow-up year. They show that the variance of this C-SITS estimator depends on (i) sample sizes, (ii) number of baseline years, (iii) follow-up year of interest, (iv) the proportion of outcome variance that lies across successive cohorts of treatment and comparison units (i.e., cohort-level intra-class correlation), and (v) how much of this variance is explained by covariates included in the model. It is important to note that these studies model the treatment effect as fixed (i.e., it is not assumed to vary across treatment units). Two limitations of the existing research on this topic are the unavailability of:

- Plausible values one can use for these critical parameters in the design stage of a study;
- Variance formulae for alternative C-SITS specifications such as models (i) with year fixed effects in lieu of group-specific time trends, (ii) that estimate an average impact estimate across all follow-up years, (iii) with cluster-level data only (as opposed to models with individual-level data nested in clusters), (iv) with various forms of baseline projections, and (v) that assume random treatment effects.

Purpose / Objective / Research Question / Focus of Research:

Description of the focus of the research.

The proposed paper aims to address the aforementioned limitations by (i) deriving expressions for the variance of the various C-SITS estimators and (ii) providing plausible values for the critical variance parameters calculated using school-level test scores from state assessments. Both of these analyses are underway and below we describe our preliminary findings.

Improvement Initiative / Intervention / Program / Practice:

Description of the improvement initiative or related intervention, program, or practice.

Not Applicable (Since this proposal is submitted under research methodology)

Setting

(Description of the research location and partners involved, if applicable.)

Not Applicable (Since this proposal is submitted under research methodology)

Population / Participants / Subjects:

Description of the participants in the research: who, how many, key features, or characteristics.

Not Applicable (Since this proposal is submitted under research methodology)

Research Design:

Description of the research design.

Without loss of generality, let us consider the application of the C-SITS design to estimate the impact of a school-level intervention on school-level test scores. Assume that schools are indexed by k and there are $K1$ treatment units and $K2$ comparison units. Further assume that time is indexed by j that there are repeated cross-sectional data are available for treatment and comparison units for $J1$ pre-treatment periods and $J2$ post-treatment periods (e.g., 4th grade average test scores available between 2006-07 and 2012-13 school years and where 2010-11 is the first year in which the intervention being examined was implemented). To simplify the presentation, we start with the following model specification implementing the simplest application of the C-SITS design, which is also known as the “baseline means projection model” or “difference-in-differences” specification (Bloom, 2003; Somers, Zhu, Jacob, Bloom, 2013):

$$(1) Y_{jk} = \beta_1(TrtGrp_k * TrtYr_{jk}) + \beta_2(TrtYr_{jk}) + \sum_{k=1}^{K1+K2} \alpha_k Sch_k + \sum_{p=1}^P \lambda_p X_p + \varepsilon_{jk}$$

where,

- Y_{jk} is the j^{th} observation on school k ,
- $TrtGrp_k$ = 1 if school is an intervention (treatment) school, 0 if comparison school
- $TrtYr_{jk}$ = 1 if observation in year j is an a post-treatment year, =0 if pre-treatment year
- X_p are other model covariates
- Sch_k are fixed dummy variables for schools
- ε_{jk} residual for j^{th} observation on school k , assumed distributed $N(0, \sigma^2)$

The coefficient $\hat{\beta}_1$ is the estimate of the treatment effect, which is the pooled effect across all treatment schools and post-treatment years. We show that the minimum detectable effect size for this parameter (assuming the outcome measure is standardized to have unit standard deviation) can be characterized as:

$$(2) MDES = (t_{\alpha/2} + t_{\beta})SE(\hat{\beta}_1)$$

where $t_{\alpha/2}$ and t_{β} are quantiles from a t-distribution. For a two-tailed tests with alpha-level criterion at the usual $\alpha = 0.05$, and if degrees of freedom are large, the value of $t_{\alpha/2} = 1.96$, and

with 80% power $t_\beta = 0.84$. $SE(\hat{\beta}_1)$, (or equivalently, $\sqrt{Var(\hat{\beta}_1)}$), is the standard error of the treatment estimate, which is given by:

$$(3) \text{Var}(\beta_1) = \frac{(1 - (R_{Y|Impact Model}^2))(AC)}{(df)\bar{T}(1 - \bar{T})(1 - R_{Predictor| All other terms in impact model}^2)}$$

where:

- $R_{Y|Impact Model}^2$ is the r-squared from the impact model shown in equation 1.
- AC is a design effect for autocorrelation. If there is autocorrelation present in the data it will inflate the variance of the treatment effect. For simplicity, we assume that there is zero autocorrelation and set this to 1.
- df is the “error degrees of freedom” from the model shown in equation 1. These degrees of freedom can be obtained as the total number of observations (i.e., all years, all schools), minus the number of terms in the model including the intercept, if present.
- \bar{T} is the proportion of the observations for which the predictor variable, which is defined as the independent variable that yields the treatment effect of interest, equals one. Specifically, for the impact model shown in Equation 1, it is the proportion of observation where $TrtGrp_k * TrtYr_{jk} = 1$.
- $R_{Predictor| All other terms in the impact model}^2$ is a measure of the squared correlation between the predictor variable and all of the other terms on the right hand side of impact model shown in equation 1. Specifically, it is the r-squared from the model $(TrtGrp_k * TrtYr_{jk}) = \beta_1(TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \sum_{p=1}^P \lambda_p X_p + \varepsilon_{jk}$

We also show that the variance formula in equation 3 can be characterized heuristically as:

$$(4) \frac{(1 - (R_{Y|TG*TYr}^2 + R_{Y|Sch(TG*TYr)}^2 + R_{Y|TrtYr(Sch,TG*TYr)}^2 + R_{Y|X(TrtYr,Sch,TG*TYr)}^2))(AC)}{(df)\bar{T}(1 - \bar{T})(1 - R_{TG*TYr|Sch}^2 + R_{TG*TYr|TrtYr(Sch)}^2 + R_{TG*TYr|X(TrtYr,Sch)}^2)}$$

Due to lack of space, we cannot provide detailed explanations of the terms in equation 4 in the main text of the proposal and simply note that the R^2 terms in this equation break down the R^2 terms in the numerator and denominator of equation 3 into more manageable components (please see table 1 in the appendix for the detailed description). As discussed in our paper, the R^2 terms in the numerator of equation 4 (or the R^2 term in the numerator of equation 3) are related to how much the outcome measure varies among schools and observation periods. If, during the design stage of a project, data from pre-treatment years are available, estimates for these terms can be obtained from those data. Also, as noted in our paper, the R^2 terms in the denominator of equation 4 are functions of the design and depend on the ratio of the treatment units to comparison units and the ratio of pre-treatment years to post-treatment years. Since these values do not depend on the outcome data, an analyst can calculate these quantities using a simulated dataset that is generated based on the intended values of the treatment/comparison and pre-treatment/post-treatment ratios.

The proposed paper derives MDES formulae for various extensions of the simple model specification in equation 1 including those that include (1) year fixed effects; (2) time trends common to the treatment and comparison units; (3) group-specific time trends; and (4) impacts that are averaged over all intervention years; and (5) separate impacts for each intervention year. The paper discusses the implications of these alternative modeling strategies for power. We also provide a generalized MDES formula which can be used for other implementations of the C-SITS design.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data or use of existing databases.

We will use school-level state assessment data obtained from school report card databases from New Jersey, California, and Texas to provide plausible values for the R^2 terms shown in equation 4 (as well as the MDES formula derived for alternative C-SITS specifications). Additionally, we will show the correspondence between the standard error of the impact estimates obtained from fitting models to actual data, and the approximate standard errors that are estimated using the formulae presented in the paper. The data collection and analysis is currently underway and will be completed in time for the conference.

Findings / Outcomes:

Description of the main findings or outcomes, with specific details.

Table 2 presents preliminary plausible parameter values for the simple C-SITS model in equation 1 and the corresponding MDES formulae in equations 3 and 4 under different combinations of the treatment/comparison and pre-treatment/post-treatment ratios. We will produce similar tables for alternative model specifications implementing different C-SITS designs.

Conclusions:

Description of conclusions, recommendations, and limitations, based on findings.

We find that when appropriate plausible values are entered into the formulae provided in our paper that there is a close correspondence between the standard errors estimated from our formulae and those obtained from fitting models to actual data. We therefore conclude that the formulae are behaving as they should. The formulae we provide can be easily programmed using widely accessible software such as Excel or R, or using statistical packages such as SAS, Stata or SPSS. The formulae and plausible values in our paper will provide analysts with a flexible basis for estimating MDES for various C-SITS designs, and will serve as a template for accumulation of relevant information can be used to build databases of plausible values that inform the designs of future studies.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Bloom, H. S. (1999). Estimating Program Impacts on Student Achievement Using “Short” Interrupted Time Series. MDRC Working Papers on Research Methodology. This paper was prepared for the 1999 Annual Meeting of the American Educational Research Association in Montreal, Canada.

Bloom, H.S. (2003). Using “Short” Interrupted Time Series-Analysis to Measure the Impacts of Whole-School Reforms. *Evaluation Review* 27(1): 3-49.

Dong, N., & Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6(1), 24-67.

Schochet, P. Z. (2008). The Late Pretest Problem in Randomized Control Trials of Education Interventions (NCEE 2009-4033). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Appendix B. Tables and Figures

Not included in page count.

Table 1. Description of the Parameters in Equation 3

$R_{Y TG*TYr}^2$	<p>is the proportion of variance of the outcome explained by the predictor variable. It is the r-squared from the model:</p> $Y_{jk} = \beta_0 + \beta_1(TrtGrp_k * TrtYr_{jk}) + \varepsilon_{jk}$ <p>During the design phase calculations of MDESs, investigators should enter the value zero for this term in the formula as they would presumably be testing the null hypothesis that the treatment effect is zero. However, when using real data to generate plausible values for future studies, it is important to account for any variation in the outcome measure attributable to the treatment effect. This can be accomplished by using a non-zero value for $R_{Y TG*TYr}^2$ along with σ_Y^2 in the MDES formula above, or by omitting the term $R_{Y TG*TYr}^2$ but substituting in a pooled treatment and comparison group variance term in place of σ_Y^2 in the MDES formula provided above.</p>
$R_{Y Sch(TG*TYr)}^2$	<p>This is the proportion of total variance that is accounted for by adding terms for schools to the model. This is the semipartial r-squared for schools¹. This term is equal to the r-squared from a larger model that includes the predictor variable and school fixed effects minus the r-squared from the smaller model which includes the predictor variable ($R_{Y TG*TYr}^2$ as described above).</p> <p>Larger model: $Y_{jk} = \beta_1(TrtGrp_k * TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk}$</p> <p>Smaller model: $Y_{jk} = \beta_0 + \beta_1(TrtGrp_k * TrtYr_{jk}) + \varepsilon_{jk}$</p> <p>In the design phase, when investigators are conceptualizing $R_{Y TG*TYr}^2$ as being equal to zero, they can conceptualize $R_{Y Sch(TG*TYr)}^2$ as being the proportion of total variance that is between-schools, while the remaining variance can be conceived of as variation over time within-schools.</p>
$R_{Y TrtYr(Sch,TG*TYr)}^2$	<p>This is the proportion of the variance in the outcome that is explained by adding the <i>TrtYr</i> variable to the model that already includes the predictor variable and school fixed effects. This term is equal to the r-squared from a larger model minus the r-squared from a smaller model, where the larger model includes the predictor variable, school fixed effects, and the indicator for post-treatment years while the smaller model only includes the predictor variable and school fixed effects. That is:</p>

¹ The terminology “semipartial r-squared” comes from Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale NJ.

	<p>Larger model: $Y_{jk} = \beta_1(TrtGrp_k * TrtYr_{jk}) + \beta_2(TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk}$</p> <p>Smaller model: $Y_{jk} = \beta_1(TrtGrp_k * TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk}$</p> <p>This term denotes the proportion of variance that is explained by a term that allows there to be a different mean of the outcome in post-treatment years relative to the pre-treatment years (due to factors other than the treatment itself). Typically, in the design phase, investigators using Equation 1 as their outcome model would assume that comparison school means would be unchanged between the pre-treatment and post-treatment years, and that, in the absence of the intervention the same would be true of the treatment schools, and therefore that the semipartial r-squared for this term would be zero.</p>
$R_{Y X(TrtYr,Sch,TG*TYr)}^2$	<p>This is the proportion of the variance of the outcome measures that is explained by adding any remaining terms (e.g., covariate Xs) to the model that already includes the predictor variable, school fixed effects and the indicator for post-treatment years. This term is equal to the r-squared from the full model in Equation 1 minus the r-squared from a smaller model that does not include the other covariates.</p>
AC	<p>is a design effect for autocorrelation. If there is autocorrelation present in the data it will inflate the variance of the treatment effect. For simplicity, we assume that there is zero autocorrelation and therefore that the design effect for autocorrelation is equal to 1.</p>
df	<p>is the “error degrees of freedom” from the impact model shown in equation 1.</p> <p>These degrees of freedom can be obtained as the total number of observations (i.e., all years, all schools), minus the number of terms in the model (including the intercept, if present).</p>
\bar{T}	<p>is the proportion of the observations for which the treatment indicator equals one.</p> <p>Specifically, for the impact model shown in Equation 1, it is the proportion of observation where $TrtGrp_k * TrtYr_{jk} = 1$.</p>
$R_{TG*TYr Sch}^2$	<p>This is a measure of the squared correlation between the predictor variable and the school fixed effects (or indicators). The value of this r-squared can be determined during the design phase in the following two-step process. First, the investigator needs to generate a data set that has the same number of treatment and comparison schools, and the same number of pre-treatment and post-treatment years as are planned for the final analysis. The data set needs to include school IDs, and indicators for $TrtGrp$ and $TrtYr$, just as they will appear in the final analysis. In the second step, the investigator fits those data to the following model</p> $(TrtGrp_k * TrtYr_{jk}) = \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk}$

	to obtain the relevant r-square, $R_{TG*TYr Sch}^2$.
$R_{TG*TYr TrtYr(Sch)}^2$	<p>This is a measure of the squared correlation between the predictor variable and the term that indicates post-treatment observations, $TrtYr$ conditional on the school indicators During the design phase, and using the data generated in the step described above, the value of this r-squared can be calculated as the difference between the r-squareds from larger and a smaller model, where the larger model is</p> $(TrtGrp_k * TrtYr_{jk}) = \beta_1(TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk}$ <p>and the smaller model is</p> $(TrtGrp_k * TrtYr_{jk}) = \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk} .$
$R_{TG*TYr X(Sch,TrtYr)}^2$	<p>This is a measure of the squared correlation between the predictor indicator and the any remaining terms (e.g. covariate Xs) conditional on school fixed effects and the indicator variable for the post-treatment years. In the absence of plausible values from prior studies, we recommend that investigators try calculating MDESs for a range of small to moderate r-squared values (e.g., ranging from 0 to around .15). For most studies with the values of the Xs should not or could not have been influenced by the treatment itself (i.e., excluding clearly endogenous variables from consideration as covariate Xs) and where comparison schools were chosen such that they would have similar values of covariate Xs to treatment schools, the r-squares are likely to be at the lower end of this range.</p> <p>If the X values are known during the design phase and can be attached to the generated design-phase data set, then the r-square shown here can be calculated as the difference between a larger model and a smaller model where the larger model is</p> $(TrtGrp_k * TrtYr_{jk}) = \beta_1(TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \sum_{p=1}^P \lambda_p X_p + \varepsilon_{jk}$ <p>and the smaller model is</p> $(TrtGrp_k * TrtYr_{jk}) = \beta_1(TrtYr_{jk}) + \sum_{k=1}^K \alpha_k Sch_k + \varepsilon_{jk}$

Table 2. Plausible Parameter Values and Corresponding MDES for the simple C-SITS model in Equation 1

N	\bar{T}	#Pre	#Post	$R_{P*T I}^2$	$R_{P*T P(I)}^2$	$R_{Y I}^2$	$R_{Y P(I)}^2$	MDES
40	0.25	3	3	0.24	0.15	0.85	0.01	0.20
40	0.5	3	3	0.25	0.23	0.85	0.01	0.19
40	0.75	3	3	0.11	0.60	0.85	0.01	0.29
40	0.25	4	2	0.44	0.06	0.85	0.01	0.22
40	0.5	4	2	0.54	0.14	0.85	0.01	0.24
40	0.75	4	2	0.31	0.34	0.85	0.01	0.26

Notes:

N= total number of treatment and comparison units

$$R_{P*T|I}^2 \equiv R_{TG*TYr|Sch}^2$$

$$R_{P*T|P(I)}^2 \equiv R_{TG*TYr|TrtYr(Sch)}^2$$

$$R_{Y|I}^2 \equiv R_{Y|Sch}^2$$

$$R_{Y|P(I)}^2 \equiv R_{Y|TrtYr(Sch)}^2$$

For simplicity, no additional covariates and related R2 terms are included in the MDES calculation.