Received: 28 February 2019 | Revised: 1 August 2019 | Accepted: 6 August 2019

DOI: 10.1111/jep.13266

ORIGINAL PAPER

WILEY Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Design, analysis, power, and sample size calculation for three-phase interrupted time series analysis in evaluation of health policy interventions

Bo Zhang PhD[1] | Wei Liu PhD[2] | Stephenie C. Lemon PhD[1] | Bruce A. Barton PhD[1] | Melissa A. Fischer MD[3,4] | Colleen Lawrence PhD[5] | Elizabeth J. Rahn PhD[6] | Maria I. Danila MD[6] | Kenneth G. Saag MD, MSC[6] | Paul A. Harris PhD[7] | Jeroan J. Allison MD, MS[1]

[1] Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts

[2] School of Management, Harbin Institute of Technology, Harbin, Heilongjiang, China

[3] Department of Internal Medicine, University of Massachusetts Medical School, Worcester, Massachusetts

[4] Meyers Primary Care Institute, University of Massachusetts Medical School, Fallon Foundation, and Fallon Community Health Plan, Worcester, Massachusetts

[5] Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee

[6] Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama

[7] Department of Biomedical Informatics and Department of Biomedical Engineering, Vanderbilt University, Nashville, Tennessee

Correspondence
Bo Zhang, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA 01605.
Email: bo.zhang@umassmed.edu

Wei Liu, School of Management, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China.
Email: liuwhit@hit.edu.cn

Funding information
National Institutes of Health, Grant/Award Numbers: 5 U01 TR001812, KL2TR01455, TL1TR01454, U24 AA026968, UL1TR001453 and 5 U01 TR001812; University of Massachusetts Center for Clinical and Translational Science, Grant/Award Numbers: KL2TR01455, TL1TR01454 and UL1TR001453; National Natural Science Foundation of China, Grant/Award Numbers: 11601106 and 91646106

Abstract

Objective: To discuss the study design and data analysis for three-phase interrupted time series (ITS) studies to evaluate the impact of health policy, systems, or environmental interventions. Simulation methods are used to conduct power and sample size calculation for these studies.

Methods: We consider the design and analysis of three-phase ITS studies using a study funded by National Institutes of Health as an exemplar. The design and analysis of both one-arm and two-arm three-phase ITS studies are introduced.

Results: A simulation-based approach, with ready-to-use computer programs, was developed to determine the power for two types of three-phase ITS studies. Simulations were conducted to estimate the power of segmented autoregressive (AR) error models when autocorrelation ranged from −0.9 to 0.9 with various effect sizes. The power increased as the sample size or the effect size increased. The power to detect the same effect sizes varied largely, depending on testing level change, trend changes, or both.

Conclusion: This article provides a convenient tool for investigators to generate sample sizes to ensure sufficient statistical power when three-phase ITS study design is implemented.

KEYWORDS

interrupted time series, policy evaluation, power, quasi-experimental design, sample size calculation, segmented regression

 © 2019 John Wiley & Sons, Ltd.

1 | INTRODUCTION

The interrupted time series (ITS) design is a strong quasi-experimental study design for evaluating the longitudinal effects of policy, systems, environmental, or other types of interventions applied to an entire population or as part of routine practice.[1-3] The advantages and assumptions underlying the use of ITS for policy decisions have been extensively described and debated.[4-6] In a conventional two-phase ITS study design, a population-level outcome is measured at repeated intervals of time, before and after the introduction of the intervention.[7-10] The statistical analysis may reveal a change in the "level" of the outcome, evidenced by an abrupt discontinuity in the stream of repeated outcome measurements surrounding the point in time when the intervention was introduced. In addition, the analysis may show a change in slope of the outcome, which represents a gradual linear trend occurring after the introduction of the intervention. Therefore, the study can be divided into a preintervention phase and a postintervention phase, and the analysis is accomplished by using segmented time series regression models with one discontinuity time point.

However, in real-world settings, it may be too simplistic to conceptualize the intervention as being implemented in its totality at a single time point. Under these circumstances, a three-phase design allows the analysis to more closely parallel the actual process of intervention implementation. There are two possible scenarios, in which a three-phase time series study design is appropriate. The first scenario is that the implementation of the intervention requires a period of time to reach its full extent and it is not fully implemented at a single initial point in time. This period is viewed as a "ramp-up" period and occurs in time between the preintervention and full postimplementation period. After the ramp-up period, the intervention is fully implemented. Therefore, this ramp-up period may be considered as the second phase of the study before the study enters its third phase of full-scale implementation. Examples of this scenario are interventions focused on organizational change, which often begin with an initial set of activities that build momentum over time. The second scenario in which a three-phased ITS design may be warranted is when a multicomponent intervention is introduced in stages, with different components of the intervention introduced sequentially or in a non-uniform manner. This period of multi-component, non-uniform roll-out may be considered the second phase, followed by third phase of full-scale implementation.

While, for a two-phase ITS design with one study arm, Zhang, Wagner, and Ross-Degnan[11] conducted simulations to estimate power and sample sizes, little guidance exists on how to design a three-phased ITS study with sufficient sample size and power. This manuscript aims at filling this knowledge gap.

2 | EXEMPLAR STUDY: STRENGTHENING TRANSLATIONAL RESEARCH IN DIVERSE ENROLLMENT STUDY

The design and analysis of three-phase ITS studies were motivated by the need to calculate sample size and to determine power for the Strengthening Translational Research in Diverse Enrollment (STRIDE) study, a 5-year study funded by the National Institutes of Health (NIH) (grant no. 5 U01 TR001812) to develop, test, and disseminate an integrated multilevel, culturally sensitive intervention to engage African Americans and Latinos in clinical trials and translational research. The STRIDE study intervention is complex and multicomponent, with a ramp-up period that is required for the research team to achieve the full-scale implementation of the intervention. Therefore, STRIDE study is an ideal exemplar for the three-phase ITS design.

The motivation for STRIDE stems from the realization that African Americans and Latinos suffer disproportionately from leading causes of death and disability, yet despite this disparity, participation in clinical trials and translational research studies remains low.[12] The STRIDE study is a multisite collaboration between the University of Massachusetts Medical School, the University of Alabama at Birmingham, and the Vanderbilt University Medical Center. It is intended to address participant, research staff, and systems barriers to African Americans and Latinos in clinical and translational research.[13-15] The STRIDE intervention consists of three components: electronic informed consent, electronic consent assistance with patient stories, and simulation training of research assistants. The primary study hypothesis is that the STRIDE intervention will increase both the recruitment and retention rates of individuals from the overall number and proportion of total research participants who are members of underrepresented racial/ethnic groups in research studies, in particular African Americans and Latinos.

To accomplish our evaluation of the STRIDE intervention, we will partner with ongoing translational research studies. The STRIDE intervention will be introduced into the protocols of the ongoing research studies. A three-phase design is appropriate for STRIDE because the multicomponent intervention will require a ramp-up period for the research teams to fully integrate the intervention as part of their routine workflow. Thus, the STRIDE intervention will be evaluated in a three-phase, two-arm ITS study that will include six ongoing translational research studies: three studies receive the intervention (treatment arm) and the rest of three studies serve as comparison studies (control arm). For each translational research study in the treatment arm, the STRIDE intervention with all three components will be fully implemented, eventually. Study outcomes will include the total number and proportion, respectively, of African American and Latino participants enrolled (recruitment) and retained (retention) in the study each week. To assess this, we plan to collect a weekly recruitment progress summary from each study, aggregated at the level of the week for each study. These recruitment summaries will be monitored to provide a baseline preimplementation. Then, the STRIDE intervention will be brought online with a ramp-up period. We will then continue to monitor the data stream during the full-scale implementation period. The study hypothesis is that the intervention has effect on the change in study outcomes from preimplementation period to ramp-up period or from ramp-up period to full-scale implementation periods.

3 | SIMULATION-BASED METHODS FOR POWER AND SAMPLE SIZE CALCULATION OF THREE-PHASE ITS STUDIES

3.1 | Design and analysis of three-phase single-arm ITS study

A three-phase single-arm ITS study is a three-phase ITS study in which all study subjects and sites are planned to be exposed to an intervention over time (see Figure 1). The data collected from a three-phase single-arm ITS study can be analysed by a segmented time series regression model with two change points:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_{t^{(1)}} + \beta_3 X_{t^{(2)}} + \beta_4 (T_t - t_1) X_{t^{(1)}} + \beta_5 (T_t - t_2) X_{t^{(2)}} + \epsilon_t$$

in which Y_t represents the aggregated outcome variable measured over time, T_t is the actual or converted study time from the start to the end of the study, $X_{t^{(1)}}$ is a binary indicator coded as 0 before the implementation of the first-level intervention and 1 after the implementation of the first-level intervention, while $X_{t^{(2)}}$ is a binary indicator coded as 0 before the implementation of the second-level intervention and 1 the implementation of after the second-level intervention, t_1 is the first time point after the implementation of first-level intervention, t_2 is the first time point after the implementation of second-level intervention, and ϵ_t is the random error term. The coefficient β_0 is the regression intercept representing the starting level of the aggregate outcome variable, β_1 is the slope or trajectory of the aggregated outcome variable before the implementation of first-level intervention, β_2 and β_3 represent the change in the level of the outcome that occurs immediately after the implementation of the first-level and second-level intervention, respectively, and β_4 and β_5 represent the difference between preintervention and first-level intervention slopes and the difference between first-level intervention and second-level intervention slopes of the aggregated outcome, respectively. The focus of the three-phase ITS analysis is to examine the significance of β_2 and β_3, or the summation of them, that indicate an immediate intervention effect of first-level and second-level intervention in terms of level change and the significance of β_4 and β_5, or their summation, that indicate the intervention effect in terms of change in trend. Note that the purpose of subtracting t_1 and t_2, the first time point after the implementation of first-level and second-level intervention, respectively, from the study time T_t is to maintain the interpretation of the corresponding regression coefficients β_4 and β_5 (see Huitema and Mckean[16] for details regarding model specification).

In the ITS analysis, the random error term ϵ_t can be specified to follow a first-order autoregressive process, which is denoted by AR(1) and specified as

$$\epsilon_t = \rho \epsilon_{t-1} + u_t$$

in which the autocorrelation parameter ρ is the correlation coefficient between adjacent random error terms and the disturbances u_t independently and identically follow a normal distribution $N(0, \sigma^2)$. The specification of the random error term ϵ_t can also be specified with a higher-order autoregressive process, an autoregressive conditional heteroscedasticity (ARCH) models, or an autoregressive integrated moving average (ARIMA) model (see Appendix A). Estimates of the regression coefficients in the three-phase ITS models are obtained using the maximum likelihood estimation procedure.

3.2 | Design and analysis of three-phase two-arm ITS study

A three-phase ITS study can be designed to include two study arms, one treatment arm (intervention group) and one control arm

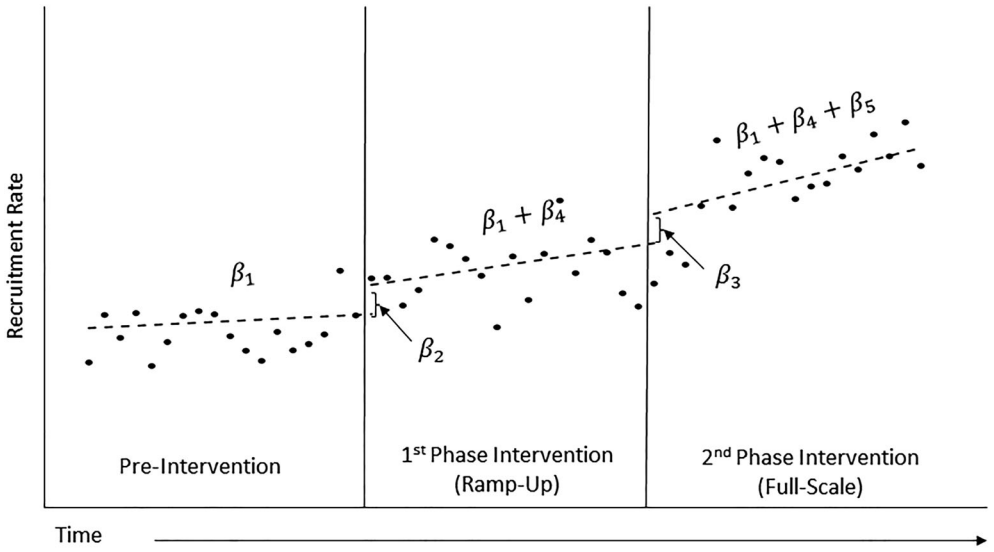

FIGURE 1 Study design and hypothetical results from a three-phase one-arm interrupted time series (ITS) trial. The hypothetical data here indicate both a change in level and differences in trend, which are represented by the upward slope of the regression line being greater in the first and second phases of intervention

(comparison group) (see Figure 2). Assignment to treatment or control arm can be randomized or not. The participants in the treatment arm receive investigated intervention, while the participants in the comparison arm receive no intervention or active control. The data collected from a three-phase two-arm ITS study can be analysed by a segmented time series regression model with the following form:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_{t^{(1)}} + \beta_3 X_{t^{(2)}} + \beta_4 (T_t - t_1) X_{t^{(1)}} + \beta_5 (T_t - t_2) X_{t^{(2)}} + \beta_6 G + \beta_7 G T_t + \beta_8 G X_{t^{(1)}} + \beta_9 G X_{t^{(2)}} + \beta_{10} G (T_t - t_1) X_{t^{(1)}} + \beta_{11} G (T_t - t_2) X_{t^{(2)}} + \epsilon_{tG}.$$

in which G is the binary indicator for treatment group ($G = 1$) versus control group ($G = 0$). For other notations, see Appendix B for detailed explanation.

3.3 | Simulation-based methods for power and sample size calculation

We conducted the power and sample size calculation through a simulation-based method for one-arm and two-arm three-phase ITS design for evaluating health policy interventions. Suppose the null hypothesis to be tested is $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, where β is a universal notation for an arbitrary regression coefficient or a vector of multiple coefficients in either one-arm or two-arm three-phase ITS models discussed above. Then, the power of this statistical hypothesis test at a fixed sample size under a prespecified significance level is equal to the probability of rejecting the null hypothesis given the alternative hypothesis is true, ie, *Prob*(Reject H_0| H_1 is true). Thus, the simulation-based method is to numerically generate a large number of data sets, say R data sets, from an ITS model with a nonzero value of β and perform the statistical hypothesis test to determine whether the null hypothesis is rejected. Then, the numerically computed power is the frequency that the null hypothesis is rejected among the R data sets. The difference between maximum likelihoods of the null hypothesis and intervention hypothesis models was examined through a chi-square test on the likelihood ratio statistic. The effect sizes that were examined in this simulation-based calculation are defined as (i) total intervention effect size, which is the sum of expected level change in two intervention phases (first-level intervention and second-level intervention) plus the expected trend change in two intervention phases over its standard deviation, (ii) effect size in total level change, which is the sum of expected level change in two intervention phases over its standard deviation, and (iii) effect size in total trend change, which is the sum of expected trend change in two intervention phases over its standard deviation. Here, the standard deviation refers to the standard deviation of the random error in the ITS segmented time series regression model. It can be estimated from fitting the model to preintervention data or relevant data from previous studies in power and sample size calculation. The total effect size represents the summation of both level and trend changes and therefore does not distinguish them. Separated hypothesis testing should be designed to detect the change in either level or trend. When the study objective is to specifically examine the change in level or the change in trend in the ITS study, the investigators should perform hypothesis testing (ii) for detecting the level change or perform hypothesis testing (iii) for detecting the trend change.

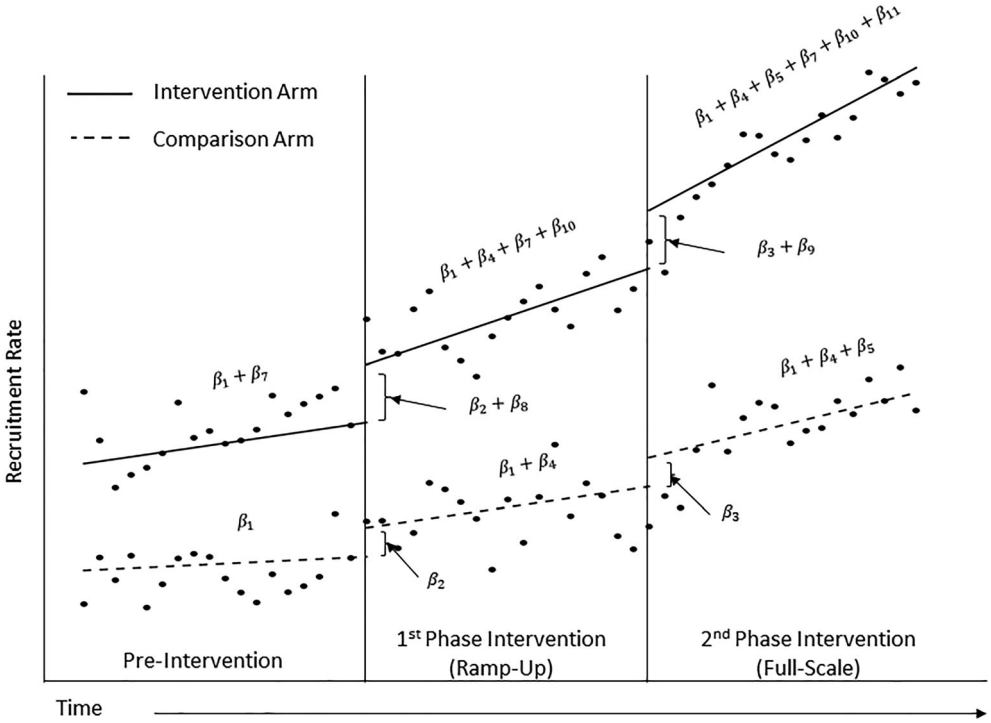

FIGURE 2 Study design and hypothetical results from a three-phase, two-arm interrupted time series (ITS) trial. The hypothetical data here indicate both a change in level and differences in trend, which are represented by the upward slope of the regression line being greater for the intervention group than the comparison group in the first and second phases of intervention

We chose the simulated effect sizes as 0.5, 1, and 2 for effect size definition (i); 2, 3, and 4 for (ii); and 0.1, 0.25, and 0.5 for (iii). The reason that we chose different effect sizes for (i), (ii), and (iii) is to ensure in all three scenarios the power can range from approximately 0.3 to 1. For hypothesis test (i), we chose equal values of expected level change and expected trend change; for hypothesis test (ii), we fixed the expected trend change to be 0, which anticipated no trend changes in either intervention period; and for hypothesis test (iii), we fixed the expected level change to be 0, which anticipated no level changes in either intervention period. Other effect sizes can also be specified, and the corresponding power can be determined by the simulation-based methods. Sample sizes (number of total time points in three study phases) of 18, 27, 36, 45, 54, 72, 81, 90, and 108, with balanced numbers of time points in three periods before and after the first-level and second-level of intervention, were considered. All scenarios used a total $R = 1000$ simulated data sets, and the model for random error term was specified as AR(1).

4 | RESULTS

Tables 1 and 2 present the estimated power of the segmented time series regression model with AR(1) random errors to detect a total change of level and trend with effect sizes 0.5, 1, and 2 and 0.05 significance level, for a one-arm ITS study (testing $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$) and for a two-arm ITS study (testing $H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$), respectively. As expected, the simulated power increased as the sample size or effect size increased. Change of power followed a U-shape pattern (the power first decreased and then increased) as the autocorrelation increased from −0.9 to 0.9. This U-shape pattern was not apparent for large sample sizes but still existed.

Tables 3 and 4 present the estimated power of the segmented time series regression model with AR(1) random errors to detect a level change with effect sizes 2.0, 3.0, and 4.0 and 0.05 significance level, for a one-arm ITS study (testing $H_0 : \beta_2 = \beta_3 = 0$) and for a two-arm ITS study (testing $H_0 : \beta_8 = \beta_9 = 0$), respectively. Tables 5 and 6 present the estimated power of the segmented time series regression model with AR(1) random errors to detect a change in trend with effect sizes 0.1, 0.25, and 0.5 significance level, for a one-arm ITS study (testing $H_0 : \beta_4 = \beta_5 = 0$) and for a two-arm ITS study (testing $H_0 : \beta_{10} = \beta_{11} = 0$), respectively. As we can observe, patterns of power change in Tables 3–6 were similar to Tables 1 and 2. Compared with Tables 1 and 2, the power in Tables 3 and 4 achieved similar level with larger effects sizes, but Tables 5 and 6 required smaller effects sizes.

5 | DISCUSSION

The ITS design has been applied to a variety of topic areas, including the evaluation of health policy, medication effectiveness and safety, quality improvement initiatives, and community screening programs, among other population-based studies.[1-3] In this article, three-phase ITS study design is discussed, with specific application when the intervention components are introduced sequentially in a ramp-up period and the intervention effect is expected to increase over time. With a three-phase study design, a corresponding analysis plan, as well as power and sample size calculation strategies, is needed. Herein, we developed a simulation-based method to estimate sample size and power for both one-arm and two-arm three-phase ITS studies. Simulation results from testing level change, trend change, and total change (sum of level and trend change) are demonstrated with diverse effect sizes and parameter specification. As anticipated, the estimated power increased as the sample size or effect size increased. Change of power has a U-shape pattern as the autocorrelation increased from −0.9 to 0.9. Comparing the power across the six tables presented here, we conclude that the power to detect the same level of effect size can vary widely, depending on whether testing level change, trend change, or testing total change are performed.

Our power and sample size calculation are conducted based upon models and hypothesis testing at the aggregated level of data. For example, the STRIDE analysis will be conducted on aggregated retention data within 1-week periods. With this analysis approach, the sample sizes required to reach certain power in the three-phase ITS studies are determined by the number of time points, not the number of data points that are aggregated at each time window. Although such aggregated analysis is common in the literature, it does entail loss of information from aggregated data across time windows since it ignores the heterogeneities between individuals. Future studies need to focus on analysing individual-level time-dependent data, with presumed mean changes occurring at the time points of policy or intervention implementation. Investigators should also pay attention to the fact that the number of subjects contributing data to the aggregated measure at each time point also affects the power of the ITS studies, although the number of time intervals likely contributes most to the power. For example, the power for 12 intervals in an ITS study consisting of only 10 individuals per interval is less than that for 12 intervals consisting of 1000 individuals per interval, because the variance of random error is less. Therefore, it is recommended enrolling enough participants in the study to ensure a sufficient power.

There are some limitations in the simulation-based modelling developed and described herein. First, during the simulation procedure, we only specify the error term as AR(1). As we discussed in the Supporting Information, there are other possible specifications for the error term (eg, autoregressive integrated moving average and ARCH). Estimated power and sample sizes can be generated and evaluated using these specifications. Second, the power and sample size calculation presented in this manuscript were conducted with a balanced ITS design (identical time points in each phase). However, our method can also be applied to calculate power for studies with unbalanced ITS designs. Third, the three-phase ITS analysis should only be applied if the ramp-up period is of adequate length or the phase in the middle is of adequate length. As suggested by Zhang, Wagner, and Ross-Degnan,[11] a minimum of eight intervals allow a separate segment to be modelled in the ITS analysis. If the ramp-up period or the period in the middle only consists of a small duration, it is advisable to censor this period in the ITS analysis or to set the

TABLE 1 Estimated power for AR(1) model with both level and trend change assuming effect size = 0.5, 1, 2 based on 1000 simulated data sets and statistical significance level 0.05, for one-arm interrupted time series study (testing $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
	Effect size = 0.5								
−0.9	0.39	0.76	0.99	1	1	1	1	1	1
−0.8	0.30	0.47	0.82	0.98	1	1	1	1	1
−0.7	0.25	0.36	0.61	0.88	0.99	1	1	1	1
−0.6	0.27	0.30	0.50	0.75	0.93	1	1	1	1
−0.5	0.26	0.26	0.40	0.65	0.87	1	1	1	1
−0.4	0.27	0.24	0.36	0.55	0.76	0.99	1	1	1
−0.3	0.26	0.22	0.31	0.48	0.66	0.96	0.99	1	1
−0.2	0.29	0.24	0.33	0.42	0.60	0.92	0.98	1	1
−0.1	0.31	0.24	0.28	0.37	0.51	0.85	0.96	0.99	1
0	0.35	0.26	0.27	0.36	0.45	0.79	0.92	0.98	1
0.1	0.40	0.27	0.25	0.34	0.42	0.73	0.83	0.94	1
0.2	0.40	0.29	0.28	0.30	0.40	0.65	0.80	0.89	0.99
0.3	0.47	0.30	0.28	0.31	0.34	0.55	0.74	0.85	0.97
0.4	0.51	0.34	0.28	0.31	0.34	0.53	0.64	0.78	0.94
0.5	0.56	0.40	0.33	0.34	0.35	0.47	0.57	0.65	0.87
0.6	0.62	0.40	0.37	0.33	0.36	0.45	0.49	0.57	0.80
0.7	0.68	0.48	0.40	0.36	0.37	0.42	0.46	0.54	0.67
0.8	0.71	0.55	0.47	0.43	0.40	0.45	0.47	0.51	0.60
0.9	0.76	0.63	0.55	0.54	0.53	0.50	0.57	0.56	0.63
	Effect size = 1								
−0.9	0.83	1	1	1	1	1	1	1	1
−0.8	0.59	0.96	1	1	1	1	1	1	1
−0.7	0.44	0.83	0.99	1	1	1	1	1	1
−0.6	0.40	0.71	0.97	1	1	1	1	1	1
−0.5	0.39	0.59	0.92	1	1	1	1	1	1
−0.4	0.37	0.55	0.84	0.99	1	1	1	1	1
−0.3	0.35	0.46	0.74	0.97	1	1	1	1	1
−0.2	0.36	0.46	0.71	0.93	0.99	1	1	1	1
−0.1	0.37	0.41	0.64	0.87	0.98	1	1	1	1
0	0.40	0.40	0.59	0.81	0.95	1	1	1	1
0.1	0.45	0.42	0.52	0.76	0.93	1	1	1	1
0.2	0.45	0.41	0.51	0.68	0.86	1	1	1	1
0.3	0.53	0.40	0.48	0.63	0.82	0.99	1	1	1
0.4	0.54	0.43	0.46	0.62	0.74	0.96	0.99	1	1
0.5	0.60	0.48	0.50	0.58	0.71	0.92	0.98	0.99	1
0.6	0.65	0.48	0.52	0.55	0.65	0.86	0.95	0.98	1
0.7	0.71	0.58	0.56	0.55	0.64	0.82	0.89	0.94	0.99
0.8	0.75	0.64	0.60	0.63	0.68	0.79	0.85	0.90	0.97
0.9	0.82	0.73	0.71	0.73	0.78	0.85	0.90	0.91	0.96
	Effect size = 2								
−0.9	1	1	1	1	1	1	1	1	1

(Continues)

TABLE 1 (Continued)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
−0.8	0.98	1	1	1	1	1	1	1	1
−0.7	0.91	1	1	1	1	1	1	1	1
−0.6	0.82	1	1	1	1	1	1	1	1
−0.5	0.73	1	1	1	1	1	1	1	1
−0.4	0.70	0.97	1	1	1	1	1	1	1
−0.3	0.63	0.93	1	1	1	1	1	1	1
−0.2	0.61	0.89	1	1	1	1	1	1	1
−0.1	0.58	0.84	0.98	1	1	1	1	1	1
0	0.62	0.83	0.98	1	1	1	1	1	1
0.1	0.60	0.78	0.95	1	1	1	1	1	1
0.2	0.63	0.75	0.93	0.99	1	1	1	1	1
0.3	0.63	0.72	0.90	0.99	1	1	1	1	1
0.4	0.71	0.73	0.87	0.97	1	1	1	1	1
0.5	0.69	0.74	0.86	0.95	0.99	1	1	1	1
0.6	0.75	0.75	0.84	0.95	0.98	1	1	1	1
0.7	0.78	0.80	0.84	0.93	0.98	1	1	1	1
0.8	0.89	0.85	0.90	0.92	0.97	1	1	1	1
0.9	0.95	0.93	0.95	0.98	0.99	1	1	1	1

TABLE 2 Estimated power for AR(1) model with both level and trend change assuming effect size = 0.5, 1, 2 based on 1000 simulated data sets and statistical significance level 0.05, for two-arm interrupted time series study (testing $H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
	Effect size = 0.5								
−0.9	0.31	0.55	0.86	0.99	1	1	1	1	1
−0.8	0.26	0.34	0.58	0.85	0.97	1	1	1	1
−0.7	0.24	0.30	0.40	0.65	0.86	1	1	1	1
−0.6	0.26	0.23	0.33	0.48	0.76	0.98	1	1	1
−0.5	0.30	0.23	0.28	0.46	0.62	0.94	0.99	1	1
−0.4	0.30	0.24	0.29	0.37	0.54	0.88	0.96	0.99	1
−0.3	0.34	0.23	0.28	0.35	0.46	0.81	0.90	0.98	1
−0.2	0.35	0.24	0.24	0.30	0.39	0.72	0.87	0.94	1
−0.1	0.41	0.26	0.25	0.28	0.37	0.62	0.78	0.88	0.99
0	0.44	0.30	0.28	0.28	0.31	0.56	0.73	0.83	0.97
0.1	0.48	0.33	0.26	0.28	0.32	0.51	0.61	0.74	0.94
0.2	0.51	0.34	0.29	0.28	0.32	0.46	0.54	0.66	0.89
0.3	0.57	0.39	0.32	0.31	0.32	0.43	0.49	0.61	0.83
0.4	0.62	0.38	0.32	0.30	0.31	0.40	0.48	0.55	0.71
0.5	0.68	0.49	0.38	0.32	0.31	0.38	0.46	0.49	0.68
0.6	0.70	0.53	0.39	0.34	0.37	0.36	0.39	0.46	0.61

(Continues)

TABLE 2 (Continued)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
0.7	0.75	0.58	0.50	0.40	0.41	0.37	0.41	0.43	0.55
0.8	0.82	0.65	0.53	0.47	0.45	0.39	0.41	0.42	0.49
0.9	0.84	0.72	0.65	0.59	0.55	0.53	0.53	0.52	0.58
	Effect size = 1								
−0.9	0.68	0.98	1	1	1	1	1	1	1
−0.8	0.47	0.84	0.98	1	1	1	1	1	1
−0.7	0.42	0.65	0.94	1	1	1	1	1	1
−0.6	0.38	0.50	0.81	0.98	1	1	1	1	1
−0.5	0.37	0.48	0.72	0.96	1	1	1	1	1
−0.4	0.37	0.42	0.67	0.87	0.99	1	1	1	1
−0.3	0.38	0.40	0.56	0.82	0.96	1	1	1	1
−0.2	0.40	0.37	0.50	0.76	0.92	1	1	1	1
−0.1	0.46	0.40	0.48	0.67	0.87	1	1	1	1
0	0.49	0.38	0.44	0.62	0.81	0.98	1	1	1
0.1	0.51	0.41	0.42	0.57	0.74	0.98	1	1	1
0.2	0.57	0.40	0.44	0.53	0.67	0.94	0.99	1	1
0.3	0.59	0.45	0.43	0.51	0.65	0.90	0.96	0.99	1
0.4	0.66	0.45	0.43	0.48	0.60	0.86	0.92	0.98	1
0.5	0.70	0.48	0.48	0.52	0.60	0.78	0.89	0.96	1
0.6	0.73	0.57	0.52	0.53	0.56	0.74	0.84	0.93	0.99
0.7	0.78	0.61	0.57	0.53	0.57	0.72	0.81	0.85	0.97
0.8	0.83	0.69	0.61	0.60	0.62	0.71	0.73	0.82	0.91
0.9	0.87	0.80	0.73	0.71	0.70	0.73	0.78	0.82	0.90
	Effect size = 2								
−0.9	1	1	1	1	1	1	1	1	1
−0.8	0.90	1	1	1	1	1	1	1	1
−0.7	0.77	1	1	1	1	1	1	1	1
−0.6	0.69	0.98	1	1	1	1	1	1	1
−0.5	0.60	0.92	1	1	1	1	1	1	1
−0.4	0.58	0.87	0.99	1	1	1	1	1	1
−0.3	0.55	0.82	0.99	1	1	1	1	1	1
−0.2	0.56	0.77	0.96	1	1	1	1	1	1
−0.1	0.58	0.72	0.93	1	1	1	1	1	1
0	0.61	0.68	0.92	0.99	1	1	1	1	1
0.1	0.62	0.64	0.87	0.98	1	1	1	1	1
0.2	0.65	0.63	0.83	0.97	1	1	1	1	1
0.3	0.68	0.66	0.78	0.95	0.99	1	1	1	1
0.4	0.72	0.69	0.78	0.90	0.98	1	1	1	1
0.5	0.73	0.69	0.76	0.87	0.97	1	1	1	1
0.6	0.79	0.70	0.76	0.86	0.94	1	1	1	1
0.7	0.82	0.77	0.78	0.84	0.92	0.99	1	1	1
0.8	0.87	0.82	0.82	0.89	0.91	0.99	1	1	1
0.9	0.91	0.90	0.90	0.94	0.97	0.99	1	1	1

TABLE 3 Estimated power for AR(1) model with a level change assuming effect size = 2, 3, 4 based on 1000 simulated data sets and statistical significance level 0.05, for one-arm interrupted time series study (testing $H_0 : \beta_2 = \beta_3 = 0$)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
	Effect size = 2								
−0.9	0.99	1	1	1	1	1	1	1	1
−0.8	0.91	0.99	1	1	1	1	1	1	1
−0.7	0.78	0.92	0.97	0.99	1	1	1	1	1
−0.6	0.64	0.82	0.91	0.96	0.99	1	1	1	1
−0.5	0.60	0.69	0.83	0.91	0.93	0.99	1	1	1
−0.4	0.54	0.64	0.77	0.83	0.89	0.97	0.99	0.99	1
−0.3	0.52	0.58	0.67	0.76	0.83	0.92	0.97	0.97	0.98
−0.2	0.52	0.56	0.61	0.72	0.77	0.87	0.91	0.91	0.97
−0.1	0.52	0.49	0.56	0.63	0.71	0.82	0.86	0.90	0.92
0	0.50	0.48	0.53	0.58	0.65	0.75	0.80	0.82	0.89
0.1	0.53	0.50	0.51	0.53	0.59	0.69	0.71	0.76	0.82
0.2	0.53	0.47	0.49	0.54	0.59	0.64	0.66	0.71	0.75
0.3	0.55	0.48	0.51	0.49	0.54	0.61	0.65	0.66	0.72
0.4	0.55	0.50	0.50	0.52	0.53	0.56	0.58	0.64	0.62
0.5	0.62	0.54	0.51	0.54	0.53	0.55	0.58	0.59	0.60
0.6	0.65	0.55	0.55	0.53	0.55	0.55	0.56	0.57	0.64
0.7	0.73	0.65	0.61	0.60	0.61	0.60	0.61	0.63	0.64
0.8	0.80	0.74	0.70	0.69	0.71	0.70	0.70	0.72	0.72
0.9	0.93	0.92	0.91	0.91	0.91	0.91	0.92	0.91	0.92
	Effect size = 3								
−0.9	1	1	1	1	1	1	1	1	1
−0.8	1	1	1	1	1	1	1	1	1
−0.7	0.97	1	1	1	1	1	1	1	1
−0.6	0.92	0.98	1	1	1	1	1	1	1
−0.5	0.86	0.96	0.99	1	1	1	1	1	1
−0.4	0.82	0.92	0.97	0.99	1	1	1	1	1
−0.3	0.75	0.86	0.95	0.99	0.99	1	1	1	1
−0.2	0.72	0.84	0.91	0.95	0.98	1	1	1	1
−0.1	0.68	0.78	0.85	0.91	0.96	0.99	0.99	1	1
0	0.73	0.75	0.83	0.86	0.91	0.97	0.99	0.99	1
0.1	0.70	0.74	0.82	0.86	0.90	0.96	0.97	0.98	0.99
0.2	0.69	0.73	0.78	0.80	0.85	0.93	0.95	0.96	0.99
0.3	0.73	0.72	0.74	0.79	0.83	0.91	0.91	0.94	0.97
0.4	0.78	0.75	0.75	0.80	0.81	0.86	0.89	0.90	0.93
0.5	0.79	0.76	0.76	0.79	0.81	0.86	0.88	0.89	0.94
0.6	0.84	0.82	0.80	0.82	0.84	0.87	0.88	0.90	0.94
0.7	0.87	0.88	0.86	0.86	0.88	0.90	0.91	0.93	0.92
0.8	0.95	0.94	0.93	0.95	0.95	0.95	0.97	0.96	0.97
0.9	1	0.98	0.99	1	1	1	1	1	1
	Effect size = 4								
−0.9	1	1	1	1	1	1	1	1	1

(Continues)

TABLE 3 (Continued)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
−0.8	1	1	1	1	1	1	1	1	1
−0.7	1	1	1	1	1	1	1	1	1
−0.6	0.99	1	1	1	1	1	1	1	1
−0.5	0.97	1	1	1	1	1	1	1	1
−0.4	0.94	0.99	1	1	1	1	1	1	1
−0.3	0.92	0.98	1	1	1	1	1	1	1
−0.2	0.90	0.96	0.99	1	1	1	1	1	1
−0.1	0.88	0.95	0.97	0.99	1	1	1	1	1
0	0.86	0.93	0.96	0.97	0.99	1	1	1	1
0.1	0.86	0.90	0.96	0.99	0.99	1	1	1	1
0.2	0.85	0.90	0.93	0.97	0.98	0.99	1	1	1
0.3	0.86	0.91	0.92	0.94	0.98	0.99	0.99	1	1
0.4	0.88	0.88	0.93	0.94	0.95	0.99	0.99	0.99	1
0.5	0.90	0.93	0.93	0.93	0.96	0.98	0.98	0.99	0.99
0.6	0.95	0.95	0.93	0.96	0.97	0.98	0.99	0.99	0.99
0.7	0.96	0.97	0.98	0.98	0.98	0.99	0.99	0.99	1
0.8	0.99	0.99	0.99	0.99	1	1	1	1	1
0.9	1	1	1	1	1	1	1	1	1

TABLE 4 Estimated power for AR(1) model with a level change assuming effect size = 2, 3, 4, based on 1000 simulated data sets and statistical significance level 0.05, for two-arm interrupted time series study (testing $H_0 : \beta_8 = \beta_9 = 0$)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
	Effect size = 2								
−0.9	0.94	0.99	1	1	1	1	1	1	1
−0.8	0.71	0.87	0.95	0.98	1	1	1	1	1
−0.7	0.58	0.71	0.83	0.89	0.96	0.99	0.99	1	1
−0.6	0.49	0.60	0.70	0.79	0.85	0.95	0.95	0.98	1
−0.5	0.46	0.52	0.59	0.67	0.76	0.86	0.88	0.94	0.97
−0.4	0.45	0.47	0.52	0.58	0.66	0.79	0.80	0.86	0.92
−0.3	0.45	0.44	0.47	0.52	0.59	0.69	0.74	0.77	0.85
−0.2	0.44	0.39	0.43	0.46	0.51	0.58	0.67	0.69	0.78
−0.1	0.45	0.38	0.39	0.40	0.47	0.55	0.58	0.62	0.70
0	0.44	0.38	0.38	0.39	0.40	0.49	0.54	0.56	0.63
0.1	0.46	0.39	0.38	0.38	0.41	0.46	0.47	0.50	0.55
0.2	0.47	0.39	0.39	0.36	0.39	0.42	0.45	0.47	0.50
0.3	0.54	0.39	0.36	0.35	0.36	0.41	0.40	0.44	0.45
0.4	0.54	0.41	0.39	0.37	0.37	0.38	0.39	0.42	0.41
0.5	0.54	0.45	0.37	0.36	0.39	0.38	0.37	0.39	0.42
0.6	0.60	0.45	0.43	0.39	0.40	0.37	0.38	0.40	0.41

(Continues)

TABLE 4 (Continued)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
0.7	0.63	0.50	0.45	0.42	0.45	0.40	0.40	0.40	0.42
0.8	0.71	0.63	0.51	0.52	0.51	0.50	0.48	0.48	0.49
0.9	0.80	0.77	0.70	0.71	0.71	0.70	0.71	0.67	0.70
	Effect size = 3								
−0.9	1	1	1	1	1	1	1	1	1
−0.8	0.95	1	1	1	1	1	1	1	1
−0.7	0.87	0.94	0.99	1	1	1	1	1	1
−0.6	0.75	0.88	0.96	0.98	0.99	1	1	1	1
−0.5	0.67	0.79	0.91	0.94	0.98	1	1	1	1
−0.4	0.67	0.73	0.82	0.90	0.95	0.99	0.99	0.99	1
−0.3	0.61	0.66	0.76	0.83	0.89	0.96	0.98	0.98	1
−0.2	0.60	0.63	0.66	0.78	0.83	0.91	0.94	0.96	0.99
−0.1	0.59	0.61	0.65	0.70	0.77	0.86	0.89	0.92	0.97
0	0.59	0.57	0.58	0.67	0.72	0.81	0.83	0.86	0.93
0.1	0.58	0.54	0.58	0.65	0.68	0.76	0.79	0.84	0.88
0.2	0.62	0.52	0.57	0.64	0.60	0.71	0.75	0.79	0.84
0.3	0.62	0.55	0.56	0.59	0.59	0.67	0.72	0.72	0.77
0.4	0.65	0.58	0.56	0.59	0.57	0.64	0.65	0.68	0.73
0.5	0.72	0.61	0.56	0.59	0.62	0.63	0.62	0.66	0.71
0.6	0.72	0.65	0.62	0.59	0.63	0.64	0.64	0.64	0.69
0.7	0.78	0.70	0.67	0.67	0.66	0.65	0.69	0.70	0.71
0.8	0.86	0.78	0.74	0.77	0.76	0.75	0.74	0.77	0.80
0.9	0.97	0.94	0.93	0.93	0.95	0.94	0.94	0.95	0.96
	Effect size = 4								
−0.9	1	1	1	1	1	1	1	1	1
−0.8	1	1	1	1	1	1	1	1	1
−0.7	0.97	1	1	1	1	1	1	1	1
−0.6	0.92	0.99	1	1	1	1	1	1	1
−0.5	0.85	0.95	0.99	1	1	1	1	1	1
−0.4	0.81	0.93	0.96	0.99	1	1	1	1	1
−0.3	0.75	0.87	0.95	0.98	0.99	1	1	1	1
−0.2	0.74	0.82	0.90	0.96	0.97	0.99	1	1	1
−0.1	0.72	0.79	0.86	0.90	0.94	0.99	0.99	0.99	1
0	0.71	0.76	0.82	0.88	0.92	0.97	0.98	0.99	1
0.1	0.70	0.73	0.80	0.83	0.86	0.95	0.95	0.97	0.99
0.2	0.72	0.72	0.76	0.80	0.83	0.90	0.94	0.95	0.97
0.3	0.71	0.72	0.74	0.79	0.83	0.86	0.90	0.93	0.96
0.4	0.77	0.74	0.75	0.76	0.81	0.86	0.87	0.89	0.92
0.5	0.78	0.75	0.77	0.80	0.82	0.84	0.84	0.87	0.89
0.6	0.83	0.80	0.79	0.83	0.79	0.86	0.85	0.87	0.87
0.7	0.90	0.86	0.84	0.85	0.86	0.88	0.88	0.89	0.91
0.8	0.94	0.94	0.92	0.94	0.93	0.95	0.94	0.94	0.95
0.9	0.99	0.99	1	1	0.99	0.99	1	1	1

TABLE 5 Estimated power for AR(1) model with a trend change assuming effect size = 0.1, 0.25, 0.5 based on 1000 simulated data sets and statistical significance level 0.05, for one-arm interrupted time series study (testing $H_0 : \beta_4 = \beta_5 = 0$)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
	Effect size = 0.1								
−0.9	0.26	0.52	0.87	1	1	1	1	1	1
−0.8	0.22	0.32	0.59	0.84	0.98	1	1	1	1
−0.7	0.22	0.25	0.44	0.69	0.91	1	1	1	1
−0.6	0.23	0.23	0.35	0.55	0.75	0.99	1	1	1
−0.5	0.24	0.21	0.30	0.47	0.66	0.96	1	1	1
−0.4	0.23	0.20	0.27	0.39	0.56	0.92	0.98	1	1
−0.3	0.24	0.18	0.23	0.36	0.50	0.85	0.94	0.99	1
−0.2	0.26	0.21	0.26	0.31	0.44	0.79	0.90	0.96	1
−0.1	0.26	0.22	0.23	0.28	0.38	0.67	0.82	0.93	0.99
0	0.32	0.23	0.23	0.27	0.36	0.63	0.75	0.88	0.98
0.1	0.35	0.24	0.22	0.28	0.36	0.60	0.69	0.81	0.97
0.2	0.36	0.28	0.26	0.24	0.33	0.52	0.65	0.76	0.93
0.3	0.44	0.29	0.27	0.26	0.30	0.45	0.60	0.70	0.88
0.4	0.47	0.32	0.26	0.27	0.29	0.44	0.53	0.65	0.84
0.5	0.51	0.35	0.31	0.31	0.32	0.41	0.52	0.57	0.78
0.6	0.59	0.39	0.34	0.31	0.33	0.41	0.47	0.52	0.72
0.7	0.62	0.46	0.40	0.35	0.35	0.41	0.45	0.54	0.65
0.8	0.66	0.56	0.47	0.45	0.45	0.48	0.50	0.54	0.64
0.9	0.73	0.63	0.55	0.55	0.54	0.55	0.58	0.62	0.70
	Effect size = 0.25								
−0.9	0.76	1	1	1	1	1	1	1	1
−0.8	0.51	0.93	1	1	1	1	1	1	1
−0.7	0.40	0.78	0.99	1	1	1	1	1	1
−0.6	0.35	0.65	0.95	1	1	1	1	1	1
−0.5	0.36	0.59	0.90	1	1	1	1	1	1
−0.4	0.34	0.49	0.80	0.98	1	1	1	1	1
−0.3	0.32	0.48	0.74	0.95	1	1	1	1	1
−0.2	0.34	0.45	0.69	0.92	0.99	1	1	1	1
−0.1	0.33	0.42	0.62	0.85	0.98	1	1	1	1
0	0.40	0.44	0.60	0.83	0.96	1	1	1	1
0.1	0.42	0.43	0.54	0.77	0.92	1	1	1	1
0.2	0.44	0.41	0.53	0.74	0.89	1	1	1	1
0.3	0.43	0.42	0.52	0.67	0.85	0.99	1	1	1
0.4	0.53	0.44	0.50	0.68	0.81	0.98	0.99	1	1
0.5	0.53	0.50	0.54	0.65	0.81	0.97	0.99	1	1
0.6	0.60	0.51	0.55	0.67	0.78	0.96	0.99	0.99	1
0.7	0.64	0.58	0.61	0.68	0.77	0.94	0.98	0.99	1
0.8	0.70	0.66	0.69	0.72	0.79	0.92	0.96	0.98	1
0.9	0.80	0.75	0.77	0.83	0.87	0.95	0.98	0.99	1
	Effect size = 0.5								
−0.9	1	1	1	1	1	1	1	1	1

(Continues)

TABLE 5 (Continued)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
−0.8	0.95	1	1	1	1	1	1	1	1
−0.7	0.84	1	1	1	1	1	1	1	1
−0.6	0.72	0.99	1	1	1	1	1	1	1
−0.5	0.68	0.99	1	1	1	1	1	1	1
−0.4	0.62	0.96	1	1	1	1	1	1	1
−0.3	0.57	0.90	1	1	1	1	1	1	1
−0.2	0.56	0.88	1	1	1	1	1	1	1
−0.1	0.57	0.85	0.99	1	1	1	1	1	1
0	0.54	0.80	0.98	1	1	1	1	1	1
0.1	0.59	0.77	0.97	1	1	1	1	1	1
0.2	0.60	0.78	0.95	1	1	1	1	1	1
0.3	0.58	0.76	0.93	0.99	1	1	1	1	1
0.4	0.63	0.76	0.92	0.99	1	1	1	1	1
0.5	0.69	0.78	0.92	0.99	1	1	1	1	1
0.6	0.75	0.81	0.91	0.98	1	1	1	1	1
0.7	0.81	0.83	0.92	0.98	1	1	1	1	1
0.8	0.85	0.87	0.95	0.99	1	1	1	1	1
0.9	0.91	0.95	0.98	1	1	1	1	1	1

TABLE 6 Estimated power for AR(1) model with a trend change assuming effect size = 0.1, 0.25, 0.5 based on 1000 simulated data sets and statistical significance level 0.05, for two-arm interrupted time series study (testing $H_0 : \beta_{10} = \beta_{11} = 0$)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
	Effect size = 0.1								
−0.9	0.22	0.34	0.61	0.89	0.99	1	1	1	1
−0.8	0.18	0.23	0.39	0.61	0.82	1	1	1	1
−0.7	0.20	0.22	0.28	0.44	0.61	0.96	0.99	1	1
−0.6	0.21	0.18	0.25	0.34	0.53	0.87	0.96	0.99	1
−0.5	0.24	0.17	0.21	0.31	0.41	0.75	0.87	0.96	1
−0.4	0.23	0.18	0.23	0.27	0.38	0.65	0.80	0.92	0.99
−0.3	0.27	0.18	0.20	0.25	0.33	0.59	0.71	0.87	0.98
−0.2	0.28	0.21	0.18	0.22	0.29	0.52	0.64	0.78	0.94
−0.1	0.31	0.20	0.19	0.21	0.26	0.44	0.59	0.69	0.90
0	0.34	0.23	0.22	0.19	0.24	0.40	0.52	0.63	0.84
0.1	0.39	0.26	0.22	0.21	0.26	0.36	0.43	0.57	0.79
0.2	0.40	0.27	0.24	0.22	0.25	0.33	0.39	0.48	0.72
0.3	0.48	0.31	0.27	0.26	0.25	0.33	0.37	0.43	0.66
0.4	0.49	0.33	0.27	0.23	0.23	0.31	0.37	0.40	0.57
0.5	0.57	0.40	0.34	0.27	0.25	0.31	0.36	0.42	0.54
0.6	0.63	0.45	0.33	0.29	0.32	0.33	0.35	0.40	0.49

(Continues)

TABLE 6 (Continued)

	Sample Size								
Autocorrelation	*N*=18	*N*=27	*N*=36	*N*=45	*N*=54	*N*=72	*N*=81	*N*=90	*N*=108
0.7	0.65	0.52	0.44	0.38	0.35	0.35	0.36	0.40	0.48
0.8	0.71	0.56	0.52	0.44	0.42	0.39	0.41	0.42	0.48
0.9	0.75	0.67	0.60	0.59	0.54	0.54	0.54	0.54	0.58
	Effect size = 0.25								
−0.9	0.56	0.95	1	1	1	1	1	1	1
−0.8	0.39	0.75	0.98	1	1	1	1	1	1
−0.7	0.35	0.58	0.89	0.99	1	1	1	1	1
−0.6	0.32	0.46	0.75	0.96	1	1	1	1	1
−0.5	0.28	0.43	0.66	0.92	1	1	1	1	1
−0.4	0.32	0.38	0.59	0.83	0.98	1	1	1	1
−0.3	0.29	0.37	0.52	0.77	0.93	1	1	1	1
−0.2	0.35	0.32	0.47	0.70	0.91	1	1	1	1
−0.1	0.36	0.34	0.43	0.63	0.82	0.99	1	1	1
0	0.39	0.36	0.41	0.61	0.79	0.99	1	1	1
0.1	0.42	0.36	0.37	0.53	0.72	0.96	0.99	1	1
0.2	0.45	0.36	0.40	0.50	0.67	0.94	0.98	1	1
0.3	0.51	0.37	0.43	0.49	0.65	0.90	0.96	0.99	1
0.4	0.53	0.41	0.41	0.49	0.61	0.87	0.94	0.98	1
0.5	0.58	0.44	0.46	0.52	0.63	0.81	0.92	0.97	1
0.6	0.63	0.50	0.47	0.55	0.61	0.80	0.89	0.95	1
0.7	0.68	0.55	0.57	0.57	0.65	0.79	0.85	0.93	0.99
0.8	0.73	0.65	0.62	0.62	0.69	0.80	0.86	0.92	0.97
0.9	0.80	0.76	0.73	0.75	0.76	0.83	0.89	0.93	0.98
	Effect size = 0.5								
−0.9	0.96	1	1	1	1	1	1	1	1
−0.8	0.82	1	1	1	1	1	1	1	1
−0.7	0.65	0.98	1	1	1	1	1	1	1
−0.6	0.59	0.93	1	1	1	1	1	1	1
−0.5	0.49	0.87	1	1	1	1	1	1	1
−0.4	0.47	0.79	0.98	1	1	1	1	1	1
−0.3	0.45	0.74	0.97	1	1	1	1	1	1
−0.2	0.46	0.72	0.93	1	1	1	1	1	1
−0.1	0.46	0.63	0.90	1	1	1	1	1	1
0	0.51	0.64	0.88	0.99	1	1	1	1	1
0.1	0.53	0.62	0.84	0.96	1	1	1	1	1
0.2	0.54	0.59	0.82	0.96	1	1	1	1	1
0.3	0.58	0.60	0.78	0.94	1	1	1	1	1
0.4	0.60	0.63	0.80	0.92	0.99	1	1	1	1
0.5	0.64	0.68	0.78	0.90	0.98	1	1	1	1
0.6	0.69	0.64	0.78	0.89	0.97	1	1	1	1
0.7	0.72	0.75	0.82	0.91	0.96	1	1	1	1
0.8	0.78	0.79	0.86	0.94	0.97	1	1	1	1
0.9	0.85	0.88	0.92	0.97	0.99	1	1	1	1

intervention indicator to fractional values between 0 and 1 to accommodate a gradual effect.

Our method provides a convenient tool, with ready-to-use computer programs, for investigators to generate estimates of sample size to ensure sufficient statistical power for three-phase ITS studies. The R and SAS programs for conducting the simulation-based calculation in this article are included in the Supporting Information. Future investigators can easily modify these programs and conduct their own power and sample size calculation when needed. The simulation-based approach proposed herein can also be extended to the studies with multiple study arms.

6 | CONCLUSION

Power and sample size calculation can be conducted for three-phase ITS studies through the simulation-based methods presented in this manuscript. Results depend on whether the study evaluates the level change, trend change, or total change of an outcome is the target of the study, as well as on specified effect sizes and simulation parameters.

ACKNOWLEDGEMENTS

The research of authors was partially supported by the STRIDE project, which was funded through the National Institutes of Health Award No. 5 U01 TR001812. Dr Wei Liu's research was partially supported by the National Natural Science Foundation of China (grant nos. 11601106 and 91646106). Dr Bo Zhang's research was partially supported by the National Institutes of Health grant U24 AA026968 and the University of Massachusetts Center for Clinical and Translational Science grant UL1TR001453, TL1TR01454, and KL2TR01455.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Bo Zhang https://orcid.org/0000-0003-1574-198X

REFERENCES

1. Biglan A, Ary D, Wagenaar AC. The value of interrupted time series experiments for community intervention research. *Prev Sci*. 2000;1(1):31-49.
2. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther*. 2002;27(4):299-309.
3. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348-355.
4. Lu CY, Penfold RB, Toh S, et al. Near real-time surveillance for consequences of health policies using sequential analysis. *Med Care*. 2018;56(5):365-372.
5. Shwartz M, Setodji CM, Allison JJ. An Introduction to the point/counter-point/reply exchange by Dr Stone and Professor Lu and colleagues. *Med Care*. 2018;56(5):373-374.
6. Lu CY, Simon G, Soumerai SB, Kulldorff M. Counter-point: Early Warning Systems Are Imperfect, but Essential. *Med Care*. 2018;56(5):382-383.
7. Linden A, Adams JL. Applying a propensity score-based weighting model to interrupted time series data: improving causal inference in programme evaluation. *J Eval Clin Pract*. 2011;17(6):1231-1238.
8. Fretheim A, Soumerai SB, Zhang F, Oxman AD, Ross-Degnan D. Interrupted time series analysis yielded an effect estimate concordant with the cluster-randomized controlled trial result. *J Clin Epidemiol*. 2013;66(8):883-887.
9. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. *Acad Pediatr*. 2013;13(6):S38-S44.
10. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis. *BMJ*. 2015;350(jun09 5):h2750.
11. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. *J Clin Epidemiol*. 2011;64(11):1252-1261.
12. Department of Health and Human Services, National Institutes of Health. Monitoring Adherence to the NIH Policy on the Inclusion of Women and Minorities as Subjects in Clinical Research Comprehensive Report: Tracking of Clinical Research as Reported in Fiscal Year 2011 and Fiscal Year 2012; 2013.
13. Salman A, Nguyen C, Lee YH, Cooksey-James T. A review of barriers to minorities' participation in cancer clinical trials: implications for future cancer research. *J Immigr Minor Health*. 2016;18(2):447-453.
14. Luebbert R, Perez A. Barriers to clinical research participation among African Americans. *J Transcult Nurs*. 2016;27(5):456-463.
15. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *Am J Public Health*. 2014;104(2):e16-e31.
16. Huitema BE, Mckean JW. Design specification issues in time series intervention models. *Educ Psychol Meas*. 2000;60(1):38-58.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Zhang B, Liu W, Lemon SC, et al. Design, analysis, power, and sample size calculation for three-phase interrupted time series analysis in evaluation of health policy interventions. *J Eval Clin Pract*. 2019;1–16. https://doi.org/10.1111/jep.13266

APPENDIX A

Higher-order autoregressive model AR(p) for the random error term ϵ_t is specified as

$$\epsilon_t = \rho_1\epsilon_{t-1} + \rho_2\epsilon_{t-2} + \cdots + \rho_p\epsilon_{t-p} + u_t,$$

where ϵ_t is stationary, $\rho_1,\cdots,\rho_p$ are constants ($\rho_p \neq 0$), and the disturbances u_t independently and identically follow a normal distribution $N(0,\sigma^2)$. Specifically, the AR(1) model is

$$\epsilon_t = \rho\epsilon_{t-1} + u_t,$$

where ρ is the correlation coefficient between adjacent random error terms.

A nonstationary model can be used for the random error term ϵ_t. It is from an autoregressive integrated moving average (ARIMA) model if dth difference of the random error term ϵ_t, denoted by ε_t, is a stationary autoregressive moving average process. The ARIMA(p, d, q) model is generally specified as

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \cdots + \rho_p\varepsilon_{t-p} + u_t - \phi_1 u_{t-1} - \phi_2 u_{t-2} - \cdots - \phi_p u_{t-q}.$$

in which the disturbances $u_{t-q},\cdots,u_{t-1}$ independently and identically follow a normal distribution $N(0,\sigma^2)$. A special case that is commonly used is the ARIMA (1, 1, 1) model

$$\epsilon_t - \epsilon_{t-1} = \rho(\epsilon_{t-1} - \epsilon_{t-2}) + u_t - \phi_1 u_{t-1}.$$

An autoregressive conditional heteroscedasticity (ARCH) model for the random error term ϵ_t is generally specified as.

$$\epsilon_t = \sigma_{t|t-1}v_t\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \cdots + \alpha_q\epsilon_{t-q}^2$$

in which $\{v_t\}$ is a sequence of independently and identically distributed random variables with zero mean and one variance (eg, standard normal distribution), $\alpha_0,\cdots,\alpha_q$ are constants, and σ_t^2 is the conditional variance of ϵ_t on $\epsilon_{t-q},\cdots,\epsilon_{t-1}$. A special case that is commonly used is the ARCH(1) model:

$$\epsilon_t = \sigma_{t|t-1}v_t \;\; \sigma_{t|t-1}^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2.$$

APPENDIX B

The data collected from a three-phase two-arm ITS study can be analysed by the segmented time series regression models with the following form

$$\begin{aligned} Y_t = {} & \beta_0 + \beta_1 T_t + \beta_2 X_{t^{(1)}} + \beta_3 X_{t^{(2)}} + \beta_4 (T_t - t_1) X_{t^{(1)}} + \beta_5 (T_t - t_2) X_{t^{(2)}} \\ & + \beta_6 G + \beta_7 G T_t + \beta_8 G X_{t^{(1)}} + \beta_9 G X_{t^{(2)}} + \beta_{10} G (T_t - t_1) X_{t^{(1)}} \\ & + \beta_{11} G (T_t - t_2) X_{t^{(2)}} + \epsilon_{tG} \end{aligned}$$

in which Y_t represents the aggregated outcome variable that is measured over time, G is the binary indicator for treatment group (G = 1) versus control group (G = 0), T_t is the actual or converted study time from the start to the end of the study, $X_{t^{(1)}}$ is a binary indicator for the second phase of the study, $X_{t^{(2)}}$ is a binary indicator for the third phase of the study, t_1 is the first time point after the onsite of first-level intervention, t_2 is the first time point after the onsite of second-level intervention, and ϵ_{tG} represent two the random error terms (ϵ_{t1} for treatment group; ϵ_{t0} for control group). The coefficients β_0 to β_5 characterize the starting intercept and slope before intervention in control group, and the change in intercept and slope after the time points of onsite of two phases of intervention in control group. Their specific interpretation is identical to the coefficients β_0 to β_5 in Section 3.1, but for control group. That means the changes, if there are any, happen spontaneously and are not triggered by the intervention. The coefficients β_6 to β_{11} represent the difference in these quantities between treatment group and control group. The coefficient β_6 represents the difference in the intercept or level of the aggregated outcome variable between treatment group and control group prior to the intervention, β_7 represents the difference in the slope or trend of the aggregated outcome variable between treatment group and control group prior to the intervention, β_8 and β_9 indicate the difference between treatment group and control group in the level of the aggregated outcome variable immediately after the onsite of the first-level and second-level intervention, respectively, and β_{10} and β_{11} represent the difference between treatment group and control group in the trend change of the aggregated outcome variable after the onsite of the first level and second level, respectively. The focus of the three-phase two-arm ITS analysis is to examine the significance of β_8 and β_9, or the summation of them, because they indicate the immediate treatment effect of first-level and second-level intervention in terms of level change and the significance of β_{10} and β_{11}, or their summation, that indicate the treatment effect in terms of change in trend. In the two-arm ITS analysis, the random error terms ϵ_{t0} and ϵ_{t1} are independent with each other, and each can be separately specified to follow a time series process that are describe in Section 3.1.