RESEARCH ARTICLE

A robust interrupted time series model for analyzing complex health care intervention data

Maricela Cruz1 | Miriam Bender2 | Hernando Ombao1,3

1Department of Statistics, University of California, Irvine, CA, USA
2Sue & Bill Gross School of Nursing, University of California, Irvine, CA, USA
3Statistics Program, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Correspondence
Hernando Ombao, University of California, Irvine and King Abdullah University of Science and Technology, Saudi Arabia.
Email: hombao@uci.edu; hernando.ombao@gmail.com

Funding information
Commission on Nurse Certification; Eugene Cota-Robles Fellowship; NSF Graduate Research Fellowship, Grant/Award Number: DGE-1321846; NSF MMS, Grant/Award Number: 1461534; NSF DMS, Grant/Award Number: 1509023

1 INTRODUCTION

Current health policy calls for greater use of evidence-based care delivery services to improve patient quality and safety outcomes. Care delivery is complex, with interacting and interdependent components that challenge traditional statistical analytic techniques, in particular, when modeling a time series of outcomes data that might be “interrupted” by a change in a particular method of health care delivery. Interrupted time series (ITS) is a robust quasi-experimental design with the ability to infer the effectiveness of an intervention that accounts for data dependency. Current standardized methods for analyzing ITS data do not model changes in variation and correlation following the intervention. This is a key limitation since it is plausible for data variability and dependency to change because of the intervention. Moreover, present methodology either assumes a prespecified interruption time point with an instantaneous effect or removes data for which the effect of intervention is not fully realized. In this paper, we describe and develop a novel robust interrupted time series (robust-ITS) model that overcomes these omissions and limitations. The robust-ITS model formally performs inference on (1) identifying the change point; (2) differences in preintervention and postintervention correlation; (3) differences in the outcome variance preintervention and postintervention; and (4) differences in the mean preintervention and postintervention. We illustrate the proposed method by analyzing patient satisfaction data from a hospital that implemented and evaluated a new nursing care delivery model as the intervention of interest. The robust-ITS model is implemented in an R Shiny toolbox, which is freely available to the community.

KEYWORDS
complex interventions, health care outcomes, intervention analysis, segmented regression, time series
patient health care outcomes in terms of research design and statistical analysis. Methodologies capable of managing data interdependency are being developed yet are still considered less robust than traditional methods, which assume that intervention causal factors can be analyzed without consideration of their participant samples and contexts. Interrupted time series (ITS) design has emerged as a quasi-experimental methodology with the strongest power to infer causality without stripping contextual and temporal factors from the analysis.

Segmented regression is the most popular statistical method for analyzing time series data of health care interventions. While powerful, there are limitations to this approach, namely, that it restricts the interruption to a predetermined time point in the series or removes the set of time points for which the intervention effects may not be realized and neglects the plausible differences in autocorrelation and variability present in the data.

In this paper, we develop the robust interrupted time series (robust-ITS) model, which is a novel model for ITS. One advantage of the robust-ITS model compared to previous methods is its ability to estimate, rather than assume a priori, the time when the effect of intervention initiates (the change point). In practice, the change point may occur either before or after the official intervention time. For instance, an intervention intended to improve care quality requiring a training over several months or weeks may already produce a change in the outcome even before the formal intervention time (before the official start of intervention) if the trainees execute their training as they learn.

We propose a method that regards the change point as variable, appropriate for situations where the data warrants such treatment. Robust-ITS allows us to test when the effect of the intervention initiates in situations when pinpointing the change is of interest. Nonetheless, if the aim is to make causal inference, it may be better to prespecify the change point or remove the set of possible change points (or the set of points for which the intervention has not fully been realized) from the analysis, as in traditional segmented regression for ITS designs.

The main contributions of robust-ITS are the formal tests for differences in the correlation structure and variability between the pre- and post-change point.

The data used for the model development come from a study aimed to determine the influence of redesigning a nursing care delivery system on nationally endorsed quality and safety metrics. Many nationally endorsed metrics must be publically reported and tracked on a monthly basis via aggregate rates, counts, or ratios. The specific data used for this modeling procedure were patient satisfaction survey scores. Patient satisfaction is an important health outcome, providing a valid measure of quality of care received and has previously been used for ITS analysis of nursing care delivery interventions. It is also a metric that is currently being used to calculate health systems reimbursement for care services, via the Center for Medicaid and Medicare Services (CMS) Value Based Purchasing Program, making it a significant focus for improvement. A time series plot of patient satisfaction scores from January 2008 to December 2012 at a number of units in a health care system is given in Figure 1.

FIGURE 1 The time series of observed average patient satisfaction for each unit, the estimated change point, estimated means, and formal intervention time. The estimated means and change point are obtained from modeling the time series with robust-ITS [Colour figure can be viewed at wileyonlinelibrary.com]
The unit of analysis in the study is the care delivery microsystem, or hospital “unit.” Patient satisfaction scores are reported as aggregate scores per month, per unit. Patient satisfaction indicators include “nurse communication,” “skill of the nurse,” and “pain management.” Patients respond to items by selecting 1 of 4 responses: never, sometimes, usually, or always.

For modeling, we chose one outcome, the average patient satisfaction over 7 patient satisfaction indicators, for 4 hospital units of the study setting. We refer to the average patient satisfaction score simply as patient satisfaction throughout the remainder of the paper. The intervention program, titled Clinical Nurse Leader (CNL) integrated care delivery, was the introduction of novel nursing care delivery policies and procedures into the hospital and its units.11 Importantly, CNL students conducted their Masters level microsystem change project, prior to the formal intervention implementation time, in the same unit they would be working on as part of the care delivery redesign intervention. This may or may not have influenced the change point of the intervention effect and was therefore considered a good test case for modeling purposes.

The remainder of this paper is organized as follows. First, we present a background of studies on interrupted time series in health care. Current statistical methods and their limitations will be discussed. Then our proposed robust-ITS model is described. Details on the estimation and inference procedure are provided. Followed by an analysis of the impact of Clinical Nurse Leader on patient satisfaction with nurse communication. Parameter estimates are presented and compared to results obtained via traditional ITS methodologies. Lastly, a summary of the robust-ITS model and a brief description of future work is provided.

2 | BACKGROUND

The traditional “gold standard” for evidence generation of health care interventions is the randomized clinical trial (RCT). The theory behind this methodology is that potential biases related to patient heterogeneity and confounding covariates are evenly dispersed across study groups and thus do not dissimilarly influence treatment effect.12 In statistical terms, the RCT design tests for the difference in outcomes between 2 groups—those exposed to the treatment and those not—completely ignoring underlying variability. However, RCTs have a narrow scope in the care delivery community since it is not feasible, and sometimes not ethical, to randomly assign the intervention. By design, explanatory RCTs do not and cannot take into account the range of dimensions of patient demographics, variations in health, and overall health care complexity.13

Interrupted time series offers a rigorous methodology to determine the effectiveness of complex health care interventions on outcomes in real-world settings that account for secular changes as part of the analytic process.6,14 When RCTs are not feasible or not applicable, ITS is considered the strongest research design in the health policy evaluation literature7 and are considered rigorous enough for inclusion into Cochrane meta analyses.15

The metrics adopted by the health policy evaluation literature to assess the effect size of an intervention via ITS are level change and trend change (change in slopes). A change in outcomes is referred to as a level change and is analogous to the difference in mean scores before and after the intervention, with independent data values. The level change is interpreted as the jump between the projected mean based on the pre–change point phase and the estimated mean post–change point. Our definition of level change is graphically depicted in Figure 2. While the level change identifies the size of an intervention’s effect, the change in trend quantifies the impact of the intervention on the overall mean. It is necessary to report both level change and change in trend to interpret the results of an ITS study accurately.16

2.1 | Limitations of segmented regression approaches

We first note that Ramsay et al “demonstrated that ITS designs are often analyzed inappropriately, underpowered, and poorly reported in implementation research.”17 We believe, along with many authors in this field, that segmented regression is most effective in analyzing ITS data. Segmented regression may be used via standard statistical packages—such as ITSA in Stata, ETS in SAS, and segmentedR in R—however, as we demonstrate in this paper, there are limitations to these current statistical packages. Segmented time series regression, or regression-discontinuity analysis, was first introduced by Quandt closely followed by Thistlethwaite and Campbell.18,19 Since then, regression-discontinuity analysis has been used in many forms, with many different parameterizations, in health services research (as already described) and other fields, such as economics and education.
We discuss the potential limitations of the current regression modeling approaches; some of which are described in a recent review by Kontopantelis et al.\textsuperscript{14} In particular, regression modeling approaches assume that the preintervention mean is linear and that the characteristics of the population remain unchanged throughout the study period. Segmented time series regression also assumes that there is a distinct separation between the preintervention and postintervention phases—by either assuming the time point at which the effect of the intervention initiates (ie, the change point) is known or removing the set of possible change points—and that there are 2 regressions present, but one overall variance and autocorrelation. These assumptions, if violated, could lead to incorrect inference and interpretation of results.

The specification of the change point as the time of intervention in segmented regression does not represent the reality that complex interventions may have varied effects and take time to manifest change and can therefore lead to incorrect measures of the intervention's effect on the system. Current statistical programs often assume an instantaneous intervention effect—a change point set to intervention time—because change point estimation involves optimization over all possible configurations, which challenges computational feasibility. Recall that the change point is the time point at which the effect of the intervention initiates. Prevalent approaches to overcoming this limitation are to remove, or censor, a specific set of time points from the analysis.\textsuperscript{6,7} This censoring not only omits data but it also potentially biases the parameter estimates, as the study team decides which time points to remove.

The assumption of a constant correlation structure (variance and autocorrelation) is not necessarily representative of complex interventions, where the system seldom reacts in an isolated way to change, and the intervention is expected to reduce variability in the system. In fact, with complex interventions, often the goal is to enhance care processes so that elements become more dependent and consistent over time; theoretically, the correlation structure should differ based on changes,\textsuperscript{20} such as an intervention. An increase in data dependency and consistency implies a difference in autocorrelation and variability. Thus, detection of differences in autocorrelation and variances between preintervention and postintervention are critical in evaluating the effectiveness of an intervention. Table 1 highlights a few popular ITS packages, some articles that describe the use of the packages for analysis of ITS data, and the limitations of each method, as already described.

### 3  THE ROBUST-ITS MODEL

#### 3.1  Preliminary analysis

Before any formal statistical modeling, the outcome should be plotted against time to illuminate the type of longitudinal mean (linear, quadratic, etc), seasonality, and the set of plausible change points. The set of possible change points should not be limited to time points solely after the intervention, for aforementioned reasons. If the longitudinal mean is not linear, an adequate transformation may be applied to obtain a linear pattern, or a different segmented regression model appropriate for the pattern present needs to be applied within the ITS design. Seasonality should be accounted for, within the mean, via traditional statistical methods concisely described in Bhaskaran et al.\textsuperscript{20} If needed, one should apply variance stabilizing transforms to the outcome variable. For the purposes of illustrating robust-ITS, the relationship between the outcome and time is assumed linear with no seasonality.
### TABLE 1

The limitations and advantages of ITS packages that focus on segmented regression and gives a few papers in which the packages have either been proposed or used

| Package                  | Papers                              | Advantages/Description                                                                 | Limitations                      |
|--------------------------|-------------------------------------|----------------------------------------------------------------------------------------|----------------------------------|
| SAS PROC AUTOREG\(^{21}\) | Penfold & Zhang\(^{7}\) Shardell et al\(^{22}\) Parienti et al\(^{23}\) | • Estimation and prediction of linear regression models with autoregressive errors.  
• Estimation and testing of general heteroscedasticity (change in variance). | • No intervention analysis.      |
| SAS PROC ARIMA\(^{21}\)  | Shardell et al\(^{22}\)             | • Analyzes and forecasts time series, transfer functions, and intervention data using ARIMA and ARMA models. | • Assumes intervention time is fixed with an immediate effect.  
• Assumes one overall correlation structure. |
| SAS ETS\(^{21}\)          | Cable,\(^{24}\) Mahamat et al,\(^{25}\) and Aboagye-Sarfo et al\(^{26}\) | • Same as the above 2 entries; SAS PROC ARIMA and SAS PROC AUTOREG are part of SAS ETS. | • Fixed intervention time with immediate effect.  
• One overall correlation structure          |
| Stata ITSA                | Linden\(^{27}\)                     | • Single and multiple group comparisons.  
• Estimates treatment effects for multiple treatment periods. | • Fixed change point. |
| segmentedR               | Muggeo\(^{28}\)                     | • Adjust for overall autocorrelation.  
• Estimates piecewise regression models with a fixed number of discontinuities, or interruptions. | • No modeling of correlation structure.  
• Assumes data are independent. |

Abbreviations: ITS, interrupted time series. 

### 3.2 Description of the robust-ITS model

One prominent feature of our approach in the robust-ITS model is the clear distinction between the time of intervention and the change point. In Penfold and Zhang,\(^{7}\) Garey et al,\(^{30}\) Ansari et al,\(^{31}\) and many more, the impact of the intervention is assumed to be instantaneous—that is, the change point is assumed to be the intervention time.\(^{7,30,31}\) Robust-ITS allows us to estimate the time point at which the effect of an intervention initiates.

The paramount contribution of robust-ITS is the modeling of the stochastic component separately between the pre– and post–change point phases. The separate modeling allows for 2 completely different data dependency and variability structures to exist prior to the intervention and post-intervention.

Denote \(t^*\) as the time point at which the intervention is introduced and \(\tau\) as the time point at which the effect of the intervention initiates (the change point). Sometimes, it may indeed be true that \(t^* = \tau\), but not necessarily. Often it is entirely possible that the time of effect of the intervention differs from the time of intervention introduction (ie, either \(\tau > t^*\) or \(\tau < t^*\)). Here, we develop a data adaptive procedure for estimating \(\tau\). There are many change-point detection methods in time series, but they often deal only with changes in the mean and variance (not the autocorrelation structure itself) and may not work well in shorter time series.\(^{32,33}\)

Define \(Y_t\) as the outcome of interest at time \(t\); for example, \(Y_t\) may be patient satisfaction at a particular hospital unit during time \(t\). The general regression is defined as

\[
Y_t = \mu_t + \epsilon_t,
\]

where \(\mu_t\) is the mean and \(\epsilon_t\) is the stochastic process. The mean component, \(\mu_t\), characterizes the mean of the outcome for the preintervention and postintervention phases. The stochastic process, \(\epsilon_t\), accounts for the outcome variability and correlation. In the following discussion, we define the mean and stochastic components for the robust-ITS model. A note on the length of the time series needed to perform the robust-ITS analysis is provided in Appendix A.

### 3.2.1 The preintervention and postintervention mean

At the first stage of modeling, the emphasis is on the mean,
\[
\mu_t = \begin{cases}
\beta_0 + \beta_1 t, & t < \tau \\
(\beta_0 + \delta) + (\beta_1 + \Delta) t, & t \geq \tau,
\end{cases}
\]

where the parameters are estimated using ordinary least squares. The parameters in \( \mu_t \) are as follows: (1) \( \beta_0 \), the intercept of the mean prior to the change point; (2) \( \beta_1 \), the slope of the outcome prior to the change point; (3) \( \beta_0 + \delta \), the intercept of the postintervention phase; and (4) \( \beta_1 + \Delta \), the slope of the postintervention phase.

Remark 1. (1) The difference between the pre–change point and post–change point intercept is \( \delta \); (2) the difference between the pre– and post–change point slopes is \( \Delta \); and (3) the difference in the mean level (preintervention minus postintervention) is \( -\delta - \Delta \tau \), the level change.

Recall that difference in mean level (or level change) is 1 of the 2 metrics in health policy evaluation literature used to measure the effect size of an intervention. Formally, the level change is defined as the difference at the change point time \( \tau \) between the extrapolated preintervention mean level and the observed intervention mean level, as is depicted in Figure 2.

Rather than impose or assume the actual onset of the change, the robust–ITS model actually estimates the change point \( \tau \) in a data-driven manner using the likelihood approach. From a set of candidate change points (set by the researcher), the procedure estimates the parameters via ordinary least squares for each possible \( \tau \), and selects the \( \tau \), and its corresponding parameters that maximize the likelihood.

Denote the length of the time series as \( T \) and let \( \theta = [\beta_0, \beta_1, \delta, \Delta, \sigma_1^2, \sigma_2^2]' \), with \( \sigma_1^2 \) and \( \sigma_2^2 \) defined as the variances prior to and post change point, respectively. As described in Section 2, one goal of interventions is to decrease variability, which leads to creating a more consistent outcome. We therefore include separate variance parameters for the pre– and post–change point phases, to allow for a change in data variability.

Let \( q \) be a candidate change point in the set of possible change points \( Q \), where \( Q = \{t^*-m, \ldots, t^*, \ldots t^*+k\} \) for positive integer values of \( m \) and \( k \) set by the researcher. For each candidate change-point \( q \) we derive the likelihood function:

\[
L(\theta|q) = \left( \frac{1}{\sqrt{2\pi\sigma_1^2}} \right)^{q-1} \exp \left( -\frac{1}{2\sigma_1^2} \sum_{t=1}^{q-1} (Y_t - (\beta_0 + \beta_1 t))^2 \right) \times \left( \frac{1}{\sqrt{2\pi\sigma_2^2}} \right)^{T-(q-1)} \exp \left( -\frac{1}{2\sigma_2^2} \sum_{t=q}^{T} (Y_t - (\beta_0 + \delta) + (\beta_1 + \Delta) t)^2 \right).
\]

Define \( L_{\text{max}}(q) = \max_{\theta} L(\theta,|q) \), then the estimated change point is \( \hat{\tau} = \arg \max_{q\in Q} L_{\text{max}}(q) \).

The estimates of the intercept and slope for each phase are obtained as in segmented regression; equivalent to estimating the slope and intercept separately for the pre– and post–change point phases as in simple linear regression. The ordinary least squares (OLS) estimates for the parameters in \( \theta \) are provided in Appendix A. The estimates for \( \sigma_1^2 \) and \( \sigma_2^2 \) depend on the stochastic process and are given for an AR(1) process also in Appendix A.

The presence of \( \tau \) does not restrict the model to a fixed interruption with an instantaneous effect and allows the design matrix and estimates to transform based on the information the data provides. This flexibility of the model can be helpful in minimizing misleading results from an assumed change point.

### 3.2.2 Stochastic properties pre– and post–change point

The stochastic component, \( \epsilon_t \), captures the correlation structure of the outcome variable across time and may change as a result of the intervention. Here, we shall develop a formal test for the difference in the correlation structure for preintervention and postintervention phases.

We use the ARIMA process to model the stochastic component, \( \epsilon_t = Y_t - \mu_t \). Since the mean function \( \mu_t \) is not known (we only have its estimate, \( \hat{\mu}_t \)), the stochastic component is not directly observed. In place of \( \epsilon_t \), we use the residuals, \( R_t = Y_t - \hat{\mu}_t \), where \( \hat{\mu}_t \) is the estimate of \( \mu_t \) obtained as described in stage 1. To use the ARIMA processes, residuals must exhibit stationary behavior, that is, the mean and variance of the residuals must be relatively constant. If the mean is not misspecified, then the residuals should be fluctuating around zero without any patterns. Moreover, the residuals should be stationary within each of the preintervention and postintervention phases.\(^{34}\)

Autoregressive conditional heteroscedasticity models may be used when the data are nonstationary; they can model the stochastic component in each phase and for the entire observational period when the variance and/or data dependency change.
is nonconstant. For our patient satisfaction data, it is reasonable to assume stationarity within each phase, and, hence, we proceed with the assumption of stationarity. See Shumway and Stoffer, Granger and Newbold, and Bollerslev in for more details on autoregressive conditional heteroscedasticity models.

Because of the impact of the intervention, the stochastic process $\epsilon_i$, preintervention might differ from the process postintervention. That is, $\epsilon_i$ for $t \in \{1, \ldots, \hat{\tau} - 1\}$ may be a different stochastic process than $\epsilon_i$ for $t \in \{\hat{\tau}, \ldots, T\}$. Hence, the autocorrelation and variance might differ pre– and post–change point. Now, the stationarity requirement is satisfied if the variance, mean, and autocorrelation are constant within each stochastic process, not constant across all time points as before.

The ARIMA parameters are estimated by maximizing the conditional likelihood and are given for the subsequent example of an autoregressive model with a lag of one. It is of most importance to understand that the lag used for the autocorrelation modeling is not an indicator of when the intervention takes effect, but, instead, it models overall data dependency; $\tau$ dictates when the intervention affects the outcome variable.

**Example 1.** A special case of a stochastic process is the first order auto-regressive [AR(1)] model:

$$R_t = \begin{cases} \phi_1 R_{t-1} + \epsilon_t^1, & 1 < t < \hat{\tau} \\ \phi_2 R_{t-1} + \epsilon_t^2, & \hat{\tau} < t \leq T. \end{cases}$$

To ensure causality in the time series sense, both $\phi_1$ and $\phi_2$ lie in the interval $(-1, 1)$.

**Remark 2.** Once the $\epsilon_i$ are appropriately modeled, the OLS estimates in Equations A1 to A4 will need to be reestimated to produce the generalized least squares estimates. If an AR(1) is fit to the overall stochastic process (across the change point), the beta parameters should be reestimated without the first time point; that is, for $t$ in $\{2, \ldots, T\}$. If different AR(1) processes are fit preintervention and postintervention, then the mean prior to the intervention should be reestimated using $t$ in $\{2, \ldots, \tau - 1\}$, and the mean postintervention reestimated using $t$ in $\{\tau + 1, \ldots, T\}$. The summation limits in Equations A1 to A4 would therefore change.

### 3.2.3 Preintervention and postintervention variance comparison

For stochastic processes in which both pre– and post–change point phases are adequately modeled by the ARIMA processes (residuals not behaving as white noise in either phase), the variances may not be easily, if at all, compared. The variances in each phase can be estimated but not statistically compared, because of the dependency of the data.

If there is no autocorrelation (or dependence), then the OLS estimates are sufficient. Nevertheless, the variance may not be the same pre– and post–change point. In situations where there is no statistically significant autocorrelation, the variances may be compared via an $F$ test. Using $r$, we can determine how many observations we have prior to and post change point, subtracting three (one for each parameter we estimate) from those values gives the degrees of freedom. For example, suppose that there are 25 and 35 time points before and after the change point, respectively, and that the estimated variances are $s_1$ and $s_2$, respectively. Then the $F$ statistic is $\frac{s_1}{s_2}$ and under the null hypothesis (assuming the variances are equal) distributed $F_{22,32}$. 


Patient satisfaction is modeled in 4 hospital units labeled as follows: Unit 1, Unit 2, Unit 3, and Unit 4. It is crucial to note that the outcome is a percentage and so restricted to lie between 0 and 100. The restriction on the outcome has imperative consequences: the time series must reach a plateau regardless of intervention introduction. The nature of the outcome must be kept in mind when interpreting the results from the analysis.

The means of the 4 time series were modeled as in Equation 1; the resulting parameter estimates are given in Table 2. The relationship between the formal intervention implementation time and the change point for the 4 units is illuminated in Tables 2 and 3, which show that the effect of the intervention is not necessarily instantaneous. In fact, Tables 2 and 3 suggest that the intervention had an anticipatory effect in 3 of the 4 units of interest. The preemptive effect is in concordance with the structure of the CNL integrated care delivery intervention, because of the CNL student inclusion into their respective units 6 months prior to the formal introduction. In Units 1, 2, and 3, the estimated change points occur, respectively, in May 2010, January 2010, and February 2010, suggesting that CNL students could have implemented the new care delivery prior to July 2010. This relationship indicates the time of change in patient satisfaction associated with the intervention may be at the mercy of CNL student behavior.

Table 2 depicts the differences in estimated means prior to and post change point, with the most informative rows of Table 2 corresponding to the 2 standardized effect sizes: change in level and change in slopes. The level change is positive and statistically significant (at the $\alpha = 0.01$ level) for Units 1 and 4, indicating that the mean drops at the change point and that the drop statistically differs from zero. Thus, the CNL integrated care delivery initially is associated with a statistically

| Parameters       | Patient Satisfaction | Unit 1  | Unit 2   | Unit 3    | Unit 4  |
|------------------|----------------------|--------|---------|----------|--------|
| Intercept pre    |                      | 64.32**| 64.67** | 68.31**  | 77.21**|
| Change point     |                      | (61.76, 66.88) | (60.07, 69.27) | (64.14, 72.49) | (74.99, 79.54) |
| Intercept post   |                      | 67.21**| 71.79** | 71.51**  | 77.42**|
| Change point     |                      | (61.86, 72.57) | (66.10, 77.49) | (64.01, 79.01) | (70.24, 84.60) |
| Change in Intercepts, $\delta$ |                | 2.89   | 7.12    | 3.20     | 0.15   |
| Change in level, $-\delta - \Delta \tau$ |           | 7.00** | -2.77   | 3.50     | 5.40** |
| Slope pre       |                      | 0.56** | 0.24    | 0.35*    | 0.28** |
| Change point    |                      | (0.41, 0.71) | (-0.08, 0.56) | (0.07, 0.63) | (0.15, 0.40) |
| Slope post      |                      | 0.22** | 0.07    | 0.09     | 0.10   |
| Change point    |                      | (0.10, 0.34) | (-0.06, 0.20) | (-0.07, 0.26) | (-0.06, 0.25) |
| Change in slope, $\Delta$ |              | -0.34** | -0.17   | -0.26    | -0.18  |
| Delay in effect of intervention, $\tau - t^*$ |         | -3     | -6      | -5       | 0      |

Since $\tau$ is discrete, only an estimate is given, no confidence interval. The asterisk, *, denotes statistical significance at the $\alpha = .05$ level.

| Parameters       | Patient Satisfaction | Unit 1  | Unit 2   | Unit 3    | Unit 4  |
|------------------|----------------------|--------|---------|----------|--------|
| Time of intervention implementation |          | Month 31, | Month 31, | Month 31, | Month 31, |
|                  |                      | July 2010 | July 2010 | July 2010 | July 2010 |
| Estimated change point, $\tau$ |        | Month 29, | Month 25, | Month 26, | Month 31, |
|                  |                      | May 2010 | January 2010 | February 2010 | July 2010 |
significant drop of patient satisfaction in Units 1 and 4. The estimated trend change or change in slopes is negative for each unit, although statistically significant (at the $\alpha = 0.01$ level) for Unit 1 only.

The slope decreases after the estimated change point in Unit 1, implying a more flattened out mean post-change point. Therefore, the CNL implementation may be associated with a flatter mean across time in Unit 1; ie, for every 1-month increase in time, there is a smaller estimated increase in patient satisfaction in the post-change point phase as compared to the pre-change point phase. However, this artifact may be present because the maximum value of the outcome variable is 100. We may be seeing some asymptote effect instead of capturing the effect of the intervention on the trend (slope).

For Units 2, 3, and 4, the estimated slope does not statistically change after the estimated change point; for Units 2 and 3, the estimated level change is also not statistically significant; and the estimated change in intercepts is not statistically significant for any of the units. Hence, for Units 2 and 3, the intervention does not seem to be associated with a change in the estimated patient satisfaction. The CNL integrated care delivery is associated with some outcome modification (either in the intercept, level change, slope change, or a combination) in Units 1 and 4.

The pre- and post-change point regressions of the 4 units are plotted in Figure 1. Figure 1 depicts that the change point occurs prior to the formal intervention time for Units 1, 2 and 3 but is equivalent to the formal intervention time for Unit 4. The estimated mean post-change point seems to flatten out in all units, and the change in level appears sizable for Units 1 and 2. Figure 1 illustrates results in concurrence with those of Table 2.

Figure 5 provides the studentized residuals after modeling the mean. The residuals seem well behaved and mostly contained between the rule of thumb $\pm 2$ and completely contained between $\pm 3$. The residuals do not exhibit any severe patterns and thus suggest robust-ITS models the mean patient satisfaction of all units adequately. Moreover, Figure 6 provides the autocorrelation function (ACF) of the residuals. The ACF plots interpreted as Shumway and Stoffer act as patterns and thus suggest robust-ITS models the mean patient satisfaction of all units adequately.

The (1) estimates and 95% confidence intervals of the AR(1) coefficients pre-change point, post-change point, and of the estimated increase in the AR(1) coefficient post-change point; (2) the estimated variances; and (3) the $F$ statistic and $P$ value corresponding to the comparison of the pre- and post-change point variances, for patient satisfaction with effective nurse communication

| Parameters                                      | Patient Satisfaction |
|-------------------------------------------------|----------------------|
| Unit 1                                          | Unit 2 | Unit 3 | Unit 4 |
| AR(1) coefficient pre-change point, $\hat{\phi}_1$ | ...    | ...    | ...    |
| $(-0.460, 0.348)$                               | ...    | ...    | ...    |
| AR(1) coefficient post-change point, $\hat{\phi}_2$ | ...    | ...    | ...    |
| $(-0.713, 0.004)$                               | ...    | ...    | ...    |
| Difference in AR(1) coefficients, $\hat{\phi}_2 - \hat{\phi}_1$ | ...    | ...    | ...    |
| $(-0.826, 0.229)$                               | ...    | ...    | ...    |
| Variance pre-change point, $\sigma^2_{1:(f-1)}$ | 10.259 | 26.474 | 23.412 | 8.127 |
| Variance post-change point, $\sigma^2_{f:T}$    | 7.976  | 13.511 | 23.965 | 12.649 |
| Variance comparison                             | 1.286  | 1.959  | 0.977  | 0.643 |
| $F$ statistic ($P$ value)                       | .248   | .035   | .516   | .88   |
result of the intervention since there will be better quality control on the fluctuations of the patient outcomes and more consistency as a result of the intervention.

4.1 Comparing robust-ITS to segmented regression

We compare the standardized effect sizes between robust-ITS and segmented regression (both with an assumed change point and the set of possible change points removed) in Table 5. The aim of Table 5 is to illustrate that the estimates of level change and trend change differ based on the type of model selected. Indeed, the estimates of level change and trend change across the 3 models differ for each of the 4 units. Segmented regression—with an assumed change point or the set of possible change points removed—may provide results that are statistically significant, or not statistically significant, in cases where the opposite is true when considering anticipatory or delayed intervention effects. Moreover, the 2 segmented regression methods may also provide opposing results.

It is important to note that there are many model specifications used for segmented regression. Two of the main models used for segmented regression in the ITS and health care literature are discussed and shown to be equivalent in Appendix A. The segmented regression models are discussed under the assumption that the change point is assumed. Nonetheless, the 2 main segmented regression models are also equivalent when the set of possible change points are removed.

The true model comparisons are provided in Table 6, intended to compare the adequacy of robust-ITS and segmented regression. Mean squared error (MSE)—the estimate of sum of squared errors, which measures the square of the deviations from the estimate mean, divided by the degrees of freedom—is provided in Table 5 for robust-ITS, segmented regression with an assumed change point, and segmented regression with the set of possible change points removed. Robust-ITS has the smallest MSE and so provides the best estimate for the mean of patient satisfaction, suggesting that robust-ITS models the data better than either of the traditional segmented regressions (with an assumed change point, or the set of possible change points removed).

4.2 Comparing robust-ITS to a quadratic model with no change point

We further compare robust-ITS to a non–change point model with quadratic time as a predictor for completeness. The model for the mean of patient satisfaction of a given unit with quadratic time as a predictor is

\[ \mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad \text{for} \quad t \in \{1, \ldots, 60\} \]
The mean squared error (MSE), with order of magnitude $10^{-5}$, of (1) segmented regression with the phase-in period removed, denoted by $+$; (2) segmented regression with an assumed change point, denoted by $++$; and (3) robust-ITS, for patient satisfaction.

| Unit | Mean squared error |
|------|--------------------|
|      | Segmented regression $+$ | Segmented regression $++$ | Robust-ITS |
| Unit 1 | 192.42 | 190.99 | 171.99 |
| Unit 2 | 428.34 | 416.50 | 406.85 |
| Unit 3 | 471.80 | 471.93 | 459.16 |
| Unit 4 | 181.44 | 161.11 | 161.11 |

Mean squared error is the estimate of sum of squared errors, which measures the square of the errors or deviations, divided by the degrees of freedom. A lower value of MSE for a model suggests a more adequate fit.

**FIGURE 3** Plots of patient satisfaction within each of the 4 units, along with the estimated means obtained by robust-ITS and a model with quadratic time as a predictor [Colour figure can be viewed at wileyonlinelibrary.com]

The estimated patient satisfaction mean curves for both robust-ITS and Model 3 are plotted in Figure 3 by unit. The parameter associated with quadratic time $\beta_2$ is only statistically significant, at the $\alpha = 0.05$ level, for Unit 1. Including quadratic time as a predictor is not necessary for Units 2, 3, and 4, since we cannot conclude that $\beta_2$ differs from zero. Adding quadratic time as a predictor is useful in Unit 1 because at the $\alpha = 0.05$ level $\beta_2$ differs from zero.

Nevertheless as shown in Table 7, the MSE (estimate of the sum of squared errors divided by the degrees of freedom) for robust-ITS is smaller than the MSE of Model 3 in all units, indicating robust-ITS fits the data better in all units.

| Unit | Mean squared error |
|------|--------------------|
|      | Model with quadratic time and no change point | Robust-ITS |
| Unit 1 | 224.91 | 171.99 |
| Unit 2 | 417.77 | 406.85 |
| Unit 3 | 460.36 | 459.16 |
| Unit 4 | 184.62 | 161.11 |

Mean squared error is the estimate of sum of squared errors, which measures the square of the errors or deviations, divided by the degrees of freedom. A lower value of MSE for a model suggests a more adequate fit.
Additionally, Model 3 assumes a continuous decline after obtaining the maximum. Suggesting Model 3 will produce a poor patient satisfaction estimate post maximum.

5 | CONCLUSION AND FUTURE WORK

There are 2 main stages that compose robust-ITS. The first is modeling the mean and the second is modeling the stochastic component. In both stages robust-ITS tests for a change in the outcome because of the intervention. To the best of our knowledge, in the ITS literature comparing and testing for a difference in the stochastic component—a change in autocorrelation and/or variance for the AR(1) case—has not been considered.

In the first stage, a set of plausible change points must be established based on the scientific question at hand. Then, based on the set of possible change points, robust-ITS estimates the mean parameters via OLS and chooses the change point whose parameter estimates maximize the likelihood. In the second stage, the residuals obtained by modeling the mean in the first stage are used to examine and determine the structure of the stochastic process. If the residuals act as white noise, (1) there is no correlation present, (2) the variances before and after the estimated change point are compared by an $F$ test, and (3) the outcome of interest is adequately modeled by the mean from stage 1. Otherwise, an ARIMA process is fit on the residuals pre– and post–change point, separately. From the ARIMA process, estimates of the correlation and variance are obtained via conditional likelihood methods. The correlation estimates are compared to determine if the stochastic process differs as a result of the change point, but the variances are not compared.

The patient satisfaction and CNL integrated care delivery analysis elucidated that the assumed change point is not always assumed adequately. Following the traditional segmented regression analysis, we would have set the change point at the same value for all units and assumed that it was equal to the formal intervention time. We estimated the change point corresponding to CNL integrated care delivery prior to the formal intervention time for 3 units and the estimated change point varied based on unit.

**FIGURE 4** The robust-ITS toolbox in R Shiny (by Cruz, Hu, Ngo, Bender, and Ombao), an interactive toolbox in which the user (1) may upload their own data in a .csv file; (2) provides basic information of the data—the toolbox requires the user to input the “theoretical executive time point (TET)” (formal time of intervention), “candidate before TET” (smallest value of the set of possible change points), “candidate after TET” (largest value of the set of possible change points), “starting month” (the month at which data collection began), and “starting year” (year at which data collection began); (3) views the output plots (after pressing the button labeled “Analyze Data”) of the fitted data, the log-likelihood at possible change points, residuals, and acf plots to determine the lag of the stochastic process; and (4) views the estimates, along with their $P$ values and standard errors, for both the mean and stochastic processes [Colour figure can be viewed at wileyonlinelibrary.com]
In 2 of the 4 units, the CNL integrated care delivery introduction was associated to a change in the mean patient satisfaction. Even though the change in mean patient satisfaction was not necessarily positive, it depicted a mean that continued towards 100%. The lack of affirmation for the CNL integrated care delivery may stem from the outcome definition as a percentage and an average. The percentage quality of patient satisfaction limits the values the outcome may take on and thus creates an asymptote effect for units that were already doing well. The averaging across 7 patient satisfaction indicators may cancel out improvements in some indicators with regressions in others.

The estimates of the autocorrelation coefficients pre- and post-change point, although not statistically significant, differed by approximately 0.25 for Units 1, 2, and 4. Since the autocorrelation was not statistically significant, the variances pre- and post-change point were compared. For Unit 2, the variance post-change point was significantly smaller than the variance pre-change point. This is a positive result of the CNL integrated care delivery, since there will may better quality control of patient satisfaction fluctuations due to the CNL intervention.

Comparing robust-ITS with traditional ITS modeling illustrates how allowing for a variable change point results in a better fit with regard to MSE. The ability to easily assess the effect of the intervention on the correlation structure, and to conduct variance comparisons when correlation is not present, allows for clearer inference on the possible effect of an intervention.

Our group has developed the robust-ITS toolbox in R Shiny (see Figure 4) that executes the methodology described here. The toolbox and its manual (in a PDF document) are located respectively at Robust-ITS and Manual. It is crucial to note that the methodology implemented in the toolbox is the methodology proposed here. Hence, any use of the toolbox should result in the citation of this paper. The robust-ITS toolbox is interactive and provides the user with graphical displays, estimates, and inference on testing for differences between the preintervention and postintervention means, correlation, and variance.

The current status of the model is only for continuous-valued outcomes. We are now in the process of expanding this to include counts and rates data (eg, infection rates and counts of accidental falls). Currently, our focus is on developing a segmented regression analog of generalized linear models for count and binary outcomes. Our aim is to expand the current toolbox or produce a new toolbox that will appropriately model count and binary outcomes in a user-friendly manner.

ACKNOWLEDGEMENTS

This study was funded in part by the Commission on Nurse Certification, and based upon work supported by the Eugene Cota-Robles Fellowship at the University of California, Irvine, the NSF Graduate Research Fellowship under Grant No. DGE-1321846, and by the NSF MMS1461534 and NSF DMS1509023 grants. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

ORCID

Miriam Bender http://orcid.org/0000-0003-2457-1652

REFERENCES

1. Jandoc R, Burden AM, Mamdani M, Lvesque LE, Cadarette SM. Leadership Commitments to Improve Value in Health Care: Finding Common Ground: Workshop Summary. Washington DC: National Academies Press; 2009.
2. Hawe P. Lessons from complex interventions to improve health. Annu Rev Public Health. 2015;36:307-323.
3. Datta J, Petticrew M. Challenges to evaluating complex interventions: a content analysis of published papers. BMC Public Health. 2013;13(1):1.
4. Clark AM. What are the components of complex interventions in healthcare? theorizing approaches to parts, powers and the whole intervention. Soc Sci Med. 2013;93:185-193.
5. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton: Mifflin and Company; 2002.
6. Taljaard M, McKenzie JE, Ramsay CR, Grimshaw JM. The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. Implement Sci. 2014;9(1):1.
7. Penfold RB, Zhang F. Use of interrupted time series analysis in evaluating health care quality improvements. Acad Pediatr. 2013;13(6):S38-S44.
8. Bender M, Murphy E, Thomas T, Kaminski J, Smith B. Clinical nurse leader integration into care delivery microsystems: quality and safety outcomes at the unit and organization level. In: Poster presented at Academy Health Annual Research Meeting; 2015; Minneapolis, MN.
9. Bender M, Connelly CD, Glaser D, Brown C. Clinical nurse leader impact on microsystem care quality. Nurs Res. 2012;61(5):326-332.
10. Kavanagh KT, Cimiotti JP, Abusaleh S, Coty M. Moving healthcare quality forward with nursing-sensitive value-based purchasing. J Nurs Scholarsh. 2012;44(4):385-395.
Appendix A: Appendices

A.1 A note on the time series length for robust-ITS

The number of time points required pre- and post-change point (or preintervention and postintervention) depends on many factors. Previously in the ITS literature, it has been suggested that a minimum of three time points is needed in both phases to adequately estimate the outcome means.\(^{16,17}\)

Estimating the intercept and slope of a straight line via regression requires at least three data points, to have sufficient degrees of freedom to estimate the variance. The constraint of 3 data points therefore makes the assumption that only an
intercept and slope need to be estimated; not true here, since we also wish to model the correlation structure. Another data point is needed for each additional parameter estimated. Ignoring the change point, since we are estimating the intercept, slope, autocorrelation, and variance of each segment, a minimum of 5 time points in each phase is needed to be able to merely estimate the parameters.

Because we estimate the change point it is necessary to obtain 5 time points in each phase separate from the set of possible change points to adequately estimate the regression lines. That is, a total of 10 (5 for the pre–change point and 5 for the post–change point phases) plus the length of the set of possible change points is required.

The number of parameters that must be estimated plus one is a severe lower bound for the number of time points needed to make inference and should not be used as a rule of thumb.

The discussion of setting a practical lower bound for the time points needed in each phase stems from the desire to have enough power to make proper inference. However, power not only depends on the length of the time series in each phase, but additionally on the distribution of the data points pre– and post–change point, variability, effect strength, and confounding. Considering solely the length of the pre– and post–change point phases is not sufficient when calculating power, many other factors must be taken into account. Little development of power calculations in the ITS setting exist.

A.2 Mean parameter estimates

The OLS estimates for the mean parameters in $\theta$ of Section 3.2.1 are as follows:

$$\hat{\beta}_0 = \frac{1}{T} \sum_{t=1}^{T} Y_t, \quad (A1)$$

$$\hat{\beta}_1 = \frac{1}{T} \sum_{t=1}^{T} (t - \frac{T}{2}) Y_t, \quad (A2)$$

$$\hat{\delta} = \frac{1}{T} \sum_{t=\hat{I}_{1}:T}^{T} (\hat{Y}_t - \bar{Y}_{1:(\hat{I}_{1}-1)}) - \hat{\beta}_1 \frac{T}{2}, \quad (A3)$$

$$\hat{\Delta} = \frac{1}{T} \sum_{t=\hat{I}_{1}:T}^{T} (\hat{Y}_t - \bar{Y}_{1:(\hat{I}_{1}-1)}) - \hat{\beta}_1 \frac{T}{2}, \quad (A4)$$

where $\bar{Y}_{a:b} = \frac{1}{b-a+1} \sum_{i=a}^{b} Y_i$ and $\bar{I}_{1:T} = \frac{1}{T} \sum_{t=1}^{T} I_t$. The estimates of $\beta_0$ and $\beta_1$ are the same as the OLS estimates obtained by fitting a linear model to the pre–change point phase alone. The estimates of $\delta$ and $\Delta$ may be obtained from fitting a linear model to the post–change point phase and subtracting the OLS estimates of the first phase from the OLS estimates (of the intercept and slope) of the second phase. The estimates for $\sigma^2_1$ and $\sigma^2_2$ depend on the stochastic process and are given in the following section for an AR(1) process.

A.3 AR(1) parameter estimates

In the AR(1) setting with a change point at $\hat{\tau}$ the autocorrelation and variance can be estimated by maximizing the conditional likelihood. The estimates are functions of the residuals $R_t$ and the residuals of the residuals $W_t$

$$\hat{\phi} = \left\{ \begin{array}{l}
\hat{\phi}_1 = \frac{\sum_{t=\hat{I}_{1}:T}^{T} (R_t - \hat{R}_{2:(\hat{I}_{1}-1)})(R_{t-1} - \hat{R}_{1:(\hat{I}_{1}-2)})}{\sum_{t=\hat{I}_{1}:T}^{T} (R_t - \hat{R}_{2:(\hat{I}_{1}-1)})^2}, \\
\hat{\phi}_2 = \frac{\sum_{t=\hat{I}_{1}:T}^{T} (R_t - \hat{R}_{2:(\hat{I}_{1}-1)})(R_{t-1} - \hat{R}_{1:(\hat{I}_{1}-2)})}{\sum_{t=\hat{I}_{1}:T}^{T} (R_t - \hat{R}_{2:(\hat{I}_{1}-1)})^2}
\end{array} \right. \quad (A5)$$
\[ \sigma^2 = \begin{cases} \sigma_1^2 = \frac{1}{T-1} \sum_{t=2}^{T} [(W_t - \bar{W}_{2:(t-1)}) - \hat{\phi}_1(W_{t-1} - \bar{W}_{1:(t-2)})]^2 \\ \sigma_2^2 = \frac{1}{T-1} \sum_{t=T+2}^{T} [(W_t - \bar{W}_{(t+1):T}) - \hat{\phi}_2(W_{t-1} - \bar{W}_{(t+1):T})]^2, \end{cases} \] (A6)

where \( R_{a:b} \) and \( W_{a:b} \) are the means of the residuals and of the residuals of the residuals, respectively, for time points \( a \) through \( b \) and

\[ W_t = \begin{cases} R_t - \hat{\phi}_1 R_{t-1}, & 1 < t < \hat{t} \\ R_t - \hat{\phi}_2 R_{t-1}, & \hat{t} < t \leq T. \end{cases} \]

### A.4 Nested \( F \) test for the equality of autocorrelation for an AR(1)

To determine whether the stochastic process differs as a result of the change point, we test the hypothesis that \( \nu \equiv \phi_2 - \phi_1 \) equals zero. If \( \nu = 0 \), there is one overarching AR(1) process for all time points and Equation 2 reduces to

\[ R_t = \phi_1 R_{t-1} + \epsilon_t, \quad 1 < t \leq T, \quad \text{(A7)} \]

otherwise, Equation 2 holds. We are comparing nested models where Equation 2 is the full model and Equation A7 is the reduced model, so an \( F \) test is appropriate. The degrees of freedom corresponding to the reduced model is \((T - 1) - 1\), to account for the lag of the AR process and the parameter in the model, \( \phi_1 \). Similarly, since the full model corresponds to 2 separately fit AR(1) processes and 2 parameters, the degrees of freedom is \((T - 2) - 2\). Denote the residual sum of squares for the reduced and full models, respectively, as

\[ \text{RSS}_R = \sum_{t=2}^{T} (R_t - \phi_1 R_{t-1})^2 \]

\[ \text{RSS}_F = \sum_{t=2}^{T-1} (R_t - \phi_1 R_{t-1})^2 + \sum_{t=T+1}^{T} (R_t - \phi_2 R_{t-1})^2. \]

Then the \( F \) statistics is

\[ F = \frac{(\text{RSS}_R - \text{RSS}_F)/(T-1) - 1}{(\text{RSS}_F)/(T - 2)} = \frac{\text{RSS}_R - \text{RSS}_F}{2} / \frac{\text{RSS}_F}{(T - 2)}, \]

and under the null hypothesis \((\nu = 0)\) is distributed \( F_{2,(T-2)} \).

### A.5 Segmented regression models

In the health care intervention literature, there are 2 main types of segmented regression models used to model the trends. The first is parametrized in the same manner as (1), with \( t \) set to the time of intervention—an assumed instantaneous effect—that is, the mean is parametrized as follows:

\[ \mu^1_t = \begin{cases} \beta_0 + \beta_1 t, & t < t^* \\ (\beta_0 + \delta) + (\beta_1 + \Delta) t, & t \geq t^* \end{cases}, \] (A8)

where \( t^* \) denotes the intervention time. For the data described in Section 1, \( t^* = 31 \). The second segmented regression model is

\[ \mu^2_t = \begin{cases} \beta_0 + \beta_1 t, & t < t^* \\ (\beta_0 + \psi) + \beta_1 t + \Psi(t - t^* + 1), & t \geq t^* \end{cases}, \] (A9)

in which the time after intervention implementation is multiplying \( \Phi \); as opposed to simply time, as in Equation A8.

Note, the trends prior to intervention introduction are the exact same for both Equations A8 and A9. Post the intervention time, the intercept increase is denoted by \( \delta \) in Equation A8 and by \( \psi - (t^* - 1) \Psi \) in Equation (A9), implying \( \delta = \psi - (t^* - 1) \Psi \). The change in slopes is denoted by \( \Delta \) and \( \Psi \) in Equations A8 and A9, respectively, and so \( \Delta = \Psi \). Although the parametrization is different, the estimates of the intercepts, slopes, and any function of the slopes and intercepts (as is the level change) are the same. Thus, the models are equivalent.
A.6 Supporting figures

**FIGURE A1** The studentized residuals of the robust-ITS estimated patient satisfaction means for each unit. The studentized residuals do not exhibit any clear patterns and seem to be closely centered around zero, indicating appropriate fits. The rule of thumb, 2, is provided in each plot [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE A2** Autocorrelation function (ACF) plots of the robust-ITS estimated patient satisfaction means for each unit. The ACF plots suggest that the residuals behave as white noise, since the autocorrelation at lags greater than zero are small and seem to get closer to zero as the lag increases [Colour figure can be viewed at wileyonlinelibrary.com]