UNIVERSITY OF CALIFORNIA,
IRVINE

Interrupted Time Series Models for Assessing Complex Health Care Interventions

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Maricela Francis Cruz

Dissertation Committee:
Professor Daniel L. Gillen, Chair
Professor Hernando Ombao
Professor Michele Guindani
Associate Professor Zhaoxia Yu
Assistant Professor Miriam Bender

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACKNOWLEDGMENTS

I thank my committee chair and academic advisor, Daniel L. Gillen, for his unrelenting encouragement to my personal development as a statistician and researcher. In addition I thank my co-advisor, Hernando Ombao, for his continued dedication to my academic success. Without either, this dissertation would not have been possible.

I thank my committee members, Michele Guindani, Zhaoxia Yu, and Miriam Bender for their valuable expertise. In addition, I thank Miriam Bender for revealing a methodological gap in the health care literature and for providing me with astonishingly clean data.

Chapter 3 and Chapter 4 of this dissertation are reprints of material as it appears in Statistics in Medicine. The co-authors of these publications directed and supervised research which forms the basis for the dissertation.

This dissertation was supported in part by the Eugene Cota-Robles Fellowship at the University of California, Irvine, the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1321846, the National Science Foundation MMS 1461534 and DMS 1509023 grants, the National Institute on Aging of the National Institutes of Health under award numbers R01AG053555 and P50AG16573, and the Commission on Nurse Certification. 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. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CURRICULUM VITAE

Maricela Francis Cruz

EDUCATION

Doctor of Philosophy in Statistics — **July 2019**
University of California, Irvine

Master of Science in Statistics — **March 2017**
University of California, Irvine

Bachelor of Arts in Mathematics — **May 2014**
Pomona College

RESEARCH EXPERIENCE

Graduate Research Assistant — **September 2014 – July 2019**
University of California, Irvine

Summer Associate — **June – September 2018**
RAND Corporation

TEACHING EXPERIENCE

Teaching Assistant — **Summer 2015, 2016, 2017**
University of California, Irvine

REFEREED JOURNAL PUBLICATIONS

Group cohesion and climate in cognitive behavioral therapy for first-time DUI offenders **2019**
Alcoholism Treatment Quarterly

System and unit-level care quality outcomes after reorganizing frontline care delivery to integrate Clinical Nurse Leaders: a quasi-experimental time series study **2019**
Journal of Nursing Administration

Assessing health care interventions via an interrupted time series model: study power and design considerations **2019**
Statistics in Medicine

A robust interrupted time series model for analyzing complex health care intervention data **2017**
Statistics in Medicine

ABSTRACT OF THE DISSERTATION

Interrupted Time Series Models for Assessing Complex Health Care Interventions

By

Maricela Francis Cruz

Doctor of Philosophy in Statistics

University of California, Irvine, 2019

Professor Daniel L. Gillen, Chair

Assessing the impact of complex interventions on measurable health outcomes is a growing concern in health care and health policy. Interrupted time series (ITS) designs borrow from traditional case-crossover designs and function as quasi-experimental methodology that enables researchers to retrospectively analyze the impact of an intervention. Statistical models used to analyze ITS data *a priori* restrict the interruption's effect to a predetermined time point or censor data for which the intervention effects may not be fully realized, and neglect changes in the temporal dependence and variability. In addition, current methods limit the analysis to one hospital unit or entity and are not well specified for discrete outcomes (e.g., patient falls). This dissertation develops novel ITS methods based on segmented regression that address the aforementioned limitations.

We propose the 'Robust-ITS' model, a single-unit model able to estimate (rather than merely assume) the lagged effect of an intervention on a health outcome. Robust-ITS accounts for plausible differences in the mean, temporal dependence and variability of an outcome pre- and post-intervention. Next, we develop the 'Robust Multiple ITS' model as an extension of Robust-ITS for multi-unit data. Alongside Robust Multiple ITS, we propose the 'supremum Wald test', able to formally test for the existence of a change point across unit specific mean functions. Lastly, we present the 'Generalized Robust ITS' model, appropriate for

outcomes whose underlying distribution belongs to the family of exponential distributions. Generalized Robust ITS expands the available methodology to adequately model multi-unit binary, count and rate ITS. The methods proposed allow researchers to test for the existence of and estimate the change point, borrow information across units in multi-unit settings, and test for differences in the mean function and correlation structure pre- and post-intervention. Throughout, the methodology is illustrated by analyzing patient centered data from a hospital that implemented and evaluated a new care delivery model in multiple units.

Chapter 1

Introduction

Evaluating the effectiveness of complex interventions on health outcomes is a growing concern in health care and health policy. The Centers for Medicare & Medicaid Services (CMS) in the United States financially incentivize health care quality reform via a value-based purchasing program for health systems care services reimbursement (Kavanagh et al., 2012). The purchasing program proliferates health care interventions aimed at bettering quality of care measures, including mortality and complications, patient safety, and patient experience (Centers for Medicare and Medicaid Services, 2018). Health care professional and policy organizations recognize the need to transform health care to better provide patient centered and team oriented care (Knebel et al., 2003; Fitzpatrick, 2003; Kohn et al., 2000). As such, these organizations often spearhead health care interventions aimed at improving care quality (American Hospital Association Commission on Workforce for Hospitals Health Systems and others, 2002; Joynt and Kimball, 2008).

Assessing the impact of health care interventions on quality of care measures is difficult with regards to research design and statistical analysis (Datta and Petticrew, 2013). Patients, providers, resources and contexts of care interact in dynamic ways to produce various

measurable health outcomes that often do not align with expectations (Hawe, 2015). This complexity and interdependency makes it challenging to assess the true impact of interventions designed to improve patient centered outcomes. Furthermore, randomized controlled trials, the "gold standard" for evidence generation of health care interventions, are often difficult to implement and not feasible in health systems with regard to health care reform (West et al., 2008). According to the 2018 Annual Review of Public Health, interrupted time series (ITS) designs may be the only feasible recourse for studying the impacts of large-scale public health policies (Handley et al., 2018).

ITS designs borrow from traditional case-crossover designs and function as quasi-experimental methodology able to retrospectively analyze the impact of an intervention and account for data dependencies (Bernal et al., 2017). Health care intervention data often present as ITS: sequences of measurements for an outcome (e.g., patient satisfaction scores) collected at multiple time points before and after an intervention. Despite this, there continue to be numerous limitations to the statistical methodology available to analyze ITS data. Namely, current methods (a) restrict the interruptions effect to a predetermined time point or remove data for which the intervention effects may not be fully realized; (b) neglect plausible differences in temporal dependence and variability; (c) restrict the analysis to a single unit; and (d) are not well specified for discrete outcomes. These are serious limitations because the methods ignore interpretable changes in the higher order moments of the response, do not take advantage of all available data that may provide information on the intervention's effect, and overlook an entire class of outcome types. In this dissertation, we propose novel ITS methodology that address these limitations.

The ITS methods developed in the subsequent chapters are motivated by our interest in estimating the lagged effect of a care delivery intervention on two patient-centered outcomes. The outcomes, recorded monthly for a five year period, are average patient satisfaction, discussed in Chapters 3-4, and patient falls, considered in Chapter 5. Time series plots of

average patient satisfaction for the Stroke and Surgical units and the log of patient falls for the Cardiac and Acute Care units are provided in Figure 1.1. These two measures are currently being used to calculate health systems reimbursement for care services via CMS's value-based purchasing program (Kavanagh et al., 2012).

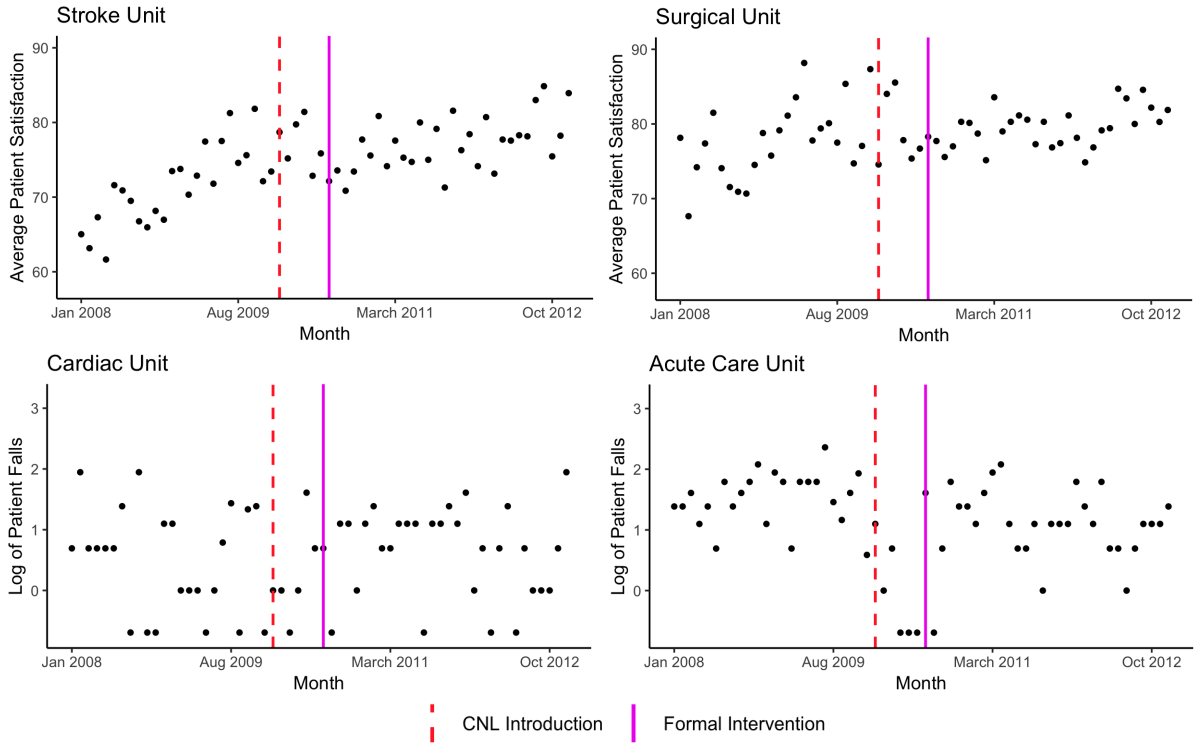

Figure 1.1: Plots average patient satisfaction for the Stroke and Surgical units, and the log of patient falls for the Cardiac and Acute Care units. Note, for the purposes of depicting the time series we add 0.5 to patient falls, making log of patient falls equal to -0.69 and giving rise to the negative points in the plots when patient falls is equal to zero.

The intervention was the implementation of Clinical Nurse Leader (CNL) integrated care delivery, a nursing model that embeds a master prepared nurse into the front lines of care (Bender et al., 2017). The nurses, referred to as Clinical Nurse Leaders (CNLs), have advanced competencies in clinical leadership, care environment management, and clinical outcomes management (Bender et al., 2017). The CNLs, conducting their master's level microsystem change project, were introduced into their respective hospital units six months prior to the formal intervention implementation time. This may or may not have influenced the 'change point' of the intervention effect. Importantly, because of this early introduction,

the estimated change point may have occurred up to six months prior to the formal intervention. We are therefore interested in estimating the time lag (or delay) between the onset of the intervention and the effect on the patient centered outcomes.

Five overall scientific aims give rise to the methodology developed in the remainder of the dissertation. The first is to determine if a change in the responses exists over a predetermined set of possible change points. Next, we aim to estimate the time lag (or delay) between the onset of the intervention and the intervention's effect. The third scientific objective is to account for changes in the mean function, temporal dependence and variability pre- and post-intervention. Then, we aim to borrow information across units in multi-unit settings. Lastly, we intend to use the knowledge gained in this setting and from our derived methodology to plan for future intervention assessments.

In the ensuing chapter, we provide a review of the current statistical methodology used for health care intervention evaluations, i.e., segmented regression. We go on to develop the 'Robust-ITS' model, a single unit model able to estimate the lagged effect of an intervention on a continuous outcome and account for differences in the mean function and correlation structure pre- and post-intervention. In Chapter 4, we propose the 'Robust Multiple ITS' model, an extension of Robust-ITS for multi-unit data, and the 'supremum Wald test', a formal test used to determine the necessity of a change point. Next, we present the 'Generalized Robust ITS' model, a generalization of Robust Multiple ITS for multi-unit binary, count and rate ITS. Alongside Robust Multiple ITS and Generalized Robust ITS, we present empirical simulations to assess the type one error, power for detecting specified change point alternatives, and accuracy of the change point estimation procedures. We conclude the dissertation with a discussion of our developed methodology and include future research directions.

Chapter 2

Segmented Regression

Health care intervention study designs are often natural experiments encouraged by policy mandates or health system innovation that are not scientifically controlled. The data that arise from these intervention studies do not typically stem from a randomized controlled trial (Craig et al., 2012). Simple comparisons of the mean pre- and post-intervention, say via a t-test, do not provide the level of statistical rigor needed to account for contextual factors and preexisting trends encountered in natural experiments (Ramsay et al., 2003). The most utilized statistical methodology for analyzing ITS data is segmented regression, a powerful methodology accounting for underlying trends, including outcome trajectories and correlation (Linden, 2015; Penfold and Zhang, 2013; Wagner et al., 2002). Segmented regression was first introduced in Quandt (1958), closely followed by Thistlethwaite and Campbell (1960). Since then, segmented regression has been used in many forms and disciplines, including health services research, economics and education.

Traditional segmented regression *a priori* sets the change point to a hypothesized value, typically at or around the formal intervention, to fully differentiate the pre- and post-intervention phases (Taljaard et al., 2014). The change point is defined as the time point at

which a change occurs in the time series, i.e., the first time point in the post-intervention phase (Wagner et al., 2002). We denote the *a priori* specified change point as τ. Next, the 'impact model' must be specified; see (Bernal et al., 2017) for a discussion on choosing the correct impact model. To synchronize this chapter with the motivating data example and the methodology proposed in this dissertation, we choose a flexible impact model that assumes linearity, and so, allow for a level (intercept) and trend (slope) change at the change point.

2.1 Description of Segmented Regression

Let y_t denote the outcome of interest at time point $t \in \{1, \ldots, n\}$. Then, the general segmented regression model is:

$$y_t = \mu_t + \epsilon_t, \tag{2.1}$$

where μ_t is the mean function and ϵ_t is the stochastic process (or error term) at time point t. The mean function is

$$\mu_t = \begin{cases} \beta_0 + \beta_1 t, & t < \tau - k \\ (\beta_0 + \delta) + (\beta_1 + \Delta)t, & t \geq \tau \end{cases}, \tag{2.2}$$

where k is equal to zero if data is not censored (i.e., removed from the analysis) and otherwise set to an integer corresponding to the number of censored time points. For consistency with the remainder of the dissertation, we set $k = 0$. There is another common parametrization of the mean function for segmented regression used in health care studies, see Appendix A.1.4 for its specification. There is a one-to-one mapping between the two parametrizations. The parameters of the mean function provided in equation (2.2) are: (1) β_0, the intercept prior to the change point; (2) β_1, the slope prior to the change point; (3) δ, the change in baseline intercept post-change point; and (4) Δ, the change in slope post-change point.

These parameters give rise to clinically important summary measures further discussed in Section 2.1.1. In existing works, the change point, τ, is assumed known. We will extend this assumption in Chapter 3.

The stochastic process, ϵ_t, accounts for the variability and correlation of the outcome at time point t. If the ITS data are independent, we may assume $\epsilon_t \overset{iid}{\sim} N(0,1)$ for all t. When the ITS measurements are not independent, the stochastic process may be modeled with an appropriate ARIMA process. Frequently, it is assumed that $\epsilon_t \overset{iid}{\sim} N(0,1)$ for all t (Bernal et al., 2017). Nevertheless, as we are interested in appropriately modelling correlation and variability in our data setting, we assume the stochastic component is modeled as an AR(1) process:

$$\epsilon_t = \phi\ \epsilon_{t-1} + e_t, \tag{2.3}$$

where $e_t \overset{iid}{\sim} N(0, \sigma_w^2)$, $\phi \in (-1,\ 1)$ and $t \in \{2, \ldots, n\}$. The correlation between any two adjacent time points is denoted by ϕ. The variance of the response at any measurement is then $\sigma^2 = \frac{\sigma_w^2}{1-\phi^2}$.

2.1.1 Intervention Impact Measures

The health care community has adopted two measures to assess the impact of an intervention: level change and trend change. It is necessary to report both level change and trend change when interpreting the results of an ITS study (Effective Practice and Organisation of Care, 2015). Level change is interpreted as the discontinuity between the projected mean based on the pre-change point phase and the estimated mean post-change point, i.e., the anchored intercept at the change point, depicted graphically in Figure 2.1. Trend change quantifies the impact of the intervention on the overall trajectory of the mean function, i.e., the change in slopes post-intervention. In equation (2.2): Δ is the trend change ; and $-\delta - \Delta\tau$ is the

level change.

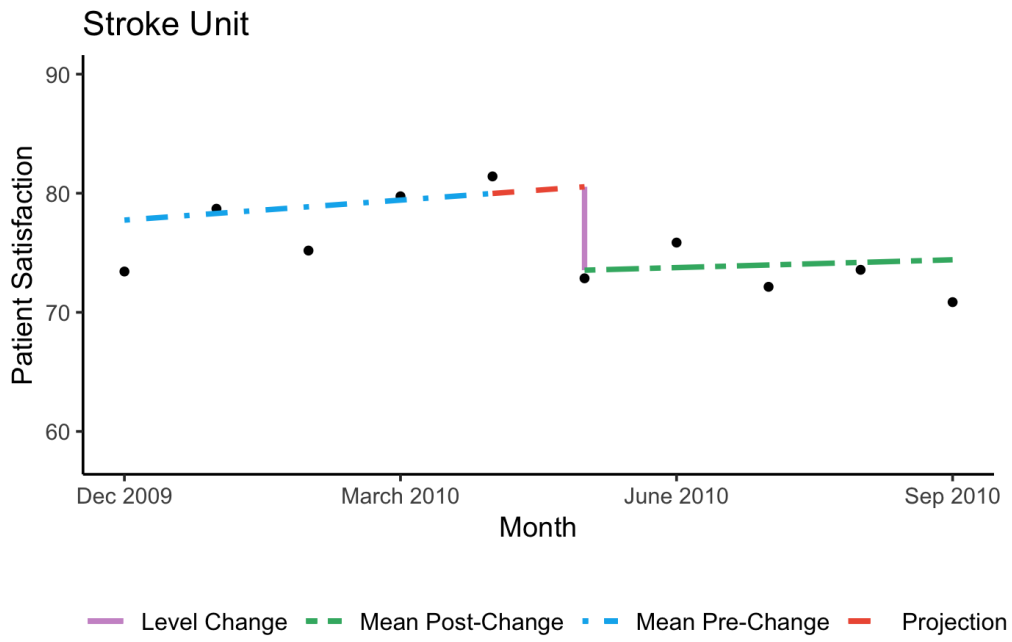

Figure 2.1: An example of an estimated mean function for segmented regression for the Stroke unit. Depicts: (1) the estimated mean function for the pre- and post-change point phases, (2) the projection of the mean function at the change point based on the pre-change point regression, and (3) the change in level as defined in this dissertation. The plot contains data from January 2010 to September 2010, instead of the entire observational period, to clearly illustrate level change.

2.2 Limitations

Segmented regression, though powerful, has its limitations. First and foremost, segmented regression assumes that there is a distinct separation between the pre- and post-intervention phases, by either assuming the change point is known or removing the set of possible change points. An instantaneous intervention effect is often assumed, i.e., the change point is often set to the intervention time, because change point estimation involves optimization over all possible configurations, which challenges computational feasibility. Specification of the change point as the time of intervention does not, however, 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. Prevalent approaches to overcoming this limitation are to remove a specific set of time points from the analysis (Penfold and Zhang, 2013; Taljaard et al., 2014). This censoring (i.e., removal of time points) not only omits data, but it also potentially biases parameter estimates, as the study team decides which time points to remove.

In addition, segmented regression assumes that there are two phases in the mean function, one pre- and another post-intervention, but one overall correlation structure. The assumption of a constant correlation structure is not necessarily representative of complex interventions where the health system seldom reacts in an isolated way to change and the intervention is expected to reduce variability in the system. Theoretically, the correlation structure should differ based on the introduction of an intervention (Cabrieto et al., 2017). In fact, with complex health care interventions, the goal is often to enhance care processes so that elements become more dependent and consistent over time. An increase in data dependency and consistency implies a difference in correlation and variability. Thus, detection of differences in correlations and variances pre- and post-intervention are critical in evaluating the effectiveness of an intervention.

Segmented regression is a single unit model, and as such, does not borrow information across units. This is a serious limitation because it does not take advantage of all available data that may provide information on the lag associated with an intervention. The current form of segmented regression is not well specified for discrete outcomes, thereby effectively overlooking an entire class of outcomes. As health care ITS are often composed of discrete measures (i.e., patient falls, unretrieved device fragment count, mortality etc.), methodology able to assess the impact of an intervention on these outcomes is also needed. The subsequent chapters of this dissertation develop methodology that address these limitations.

Chapter 3

A Robust ITS Model

In this chapter, we develop the Robust-ITS model, a novel model for ITS, that estimates the impact of an intervention on health outcomes. One advantage of the Robust-ITS model as compared to segmented regression 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. 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 phases.

We propose a method which regards the change point as variable, appropriate for situations where the data warrants such treatment. Nonetheless, if the aim is to attempt to isolate causal effects of the intervention it may be better to pre-specify 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 data motivating our 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 (Bender et al., 2015). The outcomes of interest are 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 (Bender et al., 2012). 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 (Kavanagh et al., 2012). A time series plot of patient satisfaction scores from January 2008 to December 2012 at four units (Stroke, Cardiac, Medical Surgical, and Mother Baby units) in a health care system is given in Figure 3.1.

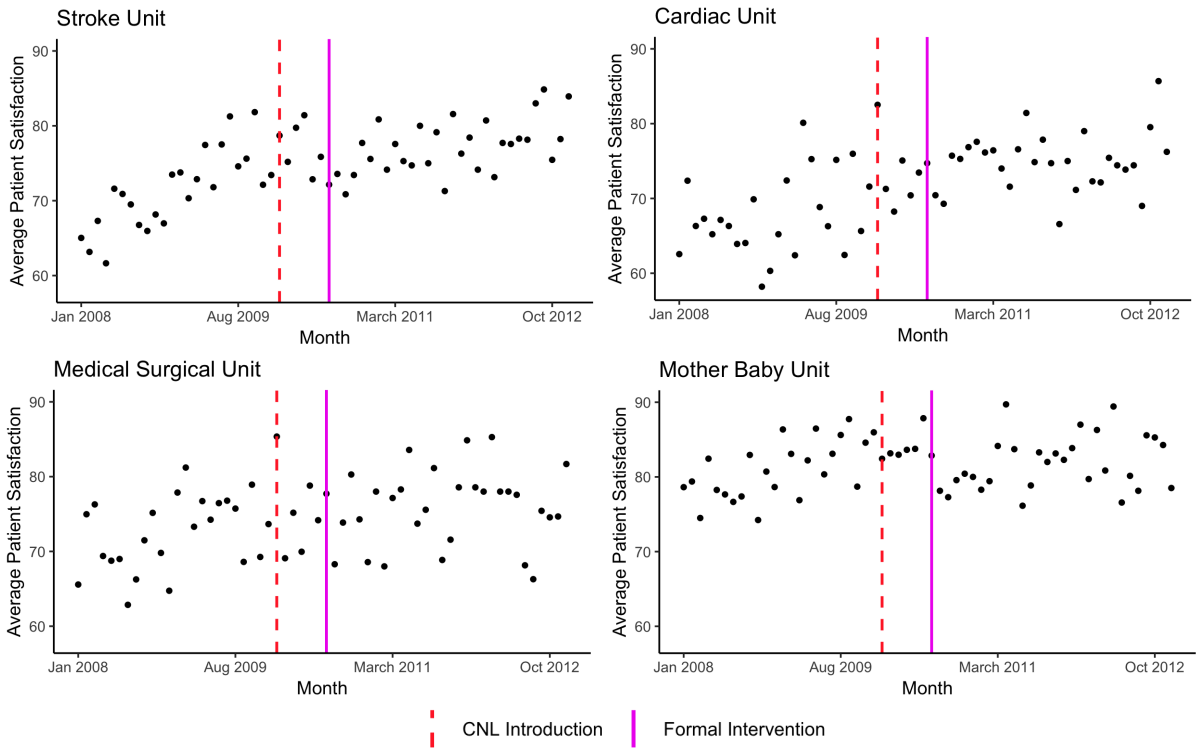

Figure 3.1: Plots the time series of observed average patient satisfaction for each unit, the nurses introduction into their respective hospital units and the formal intervention time.

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 one of four responses: *never*, *sometimes*, *usually*, or *always*.

For modeling we chose one outcome: average patient satisfaction, averaged over seven patient satisfaction indicators, for 4 hospital units of the study setting. We refer to average patient satisfaction simply as patient satisfaction throughout the remainder of the dissertation. 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 (Bender, 2014), discussed in Chapter 1.

The remainder of this chapter 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 briefly 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.

3.1 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 (Rickles, 2009). RCTs, however, 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 (Petticrew et al., 2013).

Interrupted time series (ITS) 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 (Taljaard et al., 2014; Kontopantelis et al., 2015). When RCTs are not feasible or not applicable, ITS is considered the strongest research design in the health policy evaluation literature (Penfold and Zhang, 2013), and are considered rigorous enough for inclusion into Cochrane meta analyses (Effective Practice and Organisation of Care, 2015).

We believe, along with many authors in this field, that segmented regression is most effective in analyzing ITS data. Segmented regression may be utilized via standard statistical packages — such as ITSA in Stata, ETS in SAS, segmentedR in R, etc. — however, there are limitations to these current statistical packages as there are limitations to segmented regression; see Chapter 2 for a detailed description of segmented regression limitations. Table 3.1 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.

Package	Papers	Advantages/Description	Limitations
SAS PROC AUTOREG (SAS User Guide, 2008)	Penfold and Zhang (2013), Eliopoulos et al. (2007), Parienti et al. (2011)	• 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 (SAS User Guide, 2008)	Eliopoulos et al. (2007)	• 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 process
SAS ETS (SAS User Guide, 2008)	Cable (2001), Mahamat et al. (2007), Aboagye-Sarfo et al. (2015)	• Same as the above two entries; SAS PROC ARIMA and SAS PROC AUTOREG are part of SAS ETS.	• Fixed intervention time point with immediate effect. • One overall correlation structure
Stata ITSA	Linden (2015)	• Single and multiple group comparisons. • Estimates treatment effects for multiple treatment periods. • Adjust for overall autocorrelation.	• Fixed change point. • One overarching correlation structure.
segmentedR	Muggeo (2012)	• Estimates piecewise regression models with a fixed number of discontinuities, or interruptions.	• No modeling of correlation structure. • Assumes data are independent.

Table 3.1: Denotes the limitations and advantages of ITS packages that focus on segmented regression and provides papers that either propose or utilize the packages.

3.2 The Robust Interrupted Time Series (Robust-ITS) Model

3.2.1 Preliminary Analysis

Before any formal statistical modeling, we plot the outcome against time to illuminate the type of longitudinal mean (linear, quadratic, etc.), seasonality, and the set of plausible change points. As previously noted, 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., 2013). If needed one should apply variance stabilizing transformations to the outcome variable. For the purposes of illustrating Robust-ITS, the relationship between the outcome and time is assumed linear with no seasonality.

3.2.2 Description of the Robust-ITS Model

One prominent feature of our Robust-ITS model is the clear distinction between the time of intervention and the change point. In Penfold and Zhang (2013); Garey et al. (2008); Ansari et al. (2003) 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. 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 two 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 τ 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 (i.e., either $\tau > t^*$ or $\tau < t^*$). Here we develop a data adaptive procedure for estimating τ. 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 (Davis et al., 2006; Kirch et al., 2015).

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 μ_t is the mean and ϵ_t is the stochastic process. The mean component, μ_t, characterizes the mean of the outcome for the pre-intervention and post-intervention phases. The stochastic process, ϵ_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 carry out the Robust-ITS analysis is provided in the appendices.

The Pre- and Post-intervention Mean

At the first stage of modeling the emphasis is on the mean,

$$\mu_t = \begin{cases} \beta_0^\tau + \beta_1^\tau\, t, & t < \tau \\ (\beta_0^\tau + \delta^\tau) + (\beta_1^\tau + \Delta^\tau)t, & t \geq \tau, \end{cases} \tag{3.1}$$

where the parameters are estimated using ordinary least squares. The parameters in μ_t are: (1.) β_0^τ, the intercept of the mean prior to the change point; (2.) β_1^τ, the slope of the outcome prior to the change point; (3.) $\beta_0^\tau + \delta^\tau$, the intercept of the post-intervention phase; and (4.) $\beta_1^\tau + \Delta^\tau$, the slope of the post-intervention phase.

Remark. The difference between the pre-change point and post-change point intercept is δ^τ. The difference between the pre- and post-change point slopes, i.e., the trend change, is Δ^τ and the level change is $-\delta^\tau - \Delta^\tau\tau$.

Recall, level change and trend change are the metrics used in the health policy evaluation literature to measure the effect of an intervention. Formally, the level change is defined as the difference at the change point time τ between the extrapolated pre-intervention mean function and the observed intervention mean function, as is depicted in Figure 2.1.

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

Denote the length of the time series as n and let $\theta = [\beta_0^\tau, \beta_1^\tau, \delta^\tau, \Delta^\tau, \sigma_1^2, \sigma_2^2]'$, with σ_1^2 and σ_2^2 defined as the variances prior to and post change point respectively. As described in section 3.1, 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. The set of possible change points, Q, contains t^*, the time point at which the intervention is formally introduced, and time points that allow for a lag in intervention effects (via choice of k) and anticipatory effects (via choice of m). Intervention experts chose m and k based on the intricacies of the study design and data.

For each candidate change point $q \in Q$ we derive the likelihood function:

$$L(\theta|q) = \Big(\frac{1}{\sqrt{2\pi\sigma_1^2}}\Big)^{q-1} \exp\Big(-\frac{1}{2\sigma_1^2}\sum_{t=1}^{q-1}\big[y_t - (\beta_0^\tau + \beta_1^\tau t)\big]^2\Big) \times$$
$$\Big(\frac{1}{\sqrt{2\pi\sigma_2^2}}\Big)^{n-(q-1)} \exp\Big(-\frac{1}{2\sigma_2^2}\sum_{t=q}^{n}\big[y_t - (\{\beta_0^\tau + \delta^\tau\} + \{\beta_1^\tau + \Delta^\tau\}t)\big]^2\Big).$$

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

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 θ are provided in the appendices. The estimates for σ_1^2 and σ_2^2 depend on the stochastic process, and are given for an AR(1) process also in the appendices.

The presence of τ 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.

Stochastic Properties Pre- and Post- Change Point

The stochastic component, ϵ_t, captures the correlation structure of the outcome variable across time, and may change as a result of the intervention. Here, we develop a formal test for the difference in the correlation structure for pre- and post-intervention phases.

We use the ARIMA process to model the stochastic component, $\epsilon_t = y_t - \mu_t$. Since the mean function μ_t is not known (we only have its estimate, $\widehat{\mu_t}$), the stochastic component is not directly observed. In place of ϵ_t, we use the residuals, $R_t = y_t - \widehat{\mu}_t$, where $\widehat{\mu}_t$ is the estimate of μ_t obtained as described in stage one. In order 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 pre- and post-intervention phases (Shumway and Stoffer, 2017).

Autoregressive conditional heteroscedasticity models may be used when the data is non-stationary; they can model the stochastic component in each phase and for the entire observational period when the variance and/or data dependency is non-constant. 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, 2017), (Granger and Newbold, 2014) and (Bollerslev, 1988) for more details on autoregressive conditional heteroscedasticity models.

Due to the impact of the intervention, the stochastic process ϵ_t pre-intervention might differ from the process post-intervention. That is, ϵ_t for $t \in \{1, \ldots, \widehat{\tau} - 1\}$ may be a different stochastic process than ϵ_t for $t \in \{\widehat{\tau}, \ldots, n\}$. 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; τ dictates when the intervention affects the outcome variable.

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

$$R_t = \begin{cases} \phi_1(\tau)\, R_{t-1} + e_t^1, & 1 < t < \widehat{\tau} \\ \phi_2(\tau)\, R_{t-1} + e_t^2 & \widehat{\tau} < t \leq n. \end{cases} \tag{3.2}$$

To ensure granger causality, both $\phi_1(\tau)$ and $\phi_2(\tau)$ lie in the interval $(-1, 1)$. Note, $\phi_1(\tau)$, the auto-regressive coefficient prior to the change point, is directly associated with the correlation between two time points; $\phi_1(\tau)$ is the correlation between time point t and $t+1$ where t and $t+1$ belong to the pre-change point phase (t, and $t+1 \in \{1, \ldots, \widehat{\tau} - 1\}$), and $\phi_1^{|h|}(\tau)$ is the correlation between two time points h time periods away (say t and $t+h$ both in the pre-change point phase, $\{1, \ldots, \widehat{\tau} - 1\}$). The auto-regressive coefficient post change point, $\phi_2(\tau)$ has a similar interpretation. The error terms of model (3.2) are white noise, $e_t^j \overset{iid}{\sim} N(0, \sigma_j^2)$ for $j \in \{1, 2\}$.

The variance and auto-regressive coefficients in the AR(1) setting 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, and are provided in the appendices.

To determine whether the stochastic process differs as a result of the change point, we test the hypothesis that $\nu(\tau) \equiv \phi_2(\tau) - \phi_1(\tau)$ equals zero. This can be tested by either estimating ν directly or by conducting an F-test for nested models. The F-test for nested models for this AR(1) scenario is described in the appendices.

Remark. Once the ϵ_t are appropriately modeled, the OLS estimates in equations (A.1) -(A.4) will need to be re-estimated 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 re-estimated without the first time point; that is, for $t \in \{2, \ldots, n\}$. If different AR(1) processes are fit pre- and post-intervention, then the mean prior to the intervention should be re-estimated using t in$\{2, \ldots, \tau - 1\}$, and the mean post intervention re-estimated using t in$\{\tau + 1, \ldots, n\}$. The summation limits in equations (A.1) -(A.4) would therefore change.

Pre- and Post-Intervention 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, due to 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 τ 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 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 $F_{22,32}$.

3.3 Robust-ITS Analysis of the Intervention Effect on Patient Satisfaction

Patient satisfaction is modeled in four hospital units: Stroke, Cardiac, Medical Surgical, and Mother Baby units. 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 four time series were modeled as in equation (3.1); the resulting parameter estimates are given in Table 3.2. The relationship between the formal intervention implementation time and the change point for the four units is illuminated in Table 3.2 and Table 3.3, which show the effect of the intervention is not necessarily instantaneous. In fact, Table 3.2 and Table 3.3 suggest the intervention had an anticipatory effect in three of the four 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 the Stroke, Cardiac and Medical Surgical units, the estimated change points occur respectively in May 2010, January 2010, and February 2010, suggesting 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 3.2 depicts the differences in estimated means prior to and post change point, with the most informative rows of Table 3.2 corresponding to the two measures of interest: change in level and change in slopes. The level change is positive and statistically significant (at the $\alpha = 0.01$ level) for the Stroke and Mother Baby units, 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 significant drop of patient satisfaction

Patient Satisfaction				
Parameters	Stroke	Cardiac	Medical Surgical	Mother Baby
Intercept Pre Change Point	64.32** (61.76, 66.88)	64.67** (60.07, 69.27)	68.31** (64.14, 72.49)	77.21** (74.99, 79.54)
Intercept Post Change Point	67.21** (61.86, 72.57)	71.79** (66.10, 77.49)	71.51** (64.01, 79.01)	77.42** (70.24, 84.60)
Change in Intercepts, $\widehat{\delta^{\widehat{\tau}}}$	2.89 (-2.91, 8.70)	7.12 (-0.03, 14.28)	3.20 (-5.22, 11.61)	0.15 (-7.20, 7.51)
Change in level, $-\widehat{\delta^{\widehat{\tau}}} - \widehat{\Delta}^{\widehat{\tau}}\widehat{\tau}$	7.00** (3.75, 10.25)	-2.77 (-7.92, 2.38)	3.50 (-1.72, 8.72)	5.40** (2.01, 8.78)
Slope Pre Change Point	0.56** (0.41, 0.71)	0.24 (-0.08, 0.56)	0.35* (0.07, 0.63)	0.28** (0.15, 0.40)
Slope Post Change Point	0.22** (0.10, 0.34)	0.07 (-0.06, 0.20)	0.09 (-0.07, 0.26)	0.10 (-0.06, 0.25)
Change in Slope, $\widehat{\Delta}^{\widehat{\tau}}$	-0.34** (-0.53, -0.15)	-0.17 (-0.51, 0.16)	-0.26 (-0.58, 0.06)	-0.18 (-0.38, 0.02)
Delay in Effect of Intervention, $\widehat{\tau} - t^*$	-3	-6	-5	0

Table 3.2: Provides 95% confidence intervals and estimates of the mean parameters for average patient satisfaction of the Stroke, Cardiac, Medical Surgical and Mother Baby units. Since τ is discrete, only an estimate is given, no confidence interval. The asterisk, *, denotes statistical significance at the $\alpha = .05$ level.

Patient Satisfaction				
	Stroke	Cardiac	Medical Surgical	Mother Baby
Time of Intervention Implementation	Month 31, July 2010	Month 31, July 2010	Month 31, July 2010	Month 31, July 2010
Estimated Change Point, $\widehat{\tau}$	Month 29, May 2010	Month 25, January 2010	Month 26, February 2010	Month 31, July 2010

Table 3.3: Gives the formal time of intervention implementation and the estimated time at which the effect of the intervention initiates, i.e., the estimated change point.

in the Stroke and Mother Baby units. The estimated trend change or change in slopes is negative for each unit, although statistically significant (at the $\alpha = 0.01$ level) for the Stroke unit only.

The slope decreases after the estimated change point in the Stroke unit, implying a more flattened out mean post-change point. Therefore, the CNL implementation may be associated with a flatter mean across time in the Stroke unit; i.e., for every one 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 the Cardiac, Medical Surgical and Mother Baby, the estimated slope does not statistically change after the estimated change point; for the Cardiac and Medical Surgical units 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 the Cardiac and Medical Surgical units, 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 the Stroke and Mother Baby units.

The pre- and post-change point mean functions of the four units are plotted in Figure 3.2. Figure 3.2 depicts that the change point occurs prior to the formal intervention time for the Stroke, Cardiac and Medical Surgical units, but is equivalent to the formal intervention time for the Mother Baby unit. The estimated mean post-change point seems to flatten out in all units, and the change in level appears sizable for the Stroke and Cardiac units. Figure 3.2 illustrates results in concurrence with those of Table 3.2.

Figure A.1, in the appendix, provides the studentized residuals after modeling the mean. The residuals seem well behaved and mostly contained between the rule of thumb ± 2 and completely contained between ± 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 A.2 provides the autocorrelation function (ACF) of the residuals. The ACF plots interpreted as in (Shumway and Stoffer, 2017) act as white noise, implying that the data do not exhibit autocorrelation.

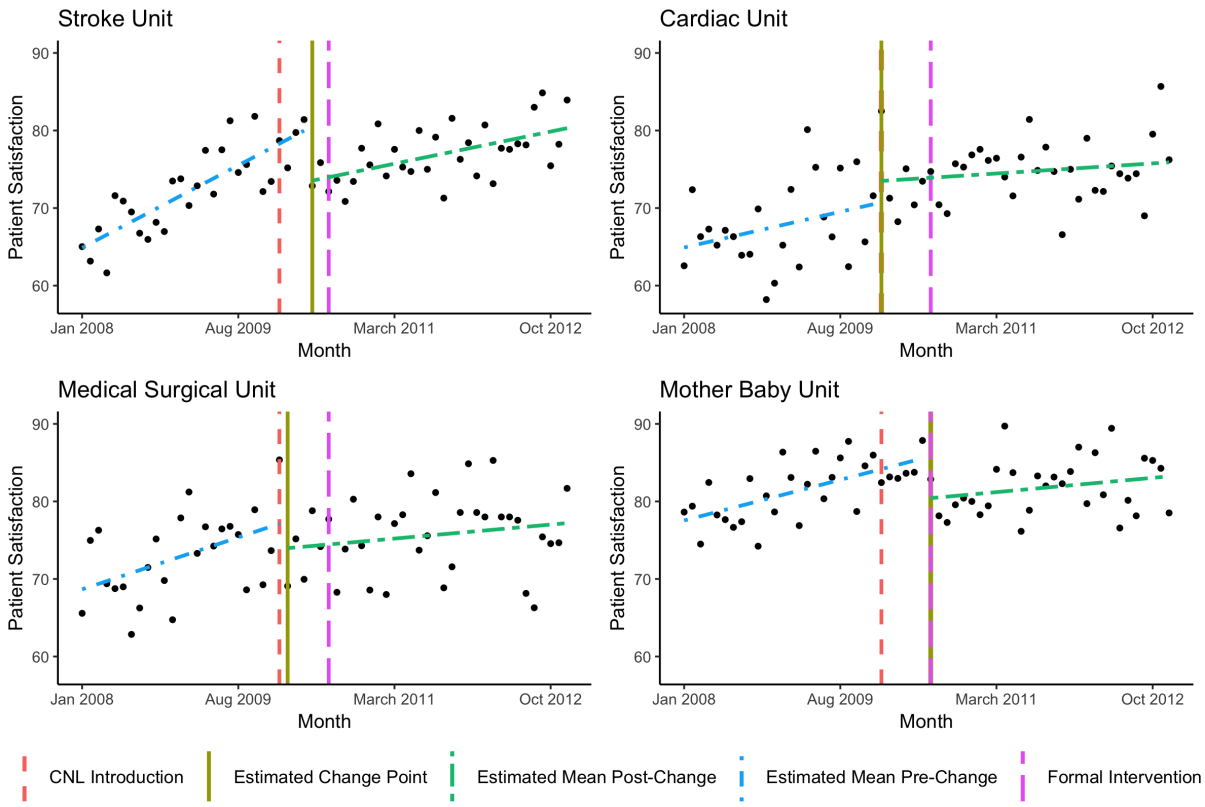

Figure 3.2: Plots the time series of observed average patient satisfaction for each unit, the estimated change point, estimated means, and formal intervention time. The estimated means and change point are obtained from modeling the time series with Robust-ITS.

Modeling the mean by equation (3.1) in stage one is sufficient because the residuals act as white noise. Nevertheless, we model the residuals pre- and post-change point with an AR(1) process separately, to provide complete information. The estimates and 95% confidence intervals of the autoregressive parameters and their difference is given in Table 3.4, along with the estimated variance prior to and post change point and their comparison. Both $\widehat{\phi}_1$ and $\widehat{\phi}_2$ do not statistically differ from zero in any of the four units, supporting our claim that the residuals act as white noise. There is no data dependency apparent in either the pre- and post-change point phases. The difference of the two autoregressive parameters, $\widehat{\phi}_2(\widehat{\tau}) - \widehat{\phi}_1(\widehat{\tau})$, also do not statistically differ from zero in the four units.

Because there is no correlation present and the stochastic component is adequately modeled by white noise (indicating independent data), there is valuable information obtained from

Patient Satisfaction				
Parameters	Stroke	Cardiac	Medical Surgical	Mother Baby
AR(1) Coefficient Pre Change Point, $\widehat{\phi_1}(\widehat{\tau})$	-0.056 (-0.460, 0.348)	-0.191 (-0.624, 0.241)	0.078 (-0.377, 0.534)	-0.271 (-0.647, 0.105)
AR(1) Coefficient Post Change Point, $\widehat{\phi_2}(\widehat{\tau})$	-0.354 (-0.713, 0.004)	0.055 (-0.266, 0.376)	0.088 (-0.264, 0.440)	-0.044 (-0.401, 0.392)
Difference in AR(1) Coefficients, $\widehat{\phi_2}(\widehat{\tau}) - \widehat{\phi_1}(\widehat{\tau})$	-0.299 (-0.826, 0.229)	0.246 (-0.278, 0.770)	0.010 (-0.551, 0.570)	0.267 (-0.267, 0.801)
Variance Pre Change Point, $\widehat{\sigma^2}_{1:(\widehat{\tau}-1)}$	10.259	26.474	23.412	8.127
Variance Post Change Point, $\widehat{\sigma^2}_{\widehat{\tau}:T}$	7.976	13.511	23.965	12.649
Variance Comparison F-statistic (p-value)	1.286 (0.248)	1.959 (0.035)	0.977 (0.516)	0.643 (0.88)

Table 3.4: Gives (a.) estimates and 95% confidence intervals of the AR(1) coefficients pre and post change point, and of the estimated increase in the AR(1) coefficient post-change point; (b.) the estimated variances and (c.) 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.

ϵ_t via the variance; the variances are compared using an F-test. The estimates of the variances are smaller post-change point for the Stroke and Cardiac units, and larger for the Medical Surgical and Mother Baby units. Nonetheless, we cannot conclude that the variance differs between the two phases for the Stroke, Medical Surgical and Mother Baby units. For the Cardiac unit the variance post-change point is statistically (at the $\alpha = 0.05$ level) smaller than the variance pre-change point. Therefore, patient satisfaction in the Cardiac unit is more predictable after the introduction of CNL integrated care delivery. A more predictable outcome, less extremely unsatisfied and satisfied patients, signifies a more controlled environment. This is a positive 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.

Comparing Robust-ITS to Segmented Regression We compare level change and trend change for Robust-ITS and two segmented regression models (one with an assumed change

point and another with the set of possible change points removed) in Table 3.5. The aim of Table 3.5 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 differs for each of the fours 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 two segmented regression methods may also provide opposing results.

Patient Satisfaction						
Unit	Change in Level $-\widehat{\delta} - \widehat{\Delta}\widehat{\tau}$			Change in Trend (slope) $\widehat{\delta}$		
	Segmented Regression+	Segmented Regression++	Robust-ITS	Segmented Regression+	Segmented Regression++	Robust-ITS
Stroke	6.04 (1.14, 10.94) 0.02*	5.7 (2.32, 9.07) 0.00**	7 (3.75, 10.25) 0.0**	-0.41 (-0.68, -0.15) 0.00**	-0.25 (-0.45, -0.06) 0.01*	-0.34 (-0.53,-0.15) 0.00**
Cardiac	-4.4 (-11.01, 2.21) 0.19	-1.8 (-6.38, 2.79) 0.44	-2.77 (-7.92, 2.38) 0.29	-0.24 (-0.61, 0.14) 0.21	-0.23 (-0.50, 0.04) 0.10	-0.17 (-0.51, 0.16) 0.30
Medical Surgical	0.94 (-6.64, 8.53) 0.8	0.59 (-4.86, 6.03) 0.83	3.50 (-1.72, 8.72) 0.18	-0.21 (-0.60, 0.18) 0.28	-0.16 (-0.46, 0.14) 0.28	-0.26 (-0.58, 0.06) 0.11
Mother Baby	5.89 (0.64, 11.14) 0.03	5.40 (1.99, 8.80) 0.00	5.40 (2.01, 8.78) 0.00	-0.28 (-0.56, -0.01) 0.05	-0.18 (-0.38, 0.02) 0.07	-0.18 (-0.38, 0.02) 0.07

Table 3.5: Provides approximate 95% confidence intervals for level change and trend change 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. Each row respectively provides the parameter estimate, confidence interval, and p-value. Note, one asterisk, *, denotes significance at the $\alpha = 0.05$ level, and two asterisks, **, denotes significance at the $\alpha = 0.01$ level.

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 the appendices. The segmented regression models are discussed under the assumption that the change point is assumed. Nonetheless, the two main segmented regression models are also equivalent when the set of possible change points are removed.

The model comparisons are provided in Table 3.6, 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 3.5 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).

Mean Squared Error			
Unit	Segmented Regression+	Segmented Regression++	Robust-ITS
Stroke	192.42	190.99	171.99
Cardiac	428.34	416.50	406.85
Medical Surgical	471.80	471.93	459.16
Mother Baby	181.44	161.11	161.11

Table 3.6: Provides 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. 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.

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

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad \text{for } t \in \{1, \ldots, 60\}. \tag{3.3}$$

The estimated patient satisfaction mean curves for both Robust-ITS and model (3.3) are plotted in Figure 3.3 by unit. The parameter associated with quadratic time β_2 is only statistically significant, at the $\alpha = 0.05$ level, for the Stroke unit. Including quadratic time

as a predictor is not necessary for the Cardiac, Medical Surgical and Mother Baby units, since we cannot conclude that β_2 differs from zero. Adding quadratic time as a predictor is useful in the Stroke unit because at the $\alpha = 0.05$ level β_2 differs from zero.

Nevertheless as shown in Table 3.7, 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 (3.3) in all units, indicating Robust-ITS fits the data better in all units. Additionally, model (3.3) assumes a continuous decline after obtaining the maximum. Suggesting model (3.3) will produce a poor patient satisfaction estimate post maximum.

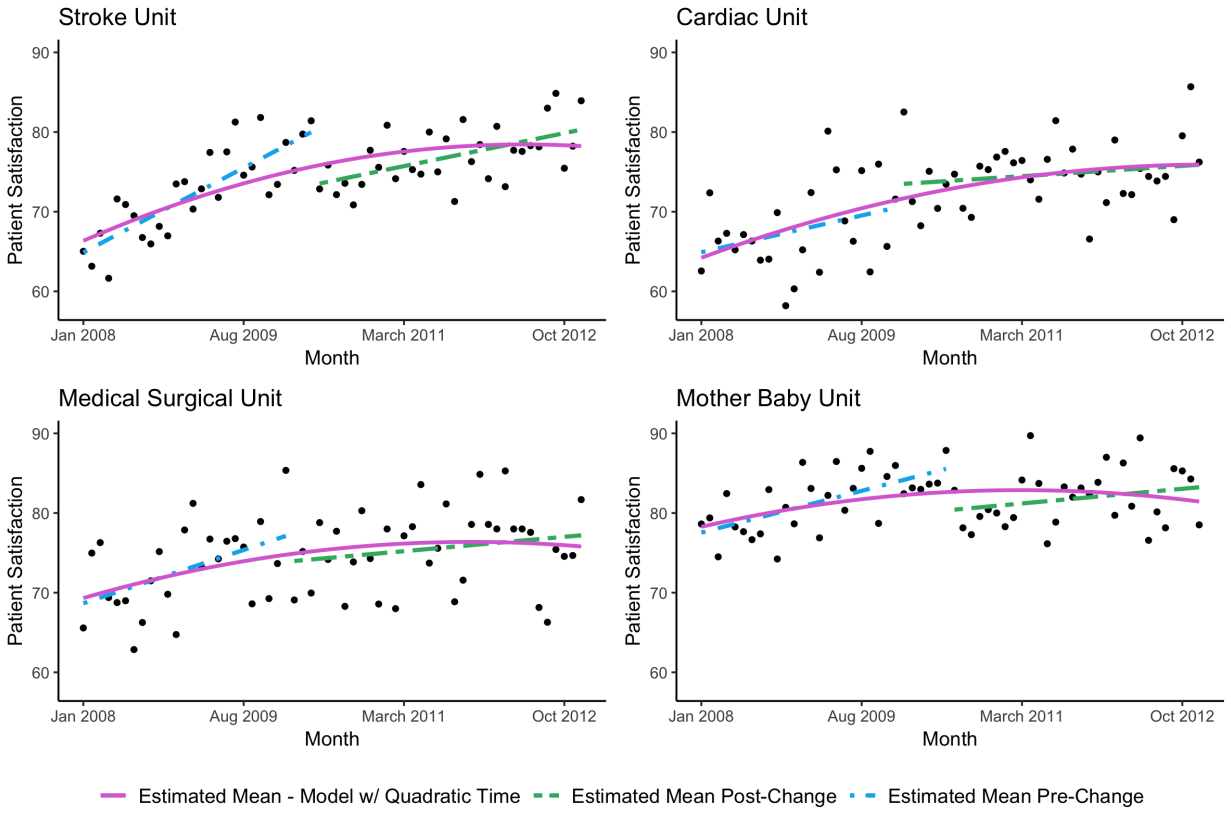

Figure 3.3: Plots of patient satisfaction within each of the four units, along with the estimated means obtained by Robust-ITS and a model with quadratic time as a predictor.

Mean Squared Error		
Unit	Model with Quadratic Time and No Change Point	Robust-ITS
Stroke	224.91	171.99
Cardiac	417.77	406.85
Medical Surgical	460.36	459.16
Mother Baby	184.62	161.11

Table 3.7: Provides mean squared error (MSE), with order of magnitude 10^{-5}, for Robust-ITS and for the non-change point model with quadratic time as a predictor, for patient satisfaction. MSE is the estimate of sum of squared errors, measuring the square of the errors or deviations, divided by the degrees of freedom. Lower MSE suggests a more adequate fit.

3.4 Summary and Conclusions

There are two main stages that compose Robust-ITS. The first is modeling the mean function and the second is modeling the stochastic component. In both stages, Robust-ITS allows for a change in the outcome. To the best of our knowledge, 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 ITS literature.

In the first stage, a set of plausible change points must be established based on the scientific question of interest. Then based on the set of possible change points, Robust-ITS estimates the mean parameters via ordinary least squares 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 one. 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 illustrated that the hypothesized change point is not always assumed adequately. Following traditional segmented regression we would have set the change point at the same value for all units, and assumed 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 three units, and the estimated change point value varied based on unit.

In two of the four 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 function 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 seven 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 Stroke, Cardiac and Mother Baby. Since the autocorrelation was not statistically significant, the variances pre- and post-change point were compared. For the Cardiac unit, 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 regards 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 3.4) 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 in this chapter. The Robust-ITS toolbox is interactive, and provides the user with graphical displays, estimates and inference on testing for differences between the pre- and post-intervention means, correlation, and variance.

The current status of the model is only for single unit analyses and continuous-valued outcomes. In the following chapters, we generalize robust-ITS to a handle multi-unti ITS and to discrete ITS (e.g., infection rates, counts of accidental falls, etc).

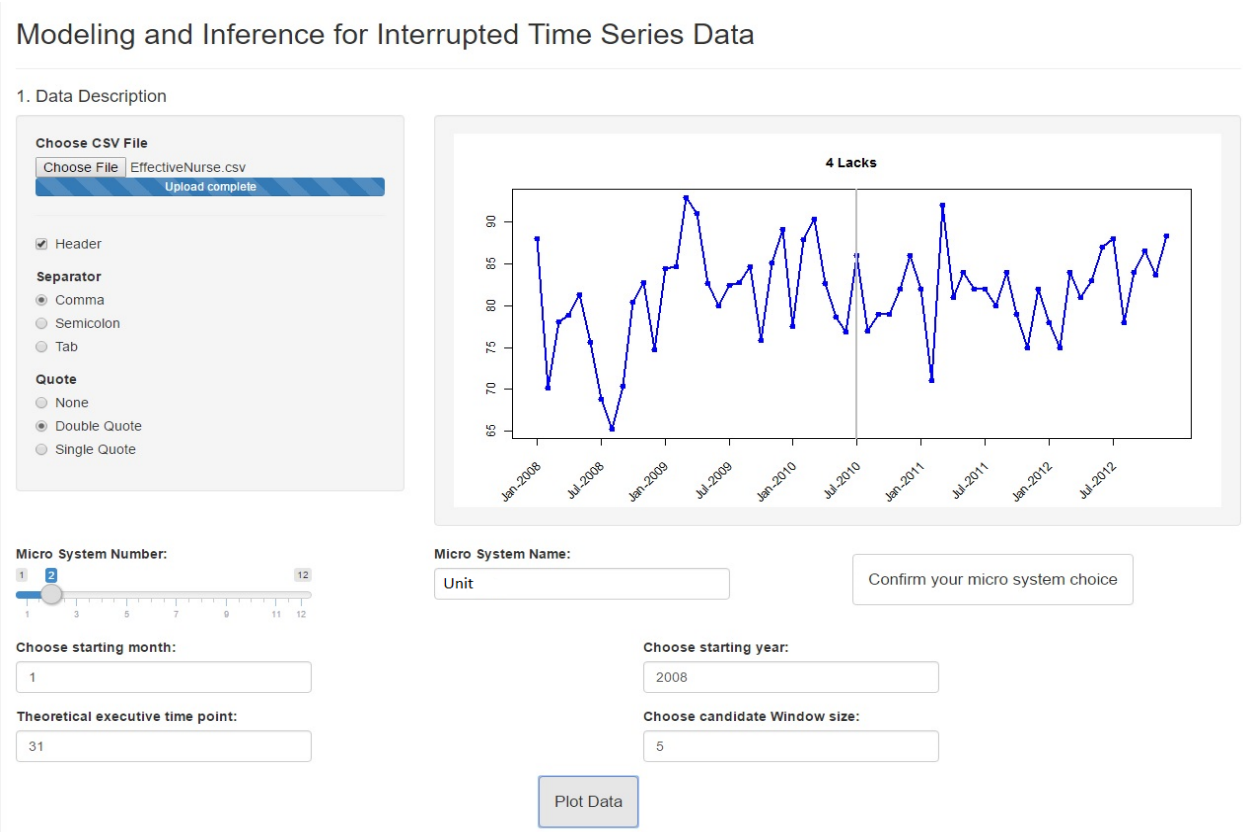

Figure 3.4: The Robust-ITS toolbox in R Shiny (by Ngo, Hu, Cruz, Bender, and Ombao), an interactive toolbox in which the user (1.) may upload their own data in a *.csv* file; (2.) provides basic information of the data — the toolbox requires the user to input the 'theoretical executive time point (TET)' (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; (4.) views the estimates, along with their p-values and standard errors, for both the mean and stochastic processes.

Chapter 4

A Multi-Unit Extension of Robust-ITS

The methods for analyzing ITS data discussed thus far do not borrow information across units. This is a serious limitation because it does not take advantage of all available data that may provide information on the lag associated with a given intervention. A main contribution of the work presented in this chapter are the empirical power studies that illustrate the gain in efficiency obtained by borrowing information across units.

The methodology presented in this chapter is motivated by our interest in estimating the lagged effect of an intervention on average patient satisfaction survey scores, recorded monthly at five clinical care units. A time series plot of patient satisfaction scores from January 2008 to December 2012 at two hospital units (the Stroke and Surgical units), is given in Figure 4.1. There seems to be a change in the mean functions of the Stroke and Surgical units around the middle of the time series, slightly before the formal implementation of the intervention on July 2010. The time series data are from a study aimed to assess the impact of a new nursing care delivery system on publicly recorded standardized quality and safety

metrics (Bender et al., 2015). These metrics are a central area for improvement because the Center for Medicaid and Medicare Services utilizes them for health systems' care services reimbursement (Kavanagh et al., 2012).

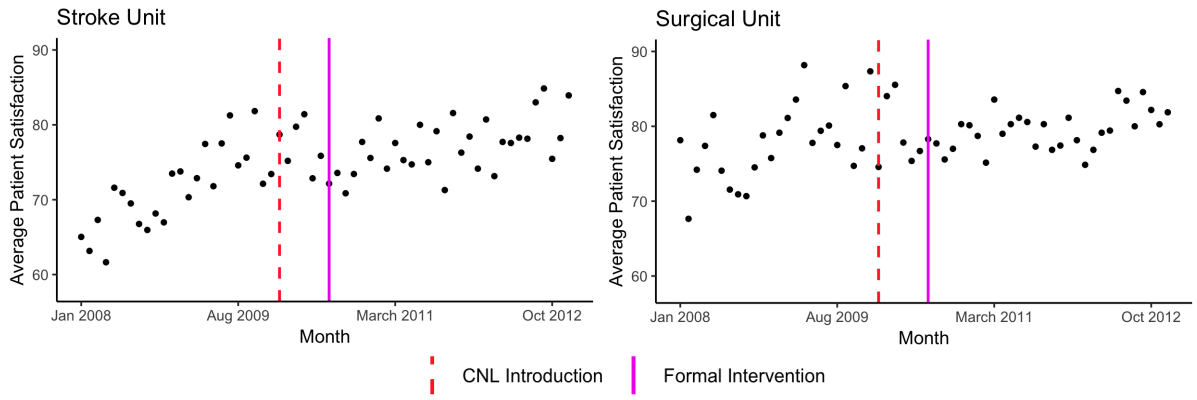

Figure 4.1: Plots the time series of observed average patient satisfaction for the Stroke and Surgical units.

The intervention was the implementation of Clinical Nurse Leader (CNL) integrated care delivery, discussed in Chapter 1. We are interested in estimating the time lag (or delay) between the onset of the intervention and the effect on patient satisfaction. Our proposed model assumes a global change point rather than unit-specific change points (1.) to pool information across hospital units and increase efficiency, and (2.) to reduce the impact of unit specific high-leverage points around the CNL and formal intervention implementation time points. Importantly, we are interested in examining whether or not a change point actually exists, thereby deducing whether or not the intervention impacts patient satisfaction. Our interest is not solely on properly modeling the CNL intervention; we are also interested in future study designs, and so, focus on power.

The most utilized statistical methodology for analyzing ITS data in the health care literature is segmented regression (Taljaard et al., 2014; Penfold and Zhang, 2013; Wagner et al., 2002; Linden, 2015). Segmented regression restricts the analysis to one health care outcome for one unit (group or cluster). In the context of assessing the above intervention, perhaps a severe drawback of segmented regression is that it restricts the interruption to a predetermined

time point in the series or censors data by removing the set of time points for which the intervention effects may not be realized. Additionally, segmented regression neglects the plausible differences in autocorrelation and variability between the pre- and post-intervention phases present in the data. The Robust-ITS model proposed in the previous chapter treats the change point as variable, appropriate for situations where the data warrants such treatment, and tests for differences in autocorrelation and variability pre- and post-change point. Nevertheless, Robust-ITS and segmented regression both neglect shared information across hospital units and inherently assume a change point *exists*.

Assessing the impact of an intervention with traditional segmented regression or Robust-ITS on these data requires a separate analysis for each individual unit. We expect many of the units to share several characteristics — i.e., abide by the same regulations, have similar schedules, hire staff based on the same criteria, etc. — because the units are housed within one hospital. Moreover, we expect the CNL 'training' or education for each of the CNL students to include commonalities, such as course work and care delivery ideology. Assessing the intervention impact on multiple units via current segmented regression methods ignores shared characteristics across units, in particular the similarity between characteristics influencing the change-point.

Inherently assuming that a change point exists, as in segmented regression and Robust-ITS, may lead to erroneous results when there is no actual a change point. Change point models will forcefully quantify a change in the outcome regardless of the presence of a true change point. This is a problem whether the change point is determined *a priori* or estimated over a set of possible change points. Assuming a change point exists when it truly does not, will force a model to provide an estimate of an artificial difference in the outcome. To avoid incorrectly specifying an unnecessary change point and regression to the mean phenomena, we focus on formally testing for the existence of a change point.

In this chapter, we develop the Robust Multiple ITS model (R-MITS), a novel extension of Robust-ITS, appropriate for multiple independent interrupted time series. Furthermore, we present the supremum Wald test, able to test for the existence of a change point across units. Importantly, we provide empirical type one error, power, and accuracy studies assessing the operating characteristics of our developed methodology. The proposed method (a.) borrows information across hospital units to increase efficiency, (b.) estimates a global change point of an instituted intervention, (c.) formally tests for the existence of a change point in the unit specific mean functions, and (d.) allows for changes in the mean functions and autocorrelation structures across units.

We go on to describe our proposed R-MITS model and provide details on the estimation and inference procedures. In our model specification we outline the supremum Wald test used to determine the necessity of a change point. Next, we present empirical simulations to assess the type one error, power for detecting specified change point alternatives, and accuracy of the change point estimation procedure. We then analyze the impact of Clinical Nurse Leader integrated care delivery on patient satisfaction. Lastly, we present a summary of our developed methodology and briefly describe future work.

4.1 The Robust Multiple ITS (R-MITS) Model

Our proposed model tests for the existence — rather than merely assume — of a change point and adequately manages multiple units/time series. A noteworthy feature of our approach is the clear distinction between the time of intervention and the change point, as in Robust-ITS. Setting the change point to a predetermined time may lead to incorrect measures of the intervention's effect on the system; particularly when set to the intervention time, because that does not necessarily represent the reality that complex interventions may have varied effects and take time to manifest change. Prevalent approaches to overcoming this limitation

are to remove, or censor, a specific set of time points from the analysis (Penfold and Zhang, 2013; Taljaard et al., 2014). R-MITS borrows information from all microsystems to estimate a global change point; i.e., determines the time point at which the effect of the intervention initiates for the entire health system. Moreover, detecting differences in autocorrelation and variances pre- and post-intervention is critical in evaluating the effectiveness of an intervention. The R-MITS model allows for two completely different data dependency and variability structures to exist prior to the intervention and post-intervention within each unit.

As in Chapter 3, to prelude model development we plot the outcome against time to (a.) illuminate the functional form of the longitudinal mean over time; (b.) determine the presence of seasonality, and (c.) further investigate the set of plausible change points and the necessity of a change point. If the functional form of the longitudinal mean is not linear, we transform the outcome to obtain a linear pattern, or apply a different segmented regression model appropriate for the pattern present within the ITS design. When needed, we account for seasonality via traditional statistical methods concisely described in (Bhaskaran et al., 2013). Although not used in the analyses here, variance stabilizing transforms can be applied on the outcomes of interest if necessary. In our interrupted time series data, the longitudinal mean functions are relatively linear in time with no apparent seasonality. Thus, no transformations are applied on the outcomes of interest.

4.1.1 Description of R-MITS

Denote t^* as the time point at which the intervention is introduced and τ as the time point at which the effect of the intervention initiates (the change point) for the outcome of interest. Sometimes it may indeed be true that $t^* = \tau$, but this may not necessarily be true for all outcomes. Often it is entirely possible that the time of effect of the intervention

differs from the time of intervention introduction (i.e., either $\tau > t^*$ or $\tau < t^*$). If $\tau > t^*$ then the effect of the intervention on the outcome is not realized until after the formal intervention time point. As might be the case when a learning effect exists with regards to the intervention, thereby leading to a delay in the realization of the full intervention impact. When $\tau < t^*$ there is an anticipatory intervention effect on the outcome. This may be the case in our motivating study, where Clinical Nurse Leaders are introduced into units prior to the formal intervention start time. We propose a data adaptive procedure for estimating and determining the existence of τ, discussed in Section 4.1.1. Many change point detection methods in time series exist, but often deal only with changes in the mean functions and variance (not the autocorrelation structure itself), and may not work well in shorter time series (Davis et al., 2006; Kirch et al., 2015). The method proposed in this chapter can suitably manage changes in the autocorrelation structure, as well as in the mean functions and volatility.

Define y_{it} as the outcome of interest for hospital unit i at time t (where $i = 1, \ldots, N$ and $t = 1, \ldots, n$). For example, y_{it} may be patient satisfaction for the Stroke unit at time t. The general regression is defined as

$$y_{it} = \mu_{it} + \epsilon_{it}, \tag{4.1}$$

where μ_{it} is the mean function and ϵ_{it} is the stochastic process that models that fluctuations around the mean function. The mean component, μ_{it}, characterizes the mean function of the response for unit j during the pre-intervention and post-intervention phases. The stochastic process, ϵ_{it}, accounts for the variability and correlation of the outcome in the i^{th} unit. In the following discussion we define the mean functions and stochastic components for the R-MITS model, and the estimation procedures.

The Pre- and Post-Intervention Mean Function

The mean function of the outcome for hospital unit i at time t is

$$\mu_{it} = \begin{cases} \beta_{i0}^{\tau} + \beta_{i1}^{\tau}\, t, & t < \tau \\ (\beta_{i0}^{\tau} + \delta_i^{\tau}) + (\beta_{i1}^{\tau} + \Delta_i^{\tau})t, & t \geq \tau \end{cases} . \tag{4.2}$$

The parameters in μ_{it} are: (1.) β_{i0}^{τ}, the intercept of the mean function prior to the change point; (2.) β_{i1}^{τ}, the slope of the outcome prior to the change point; (3.) $\beta_{i0}^{\tau} + \delta_i^{\tau}$, the intercept of the post-intervention phase; (4.) $\beta_{i1}^{\tau} + \Delta_i^{\tau}$, the slope of the post-intervention phase; for the outcome in unit i, and (5.) τ, the global over-all-unit change point of the response. Thus, $\delta_i^{\tau} = \Delta_i^{\tau} = 0$ implies there is no change in the mean structure before and after time τ. Health care specialists are primarily interested in testing for the intervention lag (delay in the effect of the intervention), and the differences in the outcome means between the pre- and post-change point phases.

Remark (1.) The metrics adopted by the health policy evaluation literature to assess the effect size of an intervention via ITS designs are the change in level and change in trend (or slopes). 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 function. It is necessary to report both level change and change in trend to interpret the results of an ITS study accurately (Effective Practice and Organisation of Care, 2015).

Remark (2.) The level change is interpreted as the change in the anchored intercept (anchored at the change point), and is therefore the jump between the projected mean function based on the pre-change point phase and the estimated mean function post-change point. In our model the unit specific change in level is defined mathematically as $\delta_i^{\tau} + \Delta_i^{\tau}\tau$, and is graphically depicted in Figure 4.2. Trend change, or slope change, is denoted by Δ_i in the mean function, equation (4.2).

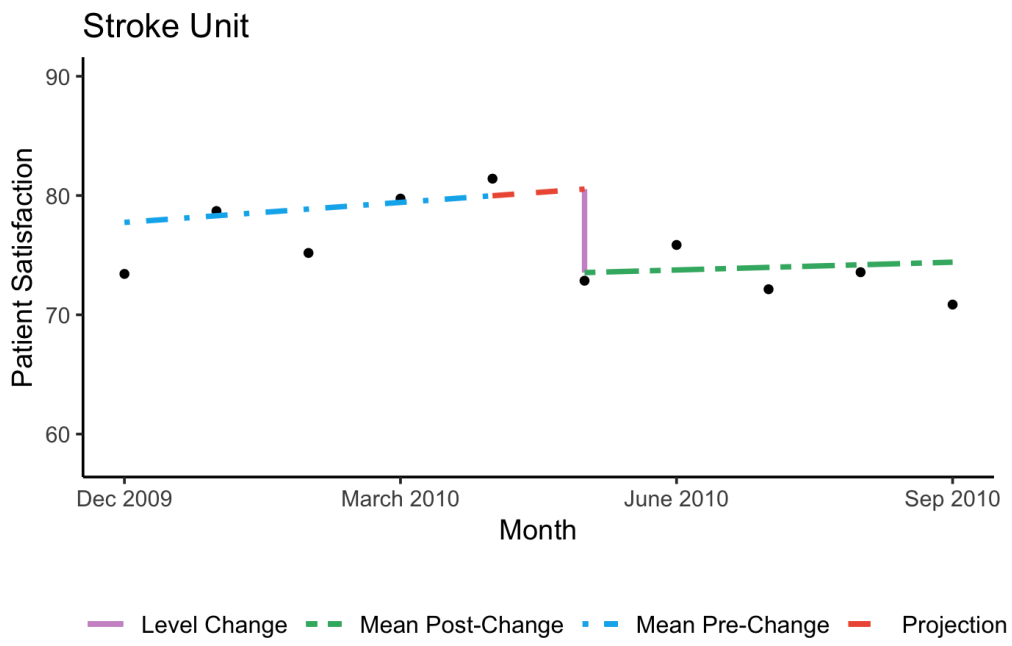

Figure 4.2: An example of a segmented regression model fit for the Stroke unit. The plot depicts (1) the segmented regression lines fit to the pre- and post-change point phases, (2) the projection of the mean at the change point based on the pre-change point regression, and (3) the change in level as defined here. The plot contains data from January 2010 to September 2010, instead of the entire observational period, to clearly illustrate the level change.

The mean function parameters are estimated simultaneously with the stochastic component parameters and change point, via maximizing the conditional likelihood given in equation (4.4) of Section 4.1.1, with the auto-regressive coefficients' estimator accounting for the volatility of the shifted series. with variance components shifted AR(1) An algorithm on how to precisely estimate the parameters is provided in Section 4.1.1.

Stochastic Properties Pre- and Post-Intervention

The stochastic component in equation (4.1), ϵ_{it}, captures the autocorrelation structure of the outcome variable across time for unit i, and may change as a result of the intervention; the ϵ_{it} are zero-mean random fluctuations around the mean function of unit i. One goal

of the CNL intervention is to increase the consistency of care delivery and hence patient assessment outcomes, (i.e., decreasing variability of the outcomes). We therefore include separate stochastic components for the pre- and post-change point phases, to allow for a change in outcome variability.

Due to the impact of the intervention, the stochastic process pre-intervention might differ from the stochastic process post-intervention. That is, ϵ_{it} for $t \in \{1, \ldots, \tau - 1\}$ may be a different stochastic process than ϵ_{it} for $t \in \{\tau, \ldots, n\}$. Note, the length of the time series is denoted by n. Hence, the autocorrelation and variance might differ pre- and post-change point. Here, the stationarity requirement is satisfied if the variance, mean function, and autocorrelation are constant within each stochastic process, not constant across all time points.

In order to fit stationary AR or ARMA processes to the stochastic components, one should first confirm that there are no striking signs of non-stationarity. That is, the mean and variance of the residuals (obtained from modeling and removing the mean function as in the previous section) must be relatively constant. If the mean function is not misspecified, then the residuals should be fluctuating around zero without trend. Moreover, the residuals should be stationary within each of the pre- and post-intervention phases (Shumway and Stoffer, 2017). Our analysis of patient satisfaction suggests that it is reasonable to assume stationarity within each phase, and hence we proceed with the assumption of stationarity.

In this work we use the AR(1) process to model the stochastic component, $\vec{\epsilon}_i = Y_i - \vec{\mu}_i$, where $Y_i = [y_{i2}, \ldots, y_{in}]'$ and $\vec{\mu}_i = [\mu_{i2}, \ldots, \mu_{in}]'$ for unit i. Note, y_{i1} is not included in Y_i and μ_{i1} is not included in $\vec{\mu}_i$, because we condition on the first observation. The AR(1) coefficient is estimated by maximizing the conditional likelihood with the denominator of the estimator averaging the volatility of the shifted AR(1) series. We therefore condition on the first observation y_{i1}. Since the mean function $\vec{\mu}_i$ is not known (we only have its estimate, $\widehat{\vec{\mu}}_i$), the stochastic component is not directly observed. Hence, we use the residuals

$R_i = Y_i - \widehat{\vec{\mu}}_i \equiv [r_{i2}, \ldots, r_{in}]'$ in place of $\vec{\epsilon}_i$. The residuals are modeled as:

$$r_{it} = \begin{cases} \phi_{i1}(\tau)\, r_{i,t-1} + e_{it,1}, & 1 < t \leq \widehat{\tau} - 1 \\ \phi_{i2}(\tau)\, r_{i,t-1} + e_{it,2} & \widehat{\tau} - 1 < t \leq n. \end{cases} \tag{4.3}$$

To ensure causality in the time series sense, $\phi_{i1}(\tau)$ and $\phi_{i2}(\tau)$ must lie in the interval $(-1, 1)$ for all i. Note, the auto-regressive coefficient prior to the change point, $\phi_{i1}(\tau)$, is the correlation between time point t and $t+1$ (the adjacent correlation or autocorrelation) where t and $t+1$ belong to the pre-change point phase (t, and $t+1 \in \{1, \ldots, \tau - 1\}$), and $\phi_{i1}^{|h|}(\tau)$ is the correlation between two time points h units away (say t and $t+h$, both in the pre-change point phase) of the outcome. The auto-regressive coefficient post-change point, $\phi_{i2}(\tau)$ has a similar interpretation. The zero-mean random fluctuations of model (4.3) are white noise, $e_{it,j} \overset{iid}{\sim} N\big(0, \sigma^2_{iw,j}\big)$ for $j \in \{1, 2\}$. The variance of the distribution of the response at any time point t is $\sigma^2_{ij} = \frac{\sigma^2_{iw,j}}{1-\phi_{ij}(\tau)^2}$ for $j \in \{1, 2\}$.

The variance and auto-regressive coefficients in the AR(1) setting can be estimated by maximizing the conditional likelihood provided in equation (4.4) of Section 4.1.1, with the auto-regressive coefficients' estimator accounting for the volatility of the shifted series. The structure of the variance-covariance matrix, and the estimators of the auto-regressive coefficients and white-noise standard deviations are given in the appendices.

To determine whether the stochastic process differs as a result of the change point for each unit, one can test the hypothesis that $\nu_i(\tau) \equiv \phi_{i2}(\tau) - \phi_{i1}(\tau)$ equals zero. This can be tested by either estimating $\nu_i(\tau)$ directly or by conducting an F-test for nested models. The F-test for nested models for the AR(1) scenario is described in Chapter 3.

Estimation of the Change-Point and Model Parameters

In this chapter we propose a conditional likelihood procedure for estimating the global change point. The set of possible change points is established by the researcher. We estimate the change point and therefore all of the parameters, both from the mean functions and stochastic components, simultaneously by obtaining the generalized least squares estimates. Then we test for the existence of a change in the mean functions — i.e., we test the null hypothesis that there is no change in any of the mean functions versus the alternative that there is a change in at least one of the mean functions — at each possible change point by applying the supremum Wald test, described in section 4.1.1.

Define the length of the time series as n, the number of units as N, the vector of mean function parameters as $\theta_i = [\beta_{i0}^{\tau},\ \beta_{i1}^{\tau},\ \delta_i^{\tau},\ \Delta_i^{\tau}]'$ and $\mathbf{\Sigma}_i$ as the variance-covariance matrix of the response in unit i. The structure of the variance-covariance matrices is included in the appendices.

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. Recall the response vector for unit j is $Y_i = [y_{i2}, \ldots, y_{in}]'$. Note, y_{i1} is not included in Y_i because we model the zero-mean random fluctuations around the mean functions as AR(1) processes. For each candidate change point $q \in Q$ we derive the conditional likelihood function, conditional on the first observations,

$$\begin{aligned} & L\left(\theta_1,\, \mathbf{\Sigma}_1\,,\, \ldots,\ \theta_N,\, \mathbf{\Sigma}_N \mid q,\, Y_1,\, \ldots, Y_N\right) \\ & \quad \equiv \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} |\mathbf{\Sigma}_i|^{-\frac{1}{2}}\ \exp\left\{ -\frac{1}{2}\big(\, Y_i - \mathbf{X}(\mathbf{q})_i\ \theta_i\,\big)'\, \big(\mathbf{\Sigma}_i\big)^{-1}\big(\, Y_i - \mathbf{X}(\mathbf{q})_i\ \theta_i\,\big)\right\}, \end{aligned} \tag{4.4}$$

where

$$\underset{(n-1)\times 4}{\mathbf{X}(\mathbf{q})_i} \equiv \begin{bmatrix} 1 & 2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & q-1 & 0 & 0 \\ 1 & q & 1 & q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & n & 1 & n \end{bmatrix}.$$

We iteratively estimate $\underset{4\times 1}{\theta_i}$ and $\underset{(n-1)\times(n-1)}{\mathbf{\Sigma}_i}$ for all i, as in Algorithm 1.

Algorithm 1 Estimating θ_i and $\mathbf{\Sigma}_i$ iteratively

1: **for** $i \in \{1, \ldots, N\}$ **do**
2: set $\zeta = 1$
3: set $\widehat{\theta}_i^0$ to OLS estimates
4: from residuals R_i^0 calculate $\widehat{\phi}_{i1}^0(q)$, $\widehat{\phi}_{i2}^0(q)$, $\widehat{\sigma}_{i1}^0$, and $\widehat{\sigma}_{iw,2}^0$ and generate $\widehat{\mathbf{\Sigma}}_i^0$
5: **while** $\zeta >$ tol **do**
6: set k to the iteration
7: calculate $\widehat{\theta}_i^i$ based on $\widehat{\mathbf{\Sigma}}_i^{k-1}$
8: use residuals R_i^k to estimate $\widehat{\phi}_{i1}^k$, $\widehat{\phi}_{i2}^k(q)$, $\widehat{\sigma}_{iw,1}^k$, and $\widehat{\sigma}_{iw,2}^k$
9: obtain $\widehat{\mathbf{\Sigma}}_i^k$,
10: set ζ to the Euclidean distance between $[\widehat{\phi}_{i1}^{k-1}(q), \widehat{\phi}_{i2}^{k-1}(q)]$ and $[\widehat{\phi}_{i1}^k(q), \widehat{\phi}_{i2}^k(q)]$
11: **end while**
12: **end for**

Define

$$L(q) = \max_{(\theta_1, \mathbf{\Sigma}_1, \ldots, \theta_N, \mathbf{\Sigma}_N)} L\left(\theta_1, \mathbf{\Sigma}_1, \ldots, \theta_N, \mathbf{\Sigma}_N \mid q, Y_1, \ldots, Y_N\right).$$

Then the estimated change point is

$$\widehat{\tau} = \arg\max_{q \in Q} L(q).$$

The estimates of $\theta_1, \ldots, \theta_N$ are the generalized least squares (GLS) estimates obtained, after the desired tolerance level is reached, conditional on $\widehat{\tau}$. The GLS estimates of θ_i and $\mathbf{\Sigma}_i$ for all i given $\widehat{\tau}$ are

$$[\,\{\widehat{\theta}_1,\ \widehat{\mathbf{\Sigma}}_1\},\ \ldots,\ \{\widehat{\theta}_N,\ \widehat{\mathbf{\Sigma}}_N\}\,] \ = \underset{[\{\theta_1,\ \mathbf{\Sigma}_1\},\ \ldots,\ \{\theta_N,\ \mathbf{\Sigma}_N\}]}{\arg\max} L\big(\theta_1,\ \mathbf{\Sigma}_1,\ \ldots, \theta_N,\ \mathbf{\Sigma}_N \,|\, \widehat{\tau}, Y_1,\ \ldots, Y_N\big).$$

The presence of τ does not restrict the model to a fixed interruption with an instantaneous effect. In fact, τ allows the design matrix and estimates to transform based on the information the data provides. Importantly, the inclusion of an over-all-unit change point allows us to utilize information from all available units to determine when the intervention begins to affect the outcome globally. This flexibility of the model can be helpful in minimizing misleading results from an assumed change point.

Multivariate Wald Test for the Existence of a Change Point

The change point is estimated by maximizing the conditional likelihood over the set Q, and thus, concurrently estimates all other model parameters at each possible change point. Since we test for the existence of a change point at each $q \in Q$, multiple testing bias exists if one utilizes standard critical values. As such it is necessary to apply a correction to control the family-wise type I error rate. To this end, we calculate the multivariate Wald test statistic for every $q \in Q$. We apply the Benjamini-Hochberg method — a procedure that compares ranked and ordered (smallest to largest) p-values to corresponding critical values individually calculated based on the total number of tests, p-value rank, and desired false discovery rate

— to adjust for the total number of tests conducted. The total number of tests is equal to the cardinality of Q. The Benjamini-Hochberg method controls the false discovery rate; control of the false discovery rate weakly implies control of the family wise type 1 error rate for an $\alpha = 0.05$ level (Benjamini and Hochberg, 1995). In this case, a binary decision of whether a change point exists or not corresponds to a rejection of the null hypothesis for any one of the tests conducted.

We focus on determining the existence of a change point across the unit specific mean functions, i.e., for each $q \in Q$ we test

$$\begin{aligned} & H_0 : \ \delta_i^q = \Delta_i^q = 0 \ \ \forall \ i, \ i = 1, \ldots, N \ \text{(no change point)} \\ \text{vs.} \ \ & H_a : \ \delta_i^q \neq 0 \ \text{and/or} \ \Delta_i^q \neq 0, \ \text{for some} \ i, \ i = 1, \ldots, i, \ \ \text{(a change point at } q\text{)}. \end{aligned}$$

Even though our model assumes a global change point to pool information across units for efficiency, a rejection of the null hypothesis for our Wald test implies a change point in at least one of the hospital units. A rejection does not imply that a change point exists across all units and is the same in all hospital units. Moreover, we do not restrict the impact of the change point at each unit — i.e., we allow the change in level and change in slope to differ across units as in R-MITS. We borrow information across units for the estimation of the global change point, but we do not force the impact on the outcome to be the same in each unit. Intervention implementation theoretically integrates care delivery policies laterally across hospital units, yet the intervention may impact each unit's underlying care processes uniquely. Clearly, if one wanted to establish the existence of a change point for a particular unit, enough data would have to be gathered within that single unit to detect and estimate (with high enough precision) the unit specific change point.

The hypotheses can be written in terms of full and reduced mean function models. Define the full and reduced mean function models as

$$\mu_{it}^1 = \beta_{i,0}^q + \beta_{i,1}^q\, t \;+\; \big(\delta_i^q + \Delta_i^q\, t\big) I(t \geq q), \tag{4.5}$$

$$\mu_{it}^0 = \beta_{i,0}^0 + \beta_{i,1}^0\, t, \tag{4.6}$$

respectively. The full model, appropriate under the alternative hypothesis, is essentially the model of equation (4.2) and models a change in the mean functions at q. The reduced model, appropriate under the null hypothesis, assumes the same mean functions across the entire observational period. Based on our model specification, testing the above hypotheses is equivalent to testing

$$H_0 : \begin{bmatrix} \delta_1^q \\ \Delta_1^q \\ \vdots \\ \delta_N^q \\ \Delta_N^q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \text{vs.} \quad H_a : \begin{bmatrix} \delta_1^q \\ \Delta_1^q \\ \vdots \\ \delta_N^q \\ \Delta_N^q \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}. \quad \text{Let } \mathbf{C} = \begin{bmatrix} 0\ 0\ 1\ 0 & & \\ 0\ 0\ 0\ 1 & & \mathbf{0} \\ & \ddots & \\ & & 0\ 0\ 1\ 0 \\ \mathbf{0} & & 0\ 0\ 0\ 1 \end{bmatrix},$$

$\vec{\beta}^1 = [\beta_{1,0}^q\ \beta_{1,1}^q\ \delta_1^q\ \Delta_1^q \ldots \beta_{N,0}^q\ \beta_{N,1}^q\ \delta_N^q\ \Delta_N^q]'$ and $\vec{\beta}^0 = [\beta_{1,0}^0\ \beta_{1,1}^0 \ldots \beta_{N,0}^0\ \beta_{N,1}^0]'$. Then the hypotheses can be written as

$$H_0 : \mathbf{C}\,\vec{\beta}^1 = \vec{0} \;\text{ vs. }\; H_a : \mathbf{C}\,\vec{\beta}^1 \neq \vec{0},$$

The multivariate Wald test statistic is given by

$$W = \big(\mathbf{C}\widehat{\vec{\beta}}^{\,1}\big)' \big[\mathbf{C}\ \widehat{\mathbf{V}}(\widehat{\vec{\beta}}^{\,0})\,\mathbf{C}'\big]^{-1}\big(\mathbf{C}\widehat{\vec{\beta}}^{\,1}\big) \overset{H_0}{\sim} \chi^2_{2N}, \tag{4.7}$$

where $\widehat{\mathbf{V}}(\widehat{\vec{\beta}}^{\,0})$ is the block diagonal estimator of the covariance matrix of $\widehat{\vec{\beta}}^{\,0}$. We specify $\widehat{\mathbf{V}}(\widehat{\vec{\beta}}^{\,0})$ in the Appendix. Note, we allow δ_i^q and Δ_i^q to differ for all i, i.e., for all units.

We calculate the multivariate Wald statistic for each $q \in Q$. Then we apply the Benjamini-Hochberg procedure to obtain corrected critical values. The Bejamini-Hochberg procedure is fully described by (Benjamini and Hochberg, 1995). If any of the multivariate Wald tests provide significant results, when compared to the corrected critical values, we conclude that a change point exists for at least one of the units. The resulting 'supremum Wald test' (SWT) is appropriate for detecting a change in any of the mean functions over a set of possible change points. Our test accounts for the heterogeneity of the mean functions and autocorrelation structures across units. In the following sections we illustrate that the SWT has empirically high power under specified change point alternatives.

4.2 Empirical Type One Error and Power Simulations

Prior to analyzing the outcome of interest, we conduct simulations to (1.) examine the type one error rate, and (2.) determine the power and accuracy of our proposed methodology to detect a global change point in the mean functions of the response. These simulations examine the operating characteristics of our proposed supremum Wald test under various conditions. We continue to test

$$\begin{aligned} & H_0 : \ \delta_i^\tau = \Delta_i^\tau = 0 \ \ \forall \ i, \ i = 1, \ldots, N \text{ (no change point)} \\ \text{vs.} \ \ & H_a : \ \delta_i^\tau \neq 0 \text{ and/or } \Delta_i^\tau \neq 0, \text{ for some } i, \ i = 1, \ldots, N, \ \ \text{(a change point at } q) \end{aligned}$$

with $q \in Q$ (the set of possible change points specified by the researcher). The full and reduced models of these simulations are those of equations (4.5) and (4.6), respectively. We have additionally examined the scenario with standardized quadratic time (and standardized linear time) in the mean functions of the reduced and full models. We focused on standardized time, as opposed to untransformed time, to avoid collinearity between the two time terms. These simulations are omitted for brevity, though we note that we obtain similar results

as those discussed in the following sections. In both sets of simulations we assume an autocorrelation structure that remains constant over the entire duration of the observational period, since the focus is on testing for the existence of a change point in the mean functions.

The outcome of interest is recorded for 60 time periods, in five units, with adjacent correlation estimates smaller than $\phi = 0.1$, and the set of possible change points equal to $\{25, \ldots, 34\}$. Thus, we chose parameters similar to these values for our simulations. We consider two values of the time series length, $n \in \{60, 120\}$, two values of the adjacent correlation, $\phi \in \{0.1, 0.6\}$, and three values for the total number of units, $N \in \{1, 3, 5\}$. When the length of the time series is $n = 60$, we allow the set of possible change points to be $Q_{60} = \{25, \ldots, 34\}$, as with the patient satisfaction data. In this situation we conduct 10 total tests, since there are 10 elements in Q_{60}. When the length of the time series is $n = 120$ we allow the set of possible change points to be $Q_{120} = \{50, \ldots, 69\}$; a total of 20 tests are conducted for Q_{120}.

We choose to compare two values of the time series length to illustrate the possible gain in efficiency longer time series provide with regards to power. We illustrate the gain that may come from doubling the length of the time series. The length of the time series can be increased in two ways: (1) increase the observational period, say from 5 years to 10 years; and/or (2.) increase the resolution of recordings, i.e., record patient satisfaction bi-monthly, as opposed to monthly. The two values of ϕ examined are larger adjacent correlation values than what we estimate for the patient satisfaction data. The largest unit-specific adjacent correlation estimates obtained for the patient satisfaction data (when information is not borrowed across units) is 0.09, and so 0.1 is an upper bound for the adjacent correlation in our setting. The value $\phi = 0.6$ represents an upper bound for the correlation between repeated measurements in the literature. The estimated adjacent correlations for patient satisfaction are smaller than either 0.1 and 0.6. Our simulation results are conservative because power decreases for ITS designs as the adjacent correlation increases (Zhang et al., 2011).

Importantly, we conduct type one error and power simulations for the supremum Wald test with one, three and five units. We examine the case with a single unit, $N = 1$, to illustrate the performance of our supremum Wald test in the traditional ITS analysis setting. We explore the value $N = 3$ to depict the healthy gain in efficiency that borrowing information across a small number of units yields. Lastly, we consider $N = 5$ because patient satisfaction is recorded at five units. Our aim is to highlight the improvement in power that borrowing information across units can provide.

4.2.1 Empirical Type One Error for SWT

We provide the empirical type one error rates when testing for the existence of a change point via the supremum Wald test. Four different scenarios are considered for one unit, three units, and five units. We generate $10,000$ time series for each scenario under the reduced model — i.e., from one overall regime where there is no change point present in either the mean functions or stochastic processes. For the case when there is only one unit we set the mean function parameters to $\vec{\beta} = [65, 0.5]'$. When there are three units or five units, the mean function parameters vary slightly across individual units. The white noise standard deviation, σ_w, is always set to 3.38, regardless of the number of units in the simulation. The value $\sigma_w = 3.38$ is approximately the average of the single unit estimates of the white noise standard deviation for patient satisfaction. The response standard deviation, σ, is 4.23 when $\phi = 0.6$ and 3.40 when $\phi = 0.1$ for all individual units. The mean function parameter values mimic results obtained from the patient satisfaction data.

The empirical type one error rates for each scenario are provided in table 4.1. As expected, the empirical type one error rate is smaller for the longer time series, and for smaller values of the adjacent correlation. Larger values of adjacent correlation imply a smaller number of effective independent statistical information. The larger adjacent correlation quantity

corresponds to higher type one error rates exclusively. In the scenario with the shorter time series and high adjacent correlation value, it is more difficult to control the family-wise type one error rate, even as we increase the number of units. The lack of type one error rate control in short time series with high correlation values is exacerbated in the simulations with quadratic time in the mean functions. In fact, for that particular setting the type one error rate becomes worse as the number of units increases. This is primarily attributable to the increased dependency in an already short time series that reduces the information in the time series. Because of this we recommend our proposed procedure when the length of the time series is at least 120 time points in cases with complex mean functions and/or highg correlation values. For all other scenarios considered, the type one error rate is well controlled, though slightly conservative because of the Benjamini-Hochberg multiplicity correction.

Adjacent	Time Series of Length 60			Time Series of Length 120		
Correlation	One Unit	Three Units	Five Units	One Unit	Three Units	Five Units
$\phi = 0.1$	0.0295	0.0291	0.0342	0.0274	0.0265	0.0263
$\phi = 0.6$	0.0460	0.0704	0.1003	0.0299	0.0318	0.0436

Table 4.1: The empirical type one error rate: the proportion of iterations for which we rejected the null hypothesis of no change point. The larger adjacent correlation quantity corresponds to higher type one error rates exclusively. The type one error rates are reasonable for almost all of the scenarios, and stay reasonable as the number of units increases. However, it is slightly difficult to control the type one error rate at the desired $\alpha = 0.05$ level with the smaller time series and high adjacent correlation value.

4.2.2 Empirical Power for SWT

We conduct simulation-based power calculations when testing for the existence of a change point via the supremum Wald test. Time series are generated under the alternative model appropriate in our setting, i.e., generated with a global change point in the mean functions. The change point is set at the middle of the time series; cases with the change point at the

boundary or close to the boundary have been considered and yield similar, yet slightly less powerful, results. We focus on providing power as a function of the slope change. Simulation-based power calculations with power as a function of the auto-regressive coefficient for ITS designs are provided in (Zhang et al., 2011).

Power is examined as a function of the slope change, with a change in baseline intercept (δ_i^{τ}) set to zero for all i. Note, estimates of δ_i^{τ} obtained from the patient satisfaction data are not statistically different from zero. The range of values for the change in slope, $\{0, 0.01, \ldots, 0.24, 0.25, 0.30, \ldots, 0.40, 0.45\}$, encompass the estimated quantities of the change in slope for the patient satisfaction data. Similar to the type one error simulations, the white noise standard deviation is set to 3.38 for all units, yielding a response standard deviation of 4.23 when $\phi = 0.6$ and 3.40 when $\phi = 0.1$.

Simulated power curves are provided in Figure 4.3, with each subfigure corresponding to a separate data generation regime. As expected, power increases as the slope change and length of the time series increases, and power decreases for the larger adjacent correlation value. Power is consistently higher for the larger number of units across the four scenarios, thereby illustrating that the supremum Wald test gains power as the number of units increases by borrowing information across units. Analyzing multiple time series data (or data from multiple hospital units) jointly, results in higher power.

Accurate Estimation of the Change Point

The power simulation results, provided in Figure 4.3, suggest that the supremum Wald test has reasonable power to detect an existing global change point, and that power increases as the number of units increase. We are not simply interested in power by itself. We are also interested in whether R-MITS will provide the correct global change point estimate when our supremum Wald test concludes that a change point exists. Figure 4.4 illustrates the

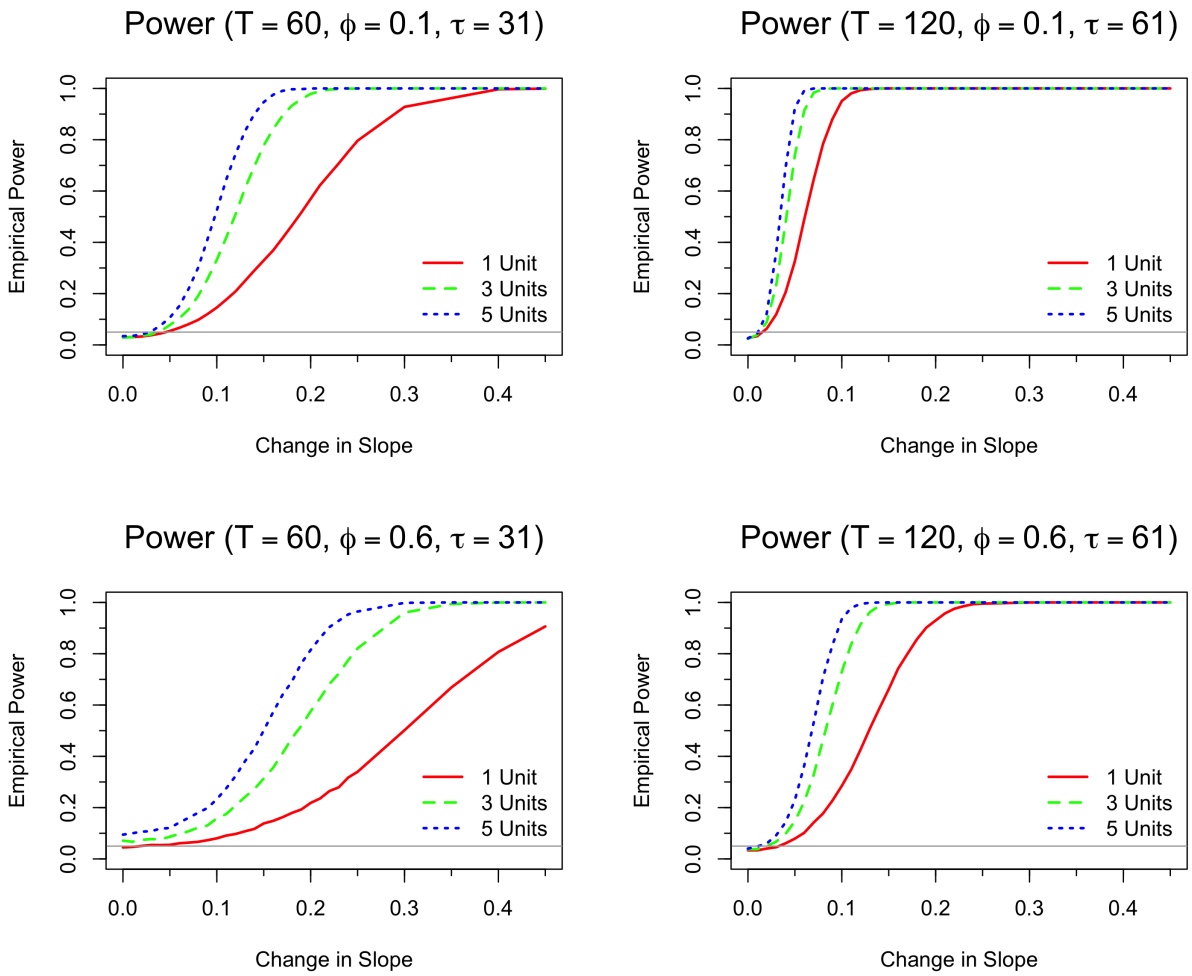

Figure 4.3: Empirical power, over 10,000 iterations, for various number of units and for 4 regimes. The empirical power increases as the number of units and the length of time series increases, and the power increases as the adjacent correlation decreases.

proportion of simulations that correctly estimate the true change point as a function of the slope change for one, three and five units. Similar to the empirical power, the proportion of correctly estimated change points increases as the number of units and the length of the time series increases. We also calculated the proportion of simulations that exactly estimate the true change point for change points not in the middle of the time series — i.e., with a change point on the boundary or near the boundary — and obtained comparable results.

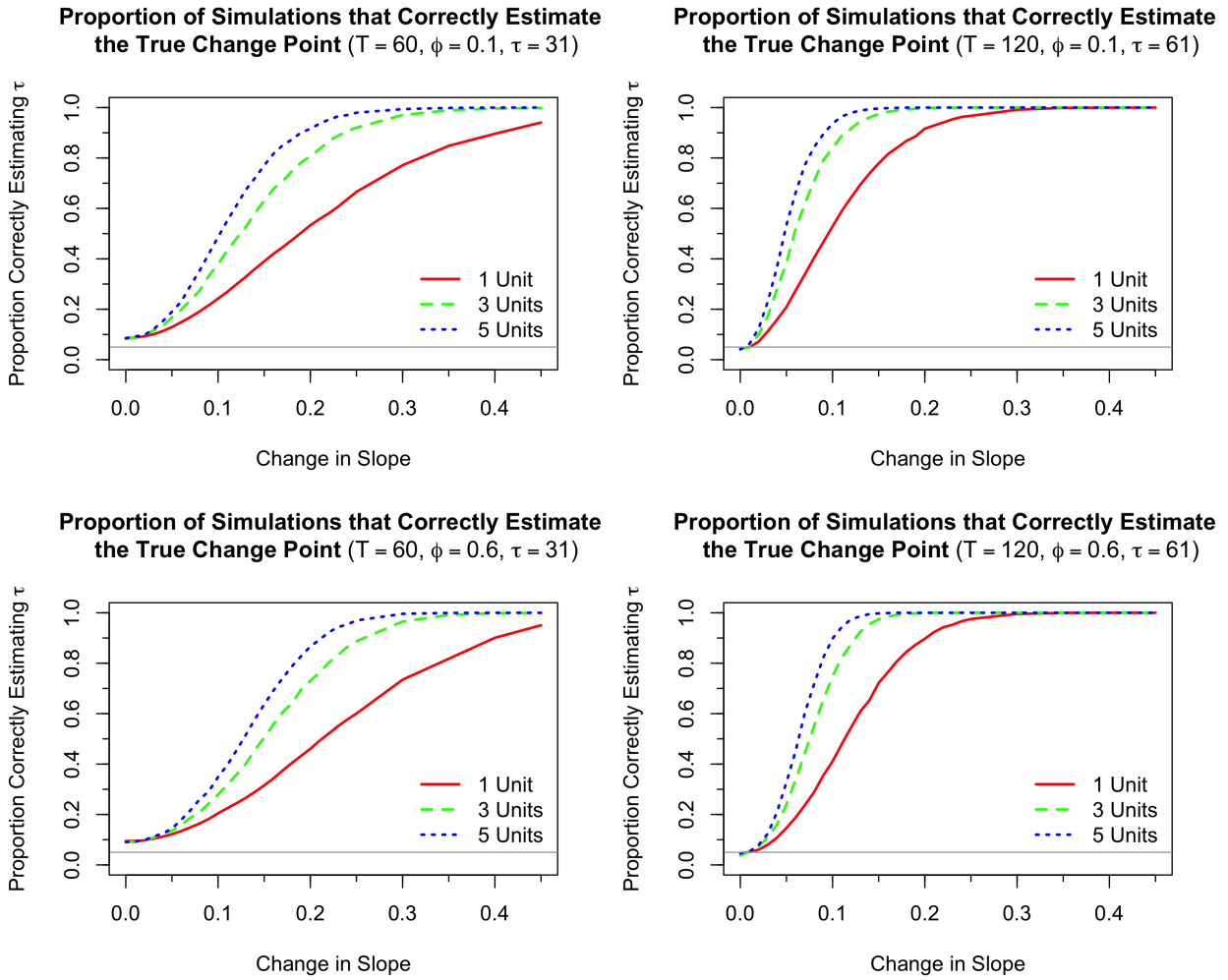

Figure 4.4: The proportion of estimated change points exactly equal to the true change point, over 10,000 iterations, for various number of units and for 4 regimes. Similar to the empirical power, the proportion of correctly estimated change points increases as the number of units and the length of time series increases.

4.3 Multi-Unit Analysis of the CNL Intervention

We assess the impact of the Clinical Nurse Leader (CNL) integrated care delivery intervention on average patient satisfaction at five hospital units. Average patient satisfaction is the mean of patient satisfaction survey scores for seven indicators, shown for the Stroke and Surgical units in Figure 4.1. The seven patient satisfaction indicators are: effective nurse communication, nurses treated me with courtesy/ respect, responsiveness of hospital staff,

effective physician communication, staff did everything to help control your pain, effective communication about medicines, and discharge information provided. We refer to the average patient satisfaction scores simply as patient satisfaction.

We are interested in estimating the time lag (or delay) between the onset of the intervention and the effect on patient satisfaction. In practice, the change point may occur either *before* or *after* the official intervention time. An intervention intended to improve care quality requiring a training over several months or weeks (such as the CNL intervention) may already produce a change in the outcome, even before the official start of intervention, if the trainees execute their training as they learn.

Inference on the global change point and time lag between the onset of the intervention and the intervention's effect is included in Table 4.2. Table 4.2 provides the (a.) global change point estimate (b.) supremum Wald test p-value, (c.) time point of Clinical Nurse Leader integration into their respective hospital units, (d.) formal intervention implementation time, and (e.) lag between formal intervention implementation and estimated change point. The supremum Wald test concludes that a change point exists over the set of possible change points for patient satisfaction in at least one of the units at the $\alpha = 0.05$ level. The p-value associated with the test for the existence of a change point is 0.003, which is less than the respective Benjamini-Hochberg corrected critical value of 0.005. R-MITS estimates a preemptive effect of the CNL integrated care delivery intervention on patient satisfaction. The global change point is estimated to occur on May 2010, while the formal intervention implementation occurs on July 2010. Estimating an anticipatory effect (from the expected and *a priori* specified change point) is not feasible with standard segmented time series regression. Segmented Regression methodology requires clearly separated pre- and post-intervention phases, often with an assumed change point greater than or equal to the formal intervention implementation time point.

CNL Introduction*	Estimated Change Point $\widehat{\tau}$	Formal Intervention Implementation	Lag**
January 2010	May 2010	July 2010	-2

Table 4.2: Provides the estimated global change point, its p-value, the month Clinical Nurse Leaders were integrated into their respective units, the formal intervention time point, and the intervention lag. We conclude that there is a change point in patient satisfaction at the $\alpha = 0.05$ level. $^{+}$ The p-value for the supremum Wald test, i.e., the p-value for the existence of a change point. * All clinical nurse leaders were integrated into their respective hospital units on January 2010. ** The intervention lag is the difference between the estimated change point and the formal intervention introduction time point.

Although the CNL integrated care delivery is officially implemented on July 2010, it was unofficially being practiced prior to July 2010. Nurses put into practice the new concepts they learned from their 'training'. It is completely realistic that many of the CNLs implemented their training prior to July 2010, particularly, if they believed it would be beneficial. Thus, the anticipatory effect of the CNL integrated care delivery intervention (of two 2 months, provided in the 'Lag' column of Table 4.2) is consistent with the integration of the CNLs on January 2010. In fact, the estimated global change point for patient satisfaction occurs four months after the Clinical Nurse Leaders introduction into their respective units. The CNL care delivery intervention requires a restructuring of patient care and care delivery, likely to manifest itself to patients after a time lag from the CNLs introduction. This time lag and the behavioral component of the intervention may explain why the global change point occurs four months after the CNLs integration into the hospital units and two months prior to the formal intervention time point.

Estimates of the R-MITS mean function parameters are provided in Tables 4.3 and 4.4, and estimates of the stochastic process parameters are included in Table 4.5. Estimates and 95% confidence intervals of the two standardized effect sizes used in the health care literature, change in level and change in trend/slope, (Effective Practice and Organisation of Care, 2015) are provided in table 4.4. The level and trend change are not statistically significant

for any unit. The estimated level change tends to be positive for the majority of hospital units, indicating an initial drop of the outcome level, as in Figure 4.2. This may be due to the adjustment period associated with the intervention. Moreover, it may occur as an artifact of the regression itself, particularly for a bounded outcome such as patient satisfaction.

Hospital Unit	Intercept Pre-Change Point $\widehat{\beta_{i0}^{\widehat{\tau}}}$			Slope Pre-Change Point $\widehat{\beta_{i1}^{\widehat{\tau}}}$		
	Estimate	95% CI	p-val	Estimate	95% CI	p-val
Stroke	64.32	(46.34, 82.31)	0	0.56	(-0.52, 1.64)	0.3
Surgical	72.8	(47.72, 97.88)	0	0.36	(-1.05, 1.77)	0.61
Cardiac	64.17	(37.08, 91.27)	0	0.31	(-1.3, 1.92)	0.7
Medical Surgical	70.19	(41.77, 98.61)	0	0.19	(-1.53, 1.91)	0.83
Mother Baby	77.1	(63.1, 91.09)	0	0.28	(-0.58, 1.15)	0.52

Table 4.3: The unit specific pre-change point intercepts and slopes.

Hospital Unit	Change in Level $-\widehat{\delta_i^{\widehat{\tau}}} - \widehat{\Delta_i^{\widehat{\tau}}}\widehat{\tau}$			Change in Slope $\widehat{\Delta_i^{\widehat{\tau}}}$		
	Estimate	95% CI	p-val	Estimate	95% CI	p-val
Stroke	6.91	(-14.65, 28.46)	0.52	-0.35	(-1.59, 0.89)	0.58
Surgical	6.17	(-20.22, 32.56)	0.64	-0.21	(-1.87, 1.45)	0.8
Cardiac	-0.15	(-34.36, 34.06)	0.99	-0.22	(-2.25, 1.82)	0.83
Medical Surgical	0.3	(-40.57, 41.18)	0.99	-0.14	(-2.53, 2.24)	0.9
Mother Baby	3.73	(-22.1, 29.56)	0.77	-0.25	(-1.72, 1.23)	0.74

Table 4.4: The unit specific change in levels and change in slopes.

Trend (slope) change is negative for patient satisfaction, suggesting a decrease in the slope of patient satisfaction post-change point. Due to the nature of patient satisfaction as a percentage — and thus as a bounded outcome — the change in slope must be interpreted with caution. Patient satisfaction cannot continue to grow at a rapid rate because the mean patient satisfaction function at the estimated change point is already relatively close to 100, the maximum patient satisfaction value. This is evident in Figure 4.5, in which the estimated mean functions for all hospital units are plotted, particularly for the Stroke, Surgical, and Mother Baby units.

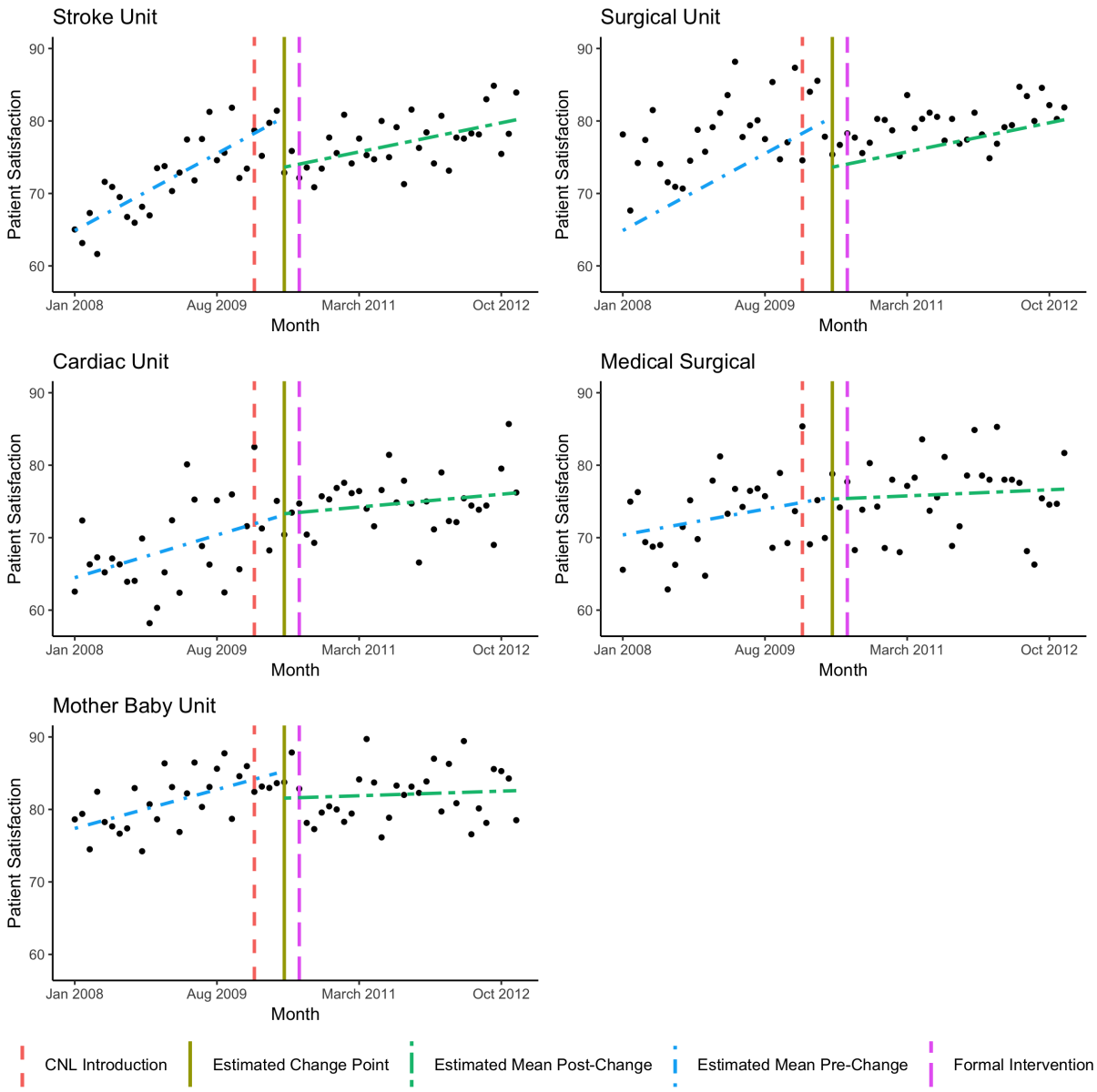

Figure 4.5: Plots the time series of observed average patient satisfaction for all hospital units, along with the estimated change point, estimated mean functions, and formal intervention time.

The estimated volatility of patient satisfaction, given by the 'Standard Deviation' column of Table 4.5, is smaller post-change point for 3 out of the 5 units, and the adjacent correlation is larger post-intervention in 4 out of the 5 units. The Medical Surgical and Mother Baby units estimated standard deviations increase post-estimated change point, increasing from 4.84 and 2.97 to 5.06 and 3.72, respectively; while in the Stroke, Surgical, and Cardiac

units the estimated standard deviation decreases from 3.15, 4.37, and 5.28 to 3.01, 2.35, and 3.76, respectively. After the estimated change point, the patient satisfaction scores are observed to be less volatile for the Stroke, Surgical, and Cardiac units, and hence may be more predictable. The adjacent correlation estimates mainly move from negative to positive post-estimated change point, indicating a more stationary patient satisfaction score post-intervention. These are positive results of the CNL intervention. It is important for hospitals to have patients that are generally satisfied over patients who range from extremely satisfied to extremely dissatisfied. Patient satisfaction scores that are more dependent, closely related, and less volatile result in a more predictable outcome.

Hospital Unit	Pre-Change Point		Post-Change Point	
	Adjacent Correlation $\widehat{\phi}_{j1}(\widehat{\tau})$	Standard Deviation $\widehat{\sigma}_{j1}$	Adjacent Correlation $\widehat{\phi}_{j2}(\widehat{\tau})$	Standard Deviation $\widehat{\sigma}_{j2}$
Stroke	-0.06	3.15	-0.35	3.01
Surgical	-0.02	4.37	0.19	2.35
Cardiac	-0.16	5.28	0.10	3.76
Medical Surgical	-0.03	4.84	0.09	5.06
Mother Baby	-0.27	2.97	0.08	3.72

Table 4.5: Estimates of the stochastic component parameters: the adjacent correlations and response standard deviations pre- and post-change points. All the adjacent correlations are relatively small, and tend to switch from negative to positive post-intervention. The response standard deviations tend to decrease post-intervention.

4.3.1 Doubly Robust ITS

R-MITS pools information across units to estimate a global change point, thereby increasing efficiency and reducing the impact of misleading influential points. Reducing the effect of influential points is desirable in our patient satisfaction data, for which the change point search space consists of only a few time points. We illustrate the gravity of influential points on the estimated change point for the single unit analyses of patient satisfaction at

the Medical Surgery and Cardiac units. To model patient satisfaction for a single unit we implement the Robust-ITS model. The estimated mean functions and change point estimates are included in Figure 4.6 for two cases. The plots on the left of Figure 4.6 correspond to single unit analyses including all observations, while the plots on the right pertain to the single unit analyses without observation $t = 25$ (January 2010). When all the observations are included, Robust-ITS estimates the change point to be February 2010 for both the Medical Surgical and Cardiac units. However, for the analyses without January 2010, the estimated change points are October 2010 and April 2010 for the Medical Surgical and Cardiac units, respectively. One single time point has the ability to perturb the estimated change point by six months in the Medical Surgical unit and by two months in the Cardiac unit. Our proposed R-MITS model guards against these influential points by borrowing information across hospital units. Pooling data across hospital units in the estimation of a global change point automatically reduces the impact of spurious influential points, resulting in robust mean function estimates.

4.4 Summary and Conclusions

Our proposed R-MITS model is appropriate for multiple time series, able to estimate a global change point rather than assume it *a priori*, and can model differences in both the mean functions and stochastic components. R-MITS borrows information across units to estimate a global change point and to estimate the mean functions and stochastic processes separately for each unit. The proposed model does not assume that the impact of the global change point on the outcome is equivalent for all units. That is, although R-MITS borrows information across units to estimate an over-all-unit change point, the level change and trend change are allowed to vary for each unit. R-MITS further allows the autocorrelation and variability during pre- and during post-intervention to differ across units.

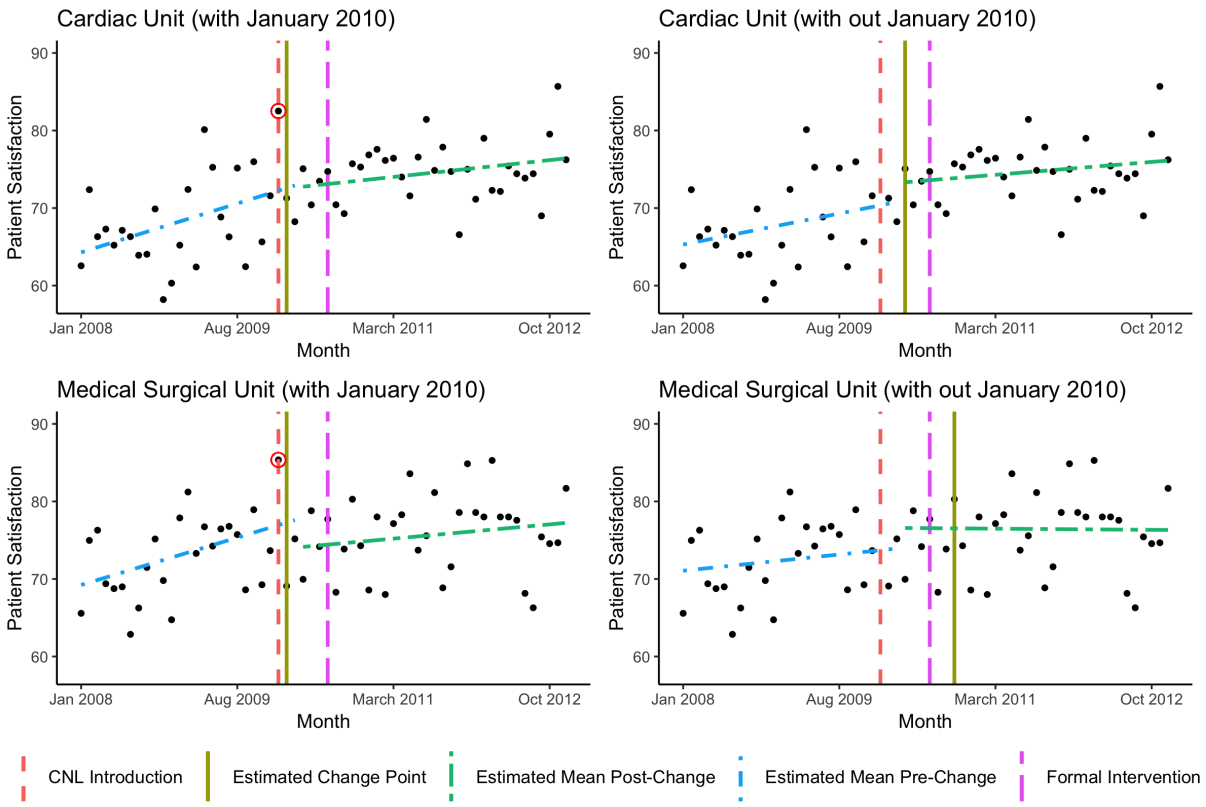

Figure 4.6: Plots the time series of observed average patient satisfaction, along with the estimated change point, estimated mean functions, and formal intervention time for the Medical Surgery and cardiac units with and without observation $t = 25$ (January 2010), obtained by using Robust-ITS to conduct the unit specific analyses. *Note, the analysis with* $t = 25$ *is on the left and the analysis without* $t = 25$ *is on the right.*

Importantly, our proposed supremum Wald test formally tests for the *existence* of a change point in at least one of the mean functions, rather than merely assuming and requiring a change. Now researchers will be able to formally test whether an intervention is associated with a change in the mean functions of a health outcome. Erroneous inference regarding the response's mean functions may result from incorrectly assuming — both the existence and placement — of the change point. Assuming a change point when no change point truly exists forces the estimation of an artificial change. Our supremum Wald test will test for the existence of a change in the response over a pre-determined set of possible change points. As demonstrated by our simulation studies the operating characteristics of R-MITS and the supremum Wald test are well behaved with regards to power and type one error.

Moreover, the empirical power of the supremum Wald test and accuracy of the change point estimates — and so the accuracy of the estimated time delay between an intervention and the intervention's effect on an outcome — increase as the number of units increases.

The R-MITS model and the supremum Wald test provide researchers with insight to re-address hypothesis generation for future study design. The methodology better informs researchers of the likely lag that may be realistic for a similar intervention. We note that in our application example nurses finishing their masters thesis project (in a program that trained them to implement the CNL intervention) were introduced into their respective hospital units six months prior to the formal intervention. The nurses integration potentially changed practice as soon as they were introduced. In fact, the estimated change point occurs between the introduction of the nurses to the hospital unit and the formal intervention. A primary utility of R-MITS is that through exploration of the change point we are able to observe this and provide direction for future study planning.

Currently the supremum Wald test focuses on changes solely in the mean functions. In the next chapter we implement a supremum Wald test that accurately detects changes in both the mean functions and stochastic components, to better handle the nuances of the autocorrelation structures across units. It is paramount to note that the current status of the R-MITS model is for continuous-valued outcomes only. We expand this class of models to discrete ITS (e.g., infection rates, counts of accidental falls, etc) in the ensuing chapter.

Chapter 5

A Generalized ITS Model for Discrete Outcomes

The Centers for Medicare & Medicaid Services (CMS) incentivize health care quality reform via a value-based purchasing program for health systems care services reimbursement (Kavanagh et al., 2012). The measures used for reimbursement include mortality and complications, patient safety, and patient experience (Centers for Medicare and Medicaid Services, 2018). Many of these quality of care measures are discrete. Recall, statistical models used to analyze ITS data are primarily based on segmented regression. Though segmented regression is presented as able to model counts, rates and proportions (Wagner et al., 2002), the methodology is not well specified for discrete outcomes where responses are bounded and there may exist dependencies between the response mean and variance. ITS methods for discrete responses remain an area of open research. As health care ITS are often composed of discrete outcome measurements (i.e., patient falls, unretrieved device fragment count, etc.) methodology able to assess the impact of an intervention on these outcome types is needed.

The methodology proposed in this chapter is motivated by our interest in estimating the lagged effect of a care delivery intervention on patient falls, recorded monthly at six clinical care units over a five year period. The intervention, as in previous chapters, was the implementation of Clinical Nurse Leader (CNL) integrated care delivery, a nursing model that embeds a master prepared nurse into the front lines of care six months prior to the formal intervention (Bender et al., 2017). ITS of the log of patient falls are included in Figure 5.1 for two clinical units (the Cardiac and Acute Care units). Our overall aim is to determine if a change in patient falls exists over a predetermined set of possible change points. Then, if a change exists, to estimate the time point at which patient falls exhibits the change and quantify that change in terms of the mean function and temporal dependence.

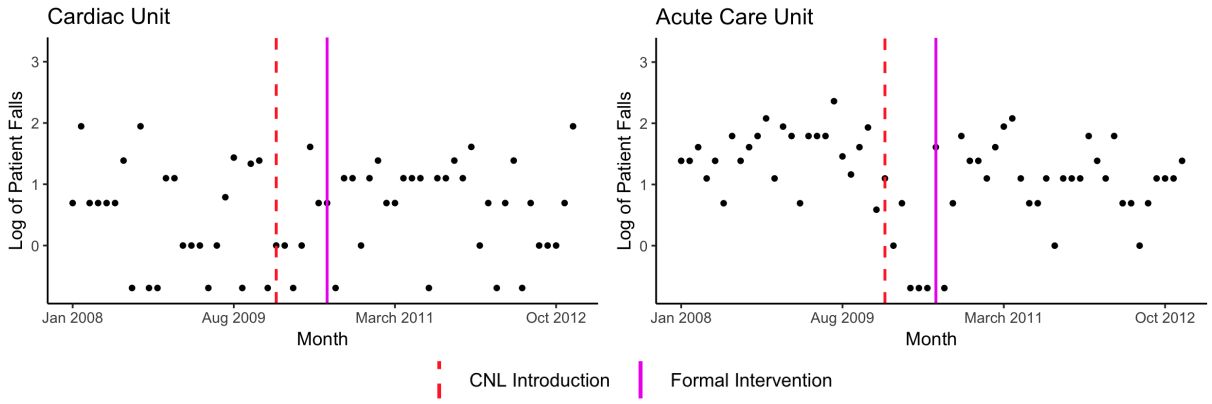

Figure 5.1: Plots the log of the time series of observed patient falls for the Cardiac and Acute Care units. Note, for the purposes of depicting the time series we add 0.5 to patient falls when patient falls is equal to zero, making log of patient falls equal to -0.69 and giving rise to the negative points in the plots.

In the subsequent sections, we develop the 'Generalized Robust ITS' (GRITS) model appropriate for outcomes whose underlying distribution belongs to the family of exponential distributions, thereby expanding the available methodology to adequately model binary and count responses. We describe our proposed GRITS model in detail and provide estimation and inference procedures. Then, we present empirical simulations that assess type one error and power for detecting specified change point alternatives, along with accuracy of our change point estimation procedure. Next, we determine the impact of the CNL integrated care delivery

intervention on patient falls via our GRITS model. To conclude, we summarize our developed model and its impact on the broader ITS literature and describe future work.

5.1 Methodology

We propose the Generalized Robust Interrupted Time Series (GRITS) model to analyze multi-unit discrete ITS. The GRITS model generalizes the Robust Multiple Interrupted Time Series model, proposed in Chapter 4, to handle discrete outcomes. As such, the GRITS model estimates an overall change point across units when appropriate and allows the mean function and correlation structure to differ pre- and post-change point. The change point is defined as the time point at which the underlying pattern of an outcome exhibits a change that may be associated with the intervention of interest. To properly account for the mean-variance relationship of discrete outcomes, we embed the traditional segmented regression approach within a broader generalized estimating equation framework.

The GRITS model makes a distinction between the change point and the formal intervention implementation time point. Denote the change point as τ and the time point at which the intervention is formally introduced as t^*. The change point τ is not necessarily equal to t^*. In fact, often $\tau > t^*$ or $\tau < t^*$ due to a delayed or anticipated intervention effect. GRITS formally tests for the existence of a change point, rather than simply assuming a change, over a predetermined set of possible change points, by implementing the supremum Wald test (Cruz et al., 2019). If the supremum Wald test concludes a change point exists, estimation is carried out via GRITS using a model with a change point at the most likely location. Otherwise, estimation of mean and correlation parameters are estimated without a change point.

5.1.1 The Generalized Robust ITS (GRITS) Model

Let y_{ij} denote the response of interest for unit i and measurement j, with $i \in [1, \ldots, N]$, $j \in [1, \ldots, n_i]$, N the total number of units and n_i the total number of measurements for unit i. We transform the measurement number to the $[0, 1]$ interval by dividing j by n_i. The transformation is done to appropriately model large valued time series. Denote the transformed j^{th} measurement number for unit i by t_{ij}. Define $Y_i = [y_{i1} \ \ldots \ y_{in_i}]'$ as the vector of all response measurements and $\mathbf{X}_i = [x_{i1} \ \ldots \ x_{in_i}]'$ as the design matrix, where $x_{ij} = [1 \ t_{ij} \ 1I(j \geq \tau) \ t_{ij}I(j \geq \tau)]'$, for all i. Let $\vec{\beta}_i(\tau) = [\beta_{i0}^\tau \ \beta_{i1}^\tau \ \delta_i^\tau \ \Delta_i^\tau]'$, the vector of mean function parameters. Observe, $\vec{\beta}_i(\tau)$ and $\mathbf{X_i}$ depend on the unit. Then, denote the conditional expectation of the response given x_{ij} as $\mu_{ij} = \mathrm{E}[y_{ij}|x_{ij}]$ for all i. We suppose

$$g(\mu_{ij}) = \beta_{i0}^\tau + \beta_{i1}^\tau t_{ij} + \left(\delta_i^\tau + \Delta_i^\tau t_{ij}\right) I(j \geq \tau) = \mathbf{X}_i' \vec{\beta}_i(\tau), \tag{5.1}$$

with $g(\cdot)$ a known link function.

Remark: The measures used to quantify the impact of an intervention on a health outcome in the ITS literature are level change and trend change. If $g(\cdot)$ is the identity link, the level change for unit i is defined as the change in anchored intercept (anchored at $t_{i\tau}$) for that unit, denoted by $\delta_i^\tau + \Delta_i^\tau t_{i\tau}$, and the trend (slope) change is denoted by Δ_i^τ.

We model the conditional working variance of y_{ij} given $\mathbf{X}_i$ as $V(\mu_{ij}) = \mathrm{Var}[y_{ij}|\mathbf{X}_i]$ for all i via a quasi-partial likelihood framework. We assume a working correlation structure that follows an auto-regressive process of order one pre- and post-change point conditional on the covariates and τ, i.e.,

$$\mathrm{corr}(y_{i,j-1}, \ y_{ij} | \ \mathbf{X}_i, \ \tau) = \begin{cases} \rho_1(\tau) \ \ j < \tau, \\ \rho_2(\tau) \ \ j \geq \tau, \end{cases} \tag{5.2}$$

with $\rho_1(\tau), \rho_2(\tau) \in (-1, 1)$, for all $i \in \{1, \ldots, N\}$ and $j \in \{2, \ldots, n_i\}$. Other working correlation structures are feasible, but we chose an AR(1) structure based on our data setting. Thus, the conditional working covariance matrix of Y_i given $\mathbf{X}_i$ can be written as

$$\mathbf{V}_i \equiv \mathbf{V}_i\big(\vec{\beta}_i(\tau), \rho_1(\tau), \rho_2(\tau)\big) = \mathbf{S}_i\big(\vec{\beta}_i(\tau)\big)^{\frac{1}{2}} \mathbf{R}\big(\rho_1(\tau), \rho_2(\tau)\big) \mathbf{S}_i\big(\vec{\beta}_i(\tau)\big)^{\frac{1}{2}},$$

with $\mathbf{S}_i\big(\vec{\beta}_i(\tau)\big) = \text{diag}\Big\{V\Big(\mu_{ij}\big(\vec{\beta}_i(\tau)\big)\Big)\Big\}$, and $\mathbf{R}\big(\rho_1(\tau), \rho_2(\tau)\big)$ an AR(1) block diagonal working correlation matrix, provided in the appendix. Then, the quasi-score function for unit i is given by

$$U_i\big(\vec{\beta}_i(\tau), \rho_1(\tau), \rho_2(\tau)\big) = \mathbf{D}_i^{'}\ \mathbf{V}_i^{-1}\ \big(Y_i - \vec{\mu}_i\big), \tag{5.3}$$

where $\mathbf{D}_i$ denotes the matrix of partial derivatives of the mean function vector, $\vec{\mu}_i \equiv [\mu_{i1}\ \ \ldots\ \ \mu_{in_i}]^{'}$, with respect to $\vec{\beta}_i(\tau)$. The change point and mean function parameters are estimated simultaneously by iteratively solving the quasi-score equation (obtained by setting (5.3) equal to zero). The adjacent correlation parameters are estimated via method of moments.

5.1.2 Supremum Wald Test (SWT)

A primary goal of our method is to test for the existence of a change in the outcome of interest over a predetermined set of possible change points. Let $\mathcal{Q} = \{t^* - m, \ldots, t^*, \ldots t^* + k\}$ denote the set of possible change points, where m and k are non-negative integers predetermined by the researchers. Then, we wish to determine whether a change point exists for any $q \in \mathcal{Q}$. To this end, we implement the supremum Wald test (SWT) proposed in Chapter 4, which calculates the multivariate Wald test statistic for every $q \in \mathcal{Q}$, implements the Benjamini-Hochberg method to adjust for the multiple comparisons, and results in a binary decision of

whether a change point exists or not (Cruz et al., 2019). Specifically, for each $q \in \mathcal{Q}$ we test:

$$H_0 : \delta_i^q = \Delta_i^q = 0 \ \ \forall\ i \ \text{ and } \ \rho_1(q) = \rho_2(q) \quad \text{(no change point)}$$
$$\text{vs. } \ H_a : \ \delta_i^q \neq 0 \text{ and/or } \Delta_i^q \neq 0 \ \text{ for some } i \text{ and } \rho_1(q) \neq \rho_2(q) \quad \text{(a change point at } q\text{)},$$

The alternative hypothesis, H_a, assumes a change point in the mean function for at least one of the units and a change in the correlation structure at q. The multivariate Wald statistic is therefore

$$W = \sum_{i=1}^{N} \Big(\mathbf{C}\widehat{\vec{\beta}}_i(q)\Big)' \Big[\mathbf{C}\widehat{\mathbf{V}}_i\big(\widehat{\vec{\beta}}_i(q)\big)\mathbf{C}'\Big]^{-1}\Big(\mathbf{C}\widehat{\vec{\beta}}_i(q)\Big) \overset{H_0}{\sim} \chi^2_{(2N+1)}, \tag{5.4}$$

$$\text{where} \quad \mathbf{C} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{V}}_i\big(\widehat{\vec{\beta}}_i(q)\big) = \big[\mathbf{D}_i'\mathbf{V}_i\mathbf{D}_i\big]^{-1}.$$

Note, $\mathbf{C}$ is a contrast matrix and $\widehat{\mathbf{V}}_i\big(\widehat{\vec{\beta}}_i(q)\big)$ is the estimated covariance matrix of $\widehat{\vec{\beta}}_i(q)$ under the alternative, assuming the working correlation structure is correctly modeled. An empirical sandwich estimator may be used to estimate $\widehat{\mathbf{V}}_i\big(\widehat{\vec{\beta}}_i(q)\big)$, but empirical results indicate poor small sample performance.

The version of the SWT discussed detects whether a change point exists in both the mean functions and correlation structure. It may be of interest to detect a change point solely in the mean functions. In this case, the Multivariate Wald test statistic can be altered accordingly, as in Chapter 4, by calculating $\widehat{\mathbf{V}}_i\big(\widehat{\vec{\beta}}_i(q)\big)$ assuming one overall correlation structure.

5.1.3 Parameter Estimation

Post-test parameter estimation depends on the conclusion of the SWT. If the SWT concludes that no change point exist, then GRITS assumes the mean function parameters and the

adjacent correlation are the same pre- and post-intervention. That is, GRITS assumes

$$g(\mu_{ij}^0) = \beta_{i0} + \beta_{i1} t_{ij} \tag{5.5}$$

for all j, all i, some link function $g(\cdot)$, and a working correlation matrix that follows an AR(1) structure, provided in the appendix. To estimate β_{i0}, β_{i1} and ρ, we implement Algorithm 2, an interative Newton-Raphson algorithm that utilizes generalized estimating equations (GEEs) to obtain estimates of the mean function parameters and the adjacent correlation. Otherwise, if the SWT rejects the null hypothesis of no change point, GRITS is expressed as described in section 5.1.1 and the mean function parameters and adjacent correlations are estimated for each $q \in \mathcal{Q}$. That is, for each $q \in \mathcal{Q}$ we implement Algorithm 2.

Algorithm 2 Estimating mean function and correlation parameters iteratively for all i

set $\zeta = 1$
set $\widehat{\vec{\beta}}_i^{0} = [0.1 \ \ldots \ 0.1]'$
while $\zeta >$ tol **do**
    set k to the iteration
    set $\widehat{\vec{\mu}}_i = \exp\{(\widehat{\vec{\beta}}_i^{k-1})^T \mathbf{X}_i\}$
    obtain Pearson residuals
    from the Pearson residuals calculate adjacent correlation(s)
    obtain $\mathbf{D}_i$ and $\mathbf{V}_i$
    calculate $\mathcal{I}_{n_i}^{-1} = \left[\mathbf{D}_i' \mathbf{V}_i \mathbf{D}_i\right]^{-1}$
    set $\widehat{\vec{\beta}}_i^{k} = \widehat{\vec{\beta}}_i^{k-1} + \mathcal{I}_{n_i}^{-1} \times U_i(\widehat{\vec{\beta}}_i^{k-1})$
    set ζ to the sum of the Euclidean distances between $\widehat{\vec{\beta}}_i^{k}$ and $\widehat{\vec{\beta}}_i^{k-1}$
end while
obtain estimated covariances of $\widehat{\vec{\beta}}_i$

The variance-covariance matrix of Y_i conditional on $\mathbf{X}_i$, $\mathbf{V}_i$, is completely specified by the mean function parameters and the adjacent correlations. The estimator of the mean parameters is provided in Algorithm 2. Estimators of the adjacent correlations are provided in the appendices in section C.2.

Next, we obtain an estimate of the change point by minimizing the quasi-likelihood information criterion under the independence model (QIC) (Pan, 2001). As an alternative, one could maximize the partial likelihood (Cruz et al., 2017, 2019) or the independence quasi-likelihood. We choose to maintain a 'likelihood free' estimation procedure, and thus, minimize the QIC. For each $q \in \mathcal{Q}$, define QIC as

$$\text{QIC}(\mathbf{R}; q) = -2\sum_{i=1}^{N} Q(\widehat{\vec{\beta}}_i(q); I, \mathbf{D}_i) \; + \; 2\sum_{i=1}^{N} \text{trace}\Big(\mathbf{\Omega}_{I,i}\widehat{\mathbf{V}}_i\big(\widehat{\vec{\beta}}_i(q)\big)\Big), \tag{5.6}$$

with $Q(\widehat{\vec{\beta}}_i(q); I, \mathbf{D}_i)$ as the quasi-likelihood and $\mathbf{\Omega}_{I,i}$ as the observed fisher information under the independence working correlation structure for unit i. Then, the estimated change point is:

$$\widehat{\tau} = \arg\min_{q \in Q} \; \text{QIC}(\mathbf{R}; q). \tag{5.7}$$

Estimates of the mean function parameters and the adjacent correlations are obtained based on Algorithm 2 conditional on the estimated change point, $\widehat{\tau}$.

5.2 Empirical Studies

We go on to study the operating characteristics of our proposed methodology. Particularly, we examine the type one error rate of the SWT, power to detect specified change point alternatives for the SWT, and accuracy of our proposed change point estimation procedure. As in Section 5.1.2, we test:

$$H_0 : \; \delta_i^q = \Delta_i^q = 0 \;\; \forall \; i \;\; \text{and} \;\; \rho_1(q) = \rho_2(q) \quad \text{(no change point)}$$

$$\text{vs.} \;\; H_a : \; \delta_i^q \neq 0 \; \text{and/or} \; \Delta_i^q \neq 0 \;\; \text{for some } i \text{ and } \rho_1(q) \neq \rho_2(q) \quad \text{(a change point at } q\text{)},$$

for each $q \in \mathcal{Q}$. The appendix provides empirical studies for the case when the alternative hypothesis assumes a change point solely in the mean functions.

We set simulated data particularities to values based on our patient falls data and generate correlated count ITS via the `GenOrd` package in R (Barbiero and Ferrari, 2015). We assume the canonical link function for a Poisson distribution, $g(\cdot) = \log(\cdot)$, and the same mean function for all units in all simulated data settings. Importantly, we considered four different values of the total number of units, $N \in \{1, 3, 5, 10\}$, in order to compare the gains in efficiency obtained by borrowing information across various units.

5.2.1 Empirical Type One Error of the SWT

To examine the type one error rate of the SWT we generated $10,000$ correlated count ITS of length $n \in \{60, 120\}$ under the null hypothesis of no change point for three values of the adjacent correlation, $\rho \in \{0.1, 0.2, 0.4\}$. We assumed $\beta_{i0} = 2$ and $\beta_{i1} = -0.2$ for all i. When $n = 60$ the set of possible change points was set to $\{25, 26, \ldots, 34\}$ and when $n = 120$ the set of possible changes points was $\{50, 51, \ldots, 69\}$. We compared two values of n to illustrate the impact of doubling the time series length on the type one error rate. With regard to the adjacent correlation values, 0.1 is a hypothesized upper bound for ρ in our patient falls data and 0.4 is a large value in the literature for count data. Type one error rates for the six scenarios are included in Table 5.1. For the cases when $N = 1$ or $\rho = 0.4$ and $n = 60$, the empirical type one error rates are large. This is likely due to a small effective sample size. For all other scenarios the type one error rates are relatively well behaved, albeit better behaved as the number of units and the length of the time series increase.

Empirical Type One Error Rate								
	$n = 60$				$n = 120$			
ρ	1 Unit	3 Units	5 Units	10 Units	1 Unit	3 Units	5 Units	10 Units
0.1	0.0730	0.0533	0.0507	0.0472	0.0450	0.0410	0.0358	0.0364
0.2	0.0788	0.0588	0.0535	0.0510	0.0473	0.0330	0.0399	0.0388
0.4	0.0859	0.0596	0.0576	0.0543	0.0567	0.0440	0.0391	0.0418

Table 5.1: Type one error rates for the SWT testing the existence of a change point in the mean function and correlation structure.

5.2.2 Empirical Power of the SWT

We generated $10,000$ correlated count ITS of length $n \in \{60, 120\}$ under the alternative hypothesis of a change point in the mean function and correlation structure. The change point was placed in the middle of the time series, at time point 31 if $n = 60$ and at 61 if $n = 120$, and the set of possible change points was assumed to be $\mathcal{Q}_{60} = \{25, 26, \ldots, 34\}$ and $\mathcal{Q}_{120} = \{50, 51, \ldots, 69\}$. As in the previous section, we considered $N \in \{1, 3, 5, 10\}$ to illustrate the gains in power obtained by borrowing information across units. We set the adjacent correlation to $\big(\rho_1(\tau)), \rho_2(\tau)\big) \in \{(0.1, 0.2), (0.2, 0.3), (0.4, 0.5)\}$, and assumed $\beta_{i0}^{\tau} = 2$, $\beta_{i1}^{\tau} = -0.2$ and $\delta_i^{\tau} = 0$ for all i. Considering dissertation time constraints and hypothesizing consistent simulation results, we omitted the case when $n = 60$ and $\big(\rho_1(\tau)), \rho_2(\tau)\big) = (0.4, 0.5)$.

For brevity, we examine power as a function of the change in slope, provided in Figure 5.2, though we expect similar results for power as a function of the change in intercept. We note that empirical power decreases as the adjacent correlations increase and increases as the length of the time series increases, as expected. Additionally, empirical power increases as the number of units increase. Therefore, there is a significant gain in power obtained by borrowing information across units and a lesser yet substantial gain in power as the length of the time series increases.

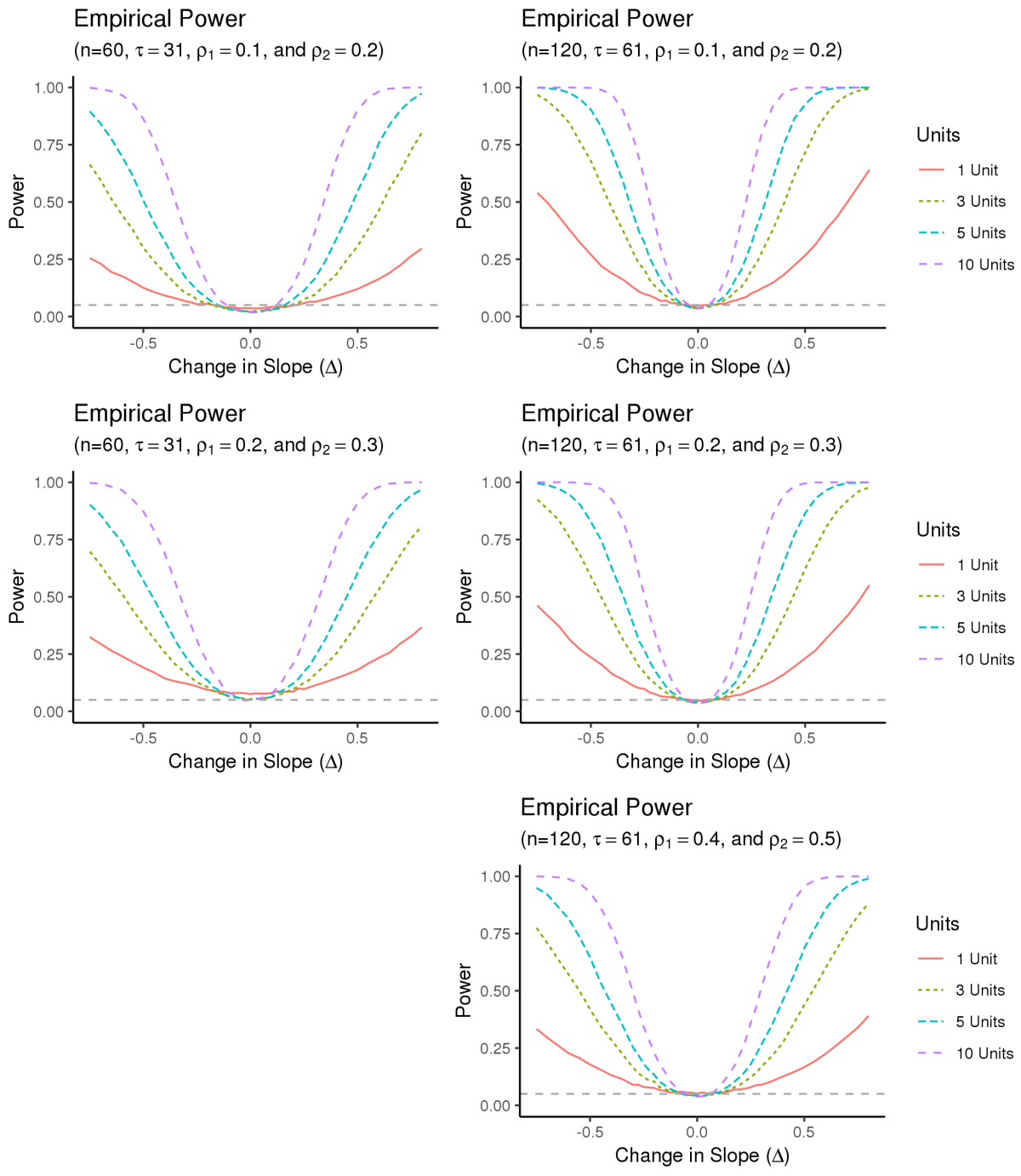

Figure 5.2: Plots empirical power of the SWT as a function of the change in slope for $n = 60$ in the first column and $n = 120$ in the second column. The values of the change in slope ranged between -0.8 and 0.8.

Accuracy of Change Point Estimation Procedure

In addition to power, we are interested in the ability of our change point estimation procedure to correctly estimate the true change point when the SWT concludes a change point does indeed exists. The proportion of simulations that correctly estimate the change point within one unit of the truth when the SWT concludes a change point exists, are included in Figure 5.3 for all scenarios considered. We again omitted the case when $n = 60$ and $\big(\rho_1(\tau)), \rho_2(\tau)\big) = (0.4, 0.5)$ due to dissertation time constraints and expected congruous simulation results. Similar to the empirical power results, accuracy of our change point estimation procedure increases as the adjacent correlation decreases and as the number of units increases. Thus, a gain in accuracy occurs when information is borrowed across units.

We note that as the length of the time series increases accuracy seems to decrease. This anomaly may be explained by the size of the set of possible change points. In our simulation studies, we double the cardinality of the set of possible change points along with the time series length, thus increasing the change point search space as n increases. The large change point search space may in turn decrease accuracy by increasing the number of plausible change points.

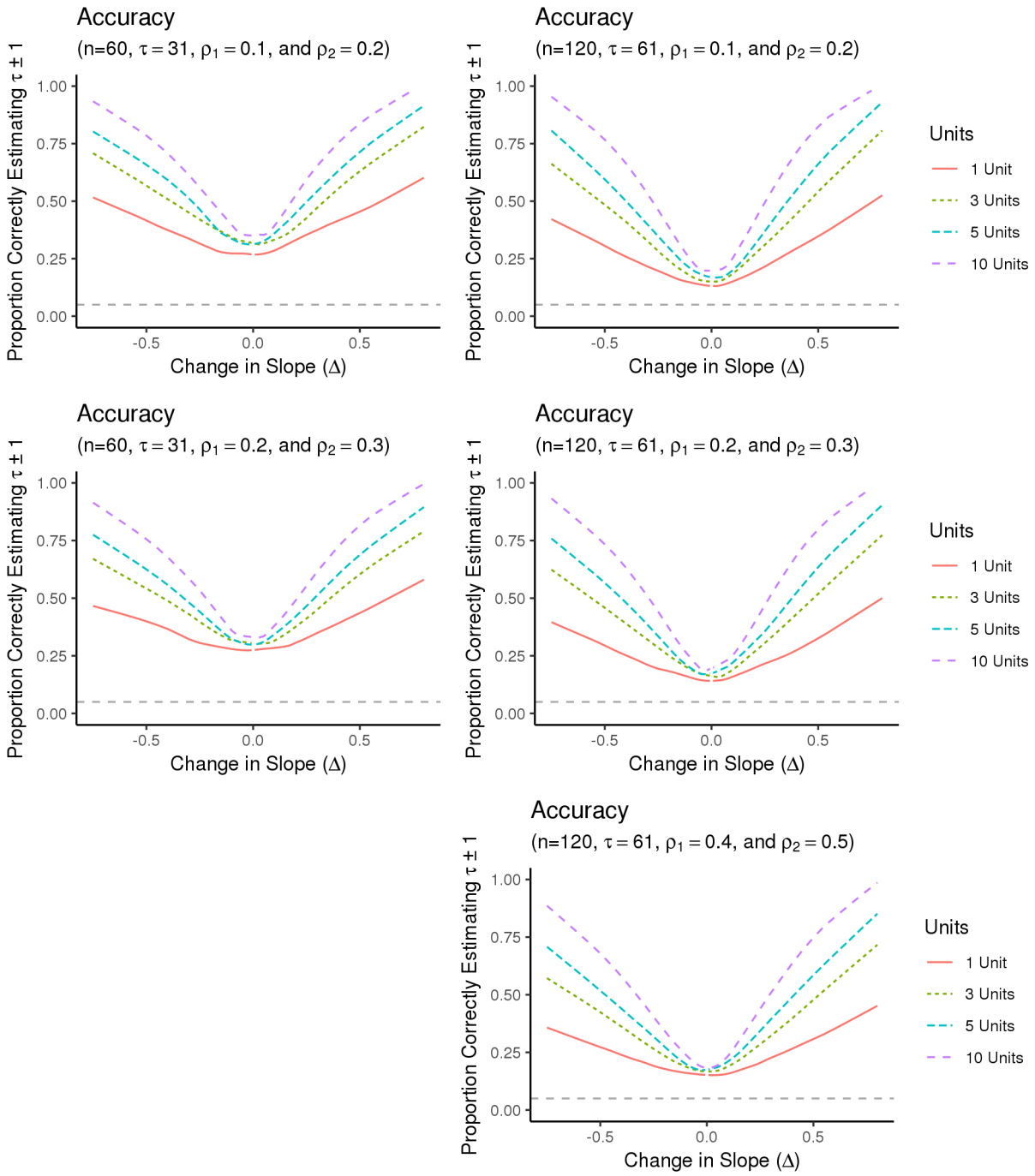

Figure 5.3: The first column plots accuracy of our change point estimation procedure as a function of the change in slope for $n = 60$ and the second column for $n = 120$. The values of the change in slope ranged between -0.8 and 0.8. Note, accuracy is defined as the proportion of simulations that estimate the change point to be within one time point of the true change point after rejecting the null hypothesis that a change point does not exist via SWT. For $\Delta^{\tau} = 0$ (the model with no change point), we did not calculate change point accuracy.

5.3 Analysis of the CNL Intervention

We assess the impact of the CNL integrated care delivery model on patient falls in six clinical care units; boxplots provided in Figure C.3 of section C.4 in the appendix. Our primary goal was to determine whether the CNL intervention was associated with a change in patient falls. To that end, our proposed GRITS model tests for the existence of a change point in patient falls between the nurses introduction into their respective hospital units (January 2010) and three months after the formal intervention (October 2010). GRITS concludes, based on the SWT, that there is a change point in patient falls between January 2010 and October 2010 for the clinical care units at the $\alpha = 0.05$ level. Thus, we model patient falls with a change point.

We are now interested in determining the time lag between the onset of the intervention and the intervention's effect. GRITS estimates a preemptive CNL intervention effect on patient falls. The estimated change point occurs one month after the nurses introduction into their respective hospital units, on February 2010. This indicates that the nurses were implementing their CNL training prior to the formal intervention and is critical knowledge with regards to future study planning.

As patient falls are count data, we assumed the canonical link for a Poisson distribution $g(\cdot) = \log(\cdot)$ throughout our modelling procedure and supposed $\mathbf{S}_i = \text{diag}(\vec{\mu}_i)^{\frac{1}{2}}$ for all i in our working correlation matrix. We, thus, discuss mean function parameters in terms of rates. Table 5.2 and Table 5.3 provide exponentiated estimates, 95% confidence intervals and p-values for the intercept and slope pre-estimated change point and for the level change, trend change and slope post-estimated change point, respectively. The 95% confidence intervals are obtained via a normal approximation with variance estimated by the inverse of the observed Fisher information. An empirical sandwich estimator may be used to estimate the variance, but, as stated in Section 5.1.2, empirical results indicate poor small sample performance.

	Intercept Pre-Change Point exp($\widehat{\beta}_{i0}^{\widehat{\tau}}$) 95% CI	Slope Pre-Change Point exp($\widehat{\beta}_{i1}^{\widehat{\tau}} \frac{6}{n_i}$) 95% CI
Stroke	1.42 (0.83, 2.42)	12.94 (1.9, 88.35)**
Surgical	2.53 (1.63, 3.94)***	2.67 (0.49, 14.62)
Acute Care	4.44 (3.15, 6.27)***	1.53 (0.39, 5.99)***
Pulmonary	6.65 (5.15, 8.59)***	7.27 (2.83, 18.68)
Cardiac	3.11 (1.95, 4.97)***	0.16 (0.02, 1.27)
Medical Surgical	2.49 (1.5, 4.13)***	0.29 (0.03, 2.55)

Table 5.2: Provides estimates of the exponentiated intercept and slope pre-estimated change point, as well as corresponding 95% confidence intervals and p-values. The pre-estimated change point slope is given (scaled) in terms of six month comparisons, thus the inclusion of the $\frac{6}{n}$ term. ** p-value $< .01$ *** p-value $< .001$

	Level Change exp($\widehat{\delta}_i^{\widehat{\tau}} + \widehat{\Delta}_i^{\widehat{\tau}} t_{i\widehat{\tau}}$) 95% CI	Trend Change exp($\widehat{\Delta}_i^{\widehat{\tau}}$) 95% CI	Slope Post-Change Point exp($[\widehat{\beta}_{i1}^{\widehat{\tau}} + \widehat{\Delta}_i^{\widehat{\tau}}] \frac{6}{n_i}$) 95% CI
Stroke	1.49 (0.87, 2.56)	0.19 (0.02, 1.68)	2.52 (0.94, 6.74)
Surgical	1.62 (0.92, 2.84)	0.5 (0.06, 3.97)	1.33 (0.41, 4.35)
Acute Care	2.14 (1.29, 3.54)**	1.26 (0.22, 7.28)	1.94 (0.64, 5.82)
Pulmonary	2.08 (1.5, 2.88)***	0.04 (0.01, 0.13)***	0.27 (0.12, 0.63)**
Cardiac	0.69 (0.34, 1.4)	9.39 (0.81, 108.52)	1.47 (0.41, 5.24)
Medical Surgical	1.48 (0.64, 3.4)	7.09 (0.43, 116.5)	2.02 (0.35, 11.6)

Table 5.3: Provides estimates of the exponentiated level change, trend change and slope post-estimated change point, as well as corresponding 95% confidence intervals and p-values. The pre- and post-estimated change point slopes are given (scaled) in terms of six month comparisons, thus the inclusion of the $\frac{6}{n}$ term. ** p-value $< .01$ *** p-value $< .001$

We estimate that the rate of patient falls at the beginning of the observational period (January 2008) ranges between 1.42 to 6.65 per 1000 patient days per month for the six clinical units. For the Stroke, Surgical and Pulmonary units, we estimate that the rate of patient falls comparing time points six months apart is larger in the pre-estimated change point phase than in the post-estimated change point phase. The Stroke and Pulmonary units' estimated rates of patient falls comparing time points six months apart are respectively 12.98 and 7.27 in the pre-intervention phase and are statistically significant, suggesting a decrease in the rate of patient falls post-intervention. For the Pulmonary unit, the estimated rate of patient falls comparing time points six months apart changes from 7.27 pre-intervention

to 0.27 post-intervention. In both phases, the estimated rate of patient falls is statistically significant, suggesting that the rate of patient falls increases pre-estimated change point and decreases post-estimated change point. Figure 5.4 plots the time series for the six clinical care units along with the unit-specific estimated mean functions.

Recall, the measures used in the ITS literature are level change and trend change. The level change in this setting is defined as $\exp(\delta_i^\tau + \Delta_i^\tau t_{i\tau})$ for unit i, which is a rate ratio. Quantifying the slope change in an informative manner in this setting is difficult, i.e., $\exp(\Delta_i^\tau)$ does not translate to a tangible quantity. Nevertheless, we include an estimate of $\exp(\Delta_i^\tau)$ in Table 5.3. With regards to level change: the estimated rate of patient falls is between 1.48 and 2.14 times higher at the estimated change point comparing the post-intervention mean function to the projected pre-intervention mean function for five out of the six units, indicating an immediate increase in the rate of patient falls. This increase in five units may be attributed to the disruption of the underlying care processes in the clinical units. In the Cardiac unit, we estimate that the rate of patient falls is approximately 31% that of the projected pre-intervention mean function at the change point.

GRITS estimates that the adjacent correlation prior to the estimated change point is $\widehat{\rho}_1(\widehat{\tau}) = -0.090$ [95% CI $(-0.317, 0.146)$] and $\widehat{\rho}_1(\widehat{\tau}) = -0.035$ [95% CI $(-0.278, 0.212)$] post-estimated change point. The estimated adjacent correlations are small and relatively close to zero, suggesting minimal temporal dependency in both phases. This is consistent with the manner in which patient falls in collected; different patients are likely sampled every month.

5.4 Summary and Conclusions

Health care ITS are often composed of non-continuous outcome measurements; i.e., many health care outcomes of interest are binary, counts, or rates (e.g., nurse turnover, number

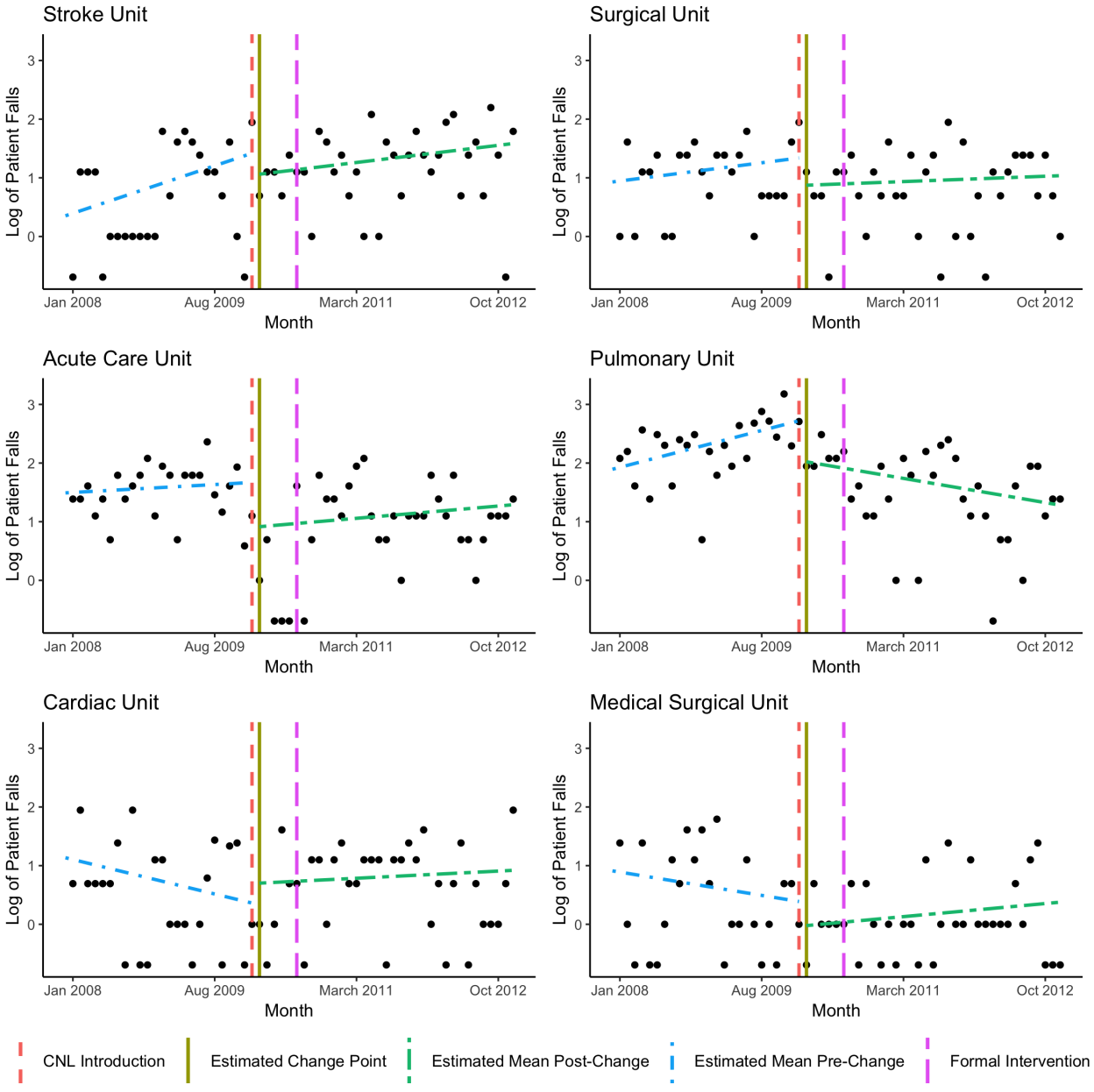

Figure 5.4: Provides estimated mean functions for the log of patient falls for the six clinical care units. Note, for the purposes of depicting the time series we add 0.5 to patient falls when patient falls is equal to zero, making log of patient falls equal to -0.69 and giving rise to the negative points in the plots. We did not use the log of patient falls in the estimation procedure, and so, this jittering does not affect model parameter estimates.

of beds, and patient falls). While segmented regression is presented as being able to model these outcomes, to the best of our knowledge, no formal statistical process is provided. The GRITS model was developed to address this deficiency in the literature. In addition to expanding the current class of ITS models to handle discrete outcomes, GRITS is able to formally test for the existence of and estimate the change point, borrow information across units in multi-unit settings, and test for differences in the mean function and correlation pre- and post-intervention.

Since GRITS exists in the ITS framework, it cannot attribute an observed effect to the intervention of interest, as is the case with most ITS designs (Bernal et al., 2018), but it can inform future study development and measures. Through our GRITS analysis of the CNL intervention we were able to discern that nurses in the six clinical care units may have *a priori* implemented their CNL training; future CNL studies may be altered to reflect this finding. As noted in the analysis section, interpreting the level change and trend change when $g(\cdot)$ is not the identity link may be difficult depending on the choice of $g(\cdot)$. Measures able to discern the intervention effects more clearly are needed for discrete ITS in the GRITS (and other non-simple linear segmented regression methods) to properly capture existing relationships.

The current state of GRITS and R-MITS (proposed in Chapter 4) assume the same change point for all units. In the future, we plan on expanding the current form of these models to account for heterogeneity of change points. This will allow researchers to fully quantify unit specific deviations from the population of hospital units.

Chapter 6

Discussion and Future Work

We developed methodology for modelling ITS data that overcome many of the limitations of segmented regression, the statistical tool of choice for health care ITS data (Penfold and Zhang, 2013; Zhang et al., 2011; Wagner et al., 2002). Specifically, the methods described in this dissertation are able to (1.) test for the existence of and estimate the change point; (2.) borrow information across units in multi-unit settings; (3.) test for differences in the mean function, correlation, and variability pre- and post-intervention; and (4.) handle continuous responses, counts, rates and binary outcomes. We have expanded the class of statistical models in the ITS literature to settings, that to the best of our knowledge, were not formally considered before.

In Chapter 3, we introduced the Robust-ITS model, a single unit model for continuous outcomes. Robust-ITS estimates the change point, rather than assuming it *a priori*, and allows for changes in the mean function and correlation structure pre- and post-intervention. Researchers using Robust-ITS can estimate the lagged effect of a health care intervention on an outcome. Moreover, researchers can determine whether outcomes are more predictable — stronger temporal dependence and smaller variability suggests a more predictable, and thus more desirable outcome — post-intervention.

We proposed R-MITS (Robust Multiple ITS), a multi-unit generalization of Robust-ITS, in Chapter 4. R-MITS borrows information across hospital units to increase efficiency, estimates a global change point, and allows for unit specific changes in the mean function, temporal dependence, and variability of an outcome. Alongside R-MITS we developed the 'supremum Wald test', able to formally test for the existence of a change point across unit specific mean functions. The main contributions of the work proposed in Chapter 4 are the empirical type one error, power, and accuracy studies illustrating the gain in efficiency (while controlling the accuracy and type one error rate) obtained by borrowing information across units. Researchers can use this methodology and the accompanying empirical studies to plan for future intervention assessment.

Lastly, in Chapter 5, we introduced the GRITS (Generalized Robust ITS) model appropriate for outcomes whose underlying distribution belongs to the family of exponential distributions, thereby expanding the available methodology to adequately model binary outcomes, counts and rates. GRITS may be used to model discrete single- and multi-unit ITS. In addition, we implemented a generalized version of the SWT that is able to test for the existence of a change point in the mean functions and correlation structure for discrete outcomes. Researchers using GRITS can adequately model many of the discrete outcomes used by the Centers for Medicare and Medicaid Services for reimbursement purposes, by taking into account the proper mean-variance relationship of the outcome. As with R-MITS, in Chapter 5, we provided empirical type one error and power studies for the GRITS model that can be used to plan for future intervention designs.

The three models described in this dissertation estimate a global over-all-units change point via a grid search over a pre-determined set of possible change points. Researchers must specify the set of possible change points with care since, as with traditional ITS designs, we must be cautious of competing intervention effects. The set of possible change points must adequately capture the time points during which the intervention of interest

plausibly impacted the outcome, yet simultaneously exclude time periods affected by another intervention. This is to avoid the risk of competing interventions. Parsing out the effect of competing interventions is a concern in general with ITS designs. Ideally, the entire observational period (both the pre- and post-intervention phases) of an ITS design should be solely affected by the intervention of interest. Although theoretically simple, in practice this requires careful consideration and expertise.

Identification of a change point via our proposed procedures relies upon detection of a difference in the mean level, the slope and/or the adjacent correlation of the response, comparing the pre- and post-intervention periods across units. As such, if no change point in the time-series truly exists this would indicate that there is no difference in the mean function and/or the correlation structure of the response over time. Most researchers would consider this absence of a difference in the mean function to be the absence of an intervention effect. One could argue that if the pre-intervention slope were positive (indicating improvement in outcomes) and if the slope remained constant during the post intervention phase, then this could have been solely attributable to the intervention. In this case the counterfactual may have revealed a decline (or an increase) in the slope if the intervention had not been instituted. Of course, such a counterfactual could never be observed in practice but certainly should be considered in theory.

Robust-ITS, R-MITS and GRITS all assume a global change point, and as such, do not provide inference on the change point of the overall population of hospital units and the unit-specific deviations. The methodology proposed in this dissertation does not account for heterogeneity of change points across units for situations where the data warrants such treatment. In the future, we will develop ITS mixed effect models as alternatives to these methods, able to detect unit specific change points and borrow information across units while allowing for change point heterogeneity. With these models, researches will be able to make

inference for the overall population of hospital units and quantify unit specific deviations from the population trajectories.

Health systems are required by CMS to record various patient centered outcomes at every hospital unit. The current ITS methodology, including the methods described in this dissertation, require each outcome to be modelled separately. In truth, ITS data are often multivariate time series. We therefore, plan to grow the class of ITS models to allow for multiple (homogeneous and mixed) outcomes, and as such, will develop multivariate versions of the models proposed in Chapters 3-5. Additionally, we will develop multivariate mixed effects robust ITS models that allow for homogeneous and mixed outcome ITS. The models we plan on developing, along with those proposed in this dissertation, expand the class of ITS methods to truthfully accommodate the intricacies of health care ITS data under various real-world circumstances.

Bibliography

Aboagye-Sarfo, P., Mai, Q., Sanfilippo, F. M., Preen, D. B., Stewart, L. M., and Fatovich, D. M. (2015). A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in western australia. *Journal of biomedical informatics* **57,** 62–73.

American Hospital Association Commission on Workforce for Hospitals Health Systems and others (2002). In our hands: How hospital leaders can build a thriving workforce. *Chicago: Author* .

Ansari, F., Gray, K., Nathwani, D., Phillips, G., Ogston, S., Ramsay, C., and Davey, P. (2003). Outcomes of an intervention to improve hospital antibiotic prescribing: interrupted time series with segmented regression analysis. *Journal of Antimicrobial Chemotherapy* **52,** 842–848.

Barbiero, A. and Ferrari, P. (2015). Genord: Simulation of discrete random variables with given correlation matrix and marginal distributions. r package version 1.4. 0. 2015.

Bender, M. (2014). The current evidence base for the clinical nurse leader: a narrative review of the literature. *Journal of Professional Nursing* **30,** 110–123.

Bender, M., Connelly, C. D., Glaser, D., and Brown, C. (2012). Clinical nurse leader impact on microsystem care quality. *Nursing Research* **61,** 326–332.

Bender, M., Murphy, E., Thomas, T., Kaminski, J., and Smith, B. (2015). Clinical nurse leader integration into care delivery microsystems: quality and safety outcomes at the unit and organization level. In *Academy Health Annual Research Meeting*.

Bender, M., Williams, M., Su, W., and Hites, L. (2017). Refining and validating a conceptual model of clinical nurse leader integrated care delivery. *Journal of Advanced Nursing* **73,** 448–464.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57,** 289–300.

Bernal, J. L., Cummins, S., and Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology* **46,** 348–355.

Bernal, J. L., Soumerai, S., and Gasparrini, A. (2018). A methodological framework for model selection in interrupted time series studies. *Journal of clinical epidemiology* **103,** 82–91.

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., and Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International journal of epidemiology* **42,** 1187–1195.

Bollerslev, T. (1988). On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* **9,** 121–131.

Cable, G. (2001). Enhancing causal interpretations of quality improvement interventions. *BMJ Quality & Safety* **10,** 179–186.

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., and Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods* **49,** 988–1005.

Centers for Medicare and Medicaid Services (2018). The hospital value-based purchasing (vbp) program. `https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/Hospital-Value-Based-Purchasing.html`. Online; accessed 13 February 2019.

Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., Ogilvie, D., Petticrew, M., Reeves, B., Sutton, M., et al. (2012). Using natural experiments to evaluate population health interventions: new medical research council guidance. *J Epidemiol Community Health* **66,** 1182–1186.

Cruz, M., Bender, M., and Ombao, H. (2017). A robust interrupted time series model for analyzing complex health care intervention data. *Statistics in medicine* **36,** 4660–4676.

Cruz, M., Gillen, D. L., Bender, M., and Ombao, H. (2019). Assessing health care interventions via an interrupted time series model: Study power and design considerations. *Statistics in medicine* .

Datta, J. and Petticrew, M. (2013). Challenges to evaluating complex interventions: a content analysis of published papers. *BMC public health* **13,** 568.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association* **101,** 223–239.

Effective Practice and Organisation of Care (2015). Interrupted time series (its) analyses. *EPOC Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services* .

Eliopoulos, G. M., Shardell, M., Harris, A. D., El-Kamary, S. S., Furuno, J. P., Miller, R. R., and Perencevich, E. N. (2007). Statistical analysis and application of quasi experiments to antimicrobial resistance intervention studies. *Clinical Infectious Diseases* **45,** 901–907.

Fitzpatrick, J. J. (2003). Joint commission on accreditation of health care organizations white paper: health care at the crossroads: strategies for addressing the evolving nursing crisis. *Policy, Politics, & Nursing Practice* **4,** 71–74.

Garey, K., Lai, D., Dao-Tran, T., Gentry, L., Hwang, L., and Davis, B. (2008). Interrupted time series analysis of vancomycin compared to cefuroxime for surgical prophylaxis in patients undergoing cardiac surgery. *Antimicrobial agents and chemotherapy* **52,** 446–451.

Granger, C. W. J. and Newbold, P. (2014). *Forecasting economic time series*. Academic Press.

Handley, M. A., Lyles, C. R., McCulloch, C., and Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health* **39,** 5–25.

Hawe, P. (2015). Lessons from complex interventions to improve health. *Annual review of public health* **36,** 307–323.

Joynt, J. and Kimball, B. (2008). Innovative care delivery models: identifying new models that effectively leverage nurses. *Prinston: The Robert Wood Johnson Foundation* .

Kavanagh, K. T., Cimiotti, J. P., Abusalem, S., and Coty, M.-B. (2012). Moving healthcare quality forward with nursing-sensitive value-based purchasing. *Journal of Nursing Scholarship* **44,** 385–395.

Kirch, C., Muhsal, B., and Ombao, H. (2015). Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association* **110,** 1197–1216.

Knebel, E., Greiner, A. C., et al. (2003). *Health professions education: A bridge to quality*. National Academies Press.

Kohn, L. T., Corrigan, J., Donaldson, M. S., et al. (2000). *To err is human: building a safer health system*, volume 6. National academy press Washington, DC.

Kontopantelis, E., Doran, T., Springate, D. A., Buchan, I., and Reeves, D. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *bmj* **350,** h2750.

Linden, A. (2015). Conducting interrupted time-series analysis for single-and multiple-group comparisons. *Stata J* **15,** 480–500.

Mahamat, A., MacKenzie, F., Brooker, K., Monnet, D., Daures, J., and Gould, I. (2007). Impact of infection control interventions and antibiotic use on hospital mrsa: a multivariate interrupted time-series analysis. *International journal of antimicrobial agents* **30,** 169–176.

Muggeo, V. (2012). Segmented relationships in regression models with breakpoints/changepoints estimation. *CRAN-R 0.2* **9,**.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics* **57,** 120–125.

Parienti, J.-J., Cattoir, V., Thibon, P., Lebouvier, G., Verdon, R., Daubin, C., Du Cheyron, D., Leclercq, R., and Charbonneau, P. (2011). Hospital-wide modification of fluoroquinolone policy and meticillin-resistant staphylococcus aureus rates: a 10-year interrupted time-series analysis. *Journal of Hospital Infection* **78,** 118–122.

Penfold, R. B. and Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. *Academic pediatrics* **13,** S38–S44.

Petticrew, M., Rehfuess, E., Noyes, J., Higgins, J. P., Mayhew, A., Pantoja, T., Shemilt, I., and Sowden, A. (2013). Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. *Journal of clinical epidemiology* **66,** 1230–1243.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the american statistical association* **53,** 873–880.

Ramsay, C. R., Matowe, L., Grilli, R., Grimshaw, J. M., and Thomas, R. E. (2003). Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *International journal of technology assessment in health care* **19,** 613–623.

Rickles, D. (2009). Causality in complex interventions. *Medicine, Health Care and Philosophy* **12,** 77–90.

SAS User Guide (2008). Sas/ets 9.2 users guide.

Shumway, R. H. and Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.

Taljaard, M., McKenzie, J. E., Ramsay, C. R., and Grimshaw, J. M. (2014). The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care. *Implementation Science* **9,** 77.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology* **51,** 309.

Wagner, A. K., Soumerai, S. B., Zhang, F., and Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of clinical pharmacy and therapeutics* **27,** 299–309.

West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., Szapocznik, J., Fishbein, M., Rapkin, B., Clatts, M., et al. (2008). Alternatives to the randomized controlled trial. *American journal of public health* **98,** 1359–1366.

Zhang, F., Wagner, A. K., and Ross-Degnan, D. (2011). Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. *Journal of clinical epidemiology* **64,** 1252–1261.

Appendix A

Robust-ITS

A.1 A Note on the Time Series Length for Robust-ITS

The number of time points required pre- and post-change point (or pre- and post-intervention) depend 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 (Ramsay et al., 2003; Effective Practice and Organisation of Care, 2015).

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 three 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 five 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 five 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- and 5 for the post-change point phases) measurements plus the length of the set of possible change points are required.

The total number of model parameters 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. Power, however, 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 (Bernal et al., 2017). 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 (Bhaskaran et al., 2013).

A.2 Mean Parameter Estimates

The ordinary least squares (OLS) estimates for the mean parameters in θ of Section 3.2.2 are:

$$\widehat{\beta}_0^{\,\widehat{\tau}} = \bar{y}_{1:(\widehat{\tau}-1)} - \widehat{\beta}_1^{\,\widehat{\tau}}\,\frac{\widehat{\tau}}{2}, \tag{A.1}$$

$$\widehat{\beta}_1^{\,\widehat{\tau}} = \frac{\sum_{t=1}^{\widehat{\tau}-1}(t-\frac{\widehat{\tau}}{2})\,y_t}{\sum_{t=1}^{\widehat{\tau}-1}(t-\frac{\widehat{\tau}}{2})\,t}, \tag{A.2}$$

$$\widehat{\delta^{\widehat{\tau}}} = \bar{y}_{\widehat{\tau}:n} - (\widehat{\beta}_1^{\,\widehat{\tau}} + \widehat{\Delta}^{\widehat{\tau}})\,\bar{t}_{\widehat{\tau}:n} - \bar{Y}_{1:(\widehat{\tau}-1)} - \widehat{\beta}_1^{\,\widehat{\tau}}\,\frac{\widehat{\tau}}{2}, \tag{A.3}$$

$$\widehat{\Delta}^{\widehat{\tau}} = \frac{\sum_{t=\widehat{\tau}}^{n}(t-\bar{t}_{\widehat{\tau}:n})\,y_t}{\sum_{t=\widehat{\tau}}^{n}(t-\bar{t}_{\widehat{\tau}:n})\,t} - \frac{\sum_{t=1}^{\widehat{\tau}-1}(t-\frac{\widehat{\tau}}{2})\,y_t}{\sum_{t=1}^{\widehat{\tau}-1}(t-\frac{\widehat{\tau}}{2})\,t}, \tag{A.4}$$

where $\bar{y}_{a:b} = \frac{\sum_{t=a}^{b} y_t}{b-(a-1)}$ and $\bar{t}_{\widehat{\tau}:n} = \frac{\sum_{t=\widehat{\tau}}^{n} t}{n-(\widehat{\tau}-1)}$. The estimates of $\beta_0^{\widehat{\tau}}$ and $\beta_1^{\widehat{\tau}}$ are the same as the OLS estimates obtained by fitting a linear model to the pre-change point phase alone. The estimates of $\delta^{\widehat{\tau}}$ and $\Delta^{\widehat{\tau}}$ 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 σ_1^2 and σ_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 $\widehat{\tau}$ 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

$$\widehat{\phi}(\widehat{\tau}) = \begin{cases} \widehat{\phi_1}(\widehat{\tau}) = \frac{\sum_{t=2}^{\widehat{\tau}-1}(R_t - \bar{R}_{2:(\widehat{\tau}-1)})(R_{t-1} - \bar{R}_{1:(\widehat{\tau}-2)})}{\sum_{t=2}^{\widehat{\tau}-1}(R_t - \bar{R}_{2:(\widehat{\tau}-1)})^2} \\ \\ \widehat{\phi_2}(\widehat{\tau}) = \frac{\sum_{t=\widehat{\tau}+1}^{n}(R_t - \bar{R}_{(\widehat{\tau}+1):n})(R_{t-1} - \bar{R}_{\widehat{\tau}:(n-1)})}{\sum_{t=\widehat{\tau}+1}^{n}(R_t - \bar{R}_{(\widehat{\tau}+1):n})^2} \end{cases} \tag{A.5}$$

$$\widehat{\sigma^2} = \begin{cases} \widehat{\sigma_1^2} = \frac{1}{\widehat{\tau}-1}\sum_{t=2}^{\widehat{\tau}}\left[(W_t - \overline{W}_{2:(\widehat{\tau}-1)}) - \widehat{\phi_1}(\widehat{\tau})(W_{t-1} - \overline{W}_{1:(\widehat{\tau}-2)})\right]^2 \\ \\ \widehat{\sigma_2^2} = \frac{1}{n-\widehat{\tau}-1}\sum_{t=\widehat{\tau}+2}^{n}\left[(W_t - \overline{W}_{(\widehat{\tau}+1):n}) - \widehat{\phi_2}(\widehat{\tau})(W_{t-1} - \overline{W}_{(\widehat{\tau}+1):n})\right]^2, \end{cases} \tag{A.6}$$

where $\bar{R}_{a:b}$ and $\overline{W}_{a:b}$ are the means of the residuals and of the residuals of the residuals, respectively, for time points a through b, and

$$W_t = \begin{cases} R_t - \widehat{\phi_1}(\widehat{\tau})\, R_{t-1}, & 1 < t < \widehat{\tau} \\ R_t - \widehat{\phi_2}(\widehat{\tau})\, R_{t-1}, & \widehat{\tau} < t \leq n. \end{cases}$$

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 $\nu(\tau) \equiv \phi_2(\tau) - \phi_1(\tau)$ equals zero. If $\nu(\tau) = 0$, there is one overarching AR(1) process for all time points, and equation 3.2 reduces to

$$R_t = \phi\, R_{t-1} + e_t, \quad 1 < t \leq n, \tag{A.7}$$

otherwise equation 3.2 holds. We are comparing nested models where equation 3.2 is the full model and equation A.7 is the reduced model, so an F-test is appropriate. The degrees of freedom corresponding to the reduced model is $(n-1)-1$, to account for the lag of the AR process and the parameter in the model, ϕ. Similarly, since the full model corresponds to two separately fit AR(1) processes and two parameters, the degrees of freedom is $(n-2)-2$. Denote the residual sum of squares for the reduced and full models, respectively, as

$$RSS_R = \sum_{t=2}^{n} (R_t - \widehat{\phi} R_{t-1})^2$$
$$RSS_F = \sum_{t=2}^{\widehat{\tau}-1} (R_t - \widehat{\phi}_1(\widehat{\tau}) R_{t-1})^2 + \sum_{t=(\widehat{\tau}+1)}^{n} (R_t - \widehat{\phi}_2(\widehat{\tau}) R_{t-1})^2.$$

Then the F-statistics is

$$F = \frac{\frac{RSS_R - RSS_F}{([n-1]-1)-([n-2]-2)}}{\frac{RSS_F}{(n-1)-1}} = \frac{(RSS_R - RSS_F)/2}{(RSS_F)/(n-2)},$$

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

A.5 Segmented Regression Models

In the health care intervention literature there are two main types of segmented regression models utilized to model the trends. The first is parametrized in the same manner as 3.1, with τ set to the time of intervention — an assumed instantaneous effect — that is, the mean is parametrized as:

$$\mu_t^1 = \begin{cases} \beta_0 + \beta_1 t, & t < t^* \\ (\beta_0 + \delta) + (\beta_1 + \Delta) t, & t \geq t^* \end{cases} \tag{A.8}$$

where t^* denotes the intervention time. For the data described in the Introduction, $t^* = 31$. The second segmented regression model is

$$\mu_t^2 = \begin{cases} \beta_0 + \beta_1 t, & t < t^* \\ (\beta_0 + \psi) + \beta_1 t + \Psi(t - t^* + 1), & t \geq t^*, \end{cases} \tag{A.9}$$

in which the time after intervention implementation is multyplying Ψ; as opposed to simply time, as in equation A.8.

Note, the trends prior to the intervention introduction are the exact same for both equations A.8 and A.9. Post the intervention time, the intercept increase is denoted by δ in equation A.8 and by $\psi - (t^* - 1)\Psi$ in equation A.9, implying $\delta = \psi - (t^* - 1)\Psi$. The change in slopes is denoted by Δ and Ψ in equation A.8 and A.9 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

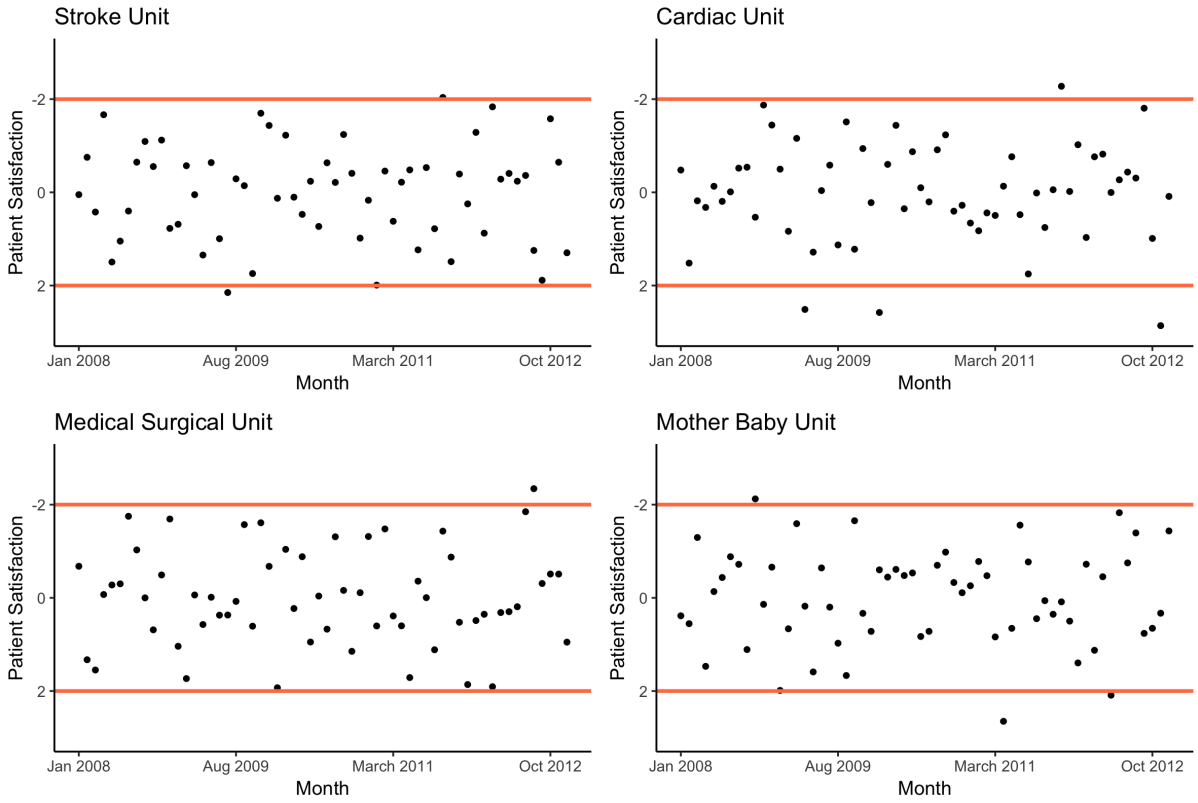

Figure A.1: Plots the studentized residuals of the Robust-ITS estimated patient satisfaction means for each unit. The studentized residuals do not exhibit any clear patterns, and seem to be closely centered around zero, indicating appropriate fits. The rule of thumb, 2, is provided in each plot.

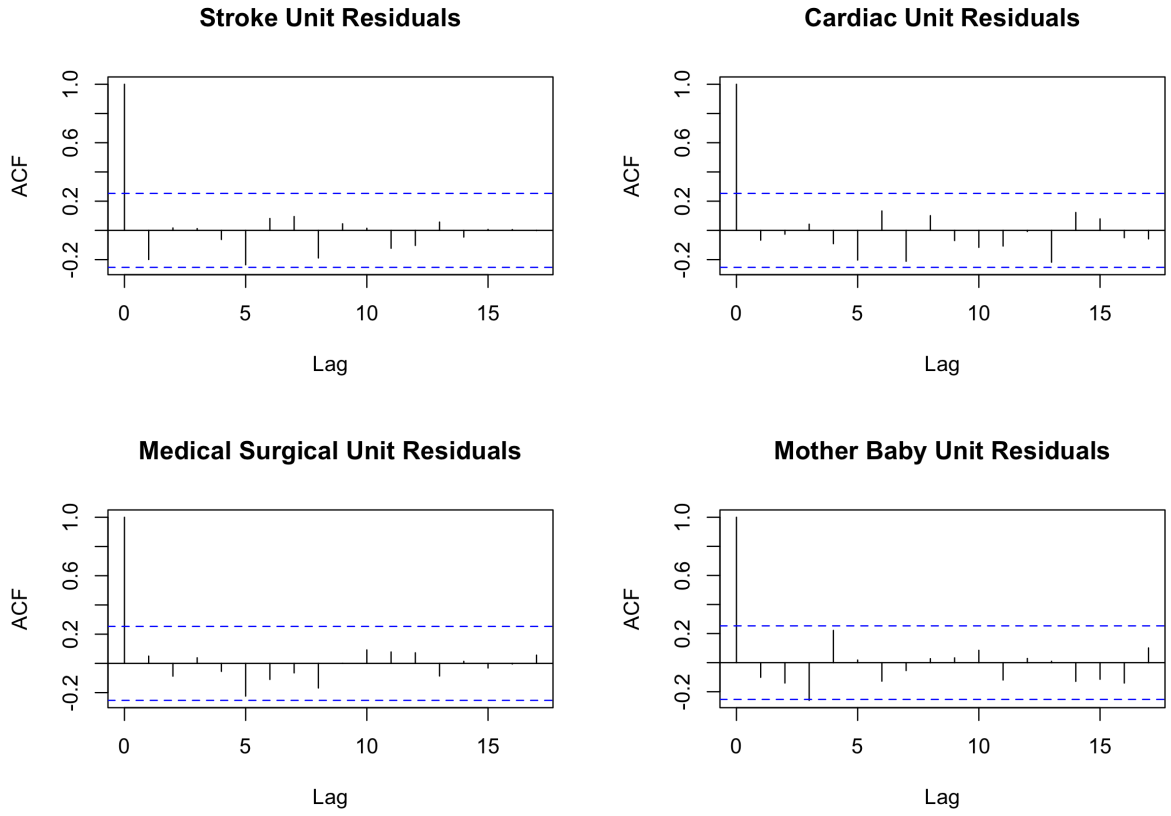

Figure A.2: Provides ACF, autocorrelation function, plots of the Robust-ITS estimated patient satisfaction means for each unit. The ACF plots suggest that the residuals behave as white noise, since the autocorrelation at lags greater than zero are small and seem to get closer to zero as the lag increases.

Appendix B

Robust Multiple ITS

B.1 Estimators of the Mean Function Parameters

The generalized least squares (GLS) estimators for the mean function parameters of $\theta_1, \ldots, \theta_N$ given $q \in Q$, obtained in step (6.) of Algorithm 1 of Section 4.1.1, are

$$\widehat{\theta}_i = \begin{bmatrix} \widehat{\beta}_{i0}^q \\ \widehat{\beta}_{i1}^q \\ \widehat{\delta}_i^q \\ \widehat{\Delta}_i^q \end{bmatrix} = \left[\mathbf{X}_i(q)' \ \widehat{\Sigma}_i^{-1} \ \mathbf{X}_i(q) \right]^{-1} \left[\mathbf{X}_i(q)' \ \widehat{\Sigma}_i^{-1} \ Y_i \right],$$

$$\text{where} \qquad \mathbf{X}_i(q) \equiv \begin{bmatrix} 1 & 2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & q-1 & 0 & 0 \\ 1 & q & 1 & q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & n & 1 & n \end{bmatrix},$$

and $\widehat{\Sigma}_i$ given in the subsequent section. Hence, conditional on the estimated change point, $\widehat{\tau}$ for unit i the estimator of (a.) the intercept pre-change point is $\widehat{\beta}_{i0}^{\widehat{\tau}}$; (b.) the slope pre-change point is $\widehat{\beta}_{i1}^{\widehat{\tau}}$; (c.) the change in level (post-estimated change point intercept anchored at $\widehat{\tau}$) is $\widehat{\delta}_i^{\widehat{\tau}} + \widehat{\Delta}_i^{\widehat{\tau}}\,\widehat{\tau}$; and the change in slope is $\widehat{\Delta}_i^{\widehat{\tau}}$.

B.2 Estimators of the AR(1) Processes Parameters

In steps (4.) and (7.) of the iterative estimation process, provided in Section 4.1.1, the residuals, $r_{it} = y_{it} - \widehat{\mu}_{it}$, are modeled as AR(1) processes:

$$r_{it} = \begin{cases} \phi_{i1}(\tau)\; r_{i,t-1} + e_{it,1}, & 1 < t < q, \\ \phi_{i2}(\tau)\; r_{i,t-1} + e_{it,2}, & q \leq t \leq n, \end{cases}$$

with $e_{it,j} \overset{iid}{\sim} N\big(0, \sigma^2_{iw,j}\big)$ for $j \in \{1, 2\}$. Recall, to ensure causality in the time series sense, $\phi_{i1}(\tau)$ and $\phi_{i2}(\tau)$ must lie in the interval $(-1, 1)$ for all i. The variance-covariance matrix,

$\boldsymbol{\Sigma}_i$, is therefore equal to

$$\begin{bmatrix} \frac{\sigma^2_{iw,1}}{1-\phi_{i1}(\tau)^2} \begin{bmatrix} 1 & \phi_{i1}(\tau) & \dots & \phi_{i1}(\tau)^{q-2} \\ \phi_{i1}(\tau) & 1 & \dots & \phi_{i1}(\tau)^{q-3} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i1}(\tau)^{q-2} & \phi_{i1}(\tau)^{q-2} & \dots & 1 \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2_{iw,2}}{1-\phi_{i2}(\tau)^2} \begin{bmatrix} 1 & \phi_{i2}(\tau) & \dots & \phi_{i2}(\tau)^{n-q} \\ \phi_{i2}(\tau) & 1 & \dots & \phi_{i2}^{n-q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i2}^{n-q} & \phi_{i2}^{n-q-1} & \dots & 1 \end{bmatrix} \end{bmatrix}$$

and completely determined by $\phi_{i1}(\tau)$, $\phi_{i2}(\tau)$, $\sigma_{iw,1}$ and $\sigma_{iw,2}$. We therefore only provide the estimators of $\phi_{i1}(\tau)$, ϕ_{i2}, $\sigma_{iw,1}$ and $\sigma_{iw,2}$, conditional on $q \in Q$.

Define

$$\overline{r}_{(1a)} = \frac{1}{q-2}\sum_{t=1}^{q-2} r_t, \qquad \overline{r}_{(1b)} = \frac{2}{q-1}\sum_{t=2}^{q-1} r_t,$$

$$\overline{r}_{(2a)} = \frac{1}{n-q-1}\sum_{t=q-1}^{n-1} r_t, \qquad \overline{r}_{(2b)} = \frac{1}{n-q-1}\sum_{t=q}^{n} r_t,$$

$$\sigma^2_{r_1} = \frac{\sum_{t=2}^{q-1}(r_t-\overline{r}_{(1b)})^2 + \sum_{t=2}^{q-1}(r_{t-1}-\overline{r}_{(1a)})^2}{2},$$

$$\text{and} \quad \sigma^2_{r_2} = \frac{\sum_{t=q}^{n}(r_t-\overline{r}_{(2b)})^2 + \sum_{t=q}^{n}(r_{t-1}-\overline{r}_{(2a)})^2}{2}.$$

Then the estimators of $\phi_{i1}(\tau)$, $\phi_{i2}(\tau)$, $\sigma_{iw,1}$, $\sigma_{iw,2}$, $\sigma_{i,1}$, and $\sigma_{i,2}$, conditional on q, are

- $\widehat{\phi}_{i1}(\widehat{\tau}) = \dfrac{\sum_{t=2}^{q-1}(r_t - \overline{r}_{(1b)})\,(r_{t-1} - \overline{r}_{(1a)})}{\sigma^2_{r_1}}$
- $\widehat{\phi}_{i2}(\widehat{\tau}) = \dfrac{\sum_{t=q}^{n}(r_t - \overline{r}_{(2b)})\,(r_{t-1} - \overline{r}_{(2a)})}{\sigma^2_{r_2}}$
- $\widehat{\sigma}^2_{iw,1} = \frac{1}{q-2}\sum_{t=2}^{q-1}\left[(r_t - \overline{r}_{(1b)}) - \widehat{\phi}_{i1}(\widehat{\tau})(r_{t-1} - \overline{r}_{(1a)})\right]^2$
- $\widehat{\sigma}^2_{iw,2} = \frac{1}{n-q+1}\sum_{t=q}^{n}\left[(r_t - \overline{r}_{(2b)}) - \widehat{\phi}_{i2}(\widehat{\tau})(r_{t-1} - \overline{r}_{(2a)})\right]^2$
- $\widehat{\sigma}_{i,1} = \dfrac{\widehat{\sigma}_{iw,1}}{\sqrt{1 - \left(\widehat{\phi}_{i1}(\widehat{\tau})\right)^2}}$
- $\widehat{\sigma}_{i,2} = \dfrac{\widehat{\sigma}_{iw,2}}{\sqrt{1 - \left(\widehat{\phi}_{i2}(\widehat{\tau})\right)^2}}$.

B.3 Covariance Matrix of the Full Model Mean Function Parameters for the SWT

The supremum Wald statistic of section 4.1.1 depends on $\widehat{\mathbf{V}}_0(\widehat{\vec{\beta}}^0)$, the block diagonal estimator of the variance covariance matrix of $\widehat{\vec{\beta}}^0$. Each block of $\widehat{\mathbf{V}}_0(\widehat{\vec{\beta}}^0)$ corresponds to $\widehat{\mathbf{V}}(\widehat{\vec{\beta}}^0_i)$, the estimated variance-covariance matrix of the mean function parameters for unit i. Note,

$$\widehat{\mathbf{V}}(\widehat{\vec{\beta}}^0_i) = \left(\mathbf{X}^{1\prime}\,(\widehat{\mathbf{\Sigma}}_i)^{-1}\,\mathbf{X}^1\right)^{-1},$$

with $\mathbf{X}^1_i$ as the design matrix of the **full** model (model of equation (4.6)) and the variance-covariance matrix under the **reduced** model (model of equation (4.5)) as $\widehat{\mathbf{\Sigma}}_i$. Since the aim of the supremum Wald test is to test the existence of a change point in the mean, we

assume an autocorelation structure that remains constant over the entire duration of the observational period. Thus, for unit i

$$\widehat{\mathbf{\Sigma}}_i = \frac{(\widehat{\sigma}_{iw})^2}{1-(\widehat{\phi}_i)^2} \begin{bmatrix} 1 & \widehat{\phi}_i & \dots & \left(\widehat{\phi}_i\right)^{n-2} \\ \widehat{\phi}_i & 1 & \dots & \left(\widehat{\phi}_i\right)^{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\widehat{\phi}_i\right)^{n-2} & \left(\widehat{\phi}_i\right)^{n-3} & \dots & 1 \end{bmatrix},$$

where $\widehat{\phi}_i$ and $(\widehat{\sigma}_{iw})^2$ are estimated under the reduced model.

Appendix C

Generalized Robust ITS

C.1 Working Correlation Matrices for GRITS

The working correlation matrix, $\mathbf{R}$, of section 5.1.1 that assumes a change point in the correlation structure at τ is:

$$\mathbf{R} = \begin{bmatrix} \begin{bmatrix} 1 & \rho_1(\tau) & \dots & \left(\rho_1(\tau)\right)^{\tau-2} \\ \rho_1(\tau) & 1 & \dots & \left(\rho_1(\tau)\right)^{\tau-3} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\rho_1(\tau)\right)^{\tau-2} & \left(\rho_1(\tau)\right)^{\tau-2} & \dots & 1 \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} 1 & \rho_2(\tau) & \dots & \left(\rho_2(\tau)\right)^{n-\tau} \\ \rho_2(\tau) & 1 & \dots & \left(\rho_2(\tau)\right)^{n-\tau-1} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\rho_2(\tau)\right)^{n-\tau} & \left(\rho_2(\tau)\right)^{n-\tau-1} & \dots & 1 \end{bmatrix} \end{bmatrix},$$

with $\rho_1(\tau)$ and $\rho_2(\tau)$ denoting the adjacent correlations in the pre- and post-change point phases, respectively.

When the SWT concludes that no change point exists, the working correlation matrix GRITS assumes is of the form:

$$\begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix},$$

where ρ is the adjacent correlation.

C.2 Estimators of the Adjacent Correlations for GRITS

Let the Pearson residuals be $r_{ij} = \frac{y_{i1} - \widehat{\mu}_{ij}}{\sqrt{V(\widehat{\mu}_{ij})}}$. Under the scenario when the supremum Wald test concludes a change point does not exists, the estimator of the adjacent correlation, ρ is:

$$\widehat{\rho}_i = \frac{\sum_{j=2}^{n} \left[(r_{i,j} - \bar{r}_{2:n})(r_{i,j-1} - \bar{r}_{1:n-1}) \right]}{\sum_{j=1}^{n} \left[(r_{i,j} - \bar{r}_{1:n})^2 \right]}, \qquad \widehat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\rho}_i.$$

Otherwise, when the SWT concludes a change point does exist, the estimators of $\rho_1(\tau)$ and $\rho_2(\tau)$ are:

$$\widehat{\rho}_{1,i}(\tau) = \frac{\sum_{j=2}^{\tau-1} \left[(r_{i,j} - \bar{r}_{2:\tau-1})(r_{i,j-1} - \bar{r}_{1:\tau-2}) \right]}{\sum_{j=1}^{\tau-1} \left[(r_{i,j} - \bar{r}_{1:\tau-1})^2 \right]}, \qquad \widehat{\rho}_1(\tau) = \frac{1}{N} \sum_{i=1}^{N} \widehat{\rho}_{1,i}(\tau),$$

$$\widehat{\rho}_{2,i}(\tau) = \frac{\sum_{j=\tau}^{n} \left[(r_{i,j} - \bar{r}_{\tau:n})(r_{i,j-1} - \bar{r}_{\tau-1:n-1}) \right]}{\sum_{j=\tau}^{n} \left[(r_{i,j} - \bar{r}_{\tau:n})^2 \right]}, \qquad \widehat{\rho}_2(\tau) = \frac{1}{N} \sum_{i=1}^{N} \widehat{\rho}_{2,i}(\tau)$$

C.3 Empirical Studies II

In section 5.2, we provide empirical studies examining the operating characteristics of our proposed methodology when testing whether a change point exists in both the mean functions and correlation structure. It may be of interest to test whether a change point exists only in the mean functions. We, therefore, go on to provide empirical studies for the case when the alternative hypothesis assumes a change point solely in the mean functions. That is, for each $q \in \mathcal{Q}$, we test:

$$\begin{aligned} & H_0 : \delta_i^q = \Delta_i^q = 0 \ \ \forall\ i \quad \text{(no change point)} \\ \text{vs.} \ \ & H_a : \ \delta_i^q \neq 0 \text{ and/or } \Delta_i^q \neq 0 \ \text{ for some } i \quad \text{(a change point at } q\text{)}. \end{aligned}$$

Since the focus is on detecting a change point in the mean functions, we assume one overall AR(1) correlation structure for the entire observational period. We again set simulated data particularities to values based on our patient falls data, generate correlated count ITS via the `GenOrd` package in R (Barbiero and Ferrari, 2015), assume the canonical link function for a Poisson distribution, $g(\cdot) = \log(\cdot)$, and the same mean function for all units. As in section 5.2, we consider four different values of the total number of units, $N \in \{1, 3, 5, 10\}$, to compare the gains in efficiency obtained by borrowing information across various number of units.

C.3.1 Empirical Type One Error of the SWT

To examine the type one error rate of the SWT, we once again generated $10,000$ correlated count ITS of length $n \in \{60, 120\}$ under the null hypothesis of no change point. We considered three values of the adjacent correlation, $\rho \in \{0.1, 0.2, 0.4\}$, and assumed $\beta_{i0} = 2$ and $\beta_{i1} = -0.2$ for all i. When $n = 60$ the set of possible change points was set to

$\{25, 26, \ldots, 34\}$ and when $n = 120$ the set of possible changes points was $\{50, 51, \ldots, 69\}$. Type one error rates for the six scenarios are included in Table C.1. Results are consistent with those obtained in Section 5.2.1: empirical type one error rates are large when $N = 1$ or $\rho = 0.4$ and $n = 60$, and relatively well behaved otherwise.

Empirical Type One Error Rate								
	$n = 60$				$n = 120$			
ρ	1 Unit	3 Units	5 Units	10 Units	1 Unit	3 Units	5 Units	10 Units
0.1	0.0674	0.0481	0.0440	0.0420	0.0426	0.0386	0.0341	0.0341
0.2	0.0705	0.0511	0.0478	0.0450	0.0449	0.0313	0.0363	0.0362
0.4	0.0750	0.0514	0.0504	0.0465	0.0532	0.0431	0.0391	0.0385

Table C.1: Type one error rates for the SWT testing the existence of a change point in the mean functions.

C.3.2 Empirical Power of the SWT

We generated $10,000$ correlated count ITS of length $n \in \{60, 120\}$ under the alternative hypothesis of a change point in the mean function. The change point was again placed in the middle of the time series, at time point 31 if $n = 60$ and at 61 if $n = 120$, and the set of possible change points were assumed to be $\mathcal{Q}_{60} = \{25, 26, \ldots, 34\}$ and $\mathcal{Q}_{120} = \{50, 51, \ldots, 69\}$. We considered $N \in \{1, 3, 5, 10\}$ to illustrate the gains in power obtained by borrowing information across units. We also considered three scenarios for the adjacent correlation, $\big(\rho_1(\tau)), \rho_2(\tau)\big) \in \{(0.1, 0.2), (0.2, 0.3), (0.4, 0.5)\}$, and assumed $\beta_{i0}^{\tau} = 2$, $\beta_{i1}^{\tau} = -0.2$ and $\delta_i^{\tau} = 0$ for all i. Due to dissertation time constraints and simulation results hypothesized to be consistent with those provided in this section, we omitted the case when $n = 60$ and $\big(\rho_1(\tau)), \rho_2(\tau)\big) = (0.4, 0.5)$.

As in Section 5.2.2, we examine power as a function of the change in slope, provided in Figure C.1. Results are consistent with those from Section 5.2.2: empirical power decreases as the adjacent correlations increase and increases as the length of the time series and the number

of units increase. Once again, there is a significant gain in power obtained by borrowing information across units and a lesser yet substantial gain in power as the length of the time series increases.

Accuracy of Change Point Estimation Procedure

We are additionally interested in the ability of our change point estimation procedure to correctly estimate the true change point when the SWT concludes a change point exists. Figure C.2 plots the proportion of simulations that correctly estimate the change point within one unit of the true change point when the SWT concludes a change point does indeed exist, for all scenarios considered. As in Section 5.2.2, accuracy of our change point estimation procedure increases as the adjacent correlation decreases and as the number of units increases, and accuracy decreases as the number of response measurements increases. The latter finding may be explained by the doubling of the cardinality of the set of possible change points when we double the length of the time series. The large search space may in turn decrease accuracy. Nonetheless, a gain in accuracy occurs when information is borrowed across units.

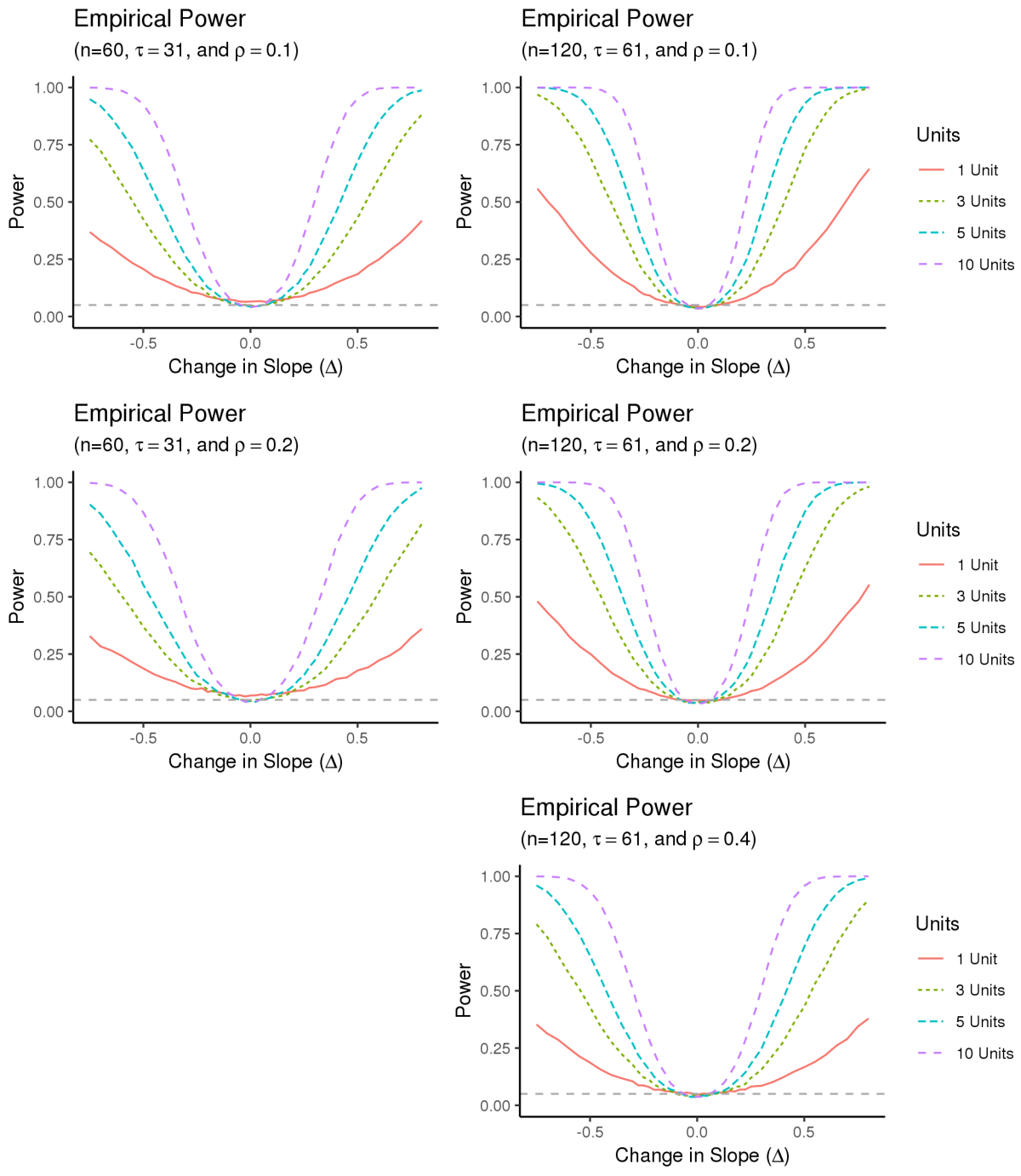

Figure C.1: Plots empirical power of the SWT as a function of the change in slope for $n = 60$ in the first column and $n = 120$ in the second column. The values of the change in slope ranged between -0.8 and 0.8.

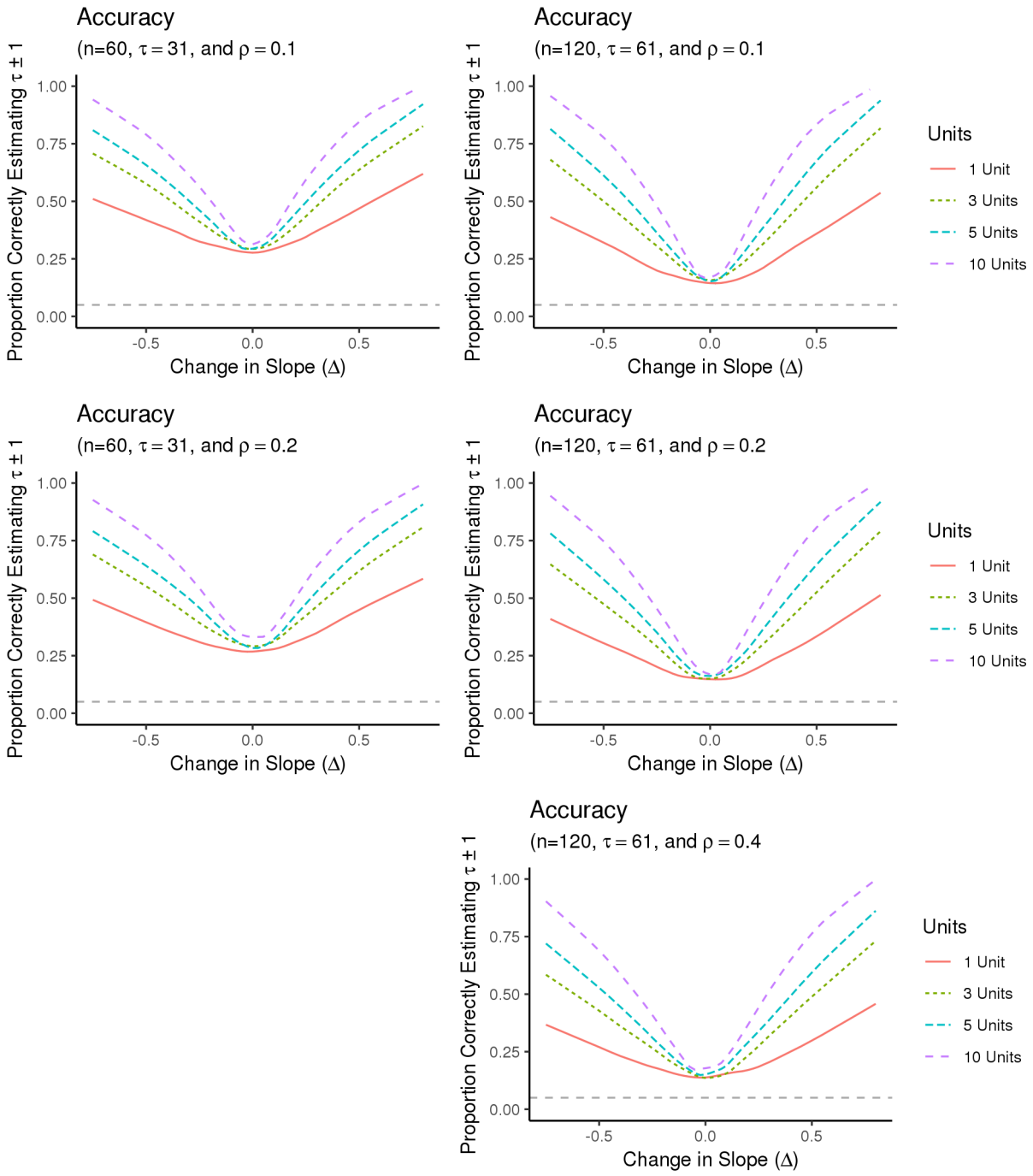

Figure C.2: The first column plots accuracy of our change point estimation procedure as a function of the change in slope for $n = 60$ and the second column for $n = 120$. The values of the change in slope ranged between -0.8 and 0.8. Note, accuracy is defined as the proportion of simulations that estimate the change point to be within one time point of the true change point after rejecting the null hypothesis that a change point does not exist via SWT. For $\Delta^{\tau} = 0$ (the model with no change point), we did not calculate change point accuracy.

C.4 Patient Falls

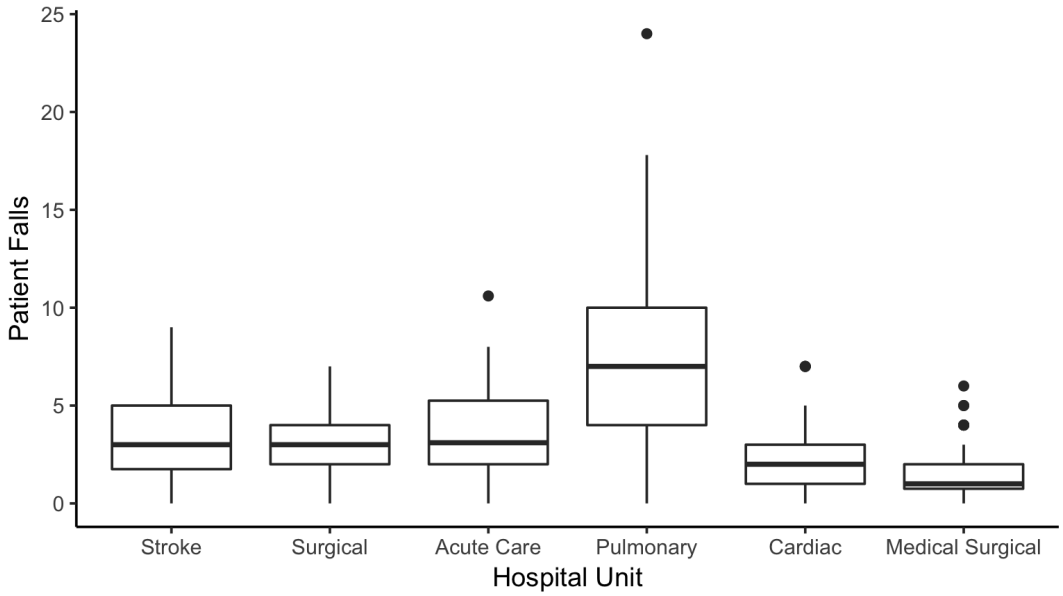

Figure C.3: Provides boxplots for patient falls for all of the clinical care units analyzed in Chapter 5.