## Biostatistics Student Research Symposium (BSRS) 2019

Abstracts



Victoria Garcia	
Alicia Johns	4
Kellen Cressell	5
Jin Liu	6
Katharine Stromberg	7
Paul M. Hargarten	
Viviana Alejandra Rodriguez	9
Christine M. Orndahl	
Spiro Stilianoudakis	11
Jonathan W. Yu	12
Rueben Retnam	13
Keighly Bradbrook	14
Erin Donahue	15
Xinxin Sun	16
Brielle Forsthoffer	17
Matt Carli	
Alicia Richards	19
Salem Rustom	20
Martin Lavallee	21

Title: Overview of a UNOS Internship: Four Years Working with the Transplant Community Presenter: Victoria Garcia Advisor: Dr. Adam Sima Abstract:

United Network for Organ Sharing (UNOS) is the private, non-profit organization that manages the nation's organ transplant system under contract with the federal government. In doing this, UNOS brings together thousands of volunteers and hundreds of organ procurement and transplant professionals - including one graduate student research intern. Research intern responsibilities vary, ranging from attending and contributing to weekly meetings, to learning how to manage various population-level datasets, to consulting with non-UNOS entities interested in transplant research, and to the preparation and submission of manuscripts, conference materials, and, on one occasion, a medical textbook chapter. The apex of the research intern responsibilities, of course, is performing statistical analyses. Two notable projects completed as the research intern were Diagnosing the Decades-Long Rise in the Deceased Donor Kidney Discard Rate and Trends and Geographic Variations in the Recovery of Organs From Drug-overdose Deaths with Donation Potential. The former modeled the increasing trend in kidney discards between 1988-2009 accounting for donor characteristics and clinical decisions using multivariable regression with GEE for same-donor clustering and propensity score analyses. The latter modified the existing process of estimating donation potential from a set of deaths to target drug-overdose (OD) deaths in 2011-2017 on varying geographic scales. Estimated drug-OD donation potential was then compared to the number of actual drug-intoxication donors recovered to identify areas of recovery improvement. This talk will discuss in greater detail the roles of the graduate student research intern and the opportunities therein.

Title: A Comparison of Electronic Medical Record Data and Medicare Claims Data Presenter: Alicia Johns Advisor: Leticia Moczygemba Abstract:

In the past two decades, the use of electronic medical records (EMR) has increased in the clinical setting as well as in the realm of research. Successful EMR implementation can benefit health care organizations by streamlining medical data to improve health outcomes and lower health care costs. Additionally, once practitioners and staff have transitioned to the EMR system, the amount of time spent locating, inputting, and organizing patient information is decreased. While there have been some studies that attempt to compare an EMR system with a claims source, our study includes linked patient data from the Carilion EMR and Medicare claims data. We wish to explore the differences between these two sources and identify areas where the EMR system can be improved. The objectives of the study were to compare the number of hospital and ED visits in an EMR with the number of hospital and ED visits in Medicare claims data in a group of Medicare beneficiaries as well as assess differences between the EMR system and Medicare claims for variables such as date of birth, sex, race, discharge date and discharge diagnosis. According to Stein and colleagues, claims data has the advantage of including "large, diverse sample sizes, longitudinal follow-up, lack of selection bias, and potential for complex, multivariate modeling". While it does provide a glimpse at patient-level diagnoses, details about unique patient visits are not quite as apparent as can be found within EMR data. This study seeks to examine whether data from electronic medical records can be a valid source for researchers to utilize.

**Title:** *TADcompare: an* R *package for differential analysis of Topologically Associated Domains* **Presenter:** Kellen Cressell **Advisor:** Mikhail Dozmorov **Abstract:** 

Recent research in chromatin conformation capture technologies (Hi-C) and its corresponding data have demonstrated the importance of topologically associating domains (TADs). TADs are considered to be relatively stable genomic structures; yet, many dynamically reorganize during development and exhibit celland condition-specific differences. Quantification of the dynamic behavior of TADs will help to better understand genome regulation. However, methods for comparing TADs between cells and conditions are highly limited. We developed 'TADcompare', a method for differential analysis of TAD boundaries between Hi-C datasets. 'TADcompare' is based on a spectral clustering-derived measure called the eigenvector gap, which provides us with a loci-by-loci comparison of TAD differences between datasets. Using this measure, we introduce a range of methods for identifying differential TADs between datasets, aggregating TADs across different conditions and tracking TAD change over time. Based on these results, we devise novel terms for classification of changes across time and between datasets. The results are validated using known biological markers and on simulated datasets demonstrating a functional role of boundary change in the genome. 'TADCompare' is available on https://github.com/cresswellkg/TADCompare. Title: Estimating knots in bilinear spline growth mixture models with time-invariant covariates in the framework of individual measurement occasions Presenter: Jin Liu Advisor: Dr. Robert Perera Abstract:

The linear spline growth mixture model (LSGMM) is a popular tool for examining nonlinear change patterns over time, in which data come from a mixture of two or more sub-populations. It approximates complex patterns by attaching at least two linear trajectories. Besides examining within-person changes and between-person differences of trajectories simultaneously, it poses interesting statistical challenges, such as estimating the location of an inflection point (or knot), the knot's variance, separating of individuals into at least two unobserved classes, examining factors which associated with those latent groups, and analyzing data with individually-varying times of observation. We developed a stepwise bilinear spline growth mixture model (BLSGMM) to cluster these linear piecewise individual trajectories as well as to investigate predictors of the latent classes. Our simulation studies demonstrate that the proposed BLSGMM is capable of clustering the nonlinear change patterns effectively. More importantly, they generally estimated the parameters of interest unbiasedly, precisely, and exhibited appropriate confidence interval coverage.

Title: Adjusting Response Adaptive Allocation for Subject Dropout Presenter: Katharine Stromberg Advisor: Dr. Adam Sima Abstract:

Response adaptive randomization methods that skew allocation towards the superior treatment arm in clinical trials have been shown to maintain power and maximize total expected response. These methods assume patients that withdraw from the study have outcomes similar to those that complete the study. However, if the consequence of withdrawal is the discontinuation from treatment, this assumption may not be valid, leading to allocation that has less than optimal characteristics. The flaws of this method are particularly prevalent if dropout rates differ between the treatment arms. We propose a response adaptive allocation ratio that adjusts for dropout, unrelated to unobserved treatment response, and factors in losses associated with discontinuation of treatment. When there is no dropout, our allocation ratio reduces to previously published response adaptive allocation ratios. We evaluate the features of our novel allocation ratio using traditional clinical trial operating characteristics.

Title: Multivariate Bayesian Multiple Imputation Using Weighted Quantile Sum Regression: A Simulation Study Presenter: Paul M. Hargarten Advisor: David C. Wheeler, PhD. Abstract:

Simultaneous exposure to a mixture of chemicals over a lifetime may increase an individual's risk of disease to a greater extent than individual exposures. Researchers have used weighted quantile sum (WQS) regression to estimate the effect of multiple exposures in a manner that identifies the important (etiologically relevant) components in the mixture. However, complications arise when an experimental apparatus detects concentrations for each chemical with a different detection limit. To account for the uncertainty in imputing below the detection limit (BDL) values when estimating health effects for exposure to mixtures of chemicals, we have integrated WQS regression into the multiple imputation framework (MI-WQS). So far, imputation models in MI-WQS examine each chemical separately and ignore the correlation between chemicals, which may result in biased estimates. In response, we introduce a multivariate Bayesian imputation method into the MI-WQS framework to jointly impute the chemical mixture by taking full advantage of the chemical mixture data. We compare this approach to several other imputation methods through a simulation study to examine the effect of various BDL percentages in a chemical mixture analysis using WQS. We demonstrate that using imputation models that are more similar to the WQS model leads to less biased and more efficient WQS estimates after repeatedly filling-in BDL values.

Title: Two-part mixed effects mixture model for zero-inflated longitudinal compositional data. Presenter: Viviana Alejandra Rodriguez Advisor: Dr. Nitai D. Mukhopadhyay Abstract:

Compositional data (CD) is mostly analyzed as relative data, using ratios of components, and log-ratio transformations to be able to use known multivariable statistical methods. Therefore, CD, where some components equal zero, represent a problem. Some methods for handling this issue have been proposed; most of them substitute the zero parts by a small quantity. These methods work properly in the presence of rounding zeros. Nevertheless, in the presence of structural zeros, the idea of substituting them for a small amount does not seem entirely adequate. For this matter, a two-part modeling approach is proposed to deal with structural zeros in longitudinal CD using a mixed-effects model. Furthermore, the model has been extended to the case where the non-zero components of the vector seem to come from two different populations. Model specification, as well as the estimation method, will be presented.

The proposed model was used to analyze the change over time of the radiation-induced lung damage (RILD) in one patient with non-small cell lung cancer (NSCLC). The extent of RILD is observed in a continuous manner, but tissue is classified into three-threshold based ordinal categories of radiographic injury, namely, dense, hazy, and none. Then, data for each patch is obtained as the proportions of voxels falling into each category. In this example, more than 99% of the patches did not show RILD. Therefore, this data set has a high presence of zeros. Results and challenges of this analysis will be discussed.

Title: Integrated Multiple Adaptive Clinical Trial Design Involving Sample Size Re-Estimation and Response-Adaptive Randomization for Continuous Outcomes Presenter: Christine M. Orndahl Advisor: Dr. Robert A. Perera Abstract:

Currently, single adaptive clinical trial designs are utilized most often, where only one adaptive design is used within the clinical trial. Consequently, only one pitfall of a fixed clinical trial design is addressed. Recently, increased interest has been developed in the area of multiple adaptive designs, incorporating more than one adaptive design within a single clinical trial. However, these multiple adaptive designs are typically performed in succession and information is not shared across the different designs. The goal of this project is to integrate multiple adaptive designs, specifically sample size re-estimation and response-adaptive randomization, into a clinical trial with a continuous outcome. In order to accomplish this, the weighted sum method for multi-objective optimization with a constraint to maintain statistical power is used to combine two objective functions. The first minimizes the sample size required while the second minimizes the total expected treatment response. These objective functions serve to adaptively adjust the sample size and the allocation ratio; this ensures that the fewest number of patients are enrolled in the trial while still maintaining adequate statistical power, and for the patients enrolled, the expected response of the sample is minimized which seeks to maximize the overall benefit of the trial. Preliminary results for applying these new methods to a clinical trial are presented.

Title: Computational Prediction of Boundaries of 3D Genomic Domains in Class Imbalance Settings Presenter: Spiro Stilianoudakis Advisor: Mikhail Dozmorov Abstract:

*Background*: Topologically Associated Domains (TADs) represent fundamental building blocks that lead to the organization and regulation of the three-dimensional (3D) genome. Several methods have been developed to leverage the growing amount of (epi)genomic annotation data for predicting TAD boundaries; however, they overlooked key characteristics of the data. Furthermore, resolution of Hi-C data remains well below the resolution of epigenomic assays (hundredths of bases for ChIP-seq) affecting the ability of Hi-C sequencing in discerning gene-level features of chromatin. This resolution gap between 1D and 3D genomic maps, as well as a paucity of chromatin features below the level of TADs, has significantly limited our understanding of epigenomic regulation of gene expression and its potential link to chromatin architecture.

*Methods*: Here we developed an ensemble framework for optimizing TAD boundary prediction using random forest classifiers by leveraging different combinations of Hi-C data resolutions, spatially defined predictor types, and re-sampling techniques, across multiple cell lines. We then aimed to predict TAD boundary location on a 1D genomic map by validating optimally built cell line specific models at base pair (bp) resolution.

*Results*: We demonstrate with our ensemble framework that cell line specific random forest classifiers built on 10 kb resolution Hi-C data, with distance-type predictors, using SMOTE (\*Synthetic Minority Oversampling Technique\*) re-sampling yield optimal predictive performances, and outperform existing methods. Additionally, TAD boundary regions predicted at bp resolution using our optimally built cell line specific models were found to be significantly closer to known enriched epigenomic marks indicating more accurate identification of TAD boundaries than existing Hi-C sequencing technologies.

*Conclusions*: Our results outline strategies for predictive modeling of 3D genomic domains using 1D genome annotation data. The accurate modeling of TAD boundaries will improve our understanding of how epigenetics shapes the 3D structure of the genome.

Title: A new estimation method for the semiparametric accelerated mixture cure model Presenter: Jonathan W. Yu Advisors: Dr. Dipankar Bandyopadhyay, Dr. Steve Chiou, Dr. Le Kang Abstract:

In clustered data such as the United Network of Organ Sharing (UNOS) database, the center size can affect patient survival time after transplant with its access to medical resources. Improper statistical procedures to handle informative cluster size can lead to biased results and misleading inferences. While the accelerated failure time (AFT) mixture cure model and the Cox proportional hazards (PH) mixture cure model are two classic models to analyze clustered survival data, the AFT has attracted less attention than its semiparametric counterpart due to the complexity of the estimation method. However, its direct physical interpretation and developments to the rank-based generalized estimating equations (GEE) provides an incentive to use for censored failure time data. We propose a new estimation method for the semiparametric AFT mixture cure model that employs a faster expectation-maximization (EM) algorithm, the SQUAREM, that can accelerate any fixed-point and smooth mapping with linear convergence rate and an induced smoothing inverse cluster size reweighting procedure to handle the informative cluster size. To evaluate the performance of the proposed method, we conducted a simulation study. The results of the simulation study demonstrate that the proposed method performs better than the existing estimation method. We apply the proposed method to UNOS data of failure times from kidney transplant patients to demonstrate that this approach has better numerical performance than existing methods in literature.

## **Biostatistics Student Research Symposium (BSRS) 2019 Abstracts**

**Title:** *Matrix-Variate Skew-t Models for Big Data with the Distributed EM Algorithm, with Applications* **Presenter:** Rueben Retnam **Advisor:** Dr. Dipankar Bandyopadhyay **Abstract:** 

The Matrix-Variate Skew-t (MVSt) distribution, introduced by Gallaugher and McNicholas in 2017, was one of the first attempts at creating a distribution that allows researchers to model skewed and heavy-tailed data in a matrix-variate setting. The distribution permits a convenient hierarchical representation, allowing it to be fit relatively easily via the Expectation-Maximization (EM) algorithm. However, the EM algorithm used to fit models that utilize this distribution can become cumbersome when applied to large matrix-variate datasets.

In our work, we develop regression models for matrix-variate data that utilize the matrix-variate skew-t distribution while scaling to tackle today's massive amounts of data. This scaling is achieved via the implementation of divide-and-conquer techniques that utilize the distributed expectation-maximization algorithm. Specifically, the E-step of the EM algorithm is run in parallel on multiple worker processes, while manager processes perform the M-step with a fraction of the results from the local expectation steps . Applications to a large observational epidemiologic dataset tracking periodontal disease in a diverse population will be presented. Further work will extend these models to other complex data structures, such as irregularly observed longitudinal data.

**Title:** Estimating Responder Statuses in Sequential Multiple Assignment Randomized Trials (SMARTs) **Presenter:** Keighly Bradbrook **Advisor:** Dr. Roy T. Sabo **Abstract:** 

Sequential, multiple assignment, randomized trials (SMARTs) are a personalized approach to investigate experimental treatment regimens in which individuals are adaptively and successively randomized to different stages of treatment based on their responsiveness to previous stages of treatment. Successful implementation of SMARTs depends on the ability of investigators to distinguish between patients who respond to particular treatments and those who do not. In some clinical situations, there may not exist acceptable mechanisms for making classifications between responders and non-responders. In such cases where there is no known basis for next-stage randomization the SMART method is difficult to use. The goal of the current work is to develop a method of sequentially and probabilistically assigning responder statuses to subjects completing the first stage of a SMART-like trial supplemented with information provided through pilot data. Two approaches to estimating these probabilities are discussed. The first approach uses a combination of cluster analysis and discriminant analysis to group subjects and update responder probabilities. The second approach uses a mixture-model approach to update responder probabilities. In both cases responder probabilities are used to randomly allocate patients between responder and non-responder classifications, thus allowing patients to continue onto the second phase of treatment in a SMART-like trial. Both approaches and all scenarios will be evaluated using simulation studies and compared in terms of group-specific misclassification error rates.

Title: A natural lead-in approach to response-adaptive allocation Presenter: Erin Donahue Advisor: Roy T. Sabo Abstract:

Response-adaptive (RA) allocation designs skew the allocation of incoming subjects toward the better performing treatment group based on the previously accrued responses. While unstable estimators and increased variability can adversely affect adaptation in early trial stages, Bayesian methods can be implemented with decreasingly informative priors (DIP) to overcome these difficulties. DIPs have been previously used for binary outcomes to constrain adaptation early in the trial, yet gradually increase adaptation as subjects accrue. We extend the DIP approach to RA designs for continuous outcomes, primarily in the normal conjugate family by functionalizing the prior effective sample size to equal the unobserved sample size. We compare this effective sample size DIP approach to other DIP formulations. Further, we considered various allocation equations and assessed their behavior utilizing DIPs. Simulated clinical trials comparing the behavior of these approaches with traditional Frequentist and Bayesian RA as well as balanced designs show that the natural lead-in approaches maintain improved treatment with lower variability and greater power. Title: Effects of microbiome filtering on downstream analyses Presenter: Xinxin Sun Advisor: Dr. Ekaterina Smirnova Abstract:

*Introduction*: The presence of contaminating microbial DNA poses a particular challenge in the study of microbial communities and their relationship with hosts. Moreover, contamination varies greatly between processing kits and labs. Filtering, which removes rare and contaminant species, can vastly reduce the differences in observed species due to sample processing protocols. Currently, there are two statistical methods for filtering the contamination, decontam and PERFect. The goal of this project was to compare the strengths and limitations of each method and make recommendations for potential improvement.

*Methods*: Both methods were applied to quality control microbiome datasets. These data sets had replicated samples processed at different conditions. Filtering identified and removed contaminant species from the observed taxa table. Results were compared in terms of alpha and beta diversity to evaluate performance using graphical assessments and non-parametric tests.

*Results*: The filtered data of PERFect had less variability in terms of species within sample diversity compared to decontam, whereas decontam lead to smaller pair-wise between-sample distances between replicate samples. Thus, PERFect was more effective in reducing alpha diversity, while decontam performed better in improving beta diversity estimation.

*Discussions*: Decontam is efficient at discriminating contaminant taxa with high abundance, but not suitable for rare taxa. In contrast, PERFect is efficient at classifying contaminant taxa with low abundance. We conclude that these methods have complimentary filtering effects, and recommend removing rare taxa using PERFect, followed by decontam to filter abundant contaminating features. Our work showed that the filtering improves the alpha and beta diversity estimation for microbial communities across samples processing in different conditions and has high potential to improve data integration from different studies.

**Title:** Testing and Validating Prospective Classification of Patients into Lymphocyte Trajectory-Based Groupings **Presenter:** Brielle Forsthoffer **Advisor:** Dr. Roy T. Sabo **Abstract:** 

Lymphoid recovery has been displayed as distinct courses individuals might take following myeloablative stem cell transplantation (SCT). Immune reconstitution has been shown to influence health outcomes such as GVHD, relapse and survival time. Classifying patients according to lymphoid recovery depends on arbitrary criteria being set prior to modeling absolute lymphocyte counts (ALCs, µL-1) over time. We aim to utilize machine learning algorithms, not only to help make such classifications more objective, but to also prospectively predict the course an individual might take following SCT. First, to avoid subjectivity, an adaptation of the group-based trajectory model (GBTM) was used to determine the clinical trajectories of any SCT recipient based off ALCs. Because of the limitations of polynomial curves not tolerating arbitrary shapes in trajectories, cubic B-splines were utilized in the GBTMs. Diagnostics were used to determine the appropriate number of groups and a suitable data cut-off. Predicting the course an individual might take early on is of utmost importance in understanding possible associations with health outcomes to help guide therapies. To represent prospectively predicting patients at each timepoint, the jackknife method was utilized. Further, time to high probability of group assignment was calculated to determine early detection of group distinction. The diagnostics indicated the three-group model with only 60 days of data gave the most accurate classification. When prospectively predicting, the low group had high classification probability between days 15 and 18, the middle group between days 15 and 26, and the high group between days 3 and 8. cGVHD was more likely to occur for participants in the middle and high groups at days 30 and 60, and DLI was more likely to occur for participants in the low group at day 60.

**Title:** *Quantifying power of differential analysis of interactions within the 3D structure of the genome* **Presenter:** Matt Carli **Advisor:** Mikhail Dozmorov **Abstract:** 

Chromatin Conformation Capture (3C) technology such as Hi-C became one of the main tools for functional genomics analyses providing information about the three-dimensional (3D) chromatin interaction patterns. Paralleling differential gene expression analysis, there is a growing interest in the differential analysis of the 3D genomic interactions. However, it is unknown how properties of Hi-C data affect the power of differential analysis of the 3D genomic interactions. The parameters investigated were the number of samples, the size of differences between groups, the effect of distance between genomic interactions, the sparsity of the data, and the resolution of the data. Both experimental and simulated data were used in our investigation. The experimental data was taken from *Rao et al., Cell, December 11, 2014,* cell line GM12878, at various resolutions. Simulated data was generated using the FIND R package and downloaded from the HiCToolsCompare repository. Utilizing the multiHiCcompare R package, changes were introduced to genomic data and power determined by the proportion of changes that were correctly identified. Permutations of these analyses were run, allowing for the systematic quantification of the effect of the parameters of interest. Power was shown to increase with an increase of samples, increase with larger differences, and decrease with an increase in sparsity and distance between interactions.

Title: Impact of Childhood Obesity on Adulthood Health via Growth Curve Modeling Given Data MAR Presenter: Alicia Richards Advisors: Yongyun Shin and Shumei Sun Abstract:

In this project, we investigate the impact of growth in adolescent BMI on the progression of systolic blood pressure (SBP) in adulthood. We analyzed repeated measurements of adolescent BMI and adulthood SBP collected from the Fels Longitudinal Study (FLS) between 1929 to 2010. We model the BMI and SBP repeatedly measured on two distinct sets of unbalanced time points that are nested within each individual using a bivariate growth curve model (GCM) with correlated random coefficients. We examine the impact of adolescent BMI on SBP in younger adulthood (22-38 years of age) and in middle age (45-65 years of age) for comparison to find how long the impact of childhood obesity lasts on adulthood health. Furthermore, we investigate the impact of time-to-time variation of BMI deviations from the growth curve within each person. Our preliminary analysis shows that the impact of childhood BMI on earlier adulthood SBP is stronger than the impact on middle age. As a future research, we will efficiently estimate the bivariate GCM by all observed data where not only the random coefficients of person specific growth curves, but also the time-to-time variability of BMI deviations from the curves may influence SBP later in time.

Title: Adaptive allocation in clinical trials using mediators. Presenter: Salem Rustom Advisor: Dr. Robert Perera Abstract:

Response Adaptive (RA) designs have been primarily developed as an ethical response to the issue of subjecting study participants to inferior treatments (by minimizing allocation to them) without compromising internal validity. However, if time-to-response is too long, then adaptive allocation cannot effectively ameliorate the issue. In scenarios where the presence and use of a mediator is plausible, the mediator could be implemented as a predictor of the outcome. Mediators are variables influenced by an independent variable which in turn influence the response. Hence using mediators for prediction allows for earlier adaptive allocation by having values for mediators much sooner than actual responses. We demonstrate this through modifying Maxwell, Cole, & Mitchell's (2011) Autoregressive Longitudinal Mediation (ALM) Model and incorporating it into a RA clinical trial design that uses Wei & Durham's (1978) Randomized Play-the-Winner (RPW) rule. Simulations of this design were completed to compare performance metrics under different conditions against the standard randomized clinical trial design and the same RA design without use of the modified ALM model. Future work to extend utility of this novel design is also discussed.

Title: Real-time Surveillance of Multi-drug resistant Infections in Hospitals: Evaluating the Bernoulli CUSUM Method for Detecting Changes in Infection Rates Presenter: Martin Lavallee Advisors: Dr. Shanshan Chen and Dr. Adam Sima Abstract:

Timely detection of outbreaks of multi-drug resistant organisms (MDROs) infection is necessary to curtail its widespread in hospitals. A potential solution for detecting outbreaks is the statistical change point (SCP) framework, which models for structural changes in a monitoring process (i.e. a time series of the infection incidence) and estimates when a change has occurred. Given the binary nature of our potential data source, we focused on a particular SCP method -- the Bernoulli CUSUM (BC) method to tackle this problem. Without having to aggregate the binary data to estimate infection rate, BC allows for prompt detection by designing and evaluating a sample-by-sample based CUSUM statistic. A structural change can then be detected when the CUSUM statistic exceeds a designated control limit (CL). Critically, the CL determines the performance of a CUSUM design, such as the delay between event and detection, and false positive/negative rates. Existing R packages either arbitrarily set the CL or empirically determine it via simulation. To seek the optimal CL, we implemented a Newton-Raphson (NR) solution, which minimizes the delay in detection. We then simulated 11,000 Bernoulli sequences with different proportion changes and evaluated the performances of the two existing R methods and the Newton-Raphson method. Results show that the NR approach for CL calculation provided the best overall performance by providing the shortest delay with slightly higher false positive rates than determining CL via simulation. Lastly, we also implemented an R-shiny app for intuitively understanding the detection process, which can also be used as a tool for choosing the optimal CUSUM design given an infection monitoring sequence in hospital.