2019

Methods for evaluating dropout attrition in survey data

Camille J. Hochheimer

Methods for evaluating dropout attrition in survey data

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Camille Jo Hochheimer

Director: Roy T. Sabo, PhD, Department of Biostatistics

Virginia Commonwealth University

Richmond, Virginia

March 6, 2019

Acknowledgments

There are so many people who have helped make this acheivement possible and I would like to thank a few here. First and foremost, I'd like to thank my parents and my sister for their unwavering support. Your positivity and belief in me has propelled me further than I ever could've imagined. To my classmates, BAMF, we have certainly walked this road together and I would not be here if it weren't for several hours locked in a conference room with you wonderful people. To my advisor, Dr. Sabo, thank you for believing in me as a researcher and as a statistician. Working with you over the past five years has been a true highlight of my time at VCU. To Alex Krist and the Family Medicine Research Team, I'm forever grateful to have spent over four years working with you all. I'm so proud of everything we have accomplished together. Thank you to my committee, from whom I have learned so much over the past five years. Thanks for always answering my endless questions, reviewing my work, and serving as co-authors. And finally, to Cameron, thank you for never letting me get caught up in the lows and for truly celebrating the highs of this process with me. I love you and am so grateful to have you as my partner.

Methods for evaluating dropout attrition in survey data

by

Camille Jo Hochheimer

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2019.

Director: Roy T. Sabo, PhD, Associate Professor, Department of Biostatistics

Abstract

As researchers increasingly use web-based surveys, the ease of dropping out in the online setting is a growing issue in ensuring data quality. One theory is that dropout or attrition occurs in phases that can be generalized to phases of high dropout and phases of stable use. In order to detect these phases, several methods are explored. First, existing methods and user-specified thresholds are applied to survey data where significant changes in the dropout rate between two questions is interpreted as the start or end of a high dropout phase. Next, survey dropout is considered as a time-to-event outcome and tests within change-point hazard models are introduced. Performance of these change-point hazard models is compared. Finally, all methods are applied to survey data on patient cancer screening preferences, testing the null hypothesis of no phases of attrition (no change-points) against the alternative hypothesis that distinct attrition phases exist (at least one change-point).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Web-based surveys have become an increasingly popular mode of collecting data on patients and consumers alike due to the ease and cost-effectiveness of delivery as well as the ability to reach large numbers of subjects. A drawback to this medium, however, is the increased potential for and evidence of dropout [1,2]. Surveys administered online are especially susceptible to dropout attrition, where a participant starts but does not complete a survey or questionnaire. For the purposes of this dissertation, both dropout and attrition will be used interchangeably to indicate this phenomenon. Survey dropout presents a missing data problem and when dropout is related to topics addressed in the survey itself, it may be considered missing at random (MAR) or missing not at random (MNAR) [3]. Researchers should take this into account when analyzing survey data in order to avoid biased results, though methods for doing so are still under development.

Eysenbach's call for a *science of attrition* introduced the idea that survey attrition occurs in three distinct phases [1]. There is an initial *curiosity plateau* at the beginning of a study where the participation rate remains high while respondents gauge their interest in the survey content. This is followed by a dropout or *attrition phase* where participants exit the survey at a higher rate, presumably due to disinterest, poorly worded survey items, or a poorly constructed survey altogether. The *stable use phase* is reached when most remaining participants are likely to complete the survey. These can be seen in Figure 4 of Eysenbach's manuscript [1], included here as Figure 1.1.

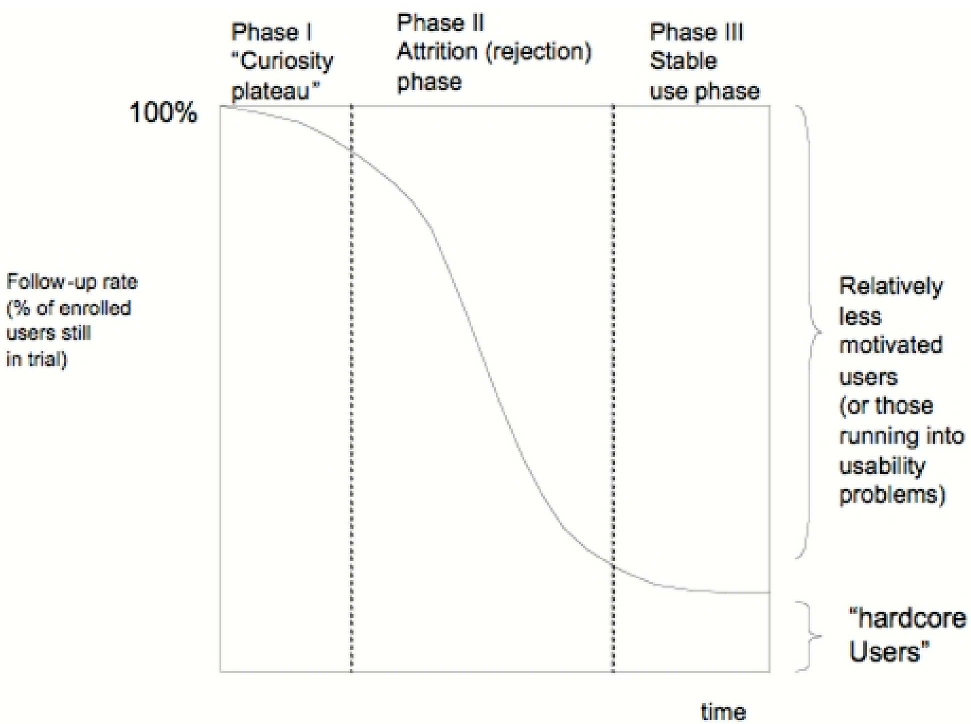


Figure 1.1: Eysenbach's attrition phases

To our knowledge, there has been no research attempting to directly estimate Eysenbach's attrition phases. The ability to directly estimate and identify phases of attrition, elucidating the point(s) in the survey at which participants start dropping out at a higher rate, may also reveal the survey content and participant characteristics influencing these patterns. If we can identify clear attrition patterns, then almost by definition we have data MNAR that require more attention [3]. In general, the evaluation of dropout patterns can be used as a tool to help improve surveys and questionnaires by highlighting "problem questions" that may be irrelevant, un-clear, or biased towards or against subsets of survey participants. In a study setting where high rates of dropout are becoming the norm, this tool could also be used to address the *science of attrition* and help manuscripts reach publication despite high (but expected) dropout rates. We began to address this problem using visual and standard statistical methods in our paper "Methods for Evaluating Respondent Attrition in Web-Based Surveys" [4].

1.2 Current methods

Several standard methods could be applied to the problem of identifying attrition phases. A generalized linear mixed model (GLMM) is one option, as one can fit the model to the entire survey with the outcome being a binary indicator of whether or not a participant drops out at a specific question and then use contrasts (specifically successive differences contrasts) to test the hypothesis of a significant difference between the proportion of dropouts between sequential questions. The issue with this model, however, is the necessity to account for dependence between questions within individuals as well as the discrete nature of the questions. When these survey characteristics are taken into account within a GLMM, the complexity of the variance-covariance structure can prohibit estimator convergence, ultimately limiting the length of questionnaire that can be analyzed using existing software. When fitting this model is possible, doing so is inefficient in terms of computation time, even in scenarios with short surveys and reasonably small sample sizes.

Eysenbach suggests the use of survival analysis, a reasonable approach because it assumes that individuals are followed over time by treating survey completion or attrition as a time-to-event outcome where questions are viewed as sequenced in time [1]. We have applied this method previously (see [4]) and in this dissertation we will incorporate the discrete nature of survey questions by using discrete time survival analysis (DTSA). Just like with the GLMM, we can apply successive differences contrasts to determine significant changes in hazard between sequential pairs of questions. An advantage of this method is its ability to accommodate surveys with more questions than the GLMM.

One drawback to both the GLMM and DSTA methods is that the successive differences method of testing for significant changes in the hazard is a crude approach to identifying phase transitions, as we are only able to zoom-in on each pair of questions instead of looking for overall patterns. While we may be able to find statistically significant differences in dropout by applying a GLMM or DTSA, the differences may not be clinically meaningful to researchers. In other words, we are linking phase transitions to hypothesis testing, which might not be an appropriate assumption. For instance, with a large sample size, formal hypothesis testing might declare differences in proportions as statistically significant, even if they are smaller than what the researcher administering the survey would consider meaningful.

Change-point models could provide a more parsimonious approach to identifying overall attrition trends. As opposed to performing a hypothesis test for every pair of questions, we can fit a model to the entire survey and then interpret change points as moments of phase transition. Whereas GLMM and DTSA models identify transitions between attrition phases by searching for the first and last instances of significant attrition, change-point models allow us to generalize Eysenbach's phases, seen in Figure 1.1, to phases of stable use and phases of high dropout [1]. This would allow for several phases of high dropout, which may be present in longer questionnaires.

Change-point models have been explored in many areas of statistics. Vexler and Gurevich lead the research on this topic within the logistic regression framework, although their models cannot account for repeated measures within individuals [5–8]. Several researchers have suggested Bayesian change-point models which calculate the posterior probability of change points, though few of these Bayesian models allow for binary outcomes and/or discrete time points [9–13].

Current research on change-point hazard models for time-to-event outcomes focuses mostly on non-parametric and semi-parametric hazard estimation and testing for one change point. These models are flexible and allow for the inclusion of covariates. This is important in order to look into factors potentially contributing to large proportions of participants dropping out. Williams proposed a parametric model, introducing a test for a single change-point within a Weibull hazard function [14]. This model is of specific interest as the Weibull model is flexible, able to incorporate covariates, and satisfies the proportional hazards assumption, allowing for a straightforward interpretation.

A few researchers have introduced tests for multiple change points [15–22]. Among them is Goodman, who proposed a test for multiple change-points within the piecewise constant hazard or exponential change-point hazard model that uses a sequential testing scheme to determine the best fitting change-point model [17]. It is imperative that any test considered accommodate multiple change-points in order to highlight multiple phases of attrition. Thus, multiple change-point hazard models are of most interest in this dissertation as they accommodate multiple phases, are reasonably easy to use, and have a meaningful interpretation to researchers who are and are not familiar with the statistical details.

1.3 Aims

We propose several approaches to estimating phases of survey attrition: existing statistical models and two types of change-point hazard models. We will compare the ability of each method to detect phases of attrition through both simulation studies and applications to an existing dataset from a survey administered by our research team.

In our framework, the overall null hypothesis is that there are no distinct phases of attrition. This does not necessarily mean that there is no attrition, no statistically significant attrition, or no clinically meaningful attrition. Indeed, we might observe a linear pattern where attrition is not present or where attrition is high throughout the survey. This pattern of constant attrition does not have distinct phases and, thus, fits our definition of the null hypothesis. The alternative hypothesis is that at least two phases of attrition are present in the survey data.

This research is ultimately intended to create a tool to identify where participants drop out at a higher rate than anticipated during both the pilot and final analysis stages of survey implementation, particularly by identifying questions that signify the beginning and end of attrition phases.

1.3.1 Aim #1: Existing statistical models

We wanted to see if existing statistical methods can be successfully applied to identify multiple attrition phases within survey data. First, we applied a user-specified threshold based on clinical meaningfulness to detect these phases. Second, we applied a generalized linear mixed model (GLMM) to the entire survey to find points of significant attrition between sequential questions. Finally, we applied discrete time survival analysis (DTSA) to find points of significant change in the hazard of dropping out.

1.3.2 Aim #2: Test within a Weibull hazard model with multiple change-points

Next, we treated survey dropout as a time-to-event process and developed a test for multiple phase transitions. In order to do so, we extended William's test to accommodate more than one change-point within a Weibull hazard model [14].

1.3.3 Aim #3: Test within an exponential hazard model with multiple change-points

We then proposed an alternate test statistic and testing scheme for multiple change-points within an exponential hazard function. This extension allowed us to bypass Goodman's sequential testing scheme in order to choose the best overall change-point model [17].

1.3.4 Aim #4: Simulation study and data application

Models in Aims #2-3 were compared through a simulation study and models for Aims #1-3 were applied to test case data. Specifically, we wanted to see whether these methods correctly identified phases of attrition when they exist and how efficiently they accomplished this goal. By applying these methods to an existing dataset on cancer screening, we demonstrated how these methods could be used to identify attrition patterns and make recommendations for improving the survey.

1.4 Summary

In order to achieve the goals set in section 1.3, we first implemented the models in Aim #1 using existing packages in the R statistical software [23]. The novelty in this aim comes from both the user-specified approach and the application of these existing models in a new setting: survey data. In order to achieve Aim #2, we generalized the test statistic proposed by Williams in order to accommodate more than one change-point [14]. We also created an algorithm for testing for more than one change-point based on optimal model fit as an alternative to Goodman's progressive testing approach [17]. We then derived an alternative test statistic for testing for multiple change-points within an exponential hazard function based on overall model fit as another alternative to Goodman's sequential approach. A simulation study was performed in order to choose an appropriate method for simulating data for a change-point distribution and to compare the performance of the proposed methods. We also present the results of the application of these methods to an existing dataset. A detailed discussion of the proposed methods to achieve each aim can be found in Chapters 2-5.

There are many ways in which this research contributes to the existing body of statistical knowledge. First, the completion of these aims helps to make Eysenbach's call for a *science of attrition* one step closer to realization. In line with this goal, the specific focus on change-point models will help researchers from any field be able to identify phases of attrition within survey data. This work also provides a way to identify survey data MAR or MNAR. The extension of the Weibull model to incorporate multiple change-points and extension of the exponential model test statistic will additionally be applicable to any situation with time-to-event data.

Chapter 2

Using practical thresholds and existing statistical methods to identify attrition phases

2.1 Introduction

2.1.1 Motivation

Survey questions are discrete in nature and dependent within survey participants. We are interested in seeing whether we can identify phases of attrition while accounting for the features of these data within the framework of existing statistical methods. Treating attrition as a response rate would allow for direct comparison of these rates within the context of a GLMM. Treating dropout as a time-to-event outcome is another possible and appealing approach, as these models already assume subjects are followed over time. DTSA is especially appropriate because questions occur as discrete time points in the process of participating in a survey or questionnaire; participants can only dropout at these distinct points. Eysenbach himself suggests applying survival analysis to attrition patterns [1].

In addition, we often see that statistical significance is not always indicative of a clinically meaningful difference, as formal statistical tests may detect very small changes (especially when there is a large sample). Investigators generally have an idea of what they consider to be meaningful and this concept can be directly applied to the identification of attrition phases. Therefore, it is of interest to develop a user-specified approach to address this issue.

2.1.2 Current methods

Eysenbach originally proposed three phases as mentioned in Chapter 1 [1]. We previously demonstrated estimating these phases visually, using a GLMM applied to each sequential pair of questions, a non-parametric log-rank test comparing Kaplan-Meier curves between groups, and a semi-parametric Cox proportional hazards regression to compare groups while adjusting for other characteristics [4]. We also mentioned the future extension of our work to DTSA in our paper and here demonstrate how this method can be applied to identify attrition phases.

2.1.3 Aim

We applied a user-specified and two existing statistical methods as our first approach to determine the phases of attrition within survey data. First, we applied a user-specified threshold based on clinical meaningfulness to detect phase transitions. Then we applied two hypothesis testing methods where significant differences were interpreted as transitions between attrition phases. A generalized linear mixed model (GLMM) was applied, using contrasts to identify significant changes in the dropout rate from one question to the next. We also applied discrete time survival analysis (DTSA) to the entire survey, using contrasts to identify significant changes in the hazard rate between questions. Finally, we conducted a simulation study to compare the performance of these three approaches.

2.2 Statistical Methods

We propose an approach motivated by establishing meaningful dropout standards and two additional approaches motivated by statistically significant attrition. The first allows researchers to set thresholds based on what they consider a meaningful or practical amount of dropout at any question. The second proposed approach is based on formal hypothesis testing where we search for instances of statistically significant differences in dropout between sequential questions. We apply two existing statistical models, the GLMM and DTSA, to survey dropout data. Statistical significance is not always indicative of a meaningful difference, especially when the sample size is large, as is often the case with web-based surveys. Thus, these serve as complimentary methods to identify inflection points of the dropout rate. The R statistical software version 3.5.0 was used to apply these methods [23].

2.2.1 User-specified attrition thresholds

For the first method, the user specifies the proportion of dropout at a single question that is considered to be clinically meaningful. Specifically, they define Δ_1 as the threshold for the start and Δ_2 as the threshold for the end of the dropout phase of attrition. We directly calculate the proportion of dropout (p_{dr}) in each survey question and compare it to both thresholds. The dropout phase begins the first time that dropout at a particular question exceeds the start threshold ($p_{dr} > \Delta_1$). Likewise, the dropout phase ends the last time that dropout at a survey question exceeds the end threshold ($p_{dr} > \Delta_2$). Different numerical values can be used for the start and end thresholds.

2.2.2 Hypothesis testing

GLMM

In order to test for statistically significant attrition, we apply a GLMM:

$$logit(\hat{p}_j) = \hat{\beta}_0 + \sum_{2}^{q} \mathrm{I}(j)\hat{\beta}_{1,j}. \tag{2.1}$$

Our binary outcome is whether or not a subject dropped out of the survey. Survey question is included as the time-varying covariate $\hat{\beta}_{1j}$ with the number of levels j equal to the number of questions q and an indicator function $\mathrm{I}(j)$ to identify the question of interest. This model accounts for within-subject dependence between response rates through the incorporation of a random effect γ_i. In other words, once a participant drops out they can no longer reenter the survey. In equation 2.1, this is represented by the random intercept $\hat{\beta}_0$ which is different for each subject. This model appropriate because we have repeated dependent outcomes within subjects due to the fact that a participant must answer the previous question in order to drop out in the current question.

The null hypothesis that the proportion of participants remaining in the survey has not changed between questions ($H_0 : \hat{p}_j = \hat{p}_{j+1}$) is rejected in favor of the alternative hypothesis that this proportion is different between sequential questions ($H_1 : \hat{p}_j \neq \hat{p}_{j+1}$) if the p-value is significant at the five percent significance level. We assumed that participants cannot re-enter the survey once they have dropped out and that skipping a question does not count as dropout if the patient returns to answer a later survey item.

We applied this model to the entire survey using the *glmer* function of the *lme4* package [24]. This function uses an unstructured variance-covariance matrix. A successive differences contrast was applied using the using the *glht* function of the *multcomp* package to test each pairwise difference in the proportion of participants remaining in the survey ($\hat{p}_j - \hat{p}_{j+1}$) [25]. A false discovery rate correction based on the Benjamini-Hochberg procedure was used to correct the p-value for multiple comparisons [26].

DTSA

We applied DTSA as another hypothesis testing method. The logit link connects the baseline hazard function θD_{ti} to the discrete time hazard function $p_{ti} = P(y_{ti} = 1|y_{t-1,i} = 0)$, where y is the event (dropout). This relationship can be written as $logit(p_{ti}) = \theta D_{ti}$, where θD_{ti} is a step function with a dummy variable for each survey question ($\theta D_{ti} = \theta_1 D_1 + \theta_2 D_2 + \ldots + \theta_n D_n$). Although we do not include covariates in this particular instance, a useful feature of the discrete time survival model is the ability to incorporate non-proportional hazards [27]. The null hypothesis that the hazard of dropping out of the survey has not changed between questions ($H_0 : \theta_j = \theta_{j+1}$) is rejected in favor of the alternative hypothesis that the hazards are different between sequential questions ($H_1 : \theta_j \neq \theta_{j+1}$) if the p-value is significant at the five percent significance level.

This model was implemented using the *svyglm* function of the *survey* package [28,29]. We again applied successive differences contrasts using the *glht* function of the *multcomp* package to determine significant differences in the hazards of pairs of questions and adjusted for these differences also using the Benjamini-Hochberg procedure [25,26].

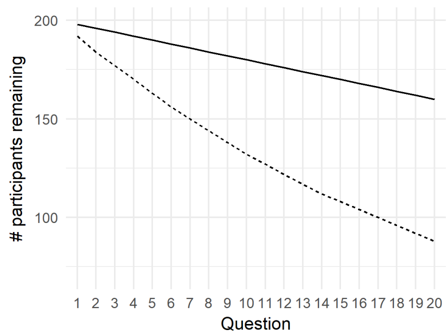
In both the GLMM and DTSA models, we searched for the first and last instances that the adjusted p-values drop below our pre-specified threshold of 0.05. These points are interpreted as the beginning and end of the dropout phase.

2.3 Simulation methods

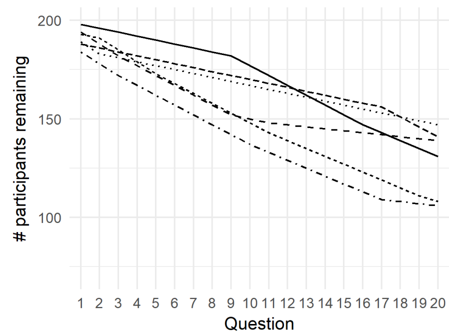
In order to demonstrate the performance of these methods in detecting attrition phases, we simulated a variety of dropout patterns. For each pattern, we simulated 10,000 datasets each with 200 simulated participants answering twenty questions. Respondents had a random chance of dropping out at any point in the survey, including the first question, and a participant could not reenter once they dropped out.

Simulated dropout patterns corresponded to constant, two-phase, and three-phase attrition. Constant attrition could be either the stable use or dropout phase of attrition throughout the survey (see Figure 2.1a). Two-phase attrition either began with the stable use phase and then transitioned into the dropout phase for the remainder of the survey or began with the dropout phase with a transition to the stable use phase (see Figure 2.1b and 2.1c). Three-phase attrition followed Eysenbach's proposed pattern in that there were stable use phases at the start and end of the survey with a dropout phase in the middle (see Figure 2.1d and 2.1e). We tested both mild and severe attrition rates for the dropout phase, where severe attrition rates demonstrated a more pronounced difference in dropout rate between phases, in order to determine the sensitivity of these methods. The location of the phase transition was varied in order to see whether these methods better identified phase transitions that occurred at the start, in the middle, or at the end of the survey.

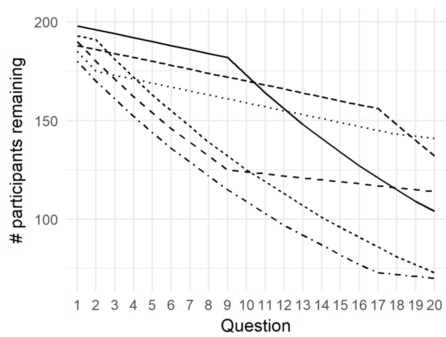
The overall null hypothesis we tested was that there are no phases of attrition, represented by the constant attrition patterns in Figure 2.1a. When the dropout rate was mild, we expected our proposed methods would not detect practically meaningful or statistically significant attrition at any point in the survey. When this rate was severe, we expected to see the first instance of meaningful attrition at question one for the user-specified method and the first instance of significant attrition



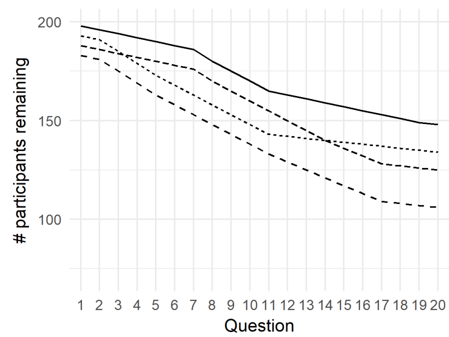
(a) Constant attrition



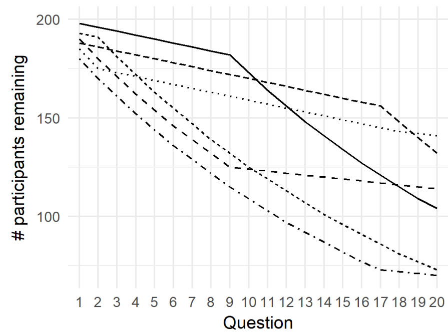
(b) Mild two phase attrition



(c) Severe two phase attrition



(d) Mild three phase attrition



(e) Severe three phase attrition

Figure 2.1: Simulated attrition patterns

between questions one and two for the GLMM and DTSA. Then, we hoped to see this attrition phase last throughout the survey with the last instance of meaningful attrition found at the final question and last instance of significant attrition found between the last two questions.

Using the constraint $\Delta = \Delta_1 = \Delta_2$ for the user-specified method, we assessed the performance of a three percent, five percent, and eight percent dropout threshold ($\Delta = 0.03, 0.05, 0.08$). We determined how well these methods achieved the goal of detecting attrition phases by calculating and comparing the type I error and sensitivity of each method. Type I error was defined as when at least two phases of attrition were detected when the underlying attrition pattern was constant (i.e., phases of attrition were detected when they did not exist). Sensitivity was defined as finding the correct number of phase transitions when they did exist. Ideally, these methods would achieve a type I error of 5% and higher sensitivity.

We also visually inspected how often each question was chosen as the start or end of the attrition phase by plotting these distributions with histograms. The histograms for the user-specified method highlight the first and last question at which the amount of dropout at a question surpassed Δ. Those for the GLMM and DTSA show the first and last comparisons between questions with a significant adjusted p-value.

2.4 Results

The resulting type I error and sensitivity for each approach and simulated attrition pattern are in Tables 2.1, 2.2 and 2.3. A selection of the histograms displaying the distribution of questions chosen as the start and end of the attrition phase are displayed in Figure 2.2.

2.4.1 User-specified thresholds

The three user-specified thresholds that we tested are compared in Table 2.1. The five percent threshold was unable to control type I error with 0% type I error in the mild constant case and 94% type I error in the severe constant case. Additionally this threshold had lower sensitivity than the three percent threshold except for three simulation scenarios, and in these cases the three percent threshold also had a high sensitivity of over 90%. The eight percent threshold had low sensitivity (often

Metric	Dropout	Attrition severity	τ location	Dropout threshold 3%	5%	8%
Type I error	Constant	Mild		0.19	0.00	0.00
		Severe		1	0.99	0.27
Sensitivity	Two phases	Mild	Middle	0.47	0.42	0.01
		Mild	Start	0.48	0.43	0.02
		Mild	End	0.72	0.2	<0.01
		Mild reverse	Middle	0.44	0.34	<0.01
		Mild reverse	Start	0.51	0.04	0.00
		Mild reverse	End	0.41	0.39	0.01
		Severe	Middle	0.82	0.48	0.38
		Severe	Start	0.79	0.47	0.44
		Severe	End	0.90	0.74	0.14
		Severe reverse	Middle	0.88	0.46	0.26
		Severe reverse	Start	0.90	0.43	0.03
		Severe reverse	End	0.88	0.42	0.39
	Three phases	Mild	Middle	0.66	0.02	0.00
		Mild	1 start	0.96	0.13	0.00
		Mild	1 end	0.96	0.16	0.00
		Mild	Ends	0.96	0.32	0.00
		Severe	Middle	0.98	0.64	0.01
		Severe	1 start	0.96	0.97	0.07
		Severe	1 end	0.96	0.98	0.11
		Severe	Ends	0.94	0.99	0.24

Table 2.1: User-specified results

Attrition severity	User-specified 3%	GLMM	DTSA
Mild	0.19	0.01	0.74
Severe	1.00	0.98	0.96

Table 2.2: Aim 1 type I error

0%) to detect any phase transition. Seeing as the three percent threshold had the best performance, we compared this threshold to the hypothesis testing methods in Tables 2.2 and 2.3.

The 3% user-specified threshold had high type I error but also high sensitivity to detect attrition phases, especially in scenarios with a severe dropout phase and three phases of attrition. This threshold achieved higher sensitivity than the hypothesis

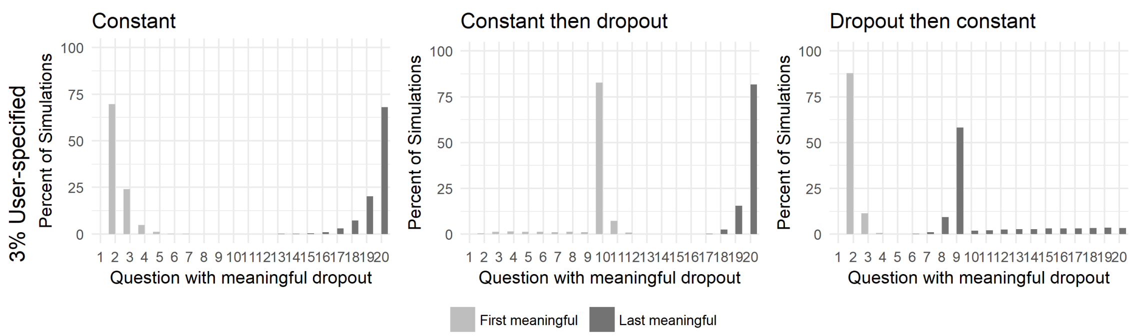
Dropout	Attrition severity	τ location	User-specified 3%	GLMM	DTSA
Two phases	Mild	Middle	0.47	0.22	0.90
	Mild	Start	0.48	0.17	0.94
	Mild	End	0.72	0.13	0.87
	Mild reverse	Middle	0.44	0.20	0.64
	Mild reverse	Start	0.51	0.01	0.69
	Mild reverse	End	0.41	0.14	0.58
	Severe	Middle	0.82	0.24	0.91
	Severe	Start	0.79	0.20	0.97
	Severe	End	0.90	0.56	0.88
	Severe reverse	Middle	0.88	0.02	0.58
	Severe reverse	Start	0.90	0.01	0.69
	Severe reverse	End	0.88	0.00	0.50
Three phases	Mild	Middle	0.67	0.12	0.08
	Mild	1 start	0.96	0.32	0.07
	Mild	1 end	0.96	0.35	0.05
	Mild	Ends	0.96	0.51	0.03
	Severe	Middle	0.98	0.84	0.08
	Severe	1 start	0.96	0.99	0.07
	Severe	1 end	0.96	0.99	0.06
	Severe	Ends	0.94	0.99	0.04

Table 2.3: Aim 1 sensitivity

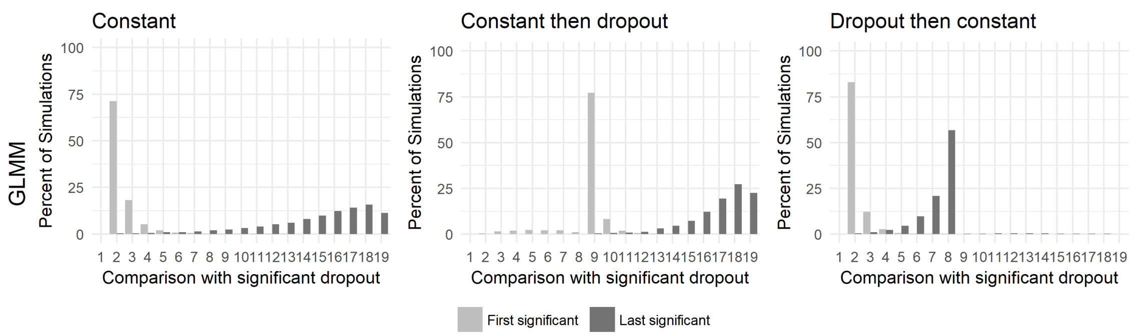
testing methods when the simulated dropout pattern for two phases began with the dropout phase and ended with the stable use phase. The user-specified method failed to identify the start of the severe constant attrition phase immediately at question one (see Figure 2.2a). In the majority of simulations, this method correctly detected severe attrition phases simulated to last from questions one to ten and from questions ten to twenty.

2.4.2 GLMM

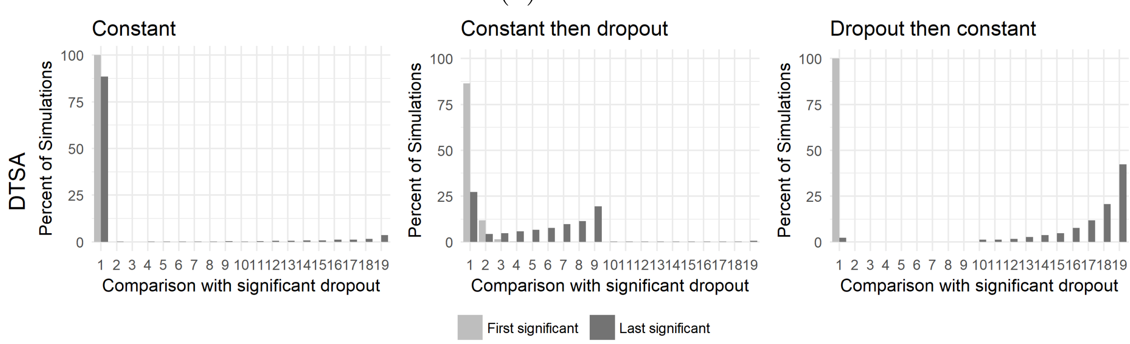
The GLMM had a conservative type I error rate in the case of constant mild attrition and high type I error in the case of constant severe attrition (Table 2.2). This method demonstrated low sensitivity to detect two phases of attrition in general and three phases with a mild dropout phase (Table 2.3). The GLMM had higher sensitivity to



(a) User-specified 3%



(b) GLMM



(c) DTSA

Figure 2.2: Question or comparison where phase transitions were detected in different patterns of severe attrition

detect all three of Eysenbach's attrition phases in simulation patterns with a severe dropout phase.

The histograms in Figure 2.2b reveal that the GLMM failed to identify the start of the attrition phase immediately between questions one and two when the simulated pattern was that of severe constant attrition. These plots also show that when a severe attrition phase was simulated to begin in the middle of the survey and continue until the end of the survey, the GLMM most often detected the start of the attrition phase correctly at question ten but detected the end of the attrition phase before the end of the survey (resulting in the low sensitivity seen in Table 2.3). When a severe attrition phase was simulated to begin at the start of the survey and last until the middle of the survey, the GLMM failed to identify an immediate start to the attrition phase but correctly distinguished the phase transition at question ten.

2.4.3 DTSA

DTSA did not control type I error and had low sensitivity to detect three attrition phases (Tables 2.2 and 2.3). For simulated patterns with two phases, DTSA demonstrated high sensitivity to detect both phases. DTSA achieved higher sensitivity than the user-specified and GLMM approaches for simulated patterns with a mild change in attrition between phases as well as when the simulated dropout pattern survey began with the stable use phase and transitioned to a severe dropout phase (Table 2.3). Overall, we observed higher sensitivity for DTSA when the simulated dropout pattern began with the stable use phase and then transitioned to the dropout phase compared to patterns beginning with the dropout phase and ending with the stable use phase. DTSA also consistently had higher sensitivity when the simulated phase transition occurred towards the start of the survey.

Histograms displaying the accuracy of DTSA can be found in Figure 2.2c. The comparison between questions one and two was correctly recognized as the first instance of significant attrition but also incorrectly chosen as the last instance when the dropout phase was simulated to last throughout the survey. Although this method detected the correct number of phases in the majority of simulations with two phases, it did not choose the correct questions as the start and end of the attrition phase. Specifically, DTSA was unable to detect an abrupt change in dropout rate in the middle of the survey. The histograms suggest a dropout phase at the beginning of the survey when the underlying simulated pattern had a dropout phase in the second half of the survey and suggest constant attrition when the simulated pattern had a dropout phase in the first half of the survey.

2.5 Discussion

In our simulation study, none of these three methods consistently detected the correct number of phases while controlling type I error. The 3% user-specified threshold had a high type I error rate but also accurately detected the phases of attrition and had moderate to high sensitivity for all simulated scenarios. While high sensitivity estimates of DTSA in the case of two attrition phases appeared promising, histograms revealed that this method consistently identified the wrong questions as the start and end of the dropout phase. This explains the low sensitivity of DTSA to detect all three phases of attrition. DTSA was extremely sensitive, finding a significant difference between the first two questions even when the simulated dropout rate was very small (e.g., 0.001).

We did not see any distinct patterns in sensitivity when the phase transition(s) occurred towards the start or end of the survey compared to the middle of the survey. Sensitivity was often higher for the hypothesis testing methods when there was a sudden increase in attrition than when there was a sudden decrease in attrition. This suggests that these methods do not consistently detect a phase transition when dropout starts off at a high rate and then levels off at a certain point in the survey. This issue persisted even when the change was more pronounced, as it was in the severe cases. While this should limit the ability of the GLMM and DTSA to detect three phases of attrition, we actually observed increased sensitivity for the GLMM when three phases were present.

Of the three thresholds tested, the three percent threshold had the best outcomes in terms of type I error, sensitivity, and accuracy, suggesting that lower user-specified thresholds perform better at identifying attrition phases than existing methods employing a formal hypothesis test.

The simplicity of the search algorithm used in these approaches is both an advantage in terms of being easy to apply and a drawback in terms of limiting the number of possible dropout phases to only one. The user-specified approach leaves it up to the discretion of the investigator to determine how sensitive the thresholds should be to capture the beginning and end of the dropout phase. It also avoids complicated statistical models and requires no further estimation other than dropout rates. These are also weaknesses of the user-specified approach, as subjective thresholds are likely not universally accepted by researchers and there is no statistical backing to these conclusions.

The largest disadvantage of the GLMM is the complexity of the covariance matrix, which poses both computational and convergence challenges. One important difference between this research and our prior work is that participants were able to drop out at the first question. Previously, having 100% compliance in the first question limited the number of questions to ten when applying the GLMM due to convergence issues. By assuming simulated participants could drop out in question one, we were able to apply the GLMM to the entire survey and compare this method to the other two discussed here.

When considering dropout as a time-to-event outcome, DTSA helps to alleviate some of these issues by automatically assuming patients are followed over time. Both hypothesis testing methods, however, search each sequential pair of questions for significant differences in survey participation. As mentioned previously, this algorithm provides only a zoomed-in look at attrition instead of looking at overall attrition patterns. These methods do not allow us to test for two dropout phases of attrition and, more generally, they do not allow us to test for more than three phases of attrition. In addition, it assumes that statistical significance implies attrition phase transition. This becomes an issue when large amounts of survey data are present and dropout that would not be considered clinically relevant might be detected by these models to be statistically significant.

These results suggests that when applying practical thresholds and existing statistical methods to the task of identifying Eysenbach's phases of attrition, sensitive user-specified methods correctly identify dropout phases at the cost of high type I error whereas hypothesis testing methods are unable to correctly identify these phases. These results strengthen the case for developing new methods to identify attrition phases and model overall attrition patterns.

Chapter 3

Testing for attrition phases using a Weibull change-point hazard model

3.1 Introduction

3.1.1 Motivation

As mentioned in Chapter 2, survival models are appealing due to the fact that they inherently assume subjects are observed over time. While testing for significant changes in hazard within the DTSA model may detect phases of attrition, this is a crude approach to finding a change in the dropout rate and cannot account for more than one dropout phase. A more practical approach might be to perform an overall test for change-points within the hazard in order to identify phase transitions.

3.1.2 Current methods

Several authors have introduced estimators for one change-point (which we will refer to as τ) within a constant hazard model, with some additionally incorporating censoring and covariates [30–34]. Testing for a change-point is unusual in that the change-point parameter τ appears only in the case of alternative hypothesis [35]. This creates a challenge in identifying the underlying distribution needed to calculate critical values for a likelihood ratio test. Different approaches to this challenge characterize the literature in this field. Matthews and Farewell were the first to propose tests for one change-point within a constant hazard model [35,36]. Researchers have since proposed several other tests, some with extensions to include both fixed and time-dependent covariates [37–42]. Of particular interest is the likelihood ratio based test for one change-point in a Weibull hazard model proposed by Williams et al. due to the flexibility of the Weibull model to fit many different shaped hazards, the fact that the Weibull hazard model satisfies the proportional hazards assumption providing a straightforward interpretation, and the direct extension to include covariates [14,43].

The Weibull hazard function with one change-point as proposed by Williams et al. is

$$h(t) = \begin{cases} \theta_1 t^{\gamma-1} & 0 \leq t < \tau \\ \theta_2 t^{\gamma-1} & t \geq \tau \end{cases} \tag{3.1}$$

The corresponding log-likelihood is

$$\begin{aligned} \log \mathscr{L}(\theta_1, \theta_2, \tau) =& G_1(\tau)\log(\theta_1) + G_2(\tau)\log(\theta_2) - T_1(\tau)\theta_1 - T_2(\tau)\theta_2 \\ &+ \sum_{i=1}^{n} (\gamma - 1)\delta_i \log t_i \end{aligned} \tag{3.2}$$

where $G_1(\tau) = \sum_{i=1}^{n} \mathbf{1}\{t_i < \tau, \delta_i = 1\}, G_2(\tau) = \sum_{i=1}^{n} \mathbf{1}\{t_i \geq \tau, \delta_i = 1\}, T_1(\tau) = \frac{1}{\gamma}\sum_{i=1}^{n} \min(t_i^\gamma, \tau^\gamma)$, and $T_2(\tau) = \frac{1}{\gamma}\sum_{i=1}^{n} \max(0, t_i^\gamma - \tau^\gamma)$. Notation has been changed slightly for consistency in this dissertation and to reflect the change from Williams et al.'s accommodations for data with staggered entry [14]. The parameters of the Weibull distribution include scale parameter θ and shape parameter γ. The shape parameter is assumed known and does not change despite the presence of a change-point. In the case of a Weibull hazard function with one change point, the scale parameters before and after the change point are θ_1 and θ_2 respectively, both unknown. Time (in our case question) is represented by t and an indicator for censoring is represented by δ taking the value $\delta = 1$ when the event (survey dropout) occurred and $\delta = 0$ when the subject was censored (here, survey completion). Williams et al. consider only fixed Type I censoring at the end of the study. This translates directly to the survey setting where participants who complete the survey are censored at the last question; in other words, only Type I right censoring is possible for survey data.

The basis of the test for one change-point is the profile log-likelihood ratio test statistic

$$\Lambda_n(\tau) = G_1(\tau)\log\left(\frac{G_1(\tau)}{T_1(\tau)} \times \frac{1}{\hat{\theta}}\right) + G_2(\tau)\log\left(\frac{G_2(\tau)}{T_2(\tau)} \times \frac{1}{\hat{\theta}}\right), \tag{3.3}$$

where $\hat{\theta} = \frac{\sum_{i=1}^{n} \mathbf{1}\{\delta_i=1\}}{(1/\gamma)\sum_{i=1}^{n} t_i^{\gamma}}$. This statistic is calculated for each potential change point within a pre-specified range $[a, b]$. Williams uses a Taylor series expansion to rewrite $\Lambda_n(\tau)$ as a composite of statistics $Z_n(\tau)$ and $R_n(\tau)$, namely

$$2\Lambda_n(\tau) = Z_n^2(\tau) + R_n(\tau), \tag{3.4}$$

where $Z_n(\tau) = \left(\frac{T_1(\tau)}{G_1(\tau)} - \frac{T_2(\tau)}{G_2(\tau)}\right)\left(\frac{G_1(\tau)G_2(\tau)}{G_1(\tau)+G_2(\tau)}\right)^{1/2}\left(\frac{G_1(\tau)+G_2(\tau)}{T_1(\tau)+T_2(\tau)}\right)$. They prove in Proposition 2.1 that $\sup_{a\le\tau\le b}|R_n(\tau)| \xrightarrow{p} 0$ and then transforms $Z(\tau)$ in order to prove in Corollary 2.1 that it is an Ornstein-Uhlenbeck process from which asymptotic critical values can be computed [14].

The null hypothesis of no change-point is rejected if $\sup_{a \le \tau \le b} Z^2(\tau)$ exceeds a the approximate critical value c^2 derived from Corollary 2.2, which is solved for by using the equation

$$(b^* - a^*)\frac{c}{\sqrt{2\pi}}e^{-c^2/2} = \alpha, \tag{3.5}$$

where

$$a^* = \log[g(b)/(g(0) - g(b))]$$

$$b^* = \log[g(a)/(g(0) - g(a))]$$

$$g(t) = P(T \ge t, \delta = 1),$$

significance level α, and by choosing the maximum root of this equation [14]. The change-point is determined to be at time τ where the maximum value of the test statistic $Z^2(\tau)$ occurred.

In order for change-point analysis to be applicable to the problem of testing for phases of attrition, it is necessary to be able to test for multiple change-points. Now, instead of τ we have τ_i where $i = 1, \ldots, K$. Only a few authors have provided such tests. Cai et al. and Kanavou propose tests specifically for two change-points [19,20]. Qian and Zhang propose a test for multiple change points using the piecewise linear hazard model that accounts for covariates and long-term survivors [18]. Although we are not focused on identifying long-term "survivors" in this dissertation, this would be an interesting avenue for future research in order to predict the survey adherence of participants with different characteristics. Han et al. suggest an approach within the piecewise exponential model allowing for small sample sizes, a challenge generally not faced in large online surveys [22]. Liu et al. propose a test for a multiple change-point model within the semi-parametric setting, however, this model relies on knowing the number of change-points a priori [15]. Goodman et al. propose a likelihood ratio and Wald-type test statistic for multiple change points within a piecewise constant hazard model (the discrete time hazard model described above is also known as a piecewise constant hazard model), the latter of which has a known limiting distribution [16,17]. Goodman et al.'s test allows for up to K change points by testing whether a model with an additional change-point has significantly better fit than the current model. He et al. propose a sequential test within the semi-parametric hazard model that allows for the comparison of groups [21]. The sequential nature of these tests limits their applicability because once a non-significant result is found, we can no longer test for a larger number of change-points even when we suspect the best fitting model should have additional change-points.

3.1.3 Aim

We applied a change-point hazard model in a new setting: survey data. We extended Williams' parametric approach using a Weibull hazard to test for multiple change-points [14]. A likelihood ratio test statistic was derived and we discuss the testing algorithm.

3.2 Methods

First, we extended Williams' approach by generalizing the likelihood ratio test statistic to account for multiple change-points in order to test our hypothesis that the current change-point model has significantly better fit than the no change-point model. The Weibull change-point pdf generalized for K change-points becomes

$$f(t) = \begin{cases} \theta_1 t^{\gamma-1}\exp\{-\frac{\theta_1}{\gamma}t^\gamma\} & 0 \le t < \tau_1 \\ \theta_2 t^{\gamma-1}\exp\{-\frac{\theta_2}{\gamma}(t^\gamma - \tau_1^\gamma) - \frac{\theta_1}{\gamma}\tau_1^\gamma\} & \tau_1 \le t < \tau_2 \\ \vdots & \\ \theta_{K+1} t^{\gamma-1}\exp\{-\frac{\theta_{K+1}}{\gamma}(t^\gamma - \tau_K^\gamma) - \frac{\theta_K}{\gamma}(\tau_K^\gamma - \tau_{K-1}^\gamma) - \ldots - \frac{\theta_1}{\gamma}\tau_1^\gamma\} & t \ge \tau_K \end{cases} \tag{3.6}$$

with corresponding survival function

$$S(t) = \begin{cases} \exp\{-\frac{\theta_1}{\gamma}t^\gamma\} & 0 \le t < \tau_1 \\ \exp\{-\frac{\theta_2}{\gamma}(t^\gamma - \tau_1^\gamma) - \frac{\theta_1}{\gamma}\tau_1^\gamma\} & \tau_1 \le t < \tau_2 \\ \vdots & \\ \exp\{-\frac{\theta_{K+1}}{\gamma}(t^\gamma - \tau_K^\gamma) - \frac{\theta_K}{\gamma}(\tau_K^\gamma - \tau_{K-1}^\gamma) - \ldots - \frac{\theta_1}{\gamma}\tau_1^\gamma\} & t \ge \tau_K. \end{cases} \tag{3.7}$$

Using Equations 3.6 and 3.7, we construct the log-likelihood for two change-points in the following way:

$$\begin{aligned} \mathscr{L}_2 = &\prod_{t<\tau_1}\Big\{[\theta_1 t^{\gamma-1}\exp\{-\frac{\theta_1}{\gamma}t^\gamma\}]^\delta[\exp\{-\frac{\theta_1}{\gamma}t^\gamma\}]^{1-\delta}\Big\} \\ &\times \prod_{\tau_1\le t<\tau_2}\Big\{[\theta_2 t^{\gamma-1}\exp\{-\frac{\theta_2}{\gamma}(t^\gamma-\tau_1^\gamma)-\frac{\theta_1}{\gamma}\tau_1^\gamma\}]^\delta[\exp\{-\frac{\theta_2}{\gamma}(t^\gamma-\tau_1^\gamma)-\frac{\theta_1}{\gamma}\tau_1^\gamma\}]^{1-\delta}\Big\} \\ &\times \prod_{t\ge\tau_2}\Big\{[\theta_3 t^{\gamma-1}\exp\{-\frac{\theta_3}{\gamma}(t^\gamma-\tau_2^\gamma)-\frac{\theta_2}{\gamma}(\tau_2^\gamma-\tau_1^\gamma)-\frac{\theta_1}{\gamma}\tau_1^\gamma\}]^\delta \\ &\times [\exp\{-\frac{\theta_3}{\gamma}(t^\gamma-\tau_2^\gamma)-\frac{\theta_2}{\gamma}(\tau_2^\gamma-\tau_1^\gamma)-\frac{\theta_1}{\gamma}\tau_1^\gamma\}]^{1-\delta}\Big\} \\ \mathscr{L}_2 = &\prod_{t<\tau_1}[\theta_1 t^{\gamma-1}]^\delta\exp\{-\frac{\theta_1}{\gamma}t^\gamma\}\prod_{\tau_1\le t<\tau_2}[\theta_2 t^{\gamma-1}]^\delta\exp\{-\frac{\theta_2}{\gamma}(t^\gamma-\tau_1^\gamma)-\frac{\theta_1}{\gamma}\tau_1^\gamma\} \\ &\times \prod_{t\ge\tau_2}[\theta_3 t^{\gamma-1}]^\delta\exp\{-\frac{\theta_3}{\gamma}(t^\gamma-\tau_2^\gamma)-\frac{\theta_2}{\gamma}(\tau_2^\gamma-\tau_1^\gamma)-\frac{\theta_1}{\gamma}\tau_1^\gamma\} \end{aligned}$$

$$\begin{aligned}
\log \mathscr{L}_2 =& \sum_{t<\tau_1}[\delta_i \log(\theta_1 t_i^{\gamma-1}) - \frac{\theta_1}{\gamma}t_i^{\gamma}] + \sum_{\tau_1\le t<\tau_2}[\delta_i \log(\theta_2 t_i^{\gamma-1}) - \frac{\theta_2}{\gamma}(t_i^{\gamma} - \tau_1^{\gamma}) - \frac{\theta_1}{\gamma}\tau_1^{\gamma}] \\
&+ \sum_{t\ge\tau_2}[\delta_i \log(\theta_3 t_i^{\gamma-1}) - \frac{\theta_3}{\gamma}(t_i^{\gamma} - \tau_2^{\gamma}) - \frac{\theta_2}{\gamma}(\tau_2^{\gamma} - \tau_1^{\gamma}) - \frac{\theta_1}{\gamma}\tau_1^{\gamma}] \\
\log \mathscr{L}_2 =& \log(\theta_1)\sum_{t<\tau_1}[\delta_i] + (\gamma-1)\sum_{t<\tau_1}[\delta_i \log(t_i)] - \frac{\theta_1}{\gamma}\sum_{t<\tau_1}[t_i^{\gamma}] + \log(\theta_2)\sum_{\tau_1\le t<\tau_2}[\delta_i] \\
&+ (\gamma-1)\sum_{\tau_1\le t<\tau_2}[\delta_i \log(t_i)] - \frac{\theta_2}{\gamma}\sum_{\tau_1\le t<\tau_2}[t_i^{\gamma} - \tau_1^{\gamma}] - \frac{\theta_1}{\gamma}\sum_{\tau_1\le t<\tau_2}[\tau_1^{\gamma}] + \log(\theta_3)\sum_{t\ge\tau_2}[\delta_i] \\
&+ (\gamma-1)\sum_{t\ge\tau_2}[\delta_i \log(t_i)] - \frac{\theta_3}{\gamma}\sum_{t\ge\tau_2}[t_i^{\gamma} - \tau_2^{\gamma}] - \frac{\theta_2}{\gamma}\sum_{t\ge\tau_2}[\tau_2^{\gamma} - \tau_1^{\gamma}] - \frac{\theta_1}{\gamma}\sum_{t\ge\tau_2}[\tau_1^{\gamma}] \\
\log \mathscr{L}_2 =& (\gamma-1)\sum[\delta_i \log(t_i)] + \log(\theta_1)\sum[\delta_i I(t_i<\tau_1)] + \log(\theta_2)\sum[\delta_i I(\tau_1 \le t_i < \tau_2)] \\
&+ \log(\theta_3)\sum[\delta_i I(t_i \ge \tau_2)] - \frac{\theta_1}{\gamma}\sum[\min(t_i^{\gamma}, \tau_1^{\gamma})] - \frac{\theta_2}{\gamma}\sum[\max(\min(t_i^{\gamma}, \tau_2^{\gamma}) - \tau_1^{\gamma}, 0)] \\
&- \frac{\theta_3}{\gamma}\sum[\max(t_i^{\gamma} - \tau_2^{\gamma}, 0)]. \qquad (3.8)
\end{aligned}$$

More generally, we can write this as

$$\begin{aligned}
\log \mathscr{L} =& (\gamma-1)\sum[\delta_i \log(t_i)] + \log(\theta_1)\sum[\delta_i I(t_i<\tau_1)] + \log(\theta_2)\sum[\delta_i I(\tau_1 \le t_i < \tau_2)] + \ldots \\
&+ \log(\theta_{K+1})\sum[\delta_i I(t_i \ge \tau_K)] - \frac{\theta_1}{\gamma}\sum[\min(t_i^{\gamma}, \tau_1^{\gamma})] - \frac{\theta_2}{\gamma}\sum[\max(\min(t_i^{\gamma}, \tau_2^{\gamma}) - \tau_1^{\gamma}, 0)] \\
&- \ldots - \frac{\theta_{K+1}}{\gamma}\sum[\max(t_i^{\gamma} - \tau_K^{\gamma}, 0)]. \qquad (3.9)
\end{aligned}$$

We substitute in the MLEs

$$\hat{\theta}_1 = \frac{G_1}{T_1}, \hat{\theta}_2 = \frac{G_2}{T_2}, \ldots, \hat{\theta}_{K+1} = \frac{G_{K+1}}{T_{K+1}}, \hat{\theta}_0 = \frac{\sum_{\ell=1\ldots K+1} G_\ell}{\sum_{\ell=1\ldots K+1} T_\ell}, \qquad (3.10)$$

where $G_\ell = \sum[\delta_i I(\tau_{\ell-1} \le t_i < \tau_\ell)]$ and $T_\ell = \frac{1}{\gamma}\sum[\max(\min(t_i^{\gamma}, \tau_\ell^{\gamma}) - \tau_{\ell-1}^{\gamma}, 0)]$, to get the log-likelihood in terms of τ, which is

$$\begin{aligned}
\log \mathscr{L} =& (\gamma-1)\sum[\delta_i \log(t_i)] + G_1 \log(\frac{G_1}{T_1}) + G_2 \log(\frac{G_2}{T_2}) + \ldots + G_{K+1}\log(\frac{G_{K+1}}{T_{K+1}}) \\
&- G_1 - G_2 - \ldots - G_{K+1}. \qquad (3.11)
\end{aligned}$$

Similarly, the log-likelihood when there are no change-points is

$$\mathscr{L}_0 = \prod[\theta_0 t^{\gamma-1}\exp\{-\frac{\theta_0}{\gamma}t^{\gamma}\}]^{\delta}[\exp\{-\frac{\theta_0}{\gamma}t^{\gamma}\}]^{1-\delta} = \prod[\theta_0 t^{\gamma-1}]^{\delta}\exp\{-\frac{\theta_0}{\gamma}t^{\gamma}\}$$

$$\log\mathscr{L}_0 = \sum[\delta_i\log(\theta_0 t_i^{\gamma-1}) - \frac{\theta_0}{\gamma}t_i^{\gamma}] = \log(\theta_0)\sum[\delta_i] + (\gamma-1)\sum[\delta_i\log(t_i)] - \frac{\theta_0}{\gamma}\sum[t_i^{\gamma}]. \tag{3.12}$$

Using Equations 3.8 and 3.12, we construct the log-likelihood ratio test statistic as

$$\begin{aligned}
\log(\frac{\mathscr{L}_2}{\mathscr{L}_0}) =& \log\mathscr{L}_2 - \log\mathscr{L}_0 = (\gamma-1)\sum[\delta_i\log(t_i)] + \log(\theta_1)\sum[\delta_i I(t_i<\tau_1)] \\
&+ \log(\theta_2)\sum[\delta_i I(\tau_1\le t_i<\tau_2)] + \log(\theta_3)\sum[\delta_i I(t_i\ge\tau_2)] - \frac{\theta_1}{\gamma}\sum[\min(t_i^{\gamma},\tau_1^{\gamma})] \\
&- \frac{\theta_2}{\gamma}\sum[\max(\min(t_i^{\gamma},\tau_2^{\gamma})-\tau_1^{\gamma},0)] - \frac{\theta_3}{\gamma}\sum[\max(t_i^{\gamma}-\tau_2^{\gamma},0)] - \log(\theta_0)\sum[\delta_i] \\
&- (\gamma-1)\sum[\delta_i\log(t_i)] + \frac{\theta_0}{\gamma}\sum[t_i^{\gamma}] \\
&\text{Let } G_1 = \sum[\delta_i I(t_i<\tau_1)], G_2 = \sum[\delta_i I(\tau_1\le t_i<\tau_2)], G_3 = \sum[\delta_i I(t_i\ge\tau_2)], \\
&T_1 = \frac{1}{\gamma}\sum[\min(t_i^{\gamma},\tau_1^{\gamma})], T_2 = \frac{1}{\gamma}\sum[\max(\min(t_i^{\gamma},\tau_2^{\gamma})-\tau_1^{\gamma},0)], \\
&T_3 = \frac{1}{\gamma}\sum[\max(t_i^{\gamma}-\tau_2^{\gamma},0)] \\
\log(\frac{\mathscr{L}_2}{\mathscr{L}_0}) =& G_1\log(\theta_1) + G_2\log(\theta_2) + G_3\log(\theta_3) - (G_1+G_2+G_3)\log(\theta_0) - T_1\theta_1 - T_2\theta_2 - T_3\theta_3 \\
&+ (T_1+T_2+T_3)\theta_0 \\
=& G_1\log(\frac{\theta_1}{\theta_0}) + G_2\log(\frac{\theta_2}{\theta_0}) + G_3\log(\frac{\theta_3}{\theta_0}) + T_1(\theta_0-\theta_1) + T_2(\theta_0-\theta_2) + T_3(\theta_0-\theta_3).
\end{aligned} \tag{3.13}$$

We substitute the MLEs for $\theta_1, \theta_2, \theta_3$, and θ_0 (see Equation 3.10) into Equation 3.13 to get the test statistic

$$\begin{aligned}
\log(\frac{\mathscr{L}_2}{\mathscr{L}_0}) =& G_1\log(\frac{G_1}{T_1}\times\frac{T_1+T_2+T_3}{G_1+G_2+G_3}) + G_2\log(\frac{G_2}{T_2}\times\frac{T_1+T_2+T_3}{G_1+G_2+G_3}) \\
&+ G_3\log(\frac{G_3}{T_3}\times\frac{T_1+T_2+T_3}{G_1+G_2+G_3}) + T_1(\frac{G_1+G_2+G_3}{T_1+T_2+T_3} - \frac{G_1}{T_1})
\end{aligned}$$

$$
\begin{aligned}
&+ T_2(\frac{G_1+G_2+G_3}{T_1+T_2+T_3} - \frac{G_2}{T_2}) + T_3(\frac{G_1+G_2+G_3}{T_1+T_2+T_3} - \frac{G_3}{T_3}) \\
=&G_1 \log(\frac{G_1}{T_1} \times \frac{T_1+T_2+T_3}{G_1+G_2+G_3}) + G_2 \log(\frac{G_2}{T_2} \times \frac{T_1+T_2+T_3}{G_1+G_2+G_3}) \\
&+ G_3 \log(\frac{G_3}{T_3} \times \frac{T_1+T_2+T_3}{G_1+G_2+G_3}).
\end{aligned} \tag{3.14}
$$

More generally, we can write Equation 3.14 as

$$
\begin{aligned}
\Lambda =&G_1 \log(\frac{G_1}{T_1} \times \frac{\sum_{j=1}^{K+1} T_j}{\sum_{j=1}^{K+1} G_j}) + G_2 \log(\frac{G_2}{T_2} \times \frac{\sum_{j=1}^{K+1} T_j}{\sum_{j=1}^{K+1} G_j}) + \dots \\
&+ G_{K+1} \log(\frac{G_{K+1}}{T_{K+1}} \times \frac{\sum_{j=1}^{K+1} T_j}{\sum_{j=1}^{K+1} G_j})
\end{aligned} \tag{3.15}
$$

When applying this method, the first step is to plot overall dropout as suggested in Hochheimer et al. [4]. We then get an initial estimate of γ by fitting a Weibull distribution with no change-points to the data. If the plot suggests a change-point model would be more appropriate, we estimate γ based on this change-point model. For example, if a bar chart of the dropout at each question suggests a two change-point model, we estimate γ by fitting the negative log-likelihood with these two change-points to the data. Using that estimate, we maximize Equation 3.11 with respect to τ_i for each potential number of change-points, choosing the model that results in the smallest BIC. The values of τ_i from this chosen change-point model are used to again estimate γ by fitting the log-likelihood corresponding to the chosen change-point model. We then repeat the process of maximizing Equation 3.11 for each potential number of change-points, choose the change-point model with the smallest BIC, calculate the test statistic in Equation 3.15 for this model, and compare it to a critical value defined through Monte Carlo simulations [44].

BIC was chosen over other modes of model comparison, such as AIC, in order to apply a stricter penalty to models with a higher number of change-points and to avoid overfitting. We define BIC as $p\log(n) - 2\log\mathscr{L}$ where p is the number of parameters maximized in Equation 3.9 (including τs, θs, and γ), n is the sample size, and $\log \mathscr{L}$ is the value of the log-likelihood at the change-points τ_i that maximize Equation 3.11.

Using this model, we test the hypothesis that the best fitting change-point model has significantly better fit than the no change-point model. Seeing as too many phase transitions would render the results difficult to interpret, we introduce a maximum number of potential phases in a survey based on the length of the survey itself. As a rule of thumb, we tested for up to $(q-1)/4$ change-points, that is the number of survey questions minus the first question and divided by four. There is no meaningful interpretation for a change-point at the first question, thus it is not eligible to be a change-point. A ten question survey, for example, would be eligible for a maximum of two change-points.

Monte Carlo simulations are applied to obtain a critical value [44]. The null distribution for this test is that with no change-points but with the same amount of overall dropout and value of γ. We simulate 10,000 replications of the null distribution and calculate the value of the log-likelihood ratio test statistic (Equation 3.15) for each one. Using these 10,000 values of the test statistic, we calculate the percentile corresponding to the alpha level to get the critical value. In order to see whether our result is significant, we compare the test statistic calculated for our test data to this critical value. If our test statistic is more extreme, we reject the null hypothesis in favor of the current change-point model. Using the resulting values of θ_i, we determine whether the phases are of higher dropout or stable use relative to the surrounding dropout phases. This method was implemented using the R statistical software version 3.5.0 [23]. R code demonstrating how this method was implemented can be found in Appendix A.

Chapter 4

Testing for attrition phases using an exponential change-point hazard model

4.1 Introduction

4.1.1 Motivation

While the Weibull method introduced in Chapter 3 allows for a more flexible parametric hazard, assuming a constant hazard (i.e., exponential) model may more effectively identify changes in the hazard of dropping out. We derived a test for up to $(q-1)/4$ change-points using the exponential distribution.

4.1.2 Current methods

Goodman proposes a likelihood ratio and Wald-type test for multiple change points within a piecewise constant hazard model (the discrete time hazard model described above is also known as a piecewise constant hazard model) [16,17]. She applies the latter due to the fact that the limiting distribution for the likelihood ratio test statistic is unknown as discussed previously. Goodman's test allows for up to K change points, represented by the hazard function

$$h(t) = \begin{cases} \theta_1 & 0 \leq t < \tau_1 \\ \theta_2 & \tau_1 \leq t < \tau_2 \\ \vdots & \vdots \\ \theta_{K+1} & t \geq \tau_K. \end{cases} \tag{4.1}$$

The corresponding log-likelihood is

$$\begin{aligned} \log \mathscr{L}(\theta_1, \ldots, \theta_{K+1}, \tau_1, \ldots, \tau_K) = & X(\tau_1) \log \theta_1 + [X(\tau_2) - X(\tau_1)] \log \theta_2 + \ldots + [n_u - X(\tau_K)] \log \theta_{K+1} \\ & - \theta_1 \sum_{i=1}^{n} \min(t_i, \tau_1) - \theta_2 \sum_{i=1}^{n} (\min(t_i, \tau_2) - \tau_1) I(t_i > \tau_1) \\ & - \ldots - \theta_{K+1} \sum_{i=1}^{n} (t_i - \tau_K) I(t_i > \tau_K), \end{aligned} \tag{4.2}$$

where n_u is the total number of events (or survey dropouts) and $X(\tau) = \sum_{i=1}^{n} I(t_i < \tau)\delta_i$, the total number of events before time τ. Notation has changed slightly for consistency in this dissertation. The Wald type test statistic for the hypothesis $H_0 : \theta_K - \theta_{K+1} = 0$ is

$$X_W = \frac{(\hat{\theta}_K - \hat{\theta}_{K+1})^2}{Var(\hat{\theta}_K - \hat{\theta}_{K+1})} \sim \chi_1^2, \tag{4.3}$$

where the variance is derived from a Hessian matrix consisting of only the parameters included in the test statistic. In order to correct for multiple comparisons and choose the most parsimonious model, the significance level for the Kth hypothesis test is $\alpha^*(K) = \frac{\alpha}{2^{K-1}}$. In this way, each additional change point is subject to a more stringent significance level.

Estimates for $\hat{\theta}_K$ and $\hat{\theta}_{K+1}$ are found by first substituting the MLEs for all $\hat{\theta}$s;

$$\hat{\theta}_1 = \frac{X(\tau_1)}{\sum_{i=1}^{n} \min(t_i, \tau_1)}, \hat{\theta}_2 = \frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n} (\min(t_i, \tau_2) - \tau_1) I(t_i > \tau_1)}, \ldots,$$
$$\hat{\theta}_K = \frac{X(\tau_K) - X(\tau_{K-1})}{\sum_{i=1}^{n} (\min(t_i, \tau_K) - \tau_{K-1}) I(t_i > \tau_{K-1})}, \hat{\theta}_{K+1} = \frac{n_u - X(\tau_K)}{\sum_{i=1}^{n} (t_i - \tau_K) I(t_i > \tau_K)}; \quad (4.4)$$

into equation 4.2. The log-likelihood becomes

$$\begin{aligned}\log \mathscr{L}(\tau_1, \ldots, \tau_K) =& X(\tau_1) \log \left(\frac{X(\tau_1)}{\sum_{i=1}^{n} \min(t_i, \tau_1)} \right) \\ &+ [X(\tau_2) - X(\tau_1)] \log \left(\frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n} (\min(t_i, \tau_2) - \tau_1) I(t_i > \tau_1)} \right) + \ldots \\ &+ [X(\tau_K) - X(\tau_{K-1})] \log \left(\frac{X(\tau_K) - X(\tau_{K-1})}{\sum_{i=1}^{n} (\min(t_i, \tau_K) - \tau_{K-1}) I(t_i > \tau_{K-1})} \right) \\ &+ [n_u - X(\tau_K)] \log \left(\frac{n_u - X(\tau_K)}{\sum_{i=1}^{n} (t_i - \tau_K) I(t_i > \tau_K)} \right) - n_u. \quad (4.5)\end{aligned}$$

Equation 4.5 is then maximized with respect to $\tau_1, \ldots, \tau_K$, with the resulting estimates $\hat{\tau}_1, \ldots, \hat{\tau}_K$ used to maximize equation 4.2 with respect to $\theta_1, \ldots, \theta_{K+1}$. Finally, the estimates for θ_K and θ_{K+1} are used to calculate the test statistic in equation 4.3.

Goodman's testing approach starts with the no change-point model and compares it to a model with one change-point. If the test statistic is significant, the one change-point model becomes the null model and is then compared to a model with two change-points. This process continues until we no longer have a significant test statistic. The model with K change points that last produced a significant test statistic is considered the final model. This approach both detects the number of change-points and estimates all unknown parameters, including the change-points themselves [16,17].

4.1.3 Aim

We assumed the dropout rate in each attrition phase is constant, thus applying the exponential change-point hazard model. Within this framework, we derived a test statistic and describe the testing scheme for multiple change-points within the exponential hazard model.

4.2 Methods

Instead of testing Goodman's hypothesis, $H_0 : \theta_K = \theta_{K+1}$, we are interested in testing the same hypotheses as in Chapter 3, with a null hypothesis of no change-points and an alternative hypothesis of between one and $(q-1)/4$ change-points. In order to derive the appropriate test-statistic, we start with the exponential pdf

$$f(t) = \begin{cases} \theta_1 \exp\{-\theta_1 t\} & 0 \le t < \tau_1 \\ \theta_2 \exp\{-\theta_1\tau_1 - \theta_2(t-\tau_1)\} & \tau_1 \le t < \tau_2 \\ \vdots & \\ \theta_{K+1} \exp\{-\theta_1\tau_1 - \theta_2(\tau_2-\tau_1) - \ldots - \theta_{K+1}(t-\tau_K)\} & t \ge \tau_K \end{cases} \tag{4.6}$$

and corresponding survival function

$$S(t) = \begin{cases} \exp\{-\theta_1 t\} & 0 \le t < \tau_1 \\ \exp\{-\theta_1\tau_1 - \theta_2(t-\tau_1)\} & \tau_1 \le t < \tau_2 \\ \vdots & \\ \exp\{-\theta_1\tau_1 - \theta_2(\tau_2-\tau_1) - \ldots - \theta_{K+1}(t-\tau_K)\} & t \ge \tau_K. \end{cases} \tag{4.7}$$

Using the 4.6 and 4.7, we construct the log-likelihood for two change-points as

$$\begin{aligned}\mathscr{L}_2 = &\prod_{t<\tau_1}\left\{[\theta_1 \exp\{-\theta_1 t\}]^{\delta}[\exp\{-\theta_1 t\}]^{1-\delta}\right\}\\ &\times \prod_{\tau_1\le t<\tau_2}\left\{[\theta_2 \exp\{-\theta_1\tau_1-\theta_2(t-\tau_1)\}]^{\delta}[\exp\{-\theta_1\tau_1-\theta_2(t-\tau_1)\}]^{1-\delta}\right\}\\ &\times \prod_{t\ge\tau_2}\left\{[\theta_3 \exp\{-\theta_1\tau_1-\theta_2(\tau_2-\tau_1)-\theta_3(t-\tau_2)\}]^{\delta}\right.\\ &\left.\times [\exp\{-\theta_1\tau_1-\theta_2(\tau_2-\tau_1)-\theta_3(t-\tau_2)\}]^{1-\delta}\right\}\\ \mathscr{L}_2 = &\prod_{t<\tau_1}\theta_1^{\delta}\exp\{-\theta_1 t\}\prod_{\tau_1\le t<\tau_2}\theta_2^{\delta}\exp\{-\theta_1\tau_1-\theta_2(t-\tau_1)\}\\ &\times\prod_{t\ge\tau_2}\theta_3^{\delta}\exp\{-\theta_1\tau_1-\theta_2(\tau_2-\tau_1)-\theta_3(t-\tau_2)\}\\ \log\mathscr{L}_2 = &\log\theta_1\sum_{t<\tau_1}\delta_i+\log\theta_2\sum_{\tau_1\le t<\tau_2}\delta_i+\log\theta_3\sum_{t\ge\tau_2}\delta_i-\theta_1\sum_{i=1}^{n}\min(t_i,\tau_1)\\ &-\theta_2\sum_{i=1}^{n}(\min(t_i,\tau_2)-\tau_1)I(t_i\ge\tau_1)-\theta_3\sum_{t\ge\tau_2}(t_i-\tau_2). \end{aligned} \tag{4.8}$$

This can be generalized to

$$\begin{aligned}\log\mathscr{L} = &\log\theta_1\sum_{t<\tau_1}\delta_i+\log\theta_2\sum_{\tau_1\le t<\tau_2}\delta_i+\ldots+\log\theta_{K+1}\sum_{t\ge\tau_K}\delta_i-\theta_1\sum_{i=1}^{n}\min(t_i,\tau_1)\\ &-\theta_2\sum_{i=1}^{n}(\min(t_i,\tau_2)-\tau_1)I(t_i\ge\tau_1)-\ldots-\theta_{K+1}\sum_{t_i\ge\tau_K}(t-\tau_K). \end{aligned} \tag{4.9}$$

We substitute the MLEs for all θs;

$$\begin{aligned}\hat{\theta}_1 = \frac{X(\tau_1)}{\sum_{i=1}^{n}\min(t_i,\tau_1)}, \hat{\theta}_2 = \frac{X(\tau_2)-X(\tau_1)}{\sum_{i=1}^{n}(\min(t_i,\tau_2)-\tau_1)I(t_i\ge\tau_1)},\ldots,\\ \hat{\theta}_{K+1} = \frac{n_u-X(\tau_K)}{\sum_{i=1}^{n}(t_i-\tau_K)I(t_i\ge\tau_K)}, \hat{\theta}_0 = \frac{n_u}{\sum_{i=1}^{n}t_i}; \end{aligned} \tag{4.10}$$

where $X(\tau)=\sum_{i=1}^{n}\delta_i I(t_i<\tau)$ and $n_u=\sum_{i=1}^{n}\delta_i$, the total number of events or dropouts. Equation 4.9 becomes

$$\begin{aligned}\log \mathscr{L} =& X(\tau_1)\log\left(\frac{X(\tau_1)}{\sum_{i=1}^n \min(t_i,\tau_1)}\right) + [X(\tau_2)-X(\tau_1)]\log\left(\frac{X(\tau_2)-X(\tau_1)}{\sum_{i=1}^n(\min(t_i,\tau_2)-\tau_1)I(t_i\geq\tau_1)}\right)\\ &+\ldots+[n_u-X(\tau_K)]\log\left(\frac{n_u-X(\tau_K)}{\sum_{i=1}^n(t_i-\tau_K)I(t_i\geq\tau_K)}\right)-X(\tau_1)-[X(\tau_2)-X(\tau_1)]-\ldots\\ &-[n_u-X(\tau_K)]. \end{aligned} \tag{4.11}$$

Similarly, the log-likelihood for no change-points is

$$\begin{aligned}\mathscr{L}_0 &= \prod[\theta_0\exp\{-\theta_0 t\}]^\delta[\exp\{-\theta_0 t\}]^{1-\delta} = \prod\theta_0^\delta\exp\{-\theta_0 t\}\\ \log\mathscr{L}_0 &= \log\theta_0\sum\delta_i-\theta_0\sum t_i. \end{aligned}\tag{4.12}$$

Using Equations 4.8 and 4.12, we construct the log-likelihood ratio test statistic

$$\begin{aligned}\log\left(\frac{\mathscr{L}_2}{\mathscr{L}_0}\right) = \log\mathscr{L}_2-\log\mathscr{L}_0 =& X(\tau_1)\log\theta_1+[X(\tau_2)-X(\tau_1)]\log\theta_2+[n_u-X(\tau_2)]\log\theta_3\\ &-\theta_1\sum_{i=1}^n\min(t_i,\tau_1)-\theta_2\sum_{i=1}^n(\min(t_i,\tau_2)-\tau_1)I(t_i\geq\tau_1)\\ &-\theta_3\sum_{i=1}^n(t_i-\tau_2)I(t_i\geq\tau_2)-n_u\log\theta_0+\theta_0\sum_{i=1}^n t_i. \end{aligned}\tag{4.13}$$

We substitute the MLEs from Equation 4.10 for $\theta_1,\theta_2,\theta_3$ and θ_0 into Equation 4.13, giving us

$$\begin{aligned}\log\left(\frac{\mathscr{L}_2}{\mathscr{L}_0}\right) =& X(\tau_1)\log\left(\frac{X(\tau_1)}{\sum_{i=1}^n\min(t_i,\tau_1)}\right)+[X(\tau_2)-X(\tau_1)]\log\left(\frac{X(\tau_2)-X(\tau_1)}{\sum_{i=1}^n(\min(t_i,\tau_2)-\tau_1)I(t_i\geq\tau_1)}\right)\\ &+[n_u-X(\tau_2)]\log\left(\frac{n_u-X(\tau_2)}{\sum_{i=1}^n(t_i-\tau_2)I(t_i\geq\tau_2)}\right)-X(\tau_1)-X(\tau_2)+X(\tau_1)-n_u+X(\tau_2)\\ &-n_u\log\left(\frac{n_u}{\sum_{i=1}^n t_i}\right)+n_u\\ =& X(\tau_1)\log\left(\frac{X(\tau_1)}{\sum_{i=1}^n\min(t_i,\tau_1)}\right)+[X(\tau_2)-X(\tau_1)]\log\left(\frac{X(\tau_2)-X(\tau_1)}{\sum_{i=1}^n(\min(t_i,\tau_2)-\tau_1)I(t_i\geq\tau_1)}\right)\end{aligned}$$

$$+ [n_u - X(\tau_2)] \log \left(\frac{n_u - X(\tau_2)}{\sum_{i=1}^{n} (t_i - \tau_2) I(t_i \geq \tau_2)} \right) - n_u \log(\frac{n_u}{\sum_{i=1}^{n} t_i}). \quad (4.14)$$

More generally, we can write the log-likelihood ratio test statistic in Equation 4.14 as

$$X(\tau_1) \log \left(\frac{X(\tau_1)}{\sum_{i=1}^{n} \min(t_i, \tau_1)} \right) + [X(\tau_2) - X(\tau_1)] \log \left(\frac{X(\tau_2) - X(\tau_1)}{\sum_{i=1}^{n} (\min(t_i, \tau_2) - \tau_1) I(t_i \geq \tau_1)} \right) + \ldots$$
$$+ [X(\tau_K) - X(\tau_{K-1})] \log \left(\frac{X(\tau_K) - X(\tau_{K-1})}{\sum_{i=1}^{n} (\min(t_i, \tau_K) - \tau_{K-1}) I(t_i \geq \tau_{K-1})} \right)$$
$$+ [n_u - X(\tau_K)] \log \left(\frac{n_u - X(\tau_K)}{\sum_{i=1}^{n} (t_i - \tau_K) I(t_i \geq \tau_K)} \right) - n_u \log(\frac{n_u}{\sum_{i=1}^{n} t_i}). \quad (4.15)$$

In order to apply this method, we start by maximizing equation 4.11 with respect to τ_i for each potential number of change-points, choosing that with the smallest BIC. As in Chapter 3, we define BIC as $p \log(n) - 2 \log \mathscr{L}$. We then compare the best fitting change-point model to the no change-point model by comparing the statistic to a critical value, which is calculated using 10,000 Monte Carlo simulations of the null distribution (no change-points but same amount of overall dropout, see Chapter 3 for more details) [44]. If the test statistic is more extreme than the critical value, this suggests significant change-points at the τ_i that maximize the likelihood. The resulting values of θ_i are then used to determine whether the resulting attrition phases are of higher dropout or stable use relative to surrounding phases. This method was implemented using the R statistical software version 3.5.0 [23]. R code demonstrating how it was implemented can be found in Appendix A.

Chapter 5

Simulation study and data application

5.1 Introduction

In order to determine how well the methods proposed in Chapters 3 and 4 detect change-points, we simulated data from the Weibull distribution with multiple change points (Equation 3.6) and the exponential distribution with multiple change-points (Equation 4.6).

Simulating data for change-point distributions was complicated by the need to simulate data assuming a constant value of γ for each phase of the Weibull change-point distribution and to incorporate the discrete nature of survey data. We considered two methods of simulating data: an inverse CDF method and a memoryless method. In the survival setting, the relationship between the CDF and cumulative hazard function can be used to simulate data [45]. The alternative technique takes advantage of the memoryless property of survival data by simulating each phase from an independent Weibull or exponential distribution. We compared these methods using a separate simulation study, the details and conclusions of which are included in this chapter.

Next, we applied the chosen simulation method to perform a simulation study evaluating the performance of our proposed tests for change-points within the Weibull and exponential hazard functions. Finally, we introduce our test case data, a survey on patients' preferences surrounding the decision to be screened for breast, prostate, or colon cancer. All methods proposed in this dissertation were applied to this survey and we compared their performance in identifying attrition phases in a real-life dataset.

5.1.1 Aim

We performed a simulation study to choose the most accurate method for simulating data from change-point hazard models. Methods proposed in Chapters 3 and 4 are compared using a simulation study implementing the chosen method. We then applied all methods proposed in this dissertation to test case data and compared their performance in identifying attrition phases in real-life survey data.

5.2 Comparing simulation methods

5.2.1 Methods

Inverse CDF

In order to determine how well the method detects change-points, we wanted to simulate data from the Weibull and exponential change-point hazard models with multiple change points (Equations 3.6 and 4.6). In order to simulate data, we use the relationship of the CDF and cumulative hazard function $F(t) = 1 - \exp(-H(t))$ where $H(t) = \int h(t)dt$ and $h(t)$ is the hazard function.

For the Weibull distribution, the hazard function corresponding to Equation 3.6

is

$$h(t) = \begin{cases} \theta_1 t^{\gamma-1} & 0 \le t < \tau_1 \\ \theta_2 t^{\gamma-1} & \tau_1 \le t < \tau_2 \\ \theta_3 t^{\gamma-1} & t \ge \tau_2 \end{cases} \tag{5.1}$$

with cumulative hazard function

$$H(t) = \begin{cases} \frac{\theta_1}{\gamma} t^{\gamma} & 0 \le t < \tau_1 \\ \frac{\theta_1}{\gamma}\tau_1^{\gamma} + \frac{\theta_2}{\gamma}(t^{\gamma} - \tau_1^{\gamma}) & \tau_1 \le t < \tau_2 \\ \frac{\theta_1}{\gamma}\tau_1^{\gamma} + \frac{\theta_2}{\gamma}(\tau_2^{\gamma} - \tau_1^{\gamma}) + \frac{\theta_3}{\gamma}(t^{\gamma} - \tau_2^{\gamma}) & t \ge \tau_2. \end{cases} \tag{5.2}$$

Noting that $F(t) = U$ where U is a uniform random variable on $(0,1)$, we derive the inverse cumulative hazard function

$$H^{-1}(x) = \begin{cases} \left(\frac{\gamma}{\theta_1}x\right)^{1/\gamma} & 0 \le x < \frac{\theta_1}{\gamma}\tau_1^{\gamma} \\ \left(\frac{\gamma}{\theta_2}(x - \frac{\theta_1}{\gamma}\tau_1^{\gamma}) + \tau_1^{\gamma}\right)^{1/\gamma} & \frac{\theta_1}{\gamma}\tau_1^{\gamma} \le x < \frac{\theta_1}{\gamma}\tau_1^{\gamma} + \frac{\theta_2}{\gamma}(\tau_2^{\gamma} - \tau_1^{\gamma}) \\ \left(\frac{\gamma}{\theta_3}\left(x - \frac{\theta_1}{\gamma}\tau_1^{\gamma} - \frac{\theta_2}{\gamma}(\tau_2^{\gamma} - \tau_1^{\gamma})\right) + \tau_2^{\gamma}\right)^{1/\gamma} & x \ge \frac{\theta_1}{\gamma}\tau_1^{\gamma} + \frac{\theta_2}{\gamma}(\tau_2^{\gamma} - \tau_1^{\gamma}), \end{cases} \tag{5.3}$$

where $x \sim Exp(1)$ and $t = H^{-1}(-\log(1-U))$. In this case, we see that $x = -log(1-U) \sim Exp(1)$ so we can simulate random variables from the exponential distribution.

The R code for implementing $H^{-1}(x)$ for the Weibull distribution with two change-points can be found in Appendix B. This derivation as well as R code for implementing this within the exponential distribution is available from Walke [45].

Memoryless method

Another way to visualize this data is the scenario where we have an independent Weibull or exponential distribution for each attrition phase. In this case, we treat the initial phase as Weibull or exponential survival with fixed type I right censoring. If the participant "survives" to phase two, the second phase is treated as left truncated with a different scale parameter θ and fixed type I right censoring. Any subsequent phase is treated similarly.

For example, if the interval of time 0-20 is split into two phases with a change-point at time 10 and a person survives until time 11 we have: $P(t > 10)P(t = 11|t > 10) = P(t > 10)\frac{P(t=11)}{P(t>10)} = P(t = 11)$. In other words, this simulation method exploits the memoryless property of survival.

Thus, we simulate each phase from an independent Weibull or exponential distribution with selected θ_i. Participants whose simulated survival time is past the end of the time interval for each phase are considered surviving to that change-point and then an additional survival time is simulated for them in the next phase. Time in the survey is calculated as the sum of time in each phase, with those who survive all phases censored at the end. Example R code implementing this method for the Weibull hazard model with two change-points can be found in Appendix B.

Simulation study methods

Data were simulated using both proposed methods including two change-points and a sample size of 500. We simulated 10,000 datasets with either 500 or 1,000 simulated participants. γ was set at 2 for the Weibull distribution and θ was chosen so that there was either a smaller or more pronounced difference in the dropout rate between attrition phases. In order to mimic survey data, simulated dropout time was truncated and simulated participants with a dropout time less than one were considered to have dropped out at the first question. The methods were assessed in terms of which more closely resembled the "true" simulated distribution, specifically the mean estimated values of γ and θ.

5.2.2 Results

Average estimates of γ and θ for the Weibull simulations using different simulation schemes, sample sizes, and attrition patterns are compared in Table 5.1. The memoryless method achieved estimates of γ closer to 2 than the inverse CDF method. Estimates of γ were most biased for this method when there was a severe dropout phase in between questions 10 and 16. The CDF method consistently underestimated γ, with values closer to 1 that resulted in much higher dropout than expected. Values of γ were most extreme for the CDF method when there was a severe dropout phase and phase transitions at either questions 4 and 10 or questions 10 and 16.

When simulating Weibull change-point data using the CDF method, values of θ were inflated. θ was particularly biased when the phase transitions occurred later in the survey at questions 10 and 16. Coupled with underestimated values of γ, this resulted in a dropout rate inflated by between 57% and 69%. While estimates of θ were less biased for simulated data using the memoryless method, θ_1 was overestimated while θ_2 and θ_3 were underestimated when there was a mild dropout phase in between the change-points. In simulation scenarios with a severe dropout phase, all values of θ were underestimated when using the memoryless method, coupled with an inflated value of γ. These deviations, however, did not affect the amount of dropout, which achieved the target dropout rate when we employed the memoryless method.

When simulating data for the exponential change-point distribution, both methods had similar performance and achieved the same amount of overall dropout (see Table 5.2).

Mild change			N=500		N=1000	
τ	Parameter	Actual	CDF	Memoryless	CDF	Memoryless
4,10	γ	2	1.30	1.74	1.30	1.73
	θ_1	0.02022	0.03476	0.02578	0.03483	0.02580
	θ_2	0.0198	0.08144	0.01437	0.08104	0.01436
	θ_3	0.00345	0.02400	0.00241	0.02380	0.00240
4,16	γ	2	1.46	1.81	1.46	1.81
	θ_1	0.02023	0.03162	0.02441	0.03161	0.02444
	θ_2	0.004958	0.01790	0.00473	0.01779	0.00471
	θ_3	0.0215	0.12967	0.00438	0.12768	0.00431
10,16	γ	2	1.37	1.87	1.36	1.86
	θ_1	0.00325	0.00954	0.00436	0.00960	0.00431
	θ_2	0.0198	0.11492	0.00705	0.11390	0.00681
	θ_3	0.0219	0.17534	0.00412	0.17175	0.00391
Severe change						
4,10	γ	2	1.09	2.10	1.09	2.09
	θ_1	0.00642	0.01224	0.00607	0.01228	0.00607
	θ_2	0.0323	0.020113	0.01238	0.19852	0.01215
	θ_3	0.00124	0.01578	0.00037	0.01535	0.00035
4,16	γ	2	1.54	2.21	1.54	2.21
	θ_1	0.00644	0.00947	0.00546	0.00945	0.00544
	θ_2	0.0081	0.02491	0.00318	0.02443	0.00312
	θ_3	0.0077	0.03406	0.00053	0.03287	0.00051
10,16	γ	2	0.98	2.45	0.97	2.43
	θ_1	0.00102	0.00518	0.00059	0.00525	0.00054
	θ_2	0.0323	0.53224	0.00367	0.52180	0.00316
	θ_3	0.0073	0.16854	0.00040	0.15975	0.00033

Table 5.1: Comparing simulation methods for the Weibull hazard with two change-points

5.2.3 Discussion

The results of this simulation study suggest that we should proceed with the memoryless simulation method for the Weibull distribution simulations. This method generated values of γ closer to the true value. This is important because this shape parameter affects the dropout rate. When γ is underestimated, the dropout rate more closely resembles exponential dropout, resulting in a higher dropout rate than desired. Furthermore, values of θ were less biased using the memoryless method for

Mild change			N=500		N=1000	
τ	Phase	θ	CDF	Memoryless	CDF	Memoryless
4,10	1	0.02022	0.02035	0.02034	0.02033	0.02031
	2	0.0198	0.02002	0.02006	0.02001	0.02001
	3	0.00345	0.00345	0.00345	0.00346	0.00346
4,16	1	0.02023	0.02040	0.02034	0.02037	0.02034
	2	0.004958	0.00496	0.00498	0.00498	0.00498
	3	0.0215	0.02174	0.02173	0.02175	0.02177
10,16	1	0.00325	0.00327	0.00326	0.00327	0.00326
	2	0.0198	0.02006	0.02004	0.02001	0.02002
	3	0.0219	0.02213	0.02213	0.02210	0.02221
Severe change						
4,10	1	0.00642	0.00642	0.00644	0.00643	0.00643
	2	0.0323	0.03290	0.03295	0.03286	0.03288
	3	0.00124	0.00125	0.00124	0.00124	0.00124
4,16	1	0.00644	0.00647	0.00646	0.00645	0.00644
	2	0.0081	0.00813	0.00815	0.00813	0.00815
	3	0.0077	0.00779	0.00778	0.00772	0.00776
10,16	1	0.00102	0.00103	0.00102	0.00102	0.00102
	2	0.0323	0.03290	0.03288	0.03279	0.03287
	3	0.0073	0.00734	0.00737	0.00733	0.00733

Table 5.2: Comparing simulation methods for the exponential hazard with two change-points

the Weibull distribution. When we are not concerned with simulating both shape and scale parameters (i.e., the exponential distribution), these methods performed similarly.

Although using the inverse CDF is a common method for generating data, this simulation study suggests that this approach is inadequate when simulating data from change-point hazard distributions with multiple change-points and at least one parameter in addition to a scale parameter, such as the Weibull or gamma distributions.

5.3 Simulation study comparing tests for change-points within the Weibull and exponential hazard models

5.3.1 Simulation methods

Data were simulated for a twenty question survey with no, one, two and three change-points using the memoryless simulation method. Although the simulation method did not affect estimates for the exponential distribution, we used the same simulation method for both distributions for consistency. The methods from Chapters 3 and 4 allow us to test for more general attrition phases and more than one attrition phase within a survey. Seeing as it is only possible to dropout at distinct points in a survey, simulated dropout time was truncated and simulated participants with a dropout time less than one were considered to have dropped out at the first question (as was done in the simulation study above). The hazard of dropping out was pre-specified based on the amount of desired dropout in each phase. Overall dropout for all simulations was 50%, representing high but realistic dropout for a web-based survey.

We varied the location of the change-points to see whether our tests consistently detected change-points when they occurred at different points throughout the survey; specifically at the beginning (question 4), in the middle (question 10), or towards the end (question 16). The differences between the θ_is, and thus the amount of dropout between phases, were varied to see if our tests detected change-points when differences in θ_i were both small and large. For each dropout pattern, we simulated 10,000 datasets for each potential sample size, 500 and 1,000. Participants could not reenter the survey once they dropped out and all simulated participants who completed the survey were censored at the last question.

Once these data were simulated, we calculated the BIC for up to four change-points (the maximum possible based on the $(q-1)/4$ rule), choosing that with the smallest BIC to compare to the null model with no change-points. Based on our simulation template, we tested only models identifying distinct change-points that were not at the first or last survey question. Thus, we ensured the values of τ_i that maximized the likelihood were distinct by verifying that $\tau_{i+1} - \tau_i \geq 1$, else we chose the model with one less change-point. Additionally, if the first or last questions were identified as change-points, we tested the model with one less change-point.

The null hypothesis was that there is only one phase, in other words that there are no change-points and θ_is are the same for all i. For each simulation, we compared the selected change-point model to the critical value from the Monte Carlo simulations (see Appendix C for R code demonstrating how this was implemented). To further penalize models with larger number of change-points, we applied Goodman's alpha spending function $\alpha^*(K) = \frac{\alpha}{2^{K-1}}$ and rejected the null hypothesis if the test statistic was more extreme than the percentile from the Monte Carlo simulations corresponding to $\alpha^*(K)$ [17]. This simulation study was implemented in R using the *optim* function to minimize the negative test statistics as opposed to maximizing the test statistics in Equations 3.15 and 4.15. A significant result, therefore, would be smaller than the $\alpha^*(K)^{\text{th}}$ percentile.

We evaluated these two methods in terms of type I error, sensitivity and power. Similar to Chapter 2, a type I error (α) was defined as when at least one change-point was detected when the known attrition pattern had no change-points (i.e. phases of attrition were detected when they do not exist) and sensitivity was defined as finding the correct number of change-points when they do exist. Power $(1-\beta)$ was defined as finding any number of change-points to be significantly better than the null model when the true model had at least one change-point. Ideally, we hoped to see a type I error level close to but not above the nominal level of 0.05, high sensitivity, and high power. We are currently unable to make scenarios to achieve a desired nominal rate of power or sensitivity (e.g., 80%). Additionally, we used histograms to visualize the distribution of τ, that is how often each survey item was chosen as a change-point, in order to assess and compare the accuracy of these methods.

When conducting this simulation study, we used a range of $(1, 20)$ to estimate one change-point and used starting points of 4 and 16 to estimate two change-points. Thus, we tested to see if a more informative starting range or value would increase the sensitivity to find exactly one or two change-points. We narrowed the range to $(\tau - 2, \tau + 2)$ for estimating one change-point and used the known change-points as the initial values when estimating two change-points. Seeing as the first step suggested in Hochheimer et al. is to visually inspect the dropout pattern, we would (and do in the next section) use more informative starting points based on this plot when applying these methods in a real life setting [4].

All simulations were completed using the R statistical software and histograms were created using the *ggplot2*, *reshape2*, *tidyr*, and *hablar* packages [23,46–49].

5.3.2 Results

Type I error for both methods and sample sizes can be found in Table 5.3. The Weibull method had slightly higher type I error than the exponential method but both achieved or were very close to the target type I error rate of 0.05. Sample size did not affect type I error.

	N=500	N=1000
Weibull	0.06	0.06
Exponential	0.05	0.05

Table 5.3: Comparing type I error for the Weibull and exponential tests

Power and sensitivity of the Weibull and exponential methods with a simulated sample size of 500 are compared in Table 5.4. The Weibull method achieved at least 80% power in 8/20 scenarios while the exponential method achieved at least 80% power in 9/20 cases. Power for the Weibull method ranged between 44% and 99% with all one change-point simulations except one achieving at least 80% power. There was a broader range in power for the exponential method from 5% to 99%.

			Weibull		Exponential	
# of change-points	τ	Dropout pattern	$1-\beta$	Sensitivity	$1-\beta$	Sensitivity
One	4	10%, 44.4%	0.86	0.38	0.05	0.03
		44.4%, 10%	DNC	DNC	DNC	DNC
		25%, 33.3%	0.98	0.74	0.99	0.56
		33.3%, 25%	0.99	0.88	0.99	0.62
	10	10%, 44.4%	0.94	0.83	0.88	0.44
		44.4%, 10%	0.95	0.53	DNC	DNC
		25%, 33.3%	0.63	0.07	0.12	0.09
		33.3%, 25%	0.87	0.29	0.58	0.52
	16	10%, 44.4%	DNC	DNC	0.99	0.96
		44.4%, 10%	0.90	0.63	0.24	0.20
		25%, 33.3%	DNC	DNC	0.99	0.92
		33.3%, 25%	0.99	0.72	0.98	0.87
Two	4, 10	5%, 44%, 6%	DNC	DNC	DNC	DNC
		15%, 30%, 16%	0.84	0.47	0.98	0.28
	4,16	5%, 44%, 6%	0.67	0.11	0.96	0.82
		15%, 30%, 16%	0.78	0.12	0.35	0.11
	10, 16	5%, 44%, 6%	DNC	DNC	0.72	0.53
		15%, 30%, 16%	0.49	0.05	0.85	0.44
Three	4, 10, 16	5%, 30%, 20%, 6%	0.44	0.38	0.90	0.26
		30%, 5%, 20%, 6%	0.49	0.36	0.65	0.64

Table 5.4: Comparing the Weibull and exponential tests, n=500

When the change-point occurred closer to the start of the survey at question 4, the Weibull method had higher sensitivity to detect exactly one change-point. When the change-point occurred later in the survey at question 10 or 16, the Weibull method had higher sensitivity to detect one change-point when there was a large difference in attrition (10% to 44% or vice versa) whereas the exponential method had higher sensitivity when there was a mild change in attrition (25% to 33% or vice versa). The exponential method had similar or better sensitivity as the Weibull method to detect two change-points except when a there was a modest dropout phase of 30% in between questions 4 and 10. When there was a large change in attrition at the beginning and end of the survey with a mild change in the middle for a total of three change-points, the Weibull method was more sensitive. When the dropout rate alternated between low and high at three change-points, the exponential method was more sensitive to detect exactly three change-points.

Table 5.5 displays the power and sensitivity of these two methods when the sample size was increased to 1,000. There was a larger range in power for the Weibull method from 19% to 98% while the exponential method again had a large range from 4% to 100%. Only 6/20 scenarios achieved at least 80% power for the Weibull method while half of the scenarios achieved at least 80% power when the exponential method was applied.

The Weibull method was more sensitive to detect exactly one change-point when it was simulated to occur at question 4. The exponential method was more sensitive to detect a change-point at later questions except when the dropout rate increased from 10% to 44% at question 10 and decreased from 44% to 10% at question 16. The exponential method was universally more sensitive to detect two change-points. When there were three change-points corresponding to a large change in attrition at the start and end of the survey and a mild change in the middle, the Weibull method was more sensitive. Meanwhile, three change-points with a dropout rate alternating between low and high were more often detected by the exponential method.

Figure 5.1 displays the distribution of τ when applying the Weibull test. In general, this method underestimated τ when the change-point occurred at question 4 whether this was the only change-point or the first of two or three change-points (see Figure 5.1a, 5.1d, 5.1e, and 5.1g). Question 10 was underestimated with some noise after the change-point (see Figure 5.1b, 5.1d, 5.1g) except when the two change-

# of change-points	τ	Dropout pattern	Weibull $1-\beta$	Weibull Sensitivity	Exponential $1-\beta$	Exponential Sensitivity
One	4	10%, 44.4%	0.56	0.03	0.04	0.02
		44.4%, 10%	DNC	DNC	1	0.98
		25%, 33.3%	0.87	0.36	0.97	0.35
		33.3%, 25%	0.98	0.72	0.99	0.53
	10	10%, 44.4%	0.50	0.44	0.69	0.12
		44.4%, 10%	0.96	0.21	0.99	0.53
		25%, 33.3%	0.37	<0.01	0.16	0.09
		33.3%, 25%	0.58	0.05	0.73	0.62
	16	10%, 44.4%	0.98	0.15	0.99	0.96
		44.4%, 10%	0.67	0.41	0.28	0.20
		25%, 33.3%	0.94	0.67	0.98	0.85
		33.3%, 25%	0.21	0.04	0.94	0.76
Two	4, 10	5%, 44%, 6%	0.96	0.10	0.99	0.97
		15%, 30%, 16%	0.72	0.24	0.94	0.44
	4,16	5%, 44%, 6%	0.51	0.02	0.92	0.83
		15%, 30%, 16%	0.56	0.04	0.49	0.17
	10, 16	5%, 44%, 6%	0.34	0.25	0.47	0.41
		15%, 30%, 16%	0.37	0.06	0.68	0.44
Three	4, 10, 16	5%, 30%, 20%, 6%	0.39	0.34	0.76	0.25
		30%, 5%, 20%, 6%	0.19	0.18	0.49	0.48

Table 5.5: Comparing the Weibull and exponential tests, n=1,000

points were simulated to occur at questions 10 and 16, in which case the method was unable to detect question 10 as the first change-point (Figure 5.1f). Question 16 was underestimated when it was the only change-point or the second of two change-points (Figure 5.1c, 5.1e, 5.1f) but overestimated when it was the third change-point (Figure 5.1g).

Although the distribution of τ had more variability using the exponential model, this method was more likely to choose the known change-point(s) (see Figure 5.2). Of the one change-point simulations, this method best estimated a change-point at question 16 (Figure 5.2c). When we simulated two change-points, the exponential method distinguished two separate change-points when they were further apart (Figure 5.2e) or later in the survey (Figure 5.2f). When two change-points occurred in the first half of the survey, this method had difficulty identifying the first change-point at question 4 (Figure 5.2d). In the three change-point simulations, the exponential test underestimated the first change-point and overestimated the last two change-points, however, it did identify three distinct change-points (Figure 5.2g).

Surprisingly, the use of more informative starting points for τ decreased the sensitivity of the Weibull method in many cases (Table 5.6). On the other hand, the sensitivity of the exponential method was almost equivalent or higher when using more informative starting points except in one case, when the dropout rate decreased slightly from 33% to 25% at question 16 (Table 5.6).

			Weibull		Exponential	
# of change-points	τ	Dropout pattern	Power	Sensitivity	Power	Sensitivity
One	4	10%, 44.4%	0.56	0.03	0.04	0.02
		44.4%, 10%	DNC	DNC	1	0.98
		25%, 33.3%	0.87	0.36	0.99	0.91
		33.3%, 25%	0.98	0.71	0.99	0.93
	10	10%, 44.4%	0.13	0.01	0.69	0.11
		44.4%, 10%	0.96	0.21	0.99	0.56
		25%, 33.3%	0.37	0.01	0.17	0.11
		33.3%, 25%	0.58	0.05	0.75	0.63
	16	10%, 44.4%	0.99	0.96	0.99	0.96
		44.4%, 10%	0.68	0.43	0.45	0.38
		25%, 33.3%	0.95	0.75	0.98	0.80
		33.3%, 25%	0.27	0.10	0.91	0.38
Two	4, 10	5%, 44%, 6%	0.84	0.74	0.99	0.99
		15%, 30%, 16%	0.79	0.40	0.98	0.74
	4,16	5%, 44%, 6%	0.51	0.02	0.92	0.83
		15%, 30%, 16%	0.56	0.04	0.49	0.17
	10, 16	5%, 44%, 6%	0.41	0.34	0.55	0.53
		15%, 30%, 16%	0.57	0.47	0.70	0.56

Table 5.6: Comparing the Weibull and exponential tests with informative starting points, n=1,000

5.3.3 Discussion

Overall, simulating and analyzing data assuming a constant hazard (the exponential model) performed better at identifying changes in the dropout rate than doing so using the Weibull model. Both methods had similar performance with a smaller sample size. When the sample size increased, the exponential method had higher sensitivity than the Weibull method, especially in simulations with two change-points in a pattern representing Eysenbach's three phases. In the online setting, it is reasonable to expect a sample size of 1,000 or more, making the exponential approach more appropriate. Although the Weibull model had higher sensitivity to detect three change-points when there was a large change in the dropout rate towards the beginning and end of the survey with a smaller change in the middle, histograms of the distribution of τ suggest that the exponential better identifies three distinct change-points (Figure 5.2g).

The histograms in Figures 5.1 and 5.2 suggest that the Weibull model better identifies a change-point in the first half of the survey for one and two change-point scenarios. The exponential model more accurately detects a change-point later in the survey when it is the only change-point but the Weibull model better identifies this later change-point when it is the latter of two change-points. The Weibull model better distinguishes two change points except when they both occur in the second half of the survey, in which case the exponential model better distinguishes these change-points. When there are three change-points, the exponential model identifies three distinct change-points whereas the distributions of the three change-points overlap using the Weibull model. This last finding is important as the goal of these tests is to allow for up to K change-points and as the number of change-points increased, the exponential model was generally more sensitive to detecting the correct number of change-points.

Although the Weibull method may appear more accurate, it is interesting to note that it rarely chose exactly question 4 or question 10 as a change-point. While the distributions of τ had higher variation when using the exponential method, the known change-points were more often chosen by this method. This translates to the higher sensitivity of the exponential model as the correct change-points were detected.

Concerned that the final model chosen would be that with the most change-points, we implemented a strict BIC penalizing a model with an additional change-point for estimating additional values of both θ and τ. We also used Goodman's alpha spending function $\alpha^*(K) = \frac{\alpha}{2^{K-1}}$ after finding that using a threshold of $\alpha = 0.05$ for all change-point models resulted in 100% power for the majority of simulations. Whereas type I error was closer to 0.10 when using $\alpha = 0.05$, the alpha spending function achieved a type I error level at or close to the nominal level for both methods.

In our simulation study, we did not simulate any phase to be shorter than four questions, thus we did not allow for any one question to be its own attrition phase (interpreted as a one-question spike or drop in attrition). Additionally, we did not simulate a phase transition later than question 16 and thus did not accept a one-question change-point at the end of the survey. These restrictions, while necessary in this case where we cannot individually inspect attrition plots for 10,000 simulations, are not reasonable when these methods are applied to actual survey data as we will see in the next section.

We also demonstrated how more informative starting values influence sensitivity. While this approach may seem biased by prior information, the first step to analyzing dropout is to plot dropout patterns. It follows that one would have an idea of where potentially significant changes in the dropout rate occurred and it would be logical to use those as starting points for where the test statistic might be maximized. Interesting enough, in our simulation study using this information was detrimental when applying the Weibull model.

While simulations for both methods had change-points at the same questions, data were simulated from different distributions (Weibull and exponential). We cannot conclude that assuming a constant hazard in each phase performs better in general before applying both methods to actual dropout data.

5.4 Data application

5.4.1 IDM module description

Our test case data are from an online survey entitled the Informed Decision-Making (IDM) module. This seventeen-question survey explored how patients approach decisions regarding screening for breast, colorectal, and prostate cancers. Here we focus exclusively on the results for colorectal cancer, where there were 1,249 participants. Questions addressed awareness of screening eligibility, screening options, primary concerns about cancer screening, and planned next steps [50]. This survey was designed by the Virginia Commonwealth University Department of Family Medicine and Population Health research team and administered from January to August of 2014 in twelve primary care practices throughout northern Virginia through the interactive online patient portal MyPreventiveCare [51–54]. More specific details regarding the survey, including screenshots of the questionnaire itself, can be found in Hochheimer et al. and Woolf et al. [4,55].

5.4.2 IDM application results

First, we inspected a plot of the number of dropouts at each question of the IDM module for colorectal cancer as suggested in Hochheimer et al., which can be seen in Figure 5.3 [4]. This plot suggests high attrition from Questions 3 to 5 with another spike at Question 9. We hypothesized that our methods from Chapter 2 would detect a dropout phase between Questions 3 and 9 while our methods from Chapters 3 and 4, which can detect more phases of attrition, would find four change-points at Questions 3, 5, 9 and 10.

Practical thresholds and existing statistical methods

As hypothesized, the 3% user-specified method detected a dropout phase between Questions 3 and 9. The GLMM was unable to detect any significant changes in dropout rate throughout the survey. Finally, DTSA only detected the start of the dropout phase. The results suggested a significant increase in the hazard of dropping out between Questions 2 and 3 but also that the dropout phase lasted until the end of the survey. This is inconsistent with the pattern seen in Figure 5.3, where we see visual proof of the stable use phase from Questions 10 to 17.

Test within a Weibull change-point hazard model

When applying the Weibull change-point model, we used the method outlined in Chapter 3 to estimate $\gamma = 2.22$. We tested for up to four change-points and found that the BIC for the four change-point model was smaller than that of the one, two and three change-point models. The change-points were detected at Questions 5, 9, 9 and 11. Whereas in our simulation study we did not allow one question to be its own attrition phase, in this case it is plausible. When looking at Figure 5.3, we see a spike in dropout at Question 9 relative to the questions around it. Using the alpha spending function, the critical value was calculated as the 0.625^{th} percentile of the test statistic calculated for 10,000 simulations of a sample with the same amount of dropout (716/1249 participants) and no change-points. The test statistic of -1029.42 was less extreme than the critical value of -1629.24, thus we failed to reject our null hypothesis and concluded that the no change-point model is the best fitting model.

Test within an exponential change-point hazard model

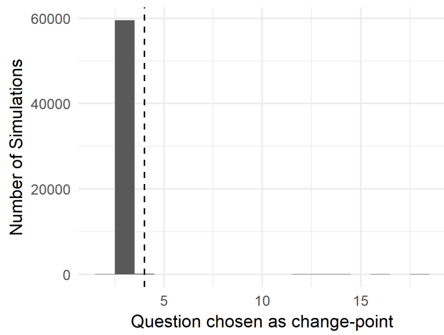
The test within the exponential model also suggested that the four change-point model had the best fit based on BIC. This method estimated change-points at Questions 6, 9, 9 and 11. As with the Weibull method, the critical value was calculated as the 0.625^{th} percentile of the test statistic calculated for simulated Monte Carlo samples with the same amount of overall dropout and no change-points. The four change-point model had a likelihood ratio test statistic of -2078.79, which was more extreme than the critical value of -1598.92. Thus, we concluded that the four change-point model is the best fit for these data.

5.4.3 Discussion

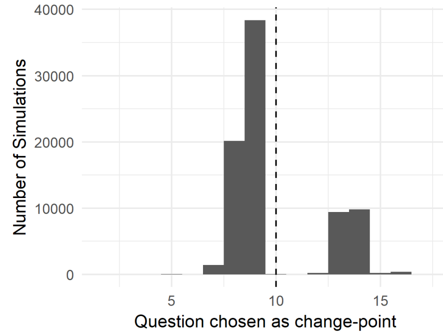
When applied to our test case data, the 3% user-specified threshold was the only method from Chapter 2 able to detect the dropout phase from Question 3 through Question 9. The GLMM was not sensitive enough to detect the attrition pattern highlighted in Figure 5.3 and DTSA was able to detect the abrupt increase in the hazard of dropping out between Questions 2 and 3 but not the abrupt decrease between Questions 9 and 10. These methods are limited in that they do not allow us to test for more than three attrition phases/two change-points.

While our methods from Chapters 3 and 4 allowed for more phases of attrition, only the exponential method allowed us to reject the null hypothesis that there were no attrition phases. The values of θ_i corresponding to change-points at Questions 6, 9, 9, and 11 suggest a phase of dropout from the start of the survey until Question 6, a spike in attrition at Question 9, and then another decrease in dropout at Question 11, with stable use lasting until the end of the survey. Using this conclusion, we suspect that there is content in Question 9 as well as Questions 1-5 driving the overall dropout rate of 57%. While we limited the number of change-points based on the survey length, a fifth change-point may be appropriate to further distinguish differences in the hazard of dropping out in the first few survey items.

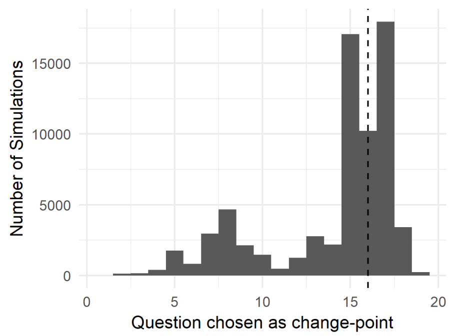
The test within the Weibull hazard function yielded a similar, although not significant, conclusion. The four change-point model had the best fit and suggested change-points at Questions 5, 9, 9 and 11.



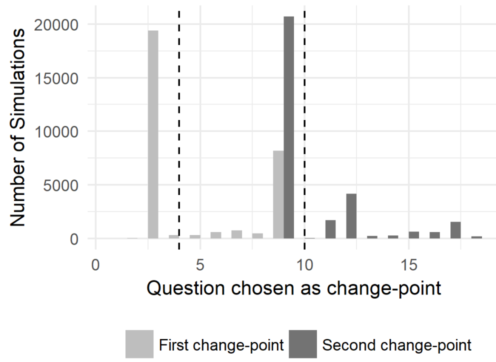
(a) Question 4



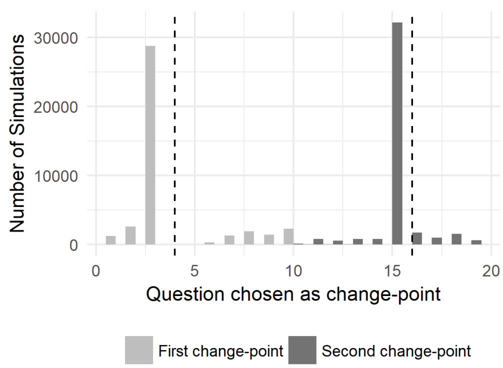
(b) Question 10



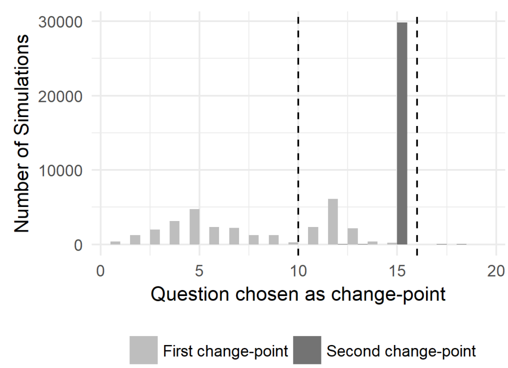
(c) Question 16



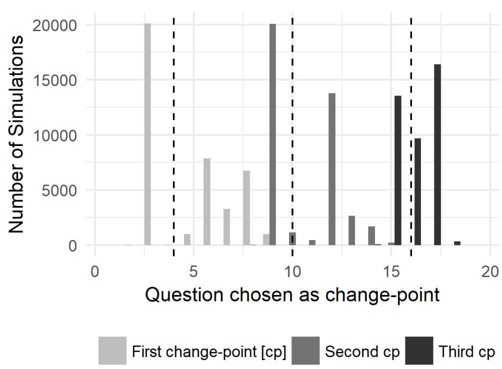
(d) Questions 4 & 10



(e) Questions 4 & 16

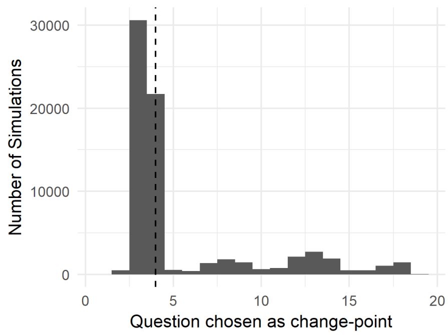


(f) Questions 10 & 16

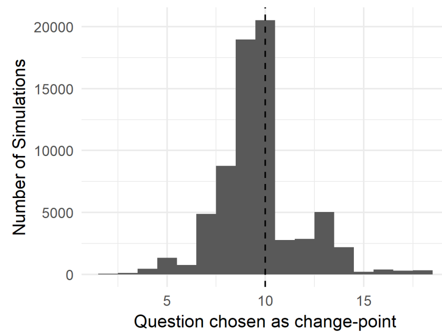


(g) Questions 4, 10, 16

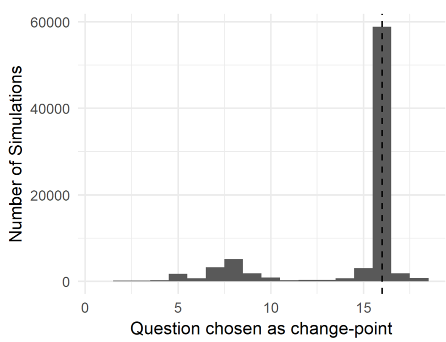
Figure 5.1: Questions chosen as change-points using the Weibull model



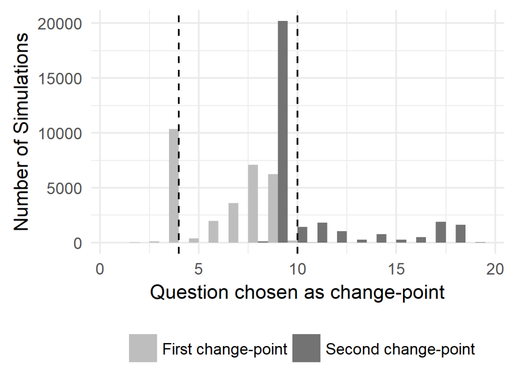
(a) Question 4



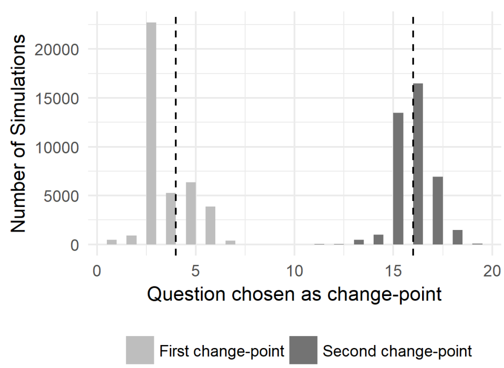
(b) Question 10



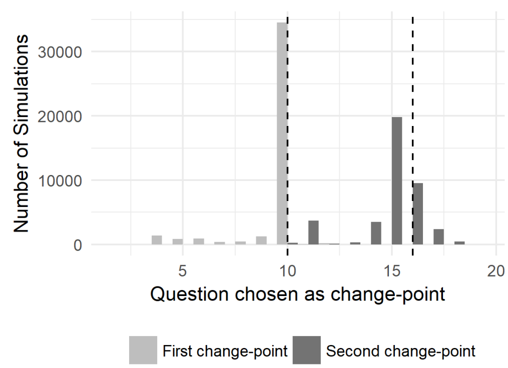
(c) Question 16



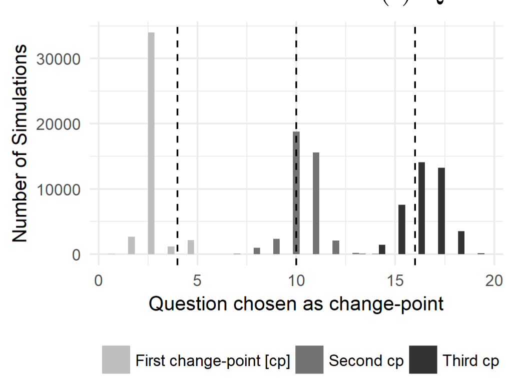
(d) Questions 4 & 10



(e) Questions 4 & 16



(f) Questions 10 & 16



(g) Questions 4, 10, 16

Figure 5.2: Questions chosen as change-points using the exponential model

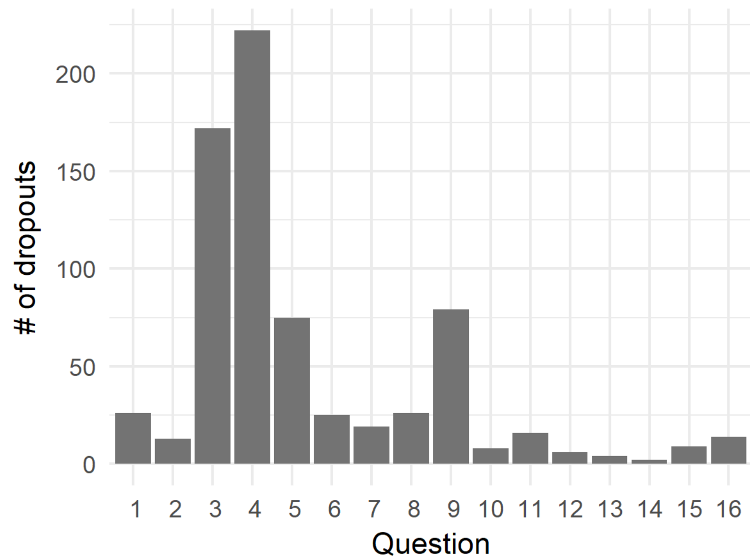


Figure 5.3: Dropouts at each question of the IDM

Chapter 6

Conclusion

6.1 Discussion

In this dissertation we proposed two novel approaches using existing methods and three novel methods to identify attrition phases in survey data. Using simulation studies and a test case data application to compare these methods, we observed variable results suggesting the exponential change-point hazard model as the best of these approaches.

The approaches in Chapter 2 were overly sensitive, resulting in high type I error. The GLMM had low sensitivity to detect exactly two phases (one change-point) and three phases (two change-points) where the second phase was a mild dropout phase. This method only had high sensitivity when there was a phase of severe attrition between two stable use phases. With the exception of the first question, the GLMM often accurately detected the start of the dropout phase but not the end of the dropout phase.

DTSA had high sensitivity to detect two but not three attrition phases (or one but not two change-points). Plots showing where phase transitions were detected, however, revealed the extreme sensitivity and inaccuracy of the method to choose the correct change-points.

Of the three methods proposed in Chapter 2, the least statistically rigorous performed the best. Choosing a low threshold of 3% dropout and searching for the first and last time the amount of dropout at a question exceeded this threshold had the highest sensitivity and accuracy to identify change-points both in the simulation study and IDM data.

Unlike the methods from Chapter 2, the tests within the Weibull and exponential change-point hazard models controlled type I error at around 0.05. When we performed a simulation study with a smaller sample size of 500, these methods had similar performance in terms of power and sensitivity whereas the exponential approach performed better than the Weibull test when the sample size was increased. This is meaningful in the field of online survey research because sample sizes are usually much larger, like in our test case data with 1,249 participants. Additionally, this result suggests that it is reasonable to assume a constant dropout rate (or hazard) in each phase.

Whereas the approaches in Chapter 2 were limited by their ability to detect high dropout at the first survey question, a change-point at this question was not interpretable nor accepted in Chapters 3 and 4. Thus, these two methods were not limited by their ability to pick up the start of a dropout phase right away since they model overall patterns instead of looking at question by question attrition.

When we plotted the distribution of τ, a first look suggested that the Weibull method did a better job of identifying one and two change-points except when the two change-points occurred in the second half of the survey. The exponential model, on the other hand, was able to clearly distinguish three distinct change-points. We realized, however, that while the distribution of τ was less variable using the Weibull method, estimated values of τ more accurately reflected the known values using the exponential method. In some cases the Weibull model never chose the known change-point even though the estimates were less variable. When applying the exponential model, the distribution of τ had higher variability but the known change-point was identified in more simulations. Finding the actual change-point may be the driving factor behind the increased sensitivity of the exponential method as these results likely had more extreme test statistics.

It is worth taking into consideration when applying these methods that the sensitivity of the Weibull method decreased when using more informative starting points. Perhaps we are better off not using information from plotting dropout a priori when applying this method. On the other hand, the sensitivity of the exponential method almost universally increased when using prior information. One should explore potential dropout patterns before using this test.

One of the biggest strengths of the tests within change-point hazard models was revealed in the test case data application. A bar plot of dropout revealed a spike at a single question and both methods correctly detected a dropout phase within this question. This question is now of interest as our team revises the IDM before fielding this survey to another sample of patients seeing as dropout had otherwise leveled out at this point in the survey.

Finally, our methods from Chapters 3 and 4 achieved reasonable power and sensitivity despite a high rate of dropout. By simulating a dropout rate of 50%, we know our methods are appropriate for the online setting where such high attrition is realistic, as in the case of the IDM module.

6.2 Limitations

We used a smaller sample size for the simulation study comparing the three methods in Chapter 2 than in Chapters 3 and 4. Seeing as these methods are very sensitive with a relatively small sample size, we suspect this would not be improved by using a larger sample size. The GLMM and DTSA specifically rely on pairwise significance tests where we would only be more likely to find a significant result with a larger sample.

All of the methods in Chapter 2 are limited by the number of change-points we are able to identify, which is only two. Thus, we looked to the methods proposed in Chapters 3 and 4 to extend our testing scheme to look for up to K change-points.

The Weibull method was originally appealing based on the flexibility of the hazard function, however, estimation of the shape parameter γ proved to be one of the largest challenges of this research. In the end, it appeared that the exponential model, which is equivalent to the Weibull model with a fixed $\gamma = 1$, had slightly better performance and was simpler to implement because we did not need to create an algorithm for estimating the shape parameter.

There are many alternative ways to estimate γ when applying the Weibull test in Chapter 3. A simple solution would've been to use the value of γ estimated from applying a no change-point Weibull distribution to the data as a starting point. Williams applies separate Weibull distributions to each phase to estimate γ for each phase [14]. An advantage of this approach is that it allows one to confirm that γ is constant across phases as assumed by the method. Estimation of these intervals becomes complicated in our case, however, by the use of discrete data. A hybrid of these approaches, where we fit a change-point model and allowed γ to vary between the phases, resulted in convergence challenges.

We suspect that more likelihood ratio tests have not been explored due to the difficulty of finding the limiting distribution for a change-point hypothesis test as mentioned in Chapter 3. In order to surpass this issue, we suggest the use of Monte Carlo simulations demonstrated in Appendix C. A drawback of this approach, however, is the computation time necessary to simulate the null distribution and estimate the test statistic for all simulations. The Monte Carlo simulations for our test case data applications took between 5-5.5 hours.

The methods discussed in this dissertation apply specifically to dropout attrition and do not address nonresponse or longitudinal attrition (see discussion in Hochheimer et al [4]). By assuming dropout monotonically accumulates throughout the survey, these methods do not account for the functionality to skip questions.

6.3 Future directions

Perhaps one of the most interesting findings of this research was the results of the simulation study comparing the CDF and memoryless simulation methods. We intend to dive deeper into this issue, perhaps exploring the effect of these different methods on estimating other parametric survival models.

A direct extension of this research would allow γ to vary by phase. Although this will likely pose convergence issues and increase computation time, this would allow for users to take full advantage of the flexibility of the Weibull distribution and may increase the accuracy of the results for this method.

While change-point models identify specific points and phases of increased or decreased attrition, characteristics of survey participants may provide insight into why rates have changed. Williams proposes the addition of covariates to his model [43]. Change-points could also be a function of these covariates and, thus, their location driven by characteristics of survey participants. We are interested in exploring the possibility of treating change-points as a function of covariates.

We could also apply Qian and Zhang's test accounting for covariates along with long-term survivors to predict the survey adherence of participants with different characteristics [18].

6.4 Conclusion

Of the methods proposed in this dissertation, our results suggest that assuming a constant hazard of dropping out within each phase and applying a likelihood ratio test using the exponential change-point hazard model is the best approach for identifying attrition phases within survey data. Although our proposed methods leave much room for improvement, our work contributes several new approaches to a relatively sparse field of research to identify attrition phases in survey data.

References

[1] G Eysenbach. The law of attrition. *Journal of Medical Internet research*, 7(1):e11, 2005.

[2] R Tortora. Reduction of respondent burden. In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Inc, 2014.

[3] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[4] CJ Hochheimer, RT Sabo, AH Krist, SH Woolf, T Day, and J Cyrus. Methods for evaluating respondent attrition in online survey data. *Journal of Medical Internet research*, 18(11):e301, 2016.

[5] G. Gurevich and A. Vexler. Change point problems in the model of logistic regression. *Journal of Statistical Planning and Inference*, 131(2):313–331, 2005.

[6] G. Gurevich and A. Vexler. Retrospective change point detection: From parametric to distribution free policies. *Communications in Statistics-Simulation and Computation*, 39(5):899–920, 2010.

[7] A. Vexler and G. Gurevich. Guaranteed local maximum likelihood detection of a change point in nonparametric logistic regression. *Communications in Statistics-Theory and Methods*, 35(4):711–726, 2006.

[8] A. Vexler and G. Gurevich. Average most powerful tests for a segmented regression. *Communications in Statistics-Theory and Methods*, 38(13):2214–2231, 2009.

[9] B. P. Carlin, A. E. Gelfand, and A. F. M. Smith. Hierarchical bayesian-analysis of changepoint problems. *Applied Statistics-Journal of the Royal Statistical Society Series C*, 41(2):389–405, 1992.

[10] P Groenewald and A Schoeman. Bayesian estimation in change-point hazard rate models: theory and methods. *South African Statistical Journal*, 35(1):95–121, 2001.

[11] S. J. Skates, D. K. Pauler, and I. J. Jacobs. Screening based on the risk of cancer calculation from bayesian hierarchical changepoint and mixture models of longitudinal markers. *Journal of the American Statistical Association*, 96(454):429–439, 2001.

[12] R. H. Loschi and F. R. B. Cruz. Extension to the product partition model: computing the probability of a change. *Computational Statistics & Data Analysis*, 48(2):255–268, 2005.

[13] J. A. Ferreira, R. H. Loschi, and M. A. Costa. Detecting changes in time series: A product partition model with across-cluster correlation. *Signal Processing*, 96:212–227, 2014.

[14] M. R. Williams and D. Y. Kim. A test for an abrupt change in weibull hazard functions with staggered entry and type i censoring. *Communications in Statistics-Theory and Methods*, 42(11):1922–1933, 2013.

[15] M. Liu, W. Lu, and Y. Shao. A monte carlo approach for change-point detection in the cox proportional hazards model. *Statistics in medicine*, 27(19):3894–3909, 2008.

[16] M. S. Goodman, Y. Li, and R. C. Tiwari. Survival analysis with change point hazard functions. *Harvard University Biostatistics Working Paper Series*, 40, 2006.

[17] M. S. Goodman, Y. Li, and R. C. Tiwari. Detecting multiple change points in piecewise constant hazard functions. *Journal of Applied Statistics*, 38(11):2523–2532, 2011.

[18] L. Qian and W. Zhang. *Multiple Change-Point Detection in Piecewise Exponential Hazard Regression Models with Long-Term Survivors and Right Censoring*, book section 18. Springer International Publishing, Switzerland, 2014.

[19] Xia Cai, Yubin Tian, and Wei Ning. Modified information approach for detecting change points in piecewise linear failure rate function. *Statistics & Probability Letters*, 125:130–140, 2017.

[20] Sofia Kanavou. Evaluation of the change-point piecewise exponential model. Master's thesis, Hasselt University Department of Biostatistics, 9 2015.

[21] Pei He, Liang Fang, and Zheng Su. A sequential testing approach to detecting multiple change points in the proportional hazards model. *Statistics in Medicine*, 32(7):1239–1245, 2013.

[22] Gang Han, Michael J. Schell, and Jongphil Kim. Improved survival modeling in cancer research using a reduced piecewise exponential approach. *Statistics in Medicine*, 33(1):59–73, 2014.

[23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[24] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[25] Torsten Hothorn, Frank Bretz, and Peter Westfall. Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363, 2008.

[26] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B:289–300, 1995.

[27] F Steele and E Washbrook. Discrete-time event history analysis, July 2013.

[28] Thomas Lumley. survey: analysis of complex survey samples, 2016. R package version 3.32.

[29] Thomas Lumley. Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19, 2004. R package verson 2.2.

[30] Y. C. Yao. Maximum-likelihood-estimation in hazard rate models with a change-point. *Communications in Statistics-Theory and Methods*, 15(8):2455–2466, 1986.

[31] I. Gijbels and U. Gurler. Estimation of a change point in a hazard function based on censored data. *Lifetime Data Analysis*, 9(4):395–411, 2003.

[32] C. Q. Wu, L. C. Zhao, and Y. H. Wu. Estimation in change-point hazard function models. *Statistics & Probability Letters*, 63(1):41–48, 2003.

[33] O.P. Rojas, F.S. Bulmaro, and J. Hernández. Modelo de riesgo con un punto de cambio y covariables dependientes del tiempo. *Revista Investigación Operacional*, 32:114–122, 2011.

[34] Jean-François Dupuy. Estimation in a change-point hazard regression model. *Statistics & probability letters*, 76(2):182–190, 2006.

[35] DE Matthews, VT Farewell, and R Pyke. Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *The Annals of Statistics*, 13(2):583–591, 1985.

[36] D. E. Matthews and V. T. Farewell. On testing for a constant hazard against a change-point alternative. *Biometrics*, 38(2):463–468, 1982.

[37] Jing Qin and Jianguo Sun. Statistical analysis of right censored failure-time data with partially specified hazard rates. *Canadian Journal of Statistics*, 25(3):325–336, 1997.

[38] L. Horvath. Tests for changes under random censorship. *Journal of Statistical Planning and Inference*, 69(2):229–243, 1998.

[39] E. Gombay and S. Q. Liu. A nonparametric test for change in randomly censored data. *Canadian Journal of Statistics-Revue Canadienne De Statistique*, 28(1):113–121, 2000.

[40] Dong-Yun Kim, Michael Woodroofe, and Yanhong Wu. Testing for a change in the hazard rate with staggered entry. *Communications in Statistics-Theory and Methods*, 33(9):2041–2058, 2004.

[41] J. F. Dupuy. Detecting change in a hazard regression model with right-censoring. *Journal of Statistical Planning and Inference*, 139(5):1578–1586, 2009.

[42] Xuan Chen and Michael Baron. Change-point analysis of survival data with application in clinical trials. *Open Journal of Statistics*, 4(09):663, 2014.

[43] Matthew R. Williams and Dong-Yun Kim. A test for abrupt change in hazard regression models with weibull baselines. *Journal of Statistical Planning and Inference*, 143(9):1566–1574, 2013.

[44] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, 2016.

[45] Rainer Walke. Example for a piecewise constant hazard data simulation in r. Report, Max Planck Institute for Demographic Research, 2010.

[46] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[47] Hadley Wickham. Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20, 2007.

[48] Hadley Wickham and Lionel Henry. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*, 2018. R package version 0.8.1.

[49] David Sjoberg. *hablar: Convert Data Types and Get Non-Astonishing Results*, 2018. R package version 0.1.0.

[50] S. H. Woolf, E. Zimmerman, A. Haley, and A. H. Krist. Authentic engagement of patients and communities can transform research, practice, and policy. *Health Aff (Millwood)*, 35(4):590–4, 2016.

[51] A. H. Krist, E. Peele, S. H. Woolf, S. F. Rothemich, J. F. Loomis, D. R. Longo, and A. J. Kuzel. Designing a patient-centered personal health record to promote preventive care. *BMC Med Inform Decis Mak*, 11:73, 2011.

[52] A. H. Krist, S. H. Woolf, S. F. Rothemich, R. E. Johnson, J. E. Peele, T. D. Cunningham, D. R. Longo, G. A. Bello, and G. R. Matzke. Interactive preventive health record to enhance delivery of recommended care: a randomized trial. *Ann Fam Med*, 10(4):312–9, 2012.

[53] A. H. Krist, S. H. Woolf, G. A. Bello, R. T. Sabo, D. R. Longo, P. Kashiri, R. S. Etz, J. Loomis, S. F. Rothemich, J. E. Peele, and J. Cohn. Engaging primary

care patients to use a patient-centered personal health record. *Ann Fam Med*, 12(5):418–26, 2014.

[54] J. M. Alpert, A. H. Krist, R. A. Aycock, and G. L. Kreps. Applying multiple methods to comprehensively evaluate a patient portal's effectiveness to convey information to patients. *J Med Internet Res*, 18(5):e112, 2016.

[55] S. H. Woolf, A. H. Krist, J. E. Lafata, R. M. Jones, R. R. Lehman, C. J. Hochheimer, R. T. Sabo, D. L. Frosch, B. J. Zikmund-Fisher, and D. R. Longo. Engaging patients in decisions about cancer screening: Exploring the decision journey through the use of a patient portal. *Am J Prev Med*, 54(2):237–247, 2018.

Appendix A

R code for applying the Weibull and exponential tests

A.1 R code for applying the test for change-points within the Weibull hazard function

```
#### pull in data ####
dta<-data.frame(time=question, censor=dropout)
#get initial estimate for gamma
gamma<-optim(c(2,0.0035),nll)$par[1]
#guess tau based on plot of dropouts
tau<-c(3,5,9,10)
gamma<-optim(c(rep(0.05,5),gamma),cp4.nll.full,control=list(maxit=10000))$par[6]

#### evaluate data ####
#LRT for one change-point
LRT.1cp.out<-optimise(LRT.1cp,c(2,6),gamma=gamma)
LRT.1cp.value<-LRT.1cp.out$objective
```

```
LRT.1cp.tau<-LRT.1cp.out$minimum
BIC.1cp.out<-BIC.1cp(tau=LRT.1cp.tau)

#LRT for two change-points
par2cp<-c(5,9)
LRT.2cp.out<-optim(par=par2cp, fn=LRT.2cp, gamma=gamma)
LRT.2cp.value<-LRT.2cp.out$value
LRT.2cp.tau<-LRT.2cp.out$par
BIC.2cp.out<-BIC.2cp(tau=LRT.2cp.tau)

#LRT for three change-points
par3cp<-c(5,8,10)
LRT.3cp.out<-optim(par=par3cp, fn=LRT.3cp, gamma=gamma)
LRT.3cp.value<-LRT.3cp.out$value
LRT.3cp.tau<-LRT.3cp.out$par
BIC.3cp.out<-BIC.3cp(tau=LRT.3cp.tau)

#LRT for four change-points
par4cp<-c(5,9,10,11)
LRT.4cp.out<-optim(par=par4cp, fn=LRT.4cp, gamma=gamma)
LRT.4cp.value<-LRT.4cp.out$value
LRT.4cp.tau<-LRT.4cp.out$par
BIC.4cp.out<-BIC.4cp(tau=LRT.4cp.tau)

#### which results are best? ####
c(BIC.1cp.out, BIC.2cp.out, BIC.3cp.out, BIC.4cp.out)
min(BIC.1cp.out, BIC.2cp.out, BIC.3cp.out, BIC.4cp.out)
```

```
#chooses 4cp


#### going back to look at gamma

tau<-LRT.4cp.tau

gamma<-optim(c(rep(0.05,5),gamma),cp4.nll.full,control=list(maxit=10000))$par[6]


#### evaluate data ####

#LRT for one change-point

LRT.1cp.out<-optimise(LRT.1cp,c(4,6),gamma=gamma)

LRT.1cp.value<-LRT.1cp.out$objective

LRT.1cp.tau<-LRT.1cp.out$minimum

BIC.1cp.out<-BIC.1cp(tau=LRT.1cp.tau)


#LRT for two change-points

par2cp<-c(5,11)

LRT.2cp.out<-optim(par=par2cp, fn=LRT.2cp, gamma=gamma)

LRT.2cp.value<-LRT.2cp.out$value

LRT.2cp.tau<-LRT.2cp.out$par

BIC.2cp.out<-BIC.2cp(tau=LRT.2cp.tau)


#LRT for three change-points

par3cp<-c(5,9,10)

LRT.3cp.out<-optim(par=par3cp, fn=LRT.3cp, gamma=gamma)

LRT.3cp.value<-LRT.3cp.out$value

LRT.3cp.tau<-LRT.3cp.out$par

BIC.3cp.out<-BIC.3cp(tau=LRT.3cp.tau)
```

```
#LRT for four change-points
par4cp<-c(6,9,9.1,11)
LRT.4cp.out<-optim(par=par4cp, fn=LRT.4cp, gamma=gamma)
LRT.4cp.value<-LRT.4cp.out$value
LRT.4cp.tau<-LRT.4cp.out$par
BIC.4cp.out<-BIC.4cp(tau=LRT.4cp.tau)

#### which results are best? ####
c(BIC.1cp.out, BIC.2cp.out, BIC.3cp.out, BIC.4cp.out)
min(BIC.1cp.out, BIC.2cp.out, BIC.3cp.out, BIC.4cp.out)
#chooses 4cp

#### simulate null distribution ####
## choose theta knowing dropout rate is 50%
nsim<-10000
theta<-0.0035
q<-
n<-
seed<-
alpha<-0.05
set.seed(seed)
dropout<-rep(NA,nsim)
thetaout<-rep(NA,nsim)
for(i in 1:nsim){
s<-n
sim.dta<-data.frame(id=c(1:n))
x1<-runif(s)
```

```
p1<-icdf(u=x1,t=theta)
sim.dta$time<-ifelse(p1>=q,q,p1)
sim.dta$censor<-ifelse(sim.dta$time==q,0,1)
thetaout[i]<-sum(sim.dta$censor)/((1/gamma)*sum(sim.dta$time^gamma))
dropout[i]<-sum(sim.dta$censor)
}
mean(thetaout)
mean(dropout)

## conduct monte carlo simulation using these values
## compare results
c(LRT.4cp.value, probs4cp.spending)
min(LRT.4cp.value, probs4cp.spending)

#getting values of theta
tau<-LRT.4cp.tau
optim(c(rep(0.05,5),gamma),cp4.nll.full,control=list(maxit=10000))$par[6]
```

A.2 R code for applying the test for change-points within the exponential hazard function

```
#### pull in data ####
dta<-data.frame(time=question, censor=dropout)

#### evaluate data ####
#LRT for one change-point
LRT.1cp.out<-optimise(LRT.1cp,c(2,6))
```

```
LRT.1cp.value<-LRT.1cp.out$objective

LRT.1cp.tau<-LRT.1cp.out$minimum

BIC.1cp.out<-BIC.1cp(tau=LRT.1cp.tau)


#LRT for two change-points

par2cp<-c(5,9)

LRT.2cp.out<-optim(par=par2cp, fn=LRT.2cp)

LRT.2cp.value<-LRT.2cp.out$value

LRT.2cp.tau<-LRT.2cp.out$par

BIC.2cp.out<-BIC.2cp(tau=LRT.2cp.tau)


#LRT for three change-points

par3cp<-c(5,8,10)

LRT.3cp.out<-optim(par=par3cp, fn=LRT.3cp)

LRT.3cp.value<-LRT.3cp.out$value

LRT.3cp.tau<-LRT.3cp.out$par

BIC.3cp.out<-BIC.3cp(tau=LRT.3cp.tau)


#LRT for four change-points

par4cp<-c(5,9,10,11)

LRT.4cp.out<-optim(par=par4cp, fn=LRT.4cp)

LRT.4cp.value<-LRT.4cp.out$value

LRT.4cp.tau<-LRT.4cp.out$par

BIC.4cp.out<-BIC.4cp(tau=LRT.4cp.tau)


#### which results are best? ####

c(BIC.1cp.out, BIC.2cp.out, BIC.3cp.out, BIC.4cp.out)
```

```
min(BIC.1cp.out, BIC.2cp.out, BIC.3cp.out, BIC.4cp.out)
#chooses 4cp

#### simulate null distribution ####
## choose theta knowing dropout rate is 50%
nsim<-10000
theta<-0.05
q<-
n<-
seed<-
alpha<-0.05
set.seed(seed)
dropout<-rep(NA,nsim)
thetaout<-rep(NA,nsim)
for(i in 1:nsim){
s<-n
sim.dta<-data.frame(id=c(1:n))
x1<-runif(s)
p1<-icdf(u=x1,t=theta)
sim.dta$time<-ifelse(p1>=q,q,p1)
sim.dta$censor<-ifelse(sim.dta$time==q,0,1)
### how many people dropping in each phase?
thetaout[i]<-sum(sim.dta$censor)/sum(sim.dta$time)
dropout[i]<-sum(sim.dta$censor)
}
mean(thetaout)
mean(dropout)
```

```
## conduct monte carlo simulation using these values
## compare results
c(LRT.4cp.value, probs4cp.spending)
min(LRT.4cp.value, probs4cp.spending)

# getting values of theta to determine type of phases
tau<-LRT.4cp.tau
optim(c(rep(0.05,5)),cp4.nll.full,control=list(maxit=10000))
```

Appendix B

R code for the CDF and memoryless simulation methods for the Weibull distribution

B.1 R code for implementing inverse hazard function

The R code for implementing $H^{-1}(x)$ with two change-points using the Weibull distribution is as follows:

```
x<-rexp(n)
first<-(theta[1]/gamma)*tau[1]^gamma #set interval 1
second<-first+(theta[2]/gamma)*(tau[2]^gamma-tau[1]^gamma) #set interval 2
t<-vector(mode="numeric",length=length(x)) #vector to hold times
for(i in 1:length(x)){
if(x[i]<first) t[i]<-((gamma/theta[1])*x[i])^(1/gamma) #if in first interval
if(x[i]>=first && x[i]<second)
t[i]<-((gamma/theta[2])*(x[i]-first)+tau[1]^gamma)^(1/gamma) #if in second interval
```

```
if(x[i]>=second)
t[i]<-((gamma/theta[3])*(x[i]-second)+tau[2]^gamma)^(1/gamma) #if in third interva
}
C <- rep(q,length(x)) #all censored at time q
time <- pmin(t,C)  #observed time is min of censored and true
censor <- as.numeric(time!=q) #if time isn't q then dropout
```

B.2 R code for implementing memoryless method

The R code for implementing this method with two change-points is as follows:

```
##user chooses these values
n<-
q<-
gamma<-
theta<-c(,,)
tau<-c(,)
seed<-1234
##memoryless simulation
icdf<-function(u,t){((-gamma/t)*log(1-u))^(1/gamma)}
set.seed(seed)
s<-n
alltime<-c(0,tau,q)
tau.diff<-alltime[2:length(alltime)]-alltime[1:(length(alltime)-1)]
#phase 1
sim.dta1<-data.frame(id=c(1:n))
x1<-runif(s)
p1<-icdf(u=x1,t=theta[1])
```

```
sim.dta1$phase1<-ifelse(p1<1,1,trunc(p1))
s<-sum(sim.dta1$phase1>=tau.diff[1])
#phase 2
sim.dta2<-subset(sim.dta1, phase1>=tau.diff[1], select=id)
x2<-runif(s)
p2<-icdf(u=x2,t=theta[2])
sim.dta2$phase2<-trunc(p2)
s<-sum(sim.dta2$phase2>=tau.diff[2])
#phase 3
sim.dta3<-subset(sim.dta2, phase2>=tau.diff[2], select=id)
x3<-runif(s)
p3<-icdf(u=x3,t=theta[3])
sim.dta3$phase3<-trunc(p3)
#combine
library(plyr)
sim.dta<-join_all(list(sim.dta1,sim.dta2,sim.dta3), by = 'id', type = 'full')
sim.dta$phase1<-ifelse(sim.dta$phase1>=tau.diff[1],tau.diff[1],sim.dta$phase1)
sim.dta$phase2<-ifelse(sim.dta$phase2>=tau.diff[2],tau.diff[2],sim.dta$phase2)
sim.dta$phase3<-ifelse(sim.dta$phase3>=tau.diff[3],tau.diff[3],sim.dta$phase3)
for(i in 1:n){
sim.dta$time[i]<-sum(sim.dta$phase1[i],sim.dta$phase2[i],sim.dta$phase3[i], na.rm=TRUE)
}
sim.dta$censor<-ifelse(sim.dta$time==q,0,1)
```

Appendix C

R code for applying the Monte Carlo simulations

C.1 R code for Monte Carlo simulations

```
### settings- user chooses these
q<-
n<-
gamma<-
theta<-
nsim<-
seed<-
alpha<-0.05
#suggest starting points for finding change-points
range1cp<-c(1,q)
par2cp<-c(,)
par3cp<-c(,,)
par4cp<-c(,,,)
```

```
### conducting simulation
# omit gamma when using exponential distribution
outstat<-matrix(NA,nrow=nsim,ncol=4)
is.error <- function(x) inherits(x, "try-error")

for(j in 1:nsim){
dta<-cp0.Msim(n=n,q=q,theta=theta)
outstat[j,1]<-optimise(LRT.1cp,range1cp,gamma=gamma)$objective

outall2<-try(optim(par=par2cp, fn=LRT.2cp, gamma=gamma),silent=T)
failed<-sum(is.error(outall2))
while(failed==1){
par2cp<-par2cp+0.01
outall2<-try(optim(par=par2cp, fn=LRT.2cp, gamma=gamma),silent=T)
failed<-sum(is.error(outall2))
}
outstat[j,2]<-outall2$value

outall3<-try(optim(par=par3cp, fn=LRT.3cp, gamma=gamma),silent=T)
failed<-sum(is.error(outall3))
while(failed==1){
par3cp<-par3cp+0.01
outall3<-try(optim(par=par3cp, fn=LRT.3cp, gamma=gamma),silent=T)
failed<-sum(is.error(outall3))
}
outstat[j,3]<-outall3$value
```

```
outall4<-try(optim(par=par4cp, fn=LRT.4cp, gamma=gamma),silent=T)
failed<-sum(is.error(outall4))
while(failed==1){
par4cp<-par4cp+0.01
outall4<-try(optim(par=par4cp, fn=LRT.4cp, gamma=gamma),silent=T)
failed<-sum(is.error(outall4))
}
outstat[j,4]<-outall4$value
}

# calculate percentiles using alpha spending function
alpha2<-alpha/2
alpha3<-alpha/4
alpha4<-alpha/8

probs1cp.spending<-quantile(outstat[,1], probs=alpha)
probs2cp.spending<-quantile(outstat[,2], probs=alpha2)
probs3cp.spending<-quantile(outstat[,3], probs=alpha3)
probs4cp.spending<-quantile(outstat[,4], probs=alpha4)
```

Vita

Camille Jo Hochheimer was born May 19, 1992 in Baltimore, Maryland and grew up in Catonsville, Maryland. She attended The University of Virginia where she graduated with honors with a BA in Spanish and a minor in mathematics. During her graduate career at Virginia Commonwealth University, she has served as a graduate research assistant for the Department of Family Medicine and Population Health.