January 2020

Some Recent Advances in Design of Bayesian Binomial Reliability Demonstration Tests

Suiyao Chen
University of South Florida

Some Recent Advances in Design of Bayesian Binomial Reliability Demonstration Tests

by

Suiyao Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Mingyang Li, Ph.D.
Tapas K. Das, Ph.D.
Devashish Das, Ph.D.
Lu Lu, Ph.D.
Qing Lu, Ph.D.

Date of Approval:
March 13, 2020

Keywords: Reliability Assurance, Optimal Design, Multiple Failure States, Acceptance Uncertainty, Multi-criteria Evaluation

Dedication

This dissertation is dedicated to my parents for their unconditional support in all my endeavors. I also dedicate this dissertation to my girlfriend Eileen for her encouragement and inspiration.

Acknowledgments

Chapter 2 of this dissertation is derived in part from my first-authored article published in Quality Engineering, Volume 29, 2017, copyright Taylor & Francis, available online: https://doi.org/10.1080/08982112.2017.1314493.

There are many people to whom I would like to express my heartfelt thanks. More specifically, I would like to thank people without whom this dissertation would not have been possible.

First of all, I would like to thank my major advisor Dr. Mingyang Li. I am more than grateful that he can give me the opportunity to join in the PhD program at University of South Florida. It is his guidance and patience that makes me go through those hard times. His innovation, passion and rigorous attitudes have set an great example for me in research.

The dissertation would not be successfully completed without my committee: Dr. Mingyang Li, Dr. Tapas K. Das, Dr. Devashish Das, Dr. Lu Lu and Dr. Qing Lu. Dr. Lu Lu has been extremely helpful and supportive for my research and she shares the authorship of my papers for the dissertation. Also, I would like to thank Dr. Tapas K. Das, Dr. Devashish Das and Dr. Qing Lu for their advice, guidance and support through my doctoral study.

Last but not the least, I would like to express my deepest gratitude to my family and friends. This dissertation would not have been possible without their warm love, continued patience, and endless support.

Table of Contents

List of Tables

List of Figures

Abstract

Reliability demonstration test (RDT) is one of important reliability assurance activities to demonstrate products' quality over time. Binomial RDT (BRDT) is one class of RDTs with appealing features, such as less failure monitoring efforts and fewer reliability modeling assumptions. Integrating with Bayesian method further allows prior knowledge incorporation for potential test sample size reduction. However, conventional designs often assume the binary failure states (i.e., success and failure) and consider a single objective of minimizing the testing cost with limited planning horizon. In this dissertation, a series of RDT designs are proposed and studied by advancing the conventional Bayesian BRDT designs from three aspects. First, the multi-state RDT designs are proposed to demonstrate product reliability either at multiple time periods or with multiple failure modes. They relax the simplified assumption of single time period or failure mode in conventional designs. Second, a BRDT design with extended planning horizon is proposed to consider both the testing acceptance uncertainty, and the anticipated cost impacts on subsequent reliability assurance activities, such as reliability growth and warranty services. Third, a multi-objective RDT design is formulated to simultaneously consider various RDT performance evaluation criteria, such as consumer's risk, producer's risk, acceptance probability and different cost components. The trade-offs among different conflicting objectives are leveraged by a Pareto Front approach. Case studies are provided to illustrate the proposed methodological frameworks and further demonstrate their superior performance and/or validity with comprehensive evaluation and analysis. The proposed work allows practitioners to develop more advanced Bayesian BRDTs to meet the increasingly complex reliability requirements from consumers and/or manufacturers.

Chapter 1

Introduction

1.1 Background

1.1.1 Reliability Definition and Significance

Reliability is defined as the probability that a component part, equipment, or system will satisfactorily perform its intended function under specified operating conditions, for a specified period of time (Meeker and Escobar, 2014). Three important elements in the reliability definition are intended function, specified operating conditions and specified period of time (Yang, 2007). Intended function is directly related with failure when the product cannot perform its intended functionality. For example, the failure of a light bulb occurs when the bulb is blown out and cannot perform its functionality of lighting. The specified operating conditions include the environmental conditions (e.g., electrical stress, thermal pressure, etc.), specified performance limits (e.g., operating time, operating frequency) and thoroughness of maintenance. The specified period of time can be linked with the designed product lifetime and the warranty service period, etc.

Reliability is of great importance because the consequences of product failure due to the lack of reliability can be catastrophic (Kapur et al., 2014). First, there can be critical safety concerns regarding the product reliability in industries such as aircraft, automobile, medical device, etc. For example, due to the failure of electronic control systems, the crash of Boeing 737-MAX8 airplane in 2018 led to the death of hundreds of passengers. Second, financial losses are often associated with a product failure from field operations such as maintenance, recall, warranty service, as well as the losses in market share, stock

prices and customer goodwill. For example, between 2009 and 2011, the world largest automobile manufacturer Toyota, had a series of recalls for over 14 million vehicles regarding their reliability issues in steering and acceleration. The company was fined by government agency for $50 million and lost around 10% growth in the U.S. market. Third, there is also legal liability associated with product reliability and failure. With customers suffering from the damages of product failures or the missing of reliability concern disclosures, the company may be sued for these charges. An example in 1999 is about the Japanese company Toshiba selling laptops with defective disk drives. Without the disclosure of risk that the disk drive controller chip might corrupt data at random, the company had to make a settlement of $2.1 billion deal under legislation. In addition to the major consequences involved for product failure and reliability, there are also many intangible issues such as company reputation, customer experience, public trust, etc. It is crucial for the public and all private parties to ensure the product reliability and prevent/mitigate the negative consequences caused by failures.

1.1.2 A Product Life Cycle View of Reliability Engineering Activities

Reliability engineering is a discipline to ensure the reliability of product when it is in specified operating conditions (O'Connor and Kleyner, 2012). It aims to apply engineering knowledge and analytical models to minimize the likelihood of product failures, identify causes of failure, estimate the reliability of products and designs, and take structured actions to prevent failures. The product life cycle for reliability engineering starts from product specification and planning to disposal/end of service. There will be three major stages including product design & development, production and field operation, which involves various reliability engineering activities. More specifically, in the design & development stage, there will be conceptual design for reliability, detailed design for reliability, reliability development and testing. As illustrated in Figure 1.1, there will be

different reliability engineering activities involved in different stages of the product life cycle (Kapur et al., 2014).

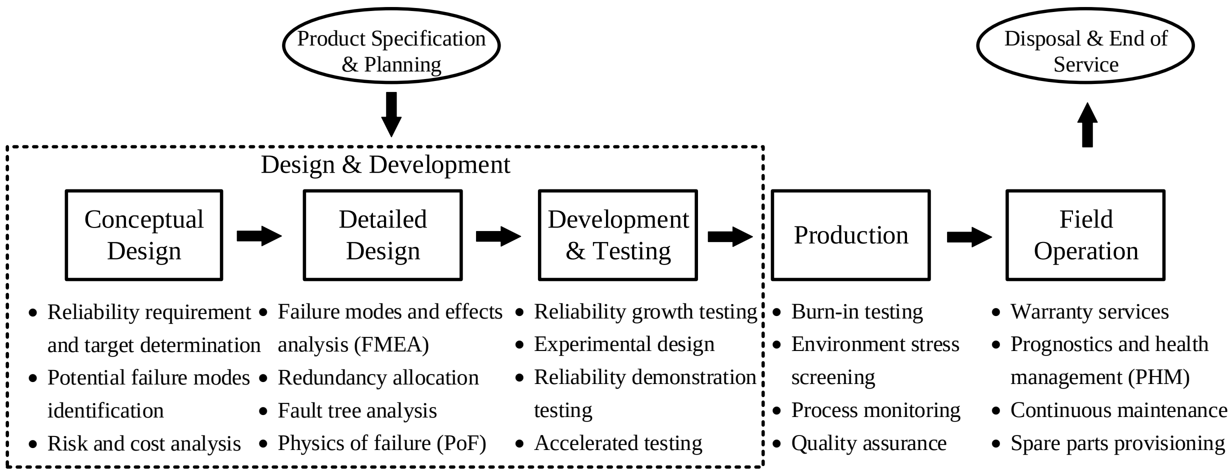


Figure 1.1: Reliability engineering activities throughout the product life cycle

Specifically, after the product specification and planning, the customer needs for product reliability should be understood and the business trends on product reliability requirements should be analyzed. Then, the design and development stage will be an essential phase to determine the detailed design for prototypes and testing for reliability assurance. The whole Design for reliability (DfR) (Crowe and Feinberg, 2001) process is to support product design and development from early conceptual design stage and ensure the satisfaction of customer expectation on product reliability throughout the product life cycle. In conceptual design, the reliability requirement and target will be determined and potential failure modes of product will be identified. In detailed design, different reliability engineering techniques are available to identify failure modes, investigate failure causes and mechanisms, and evaluate the anticipated effects of failure causes on product performance over time. For instance, failure modes and effects analysis (FMEA) is a method to review the components and substructures in the system as many as possible, identifying possible failure modes and justifying their causes and effects. Redundancy allocation can effectively reduce the impact of potential failures and improve the performance of products. Fault tree analysis is another method to analyze system safety and

reliability with identification of all possible combinations of component failures which will lead to a fault event, namely an abnormal and unexpected system state. After different failure causes are qualitatively and/or quantitatively identified, the physics of failure (PoF) method utilizes physical models and simulation to leverage the knowledge of failure mechanisms and causes to predict reliability and improve product performance.

Once the design is successfully completed, the prototypes will be built and delivered for design validation and testing to make sure all acceptance criteria can be satisfied and the product can be ready for production. In order to determine whether the product can meet the reliability requirements, a series of tasks need to be conducted through product development and reliability testing (Wasserman, 2002). Different reliability tests have different objectives. For example, when testing the prototypes to assure the product reliability, there may be deficiencies in product design and manufacturing/engineering procedures. Then, the product will be put back to the development stage for reliability growth test to identify the problems and take corrective actions so that the product reliability can be improved until it meets the requirement. Reliability demonstration test, which is the focus of this dissertation, is conducted after the reliability growth test to further demonstrate the capability of the product to meet the reliability requirement. Once the product passes the reliability demonstration test and goes to production stage, burn-in test and environmental stress screening may be conducted to detect and eliminate the early failures in trial production. Test plans are usually conducted under specific test conditions (e.g., environmental conditions such as stress and temperature, operating conditions such as testing time, frequency of operation, etc.) and the accelerated tests may be employed for all the reliability testing activities to shorten the overall testing period and increase efficiency. More details will be provided to review the tasks in reliability testing in the following section, which defines the scope of this dissertation.

Once the product reliability is demonstrated and the production process is validated after trail production, mass production can begin and products can finally be released to

the market. Several activities such as maintenance and warranty services need to be taken to ensure the product performance and monitor the product reliability in the field operation. For instance, maintenance activities retain or restore the functionality of product in a cost-effective manner, which includes preventative and predictive maintenance (Smith, 2017; Kapur et al., 2014) to routinely inspect and fix the minor problems in the product in order to prevent the major failures that may affect the product performance. These activities can help enhance the product service quality, reduce the critical breakdowns and minimize the loss due to product failures. Warranty services are usually provided by the manufacturers regarding the services for customers against product failure or defects such as replacement and repair (Murthy, 2006). A warranty is a term of a contract between customers and manufacturers or third-party insurers that assures the product performance under the covered period with additional cost at the time of purchase. Since a warranty involves additional services for the product, it will result in additional warranty service cost, which can be influenced by the product reliability. There are various types of warranties sold along with different products (Blischke and Murthy, 1992) and the analysis of warranty data (Murthy and Blischke, 2006) and its association with product reliability can be crucial for better customer experience, cost control and continuous improvements. At the end of service, the product will be disposed or recycled based on the user guide.

1.1.3 Tasks in Reliability Testing

This dissertation focuses on the design of a specific type of reliability testing activities. Before explicitly positioning the scope of the proposed work, it is necessary to briefly introduce different types of reliability tests and studies in the reliability engineering literature and practice, namely reliability growth test, reliability demonstration test, burn-in tests, environmental stress screening and accelerated tests.

In product design and development stage, the reliability growth test (RGT) (Kececioglu, 2002) will be performed to continuously improve reliability in system and components. A systematic process (e.g., test, analyze and fix(TAAF)) will be conducted to disclose deficiencies, identify failure modes and design weaknesses, and verify corrective actions to prevent recurrence in future operations. A reliability target is usually set for the reliability growth program and should be achieved during the testing procedure with proper planning and allocations of resources. The management strategy is used to determine whether to take corrective actions on specific failure modes or mechanisms. Three different approaches are used to implement the corrective actions including Test-Fix-Test (failure modes identified with fixes implemented after test phase.), Test-Find-Test (failure modes identified with fixes implemented during test phase.), and Test-Fix-Find-Test (failure modes identified with some fixes implemented during test phase and the others after test phase.) Many work has been accomplished in the existing literature to identify the reliability growth patterns and develop different modeling approaches, which may affect the analysis results as well as the decision making process. Duane (1964) were among the first to systematically describe the reliability growth patterns from empirically derived linear relationship between cumulative failure rate and test time on log-log scale. Crow (1982, 1984) developed practical methodologies for estimation of maximum potential reliability growth for specific product lifetime distributions such as Weibull and Exponential distributions and provided the confidence intervals for the estimation. The reliability growth with potential estimations would result in effective management strategy. Fries and Sen (1996) described the comprehensive survey of discrete reliability growth models (DRGM) which focused on the reliability data with success-failure outcomes. The characterizations and underlying assumptions of different models were discussed and compared. Ellner and Hall (2006) explored the functional expression for failure modes from corrective actions to develop reliability growth plan and growth path.

In order to test whether product reliability reaches the required level after reliability growth in product development stage, reliability demonstration test (RDT) (Hamada et al., 2008; Meeker and Escobar, 2014) will be conducted to demonstrate the capability of product to meet the reliability requirements of customers. To better design the RDT, various aspects, including test sample size, test duration, and testing process for different products, need to be considered. A more detailed review of RDT designs and applications will be provided in the next section.

Before final products being released to the market, they may contain a sub-population of products with early failures due to material defects, unverified design changes and/or manufacturing errors. To eliminate those product units from weak sub-populations, burn-in tests (Smith, 2017) and Environmental Stress Screening (ESS) (Kececioglu and Sun, 2003) will be performed to precipitate and eliminate the latent defects in products through the early manufacturing process. These tests can be failure-based to observe early failures of products or degradation-based to observe initial degradation for highly reliable products that are hard to fail. The stress levels, lifetime distributions and degradation patterns will be taken into consideration. For burn-in tests, Kuo and Kuo (1983) and (Kuo, 1984) provided comprehensive reviews of burn-in as an important screening method to predict and enhance field reliability for electronic equipment manufacturing and offer cost-effective opportunities. It pointed out the applicability of burn-in for both entire system and individual components and the feasibility to achieve optimal cost decisions. It also learned the side effects of burn-in where new problems might be created. Leemis and Beneke (1990) developed the taxonomy for burn-in and systematically discussed the lifetime distributions, data and logistics used in burn-in tests. Ye et al. (2012) proposed a degradation-based burn-in approach and jointly considered burn-in and maintenance activities to minimize the long-term average cost of Micro-Electro-Mechanical Systems. Tsai et al. (2010) discussed the optimal burn-in process for products with high reliability and rare failures by assuming the degradation path of products to follow a Wiener process.

Similarly, studies on ESS have also been extensively conducted. Wong (1990) introduced the method of using ESS data for reliability demonstration and reliability growth, which improved the accuracy and efficiency for reliability assurance. Reddy and Dietrich (1994) developed the models for 2-level ESS with mixed distributions of time-to-failure and determined the optimal screening duration with minimum life-cycle cost. Pohl and Dietrich (1995) considered the ESS method for multi-component systems and used mixed Weibull distribution to examine the change of screening strategies without identifying all the failures precipitated by screening. Yang (2002) extended the methodology of ESS with degradation measurements to screen the products with short duration and minimum life-cycle cost. The major difference between ESS and burn-in is that ESS is an accelerated process of stressing product in continuous cycles between environmental extremes (e.g., thermal cycling, random vibration) and is usually conducted at the assembly, module and system levels. Burn-in is a lengthy process of stressing a product at a specified constant temperature and usually conducted at component levels.

In all above reliability tests, it is often the case that if testing product units at normal operating conditions, it will take unexpected long time and significant resources to observe failures. To improve the efficiency and expedite the testing process, another important class of reliability tests is accelerated tests. Under the motivation that few product failures may happen at normal environmental or experimental conditions or within a reasonable testing period, it is necessary to find the way to assess the product reliability quickly and more cost-efficiently. It is also beneficial to detect the faults and possible failure modes of product within a short period of time. Therefore, the methods of accelerated tests (Nelson, 2009) have been developed to measure the product lifetime reliability (e.g., accelerated life tests (ALT)) or degradation (e.g., accelerated degradation tests (ADT)). The accelerated tests usually consider high stress testing conditions (e.g., usage rate, temperature, voltage, pressure, etc.) with shorten product lifetime and hasten degradation of product performance, which also apply to other reliability test designs. Nelson (1980)

provided the statistical models and methods to analyze ALTs with Weibull distribution and inverse power law and applied to step-stress data. Meeker and LuValle (1995) further developed the ALT models based on chemical models and reliability kinetics for failures of printed-circuit-board. Meeker et al. (1998) described the ALTs for high reliability products and connected degradation reliability models with acceleration models to quantify the failures in chemical reactions. Accelerated tests can be integrated into other reliability testing activities such as RGT, RDT, burn-in and ESS. Krasich (2007) discussed the idea of applying reliability growth methodology to ALT to calculate the cumulative time to failure regardless of the sequence of applied stresses. Luo et al. (2015) addressed the problem of apply RDT when facing competing failure modes and developed the accelerated life reliability demonstration method with reliability target allocation to each failure mode. Krasich (2004) also developed the accelerated method for reliability demonstration of product lifetime. Burn-in and ESS are usually associated with acceleration strategies to improve the test efficiency (Kececioglu and Sun, 2003).

1.1.4 Reliability Demonstration Test

The scope of this dissertation is focusing on the design of RDT to advance existing work and provide optimal RDT plans for practitioners. RDT is one of the key components in product design & development stage to validate that a product has the capability to meet the requirements from customers for achieving good performance over time. In order to assure the product reliability, a RDT is conducted with the goal to control the consumer's risk of receiving inferior products which actually pass the test. RDT has various applications (Kleyner, 2008) in many industries such as aircraft, medical device, automobile, electronics. It has been critical for many companies and industries to employ RDTs to secure the competitiveness of their products. Existing literature (Meeker and Escobar, 2014; Kececioglu, 2002; Hamada et al., 2008; Lee et al., 2015) has been working on RDT designs with different types of reliability data and statistical models, different evaluation

criteria (e.g., consumer's risk, producer's risk, etc.) to quantify the probabilistic measures of failure, and different learning paradigms (e.g, Sequential, Non-sequential) to conduct the RDT with improved efficiency and cost-effectiveness.

Optimal RDT design needs to be determined before conducting the test in order to assure the reliability requirements and achieve the cost-effectiveness. The flow chart illustrated in Figure 1.2 shows the considerations in optimal RDT design and details in design, implementation and decision making. After the product design & development stage, the product reliability has been developed to a certain level and needs to be tested to assure that the current reliability meets the requirement. The first step is to determine the RDT design learning paradigms. For example, whether to conduct the test sequentially or all at once will make a difference. Non-sequential testing needs to decide the test sample size at the beginning and observe the failures at the end of the test. Sequential testing requires the product units to be tested sequentially so that the reliability information can be updated through monitoring and inspection until the required reliability requirement can be demonstrated. The second step is to determine the evaluation criteria, which will be used for test implementation and decision making. For example, consumer's risk is linked with type-II error and need to be controlled within an acceptance level to reduce the probability of receiving inferior products by customers. Other criteria such as producer's risk, acceptance probability and test sample size may also be considered. Depending on whether prior knowledge can be incorporated, the evaluation criteria can be obtained under either non-Bayesian (e.g., Frequentist) or Bayesian framework. Compared with non-Bayesian framework, Bayesian framework may have the advantages of reducing the test sample size and costs when incorporating prior knowledge of product reliability. Then, different types of data will be collected based on the characteristics of products, such as failure count data, failure time data and degradation data. For different types of data, the statistical models should be selected to analyze the specific data structure and estimate the product reliability (Guo and Liao, 2011). For example, suc-

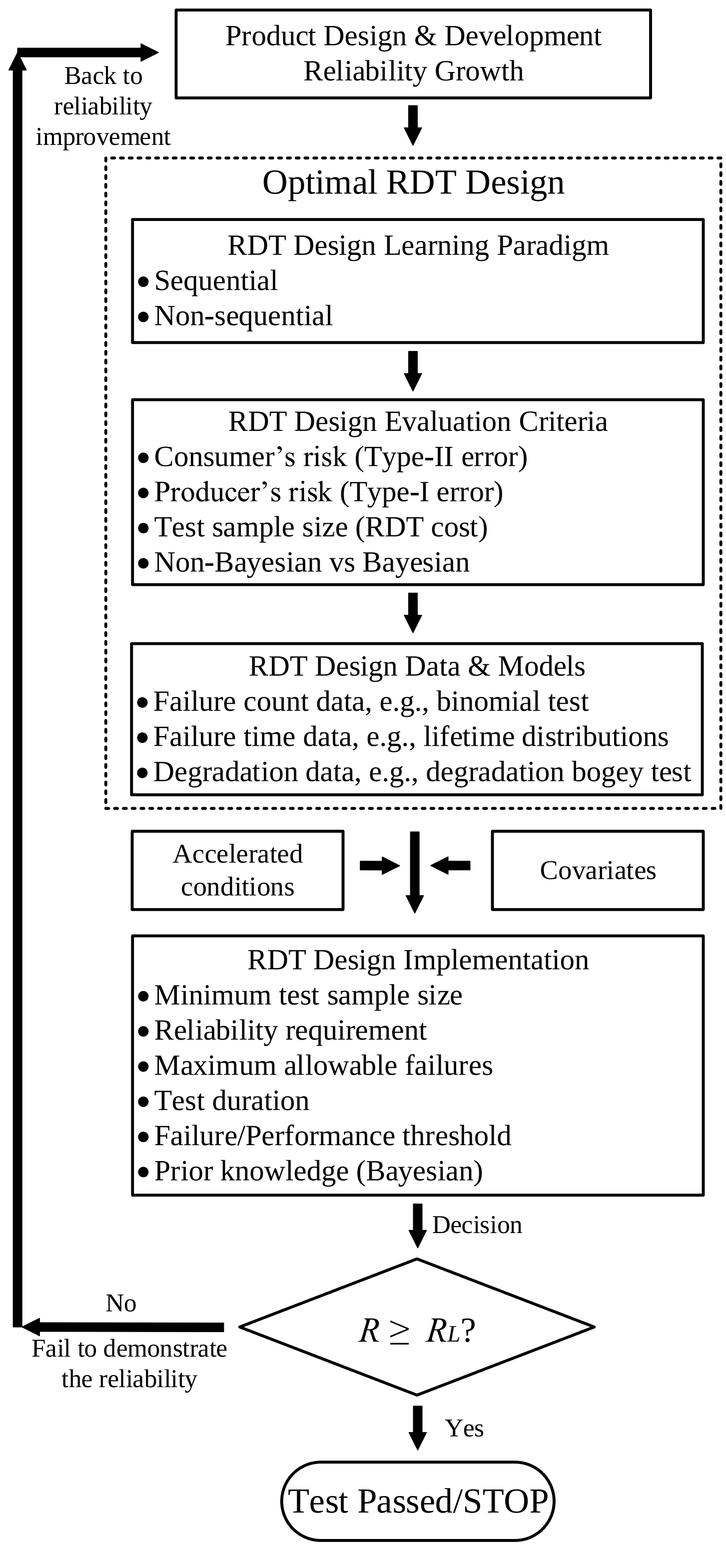


Figure 1.2: Flow chart of optimal RDT design and implementation

cess run test is usually employed for failure count data with zero failure allowed in the test. Lifetime distributions are usually assumed from failure time data such as Weibull and Lognormal distributions. Degradation models usually set the performance threshold to evaluate the product failures. Most of the tests may need to be conducted under accelerated conditions in order to shorten the testing period and improve test efficiency and cost-effectiveness. Acceleration models can be integrated with the RDT design models and incorporate different environmental stresses (e.g., pressure, temperature, voltage, etc.) and operating conditions (e.g., frequency of operation, usage rate, etc.). Optimal RDT design will not only decide how many units to be tested, how long the test will be or how the failures are measured, it will also determine the optimal accelerated conditions for the test. Additional design information may also be added to the statistical models with covariates. In design implementation step, the test will be conducted based on the design details such as minimum test sample size, reliability requirement, maximum allowable failures, test duration, failure/performance threshold, etc. If Bayesian framework is employed, the prior knowledge can also be incorporated. In the decision making step, if the lower level reliability requirement can be demonstrated, the test will be passed. Otherwise, the product should be sent back to the development stage for further reliability improvement such as taking corrective actions and performing reliability growth.

In the following part, a comprehensive review on existing literature of RDT design will be provided to discuss designs with different data and models, evaluation criteria, learning paradigms and diverse applications. The characteristics as well as the strength and weakness of different types of RDT designs will be discussed and compared. Illustrations of applying different RDT designs in real-world practice will also be provided.

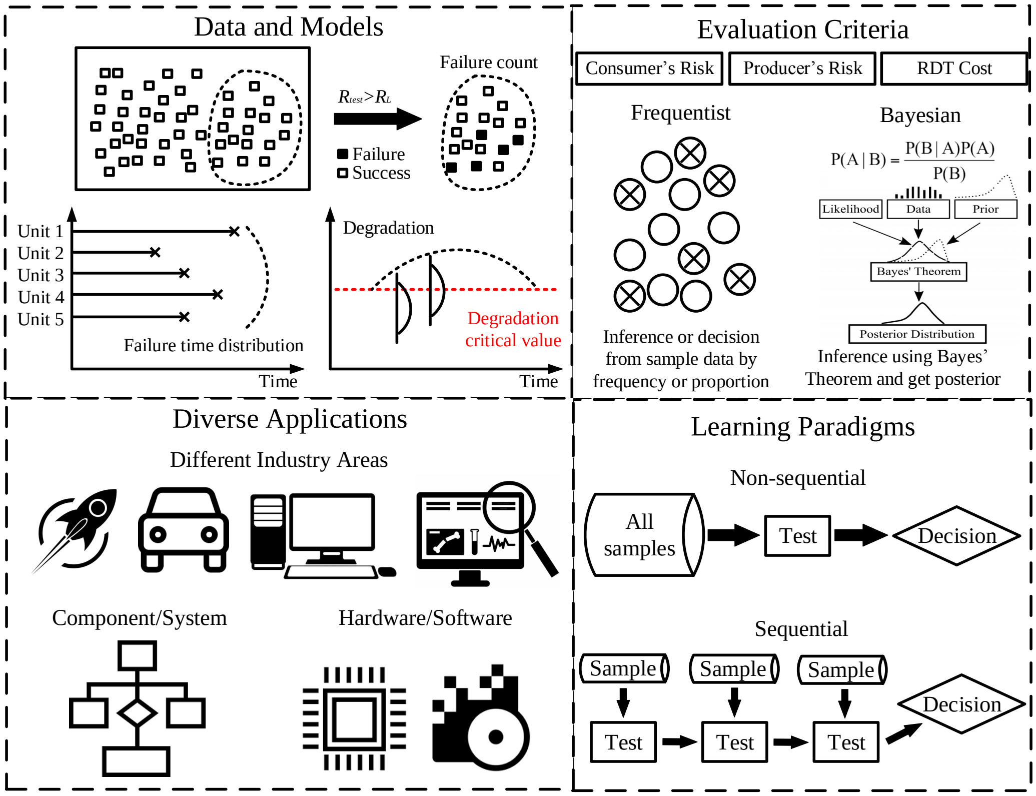


Figure 1.3: Summary of RDT designs in existing literature

1.2 Literature Review

This section will provide a comprehensive literature review of existing studies in RDT designs and summarize them from the following four aspects. First, based on different types of tests implemented and reliability data collected, such as failure count data, failure-time data and degradation data, different RDT designs have been developed with different statistical models. Second, the evaluation criteria used to evaluate the performance of RDT also vary in the existing studies, which include consumer's risk, producer's risk, acceptance probability, etc. They can be constructed under either non-Bayesian framework with Frequentist approach or Bayesian framework with prior knowledge incorporation, both of which can provide various statistical interpretations of product reliability. Third, there are two different learning paradigms in RDT designs, namely sequential RDT design and non-sequential RDT design. Depending on the physical constraints

of testing facilities, testing costs and availability of prior knowledge, the proper learning paradigm for RDT design can be chosen. Finally, there are diverse industrial applications of RDT design from different areas such as military, automobile, electronics, etc. Depending on the complexity of products, the reliability demonstration can be conducted on component or system level. In additional to hardware reliability demonstration, RDT designs for software are also reviewed. Figure 1.3 further summarizes the above four aspects of existing RDT design studies.

1.2.1 RDT Designs Based on Different Reliability Data and Models

Based on different types of reliability data (Guo and Liao, 2011; Crk, 2000; Lee et al., 2015) used in the test designs, existing RDT designs can be generally categorized into three types: (1) RDT designs based on failure count data, (2) RDT designs based on failure-time data, and (3) RDT designs based on degradation data. Due to the properties of different reliability data structures, the methods employed in each category also differ in their models and applications.

For failure count data, most methods in RDT designs are based on the binomial equation (Kececioglu, 2002), which assumes that the product failure probability follows a binomial distribution. The desired product reliability is usually pre-specified as the lower level reliability requirement in the RDT design. A set of product units will be tested for a given testing period. If the observed number of failures doesn't exceed the maximum allowable failures, the test can be passed and the reliability requirement can be demonstrated. For example, the manufacture may want to test a sample of product units to demonstrate the lower level reliability of 97%. With the maximum allowable failures given by 5, the test can be passed if no more than 5 failures are observed at the end of the testing period. Due to the limits in budget and resources, most RDT designs aim to find the optimal test plan with a minimum test sample size and its corresponding maximum allowable failures, which can minimize the testing cost with the control of evaluation cri-

teria (e.g., consumer's risk) at certain acceptance level. For example, if the manufacture wants to demonstrate the lower level reliability 97% with consumer's risk controlled at 0.05, the minimum test sample size can be determined (e.g., 50) when maximum allowable failures is given (e.g., 5). Since no failure time distribution is assumed and the test will be passed only depending on the success/failure of test units, it is usually referred to as binomial test (Kleyner, 2008) or attribute test (Lee et al., 2015). Gerokostopoulos et al. (2015) provided the practical experience in determining the optimal test sample size by exploring the changes of demonstrated reliability when the maximum allowable failures and test sample size changed. Jensen (2015) considered the method of determining test sample size when the reliability data was dependent, using Markov dependence model to calculate sample size under different reliability requirements. In practice, in order to minimize the test sample size, the maximum allowable failure is usually set as zero, which corresponds to the zero-failure test (Guo et al., 2010a) or success run test (Kececioglu, 2002). Success run test is known to be one of the most commonly utilized RDTs in the industry. It can be applied to one-shot systems effectively (e.g., missiles that are classified into success or failure right after being launched without time factor defined), which may have few failure and high reliability requirement. Guo et al. (2010a,b) described the details of a flexible and practical method to estimate the one-shot system reliability and inference on the confidence boundaries when zero failure was observed from components. Kleyner et al. (2004) used the actual failures data from automobile validation tests to minimize the life cycle costs from testing and warranty. For some products which the failures can hardly be observed in the specified testing period, binomial test may be ineffective since it may require excessive test time to observe failures and also need a larger test sample size to demonstrate the required reliability. Alternatively, the degradation zero-failure test (Yang, 2009) can be used to measure the degradation performance of the product, which can help determine the failure status once the product degradation reaches a specified threshold. The advantages of binomial test, which is the scope of this

dissertation, include: (1) the easy implementation without specifying the product lifetime distributions; and (2) less efforts in monitoring and inspection during the test, which can be simple and convenient for practitioners in real applications.

Failure time data (Meeker and Escobar, 2014) is usually collected from testing a sample of units at a given testing period and measures the failure time of those units which fail within the testing period. The units that still survive by the end of the testing period are referred to as right-censored units. Based on the failure time data, different RDT designs can be developed by considering different underlying failure time distributions and reliability evaluation methods (Kececioglu, 2002; Hamada et al., 2008). Weibull distribution is most commonly used for failure time distribution in RDT designs based on failure time data. It has the versatility to describe the characteristics of distributions with different shape and scale parameters. To conduct statistical inference on Weibull distribution parameters and determine the optimal RDT design, existing methods can be mainly classified into two subcategories: (1)simulation-based methods (Ramírez-Márquez and Jiang, 2006; Billinton and Wang, 1999) and (2) analytical methods (Meeker and Nelson, 1977; Meeker and Escobar, 2014). Simulation based on Monte Carlo methods vary the test sample sizes and test duration to find a proper test design that satisfies the precision requirement on reliability estimation. (Ramírez-Márquez and Jiang, 2006) used component level reliability to estimate system level reliability with Monte-Carlo Simulation. Similarly, Billinton and Wang (1999) proposed a time sequential Monte Carlo Simulation technique to be used in evaluating complex distribution system. The pro of simulation is that the precision can be effectively improved with more test units and longer test duration. However, when the set of options is large, simulation can be time consuming and only suggested for small samples. Analytical methods (Kleyner, 2008; Lee et al., 2015; Jin and Coit, 2008) include point estimation methods (e.g., Median Rank Regression (MRR), Maximum Likelihood Estimation (MLE), etc.) and interval estimation methods (e.g., likelihood ratio method, Fisher information bounds, etc.). Analytical methods overcome the

drawback of simulation-based methods to reduce the computation complexity and cost. Once the actual failure times or censoring times are observed, the parameters of Weibull distribution and variance-covariance matrix can be approximated (Meeker and Nelson, 1977; Lawless, 1978). The ultimate goal is to reduce cost-related test sample size and test duration while demonstrating the reliability requirement at certain confidence level. The RDT design based on failure time data has many advantages such as the analytical description of reliability as a function of time, and better understanding of possible test outcomes based on simulation results. However, it comes with the prize of requiring more monitoring efforts at additional testing costs. Further, Weibull distributions may not be universally applicable (Kleyner, 2008) and many other failure-time distributions, such as Gamma (Fernández, 2011) and Log-normal distributions (McKane et al., 2005), can be considered for RDT designs. Additionally, RDT designs based on failure time data are not applicable to one-shot systems which assume no time factor.

For many highly reliable products, it is often unable to observed failures within a reasonable testing period, which makes RDT design based on failure time data invalid. Alternatively, by measuring certain degradation performance, the degradation data can be a helpful information source to reflect the product reliability. The degradation data can be utilized (Crk, 2000) to measure the performance characteristics at different time during the test. The degradation measurements can be used to estimate the time of failure before observing the failures. In other cases, the product failure can be said to occur when the degradation performance reaches a specified threshold. In this way, the test duration can be reduced if degradation measurements can provide the prediction of reliability at a high level of confidence. Degradation-based RDT designs are mainly used for bogey test (Wang, 1991; Yang, 2009), which requires a sample of units to be tested for a specified test duration and demonstrates the reliability when no failure occurs in the testing period. The utilization of degradation data can improve the bogey test efficiency by precisely estimating the reliability and determining whether or not a test unit

will fail by the end of the test without actually testing to failure. Yang (2013) proposed the heuristic degradation test plans to significantly reduce the test sample size and testing period and to further ensure the robustness to quantify uncertainty from model parameters estimation. Wang and Coit (2004) considered the general modeling approach for reliability prediction using multiple degradation measures to assess the system level reliability. However, the accuracy of reliability prediction may rely on the specification of degradation models (Gorjian et al., 2010). For example, the general degradation path model can be both statistically and computationally simple for reliability prediction, however, the modeling assumptions or flexibility in the degradation process can be restrictive. The random process model can estimate reliability with more modeling flexibility, however, the estimation may be inaccurate due to insufficient data points.

Due to the fact that many products may have a relatively long lifetime over years, it is not practical to conduct the RDTs for such long testing period. In order to obtain the reliability information more efficiently, all the aforementioned RDTs are often implemented under certain accelerated conditions. For example, if the manufacturer wants to test a sample of lamps, under the accelerated conditions such as high temperature, high frequency of usage, etc., the actual testing period may be 5 days, which corresponds to a 5-year lifetime. However, those RDT designs mainly focus on determining the optimal minimum test sample size, test duration, etc., without determining the optimal levels of accelerated conditions. In order to optimize both test sample size/test duration together with levels of accelerated conditions based on reliability data, accelerated tests are commonly employed (Nelson, 2005, 2009). For example, after taking covariates (e.g., usage rate, temperature, pressure, etc.) into consideration, accelerated tests may vary the levels of covariates to determine the optimal test sample size/ test duration and the optimal accelerated conditions simultaneously. Accelerated tests can be applied to different types of RDT designs (Meeker and Escobar, 2014) (e.g., ALTs, ADTs, etc.) and can be employed in either Bayesian (Fan and Chang, 2009) or non-Bayesian methods (Porter, 2004). The

major benefits of accelerated tests include significant reduction of test duration and costs as well as the efforts in monitoring accelerated variables, which can provide more assessments and estimation information (e.g., detection of failure modes (Luo et al., 2015)) of the product reliability.

1.2.2 RDT Designs Based on Different Evaluation Criteria

There are many different evaluation criteria involved in RDT designs. Specifically, consumer's risk is the probability that customers receive inferior products that pass the test. High consumer's risk will lead to dissatisfaction of customers and negative impact on product reputation. Producer's risk is the probability that manufacturers reject the test units that suppose to pass the test. High producer's risk may result in additional cost for manufacturers. Another criterion which quantifies the probability of passing/failing the test is denoted as acceptance probability. Acceptance probability can be used to evaluate the impact of RDT decision uncertainty on subsequent reliability assurance activities such as reliability growth and warranty services. Cost components from different activities are also considered as criteria to evaluate the cost-effectiveness of the RDT designs, such as RDT cost (e.g., test sample size), expected warranty services cost, etc. Depending on whether prior knowledge can be incorporated, different evaluation criteria including consumer's risk, producer's risk, acceptance probability, can be conducted under classical (i.e. Frequentist) approach or Bayesian approach.

Using Frequentist approach, the estimation of unknown parameters is associated with certain level of confidence, which can be used to describe how confidently we trust the estimation parameters. The obtained evaluation criteria such as consumer's risk, producer's risk and acceptance probability will be objective, depending on the assumptions of specific models and available test data. In (Tobias and Trindade, 2011; Hamada et al., 2008), the classical approach have been illustrated to calculate different evaluation criteria. It can be intuitively applied to different types of RDT designs. For example, McKane

et al. (2005) developed the design for RDT with location-scale-based distributions and censored reliability data, to determine the optimal test sample size and maximum allowable failures. Guo and Liao (2011) compared the methods for RDT designs between failure time data with lifetime distributions and failure count data using binomial distribution for number of failures. However, Frequentist approach has the limitation that the probabilistic distribution of the estimation is not available, which may have the weakness in quantifying the overall estimation uncertainty. It is also limited when no additional information can be incorporated and cannot further improve the reliability estimation performance.

Bayesian approach, which is employed in this dissertation, considers probabilistic measures using prior knowledge to quantify the uncertainty from test decision and assess the risks for different stakeholders (e.g., manufacturers, consumers). The parameters in the model are assumed to follow certain prior distributions (elicited from prior knowledge) and can be integrated with new test data to obtain the posterior distribution, which can reflect the updated degree of belief on unknown parameters. Bayesian RDT designs (Tillman et al., 1982) distinguish itself from non-Bayesian (i.e., Frequentist) RDT designs by incorporating prior information to explore the potential for sample size and test duration reductions. Prior information can be incorporated from useful expert information, domain knowledge and historical test data. For complex products with different components, the test information on components can also be incorporated to describe the reliability of entire system. One typical example is one-shot system (Guo et al., 2010a) such as missiles and satellites. Since RDTs for the entire one-shot system will lead to its destruction, which results in extensive costs, it is practical to incorporate prior test information of components using Bayesian approach. As inheriting from non-Bayesian RDT designs, a variety of studies using Bayesian RDT design have been conducted for different reliability demonstration problems (Hamada et al., 2008; Coolen and Coolen-Schrijner, 2008; Fan and Chang, 2009). One basic idea is to incorporate prior information of failure distribu-

tion (either failure occurrence or failure time) to get the posteriors of different reliability criteria (e.g., consumer's risk, producer's risk), which can help determine the details for RDT design implementation such as minimum test sample size, test duration or other specifications. In binomial test, the prior distribution of reliability can be assumed as beta distribution, which is a conjugate prior for binomial distribution (Kececioglu, 2002). For example, Kleyner et al. (1997) used Bayesian method to reduce test sample size for success run test in automotive electronics and incorporated the mixture prior derived from information of similar products. Lu et al. (2016) further considered multiple objectives and balanced their trade-offs to select the optimal test plans. When failure time distribution is assumed, the prior information of the shape parameters can be incorporated using Bayesian approach. For instance, two-parameter Weibull distribution is commonly used for assumption of product lifetime failure distribution, where each of the shape parameters can have the prior distribution of gamma distribution with hyper-parameters (Hamada et al., 2008). For system reliability demonstration, Ten and Xie (1998) proposed the Bayesian RDT method for series system with binomial subsystems. Mann's Approximately Optimum Lower Confidence Bound model is used to derive prior knowledge from subsystems. One concern for Bayesian RDT design is the choice of prior distribution (Coolen and Coolen-Schrijner, 2008), which takes into account subjective information and can greatly influence the design settings. In practice, with sufficient data observed, the posterior distribution of reliability criteria can be less sensitive to the choice of prior distribution.

1.2.3 RDT Designs Based on Different Learning Paradigms

When designing the RDT, it is important to control the testing costs regarding the test sample size, test duration and other related expenses. With reasonable test sample size and test duration, the RDT may be conducted with all the units tested at one time to demonstrate the product reliability. However, for some products with high reliability,

which may require significantly large test sample size and much longer test duration, it may not be practical test all the units at once over a long period. Alternatively, another learning paradigm can be followed to test the unit sequentially and update the reliability information until the reliability requirement can be satisfied. Based on different considerations in test efficiency, cost effectiveness and reliability estimation performance, etc., there will be two different learning paradigms in RDT designs, non-sequential RDT designs and sequential RDT designs.

Non-sequential RDT designs typically need to determine a fixed test sample size and other test settings before conducting the test. The test results will be evaluated after all units have been tested at the end of the test duration. For example, Jensen (2015) explored the dependence of reliability data and how to model it for binomial test and determine the sample size for different criteria thresholds and reliability requirements. Coolen and Coolen-Schrijner (2005) used a non-parametric predictive approach to determine the number of tests needed for failure free products. Martz Jr and Waller (1979) developed the Bayesian method to determine the test sample size for zero-failure tests using prior knowledge. However, due to the constraints in cost budget and test duration, non-sequential RDT design may not be applicable sometimes. In this dissertation, non-sequential RDT design is employed, which can explicitly explore the variations in determining minimum test sample size for demonstration of various product reliability requirements.

As an alternative, a sequential RDT design (Sun and Berger, 1994) is to test each sample to failure (product fails or predetermined test time elapsed) and evaluate the product performance and reliability sequentially. It is based on a sequential hypothesis testing paradigm in which the test statistic is reevaluated sequentially when each new failure is observed. Once a test failure is observed, the sequentially performed inspection procedure based on established criteria will determine if the required reliability is demonstrated or whether one more sample/longer time need to be tested. A typical sequential

RDT design is based on Sequential Probability Ratio Test (SPRT) (Siegmund, 2013) and both cumulative test duration and number of failures can be compared with established criteria (e.g., ratio of upper and lower bound for failure probability or distribution parameters) to decide whether to accept, reject or continue testing the item. It can be seen that the test sample size required to reach a conclusion is a random number and cannot be predetermined. The sequential nature of this type of RDT design can actually utilize the information from previous stages of the test, which help reduce the required test sample size or test duration. Sequential RDT designs can be applied under either non-Bayesian (Li et al., 2009) or Bayesian framework (Rodríguez-Narciso and Christen, 2016), depending on whether to utilize available prior information. Lu and Rudy (2001) also provided more descriptions and comparisons on using sequential sampling with both Bayesian and non-Bayesian methods to determine the test sample size. In some applicable cases, sequential RDT designs may have the advantage over non-sequential RDT designs to require minimum observed failures to demonstrate the product reliability. However, since there is no predetermined maximum constraints for the test sample size, test duration or other subsequent costs, the variations of those design settings in a broad range for sequential RDTs may cause issues (e.g., administration, equipment limitations, inspection) in practice.

1.2.4 RDT Designs with Diverse Applications

RDTs have been widely employed in various industries to demonstrate the reliability of product so that it can meet the requirement of customers. Depending on the characteristics of products and application scenarios, RDT designs have been developed in many industries with diverse applications such as missiles in military, satellites and rockets in space craft industry, computers and smartphones in high-tech electronics, automobiles, etc. For example, Tam (1995) described the reliability demonstration for the component in missile applications called Plastic-Encapsulated Microcircuits to ensure its long-term

storage for military purpose. Sun et al. (2012) provided the approach for one type of rare earth permanent magnets which was used in space industry and might experience zero failure in the degradation process. Sarson-Lawrence et al. (2014) provided the approach to quantify the reliability of servomotor in unmanned aircraft applications and demonstrated their reliability with time-to-failure records. RDTs are also commonly used to ensure the product reliability and customer experience. Yang (2009) discussed the method to demonstrate the reliability of electromagnetic valve with degradation measurements in bogey test. Yang and Sun (1999) provided a comprehensive review of reliability analysis methods for hard disk and the approaches for reliability demonstration in a fast-path development cycle. Automobile industry has huge involvement in RDTs to ensure the reliability of vehicle, the driving experience and the long-term service quality. Kleyner and Sandborn (2008) provided the life cycle cost methods to validate the automobile reliability and minimized the costs in testing and warranty. Allmen and Lu (1993) designed the methods for failure detection and reliability demonstration for automotive applications to determine the test sample size and test duration.

Depending on the complexity of products where RDT is applied, the availability of data and practical applications, RDT designs can be classified into component or system level. For example, for product with simple functionality, such as a lighting bulb, the RDT design will be intuitive to directly test the bulb till failure. Zhang et al. (2015) developed efficient RDT design method with ADT for long-service product such as spherical plain bearing. However, for products with complicated components such as laptop or automobile, testing all the products to failures may not be cost-effective. Especially for one-shot systems like missiles or rockets, it is impossible to test all the product to failure due to the destructive nature of products and extremely high expenses. In this situation, many methods have been developed to test and demonstrate the reliability of components of the products and further demonstrate the reliability for the overall system. Mazumdar (1977) provided the optimal procedure to test the components and determined the reli-

ability of series systems. Hsieh and Ling (1999) developed the Bayesian framework to incorporate prior knowledge from component analysis for system development and system reliability demonstration. Guo et al. (2010b) specifically discussed the RDT designs for one-shot systems to estimate the confidence interval for zero-failure system's reliability, in which Bayesian method is used to derive the system reliability from component estimations. Fernández (2011) explored the optimal RDT decision making for k-out-of-n system with Gamma distributed components to obtain the test sample size and acceptance factors through mixed integer optimization. Ramírez-Márquez and Jiang (2006) discussed the analytical methods to assess the system reliability from component analysis with improved confidence intervals.

In addition to demonstrating the reliability of hardware components/systems, there is a need of reliability assurance for software products (Cai et al., 1991) as well, because the downtime or failure of software may cause negative impacts on financial conditions, service experience, etc. Software is usually used for controlling production lines and operating systems of computers, there is no physical failures of the software product. Therefore, the reliability is discussed in terms of mean time to breakdown, the downtime between recovery, software failure rate and number of bugs in the software. We can also count the number of broken software applications from the sample to determine the product reliability. (Sandoh, 1991) transferred the methods from hardware to software and developed the RDT design to determine the test duration and test sample size for software applications based on consumer's risk and producer's risk. The zero-failure reliability demonstration for software has also been developed. Tal et al. (2001) further developed the RDT for safety-critical software applications in a sequential manner.

1.3 Overview of Dissertation

This dissertation focuses on advancing the existing design of RDTs by tackling the complexities from the following three aspects, namely data complexity, planning com-

plexity and evacuation criteria complexity. We use RDT designs based on failure count data to illustrate the proposed frameworks, and they can be readily generalized to design of RDTs based on other types of reliability data. Further, the proposed work is established under the Bayesian framework and focuses on the non-sequential learning paradigm. The first advancement is to relax the conventional assumption of binary failure states in RDT design and considering multiple failure states instead. The second advancement is to expand the planning horizon of RDT design in a more global sense by considering the uncertainty after RDT as well as the effect of RDT on its subsequent reliability activities. The third advancement is to investigate multiple objectives and their trade-offs in the optimal RDT design instead of assuming a single objective in the conventional RDT design.

1.3.1 Data Complexity

In conventional RDT design, the failure count data will be collected from a single testing period for single failure mode under a predetermined test sample size. However, such RDT design with binary failure states may not be able to meet the more diverse requirements of customers for demonstrating product reliability in multiple testing periods. Customers may demand higher reliability during the earlier time period because they are averse to early failures of the product. Considering an example with two companies, both of them are conducting the RDTs based on the same design settings (e.g., the same sample size and the same maximum allowable failures of 5) to demonstrate product reliability in a 5-year mission time. Company I observes 1 failure in the first two years and 4 failures in the last three years. Company II observes 4 failures in the first two years and 1 failure in the last three years. Both of the companies will pass the tests according to the design. However, customers may prefer the products from company I if their dissatisfaction is more related to early failures than late failures. The conventional RDT design assuming a single testing period cannot differentiate such preference of customers and there is need to develop RDT design which allows reliability demonstration for multiple time periods.

In addition, conventional RDT design does not differentiate multiple failure modes. But in practice, product failure often consists of multiple failure modes with different criticality and negative consequences, which affect the customer's satisfaction differently. For instance, compared to the failure of a video card or a key board with lower replacement cost, customers may be more concerned with CPU failure of a computer. In other words, customers may have different expectations of the reliability requirement of different failure modes.

To meet with the increasingly complex and diverse demands of customers, there is a need to better design RDT which takes into account the aforementioned complexities of reliability requirements. This line of research will investigate the multi-state RDT design. When multiple states refer to multiple time periods, the proposed RDT design is able to demonstrate the product reliability with different reliability requirements over different time periods. When multiple states refer to multiple failure modes, the proposed RDT design is able to demonstrate product reliability with multiple failure modes.

1.3.2 Planning Complexity

Conventional RDT design has a relatively restrictive planning horizon which mainly focuses on the RDT alone. Using binomial RDT as an example, in order to find the optimal design, the focus will be minimizing the RDT testing cost by selecting the optimal plan with minimum test sample size. However, few research has incorporated the relationship between RDT and other relevant reliability assurance activities and further investigated how the relevant reliability activities will affect the RDT design. When a RDT is accepted, product will be released to market and warranty services will be initiated. When a RDT is rejected, the product will be sent back to the design and development stage for further improvement via reliability growth. Thus, accepting or rejecting the RDT will directly determine which subsequent reliability activity will be triggered. However, at the design stage of RDT, the actual RDT results are random and there is uncertainty of accepting or

rejecting the RDT. Different configurations of the test design will influence such uncertainty and further influence the probability of initiating subsequent reliability activities of reliability growth and warranty services. Meanwhile, different levels of reliability requirement to be demonstrated by the RDT will also affect the potential costs of the subsequent activities. For example, demonstrating a higher reliability requirement may help reduce the warranty service costs since the number of warranty claims may decrease. When the level of reliability requirement changes, both the RDT design and the probability of triggering reliability growth and warranty services will also change. Conventional RDT design without a broader planning horizon and a more global view tends to overlook the aforementioned planning complexity and lead to sub-optimal RDT design.

To expand the planning horizon, this line of research aims to develop an optimal RDT design strategy by explicitly quantifying the RDT acceptance decision uncertainty after the test has been implemented and further considering the cost components of subsequent reliability assurance activities including reliability growth and warranty services. The impacts of different cost components from different reliability assurance activities on acceptance decision as well as RDT design inputs will be comprehensively investigated. The proposed RDT design tends to be more economical than conventional ones from a global perspective.

1.3.3 Evaluation Criteria Complexity

Conventional RDT design often minimizes the RDT cost by controlling a single performance evaluation criterion under a single objective optimization formulation. For instance, in binomial RDT design, given the pre-specified reliability requirement and maximum allowable failures, under the controlled threshold of consumer's risk, the optimal RDT design supposes to obtain the minimum test sample size.

However, in addition to the testing cost and consumer's risk, there are also other performance evaluation criteria. For instance, manufacturers may be more concerned about

rejecting the reliable products in the test and thus, producer's risk can be an important evaluation criterion as well. A low product's risk design may also save the costs of refreshing production line or disposing the rejected units for the manufacturers. In addition, they may be concerned about whether the actual RDT test can be more easily accepted. Then, acceptance probability can be an important evaluation criterion taken into consideration. If the test cannot be accepted, manufacturers may need to conduct the test again or take corrective actions to improve the product reliability, which results in additional costs as well. Manufacturers may also be concerned about subsequent costs affected by RDT design. For example, the expected warranty services cost will depend on the level of demonstrated reliability. The lower reliability is demonstrated, the more warranty services cost will be expected because of more customer dissatisfaction and warranty claims. Some of the evaluations criteria may be conflicting, e.g., consumer's risk vs producer's risk. When manufacturers try to reduce the consumer's risk by testing more units and demonstrating higher reliability, the producer's risk may be unacceptably high and lead to additional costs.

To simultaneously address the aforementioned objectives on controlling multiple evaluation criteria without pre-determining a specific objective, there is a need to investigate the RDT design under a multiple-objective optimization framework. This line of research will investigate multiple potential conflicting objectives and comprehensively investigate their relationships and trade-off patterns. A multi-objective RDT design will be developed with a set of non-dominant test plans obtained to balance the trade-offs among multiple conflicting objectives. How different design inputs and prior setting will affect the objectives will be investigated as well to provide more practical guidance to practitioners in reliability and quality engineering.

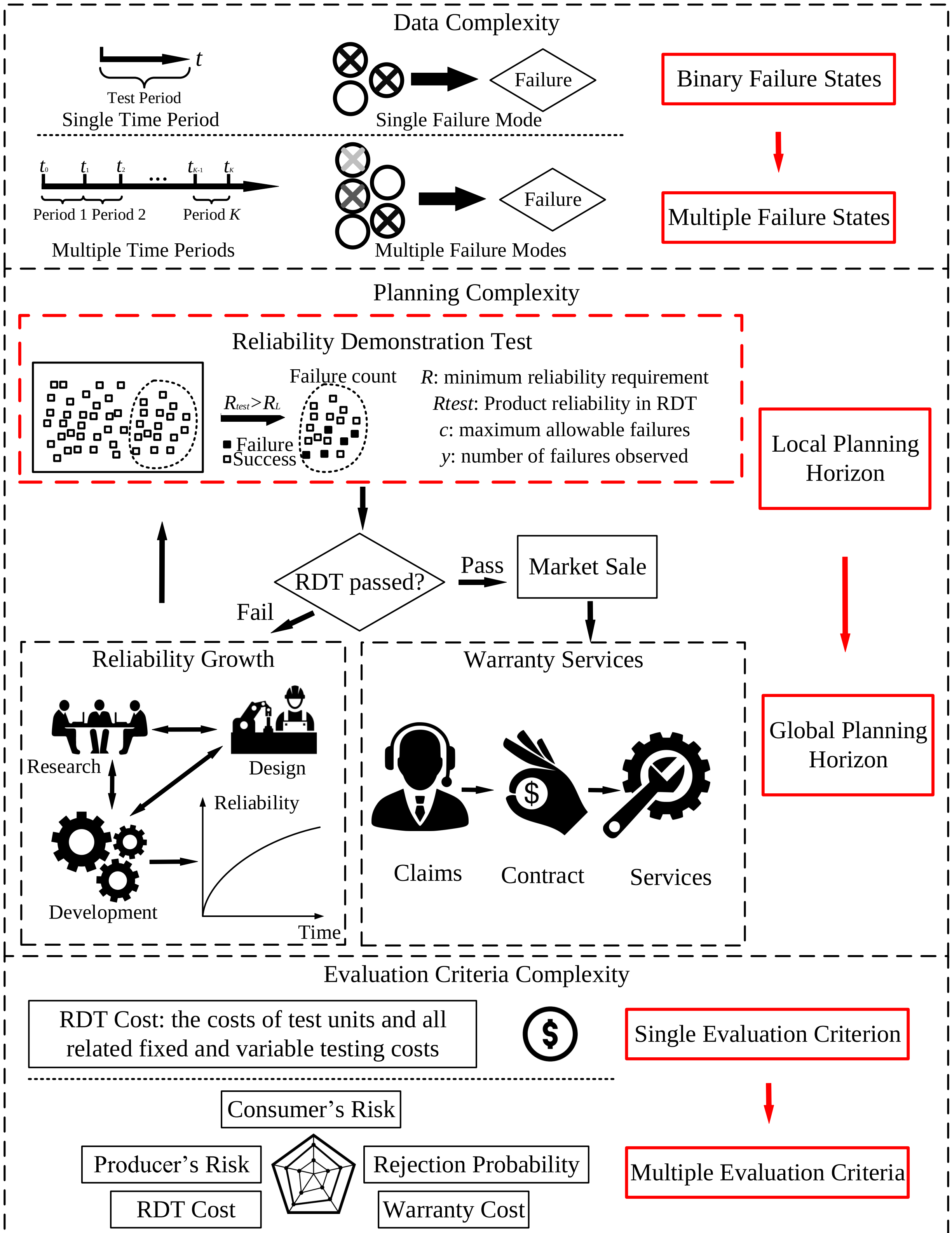


Figure 1.4: Three advancement aspects of RDT design studied in the dissertation

1.4 Organization of this Dissertation

The dissertation can be organized as follows. Chapter 1 introduces the background of reliability, reliability activities on product life cycle and tasks in reliability testing, and the scope of this dissertation is defined to be reliability demonstration test. Chapter 2 proposes two Bayesian multi-state reliability demonstration tests (MSRDTs) for demonstrating reliability at multiple time periods as well as with multiple failure modes. Chapter 3 proposes a Bayesian optimal BRDT design with explicit quantification of the acceptance uncertainty and further integration of the BRDT testing cost with the expected reliability growth and warranty service costs. Chapter 4 proposes a multi-objective optimal RDT design framework for multiple time periods by simultaneously investigating multiple and potential conflicting objectives, such as consumer's risk, producer's risk and expected warranty service cost and comprehensively investigating their interrelationships and potential trade-off patterns. Chapter 5 concludes the dissertation with some directions for future research. Figure 1.5 gives an organization diagram of the dissertation.

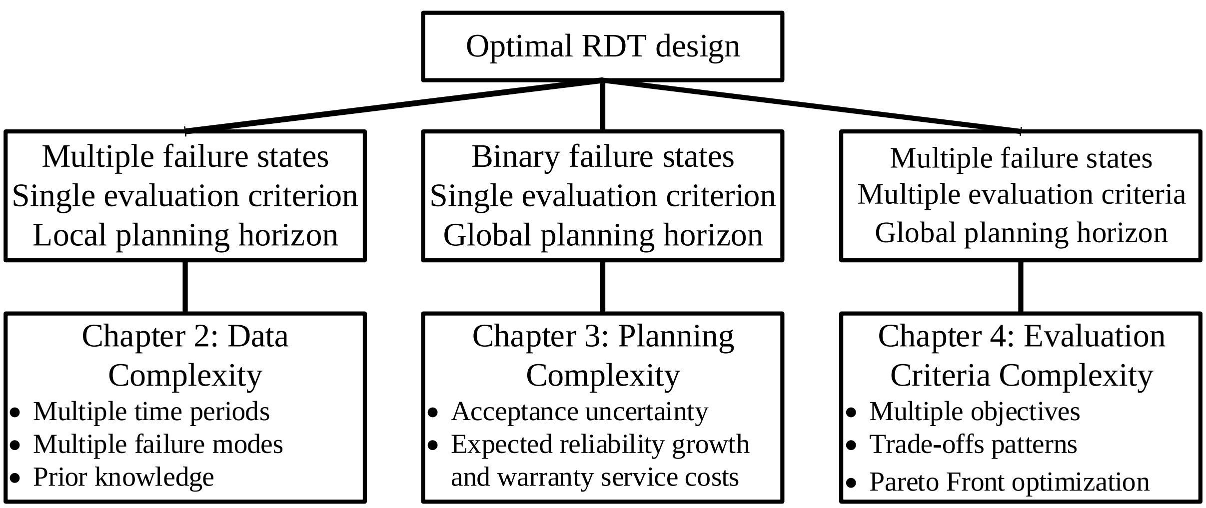


Figure 1.5: Organization of the dissertation

Chapter 2

Multi-State Reliability Demonstration Tests

2.1 Introduction

Reliability of a product is the probability that the product can perform its required function at a given time point. As a time-dependent characteristic, reliability is an important measure of the product quality and safety over time, which has a great impact on the satisfaction of customers and can influence their purchase decisions linked with the revenue of manufacturers. In order to succeed in the market competition, manufacturers need to produce products with high reliability over their expected lifetime. Reliability demonstration tests (RDTs) are often conducted by manufacturers to demonstrate the capability of their products to meet the requirements from customers for achieving good quality and performance over time. Given the budget and time constraints, manufacturers need to determine the number of test units, the time duration of the test, and the maximum number of failures allowed to pass the test. These choices are usually made to ensure the consumer's risk (CR) on having a product that has passed the test but fails to meet the reliability requirement is controlled. Controlling the CR at an acceptable level can take the burden off the customers on bearing a high risk of receiving inferior products which are claimed to have met the requirements on reliability, and hence can help improve customers' satisfaction.

Different categories of RDTs have been studied in the literature based on different types of reliability data, such as failure counts data (Guo et al., 2010b; Li et al., 2016; Lu et al., 2016), failure time data (Guo and Liao, 2011; McKane et al., 2005) and degradation

data (Yang, 2009). Failure counts data reports the number of failures that occur during a fixed test period. The RDTs based on failure counts data (Wasserman, 2002, pp. 208-210) are also called binomial RDTs (BRDTs) since failure counts are modeled with binomial distributions.

In a BRDT, within a given testing period, if the number of failures does not exceed the maximum number of allowable failures, the test is passed. The maximum number of allowable failures c and the minimum number of test units n are determined to ensure a certain minimum acceptable reliability requirement, R, is met with the controlled CR at or below β by the end of the test duration. The BRDTs are broadly applied in reliability engineering practices because (i) they require less monitoring efforts in the middle of the test duration; and (ii) they are simple and straightforward to be implemented and analyzed. However, with the increasing needs from customers, the BRDTs are no longer able to meet all requirements in many applications. For example, customers may have varied requirements on reliability performance over different time periods. It is common that many customers have little tolerance of early failures and hence require high reliability during early lifetime and lower reliability for later time. In this case, a BRDT for demonstrating reliability within a single time period is inadequate to meet all requirements.

Consider a scenario when two companies run BRDTs with the same testing period of 5 years and use the maximum number of allowable failures as $c = 5$. Products from company I had 1 failure in the first two years and 3 failures in the last three years. Products from company II had 3 failures in the first two years and 1 failure in the last three years. Even though the products from both companies can pass the demonstration tests, their underlying reliability performance indicated from the failure counts data can be different. If a customer needs products with high reliability in early lifetime (corresponding to allowing no more than 2 failures during the first two years), the risk of the product from company II failing to meet the requirement can be much higher than that of the product from company I. A typical BRDT with five-year testing period cannot demonstrate the

performance over the early two years, and hence raises the CR on accepting an inferior product that fails to meet all requirements.

Another limitation of the BRDTs is that they are often used for testing the pass or failure of a product without distinguishing the causes and consequences of different failure modes. A product with a complex system is often composed of multiple key components which may have different failure modes associated with varied consequences. Their failures can have different negative effects on the functionality of the entire product. For instance, the failure of the central processing unit (CPU) of a computer is much more crucial than the failure of a video card. Customers may also have different expectations for different components according to their values or costs of replacement. The cost of replacing a CPU or a motherboard is much higher than replacing a keyboard or a mouse. As a result, customers can have much higher expectation on the reliability of the more valuable and critical parts than the reliability of other parts or accessories. A typical BRDT cannot demonstrate separate reliability requirements for multiple failure modes.

To meet the ever-increasing demands of customers, more versatile RDTs with more tailored plans for testing multiple reliability requirements can better serve the customers with enriched information on product reliability. This paper proposes RDT strategies for two categories of reliability demonstration tests over multiple time periods and for multiple failure modes, both of which are referred to as multi-state RDTs (MSRDTs) throughout the rest of the paper. Alternative test plans within each category are also explored and compared with the conventional BRDTs for demonstrating multiple reliability requirements. Bayesian analysis is used for quantifying the CR associated with various test plans. The Bayesian method offers more flexibility on incorporating prior information of product reliability from either subject matter expert knowledge or historical data (Pintar et al., 2012; Weaver and Hamada, 2008; Wilson and Fronczyk, 2017). The impacts of different test strategies and different prior elicitations on the minimum test sample size (i.e. the number of test units required) will be studied to provide more insights on guiding de-

cisions on demonstration test plans. If there exist historical data which supports higher reliabilities compared to the requirements, then using Bayesian approach to incorporate prior information has the potential to reduce the minimum test sample size required for the MSRDTs.

The remaining of the paper is organized as follows. In the next section, the conventional BRDT plans are reviewed with discussions of their benefits and limitations. Then the new MSRDTs for demonstrating reliability requirements over multiple time periods are proposed. Two different design strategies are proposed and compared under different prior elicitation settings. In the following section, another category of new MSRDT designs for demonstrating reliability requirements involving multiple failure modes are proposed and their performances are evaluated and compared with the conventional BRDTs. Case studies on two categories of MSRDTs for multiple time periods and multiple failure modes are provided to illustrate the proposed test plans and demonstrate their performances. Conclusions and discussions are provided in the end.

2.2 Binomial RDTs

For many single use or "one-shot" product units, the testing procedure can be destructive. In this case, binomial RDTs (BRDTs) are the common choices to obtain the failure count data at the end of a predetermined test period (Kececioglu, 2002, pp.759-768). Let π denote the probability of failure over the test period, and R denote the minimum acceptable reliability at the end of the test duration. In Bayesian analysis, for a chosen number of test units, n, and a maximum number of allowable failures, c, the CR is measured by the posterior probability of the product failing to meet the reliability requirement given that the product has passed the test, which can be calculated as

$$\begin{aligned}\text{CR}_{\text{binomial}} &= P(\text{Failure probability fails to meet the reliability requirement}|\text{Test is passed}) \\ &= P(\pi > 1 - R|y \leq c)\end{aligned}$$

$$= 1 - P(\pi \leq 1 - R | y \leq c)$$

$$= 1 - \frac{\int_0^{1-R} [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}{\int_0^1 [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}. \quad (2.1)$$

Note that in Eq. (2.1), $p(\pi)$ denotes the prior distribution of π which can be specified based on subject matter expert knowledge or historical data and y denotes a random variable of the number of failures observed in the test period. Let β denote the maximum acceptable value for the CR, then a BRDT is determined by choosing the (n, c) combination such that the corresponding $\text{CR}_{\text{binomial}} \leq \beta$. According to (Lu et al., 2016), for any fixed choice of c, $\text{CR}_{\text{binomial}}$ increases as the test sample size n increases. We use n_{b} to denote the minimum test sample size that is required to control the CR within an acceptable range $\text{CR}_{\text{binomial}} \leq \beta$.

In Bayesian analysis, the $\text{CR}_{\text{binomial}}$ in Eq. (2.1) can be calculated using Monte Carlo integration (Robert and Casella, 2013, pp.71-131), where a large number of samples of π of size $M = 15000$ are generated from the specified prior distribution $p(\pi|x)$, and $\text{CR}_{\text{binomial}}$ is calculated approximately by

$$\text{CR}_{\text{binomial}} \approx 1 - \frac{\sum_{j=1}^{M} [\sum_{y=0}^{c} \binom{n}{y} (\pi^{(j)})^y (1-\pi^{(j)})^{n-y}] I(\pi^{(j)} \leq 1 - R)}{\sum_{j=1}^{M} [\sum_{y=0}^{c} \binom{n}{y} (\pi^{(j)})^y (1-\pi^{(j)})^{n-y}]}, \quad (2.2)$$

where $\pi^{(j)}$ is the jth generated sample of failure probability for the specified prior distribution.

Table 2.1 shows an example of BRDT plans with different choices of prior distributions of π. The mean and standard deviation (i.e., the square of variance) values are provided

to give some intuitions about the center and the spread of the prior distributions. For example, $\pi \sim \text{Beta}(1,1)$ is centered at 0.5 but has large standard deviation at 0.2893. While $\pi \sim \text{Beta}(2,18)$ has the mean failure probability of 0.1 but much smaller standard deviation (0.0647) around its mean. The minimum acceptable reliability from the consumers requirement was set at $R = 0.8$ and the maximum tolerable CR is chosen to be $\beta = 0.05$. When no historical data or prior information is available, a non-informative prior $\pi \sim \text{Beta}(1,1)$ can be used. For any assumed prior distribution of π, manufacturers can choose a test plan determined by (n_b, c) using the minimum sample size n_b for any chosen maximum number of allowable failures c. For instance, when $c = 0$ and a non-informative prior $\pi \sim \text{Beta}(1,1)$ is assumed, the minimum sample size which can ensure the CR calculated in Eq. (3.12) is no more than $\beta = 0.05$ is calculated to be $n_\text{b} = 13$. Hence, at least 13 units need to be tested if the test can only be passed when no failure is observed. However, as larger maximum number of allowable failures being set for passing the test, the CR increases as it becomes easier to pass the test for a given sample size n. Hence, to control the CR at or below $\beta = 0.05$, more test units need to be tested to pass the test due to more allowable failures.

Table 2.1: Minimum sample sizes required by BRDTs with different choices on c and prior distributions of π

$\pi \sim$ Beta	$(1,1)$	$(2,18)$	$(4,16)$	$(10,15)$	$(10,10)$
Mean(π)	0.5	0.1	0.2	0.4	0.5
SD(π)	0.2893	0.0647	0.0873	0.0965	0.1086
c	n_b				
0	13	4	18	45	53
1	18	7	23	51	58
2	24	11	28	57	64
3	29	15	34	62	69
4	34	19	39	68	74
5	39	22	44	74	80
6	44	26	49	80	85
Settings: $M = 15000, R = 0.8, \beta = 0.05$					

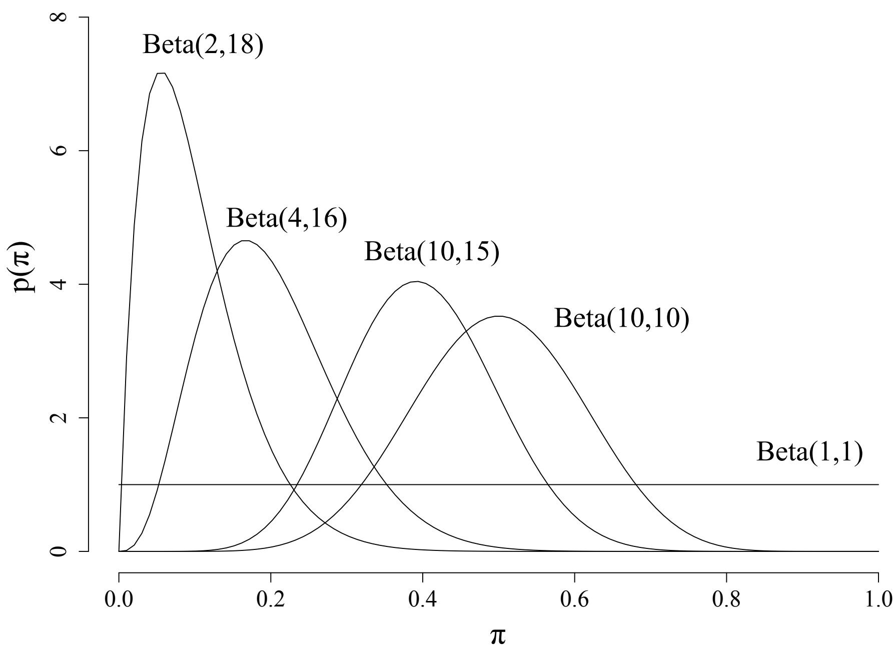


Figure 2.6: Density curves of different prior distributions explored in Table 1

When more informative priors are available from historical data or expert knowledge, they can affect the selection of test plans. Table 2.1 has explored the impacts of different prior distributions $p(\pi)$ on the selected test plan for different tolerances on the maximum number of allowable failures, c. Figure 3.23 shows the five prior distributions explored in Table 2.1. The flat density curve corresponds to the non-informative prior $\text{Beta}(1, 1)$ which assumes that all possible values for $\pi \in (0, 1)$ have equal probability. Other prior distributions from $\text{Beta}(10, 10)$ to $\text{Beta}(2, 18)$ become more informative with reduced spread (corresponding to smaller standard deviation in Table 1) and provide stronger support for smaller failure probability π. For any given c, the minimum sample size required can be reduced if the prior distribution from historical data supports the reliability requirement. For example, when a prior distribution $\pi \sim \text{Beta}(2, 18)$ is used, which supports high reliability around $1 - 2/(2 + 18) = 0.9 > R = 0.8$, fewer units need to be tested to demonstrate the reliability requirement (e.g., $4 < 13$ when $c = 0$). However, if the specified prior distribution is not in favor of the reliability requirement, as

illustrated with prior distributions $\text{Beta}(4, 16)$, $\text{Beta}(10, 15)$ and $\text{Beta}(10, 10)$, which favor incrementally lower reliability, more units are required to be tested to demonstrate the same reliability requirement.

On the other hand, Table 2.2 demonstrates the impact of different requirements on re-

Table 2.2: Minimum sample sizes required by BRDTs with different choices on c and reliability requirements

	n_{b}		
c	$R = 0.9$	$R = 0.8$	$R = 0.6$
0	28	13	5
1	39	18	8
2	50	23	11
3	61	28	13
4	71	33	16
5	81	38	18
6	91	43	20
Settings: $M = 15000, \beta = 0.05$ $\pi \sim \text{Beta}(1, 1)$			

liability. For a given choice on the prior distribution, as R decreases corresponding to reduced requirement on reliability, the minimum sample size, n_{b}, decreases for a fixed choice on c. This matches our intuition that fewer units need to be tested for demonstrating lower requirement on reliability.

The BRDTs are useful for demonstrating reliability requirements for binary tests. For example, a test plan $(n_{\text{b}} = 81, c = 5)$ for a predetermined test period of 5 years can demonstrate no less than 0.9 reliability in 5 years with the CR controlled by 0.05. However, it offers no capability of demonstrating reliability before the end of the test period. For example, if the customers are particularly concerned about the reliability in the first two years in addition to the reliability by the end of the five years, the conventional BRDTs are unable to demonstrate all requirements over multiple time periods. In addition, BRDTs are unable to differentiate and demonstrate reliability requirements involving multiple failure modes associated with different consequences. In the next two sections, two categories of new MSRDTs are proposed to demonstrate reliability require-

ments over multiple time periods and for multiple failure modes, respectively. Alternative designs are also proposed and their performances are evaluated and compared under different prior elicitation.

2.3 MSRDTs over Multiple Time Periods

Conventional BRDTs often demonstrate the product reliability within a single time period, such as during the mission time or the service life, to meet with the customers' requirements. However, customers' satisfaction in different time periods may differ. For instance, upon the purchase of products, customers may expect higher reliability during the early lifetime. The occurrence of early failures may have stronger negative impact on customers' satisfaction and company's reputation than failures occurred in the later stage of the service period. To explicitly demonstrate different product reliability requirements over multiple time periods rather than a single time period, the strategies of MSRDTs, i.e., multi-state RDTs, are proposed in this section to meet customers' expectation on reliability over multiple time periods.

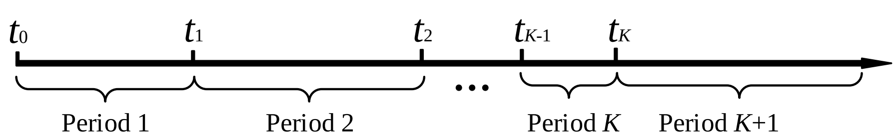


Figure 2.7: Illustration of multiple time periods in K periods between $(t_0, t_K]$

Consider a finite testing period with the start time at t_0 and end time at t_K. The testing time duration $(t_0, t_K]$ is exclusively partitioned into K non-overlapping time periods, $(t_{i-1}, t_i], i = 1, ..., K$, as illustrated in Figure 4.25. Let π_i and y_i denote the probability of failure and the number of observed failures within the i^{th} time period $(t_{i-1}, t_i]$, respectively. Then the number of units that survive the entire test duration (right-censored at the end of the test duration t_K) can be expressed as $n - \sum_{i=1}^{K} y_i$, where n is the total number of test units. The probability of surviving the test is given by $\pi_{K+1} = 1 - \sum_{i=1}^{K} \pi_i$. The objective of a MSRDT over multiple time periods is to simultaneously demonstrate

the product reliability at multiple time points satisfying a set of lower reliability requirements, $R_i, i = 1, \cdots, K$, with the assurance level controlled at $(1 - \beta)$. Here, R_i is the minimum acceptable reliability in the first i cumulative time periods, $(t_0, t_i]$, β is the maximum acceptable consumer's risk and assurance level can be explained as the minimum level of probability that the test can be passed (Hamada et al., 2008, pp.343-346). Two different scenarios of acceptance criteria are proposed as follows: Scenario I is the MSRDT that will be passed if the cumulative number of observed failures $\sum_{k=1}^{i} y_k$ at each cumulative time period $(t_0, t_i]$ is no more than its corresponding cumulative maximum number of allowable failures $\sum_{k=1}^{i} c_k$ for all cumulative time periods $(t_0, t_i]$, at $i = 1, ..., K$. For example, consider a two-period MSRDT with tests conducted at the end of the second and fifth year. For 100 test units, the MSRDT will be passed if the number of observed failures in first two years do not exceed 1 and the number of observed failures at the end of the fifth years do not exceed 5. Scenario II is the MSRDT that will be passed if the number of observed failures y_i at each non-overlapping time period $(t_{i-1}, t_i]$ is no greater than its corresponding maximum number of allowable failures c_i for all time periods $(t_{i-1}, t_i]$, at $i = 1, ..., K$. For such two-period test, the MSRDT will be passed if the number of observed failures in first two years do not exceed 1 and the number of observed failures in the next three years do not exceed 4. It is noticed that the major difference between the two scenarios is that scenario II plans the tests for non-overlapping time periods while scenario I considers the cumulative time-periods instead.

For each acceptance criterion, the design of MSRDT over multiple time periods aims to determine (i) the minimum sample size, denoted by n_{I} and n_{II} for scenarios I and II, respectively, and (ii) the cumulative maximum number of allowable failures at time t_i, $\sum_{k=1}^{i} c_k$, for scenario I and the maximum number of allowable failures within i^{th} time period, c_i, $i = 1, ..., K$ for scenario II. For either scenario, the MSRDT is selected by choosing the test plans which control the CR at or below β. It is noticed that the proposed MSRDT strategies are suitable for demonstration tests that generate failure counts data (Li et al.,

2016; Guo et al., 2010b) over multiple time periods, and do not make any assumptions on the failure time distribution. The advantages of the proposed methods are to fulfill the reliability requirements of customers over different testing periods (e.g., either cumulative time periods from scenario I or separate periods from scenario II) simultaneously and provide different testing strategies that require different minimum testing sample sizes based on different maximum numbers of allowable failures. Assuming a certain failure time distribution over multiple time periods or for multiple failure modes may considerably limit the use of the proposed strategies because the whole testing period needs to be considered and only expected number of failures can be obtained, which is not commensurate with the objectives of proposed strategies as mentioned above. Alternative RDT designs such as Weibull testing which is more suitable for failure time data, is out of the scope of this paper, but is of interest for future work.

To illustrate the proposed MSRDTs over multiple time periods and further investigate the difference between two scenarios of acceptance criteria, the MSRDTs over two time periods (i.e., $K = 2$) are considered without loss of generality. Let R_1 and R_2 denote the minimum acceptable reliabilities over the time periods $(t_0, t_1]$ and $(t_0, t_2]$ with $R_2 < R_1$. The probabilities of failure for each time period meet the requirements if $\pi_1 \leq 1 - R_1$ and $\pi_1 + \pi_2 \leq 1 - R_2$. For acceptance criterion in scenario I, the test plan of MSRDT is to determine $\{n_{\mathrm{I}}, c_1, c_1 + c_2\}$, and the probability of accepting the test plan for any given (π_1, π_2), denoted by $H_{\mathrm{I}}(n, c_1, c_2)$, can be explicitly written as

$$H_{\mathrm{I}}(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=0}^{c_1+c_2-y_1} [(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!}) \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n-y_1-y_2}]$$

and with $p(\pi_1, \pi_2)$ denoting the joint prior distribution of $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2)$, the corresponding CR_{I} is controlled at or below β by

$$\mathrm{CR}_{\mathrm{I}} = 1 - \frac{\int_0^{1-R_1} \int_0^{1-R_2-\pi_1} H_{\mathrm{I}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_{\mathrm{I}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1} \leq \beta. \tag{2.3}$$

For the acceptance criterion in scenario II, the MSRDT plan can be determined by specifying $\{n_{\text{II}}, c_1, c_2\}$, and the probability of accepting the test plan for any combination of (π_1, π_2), denoted by $H_{\text{II}}(n, c_1, c_2)$ is given by

$$H_{\text{II}}(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=0}^{c_2} [(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!}) \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n - y_1 - y_2}],$$

and the corresponding CR_{II} is controlled by

$$\text{CR}_{\text{II}} = 1 - \frac{\int_0^{1-R_1} \int_0^{1-R_2-\pi_1} H_{\text{II}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_{\text{II}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1} \leq \beta. \tag{2.4}$$

A case study is shown below for illustrating the proposed MSRDT strategies for a two-period test. The reliability requirements are set as $R_1 = 0.8$ and $R_2 = 0.6$ over the time periods $(t_0, t_1]$ and $(t_0, t_2]$ with $t_2 < 2t_1$, which indicates longer time interval of $(t_0, t_1]$ than $(t_1, t_2]$. Hence, a higher reliability requirement R_1 is desired for the early cumulative time period $(t_0, t_1]$ because the customers are averse to early failures. The CR is controlled at $\beta = 0.05$, indicating that the probability of accepting the test when the actual reliability requirements are not met is controlled at or below 0.05. To evaluate the complex integration in either Eq. (2.3) or Eq. (2.4), Monte Carlo sampling is performed with the sample size of $M = 15000$ to maintain the evaluation accuracy. The Dirichlet distribution, denoted by Dirichlet$(\alpha_1, \alpha_2, \alpha_3)$, is used as the prior distribution for $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2)$, where $\alpha_1, \alpha_2, \alpha_3$ are hyper-parameters to be elicited based on the prior knowledge. The Dirichlet distribution is a family of continuous multivariate probability distribution parametrized by the vector of positive hyper-parameters $\alpha_i, i = 1, \cdots, K$ for K categories of outcomes. The advantage of using Dirichlet distribution is two folded. First of all, it is the conjugate prior for the multinomial distribution, and hence can facilitate a convenience of updating knowledge as new data are observed because the posterior distribution of the failure probabilities also follow a Dirichlet distribution. Second, the hyper-parameters in the Dirichlet distribution are associated with more intuitive practical implications as they

are directly connected with the failure probabilities for each category of outcomes based on the prior knowledge in the form of $\alpha_i / \sum_{i=1}^{K} \alpha_i$. A few different settings of hyper-parameters will be explored later to investigate the impact of prior knowledge on the performance of the proposed test plan.

Table 2.3: Comparison between scenarios I & II and BRDT, with non-informative prior

Scenario I			Scenario II			BRDT	
c_1	$c_1 + c_2$	n_{I}	c_1	c_2	n_{II}	c	n_{b}
0	0	12	0	0	12	0	5
0	1	13	0	1	13	1	8
1	1	15	1	0	17		
0	2	14	0	2	14	2	11
1	2	17	1	1	18		
2	2	19	2	0	22		
0	5	20	0	5	20	5	18
1	5	22	1	4	21		
2	5	24	2	3	23		
3	5	26	3	2	28		
4	5	28	4	1	33		
5	5	30	5	0	37		
0	6	22	0	6	22	6	20
1	6	24	1	5	23		
2	6	26	2	4	24		
3	6	28	3	3	28		
4	6	30	4	2	33		
5	6	32	5	1	37		
6	6	34	6	0	42		
Settings: $p(\pi_1, \pi_2) \sim \text{Dirichlet}(1, 1, 1)$							
$R_1 = 0.8, R_2 = 0.6, M = 15000, \beta = 0.05$							

When no prior information is available, a non-informative prior distribution, given by $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2) \sim \text{Dirichlet}(1, 1, 1)$ can be used for indicating the lack of prior knowledge. The selected test plans under the acceptance criteria of two scenarios with different choices on the maximum number of allowable failures are illustrated in Table 2.3. The test plans are grouped based on the total number of failures allowed during the entire test duration. Several features are observed. First of all, under both scenarios I & II, given a fixed choice of c_2, the minimum sample size n_{I} or n_{II} increases as c_1 increases. Similarly,

given a fixed c_1, n_{I} and n_{II} also increase with c_2. As for a given fixed number of test units, allowing more failures (i.e. increasing c) can make it easier to pass the test and thus increase the CR. Hence, it requires to test more units to control the CR at a predetermined maximum acceptable level. The patterns of minimum sample sizes can be observed more clearly in Figure 2.8 and Figure 2.9.

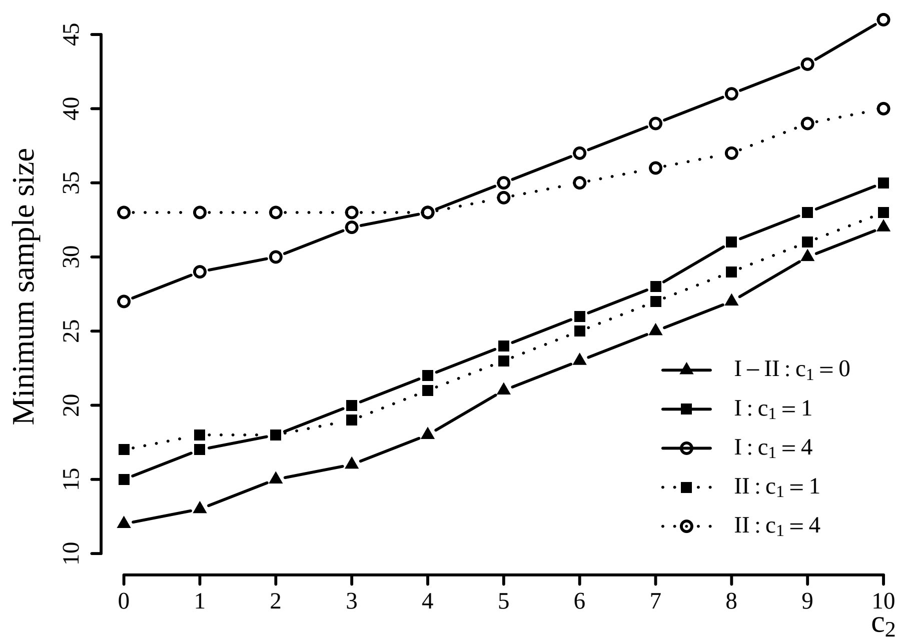


Figure 2.8: Comparison between scenario I & II based on the minimum sample size as c_2 increases for some fixed c_1 values

Figure 2.8 shows the change in the minimum sample size as c_2 increases for a few selected c_1 values under both scenarios. Solid lines are used for showing scenario I and dash lines are used for scenario II. Different symbols are used for displaying different c_1 values. For a fixed c_1 value, the minimum sample sizes under both scenarios I & II increase as c_2 increases. For example, when $c_1 = 0$, two scenarios are essentially the same in terms of the acceptance criteria. Hence, the same minimum sample size is required for both scenarios, which is shown with the solid line with the triangles and increases as c_2 increases. When $c_1 > 0$, the minimum sample size still generally increases as c_2

increases. However, the trend is slightly different between the two scenarios. The n_{I} increases monotonically with c_2, while the n_{II} starts off with similar sample sizes for small c_2 values to a certain point and then starts to increase more prominently as c_2 increases. For example, when $c_1 = 4$, the minimum sample size for scenario II (shown with a black dash line with the open circles) is relatively flat for $c_2 \leq 4$ and then increases for $c_2 > 4$. This is because under scenario II, the maximum number of allowable failures for the two non-overlapping periods determines their corresponding minimum required test units, which then jointly determine the overall minimum sample size for the entire test. Therefore, the overall sample size can be dominated by the maximum number of allowable failures for one of the test periods if one of the c_i is considerably larger compared to its failure probability under the reliability requirements to be demonstrated. Thus, when c_2 is small, c_1 plays an dominating role in determining the overall sample size for the entire test, which corresponds to the flat portion of the minimum sample size curve for each fixed c_1. However, as c_2 becomes larger than c_1, the overall minimum sample size is dominated by the requirement from period 2 and hence resumes an increasing pattern as c_2 increases. To compare between the two scenarios, it appears that n_{I} is usually larger than n_{II} for small c_2 values, but becomes smaller than n_{II} when c_2 becomes larger than a certain value. This is because for the same required c_i values, the test plans in scenario I generally can allow larger maximum number of allowable failures for period 2 (when the maximum number of allowable failure is not reached during period 1) and hence request to test more units when c_2 has dominating impact on the overall minimum sample size.

Figure 2.9 shows how the minimum sample size changes with c_1 for fixed c_2 values under both scenarios. Generally, for any fixed c_2, the minimum sample size increases as c_1 increases under scenario I. Also, a larger c_2 value requires to test more units and the difference in n_{I} among different c_2 values are similar across different c_1 values, which is evidenced by the almost parallel lines observed for Scenario I in Figure 2.9. However, for scenarios II, even though n_{II} increases monotonically with c_1, there are diminishing

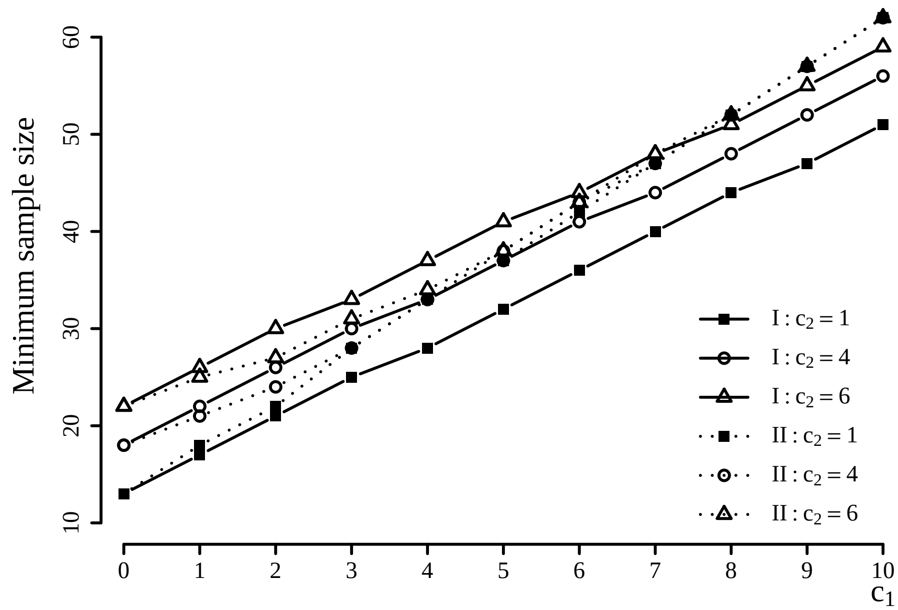


Figure 2.9: Comparison between scenario I & II based on the minimum sample size as c_1 increases for some fixed c_2 values

differences in n_{II} at different c_2 values as c_1 increases. This is because under scenario II, increasing c_1 will affect n_{II} by increasing the minimum sample size needed to demonstrate the reliability requirement in Period 1 and hence leads to a dominating effect on the size of n_{II} (which is equivalent to a diminishing impact of the difference in c_2 values). While under scenario I, increasing c_1 will result in increases in the minimum sample sizes needed for demonstrating both reliability requirements at the end of the two non-overlapping time periods, and hence has a consistent impact on the overall minimum sample size n_{I}.

It is also interesting to compare the two scenarios given the same total maximum number of allowable failures $c_1 + c_2$ in the entire test duration. Figure 2.10 compares the minimum sample sizes for both scenarios given a fixed total maximum number of allowable failures $c_1 + c_2$. Two cases with $c_1 + c_2 = 15$ and $c_1 + c_2 = 20$ are investigated, which are shown in Figure 2.10 with the solid and dash lines, respectively. The bottom and the top axes display all combinations of c_1 and c_2 values. A few patterns can be observed. First,

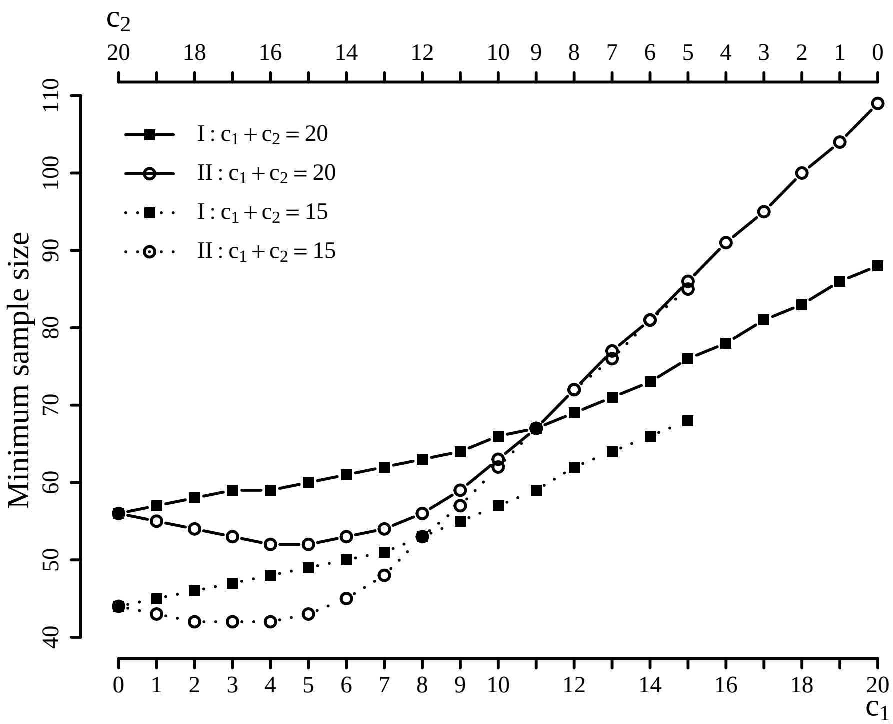


Figure 2.10: Comparison between scenario I & II based on the minimum sample size for fixed $c_1 + c_2$ values

both n_{I} and n_{II} increase as $c_1 + c_2$ increases. This matches with the pattern for the conventional BRDTs in that it generally requires to test more units to ensure the same assurance level if a more relaxed criterion has been used for passing the test by allowing more failures to be observed during the entire test duration. Second, increasing c_1 (at the same time reducing c_2) will consistently increase n_{I} but reduce n_{II} first for small c_2 values and then increase n_{II} after c_2 reaches a certain value. Third, in terms of the relative performance of the two strategies, scenario II is associated with smaller overall minimum sample size for large c_1 and small c_2 values. As c_2 increasing to about the same size as c_1, scenario I starts to have a smaller minimum sample size and the difference becomes larger as c_1 continues to increase. This can be evidenced by the crossover pattern between the monotonically increasing line with the squares for scenario I and the U-shaped curve with the open circles for scenario II. Brief analytical explanations are provided in Appendix for Chapter 2 to improve understanding of the observed differences between two scenarios.

Under the same maximum number of allowable failures $c_1 + c_2$ for the entire test duration, scenario II is expected to have stricter requirements ($y_1 \leq c_1, y_2 \leq c_2$) than that of scenario I ($y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2$), meaning that any tests that pass in scenario II will also pass in scenario I. Intuitively, scenario II will be preferred if minimizing the CR is the only criterion of interest, which on the other hand generally requires larger minimum sample size. However, smaller test sample is also generally preferred in RDT plan from the manufacturer's point of view. Hence, the tests with minimum sample size after controlling the CR are generally preferred. As illustrated in Figure 2.10, the two test scenarios may have varied performance in the required minimum sample size for different settings and scenario II does not consistently outperform scenario I based on the minimum sample size. It is also noticed in Table 2.3 that the difference between the two scenarios when c_1 is small becomes smaller for small $c_1 + c_2$ values, and is almost negligible for $c_1 + c_2 \leq 6$. On the other hand, Scenario I can be preferred for relatively large c_1 values when $c_1 + c_2$ is large or when only small $c_1 + c_2$ is allowed. It is also noted that for tests using more strict passing conditions, they are generally associated with smaller probabilities of passing the test (i.e. low acceptance probability) and often higher probabilities for manufacturers to reject the products that actually have met the reliability requirements (Lu et al., 2016). Hence, the decision on selection of scenarios should be catered for a particular application to meet the objectives of specific demonstration tests.

In addition, Table 2.3 also shows the comparison between the MSRDT strategies over two time periods with the conventional BRDTs when non-informative prior is used. The last two columns in Table 2.3 give the maximum number of allowable failures and the minimum sample size for demonstrating the reliability requirement at the end of test duration (i.e. the end of period 2). For any given total maximum number of allowable failures over the entire test duration, $c = c_1 + c_2$, the conventional BRDTs require to test fewer units for demonstrating only a single reliability at the end of the test. The MSRDTs, on the other hand, gain the capability of demonstrating multiple reliability requirements

at different time points by paying slightly more price for testing a few more units. However, as $c = c_1 + c_2$ increases, fewer extra units are required to be tested for demonstrating more reliability requirement at multiple time points. For example, for $c = 5$, the conventional BRDT requires to test 18 units to demonstrate reliability at the end of the two-year period as 0.6. To demonstrate an additional higher reliability at the end of the first year at 0.8, both MSRDT strategies require to test at least 20 units with no failure allowed to be observed during the first year. More units need to be tested if more failures are allowed to be observed during the first year.

Table 2.4: Minimum sample sizes required by the two-period MSRDT using the acceptance criterion in scenario I for different prior distributions

Dirichlet		$(1,1,1)$	$(3,3,24)$	$(6,6,18)$	$(12,3,15)$	$(3,12,15)$	$(6,12,12)$	$(12,6,12)$
c_1	$c_1 + c_2$				n_{I}			
0	0	12	1	22	58	22	30	67
0	1	13	1	22	58	25	32	66
1	1	15	2	24	62	25	33	69
0	2	14	2	23	58	28	35	65
1	2	17	3	26	62	28	35	69
2	2	19	4	27	66	28	36	72
0	5	20	5	27	60	35	42	66
1	5	22	5	29	65	36	43	69
2	5	24	5	32	69	36	43	73
3	5	26	8	34	73	36	44	76
4	5	28	11	36	76	36	44	79
5	5	30	13	37	79	37	44	81
0	6	22	6	28	60	38	45	67
1	6	24	6	31	65	39	45	70
2	6	26	6	33	70	39	46	73
3	6	28	8	36	74	39	46	77
4	6	30	11	38	77	39	47	80
5	6	32	14	39	81	39	47	82
6	6	34	16	40	83	39	47	84
Settings: $M = 15000, R_1 = 0.8, R_2 = 0.6, \beta = 0.05$								

It is well known that incorporating different prior information may have large impacts on the results in Bayesian analysis. Next we explore the impact of different prior elicitations on the selected MSRDT plans under both scenarios. Table 2.4 and Table 2.5 summa-

Table 2.5: Minimum sample sizes required by the two-period MSRDT using the acceptance criterion in scenario II for different prior distributions

Dirichlet		(1,1,1)	(3,3,24)	(6,6,18)	(12,3,15)	(3,12,15)	(6,12,12)	(12,6,12)
c_1	c_2	n_{II}						
0	0	12	1	22	58	22	30	67
0	1	13	1	22	58	25	32	66
1	0	17	2	27	64	24	33	73
0	2	14	2	23	58	28	35	65
1	1	18	2	27	63	27	35	71
2	0	22	5	32	69	27	37	79
0	5	20	5	27	60	35	42	66
1	4	21	5	29	65	35	42	70
2	3	23	5	33	69	34	43	74
3	2	28	7	38	74	34	44	80
4	1	33	10	43	79	34	45	87
5	0	37	14	48	86	34	47	97
0	6	22	6	28	60	38	45	67
1	5	23	6	30	65	37	45	71
2	4	24	6	34	70	37	45	74
3	3	28	7	38	74	36	46	79
4	2	33	9	43	79	36	47	85
5	1	37	13	48	84	36	49	93
6	0	42	18	54	92	36	51	104
Settings: $M = 15000, R_1 = 0.8, R_2 = 0.6, \beta = 0.05$								

rize the required minimum sample sizes for the MSRDT plans over two test periods with different choices of prior distributions for scenarios I & II, respectively. Seven different prior distributions including the non-informative prior, Dirichlet$(1,1,1)$, are explored. The patterns are rather consistent across Table 2.4 and Table 2.5. Under both scenarios, when the prior distribution support higher reliabilities than the minimum requirements, such as Dirichlet$(3,3,24)$ shown in the fourth column in both tables, the minimum sample size can be substantially reduced for any given combinations of c_1 and c_2 values than using the non-informative prior (shown in the third column in both tables).

On the other hand, if the prior distribution supports reliabilities at or below the requirements, more units need to be tested to demonstrate the requirements than using the non-informative prior. This can be observed in Figure 2.11 and Figure 2.12 which show

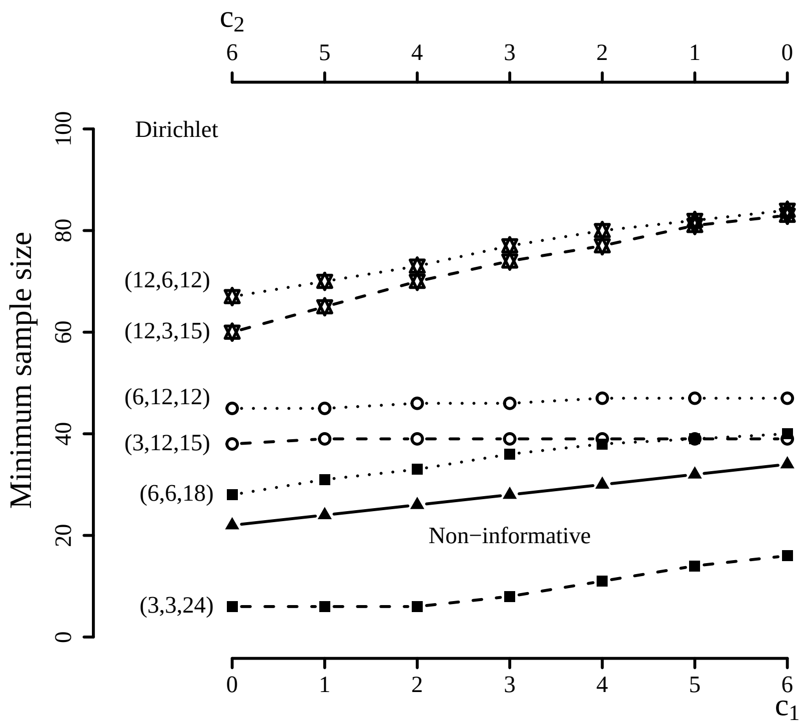


Figure 2.11: Minimum sample sizes required in scenario I with fixed $c_1 + c_2 = 6$ for different prior distributions

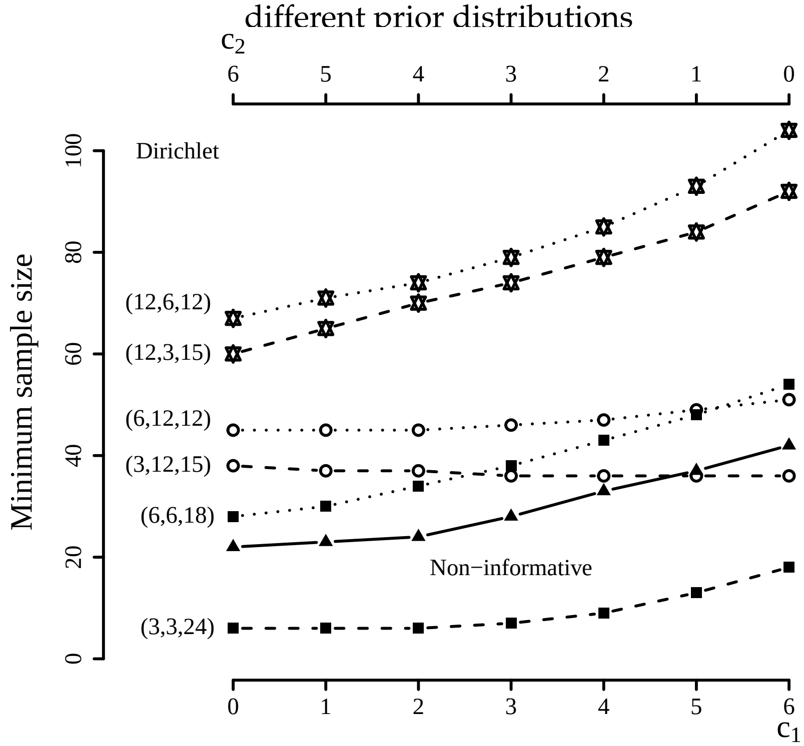


Figure 2.12: Minimum sample sizes required in scenario II with fixed $c_1 + c_2 = 6$ for different prior distributions

the minimum sample size for fixed $c_1 + c_2$ under Scenario I & II, respectively. In both figures, the solid lines with triangles represent the sample sizes for different (c_1, c_2) combinations using a non-informative prior. The dash lines with squares show the sample sizes for a prior distribution Dirichlet$(3, 3, 24)$ that supports higher reliabilities than the requirements, which are consistently below the non-informative line. All other prior distributions support reliabilities at or below the requirements, and hence all require to test more units with the corresponding lines located above the non-informative line. The farther the specified prior distribution is to the reliability requirements, the more test units are needed in the MSRDTs over multiple time periods. One special case is the dash line observed in Figure 2.12 for a prior distribution Dirichlet$(3, 12, 15)$, which is consistently below the non-informative line indicating smaller minimum sample sizes are required for all (c_1, c_2) combinations. Since the prior distribution regarding period 1 supports higher reliabilities than the requirements, while the prior distribution regarding period 2 supports reliabilities below the requirements, the effects of sample size reduction from period 1 and sample size increase from period 2 may jointly determine the overall minimum sample size, and hence lead to slightly different pattern than what has been observed for other prior distributions.

2.4 MSRDTs for Multiple Failure Modes

In the previous section, the MSRDT strategies consider each time period as an individual state for demonstrating specific reliability requirement within the time period. This section proposes a different category of MSRDTs which considers different failure modes as individual states that are often associated with different consequences of failures and different costs of replacement. The conventional BRDTs report dichotomous outcomes (i.e., success and failure) for each test unit, in which case different failure modes of the product are not differentiated and the severity levels of different consequences associated with different failures modes are overlooked. In real applications, a product often has

multiple failure modes in varied levels of severity, which can lead to different impacts on customers' dissatisfaction.

For instance, the failure of a CPU or a hard drive of a computer system is much more critical than the failures of some accessory parts such as a keyboard or a microphone, since the former can lead to a complete break down of the overall system, a loss of valuable information and/or a high repair/replacement cost while the latter usually only results in system under-performance and a low repair/replacement cost. Consequently, the failures of critical or valuable parts will lead to stronger dissatisfaction of customers, and hence result in higher expectation on reliability for these components. It is desirable to develop test strategies that allow demonstrating separate reliability requirements for multiple failure modes.

Assume a product has J independent failure modes. For each test unit, it will either have failed in mode $j, j = 1, ..., J$ or remain working by the end of the testing period. Let π_j and y_j denote the probability of failure and the number of observed failures in failure mode j within the test period (or an equivalent mission time period), respectively. Then, $\pi_{J+1} = 1 - \sum_{j=1}^{J} \pi_j$ and $n - \sum_{j=1}^{J} y_j$ denote the probability of success and the number of survived units by the end of the test. The MSRDTs for multiple failure modes aim to demonstrate at an assurance level at $(1 - \beta)$ that the product reliability will meet multiple minimum reliability requirements for each of the different failure modes, denoted by $R_j, j = 1, \cdots, J$. Here, β is the CR on having a product that has passed the demonstration test but fails to meet all reliability requirements for different failure modes. Note that all failure modes are defined in the same test period. For any specified reliability requirements R_j's and the maximum acceptable CR controlled at or below β, the MSRDTs for multiple failure modes are designed to determine the minimum sample size n_{m} as well as the maximum number of allowable failures c_j in the jth failure mode for $j = 1, ..., J$.

Without loss of generality, considering two failure modes with $J = 2$ for illustrating the proposed MSRDT strategy. Let R_1 and R_2 denote the minimum acceptable reliabilities

for failure modes 1 and 2, respectively. The test is passed if the number of observed failures y_j is less or equal to the maximum number of allowable failures c_j for both failure modes, and the test plan is to determine the choice on $\{n_{\mathrm{m}}, c_1, c_2\}$. For independent failure modes, the acceptance probability $H_{\mathrm{m}}(n, c_1, c_2)$ for certain (π_1, π_2) values can be written as

$$H_{\mathrm{m}}(n, c_1, c_2) = \sum_{y_1=0}^{c_1} [(\frac{n!}{y_1!(n-y_1)!})\pi_1^{y_1}(1-\pi_1)^{n-y_1}] \sum_{y_2=0}^{c_2} [(\frac{n!}{y_2!(n-y_2)!})\pi_2^{y_2}(1-\pi_2)^{n-y_2}]$$

and the corresponding CR, denoted by CR_{m}, is calculated by

$$\mathrm{CR}_{\mathrm{m}} = 1 - \frac{\int_0^{(1-R_1)} \int_0^{(1-R_2)} H_{\mathrm{m}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_{\mathrm{m}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}, \tag{2.5}$$

where $p(\pi_1, \pi_2)$ is the joint prior distribution of (π_1, π_2). For independent failure modes, there is $p(\pi_1, \pi_2) = p(\pi_1)p(\pi_2)$. The minimum sample size is determined by controlling the CR_{m} obtained in Eq. (2.5) to be at or below β. Simulation case studies are conducted for exploring different reliability requirements, maximum numbers of allowable failures for different failure modes, as well as different prior elicitations and their impacts on the required minimum sample size for the MSRDTs for two failure modes. The results are summarized in Tables 2.6 and 2.7 for two cases with similar or different reliability requirements for the two failure modes. In Table 2.6, identical minimum reliability requirements are assumed for the two failure modes, where $R_1 = R_2 = 0.8$ indicates that the customers have the same expectation on reliability for both failure modes. Table 2.7 assumes different reliability requirements with $R_1 = 0.8$ and $R_2 = 0.6$. Here, failure mode 1 is considered more critical and/or have more severe consequences associated with its failure, and hence is required for a higher reliability. The CR_{m} is still controlled at $\beta = 0.05$ and the sample size for Monte Carlo sampling is chosen as $M = 15000$ to maintain the

simulation accuracy. Beta distributions are used for specifying the prior distributions for both π_1 and π_2 for the two failure modes.

Table 2.6: Multiple failure modes with the same reliability requirements for different prior distributions

Beta		π_1 / π_2	$(1,1)$ / $(1,1)$	$(2,18)$ / $(2,18)$	$(4,16)$ / $(4,16)$	$(10,15)$ / $(10,15)$	$(2,18)$ / $(4,16)$	$(2,18)$ / $(10,15)$	$(4,16)$ / $(10,15)$
c_1	c_2					n_{m}			
0	0		16	7	22	70	18	48	60
0	1		20	9	25	75	23	55	66
1	0		20	9	25	71	19	45	57
0	2		25	12	29	81	29	61	73
1	1		22	11	28	75	24	52	62
2	0		25	12	30	74	20	45	56
0	5		40	22	43	103	44	81	94
1	4		35	19	39	93	39	70	79
2	3		31	18	36	84	34	62	72
3	2		31	19	36	80	30	56	67
4	1		35	20	40	82	27	50	61
5	0		40	23	45	88	27	45	57
0	6		45	25	47	112	49	88	102
1	5		40	22	43	100	44	76	85
2	4		35	21	40	90	39	68	77
3	3		34	21	39	84	34	61	72
4	2		35	21	40	83	31	55	66
5	1		40	23	45	86	29	50	62
6	0		45	27	50	94	30	46	59
Settings: $M = 15000, R_1 = 0.8, R_2 = 0.8, \beta = 0.05$									

When two failure modes have the same reliability requirements at $R_1 = R_2 = 0.8$, Table 2.6 summarizes the minimum sample size with different choices of the maximum allowable failures and different prior settings. When no prior information is available, a non-informative prior distribution of Beta$(1,1)$ is assigned for both π_1 and π_2. Similar patterns can be observed as for the MSRDTs over multiple time periods. When c_1 is fixed, the minimum sample size n_m increases as c_2 increases; when c_2 is fixed, n_m increases with c_1. This is intuitive as having more allowable failures makes it easier to pass the test and thus increases the CR. To control a reasonable CR, a larger number of test units need to be tested by allowing more failures to be observed during the test. When $c_1 + c_2$ is fixed, the

Table 2.7: Multiple failure modes with different reliability requirements for different prior distributions

Beta		π_1	$(1,1)$	$(2,18)$	$(10,10)$	$(4,16)$	$(2,18)$	$(10,10)$
		π_2	$(1,1)$	$(2,18)$	$(10,10)$	$(10,15)$	$(10,10)$	$(2,18)$
c_1	c_2				n_{m}			
0	0		13	3	105	19	17	59
0	1		14	3	101	20	20	63
1	0		19	7	116	25	18	64
0	2		14	3	97	21	23	66
1	1		19	7	111	24	20	68
2	0		24	10	126	31	18	68
0	5		19	5	85	26	31	70
1	4		20	7	99	27	28	74
2	3		24	11	113	30	26	77
3	2		29	14	128	35	24	79
4	1		34	18	143	44	23	80
5	0		38	21	158	56	22	80
0	6		21	6	82	28	33	69
1	5		21	7	95	28	31	75
2	4		24	11	109	30	28	79
3	3		29	14	123	35	26	82
4	2		34	18	138	42	25	84
5	1		39	22	154	52	24	85
6	0		43	25	169	65	24	84
Settings: $M = 15000, R_1 = 0.8, R_2 = 0.6, \beta = 0.05$								

minimum sample size n_m exhibits a symmetric pattern under the non-informative prior setting due to the identical reliability requirements for both failure modes. For example, when $c_1 + c_2 = 6$, the minimum sample sizes for $c_1 = 0, c_2 = 6$ and $c_1 = 6, c_2 = 0$ are identical. In addition, when c_1 and c_2 become more similar in size (e.g., $c_1 = 2, c_2 = 4$ compared to $c_1 = 0, c_2 = 6$), it requires smaller minimum sample size to remain the same assurance level for demonstrating the requirements on both failure modes. This makes sense as when the maximum number of allowable failures is considerably larger for one failure mode given the same reliability requirement, it requires to test more units for demonstrating the requirement for this failure mode, which then inflates the overall minimum sample size needed in the MSRDT for demonstrating reliability requirements for both failure modes.

Different prior elicitations also have large impacts on the selected test plan, as shown in Table 2.6. When prior knowledge supports higher reliability than the requirements to be demonstrated, fewer units need to be tested and vice versa. For instance, the prior distributions of $\pi_1 \sim \text{Beta}(2, 18)$ and $\pi_2 \sim \text{Beta}(2, 18)$ indicate that there is a strong belief of lower failure probabilities than the requirements within the test period for both failure modes. Thus, the corresponding minimum sample size is smaller than what is needed for using the non-informative prior. On the other hand, when the prior distributions of $\pi_1 \sim \text{Beta}(10, 15)$ and $\pi_2 \sim \text{Beta}(10, 15)$ are used, which indicates a moderately strong belief in larger failure probabilities than the requirements for both failure modes, more units need to be tested to demonstrate the higher reliability requirements compared to what is needed when no prior information is available.

When $c_1 + c_2$ is fixed, the required minimum sample size is also sensitive to the specified prior distribution. Figure 2.13 illustrates the change in the n_m for different (c_1, c_2) combinations given fixed $c_1 + c_2 = 6$. When the non-informative priors are assumed, the curve for n_m (solid line with the triangles) shows a symmetric pattern with the minimum sample size achieved at $c_1 = c_2 = 3$. When informative priors indicating lower failure probabilities than requirements for both failure modes (such as $\pi_1 \sim \text{Beta}(2, 18), \pi_2 \sim \text{Beta}(2, 18)$ corresponding to the dash line with the open circles) are assumed, the minimum sample size curve is below the non-informative curve. As the prior belief indicates higher failure probability for at least one of the failure modes (such as $\pi_1 \sim \text{Beta}(2, 18), \pi_2 \sim \text{Beta}(10, 15)$ corresponding to the dotted line with the solid circles or $\pi_1 \sim \text{Beta}(10, 15), \pi_2 \sim \text{Beta}(10, 15)$ corresponding to the dash-dotted line with the open circles), the corresponding minimum sample size curve shifts upwards on at least one side of tails or on both sides.

Table 2.7 shows the test plans when different reliability requirements are used for the two failure modes with $R_1 = 0.8$ and $R_2 = 0.6$. When the non-informative priors are used, the symmetric pattern is no longer observed due to different requirements on relia-

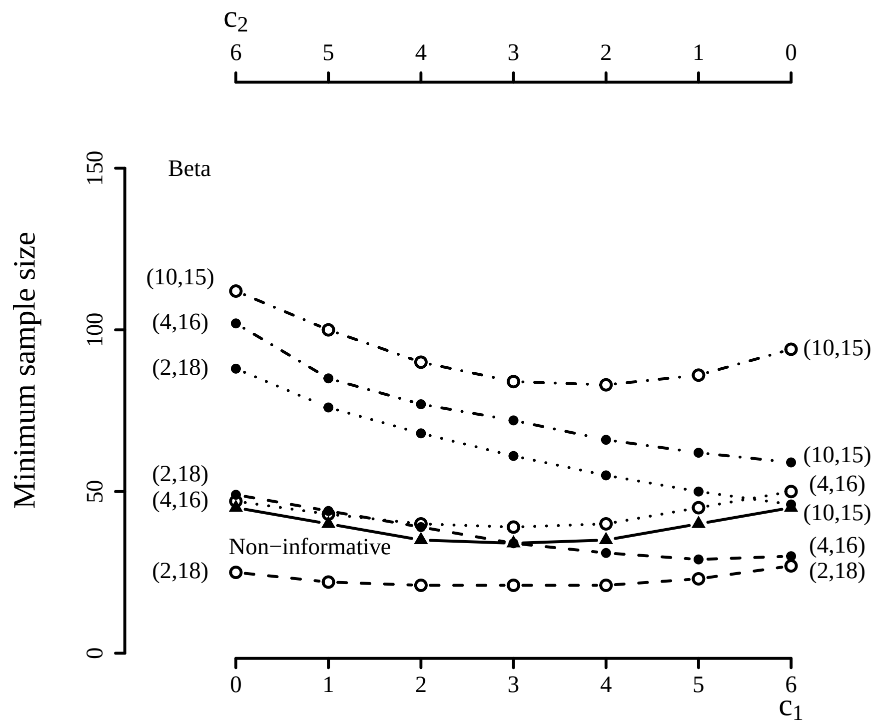


Figure 2.13: Multiple failure modes with the same reliability requirements for fixed $c_1 + c_2$ and different prior distributions

bility for the two failure modes. Particularly, n_m is larger when c_1 is large since more units need to be tested to demonstrate higher reliability requirement for failure model 1 while allowing more failures to be observed during the test period. Also, for the same c_1 and c_2 settings, the minimum sample size for demonstrating $R_1 = R_2 = 0.8$ is smaller than what is required for demonstrating $R_1 = 0.8$ and $R_2 = 0.6$ since fewer units can be tested to demonstrate a lower reliability requirement for failure mode 2. When more informative priors are used, similar patterns are observed from both Table 2.7 and Figure 2.14. A potential sample size reduction can be achieved when the prior knowledge supports higher reliability than what is required to be demonstrated by the MSRDT.

2.5 Concluding Remarks

Conventional binomial RDTs, which focus on demonstrating a single reliability requirement within a single test period, have limited use when multiple reliability require-

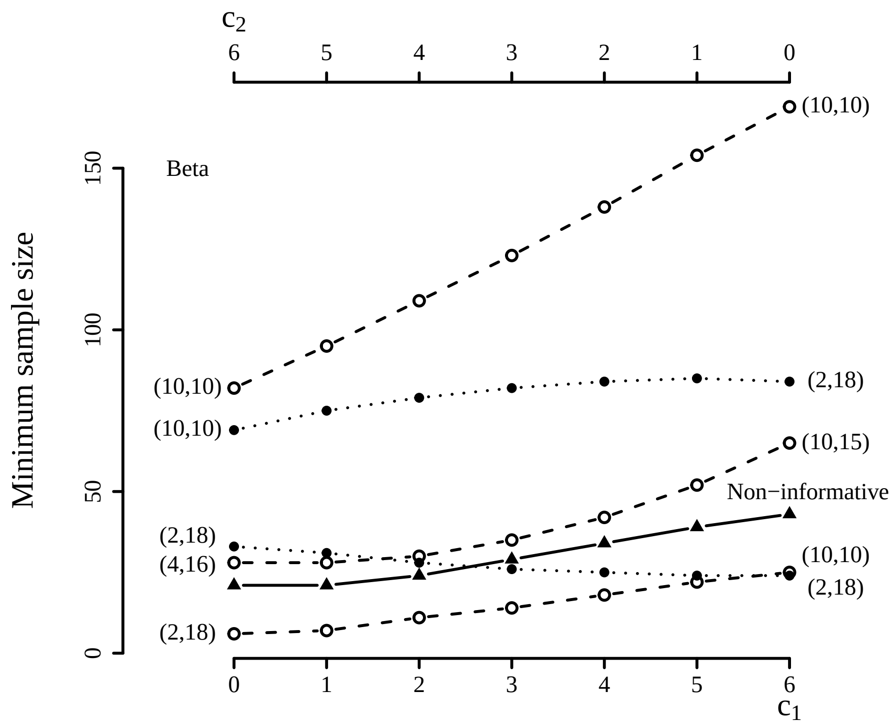


Figure 2.14: Multiple failure modes with different reliability requirements for fixed $c_1 + c_2$ and different prior distributions

ments need to be met. This paper proposes two types of RDTs for demonstrating reliability over multiple time periods and for multiple failure modes. These RDTs with multiple reliability requirements are all referred to as multi-state RDTs (MSRDTs).

In the MSRDTs over multiple time periods, every time period of interest is treated as a state, and the joint distribution of failure counts over the non-overlapping time periods can be modeled by a multinomial distribution. Two different test strategies are proposed for demonstrating multiple requirements over different time periods. One strategy uses the cumulative failure counts at the end of all periods as the criteria for passing the test; while the other uses separate failure counts over non-overlapping time intervals as the criteria for passing the test. Simulation studies were conducted for comparing the two strategies by considering two-period MSRDTs. It was founded that the strategy based on cumulative failure counts (scenario I) is generally preferred for cases that allow fewer total failure counts over all time periods or when a larger maximum number of allowable

failures is allowed for the early cumulative time period. The strategy using separate failure counts (scenario II) is only preferred for requiring smaller minimum sample size when a smaller maximum number of allowable failures is allowed for the early separate time period.

In the MSRDTs for multiple failure modes, each failure mode is treated as a state and all reliability requirements for the multiple failure modes that may be associated with different consequences in varied levels of severity and/or costs of repair/replacement can be simultaneously demonstrated. The required minimum sample size is usually determined mainly by the failure mode that has the highest reliability requirement and/or least stringent criterion for passing the test (i.e. allowing a larger maximum number of allowable failures for a particular failure mode).

The impacts of incorporating different prior distributions are also explored for both categories of MSRDTs. The patterns are consistent regardless of which test strategy is considered. When the prior knowledge supports higher reliability than the requirements to be demonstrated, fewer units can be tested compared to using the non-informative priors for demonstrating the same reliability requirements. However, if the historical data supports lower reliability than what are required to be demonstrated, then more units need to be tested to override the effects of the prior distribution for demonstrating higher reliability than what has been indicated from existing data. For future work, it is expected to develop thorough mathematical justifications with theoretical formulations and derivations to validate the discussed patterns using both non-informative and informative prior distributions.

Chapter 3

Optimal Binomial Reliability Demonstration Tests Design Under Acceptance Decision Uncertainty

3.1 Introduction

Reliability is the probability that a product will satisfactorily perform its intended function at a specified time point (Meeker and Escobar, 2014, pp.1-3). Because reliability measures the product quality over time, it plays an important role throughout the product life cycle, from product design and development to operations and warranty phases. Reliability demonstration test (RDTs) are one class of reliability assurance activities at the design and development phases for demonstrating a product's ability to meet an anticipated lower level reliability requirement at a desired level of confidence (Kleyner and Sandborn, 2008). To assure that a product has met the consumer's expectation, a RDT is conducted by the manufacturer to demonstrate the product's before it can be released to the market. The goal is to control the consumer's risk (CR) of receiving inferior products, and thus improve customer satisfaction and business reputation. As the global market becomes more and more competitive, the RDTs have become critical testing strategies for many companies to safeguard their market positions and improve the competitiveness of their products.

Based on different reliability data types and model assumptions (Meeker and Escobar, 2014, pp.3-15), the RDTs can be mainly classified into three categories: 1) RDT based on failure count data (Guo and Liao, 2011; Jensen, 2015; Chen et al., 2017; Li et al., 2016); 2) RDT based on failure time data (Hamada et al., 2008; Lee et al., 2015; Xu et al., 2017);

and 3) RDT based on degradation data (Yang, 2009, 2013; Jin and Matthews, 2014). The first type of RDT tests a set of units for a predetermined time period with the number of failed units observed(e.g., 5 failures out of 100 test units). In this case, the number of failures is often modeled with a discrete distribution, such as a binomial distribution for single-failure units or a Poisson distribution for recurrent failure units. The second testing scheme considers failure time data where time to failure is observed for failed units and the censoring time is observed for units that have not failed by the end of the testing duration. A lifetime distribution, e.g., the Weibull or Lognormal distribution, is often assumed for modeling failure time data. In the last testing scheme, the RDTs are designed based on the degradation data, which mainly use the repeated measurements of the deteriorating performance over time to infer and demonstrate product reliability based on specified critical failure thresholds.

This paper focuses on the design of RDTs based on the failure count data which are modeled with a binomial distribution, and hence are also referred to as binomial RDTs (BRDTs) (Lee et al., 2015). The BRDTs have been widely employed in reliability engineering and practices due to its convenience in actual implementation and analysis. Compared to RDTs based on failure time or degradation data, which require either periodic inspection or continuous monitoring of the failed/deteriorating states of product units, the BRDTs only need to inspect the units at the end of the test and report the number of failures (Guo and Liao, 2011), and hence require much less monitoring efforts. In addition, the BRDT designs do not require any assumption on the lifetime distribution or degradation processes. They are particularly useful for one-shot systems with destructive testing (Guo et al., 2010b) (e.g., the missiles).

In the conventional BRDT design (Jensen, 2015; Kleyner et al., 2015; Lu et al., 2016) with n test units, given a lower level reliability requirement, R_L, at the end of a predetermined test period, the test is passed if the number of observed failures, y, does not exceed the maximum number of allowable failures, c. The BRDT design aims to determine the

design parameters (n, c), with the actual product reliability satisfying $R \geq R_L$ and an acceptable consumer risk (CR) level β. β quantifies the probability that a unit mistakenly passes the test but fails to meet the specified R_L. To find the optimal BRDT design, the conventional approach focuses on minimizing the BRDT testing cost by selecting an optimal plan determined by $(n^*, c = 0)$, where n^* is the minimum sample size when no failure is allowed to pass the test. These zero-failure tests, are also called the success run tests (O'Connor and Kleyner, 2012, pp.357-358). However, there has been limited work considering the costs generated from broader reliability relevant activities. Kleyner and Sandborn (2008) was among the first to consider a life cycle cost for reliability demonstration. However, in this paper, the discussion of trade-offs between the BRDT testing cost and warranty cost still stay at a conceptual level without taking into consideration of the uncertainty of the BRDT testing results, which is associated with the probabilistic relationships between the BRDT and its subsequent reliability assurance activities. To the best of our knowledge, there has been no study that explicitly incorporates the acceptance decision uncertainty for the optimal BRDT design and investigates the cost-effectiveness of such optimal design by integrating its subsequent reliability assurance activities, including the reliability growth (RG) and warranty services (WS).

To further explain the acceptance decision uncertainty of the BRDT design and how it will affect subsequent reliability assurance activities, Figure 3.15 provides a flowchart of the three related activities, i.e., RG, BRDT and WS. Based on the observed failures y and the design parameter c, a binary decision of rejecting or accepting the BRDT will be made, associated with initiating either the RG or WS. At the design stage of BRDT, the number of observed failures y is random, resulting in the uncertainty of such a binary decision. Different configurations of the BRDT design will further influence the probability of the binary decision and whether RG or WS being initiated. Moreover, different BRDT designs will also affect the potential costs of RG and WS. For instance, if the BRDT is accepted, then designing a BRDT for demonstrating a higher reliability may reduce the WS cost

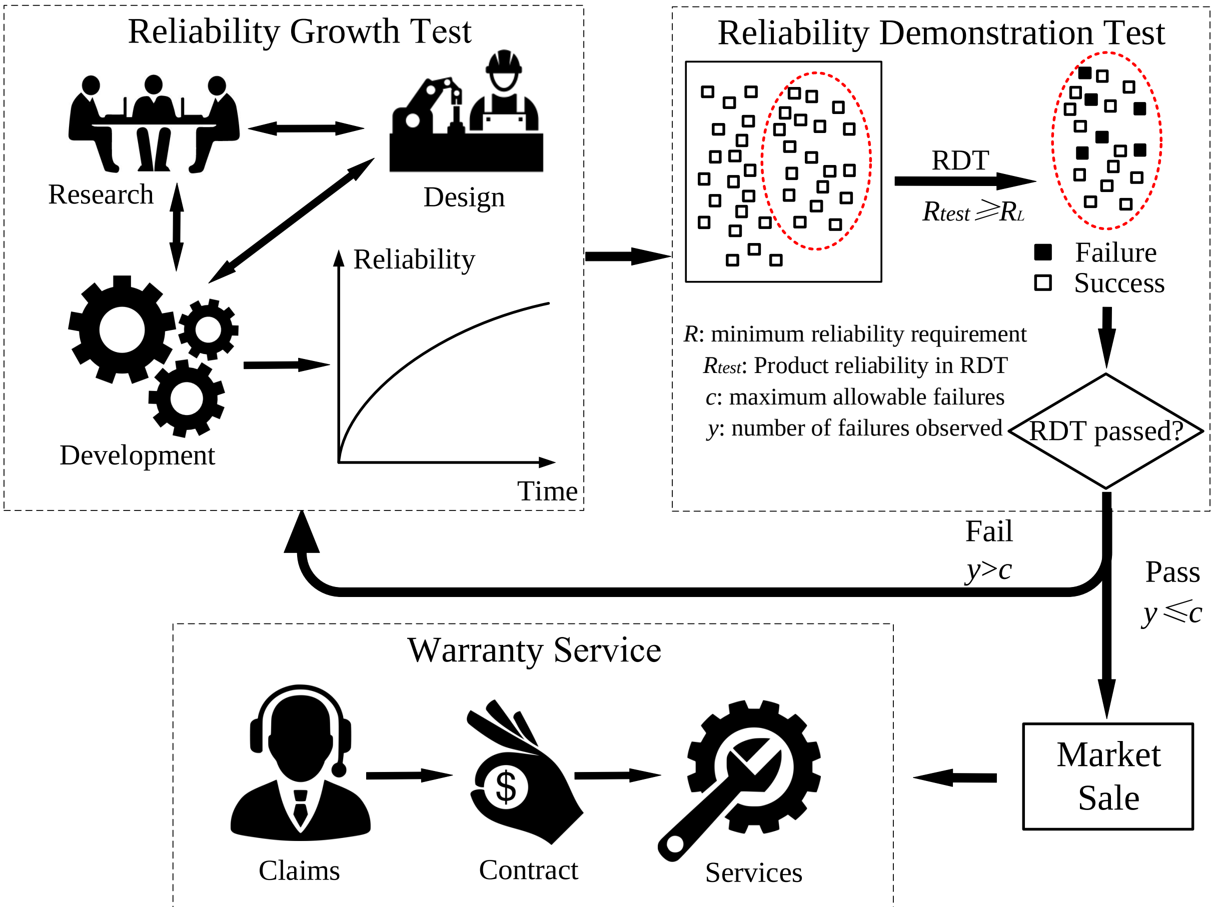


Figure 3.15: Reliability assurance activities in the product life cycle

since the number of warranty claims is expected to decrease for more reliable products. But on the other hand, if the BRDT is rejected, then improving product to achieve such higher reliability goal may also increase the RG cost. However, conventional optimal BRDT designs, such as success run tests, do not take into account the acceptance decision uncertainty or its cost impacts on subsequent reliability activities. In fact, success run tests often have a low probability of accepting the product, and then the RG will be more likely to be initiated.

To address the aforementioned gaps in the existing literature, there is a need to develop an optimal BRDT design strategy which explicitly considers the acceptance decision uncertainty as well as the cost impacts on the RG and WS. To achieve this objective, there are three aspects needing to be investigated. First, it is necessary to quantify the ac-

ceptance decision uncertainty and examine the nonlinear relationship between the design parameters and the acceptance uncertainty as well as the costs of BRDT testing, RG and WS. Second, different products may have different cost components throughout the product life cycle, and hence requires different optimal designs for optimizing the overall cost. For example, mission-critical high-tech systems, such as missiles and space aircrafts, may require extensive costs in RG for technology improvement while daily-used consumer products, such as telecommunication devices, may consume more costs in WS to safeguard customer satisfaction and market competitiveness. Third, in the existing Bayesian BRDT design literature, the impact of prior elicitation on the test design has been explored (Hamada et al., 2008; Chen et al., 2017; Weaver and Hamada, 2008). When the decision uncertainty and different cost components are incorporated in the Bayesian BRDT design, it is necessary to understand how the choices of prior distributions will influence the selected test designs. To answer these questions and address the various design complexities, this paper proposes an optimal BRDT design framework by explicitly quantifying and incorporating the acceptance decision uncertainty and minimizing the overall cost of BRDT testing, RG and WS. The relationships between the design parameters and the various aspects of the test performance will be studied, and the trade-offs between the reliability characteristics and the different cost components will be evaluated. A comprehensive case study is carried out to systematically investigate the performance of proposed BRDT design under different cost scenarios and extract some general patterns. The influence of prior elicitation on BRDT design is also studied to understand its impacts on the selected sample size and overall cost.

The remaining of the paper is organized as follows. In the next section, the conventional BRDT designs are reviewed under both the Frequentist and Bayesian frameworks. Then, the optimal BRDT design is proposed by quantifying of the acceptance decision uncertainty and the trade-offs among different cost components. A comprehensive case study is followed to illustrate the proposed method and compared with conventional de-

signs through a real-world example. In the following sections, different cost scenarios will be investigated, and a sensitivity analysis on the specified prior distribution and its influence on the selected test plan and the overall cost will be explored. Conclusions are provided in the end.

3.2 Conventional BRDT Design

In the conventional design of BRDT, design parameters (n, c) can be determined under both the Frequentist and Bayesian approaches (Kececioglu, 2002, pp.415-429), where n is the number of units to be tested and c is the maximum allowable failures for accepting the test. Under the Frequentist approach, (n, c) are determined by controlling the $\mathrm{CR_F}$ calculated by

$$\mathrm{CR_F}(n, c) = \sum_{y=0}^{c} \binom{n}{y} (1 - R_L)^y (R_L)^{n-y}, \tag{3.6}$$

where y is the number of observed failures during test, which follows a binomial distribution, $y \sim \mathrm{Bin}(n, \pi)$ with a true product failure probability π. The set of values (n, c) are chosen to ensure $\mathrm{CR_F}(n, c) \leq \beta$ for all $1 - \pi \geq R_L$, where β and R_L are the tolerance level of CR and the lower level reliability requirement from the consumers, respectively. There are many choices of (n, c) values satisfying the above constraints. The conventional optimal BRDT design aims to minimize the cost of the demonstration test, or equivalently, the testing sample size n. Since the minimum testing sample size will increase as c increases, the optimal design parameters become $(n^*, c = 0)$, where n^* is the smallest sample size required for a zero-failure test (i.e., success run test).

Under the frequentist framework and the constraints in zero failure tests, once the R_L and β are specified by the users, the optimal design parameters is uniquely determined. In contrast, the Bayesian method allows the incorporation of prior knowledge of product reliability (i.e., $1 - \pi$), which can potentially reduce the testing sample size if the prior knowledge is in favor of the specified R_L (Hamada et al., 2008; Pintar et al., 2012; Weaver

and Hamada, 2008; Wilson and Fronczyk, 2017). By using the Bayesian approach, the design parameters (n, c) are determined by controlling the posterior CR calculated in the following form by conditioning on the prior knowledge of passing the test

$$\begin{aligned} \mathrm{CR_B}(n, c) &= \Pr(1 - \pi < R_L | y \leq c) \\ &= \frac{\int_{1-R_L}^{1} [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}{\int_{0}^{1} [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}. \end{aligned} \tag{3.7}$$

In the above equation, $\mathrm{CR_B}(\cdot)$ represents the posterior probability of the product reliability not meeting the lower level reliability requirement (i.e., $1 - \pi < R_L$) given the test is accepted (i.e., $y \leq c$), and $p(\pi)$ denotes the prior probability density of the true product failure probability π under the Bayesian framework. By controlling the posterior CR at or below β, the minimum sample size n^* can be determined. It has been demonstrated in existing work that incorporating prior information sometimes can help reduce testing sample size (Chen et al., 2017; Lu and Rudy, 2001). In this paper, we will also use the Bayesian method to leverage information from historical data or subject matter expertise for designing optimal BRDT plans.

The conventional optimal BRDT design using either the Frequentist or Bayesian approach can be expressed as an optimization problem in a unified framework given by

$$\begin{aligned} \underset{n,c}{\text{minimize}} \quad & D = C_f + C_v n \\ \text{subject to} \quad & \mathrm{CR}(n, c) \leq \beta, \\ & n \in \mathbb{Z}^{>}, c \in \mathbb{Z}^{\geq}, \end{aligned} \tag{3.8}$$

where D is the BRDT testing cost with C_f representing the total fixed cost (e.g., testing facility and equipment costs, training and labor costs of personnel, administrative costs,

etc.) independent of testing sample size, and $C_v n$ representing the total variable cost (e.g., material costs, manufacturing and assembling costs of each test unit) for n test units. C_v is the average variable cost per test unit. $\mathbb{Z}^{\geq}$ is the set of all non-negative integers and $\mathbb{Z}^{>}$ is the set of all positive integers. $\mathrm{CR}(\cdot)$ can be calculated based on either Eq. (4.14) or Eq. (4.15). The optimal BRDT is essentially a combinatorial optimization problem aiming to find the optimal integer combinations of (n, c) in order to minimize the BRDT testing cost D, or equivalently, the number of testing units n.

3.3 Optimal BRDT Design Under Acceptance Decision Uncertainty

Conventional optimal BRDT design described in Eq. (3.8) has several limitations. First, it only focuses on reducing the BRDT testing cost while neglects the cost impacts on subsequent reliability assurance activities, such as the RG and WS. Second, as described in Figure 3.15, the binary decision of accepting or rejecting the test is uncertain at the BRDT design stage, which further results in the uncertainty of initiating the following activities of either RG or WS. Conventional optimal design fails to incorporate such uncertainty in test planning. The proposed cost-based optimal BRDT design will consider the integrated overall cost from the BRDT testing, RG and WS subject to the acceptance decision uncertainty.

Let ξ be a binary variable to represent the binary decision of BRDT, namely $\xi = 1$ when test is passed and $\xi = 0$, otherwise. The acceptance criterion for a BRDT is that a test will be passed (i.e., $\xi = 1$) if the number of failures y is not greater than the maximum allowable failures c. At BRDT planning stage, the actual realization of y is unknown before a test is implemented and y is a random quantity following a binomial distribution with parameters n and π, whose prior density is $p(\pi)$. Thus, when designing a BRDT, there is an acceptance decision uncertainty associated with the actual test to be implemented. Such uncertainty can be explicitly quantified by the acceptance probability (AP) (Lu et al., 2016), i.e., $\Pr(\xi = 1) = \Pr(y \leq c)$. Under the Bayesian framework, AP can

be evaluated based on the prior predictive distribution of y by "averaging" over all the possible values of π based on its prior density as:

$$
\begin{aligned}
\mathrm{AP} &= \Pr(y \leq c) \\
&= \int_0^1 [\sum_{y=0}^{c} \binom{n}{y} (\pi)^y (1-\pi)^{n-y}] p(\pi) d\pi.
\end{aligned} \tag{3.9}
$$

It is noticed that AP essentially quantifies the prior belief probability of future BRDT acceptance uncertainty while AP and (1-AP) determine the likelihood of releasing the product into the market with the WS initiated and the chance of being sent back for RG, respectively. Since AP is influenced by both design parameters (n, c) in Eq. (3.9), and the choices of (n, c) are also influenced by R_L from controlling CR in Eq. (4.15), it will be interesting to understand how AP changes with different design parameters, which is also helpful for understanding its cost impacts on different reliability activities. Figure 3.16 shows the nonlinear relationships among AP, c, n^* and R_L by controlling CR under 0.05, where n^* is the optimal solution (i.e., minimum testing sample size) obtained from Eq. (3.8) for any fixed c value. Several patterns can be observed from Figure 3.16. First, when c is fixed and R_L increases, n^* will increase but AP will decrease. It is because demonstrating a higher product reliability requires testing more units. Then more units are likely to fail, which makes the passing criterion $y \leq c$ less likely to be satisfied. Second, when R_L is fixed and c increases, both n^* and AP will increase. This is intuitive as increasing the maximum allowable failures makes it easier to pass the test (i.e., increased AP) but also requires testing more units to ensure the same R_L is met. Also, given a fixed lower level reliability requirement R_L specified, the conventional success run tests with $(c = 0)$ generally have the lowest AP values as compared with other test plans with higher c values.

Then the proposed optimal BRDT design can be explicitly formulated as follows by incorporating AP and calculating the expected costs of subsequent reliability activities,

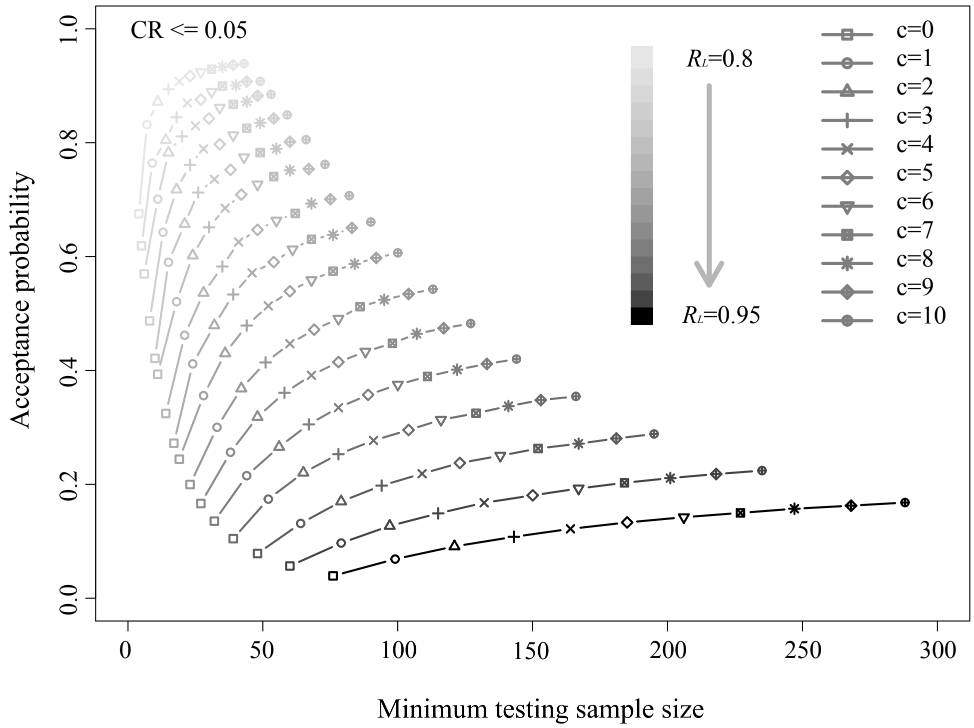


Figure 3.16: Trade-offs among R_L, c, n^* and AP, with controlled CR

namely WS and RG,

$$
\begin{aligned}
\underset{n,c,R}{\text{minimize}} \quad & D + E_{\xi}[\xi W + (1-\xi)G] \\
& = \underbrace{C_f + C_v n}_{\text{BRDT testing cost}} + \underbrace{\text{AP} \times E_{\pi|y\le c}(C_w * N * \pi)}_{\text{expected WS cost}} + \underbrace{(1-\text{AP}) \times G}_{\text{expected RG cost}} \\
\text{subject to} \quad & \text{CR}(n,c) \le \beta, \\
& \underline{R}_L \le R_L \le \overline{R}_L, \\
& n \in \mathbb{Z}^{>}, c \in \mathbb{Z}^{\ge},
\end{aligned}
\tag{3.10}
$$

where $\mathbb{Z}^{>}$ and $\mathbb{Z}^{\ge}$ represent the sets of positive and non-negative integers, respectively. The objective of the optimal design is to minimize the BRDT testing cost, D, together with the expected costs of WS and RG, where W and G denote the total WS and RG costs, respectively. An acceptable range $[\underline{R}_L, \overline{R}_L]$ is provided for R_L, with $\underline{R}_L$ denoting the lower

bound of R_L that corresponds to the pre-specified lower level reliability requirement in conventional BRDT design, and $\overline{R}_L$ denoting the upper bound of R_L that can be technically achieved in practical applications. If neglecting AP, excluding the last two cost components and pre-specifying $R_L = \underline{R}_L$, the conventional optimal design in Eq. (3.8) can be viewed as a special case under the proposed BRDT design. Given a BRDT test is accepted, the anticipated total WS costs to be involved becomes $W = E_{\pi|y \leq c}(C_w * N * \pi)$, where N the expected sales volume, C_w is the average cost per warranty claim and $\pi|y \leq c$ is the true product failure probability conditioning on a test has been passed. W can be explicitly calculated as

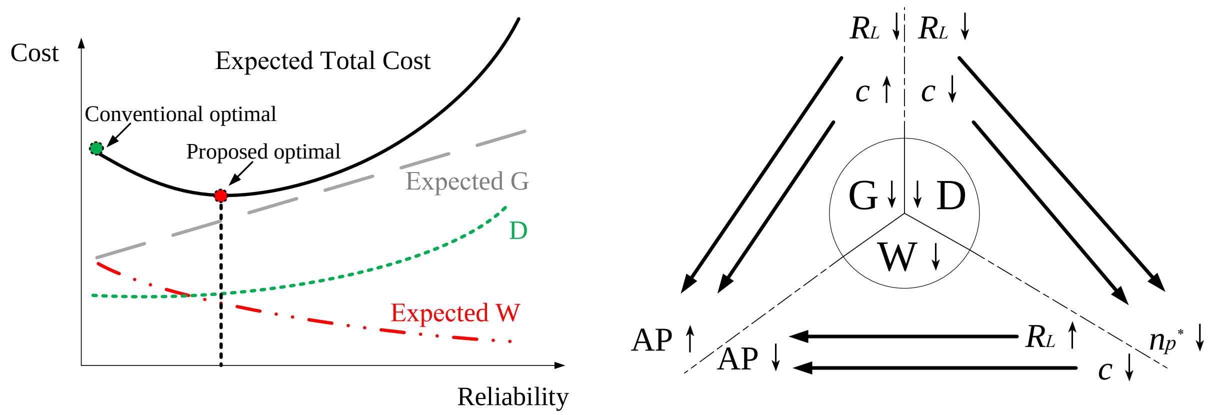
$$\begin{aligned} E_{\pi|y \leq c}(C_w * N * \pi) &= C_w N E(\pi|y \leq c) \\ &= C_w N \frac{\int_0^1 [\pi \sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}{\int_0^1 [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}. \end{aligned} \tag{3.11}$$

G is the total anticipated RG costs in the single redesign cycle before performing the next round of BRDT, which includes the resources (e.g., time, personnel, equipment, materials, etc.) spent in identifying failure modes in the existing design and/or manufacturing process and further making corrective actions (e.g., design changes, quality control, etc.) to eliminate such failure modes and strengthen the product reliability. Note in the Eq. (3.10), we use aggregate cost structure for the WS and RG costs to illustrate the general methodology as well as evaluate their impacts on BRDT design. In a specific application context, both WS and RG costs may vary significantly and need to be modeled more explicitly. To facilitate more general application of the proposed method, in the case study section, we will investigate a variety of cost scenarios for BRDT, WS and RG to gain a comprehensive understanding of the general patterns and offer insights for general applications.

It is noticed that in practice, to shorten the testing duration at the BRDT testing stage, BRDT test is often carried out under the accelerated test conditions (e.g., usage rate, temperature, voltage, etc.) to ensure that the product reliability in the test period is equivalent to the product reliability in the warrant period at the normal operating condition. In other words, given the warrant period t_{w} under the normal operating condition x_{w} and the specified BRDT testing duration t_{a}, an accelerated condition x_{a} for BRDT testing can be determined by satisfying $T(x_{\mathrm{a}}) = T(x_{\mathrm{w}})/\gamma$, where $T(x_{\mathrm{a}})$ and $T(x_{\mathrm{w}})$ are random quantities of product lifetime under accelerated and normal operating conditions, respectively, and $\gamma = T(x_{\mathrm{w}})/T(x_{\mathrm{a}})$ is an accelerating factor. The product reliability then becomes $1 - \pi = \Pr(T(x_{\mathrm{a}}) > t_{\mathrm{a}}) = \Pr(T(x_{\mathrm{w}}) > t_{\mathrm{w}})$. and the test duration of BRDT can be shorten as $t_{\mathrm{a}} = t_{\mathrm{w}}/\gamma$. Based on a specific accelerated life testing (ALT) model, $T(x_{\mathrm{a}})$ and $T(x_{\mathrm{w}})$ can be transformed accordingly with $\Pr(T(x_{\mathrm{a}}) > t_{\mathrm{a}})$ being equivalent to $\Pr(T(x_{\mathrm{a}}) > t_{\mathrm{a}})$. For instance, considering a log-normal ALT model at operating condition x, i.e., $T(x) \sim \mathrm{Lognormal}(\mu(x), \sigma^2)$, where location parameter $\mu(x)$ depends on the operating condition x and scale parameter σ is fixed, there is $T(x_{\mathrm{a}}) = T(x_{\mathrm{w}})/\gamma$ when $\mu(x_{\mathrm{a}}) = \mu(x_{\mathrm{w}}) - \log(\gamma)$. Similarly, when considering a Weibull ALT model at operating condition x, i.e., $T(x) \sim \mathrm{Weibull}(\lambda(x), k)$, where scale parameter $\lambda(x)$ depends on the operating condition x and shape parameter k is fixed, there is $T(x_{\mathrm{a}}) = T(x_{\mathrm{w}})/\gamma$ when $\lambda(x_{\mathrm{a}}) = \lambda(x_{\mathrm{w}})/\gamma$. For additional details of ALT models, please refer to (**?**Nelson, 2009).

In addition to the consideration of the expected costs of subsequent reliability activities, another new contribution of this paper is that we consider the impact of the choice of R_L on optimizing the test plan. In the conventional BRDT design, the lower level reliability requirement R_L is fixed at a specific value, e.g., $R_L = \underline{\mathrm{R}}_L$. However, when the WS and RG costs are considered, the choice of R_L will also affect these costs. For example, a higher R_L will request testing more units and hence an increased BRDT testing cost in the first cost component of Eq. (3.10). This will also result in a lower AP and a higher chance to require further RG activities with an increased expected RG cost in the third cost compo-

nent. On the other hand, a lower R_L will generally lead to an increased expected WS cost in the second cost component. It is because the second cost component can be reduced into $C_w N \int_0^1 [\pi \sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi$ based on Eq. (3.9) and Eq. (3.11). Given a fixed c and a lower R_L, the sample size n required to demonstrate a less demanding R_L will decrease, which will further increase the above term. Thus, we consider R_L as an additional design parameter. Instead of fixing $R_L = \underline{R}_L$, we consider a range of acceptable values for R_L, debited by $[\underline{R}_L, \overline{R}_L]$, and seek the set of (n, c, R_L) design parameters that minimizes the overall cost expressed in Eq. (3.10). As the relative size of different cost components varies, different optimal test plans will be selected. For instance, when the WS cost of a product is dominating the total cost, a higher R_L within the acceptable range will be preferred to reduce the WS cost.



(a) Changing patterns of cost components and optimal decisions

(b) Trade-offs among three cost components and the determination of design parameters

Figure 3.17: Illustrations of changing patterns and trade-offs of cost components

Figure 3.17 illustrates the trade-offs between different cost components considered in the proposed work. In Figure 3.17a, as R_L increases, both the BRDT testing cost and expected RG cost will increase while the expected WS cost will decrease. As a result, the overall cost exhibits a convex changing pattern. Given an acceptable range $[\underline{R}_L, \overline{R}_L]$ of R_L, the proposed BRDT design by simultaneously determining n, c and R_L values can

yield the smallest overall cost. Figure 3.17b summarizes the cause-and-effect relationship among design parameters, AP and three cost components. For example, to reduce D alone, BRDT can be designed via adjusting R_L and/or c towards lower value(s), which will reduce the total testing sample size n and BRDT testing cost D. As shown in Figure 3.17b, three cost components cannot be minimized simultaneously by adjusting c or R_L, which demonstrates the trade-off among different cost components. The proposed optimal BRDT design will minimize the overall cost by identifying the optimal design parameters (n_p^*, c^*, R_L^*), where n_p^* denotes the optimal testing sample size for the proposed work and differentiates from n^* in conventional work.

Numerical examples will be illustrated in case study section to demonstrate the effectiveness of the proposed work in determining optimal BRDT designs. Both CR and AP in Eq. (3.10) can be calculated under Bayesian framework using Monte Carlo integration (Robert and Casella, 2013, pp.71-131), where a large number of samples of π of size M (e.g., 5000) are generated from the specified prior distribution $p(\pi|x)$. Then, CR and AP can be calculated approximately by

$$\text{Posterior CR} \approx 1 - \frac{\sum_{j=1}^{M}[\sum_{y=0}^{c}\binom{n}{y}(\pi^{(j)})^{y}(1-\pi^{(j)})^{n-y}]I(\pi^{(j)} \leq 1 - R_L)}{\sum_{j=1}^{M}[\sum_{y=0}^{c}\binom{n}{y}(\pi^{(j)})^{y}(1-\pi^{(j)})^{n-y}]}, \tag{3.12}$$

$$\text{AP} \approx \frac{1}{M}\sum_{j=1}^{M}[\sum_{y=0}^{c}\binom{n}{y}(\pi^{(j)})^{y}(1-\pi^{(j)})^{n-y}], \tag{3.13}$$

where $\pi^{(j)}$ is the jth generated sample of true product failure probability for the specified prior distribution.

3.4 Case Study

This section illustrates the proposed optimal BRDT design through a case study from a real application, and demonstrates its advantages over the conventional designs for minimizing the overall reliability related cost. In addition, we will explore different scenarios with varied sizes of the cost components (e.g., costs of BRDT testing, WS and RG) to understand their impacts on the trade-offs among different cost components, how design parameters drive changes in different cost components, different choices of optimal designs under different cost structures, and the impact of prior information on selected BRDT designs.

3.4.1 Proposed vs Conventional – Real-world Example

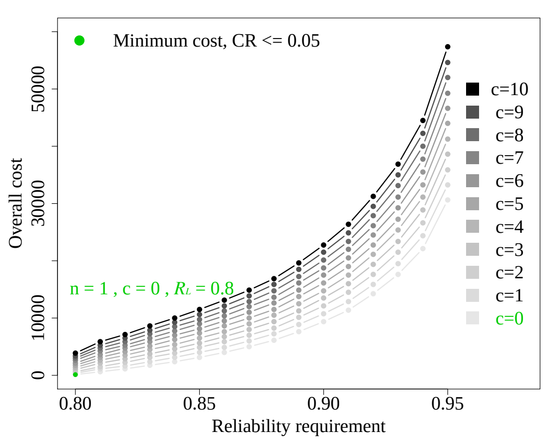
Pham and Turkkan (1992) provides a real-world example regarding the reliability of fuel pump. Based on historical testing data, and the engineers' belief that there is 0.9 probability that the reliability for the fuel pumps to operate in the next 1000 hours will fall between 0.80 and 0.93. This translates into a prior distribution of the fuel pump's true failure probability as $\pi \sim \text{Beta}(9, 60.86)$ (Duran and Booker, 1988). The beta prior distribution was selected because (i) it is the conjugate prior for the binomial distribution, which facilitates computation efficiency; and (ii) it is flexible in representing different shapes (e.g., flat, symmetric, right-/left-skewed) of prior densities.

The prior distribution of the fuel pump true failure probability described above is incorporated into both the conventional and proposed BRDT designs under the Bayesian framework. The CR is controlled at the acceptable level at $\beta = 0.05$. First, we assume the product has a comparable magnitude of aggregated costs in RG and WS, which are tremendously higher than the BRDT testing cost. This cost relationship can be expressed in the form $D \ll W \approx G$. This could happen when the product reliability growth process is complex and time-consuming and thus is associated with high expenses, and the WS

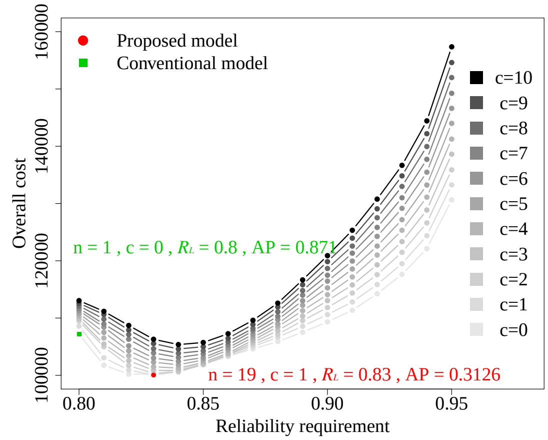
cost is also high due to either the high sales volume or high repair or replacement cost per warranty claim. In these cases, the BRDT testing cost is often negligible compared to costs of RG and WS due to the small number of units to be involved in the tests. Other scenarios with different relative sizes of D, W and G will be explored in the next section. In addition, we consider a reasonable acceptable range to be $0.8 \leq R_L \leq 0.95$ and we can allow c to range between 0 and 10. The minimum sample sizes, n^* and n_p^*, under conventional and proposed BRDT designs can be obtained for each combination of (R_L, c) under the controlled CR to minimize the BRDT testing and the overall costs in Eq. (3.8) and Eq. (3.10), respectively.

Figure 3.18 compares the cost of the conventional and proposed optimal BRDT designs for different combinations of (R_L, c) values. Figure 3.18a shows the overall cost of the conventional designs based on minimizing the BRDT testing cost. For each fixed c value, the overall cost increases as R_L rises since more units are required when the lower level reliability requirement to be demonstrated is higher. At any fixed R_L, the overall cost also increases as more failures are allowed to pass the test since this will also require testing more units to maintain an acceptable CR level. Additionally, the overall cost does not increase linearly with R_L. When R_L is at or above 0.9, the overall cost increases at a rising speed as reliability increases because a much larger number of units is required to be tested to suppress the impact of the prior distribution for demonstrating a lower level reliability requirement higher than what is supported by the prior knowledge (the reliability is centered around 0.87 with a high probability).

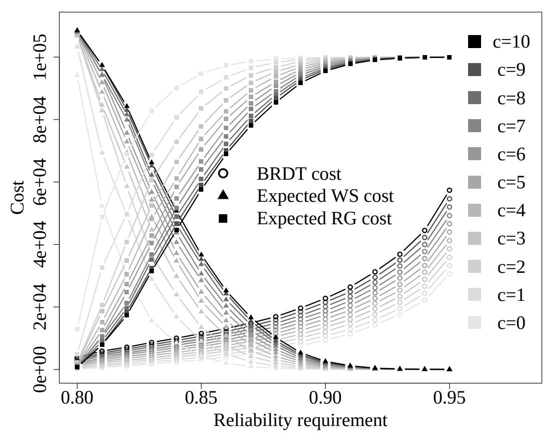
In contrast to the conventional designs, the overall cost of the proposed optimal BRDT design does not always increase monotonically with R_L or c. As shown in Figure 3.18b, we observe U-shaped patterns of the overall cost as R_L increases at different c values. For example, for any given c, the overall cost decreases with R_L at lower reliability values and then increases as R_L further increases i.e. exhibiting a U-shaped pattern. Since the overall cost in the proposed design is a combination of the BRDT testing cost as well as



(a) Cost changing patterns in conventional design considering only BRDT testing cost



(b) Cost changing patterns in proposed framework considering costs and AP



(c) Separate cost changing patterns of three cost components in the proposed framework

Figure 3.18: Cost comparison between conventional and proposed optimal BRDT designs

the expected WS and RG costs, different cost changing patterns essentially result from the various changing patterns of its composing cost components as well as AP. Similar to conventional designs, the BRDT testing cost is expected to increase with both R_L and c. The expected WS cost reduces as R_L increases at any fixed c value because fewer warranty claims are expected with an increased product reliability and there is also a smaller probability of having a successful test (i.e. AP reduces as R_L increases). On the other hand, at any fixed R_L, increasing c will increase the expected WS cost because there is a higher probability of passing the test (i.e. AP increases with c as shown in Figure 3.16). The

impact of c value on the expected WS cost is larger when $R_L < 0.9$ and is maximized at around $R_L = 0.83$. On the contrary, the expected RG cost is a decreasing function of c but an increasing function of R_L. The trade-off between the expected WS and RG costs leads to the U-shaped pattern of the overall cost in Figure 3.18b. It is also noticed that as R_L becomes greater than 0.9, both the expected WS and RG costs do not change much as AP becomes closer to 0. In that case, the changing patterns of the overall cost will resemble the patterns observed in the conventional design.

In addition to different cost changing patterns, the optimal test plans between conventional and proposed BRDT designs also differ, as shown in Figure 3.18b. Conventional design which minimizes the BRDT testing cost alone yields an optimal test plan of $(n^* = 1, c = 0)$ with pre-specified lower level reliability requirement $R_L = 0.8$. However, it gives a higher overall cost after considering the expected WS and RG costs. The proposed optimal test plan will select $(n_p^* = 19, c = 1, R_L = 0.83)$ to minimize the overall cost R_L selected from an acceptable range of [0.8,0.9]. The overall cost impacts are realized from two aspects. First, instead of pre-specifying a fixed lower level reliability requirement, adjusting R_L towards a higher level from an acceptable range of R_L will reduce the expected WS cost (with reduced AP and number of warranty claims) and provides good potential of further realizing the overall cost reduction (when WS cost is a dominant cost component). Second, instead of the smallest maximum allowable failures, higher c can be achieved as well to further decrease the probability of passing the test and the expected WS cost. In short, conventional design can be viewed as a special case of the proposed design by pre-specifying R_L and neglecting cost components of subsequent reliability assurance activities, such as WS and RG. The corresponding optimal test plan is therefore locally optimal in the context of minimizing the BRDT testing cost alone, rather than a global optimal test plan in a wider context by considering acceptance uncertainty and the overall cost with a combination of various cost components.

As described above, the overall cost changing pattern depends on both the trade-offs among different cost components as well as the specified prior distribution of product reliability. In the above example, we consider the relative sizes of the different cost components as $D \ll W \approx G$ (e.g., W and G are at least 10 times higher than D). As discussed above, the relative magnitude of D, W, and G as well as AP jointly the optimal test plan for minimizing the overall cost. Therefore, in the next section, we provide a comprehensive evaluation of other scenarios with varied size relationship among the three cost components, and examine how the overall cost changing patterns may vary among different scenarios. In addition, we also conduct a sensitivity analysis to understand the impact of different prior distributions on the selected optimal test plans.

3.4.2 Comprehensive Scenarios

In practice, the WS cost may vary considerably among different products with different costs of the warranty services, different warranty periods (e.g., 3 months, 2 years) and warranty policies (e.g., free replacement vs pro-rata) adopted by the manufacturers (Blischke, 1993, pp.40-90,131-205). In addition, the RG cost may also vary substantially among different products due to their diverse characteristics as well as various design, development and manufacturing efforts/technologies involved. For instance, compared to a simple component/device with matured technology and/or evolutionary design, the RG cost for a complex system with evolving technology and/or revolutionary design will be much higher. For RG with a single redesign cycle considered in this paper, RG cost is also influenced by the cost-effectiveness of corrective actions, the budget/resources constraints as well as management strategies in the current redesign cycle before performing the next round of BRDT, which will further lead to significant cost variation among diffident products. To accommodate different applications with varied WS and RG cost scenarios at the aggregate level, we will comprehensively investigate different cost sce-

narios in this section to obtain some general insights on how different RG and WS costs will affect the proposed BRDT design.

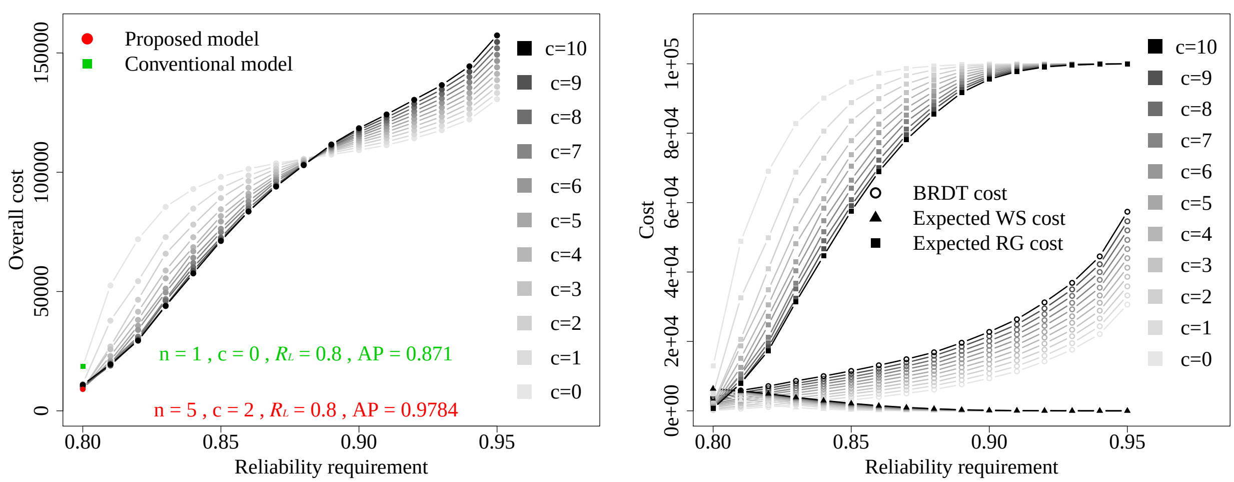
"To denote the relative size of two cost components, we use "$\approx$", "$<$" and "$\ll$" to represent "similar", "smaller" and "substantially small" in magnitude, respectively. Specifically, "$<$" represents a relation between two costs with approximately 1-3 times difference in magnitude while "$\ll$" represents a relation with at least 10 times difference in magnitude. Instead of exhausting all possible scenarios, we further restrict ourselves to consider only the cases where $D \ll G$ because for most of the manufacturing products, demonstration tests are usually a lot cheaper than RG as the latter often involves a lengthy and labor intensive efforts. For example, in automobile industry (Kleyner and Sandborn, 2008), the cost of new car design and development can be extremely expensive, which includes huge labor costs and operation costs up to $1 billion over a long time period for product design and RG. In the contrast, D can be relatively low, which may require testing hundreds of cars in a few months and cost several million dollars.

Table 3.8: Summary of different cost changing patterns and scenarios

Cost pattern	Scenario	n_p^*	R_L	c	AP	Costs (proposed)	Costs (conventional)
Increasing	I : $W << D << G$	14	0.8	5	0.9887	5.344E+03	2.113E+04
	II : $W < D << G$	14	0.8	5	0.9887	1.613E+04	3.055E+04
	III : $W \approx D << G$	11	0.8	4	0.9858	2.565E+04	3.886E+04
	IV : $D < W << G$	14	0.8	5	0.9887	3.574E+04	5.362E+04
	V : $D << W << G$	27	0.8	9	0.9935	2.047E+05	3.598E+05
U-shape	VI : $D << W < G$	27	0.8	9	0.9935	7.794E+05	8.582E+05
	VII : $D << W \approx G$	75	0.88	3	0.0474	1.504E+06	2.132E+06
	VIII : $D << G < W$	61	0.89	0	0.0026	1.508E+06	2.963E+06
Decreasing	IX : $D << G << W$	91	0.91	0	0.0004	1.514E+06	1.681E+07
Settings: $C_f = 1000, C_v = 125, C_w = 5, \pi \sim \text{Beta}(9, 60.86), M = 5000$(simulation sample size)							

Table 3.8 summarizes the cost changing patterns among the nine different ordering relationship scenarios with R_L ranging from 0.8 to 0.95 and c ranging from 0 to 10. The cost patterns can be classified into 3 categories, namely increasing, U-shaped and decreasing patterns. When G is a dominating cost component, the overall cost pattern will exhibit

an increasing trend as R_L increases, as shown by scenarios I to V. High-tech and mission-critical products/systems, such as integrated circuits and automobiles, requires extensive investments in R&D and can achieve higher reliability during warranty period, which leads to relatively less WS costs (O'Connor and Kleyner, 2012, pp.225-227)(Kleyner and Sandborn, 2008; Kleyner et al., 2004). When W becomes comparable with G, the trade-off between two expected cost components leads to the U-shaped patterns of the overall cost, as indicated by scenarios VI, VII and VIII. High performance components, such as durable car batteries, may have comparable costs in product development and warranty services. When W becomes a dominating cost component, the overall cost changing pattern will be a decreasing function with respect to R_L, as indicated by Scenario IX. Low-tech consumable products, such as light bulbs, may fit into this category due to its inexpensive design and mature manufacturing but a large sale volume, which leads a significant number of warranty claims during the warranty period if the lower level reliability requirement is less demanding.



(a) (b)

Figure 3.19: Increasing patterns of the overall cost

Figures 3.19 to 3.21 further show the three representative cost changing patterns and their separate cost components, respectively. In Figure 3.19, the increasing pattern of the expected RG cost dominates the others and leads to the increasing pattern of the overall

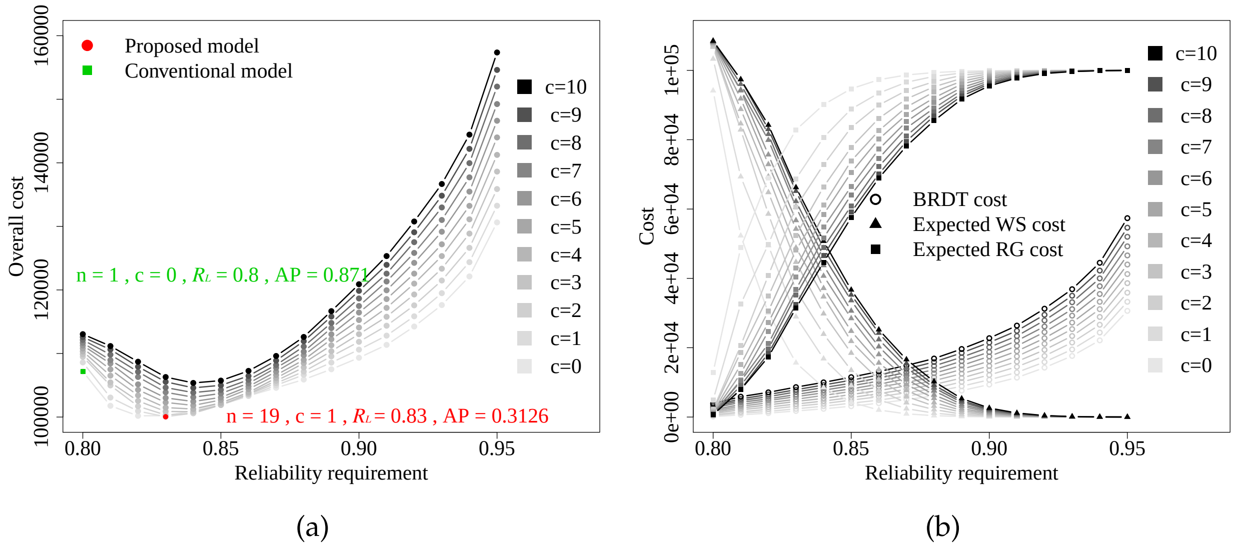


Figure 3.20: U-shape patterns of the overall cost

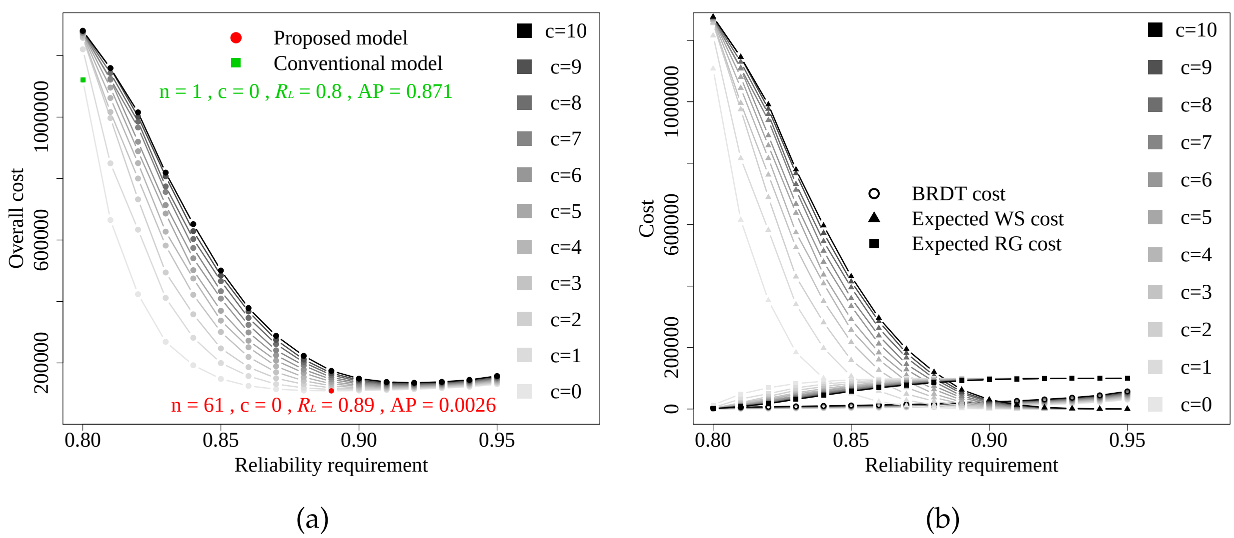


Figure 3.21: Decreasing patterns of the overall cost

cost. Compared to conventional optimal test plan of $(n^* = 1, c = 0, R_L = 0.8)$, the proposed optimal test plan, i.e., $(n_p^* = 5, c = 2, R_L = 0.8)$, can be achieved at the lower bound of R_L but a larger value of maximum allowable failure. A larger c value will reduce (1-AP) to mitigate the dominant effect of cost component G. In Figure 3.20, as G and W become comparable, their trade-off leads to a U-shape pattern of the overall cost. Similar patterns have been discussed in the real-world example. In Figure 3.21a, the overall changing pattern exhibits a decreasing trend as the expected WS cost dominates

the overall cost. Compared to conventional optimal test plan of $(n^* = 1, c = 0, R_L = 0.8)$, the proposed optimal test plan, i.e., $(n_p^* = 61, c = 0, R_L = 0.89)$, can be achieved at the smallest c value but a higher R_L than the minimum acceptable value. A larger R_L value will simultaneously reduce AP to mitigate the dominant effect of cost component W.

Table 3.9: Optimal BRDT designs by varying W / G ratio

W/G	n_p^*	R_L	c	AP	Cost
0.05	14	0.8	5	0.9887	1.614E+04
0.25	5	0.8	2	0.9784	6.035E+04
0.5	5	0.8	2	0.9508	1.199E+05
0.75	19	0.84	0	0.0988	1.513E+05
1	32	0.86	0	0.0270	1.543E+05
1.5	40	0.87	0	0.0132	1.559E+05
5	61	0.89	0	0.0026	1.589E+05
10	75	0.9	0	0.0010	1.604E+05
Settings: $C_f = 1000, C_v = 125, C_w = 5, G = 1.5\text{E}+05, M = 5000, \pi \sim \text{Beta}(9, 60.86)$					

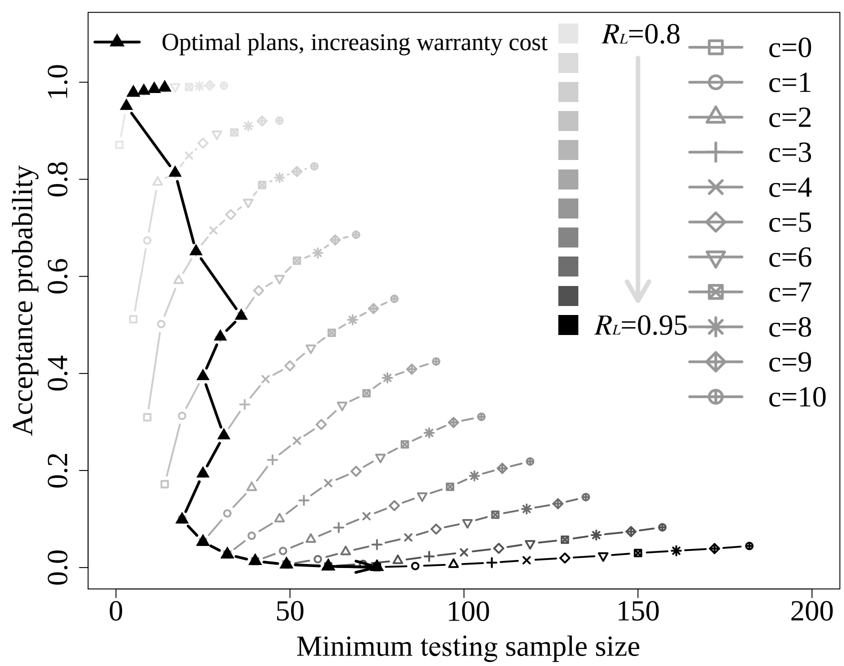


Figure 3.22: Pathway of optimal BRDT designs with increasing W/G ratio

To further understand the second category where the overall cost shows a U-shaped pattern, which occurs when W and G are comparable in magnitude, Table 3.9 shows the optimal BRDT design parameters at varied W/G ratios. When W/G ratio is small

(e.g., 0.05), W value is negligible and G is dominant. The optimal design parameter will select the same lower level reliability requirement of 0.8 specified in conventional optimal design. The reduced overall cost is realized through selecting a higher value of c to reduce (1-AP) as well as the expected RG cost. As W/G ratio increases, the influence of W will become increasingly significant in contributing the overall cost. To shrink the effect of more costly W, the optimal design tends to select the R_L towards a higher level. As the same time, c value is selected at a lower level to reduce AP as well as the expected WS cost. Figure 3.22 further visualizes the changing pathway of optimal BRDT designs as W/G ratio increases. Note that for conventional design, the corresponding optimal design parameters are always $(n^* = 1, c = 0, R_L = 0.8)$, regardless of the changes of W/G ratios.

3.4.3 Impact of Prior Distributions

To investigate the influence of different prior settings, Figure 3.23 shows seven different beta priors of π under various prior mean and variance specifications. For example, with the same prior mean of 0.1 (equivalently, the average prior belief of product reliability of 0.9), $\text{Beta}(1, 9)$, $\text{Beta}(2, 18)$, $\text{Beta}(4, 36)$ are associated with high, medium and low prior variances, respectively. Similarly, as the prior mean of π increases to 0.2 (with the average prior belief of product reliability reduced to 0.8), $\text{Beta}(0.5, 2)$, $\text{Beta}(7.5, 30)$, $\text{Beta}(15, 60)$ are associated with decreasing prior variances. The flat density curve refers to the non-informative prior $\text{Beta}(1, 1)$, which assumes the that π has an equal probability of taking any value from $[0, 1]$.

Table 4.12 shows the optimal BRDT design parameters (n_p^*, c, R_L) as well as the corresponding AP for different prior distributions shown in Figure 3.23, where the maximum tolerance of CR is controlled at $\beta = 0.05$. Several findings can be concluded. First, for the fixed R_L, c and similar prior variance, a prior distribution whose prior mean is more supportive of reliability to be demonstrated can lead to a smaller n_p^* and a higher AP.

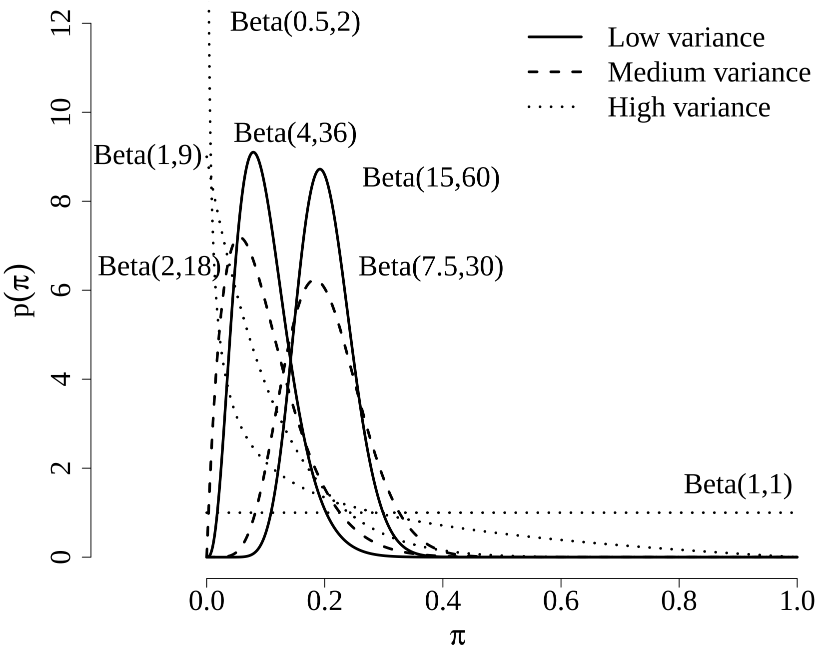


Figure 3.23: Density curves of different prior distributions

For instance, with similar prior variances, Beta$(2, 18)$ has a lower prior mean of π than Beta$(7.5, 30)$, indicating an average prior belief of a higher product reliability. It further implies that prior knowledge elicited by Beta$(2, 18)$ is always more supportive of demonstrating the same lower level reliability requirement than Beta$(7.5, 30)$. Thus, given any fixed R_L and c, Beta$(2, 18)$ leads to a smaller testing sample size and a higher AP, compared to Beta$(7.5, 30)$.

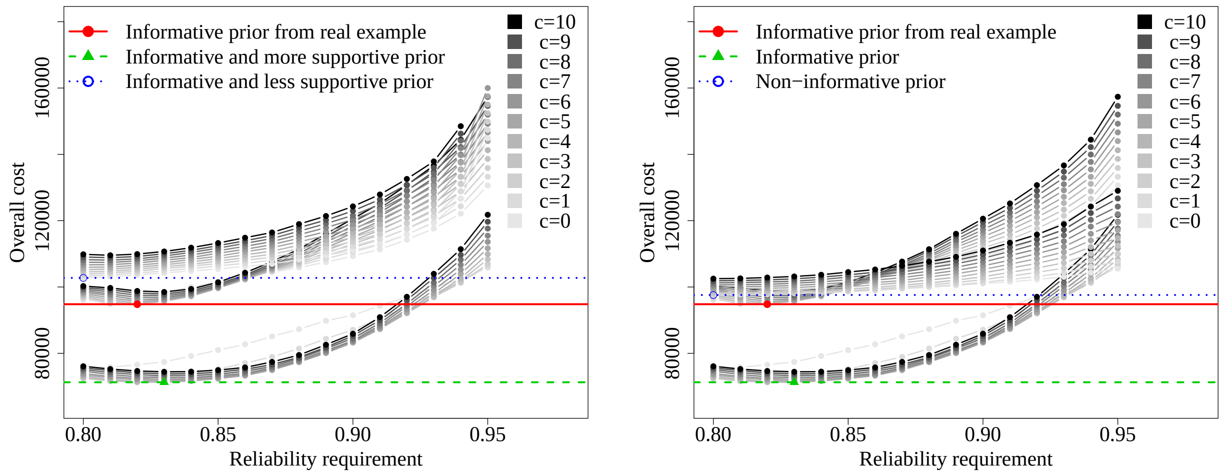
Second, with a similar prior mean that is more supportive of reliability to be demonstrated, a prior distribution with a smaller prior variance (i.e. better precision) can lead to a smaller n_p^* and a higher AP. For instance, when $R_L = 0.8$, Beta$(4, 36)$, Beta$(2, 18)$ and Beta$(1, 9)$ have the same average prior belief that product reliability is around 0.9, which supports for demonstrating the lower level reliability requirement at 0.8. For any fixed c, as prior variance decreases (corresponding to a more informative prior), a smaller testing sample size and a higher AP can be achieved. It is because a smaller prior variance indicates a stronger prior belief in supporting the reliability to be demonstrated. As a result, fewer units need to be tested and the test is also more likely to be successful.

Third, when the prior mean is less supportive of reliability to be demonstrated, a prior distribution with higher prior variance can lead to a smaller n_p^* and a higher AP. For instance, when $R_L = 0.85$, $\text{Beta}(15, 60)$, $\text{Beta}(7.5, 30)$ and $\text{Beta}(0.5, 2)$ have the same average prior belief that product reliability is 0.8, which is less supportive to demonstrating the lower level reliability requirement at 0.85. For any fixed c, as prior variance increases, a smaller testing sample size and a higher AP can be achieved. It is because a less supportive prior belief can increase the number of product units to be tested with a decreased AP by utilizing additional test samples to suppress the impact of the prior knowledge. When prior variance increases, there is less strong belief on the lower reliability, which essentially mitigate the effects of having a prior belief in lower reliability. For non-informative priors (e.g., $\text{Beta}(1, 1)$), they can be viewed as prior distributions in favor of lower reliability than what to be demonstrated (with prior mean at 0.5) but much higher prior variance (e.g., 0.0833), compared to informative priors above. When the informative priors are supportive of reliability to be demonstrated, the non-informative prior will lead to a larger n_p^* and a lower AP. When the informative prior supports rejecting the reliability to be demonstrated, the associated n_p^* and AP of the non-informative prior may be greater or smaller, depending on the prior variance of the informative prior. As its prior variance decreases, the prior belief of rejecting the reliability to be demonstrated is at a higher level of assurance, which tends to increase the sample size of BRDT (in order to mitigate the effect of such strong prior belief) as well as to decrease the probability of passing the test (i.e., AP). As n_p^* increases, it will first have a smaller sample size as compared to that of the non-informative prior, e.g., $\text{Beta}(1, 9)$ v.s. $\text{Beta}(1, 1)$, but later increase to a level which has a larger sample size than that of the non-informative prior, e.g., $\text{Beta}(4, 36)$ v.s. $\text{Beta}(1, 1)$. Similarly, as AP decreases, it will first have a larger value than that of the non-informative prior, e.g., $\text{Beta}(1, 9)$ v.s. $\text{Beta}(1, 1)$, but later decrease to a level which has a smaller AP than that of the non-informative prior, e.g., $\text{Beta}(4, 36)$ v.s. $\text{Beta}(1, 1)$.

Table 3.10: Minimum testing sample sizes under different prior settings

$\pi \sim$ Beta	$(1,9)$			$(2,18)$			$(4,36)$			$(0.5,2)$			$(7.5,30)$			$(15,60)$			$(1,1)$		
Mean(π)	0.1			0.1			0.1			0.2			0.2			0.2			0.5		
Var(π)	0.0082			0.0043			0.0022			0.045			0.0042			0.0021			0.0833		
R_L	c	n_p^*	AP	c	n_p^*	AP	c	n_p^*	AP	c	n_p^*	AP	c	n_p^*	AP	c	n_p^*	AP	c	n_p^*	AP
0.7	0	1	0.8995	0	1	0.8995	0	1	0.8997	0	4	0.5498	0	3	0.5214	0	1	0.7995	0	8	0.1138
	1	2	0.9818	1	2	0.9854	1	2	0.9877	1	7	0.6404	1	6	0.6550	1	2	0.9576	1	12	0.1578
	2	3	0.9954	2	3	0.9973	2	3	0.9982	2	10	0.6770	2	8	0.7811	2	3	0.9906	2	15	0.1925
	3	4	0.9986	3	4	0.9994	3	4	0.9997	3	13	0.6966	3	11	0.8135	3	4	0.9978	3	19	0.2056
	4	5	0.9995	4	5	0.9998	4	5	0.9999	4	16	0.7089	4	14	0.8365	4	5	0.9995	4	22	0.2233
	5	6	0.9998	5	6	0.9999	5	6	0.9999	5	19	0.7172	5	16	0.8803	5	6	0.9999	5	25	0.2370
0.8	0	5	0.6405	0	3	0.7394	0	1	0.8997	0	7	0.4454	0	23	0.0193	0	31	0.0033	0	13	0.0732
	1	9	0.7628	1	7	0.8316	1	2	0.9877	1	12	0.5190	1	28	0.0477	1	36	0.0105	1	18	0.1080
	2	13	0.8131	2	10	0.8933	2	3	0.9982	2	17	0.5489	2	33	0.0785	2	42	0.0186	2	24	0.1234
	3	17	0.8402	3	14	0.9093	3	4	0.9997	3	21	0.5761	3	39	0.1004	3	47	0.0301	3	29	0.1375
	4	21	0.8571	4	18	0.9196	4	5	0.9999	4	26	0.5843	4	44	0.1274	4	52	0.0430	4	34	0.1477
	5	25	0.8686	5	22	0.9267	5	6	0.9999	5	30	0.5980	5	49	0.1524	5	57	0.0566	5	39	0.1553
0.85	0	10	0.4702	0	11	0.3934	0	11	0.3566	0	10	0.3846	0	44	0.0018	0	67	3.115E-05	0	18	0.0538
	1	17	0.5771	1	17	0.5426	1	17	0.5160	1	17	0.4479	1	52	0.0050	1	74	1.358E-04	1	25	0.0784
	2	23	0.6383	2	24	0.6018	2	23	0.6059	2	23	0.4827	2	59	0.0096	2	81	3.487E-04	2	32	0.0928
	3	29	0.6736	3	30	0.6515	3	29	0.6633	3	29	0.5024	3	66	0.0150	3	88	6.906E-04	3	39	0.1024
	4	35	0.6965	4	36	0.6849	4	35	0.7028	4	34	0.5216	4	73	0.0209	4	95	1.168E-03	4	46	0.1092
	5	41	0.7126	5	42	0.7088	5	41	0.7317	5	40	0.5296	5	80	0.0269	5	103	1.656E-03	5	52	0.1165
0.9	0	20	0.3072	0	26	0.1739	0	36	0.0675	0	17	0.3048	0	88	5.401E-05	0	149	1.661E-08	0	28	0.0353
	1	31	0.3993	1	37	0.2643	1	46	0.1329	1	27	0.3644	1	99	1.965E-04	1	160	9.781E-08	1	40	0.0493
	2	40	0.4582	2	47	0.3259	2	57	0.1807	2	36	0.3950	2	110	4.346E-04	2	170	3.493E-07	2	50	0.0592
	3	50	0.4878	3	57	0.3677	3	67	0.2238	3	46	0.4084	3	120	7.950E-04	3	181	8.644E-07	3	61	0.0648
	4	60	0.5075	4	67	0.3979	4	77	0.2587	4	55	0.4204	4	131	1.202E-03	4	191	1.891E-06	4	71	0.0698
	5	69	0.5264	5	77	0.4206	5	87	0.2874	5	64	0.4288	5	142	1.662E-03	5	201	3.617E-06	5	82	0.0726
0.95	0	50	0.1512	0	75	0.0410	0	114	0.0034	0	36	0.2154	0	359	1.134E-10	0	>500	–	0	55	0.0188
	1	71	0.2118	1	98	0.0709	1	136	0.0086	1	55	0.2613	1	379	8.454E-10	1	>500	–	1	77	0.0264
	2	92	0.2443	2	119	0.0946	2	157	0.0145	2	74	0.2815	2	399	3.475E-09	2	>500	–	2	98	0.0308
	3	112	0.2663	3	141	0.1114	3	179	0.0201	3	93	0.2931	3	419	1.043E-08	3	>500	–	3	118	0.0338
	4	132	0.2815	4	162	0.1253	4	200	0.0257	4	111	0.3019	4	440	2.469E-08	4	>500	–	4	138	0.0360
	5	152	0.2925	5	183	0.1364	5	221	0.0309	5	129	0.3081	5	460	5.249E-08	5	>500	–	5	158	0.0376
Settings: $M = 5000, \beta = 0.05$																					

Figure 3.24 further compares the overall cost patterns among different prior settings, including the informative prior elicited from real data (e.g., Beta$(9, 60.86)$), the informative priors with different supportive prior beliefs (e.g., Beta$(2, 18)$, Beta$(7.5, 30)$) with different prior means in Figure 3.24a, and the non-informative prior with flat prior belief of product reliability in Figure 3.24b, respectively. With the informative prior (e.g.,



(a) Informative priors with different levels of supportive prior belief

(b) Informative prior v.s. non-informative prior, compared with real example

Figure 3.24: Cost changing patterns under different prior settings

Beta$(2, 18)$) that is more supportive to the lower level reliability requirement to be demonstrated, the overall cost in optimal BRDT design (e.g., solid red) can be effectively reduced (e.g., dash green).

3.5 Concluding Remarks

The conventional BRDT designs which focuses on minimizing the BRDT testing cost alone, can be limited without considering the uncertainty of the BRDT decision and its cost impacts on subsequent reliability assurance activities. This paper proposes a new optimal BRDT design framework with a longer planning horizon to explicitly quantify the acceptance uncertainty of the test and selects the optimal test plan that minimizes the overall anticipated cost from the BRDT to the WS phase. Specifically, the acceptance probability is calculated to quantify the likelihood of initiating subsequent activities of either RG or WS. The nonlinear relationships between the design parameters and different cost components, are investigated through a case study.

In the case study, a real-world example is used to illustrate the proposed optimal test plan and demonstrate its superior performance over conventional BRDT designs. Fur-

ther, different cost scenarios were explored to extract some general patterns on how the overall cost changes with the lower level reliability requirement, the minimum sample size, and the maximum allowable failures. Finally, the influences of the specified prior distribution on the choice of the optimal BRDT design and the overall cost have been also investigated and summarized. In general, a test plan based on incorporating prior knowledge in favor of the lower level reliability requirement is more economical by requiring a smaller number of testing units and increasing the probability for having a successful test.

To conclude, the proposed optimal BRDT design framework is the first to incorporate the acceptance decision uncertainty in the test planning stage. It advocates a more global and holistic view for reliability engineering practitioners in designing RDTs. Although the scope of this paper is within the BRDT, similar concepts can be well extended to other types of RDTs, such as RDTs based on failure time or degradation data, which will be investigated as future works. In addition, more detailed modeling of WS and RG costs can be realized given detailed cost information of RG and WS in a specific application context. It will be also interesting to further expand the planning horizon of current BRDT design by considering multi-stage uncertainty in the subsequent activities, including both WS, RG with multiple redesign cycles and future BRDT test(s) after each RG redesign cycle.

Chapter 4

Multi-objective Optimal Design for Binomial Reliability Demonstration Tests with Multiple Time Periods

4.1 Introduction

Reliability demonstration tests (RDTs) have been widely used for reliability assurance during the design, development and validation stages of product life cycle. The purpose of RDTs is to assess whether the product reliability can satisfy the minimum requirement before being delivered to customers or put into services. If the test is passed with the required reliability demonstrated, manufacturers can move forward to production and the product can be released to the market. Otherwise, failing the test will lead back to the product development stage for further improvement through reliability growth. RDTs have been broadly employed in various industries with products ranging from high-tech systems/components with demanding reliability requirements, such as aircraft, medical device, semiconductor and microelectronic, to high-volume consumer products, such as lamps or toys. With the global market competition becoming fierce, more and more manufactures would like to have sophisticated RDT design strategies to improve the competitiveness of products and safeguard market shares.

To implement the RDTs in practice, the practitioners need to carefully consider which type of reliability data is associated with the product, which model will be suitable for the RDT design and whether the assumptions are satisfied (Meeker and Escobar, 2014, pp.3-15). There are three major categories including 1) RDT designs based on failure count data (Guo and Liao, 2011; Jensen, 2015; Chen et al., 2017; Li et al., 2016), which

usually model the number of observed failures with discrete distributions (e.g., Binomial distribution, Poisson distribution); 2) RDT designs based on failure time data (Hamada et al., 2008; Lee et al., 2015; Xu et al., 2017), which consider the time to failure or censoring time for product failure or success respectively and use lifetime distributions such as Weibull or Lognormal distribution to model the failure time data; and 3) RDT designs based on degradation data (Yang, 2009, 2013; Jin and Matthews, 2014), which measure the deteriorating performance of products over time with specification of failure thresholds. Depending on the characteristics of different products, RDTs can be performed at both system and component levels, with the binary test outcomes of pass/fail. Due to the limitations on resource and budgets, there will be many planning decisions need to be analytically determined such as how many units need to be tested, how long is the test duration, whether to test them all at once or sequentially, how to determine the experimental conditions for the testing, how to determine the failure of units, and what will be the decision rules for success/failure of the RDTs. It requires the RDT design to be efficient enough to achieve the goal of demonstrating the required reliability performance with a desired level of confidence.

In this paper, the scope of the proposed work is within the optimal RDT design for failure count data. Failure count data is usually collected from testing a sample of product units and observing the number of failures over a certain testing period. The test will be passed with lower level reliability requirement (pre-specified by the manufacturer) satisfied if the observed number of failures is no more than a given design setting of the maximum allowable failures. With all design settings provided, the consumer's risk (CR), which is the probability of passing the test when the product cannot meet the lower level reliability requirement, will be controlled under an acceptable threshold to ensure that the product will meet the expectation of customers. Ultimately, the minimum test sample size can be obtained for the test design so that the RDT cost can be minimized. The optimal settings in conventional optimal RDT design are determined by the binomial equation

(Kececioglu, 2002) and thus such tests are called binomial RDTs (BRDTs). Because of the convenience in practical implementation, BRDTs have been widely applied in reliability assurance activities. It requires no periodic inspection or continuous monitoring of the product performance and only needs to report the failures at the end of testing periods (Guo and Liao, 2011). BRDTs can also be particularly useful for one-shot systems (Guo et al., 2010b) with destructive testing nature and high-reliability requirement (e.g., missiles, rockets, etc.).

Conventional BRDT designs have several limitations. First, they often focus on demonstrating reliability requirement of products over a single time period (e.g., mission time), which may not be able to meet the increasing demand of customers with more detailed reliability requirements. For example, for a product with five-year expected lifetime, customers may be more adverse to early failures in first two years after purchase with higher reliability requirement, and may be less critical to the same number of failures in the last three years. The conventional BRDT design cannot deal with varied reliability requirements specified during different time periods (e.g., first two years and last three years). Chen et al. (2017) proposed the RDT strategies by extending conventional BRDT design to multi-state RDT (MSRDT) design to demonstrate the product reliability over multiple time periods, in order to meet the increasing demand of customers. To illustrate, the MSRDT design with multiple time periods considers the distribution of number of failures in each continuous and non-overlapping testing period (e.g., first two years, late three years) and demonstrates the lower level reliability requirements for cumulative testing periods (e.g., first two years and five years) simultaneously. For each separate/cumulative period, the maximum allowable failures will be given and the lower level reliability requirements will be pre-specified. Hence, the minimum test sample size can be obtained with controlled CR under certain acceptable level.

The second limitation of BRDT designs as well as many RDT designs in general is that they mainly focus on optimizing a single objective, namely minimizing the test sample

size at the minimum testing cost. In fact, there are many other performance evaluation criteria of RDT, if not addressed appropriately, may affect manufacturers and/or customers. For example, producer's risk (PR), the probability of failing to accept a test when the product can actually meet the lower level reliability requirement, is usually overlooked. When PR is too high, manufacturers may have high expenses on rejecting the reliable product units, which may be too costly in practice. The acceptance probability (AP) is the probability of passing the actual test, which quantifies the RDT acceptance decision uncertainty. When the minimum test sample size corresponds to a extremely low AP, the test becomes more difficult to be passed, which may require more resources and expenses on the test and become less beneficial to manufacturers. In both cases above with either high PR or low AP, a RDT design solely based on minimizing the testing sample size will not optimize the overall benefits of manufacturers when additional costs may occur. To effectively balance different RDT performance evaluation criteria and leverage the diverse needs of different beneficiaries (e.g., manufacturers, customers) from the test, there is a need to simultaneously incorporate multiple objectives in the RDT design. Lu et al. (2016) is among the first to put forward the idea of multiple objectives optimization using Pareto Front to balance the trade-offs of multiple and potentially conflicting objectives in BRDT design and select the optimal plans with Pareto efficiency (i.e., not all of the objectives in each selected plan dominated by those of other plans). When CR is controlled under an acceptance level, we may have an unacceptably higher PR, affected by the trade-off between CR and PR. The trade-offs from four objectives, including CR, PR, AP and test sample size, can be balanced and the Pareto-optimal test plans will be selected given the specifications on lower level reliability requirements and different design settings such as maximum allowable failures. Since the work of (Lu et al., 2016) only considered reliability demonstration for a single time period, the customers with increasing demands on reliability for multiple failure states, as considered by the work of MSRDT design in (Chen et al., 2017).

Another limitation of most of the aforementioned BRDT designs as well as RDT design in general is that they only focus on the RDT itself without considering its impact on subsequent reliability activities. However, due to the uncertainty of RDT test, whether passing or failing the test may result in whether the product will be released to market and whether or not the subsequent reliability assurance activities of warranty services will be initiated. For example, when lower values of product reliability need to be demonstrated, the test will be easier to pass with higher AP, which also requires less test samples. However, with a lower product reliability requirement, it is expected to observe more warranty claims (due to failures) in the field, which will increase the expected warranty services cost. There will be a potential trade-off between the RDT test cost and the warranty services cost. Therefore, there is a need to consider both RDT testing cost as well as warranty services cost as two separate objectives (in reflecting immediate as well as long-term economic impact of RDT) during the product life cycle. Additionally, it will be desirable to include the lower level reliability requirement as an additional design setting to better balance the trade-off among different design objectives. Kleyner and Sandborn (2008) is among the first to mention that there is a conflicting relationship between RDT cost and warranty services cost, and such relationship can be manipulated through the selection of lower level reliability requirement. Instead of assuming a fixed value of lower level reliability requirement, we will optimally determine it from an acceptable range together with determining other design parameters, such as test sample size and maximum allowable failures.

To fill the gaps in existing literature, a multi-objective optimal design of RDT with multiple time periods is proposed in this paper under the Bayesian framework. The Bayesian framework is considered due to its effectiveness in incorporating prior knowledge, which may improve the estimation performance and reduce the overall cost. Multiple objectives, such as CR, PR, AP (or rejection probability(RP)) and different cost components (e.g., RDT testing cost as well as its expected cost impacts on subsequent warranty services),

are simultaneously considered in the proposed design framework. The assumption of a pre-specified lower level reliability requirement is further relaxed, which provides richer design space for determining the optimal RDT plans. The pairwise relationship among these objectives will be comprehensively explored and the explicit trade-off patterns will be identified. The influence of different design parameters, such as lower level reliability requirements, maximum allowable failures and test sample size, on those trade-off patterns will be further investigated. In order to balance the trade-offs of multiple conflicting objectives, Pareto Front will be obtained from multi-objective optimization to identify a set of test plans that are non-dominated to each other. Based on the Pareto Front, a screening process based on users preference will be provided to facilitate practitioners in selecting the most appropriate optimal test plans. Moreover, the influence of different prior settings on the proposed optimal RDT plans and their trade-off patterns are also comprehensively investigated with various insights obtained.

The remaining of the paper is organized as follows. In the next section, the existing RDT designs, including conventional BRDT with a single objective and a single testing period and RDT with a single objective and multiple testing periods,are introduced. Then, the optimal RDT design with multiple objectives for multiple testing periods is proposed under Bayesian framework, with Pareto Front method introduced for multi-objective optimization. A comprehensive case study is followed to illustrate the proposed framework and demonstrate its effectiveness through a real-world example. The impact of prior knowledge trade-off balance and optimal test plans selection is also explored. Conclusion and future work directions are provided in the end.

4.2 Single Objective RDT Designs

In existing RDT designs, single objective, namely the RDT testing cost, is mainly considered to select the optimal test plan that can minimize the test sample size in a single testing period. With the pre-specified design parameters (e.g., lower level reliability

requirement) or controlled risk criteria (e.g., CR), the single objective is to reduce the test sample size so that the RDT cost can be minimized. In this section, both BRDT and MSRDT designs with single objective optimization will be reviewed.

4.2.1 Binomial RDT Design

In the conventional BRDT design (Kececioglu, 2002), the single objective is to minimize the number of test units n while simultaneously determine the maximum allowable failures c so that the overall BRDT cost can be optimized over the single testing period. Given the pre-specified lower level reliability requirement R_L, both Frequentist and Bayesian approach (Hamada et al., 2008) can be used to determine the design parameters (n, c). In Frequentist approach, (n, c) are determined by controlling the risk criterion $\mathrm{CR_F}$ calculated as follows

$$\mathrm{CR_F}(n, c) = \sum_{y=0}^{c} \binom{n}{y} (1 - R_L)^y (R_L)^{n-y}, \tag{4.14}$$

where the number of observed failures y during the testing period follows a binomial distribution, i.e., $y \sim \mathrm{Bin}(n, \pi)$ with product failure probability π. The criterion is also constrained by $\mathrm{CR_F}(n, c) \leq \beta$ for all $1 - \pi \geq R_L$, where β is the maximum tolerance level. Theoretically, infinite sets of (n, c) can be selected to satisfy the constraint of the risk criterion CR, which will result in an infinite number of feasible BRDT designs. Given fixed c, $\mathrm{CR_F}(n, c)$ will decrease as n increase and n^* is the minimum test sample size that satisfies the constraint to control $\mathrm{CR_F}(n, c)$ below β. If c increases, n^* will also increase. Therefore, to achieve the single objective with minimum BRDT cost, the optimal design parameters usually become $(n^*, c = 0)$ with zero failure allowed in the test, which is also commonly known as the zero-failure test or success run test.

One drawback of frequentist approach is that it limits the opportunity for further cost reduction since the optimal design parameters are directly chosen once R_L and β are pre-specified. With the advantage of prior knowledge incorporation, Bayesian method can

be a better alternative to potentially further reduce the test sample size when appropriate prior information of product reliability (i.e., $1-\pi$) can be elicited (Hamada et al., 2008). The Bayesian approach determines (n,c) by controlling the posterior CR calculated as follows

$$\begin{aligned} \mathrm{CR_B}(n,c) &= \Pr(1-\pi < R_L | y \leq c) \\ &= \frac{\int_{1-R}^{1} [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}{\int_{0}^{1} [\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y}] p(\pi) d\pi}, \end{aligned} \tag{4.15}$$

where $\mathrm{CR_B}(\cdot)$ represents the posterior probability of product reliability not satisfying the lower level reliability requirement (i.e., $1-\pi < R_L$) given the test can be accepted (i.e., $y \leq c$). Additionally, based on expert domain knowledge of expertise and/or historical testing data, the prior distribution density of π, i.e., $p(\pi)$, can be specified. With the posterior CR controlled for any pre-specified R_L and β, the single objective can be achieved with minimum test sample size n^*. The benefits of prior knowledge incorporation have been demonstrated in existing work (Chen et al., 2017; Lu and Rudy, 2001) to further reduce test sample size if the prior elicitation is appropriate. In this paper, Bayesian method will be employed to explore and demonstrate the effectiveness of prior elicitation on cost reduction and decision making.

4.2.2 Multi-state RDT Design

The multi-state RDT design strategies (Chen et al., 2017) is a direct extension of conventional BRDT design by explicitly demonstrating different product reliability requirements over multiple testing periods rather than a single testing period. In this way, customer's expectations on product reliability in multiple time periods can be met simulta-

neously. For example, both of the two companies are conducting the RDT with the same test sample size 100 for a 5-year equivalent demonstration period for their products and require 5 maximum allowable failures. Company I may observe 1 failure in the first two years and 4 failures in the last three years. Company II may observe 4 failures in the first two years and 1 failure in the last three years. Even though both of them meet the requirements to pass the test, customers may prefer to buy from company I rather than company II because of less early failures after purchase and better usage experience. Figure 4.25 illustrates the MSRDT for multiple time periods. Within the finite testing period $(t_0, t_K]$, K non-overlapping time periods, $(t_{i-1}, t_i], i = 1, ..., K$ are exclusively partitioned. The failure probability $\pi_i, i = 1, ..., K$ is defined to be independent from each other for each separate testing period, which follows the multinomial distribution to describe the overall failure probability distribution of the test. The maximum allowable failures c_i and the number of observed failures y_i in each period are defined, respectively. Multiple lower level reliability requirements $R_{Li}, i = 1, \cdots, K$, where R_{Li} is the minimum acceptable reliability over the first i cumulative time periods, i.e., $(t_0, t_i]$, are pre-specified and can be demonstrated simultaneously. The risk criterion $\mathrm{CR}_{\mathrm{MS}}$ will be controlled at β), which is the maximum acceptable level.

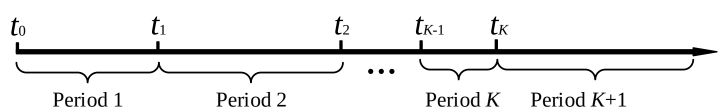


Figure 4.25: Illustration of multiple time periods in K periods between $(t_0, t_K]$

There are two different scenarios of acceptance criteria and the major difference is that scenario I considers cumulative time periods while scenario II considers separate periods. Under scenario I, we will pass the test if the cumulative number of observed failures $\sum_{k=1}^{i} y_k$ corresponding to each cumulative time period $(t_0, t_i]$ is no more than its cumulative maximum allowable failures $\sum_{k=1}^{i} c_k$ for all cumulative time periods $(t_0, t_i]$, at $i = 1, ..., K$. Under scenario II,we may pass the test if the number of observed failures y_i at each non-overlapping time period $(t_{i-1}, t_i]$ is less or equal to its maximum allowable

failures c_i for all separate time periods $(t_{i-1}, t_i]$, at $i = 1, ..., K$. For example, in the two-period MSRDT, at the end of first year and fifth year, the maximum allowable failures are $c_1 = 1$ and $c_2 = 4$, respectively. With 100 test units, the MSRDT will be passed if $y_1 \leq 1, y_1 + y_2 \leq 5$ in scenario I and $y_1 \leq 1, y_2 \leq 4$ in scenario II.

The MSRDT design extends to multiple failure states in the test. However, it still focuses on a single objective to minimize the MSRDT cost by reducing the test sample size. For example, in a MSRDT over two time periods (i.e., $K = 2$), given the pre-specified lower level reliability requirements R_{L1} and R_{L2} over the time periods $(t_0, t_1]$ and $(t_0, t_2]$, respectively, the corresponding $\mathrm{CR}_{\mathrm{MS}}$ is controlled at or below β by

$$\mathrm{CR}_{\mathrm{MS}}(n, c_1, c_2) = 1 - \frac{\int_0^{1-R_{L1}} \int_0^{1-R_{L2}-\pi_1} H_{\mathrm{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_{\mathrm{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1} \leq \beta, \tag{4.16}$$

where $p(\pi_1, \pi_2)$ denotes the joint prior distribution of failure probabilities $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2)$ and $H_{\mathrm{MS}}(n, c_1, c_2)$ is the probability of accepting the test plan for any given (π_1, π_2), which varies from different scenarios. The optimal plan can be achieved with (n^*, c_1, c_2) where n^* can minimize the test sample size for the overall period of $(t_0, t_2]$. Similar to conventional BRDT design, zero-failure tests with $(n^*, c_1 = 0, c_2 = 0)$ usually become the optimal choices for ultimate MSRDT cost minimization.

4.3 Multiple Objectives RDT Design

Singe objective RDT designs only focus on minimizing the test sample size with controlled CR and pre-specified lower level reliability requirements. However, such test designs often ignore the higher risk of manufacturers themselves (e.g., unacceptable high PR), the low probability of passing the test (e.g., low AP) or costs from subsequent reliability activities (e.g., warranty services), which may affect the decision making in selecting the optimal test plans that minimize the overall benefits for manufacturers and customers. Specifically, there are multiple risk criteria and cost components involved in

the testing periods and product life cycle. Customers are facing the risk of receiving inferior products that pass the test (i.e., CR), which may affect their satisfaction and future purchase of the product. Manufacturers need to control the risk of rejecting the reliable products in the test (i.e, PR), which may increase the test cost. The probability of passing the RDT actually determines the subsequent outcomes of different reliability assurance activities, which will be directly associated with the corresponding cost components in each activities. For example, if the RDT can be passed with lower product reliability, the warranty services cost may increase because of the claims from product failures or unsatisfied performance. Meanwhile, there may be different trade-off patterns among these objectives that will have impacts on the selection of optimal plans. For example, when CR is controlled at an acceptable level to meet the expectation of customers, PR may be too high to be acceptable, which may result in additional costs to the manufacturers. When the test becomes more difficult to pass, the potential expense of reliability growth cost will be high. Obviously, there is a research need to develop a comprehensive optimal RDT design framework to simultaneously balance multiple objectives and optimize the overall costs from all related activities.

(Lu et al., 2016) was among the first to discuss about the BRDT design with multiple objectives (MO-BRDT), including the test sample size, CR (i.e., the probability of the product passing the test when it shouldn't be passed), PR (i.e., the probability of the product failing the test when it should be passed) and AP (i.e., the probability of passing the test) can be balanced simultaneously to select the optimal test plan. The four objectives are balanced through a two-stage decision making process using the Pareto Front approach. First, a collection of non-dominating test plans are selected as Pareto optimal solutions by considering all objectives simultaneously. Second, based on the particular needs of users, a reduced number of choices can be identified to best match their specific goals. However, there are several limitations which may affect the practicability of the MO-BRDT design. First, the design still focuses on the single testing period or single failure state, which

may not be able to satisfy the increasing demand of customers (Chen et al., 2017). Second, the objectives considered in MO-BRDT may not be adequate to optimize the overall costs from all related reliability assurance activities such as warranty services in a wider planning horizon. There is still a research gap to be filled by developing an optimal RDT design framework which can resolve the complexities from both multiple states of reliability data and requirements as well as a wider planning horizon for all related reliability assurance activities.

4.3.1 Multiple Objectives Multi-State RDT Design

In order to consider the complexities of reliability data from multiple failure states and the expanded planning horizon covering all related reliability assurance activities, the proposed work of multi-objective MSRDT (MO-MSRDT) can be regarded as the direct extensions to conventional BRDT designs (Kececioglu, 2002; Guo and Liao, 2011) and MSRDT design (Chen et al., 2017) with single objectives as well as BRDT design with multiple objectives (Lu et al., 2016). Specifically, the propose work simultaneously considers multiple states and explores the trade-offs among multiple objectives to develop a comprehensive decision making framework for different user preferences and requirements. Multiple testing periods are considered with flexible inputs of design parameters (e.g., R_{Li}, c_i, $i = 1, \ldots, K$ etc.) to meet the needs of customers with various complicated lower level reliability requirements. The trade-offs among multiple objectives including n, CR, PR, RP (i.e., 1-AP, the probability of rejecting the test) and warranty service cost (W) are explicitly evaluated.

To illustrate the proposed work, the two-period MO-MSRDT with five objectives are considered without the loss of generality. Including CR_{MS} which has been mentioned before, all related objectives including PR_{MS}, AP_{MS}, RP_{MS} can be formulated as

$$\mathrm{CR_{MS}}(n, c_1, c_2, R_{L1}, R_{L2}) = 1 - \frac{\int_0^{1-R_{L1}} \int_0^{1-R_{L2}-\pi_1} H_{\mathrm{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_{\mathrm{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}, \quad (4.17)$$

$$\mathrm{PR_{MS}}(n, c_1, c_2, R_{L1}, R_{L2}) = \frac{\int_0^{1-R_{L1}} \int_0^{1-R_{L2}-\pi_1} (1 - H_{\mathrm{MS}}(n, c_1, c_2)) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 (1 - H_{\mathrm{MS}}(n, c_1, c_2)) p(\pi_1, \pi_2) d\pi_2 d\pi_1}; \quad (4.18)$$

$$\mathrm{AP_{MS}}(n, c_1, c_2) = \int_0^1 \int_0^1 (H_{\mathrm{MS}}(n, c_1, c_2)) p(\pi_1, \pi_2) d\pi_2 d\pi_1; \quad (4.19)$$

$$\mathrm{RP_{MS}}(n, c_1, c_2) = \int_0^1 \int_0^1 (1 - H_{\mathrm{MS}}(n, c_1, c_2)) p(\pi_1, \pi_2) d\pi_2 d\pi_1; \quad (4.20)$$

The objective $\mathrm{W_{MS}}$ for evaluation on warranty service cost is defined as the expected cost ratio between average warranty claim cost and product unit cost. One of the scenario in MO-MSRDT will be used for illustration with cumulative requirements on the maximum allowable failures for each cumulative testing period. The corresponding expected cost ratio can be formulated as

$$\begin{aligned} \mathrm{W_{MS}}(n, c_1, c_2) = & \frac{C_{w_1}}{C_p} E(\pi_1 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) \\ & + \frac{C_{w_2}}{C_p} E(\pi_2 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2), \end{aligned} \quad (4.21)$$

where C_{w_1} and C_{w_2} are the average warranty claim costs from for the corresponding separate testing periods $(t_0, t_1]$ and $(t_1, t_2]$, respectively, and C_p is the product unit cost. Since the average warranty claim cost is usually less or equal to the product unit cost, depending on the specific warranty policies (e.g, free replacement, pro-rata, etc.), the ratios of $\frac{C_{w_1}}{C_p}$

and $\frac{C_{w_2}}{C_p}$ will be no greater than 1. To approximately estimate the expected cost ratio for each separate testing period, the means of true product failure probability for each corresponding period conditioning on passing the test, i.e., $E(\pi_1|y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2)$ and $E(\pi_2|y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2)$, can be calculated respectively and shown as follow,

$$E(\pi_1|y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2) = \frac{\int_0^1 \int_0^1 \pi_1 H_{\text{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}{\int_0^1 \int_0^1 H_{\text{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}; \tag{4.22}$$

$$E(\pi_2|y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2) = \frac{\int_0^1 \int_0^1 \pi_2 H_{\text{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}{\int_0^1 \int_0^1 H_{\text{MS}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}; \tag{4.23}$$

The probability of accepting the test plan $H_{\text{MS}}(n, c_1, c_2)$ for the selected scenario can be formulated as

$$H_{\text{MS}}(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=0}^{c_1+c_2-y_1} [(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!}) \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n-y_1-y_2}] \tag{4.24}$$

The derivations of formulas and example of another scenario can be found in Appendix.

Numerical examples will be illustrated in case study section to demonstrate the effectiveness of the proposed work. The objectives are calculated under Bayesian framework using Monte Carlo integration (Robert and Casella, 2013, pp.71-131), where a large number of samples of failure probability $\pi_i, i = 1, ..., K$ of size M (e.g., 5000) are generated from the specified prior distribution. The Dirichlet distribution can be used as the prior distribution for $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2)$ for two-period MO-MSRDT design. Denoted as Dirichlet$(\alpha_1, \alpha_2, \alpha_3)$, $(\alpha_1, \alpha_2, \alpha_3)$ are hyper-parameters for prior knowledge elicitation. As a family of continuous multivariate probability distribution, the Dirichlet distribution can be parameterized with the vector of positive hyper-parameters $\alpha_i, i = 1, \cdots, K$, where K is directly corresponding to the K separate testing periods in MO-MSRDT de-

sign. This advantage of Dirichlet distribution can enable more intuitive applications associated with each testing period. Another advantage is the capability of conjugate prior for multinomial distribution (Meeker and Escobar, 2014) can facilitate the convenience updating based on new information. The posterior distribution of probabilities after updating also follows the Dirichlet distribution. Different settings of hyper-parameters will also be explored to investigate the impact of prior knowledge on the MO-MSRDT design performance. Then, the objectives and failure probabilities can be approximated as

$$\begin{aligned}\mathrm{CR}_{\mathrm{MS}}(n, c_1, c_2, R_{L1}, R_{L2}) \approx & 1 - \frac{1}{\sum_{j=1}^{M} H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}} \cdot \\ & [\sum_{j=1}^{M} [H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}} \cdot \\ & I(\pi_1^{(j)} \leq 1 - R_{L1}, \pi_1^{(j)} + \pi_2^{(j)} \leq 1 - R_{L2})]],\end{aligned} \tag{4.25}$$

$$\begin{aligned}\mathrm{PR}_{\mathrm{MS}}(n, c_1, c_2, R_{L1}, R_{L2}) \approx & [\sum_{j=1}^{M} [(1 - H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}) \cdot \\ & I(\pi_1^{(j)} \leq 1 - R_{L1}, \pi_1^{(j)} + \pi_2^{(j)} \leq 1 - R_{L2})]] \cdot \\ & \frac{1}{\sum_{j=1}^{M} (1 - H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}})},\end{aligned} \tag{4.26}$$

$$\mathrm{RP}_{\mathrm{MS}}(n, c_1, c_2) \approx \sum_{j=1}^{M} (1 - H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}), \tag{4.27}$$

$$E(\pi_1 | y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2) \approx \frac{\sum_{j=1}^{M} \pi_1^{(j)} H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}}{\sum_{j=1}^{M} H_{\mathrm{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}}; \tag{4.28}$$

$$E(\pi_2|y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2) \approx \frac{\sum_{j=1}^{M} \pi_2^{(j)} H_{\text{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}}{\sum_{j=1}^{M} H_{\text{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}}}; \tag{4.29}$$

$$\begin{aligned} H_{\text{MS}}(n, c_1, c_2)|_{\pi_1^{(j)}, \pi_2^{(j)}} = \sum_{y_1=0}^{c_1} \sum_{y_2=0}^{c_1+c_2-y_1} [(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!}) \cdot \\ (\pi_1^{(j)})^{y_1} (\pi_2^{(j)})^{y_2} (1 - \pi_1^{(j)} - \pi_2^{(j)})^{n-y_1-y_2}]. \end{aligned} \tag{4.30}$$

where j is the jth sample of $\pi_i, i = 1, 2$ from the simulated sample of prior distribution with size M.

4.4 Pareto Front Optimization

This section will provide a brief introduction of Pareto Front approach in multiple objective optimization. Considering the optimization problem in optimal RDT design with K objectives and $f_k(x), k = 1, 2, \ldots, K$ as the objective function for the kth objective (e.g., CR, PR, W, etc.), the multi-objective optimization problem can be formulated as

$$\begin{aligned} &\text{minimize} \quad (f_1(x), f_2(x), \ldots, f_K(x)) \\ &\text{subject to} \quad x \in X, \end{aligned} \tag{4.31}$$

where X can be the feasible set of different decision variables (e.g., design parameters c, R, etc.) and each objective function $f_k(x)$ may be subject to the influence from different decision variables $x \in X$. The first step in solving the optimization problem is to determine the feasible solution space based on different x. For example, in order to demonstrate the product reliability, a acceptable range of lower level reliability requirements can be provided (e.g., $R_{L1} = 0.85 \sim 0.95, R_{L2} = 0.8 \sim 0.9$) and the values of corresponding

objectives for each combination of design parameters can be obtained, which process can be expressed by the function,

$$f : X \to \mathbb{R}^K, f(x) = (f_1(x), f_2(x), \ldots, f_K(x)). \tag{4.32}$$

In order to find the optimal test plan, there are two major approaches to tackle the multi-objective optimization problem (Kasprzak and Lewis, 2001). One is trying to find the single “best" plan which can provide the best possible outcome. This type of approaches usually involves aggregation of some objective functions or evaluation of decision variables to create more aggregate objective(s) or constrain(s). In this way, the complexity of solving the multi-objective optimization problem in high dimension can be simplified. For example, Messac et al. (2000) discussed the properties of some aggregate objective function and provided insights on developing suitable aggregate objective functions to capture the best optimal solution. Another example is the desirability approach (Derringer and Suich, 1980), which creates scores to evaluate the “desire" on different decision variables for each objective and optimizes the overall score.

Due to the constraints of budget and resource and specific user preferences, it is difficult to find the ideal solution that can optimize all the criteria and provide the best possible outcome. Because of the trade-offs among different objectives, a balanced decision is needed with some of the criteria being compromised to meet the goals in practice. The other commonly used approach for multi-objective optimization is to determine the set of non-dominated solutions along the Pareto Frontier (Kasprzak and Lewis, 2001) and select the plans corresponding to the values of objectives. Considering a set of feasible solutions to a multi-objective optimization problem $Y = \{y \in \mathbb{R}^K : y = f(x), x \in X\}$, one feasible solution will dominate the other if at least one objective is strictly better and all others are as good as the objectives from the other solution. If one solution is non-dominated, it means improving any of the objectives may inevitably deteriorate the other objectives.

The objectives from non-dominated solutions form a Pareto Front and the set that includes all non-dominated solutions $P(Y) = \{y' \in Y : \{y'' \in Y : y' \succ y'', y' \neq y''\} = \emptyset\}$, which provides the solution space of superior test plans from which the practitioners can make the final selection, are said to be Pareto optimal. Efficient search algorithms have been studied to Pareto Front such as the evolutionary algorithms (Deb, 2001), the structured two-stage decision making (Lu et al., 2011), etc. In this paper, the employed algorithm is called Block-nested-loops (BNL) from Skyline operator (Borzsony et al., 2001), which is popular in database query to find the Pareto optimal solutions efficiently, and implemented in R (Roocks, 2016).

4.5 Case Study

This section illustrates the proposed MO-MSRDT design through a case study from a real application, and demonstrates its comprehensiveness and effectiveness in meeting the increasing reliability demand of customers and determining the optimal test plans with balanced risks and costs. The trade-off patterns among multiple objectives will be identified and explored. Meanwhile, the impact of different design settings (e.g., flexible lower level reliability requirement, maximum allowable failures, etc.) on the change of trade-off patterns will be evaluated. The proposed framework will also be compared with existing designs to demonstrate the advantages in determining the optimal test plans. Based on the exploration results, the comprehensive screening strategy for practitioners to select the optimal test plans will be proposed. In addition, the impact of prior information on trade-off patterns as well as the determination of optimal test plans will be evaluated.

4.5.1 Multiple Objective Interrelationships – Real-world Example

In real-world practices, electronic products like printers, laptops, kitchen applications, etc., are designed to have the product lifetime over years and usually have the limited time warranty (e.g., one year limited warranty for Apple products) associated with their

initial purchases. Within the limited warranty period, the manufacturers usually provide a comprehensive coverage to deal with the defects or failures of the product. The most common limited time warranty policy is the free replacement warranty, i.e., the product with failure can be replaced/repaired at no cost during the limited warranty period. For example, if an iphone cannot start properly or the touchscreen is broken within one year after purchase date, we can always take it to the Apple store and the staff will highly likely replace it with a new one or repair the parts with no additional charge.

Some manufacturers may offer additional options to provide supplemental coverage for the rest of product lifetime after the initial period with free replacement. One option is the limited time pro-rata warranty policy. This warranty will provide limited time of coverage for the product, however, the cost of replacement/repair will be based on the usage years after purchase. For example, for a product with 5-year pro-rata warranty, if the product fails at the end of year 3, then the replacement cost will be 60% of the product price and needs to be paid by the customers. Another option is extended warranty policy. This policy can be purchased in additional to the original purchase of product with a premium paid by the customers. It usually provides a comprehensive coverage (e.g., free replacement) for the product for an extended warrant period. For example, in addition to the one-year free replacement warranty for an iphone, we can also purchase the AppleCare service which can add multiple years to the current warranty period.

From manufacturers' perspective, the warranty service cost for the original free replacement period will be the expenses associated with dealing the warranty claims over a limited time period. The expenses usually include customer calling center services, a new replacement product, recycle of defected product, shipping cost, etc. In the second stage of warranty coverage with pro-rata warranty policy or extended warranty policy, the warranty service cost will be the expense of warranty services over the period less the payment based on pro-rata costs or the premium collected for the extended warranty.

To illustrate, considering a laptop sold with a one-year free replacement warranty policy. An additional premium paid for a two-year extended warranty policy is \$30. Assuming that the averaged expense of free replacement of the laptop will be \$100, the warranty costs for initial one year and the extended two years will be \$100 and \$70, respectively. From the failure rate survey in laptop industry (Sands, 2009), the first year failure rate from malfunction is 4.7% while the three-year failure rate is 20.4%. Therefore, the reliability information can be generalized for two cumulative periods as 95.3% for the first year and 79.6% for the first three years. Then, the prior knowledge of the laptop failure probability over multiple periods can be generalized by the Dirichlet distribution, $\pi \sim \text{Dir}(4.7, 15.7, 79.6)$, where each number corresponds to the failure rate in each separate period overtime (e.g., 4.7% in first year, 15.7% in second and third years, 79.6% after three years). With the product reliability decreasing overtime, the warranty costs for manufacturers will also decrease when comparing the initial warranty period and the extended warranty period. This real example will be used for illustration.

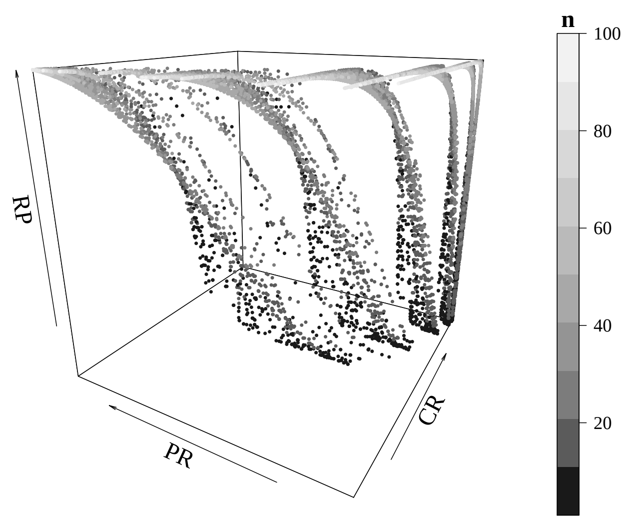


Figure 4.26: General trade-off patterns of objectives

The proposed optimal design of MO-MSRDT considers three risk criteria (e.g., CR, PR, RP) and two cost components (e.g., n for RDT cost, W for warranty service cost) as objectives for decision making. The interrelationships among different objectives need to be explored and compared in order to gain intuitive understanding about their trade-

offs and relationships. For example, Figure 4.26 visualizes the points in the 3D plot that correspond to the test plans with different design parameters (e.g., R_{L1}, R_{L2}, c_1, c_2) and test sample size n, which result in the changing patterns of three risk criteria CR, PR and RP. The general pattern is twisted for different values of three risk criteria and changes with n as well. For each pair of two objectives, it will be interesting to explore whether there exists a trade-off relationship, which can generate insights for balancing the objectives in decision making process. To illustrate, when CR is small, the test plans usually have higher RP and PR, and the test sample size will increase. There may be trade-offs between CR and RP or CR and PR.

The next step is to explore the relationships of each pair of two objectives and identify the trade-off patterns among different objectives. Recall the existing RDT designs in literature (Kececioglu, 2002; Lu et al., 2016; Chen et al., 2017) and relevant statistical formulations of all the objectives, some general facts can be clarified in the optimal RDT design. First, if we want to demonstrate $R \geq R_L$ with an assurance level of 0.95 (i.e., CR is 0.05), we can get a minimum test sample size. When the test sample size increases given other design settings unchanged, the assurance level of $R \geq R_L$ will increase due to the statistical characteristic in hypothesis testing. More specifically, in the test with a selected n and CR, if the test can still be passed with a higher n (which is more difficult because of the uncertainty in product failures), the belief of the products to meet the lower level reliability requirement will be assured at a higher level (e.g., from 0.95 to 0.99). Consequently, CR will decrease when n increases and consumers may have more confidence in not receiving inferior products that have passed the RDT since they are assured to have satisfied R_L. PR and RP may increase because the risk of rejecting the reliable products will increase from the producer's perspective and the test will become more difficult to be accepted. W, as the weighted warranty cost ratio, will decrease as n increase. The mean of true product failure probability is affected by (n, c) and less failures are expected if the test can be passed by testing more samples (higher n) with the same c. Similarly, when

c increases, given other settings the same, the assurance level of $R \geq R_L$ will decrease because it results in less difficulty to pass the test and more failures can be expected. As a consequence, CR will increase as well as W. PR and RP will decrease when producers are faced with less risk of rejecting the reliable products and the test is easier to be passed. Third, when the lower level reliability requirement increases, with the same (n, c), the assurance level of $R \geq R_L$ will decrease and CR will increase. W will also increase since more failures can be expected even though the test is passed. RP may increase since it becomes harder to meet the lower level reliability requirement. However, PR may reduce since more products become less likely to satisfy R_L and will not be passed in the test.

Based on the discussions above, the trade-off patterns among different pairs of objectives can be illustrated in Figures 4.27 and 4.28 for the example of MO-MSRDT with two testing periods. When the lower level reliability requirement is fixed (i.e., fixed R_{L1},R_{L2}) or the maximum allowable failures is fixed (i.e., fixed c_1,c_2), trade-off patterns have been identified in six pairs of objectives including CR vs PR, CR vs RP, CR vs n, W vs PR, W vs RP and W vs n. For RDT plans with varied c_1, c_2, n and fixed R_{L1},R_{L2}, stringent trade-off patterns can be found for pairs of CR vs PR, CR vs RP, W vs PR and W vs RP, with concave curves. Relaxed trade-off patterns can be found for pairs of CR vs n and W vs n, with nearly linear or convex curves. For RDT plans with varied R_{L1},R_{L2} and fixed c_1, c_2, Three pairs of CR vs PR, CR vs RP and W vs PR present the concave curves along the change of lower reliability requirements. W and RP show stringent trade-off pattern with a single concave curve. CR and n show various trade-off patterns along with the change of R_{L1},R_{L2}, such as concave, convex and nearly linear curves. W and n show the relaxed trade-off pattern with a single convex curve. The trade-off between CR and PR can be intuitive because of their links with type-II and type-I errors, respectively. For instance, if producers want to reduce the PR by always accepting the RDT given any test samples, CR will be higher since any inferior products will pass the test. In the opposite, if CR is to be reduced by decreasing the lower level reliability requirement and maximum allow-

able failures or increasing test sample size, the producer will have higher risk to reject the reliable products without passing the test and PR will increase. Similarly, for CR and RP, as CR being reduced with increased n or decreased maximum allowable failures, the difficulty of passing the test may increase and RP may increase. The trade-off between CR and n can be more intuitively understood based on the fact that more test samples can improve the assurance level of demonstrating the lower level reliability requirements. W from Eq. (4.21) is constructed from the mean of true product failure probability conditioning on passing the test, which shares similar interpretations and changing patterns with CR (i.e., probability of $R < R_L$ conditioning on passing the test). Therefore, W also have trade-off patterns with the other three objectives including PR, RP and n. For the other four pairs of objectives including CR vs W, PR vs RP, PR vs n and RP vs n, no trade-off patterns have been observed. All those pairs may change in the same direction when other design settings change. As shown in Appendix, under either fixed c_1, c_2, or R_{L1},R_{L2}, these objectives will show positive relationships without obvious trade-off patterns.

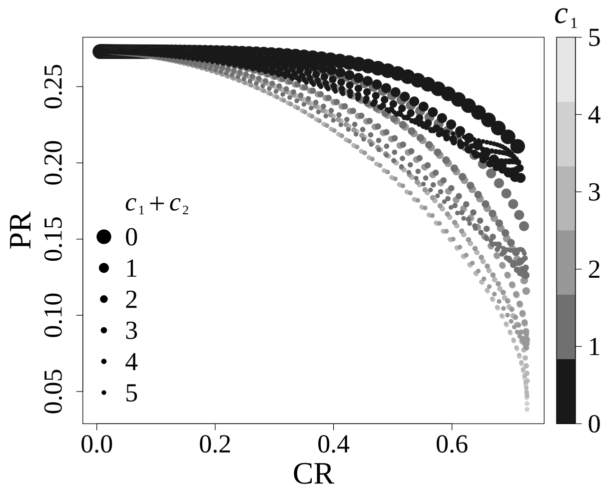
4.5.2 Impact of Design Parameters

After exploring the trade-offs among all pairs of five objectives, all the six pairs with trade-off patterns have been identified, including CR vs PR, CR vs RP, CR vs n, W vs PR, W vs RP and W vs n. It is noticed that some of the trade-off patterns may change with different design settings such as the maximum allowable failures and the lower level reliability requirements. In order to understand the impact of design parameters on the change of trade-off patterns for different pairs of objectives, several questions need to be answered including how(or why) the objective values(or the trade-off patterns) change with different inputs of design parameters and which inputs play dominating(or negligible) roles in affecting the change of trade-off patterns. The underlying reasons of these effects should also be intuitively explained.

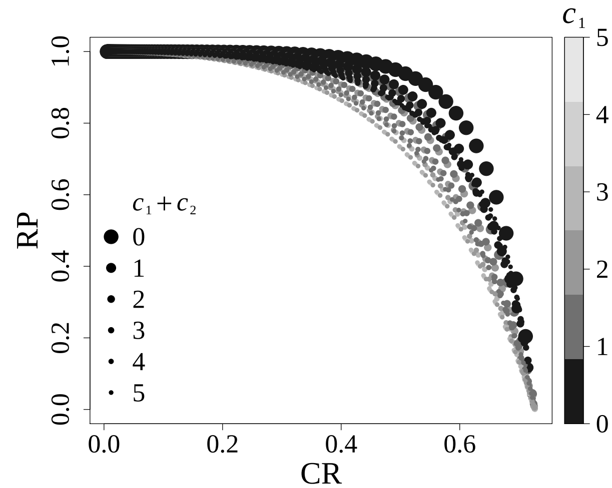
The impacts of different maximum allowable failures (e.g., c_1, c_2) will be explored by fixing the lower level reliability requirements (e.g., R_{L1},R_{L2}). Since c_1 and c_2 correspond to two separate demonstration periods respectively, $c_1 + c_2$ will represent the overall requirement on maximum allowable failures. For CR and PR in Figure 4.27a, the increase of $c_1 + c_2$ leads to the reduction of assurance level of demonstrating the lower level reliability requirement as CR increases. PR decreases because reliable products become less likely to be rejected when the test becomes easier to pass. The trade-off curves rotate clockwise towards higher CR and lower PR. The separate impact of changing c_1 may be more obvious than c_2, which shows consistent color changes along with the swinging curves. Since the test becomes easier to pass with more failure allowed, RP tends to decrease. In Figure 4.27b, the trade-off curves also swing clockwise towards higher CR and lower RP. However, for both CR vs PR and CR vs RP, the changes in trade-off strength and trade-off patterns are not significant. In the contrast, the impacts of (c_1, c_2) on trade-off between CR and n can be significantly observed in Figure 4.27c. For example, when $c_1 + c_2$ increases, CR will increase for any given test sample size and the curves will shift towards higher CR and changes from convex to nearly linear, which means the trade-off between CR and n becomes stronger. The impacts of (c_1, c_2) on W will be similar with CR, which leads to higher W when $c_1 + c_2$ increases. Therefore, for W vs PR in Figure 4.27d and W vs RP in Figure 4.27e, we can observe the clockwise rotating trade-off curves without much change in the trade-off patterns and strength. In Figure 4.27f, the trade-off curves shift towards higher W for any given n and the trade-off between W and n becomes stronger.

Similarly, the impacts of different lower level reliability requirements (e.g., R_{L1},R_{L2}) can be explored with fixed maximum allowable failures (e.g., c_1, c_2). When either $(R_{L1}$ or $R_{L2})$ increases, meaning the reliability requirement for the corresponding testing period has been increased, CR will increase given the same test sample size because the assurance level of demonstrating the lower level reliability requirements increases. PR will decrease since more products will become less likely to meet the lower level reliability re-

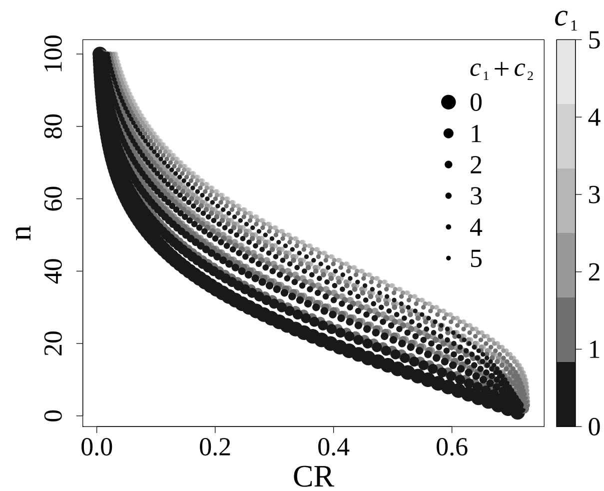
quirement when the test is more likely to be rejected. As shown in Figure 4.28a, the trade-off between CR and PR will become weaker and the curves are shifting from top left with higher PR lower CR to bottom right with lower PR and higher CR. Comparing R_{L1} and R_{L2} separately, the changes of R_{L1} lead to corresponding CR and PR at different curves more randomly. However, the changes of R_{L2} may result in more consistent shifting of curves. Since R_{L2} covers both time periods, it will have more impact on determining the overall lower level reliability requirement change. Meanwhile, the trade-off between CR and RP becomes stronger as shown in Figure 4.28b. The trade-off curves expand towards top right with higher CR and RP. R_{L2} still has more consistent impact on the change of trade-off patterns. However, the increase of either (R_{L1}, R_{L2}) may not change the values of RP, which is only relevant to the test sample size, maximum allowable failures and prior knowledge of true product failure probability. For CR and n in Figure 4.28c, there will be significant pattern changes when (R_{L1}, R_{L2}) increases. When the lower level reliability requirement has lower values, the trade-off curves are convex with lower CR as n increases. When (R_{L1}, R_{L2}) increases, the trade-off curves shift towards higher CR and have pattern changes from convex curves to nearly linear curves then to concave curves. The trade-off between CR and n becomes stronger. For example, when $R_{L2} = 0.8$, to achieve an acceptable CR below 0.1, we may need to $n = 20$. However, if $R_{L2} = 0.9$, to achieve the same level CR, we may need a significant increase of $n = 200$. W, which calculated from the mean of true product failure probability conditioning on passing the test is not affected by the change of lower level reliability requirement. Therefore, for W and PR in Figure 4.28d, the increase of (R_{L1}, R_{L2}) shifts the curves down towards lower PR and leads to weaker trade-off. Since RP and n are not affected by the change of lower level reliability requirement neither, the single concave curve for W vs RP in Figure 4.28e and single convex curve for W and n in Figure 4.28f show strong trade-off patterns for these two pairs respectively.



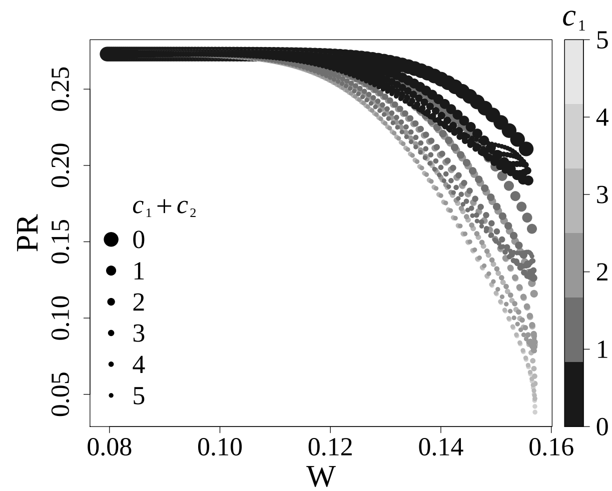
(a) Trade-off : CR vs PR, fixed R_{L1},R_{L2}



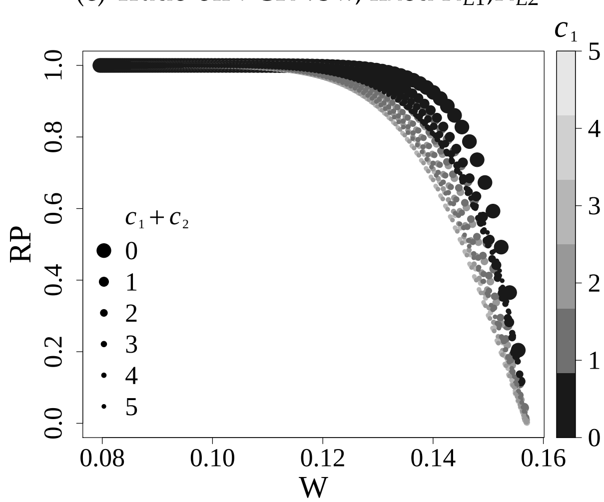
(b) Trade-off : CR vs RP, fixed R_{L1},R_{L2}



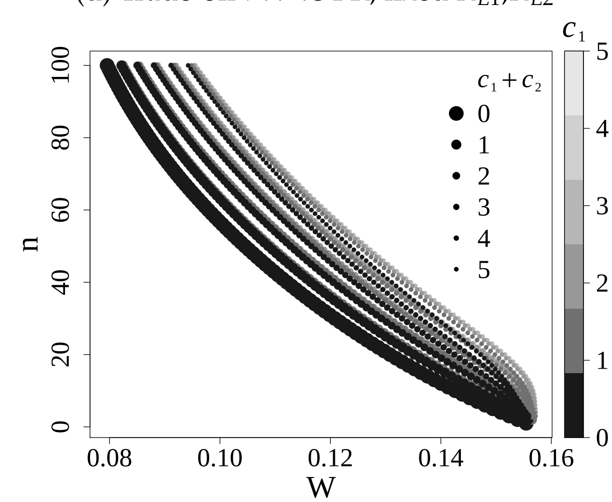
(c) Trade-off : CR vs n, fixed R_{L1},R_{L2}



(d) Trade-off : W vs PR, fixed R_{L1},R_{L2}

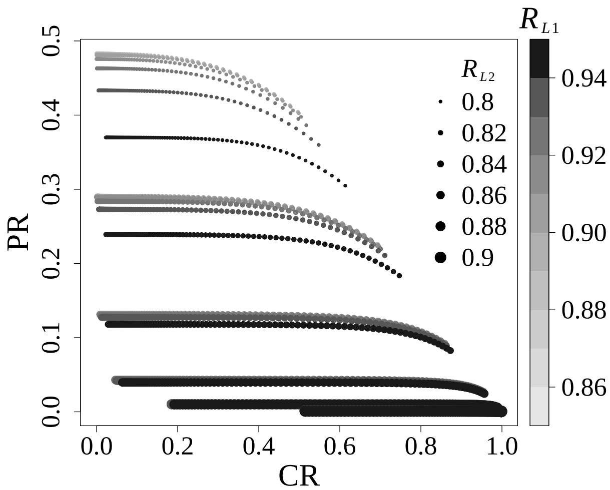


(e) Trade-off : W vs RP, fixed R_{L1},R_{L2}

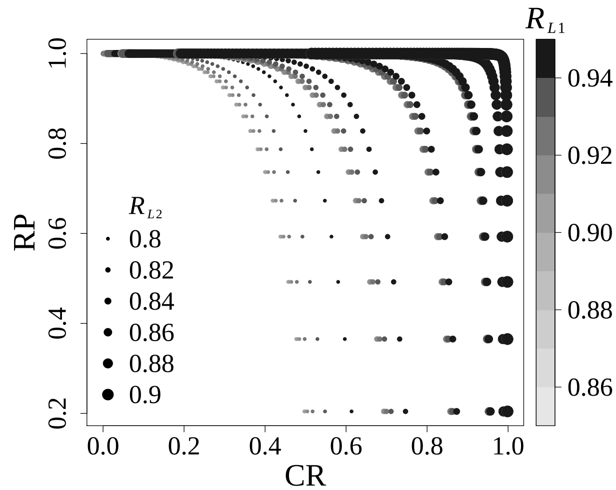


(f) Trade-off : W vs n, fixed R_{L1},R_{L2}

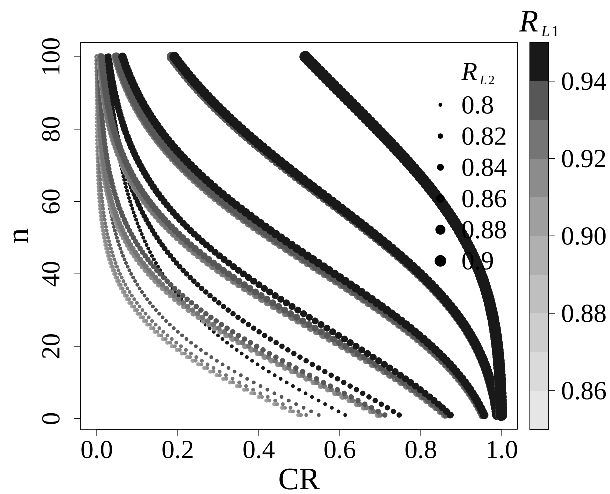
Figure 4.27: Trade-offs among CR, PR, RP, n, W with fixed R_{L1},R_{L2}



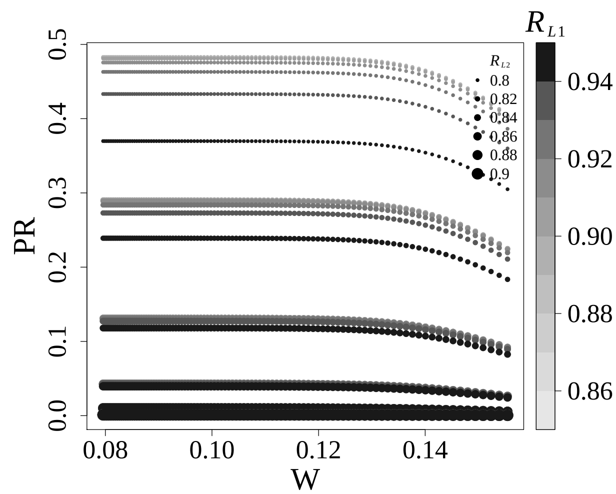
(a) Trade-off : CR vs PR, fixed fixed c_1, c_2



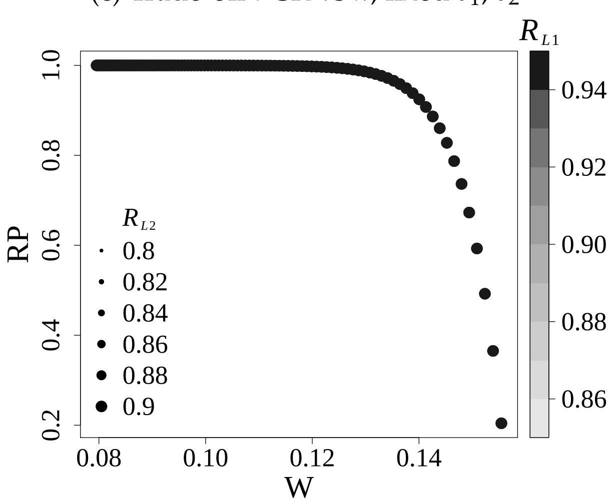
(b) Trade-off : CR vs RP, fixed c_1, c_2



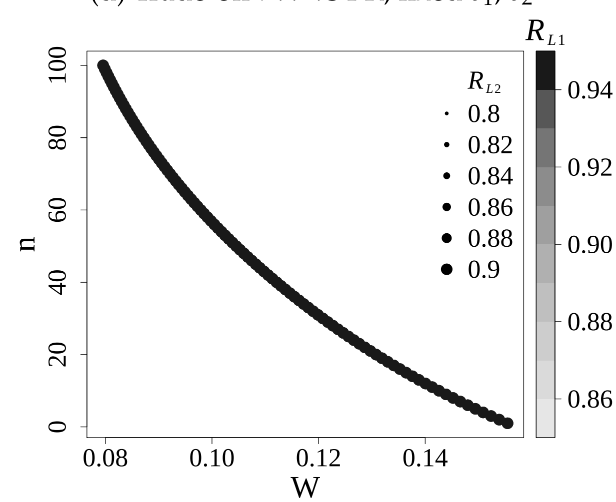
(c) Trade-off : CR vs n, fixed c_1, c_2



(d) Trade-off : W vs PR, fixed c_1, c_2



(e) Trade-off : W vs RP, fixed c_1, c_2



(f) Trade-off : W vs n, fixed c_1, c_2

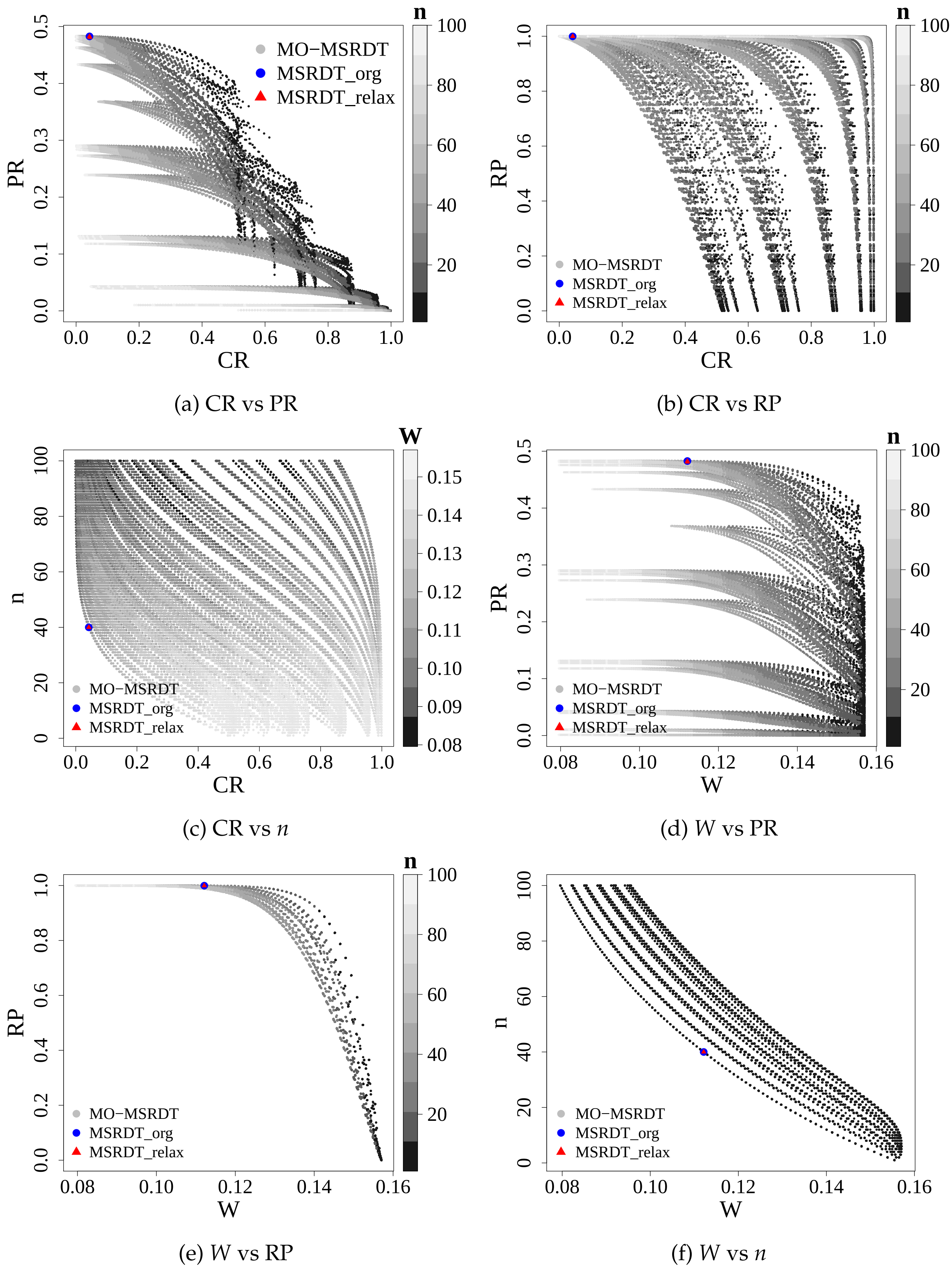
Figure 4.28: Trade-offs among CR, PR, RP, n, W with fixed c_1, c_2

4.5.3 Proposed vs Single Objective Designs

The proposed MO-MSRDT design advances conventional single objective design in several perspectives such as the multi-objective consideration, the relaxed assumption of pre-specified lower level requirements, etc. The comparison will be made to illustrate the advantages of the proposed MO-MSRDT design over existing single-objective RDT designs including BRDT and MSRDT. Conventional BRDT design aims to find the plan (n, c) that minimizes the BRDT cost under controlled CR and pre-specified R_L. In practice, zero allowable failure $c = 0$ may be selected since it leads to the minimum test sample size. However, since BRDT only focuses on a single testing period or multiple failure states, it is not comparable with the proposed MO-MSRDT which considers multiple states in the test. Existing single-objective MSRDT will be compared with the proposed work and illustrated with two time periods. There will be two variations in comparison. The first one is the original design (MSRDT-org) that will find the test plan (n, c_1, c_2) and minimize the overall test sample size n. The reliability R_1, R_2 for each cumulative time period will satisfy the lower level reliability requirements $R_1 \geq R_{L1}$, $R_2 \geq R_{L2}$. Single optimal plan will be obtained for MSRDT-org with controlled CR and pre-specified R_{L1}, R_{L2}. The second variation is MSRDT-relax which relaxes the assumption on pre-specification of lower level reliability requirements. For each cumulative time period, there will be a acceptable range of R_{L1}, R_{L2} to be considered and the test plan $(n, c_1, c_2, R_{L1}, R_{L2})$ will be determined to minimize n. It will be interesting to explore the relationships between the optimal plans from two variations of existing MSRDT designs and the Pareto Front from the proposed MO-MSRDT design framework.

In Table 4.11, the differences among BRDT, MSRDT-org, MSRDT-relax and the proposed MO-MSRDT have been summarized. Because of the relaxation, there will be several optimal plans being selected in MSRDT-relax that will minimize the test sample size, compared with the single optimal test plan in MSRDT-org. The Pareto Front in MO-MSRDT is composed of the test plans with balanced trade-offs for all five objectives with

the purpose to reduce each objective as much as possible. The proposed MO-MSRDT design will have significantly larger solution space (e.g., 109620) as well as richer information of the optimal plans in Pareto Front (e.g., 95863). Meanwhile, the solution space of the proposed design may be subject to the changes of user preferences on different design settings. For example, if the user requires narrowing the acceptable ranges of lower level reliability requirement, the solution space can be reduced. Several observations can be made from the visualization plots in Figure 4.29. First, the two variations of MSRDT share the same minimum test sample size for the optimal plans and do not have much difference in other objective values. Second, the Pareto Front of six pairs of objectives still follow the trade-off patterns identified before. Third, for each pair of objectives, the corresponding values in optimal plans from MSRDT-org or MSRDT-relax will be either dominated by some solutions from Pareto Front of MO-MSRDT (e.g., CR vs PR in Figure 4.29a, CR vs RP in Figure 4.29b, W vs PR in Figure 4.29d, W vs RP in Figure 4.29e), or becomes non-dominated plans along with the Pareto Front (e.g., CR vs n in Figure 4.29c, W vs n in Figure 4.29f). In the proposed framework, after considering the trade-offs among multiple objectives, the Pareto Front can provide more choices with dominated or non-dominated values in the objectives. In other words, the optimal solution from existing MSRDT designs can be some special cases in the Pareto Front of the proposed MO-MSRDT design. We can also trace back to the original solution space of the proposed MO-MSRDT, as shown in Figures 4.30 and 4.31. The illustrated examples in Figure 4.30 has fixed $R_{L1} = 0.94$ and $R_{L2} = 0.82$ for the selected test plans with varied c_1, c_2 and n, and in Figure 4.31 has fixed $c_1 = 0$ and $c_2 = 0$ with varied R_{L1}, R_{L2} and n, both of which represent the selection procedure of Pareto Front from original solution space of each pair of objectives with trade-offs. The test plans in Pareto Front (green) consist of a large proportion and only a small portion of original plans are selected.

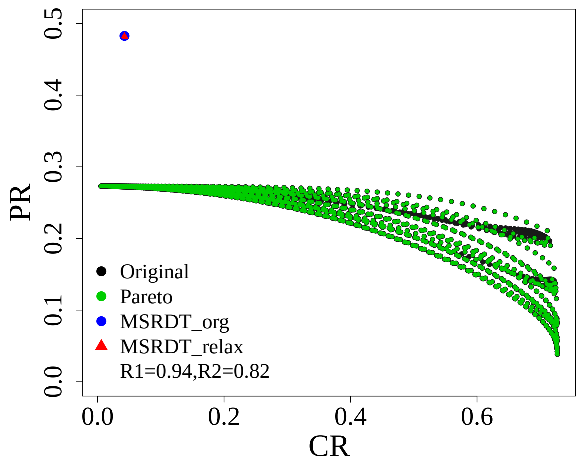


(a) CR vs PR (b) CR vs RP

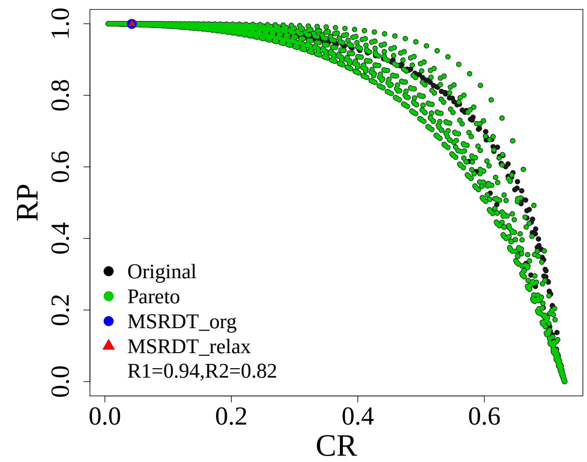
(c) CR vs n (d) W vs PR

(e) W vs RP (f) W vs n

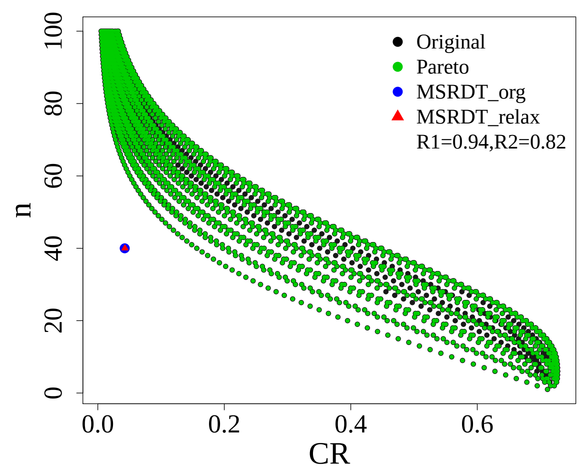
Figure 4.29: Proposed MO-MSRDT Pareto Front vs Existing MSRDT optimal plans



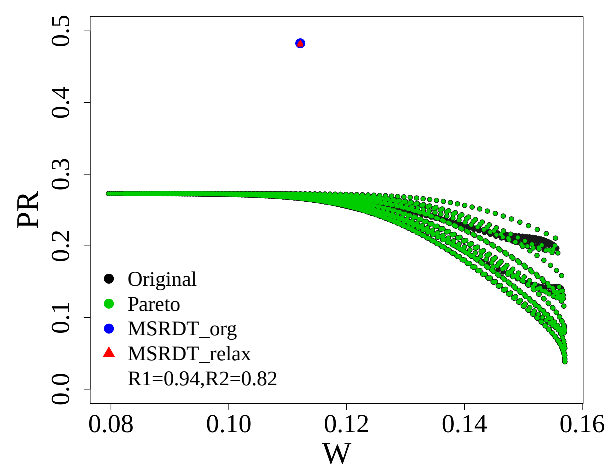
(a) CR vs PR



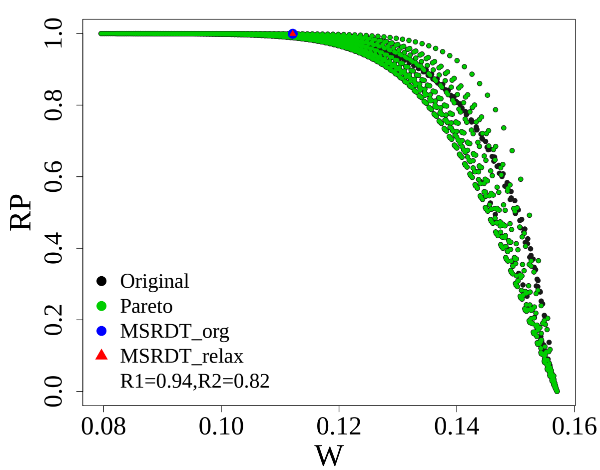
(b) CR vs RP



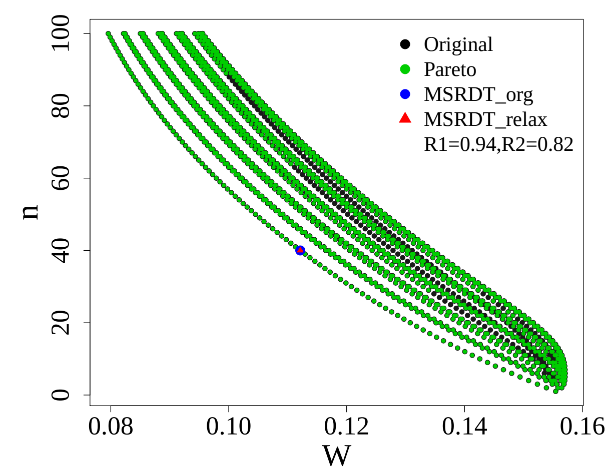
(c) CR vs n



(d) W vs PR

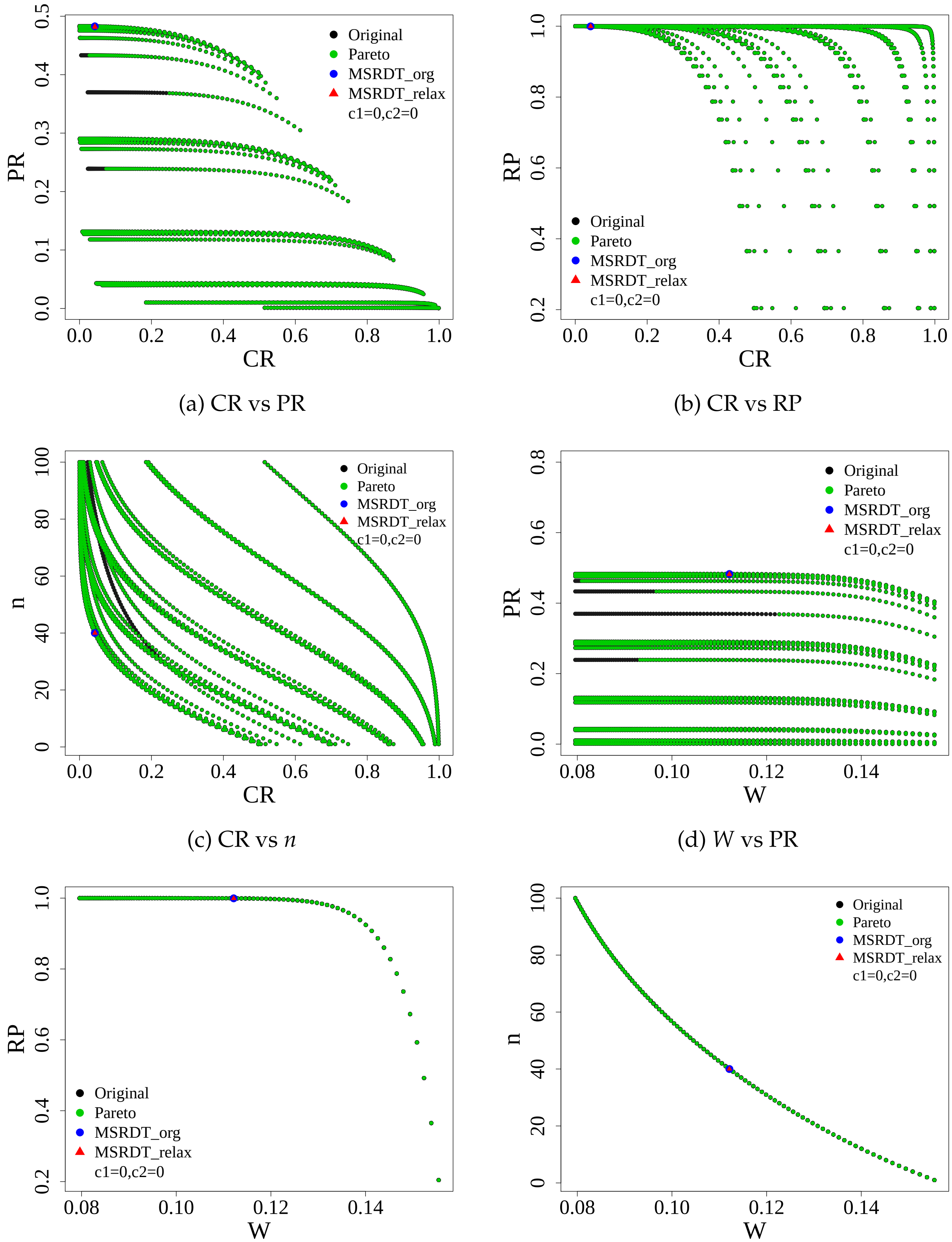


(e) W vs RP



(f) W vs n

Figure 4.30: Proposed MO-MSRDT vs Existing MSRDTs, fixed R_{L1},R_{L2}



(a) CR vs PR

(b) CR vs RP

(c) CR vs n

(d) W vs PR

(e) W vs RP

(f) W vs n

Figure 4.31: Proposed MO-MSRDT vs Existing MSRDTs, fixed c_1,c_2

Table 4.11: Comparison with single objective design

Design	Objectives	Parameters	Plans space	Optimal plan	Specified
BRDT	minimum n	c	NaN	Single	CR, R
MSRDT-org	minimum n	c_k	21	Single	CR
MSRDT-relax	minimum n	c_k, R_k	1134	Multiple(7)	CR, R_k
MO-MSRDT	CR, PR, RP, n, W	c_k, R_k	109620	Pareto Front(95863)	User preferences

Settings: Original data rows = 109620, $\frac{C_{w_1}}{C_p} = 1$, $\frac{C_{w_2}}{C_p} = 0.7$, $\pi \sim \mathrm{Dir}(4.7, 15.7, 79.6)$

$n = 0 \sim 100; R_{L1} = 0.85 \sim 0.95$ by $0.01; R_{L2} = 0.8 \sim 0.9$ by $0.02; c_1 = 0 \sim 5; c_2 = 0 \sim 5; c_1 + c_2 \leq 5$

4.5.4 Optimal Plans Selection with User Preferences

The generation process of optimal test plans can be strategically planned, as shown in Figure 4.32. From the baseline knowledge of the design settings such as the cost budget, failure and risk tolerance, lower level reliability requirements and prior information of product reliability, the design settings can be determined. For example, based on the budget in RDT, we can limit the test sample size (e.g., $n \leq 100$). Based on the characteristic of product, the minimum and maximum values of lower level reliability to be demonstrated for the each period can be limited in acceptable ranges (e.g., $0.85 \leq R_{L1} \leq 0.95, 0.8 \leq R_{L1} \leq 0.9$). We can also refer to the previous analysis of trade-offs among multiple objectives and determine the design settings with specific focuses or controls on trade-offs of selected objectives. For example, in order to control the trade-off between CR and PR, the acceptable ranges of lower level reliability requirements can be adjusted. With reduced values of lower level reliability requirements, the test plans may have lower CR and higher PR, which can be controlled based on the user preferences.

With the design settings settled, the original pool of RDT plans can be obtained, which includes all the combinations of design parameters from the predetermined ranges. The next step is to balance the trade-offs using Pareto Front method and obtain the set of RDT plans with Pareto efficiency. The plans in the Pareto Front are non-dominated by each other. The Pareto Front will contain a substantially smaller amount of RDT plans compared with the original pool, however, it may still be too dense to easily identify a few optimal plans. We also notice that users may have different preferences or tolerance on specific objectives, which may be subject to the cost budget or risk aversion. For example, if the manufacturers care the most about customer satisfaction, they may want to control the CR below 0.05 so that they have sufficient assurance on the demonstrated product reliability. The user-preference-based screening can be performed on the Pareto Front to help practitioners to further narrow down the set of optimal plans and select the ones with best fit.

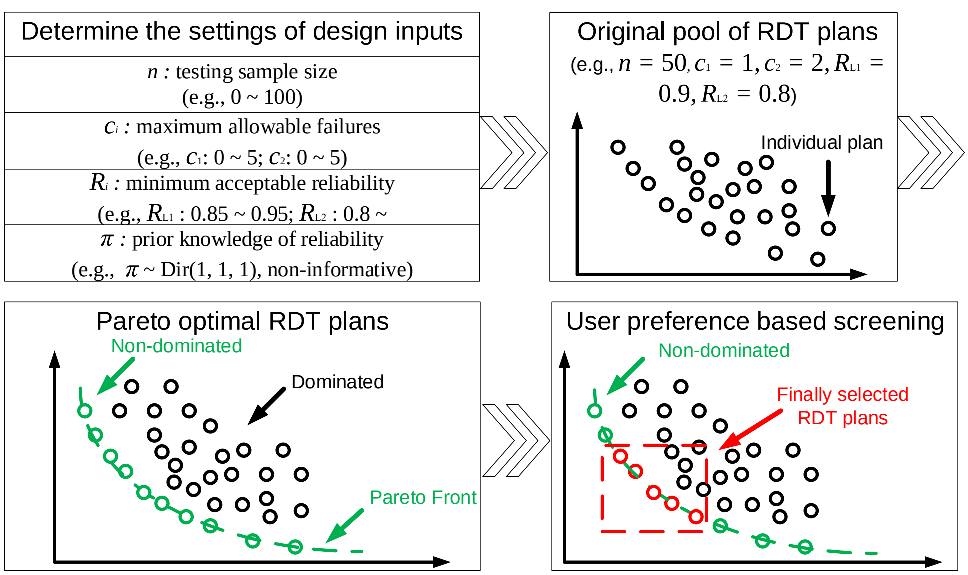


Figure 4.32: Flow chart of decision process

In order to narrow down the selections from Pareto Front and make it easier for practitioners to choose the optimal plans, the screening based on multiple objectives can be performed. The selection ratios of Pareto Front by truncating different objectives have

been compared in Figure 4.33. By setting the upper bound tolerance for each of the objective, the corresponding selection ratio of Pareto Front can be obtained. The objectives have different changing patterns when the upper bound tolerance changes. For example, the slope of selection ratio change will be high when CR's tolerance is low (e.g., CR$\leq$ 0.05). When tolerance is above 0.2, increasing the tolerance will lead to linear increase of selection ratio. However, the selection ratio for RP will increase linearly as the tolerance increase (e.g, RP $\leq$ 0.8) and start to grow exponentially. The differences in the selection ratios after screening based on different objectives may result from the distributions of each objectives as well as the trade-offs among different objectives. In practice, the screening can be performed not only on a single objective but also on different combinations of objectives in order to satisfy the user preferences from different perspectives (e.g., cost budget, risk tolerance, reliability requirement, etc.). When the objectives are controlled to a lower level, the selection of optimal plans will be reduced to a small portion from the Pareto Front.

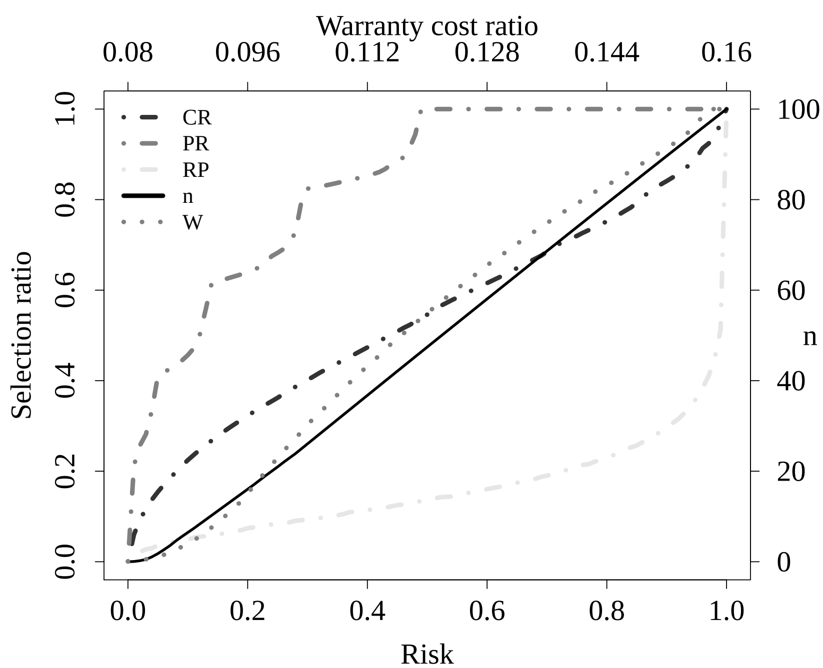


Figure 4.33: Selection ratios of Pareto Front by truncating different objectives

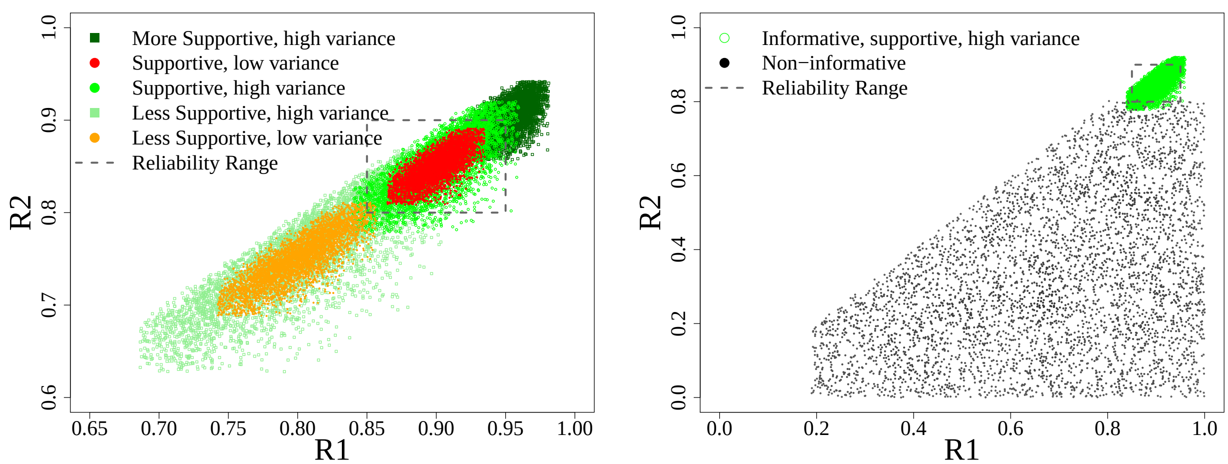
4.5.5 Impact of Prior Information

Table 4.12: Different prior settings

	Supportive		Non-informative
$\pi \sim$ Dirichlet	$(10, 5, 85)$	$(10, 5, 85) * 3$	$(1, 1, 1)$
Mean(π)	$(0.1, 0.05, 0.85)$	$(0.1, 0.05, 0.85)$	$(0.33, 0.33, 0.34)$
Var(π)	$(0.0000, 0.0833, 0.0556)$	$(0.0000, 0.0139, 0.0013)$	$(0.0000, 0.0072, 0.006)$
	More supportive	Less supportive	
$\pi \sim$ Dirichlet	$(5, 5, 90) * 2$	$(20, 5, 75) * 0.5$	$(20, 5, 75) * 2$
Mean(π)	$(0.05, 0.05, 0.9)$	$(0.2, 0.05, 0.75)$	$(0.2, 0.05, 0.75)$
Var(π)	$(0.0000, 0.0119, 0.0004)$	$(0.0000, 0.0119, 0.0037)$	$(0.0000, 0.0031, 0.0009)$
	Settings: n: $0 \sim 100$ $C_{w_1} = 100, C_{w_2} = 70, C_p = 100$ R_1: $0.85 \sim 0.95$ by 0.01; R_2: $0.8 \sim 0.9$ by 0.02 c_1: $0 \sim 5$ by 1; c_2: $0 \sim 5$ by 1; $c_1 + c_2 \leq 5$		

To explore the impacts of different priors on the change of trade-off patterns, several different Dirichlet priors of true product failure probability over multiple time periods are used with various prior mean and variance specifications, as summarized in Table 4.12 and visualized in Figure 4.34. For example, the average supportive prior Dir(10, 5, 85) has the mean failure probabilities (0.1, 0.05, 0.85) for the first period, the second period and the rest of product life time, respectively. Equivalently, the average prior belief of product reliability can be 0.9 for first period and 0.85 for the first two periods. Dir((10, 5, 85)*2) has the same mean failure probabilities but lower variances for each period. As the mean failure probabilities decrease, Dir((5, 5, 90)*2) is an more supportive prior with high variance corresponding to the average prior Dir(10, 5, 85). Dir((20,5,75)*0.5) and Dir((20,5,75)*2) are less supportive priors with increasing mean failure probabilities, with

low and high variances respectively. Dir(1, 1, 1) is a non-informative prior which assumes each separate period has an equal mean failure probability of 0.33, with much higher variance than the other informative priors. For each pair of the objectives with trade-off patterns, the comparisons among different priors will be made regarding supportive and not supportive or informative and non-informative, with the investigation of impacts under different failure probability means and/or variances.



(a) Informative priors

(b) Non-informative priors

Figure 4.34: Difference priors

The exploration of trade-offs for each pair of objectives will be two-folded, with either fixed lower level reliability requirements in Figure 4.35 or fixed maximum allowable failures in Figure 4.36. To illustrate, when R_{L1} and R_{L2} are fixed, the trade-off between CR and PR (Figure 4.35a) may change significantly with different priors. For supportive priors sharing the same mean of failure probability, the prior with lower variance can increase the assurance level of the product meeting the reliability requirement and CR will decrease. PR tends to increase since the rejection of test will lead to the rejection of more reliable products under the belief of prior knowledge. The trade-off curves are shifting upper left towards lower CR and higher PR. With similar level of prior variance (e.g., high variance), more supportive prior will result in lower CR and higher PR and the changing patterns when reducing the mean of failure probability is the same as reducing the vari-

ance. In the opposite, when the means of failure probability increase, the less supportive priors will shift the trade-off curves towards higher CR and lower PR. For less supportive priors with same mean, lower prior variance will indicate a stronger belief of unreliable product, which leads to the increase of CR, but the decrease of PR due to the fact that most products are unreliable and may be rejected correctly. Non-informative prior with higher mean of failure probability and higher variance will have lower level of PR and a wider range of CR values. However, when c_1 and c_2 are fixed in Figure 4.36a, the changing patterns of trade-off may not be obvious due to the overlapped curves corresponding to different lower level reliability requirements.

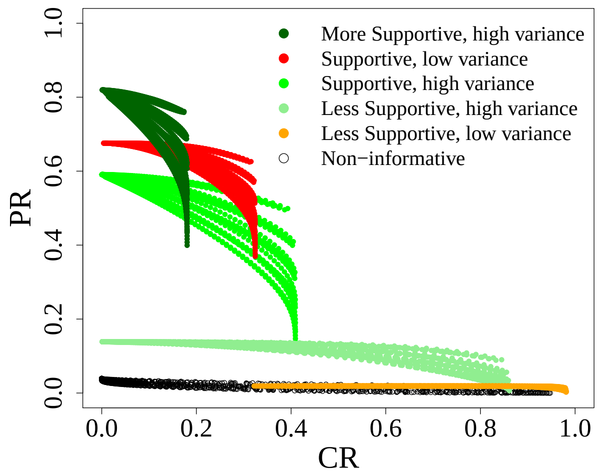
With fixed R_{L1} and R_{L2}, for CR and RP, in Figure 4.35b, supportive priors with higher means of failure probability or lower prior variance may result in lower CR and rotate the trade-off curves clockwise. For less supportive priors with higher means of failure probability or lower prior variance, the trade-off between CR and RP becomes stronger and the curves will expand towards higher CR and RP. For CR vs n in Figure 4.35c, supportive priors with lower means of failure probability or lower variance may result in lower CR, given any test sample size and the curves will rotate clockwise with reduced trade-offs. Less supportive priors tend to shift the curves towards higher CR and lead to stronger trade-off. The shape of the curves may also change from convex curves to concave curves, with higher means of failure probability and lower variance. Non-informative prior tends to have a wider range of CR values with the spreading convex trade-off curves. Similarly, for CR vs PR with fixed c_1 and c_2, the changing patterns of CR vs RP in Figure 4.36b and CR vs n Figure 4.36c with different priors are not obvious.

For the trade-offs between W and either PR, RP or n, the changing patterns will be similar to that of CR and either PR, RP or n, respectively, since both W and CR measures the failure probability given the test is passed from different perspectives. For example, when R_{L1} and R_{L2} are fixed, for W vs PR in Figure 4.35d, supportive priors with lower means of failure probability or lower variance may shift the trade-off curves towards lower W

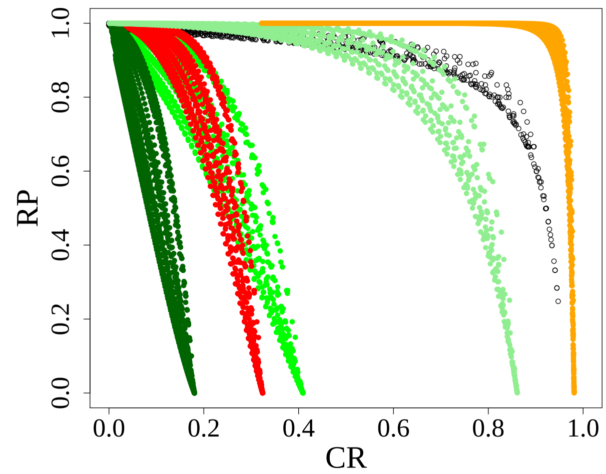
and higher PR. Less supportive priors with higher means of failure probability or lower variance may lead to higher W and lower PR. Non-informative prior tends to have lower PR with a wide range of W. The changing patterns becomes unobvious when c_1 and c_2 are fixed, as shown in Figure 4.36d. W vs RP, W vs n also shares similar trade-off changing patterns with CR vs RP and CR vs n respectively, with fixed R_{L1} and R_{L2}. However, when c_1 and c_2 are fixed, the changing patterns of trade-off curves can be observed for W vs RP in Figure 4.36e and W vs n in Figure 4.36f. This is because the objectives W, RP and n are not affected by different lower level reliability requirements and the individual curves with fixed maximum allowable failures in Figures 4.36e and 4.36f will be among the curves from Figures 4.35e and 4.35f when R_{L1} and R_{L2} are fixed, which shares similar trade-off changing patterns.

In addition, the prior knowledge may also have impacts on the density of Pareto Front as shown in Figure 4.37. For the priors with the same means of failure probability, lower variance of the prior distributions may help reduce the number of plans in Pareto Front. For supportive priors, with similar variances, more supportive priors will help narrow down the selection of plans as well. Since the belief of reliability is higher, the Pareto Front will be concentrating towards the plans that can meet the reliability requirements. Oppositely, if the prior information is less supportive, the Pareto Front will also be reduced since more plans are believed not able to meet the reliability requirements.

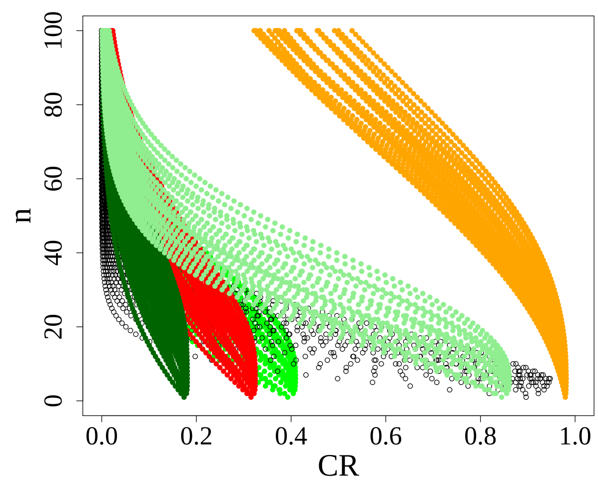
In summary, the exploration of the impact of different prior distributions on the trade-off patterns among multiple objectives can be beneficial in two perspectives. First, based on the prior knowledge of specific product, practitioners can determine which objectives to focus on in order to satisfy their needs. For example, if the prior knowledge supports high reliability, we will have more control in CR and W, from which lower CR and W may indicate better customer satisfactions and less complains in warranty services. Second, the prior knowledge can help control the richness of the Pareto Front, which may help better inform the practitioners to select the optimal test plans more efficiently.



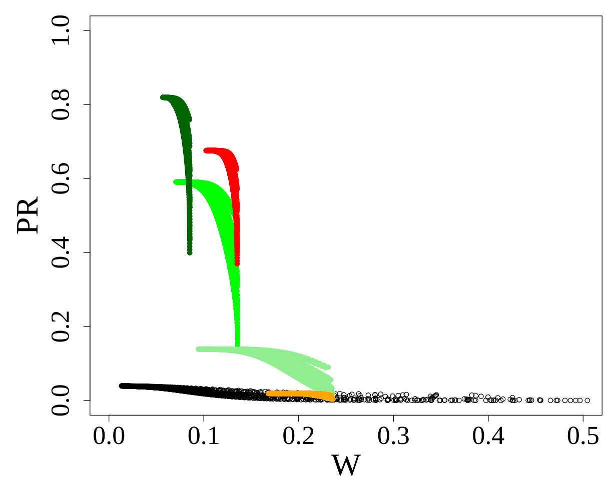
(a) Priors : CR vs PR



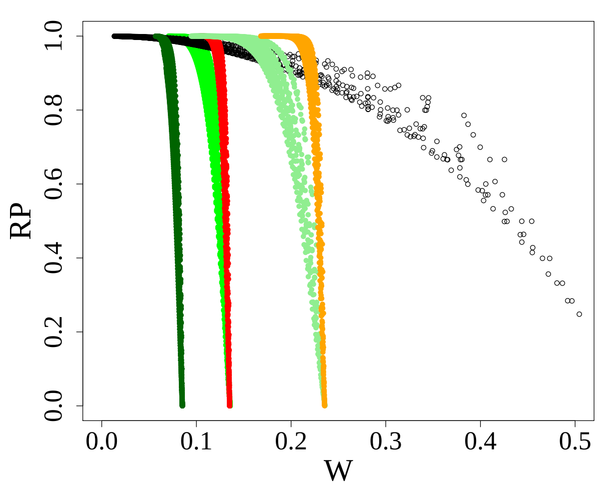
(b) Priors : CR vs RP



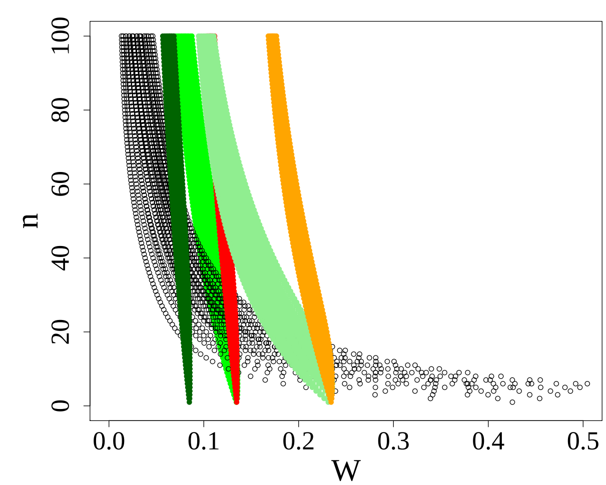
(c) Priors : CR vs n



(d) Priors : W vs PR

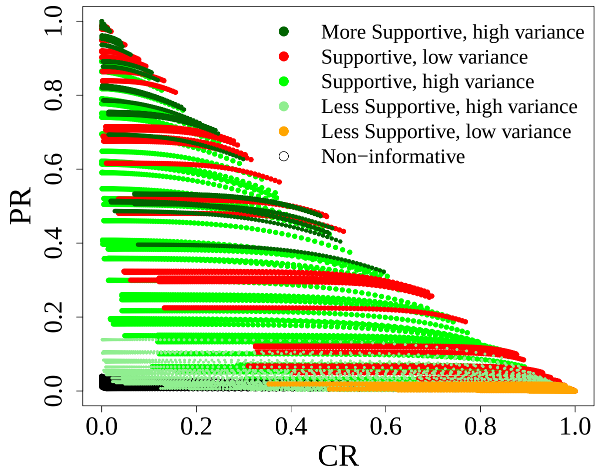


(e) Priors : W vs RP

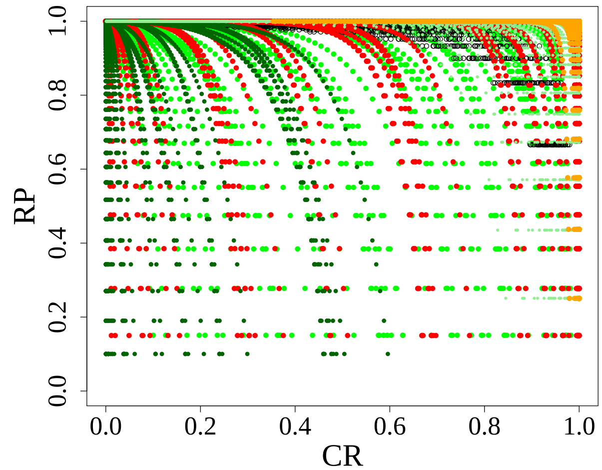


(f) Priors : W vs n

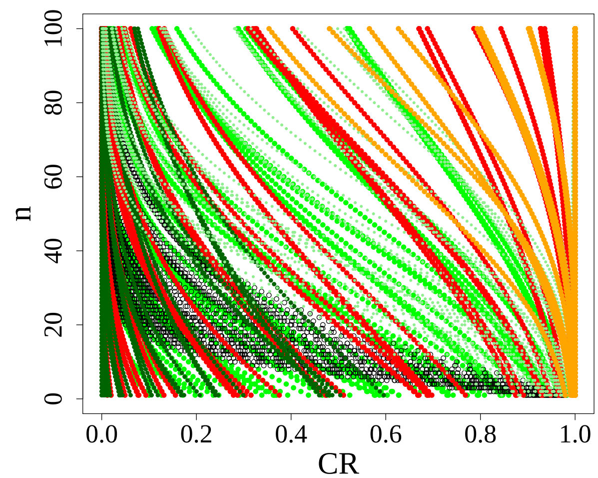
Figure 4.35: Trade-offs with different priors, fixed R_{L1},R_{L2}



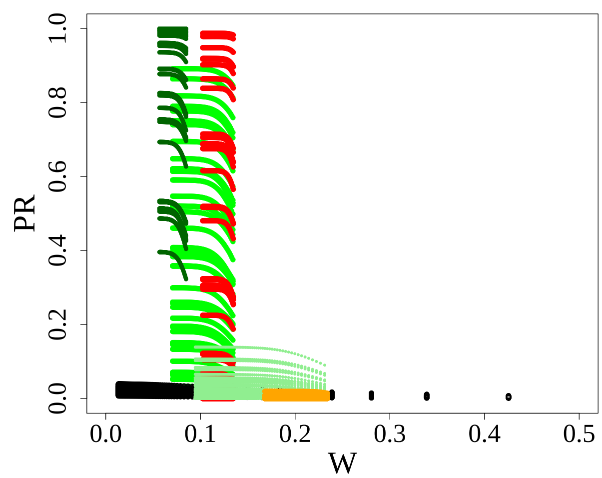
(a) Priors : CR vs PR



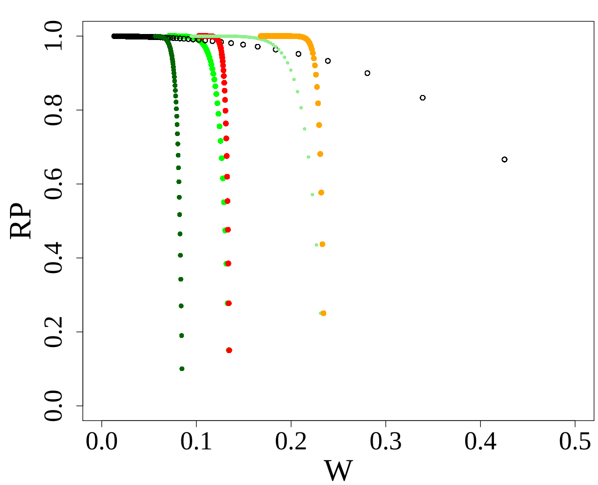
(b) Priors : CR vs RP



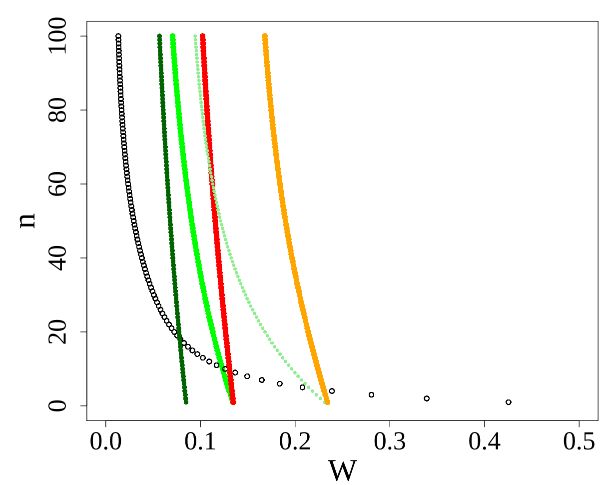
(c) Priors : CR vs n



(d) Priors : W vs PR



(e) Priors : W vs RP



(f) Priors : W vs n

Figure 4.36: Trade-offs with different priors, fixed c_1,c_2

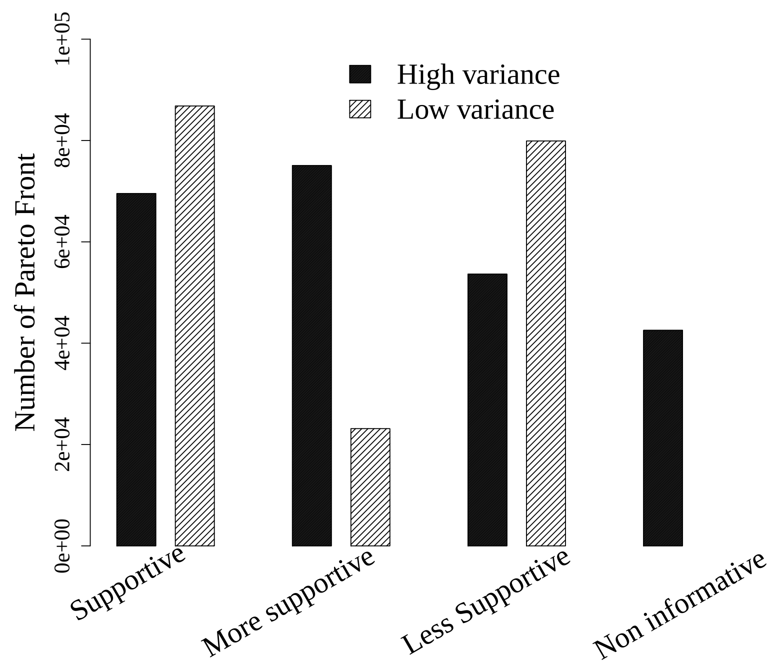


Figure 4.37: Number of Pareto Front for different priors

4.6 Conclusion Remarks

Existing RDT designs are limited in several perspectives regarding the failure states, objectives and design settings. In conventional BRDT design, single objective of minimizing the RDT cost is usually achieved by minimum test sample size with controlled risk criterion CR and pre-specified design parameters such as lower level reliability requirements. Meanwhile, it only focuses on single testing period or single failure state of the product, which may not be capable to meet the increasing reliability demand of customers. The MSRDT design in (Chen et al., 2017) extended the conventional BRDT to multiple failure states, such as multiple testing periods. However, it still focused on a single objective of RDT cost minimization. The BRDT design with multiple objectives consideration in (Lu et al., 2016) extended the conventional BRDT with the consideration of multiple objectives such as CR, PR, AP and n. By balancing the trade-offs among multiple objectives through Pareto Front method, optimal test plans can be selected based on different focuses of each objectives. However, the multiple objectives BRDT design has

no consideration for multiple failure states. Also, it only focused on the objectives for the test itself, without considering related reliability assurance activities such as warranty services cost. In addition, all the existing work assumes pre-specification of lower level reliability requirements, which may limit the flexibility and effectiveness in determining the optimal test plans.

This paper proposes the MO-MSRDT optimal design framework which comprehensively advances the existing RDT designs by considering multiple objectives for multiple states with relaxed design settings such as lower level reliability requirements. Five objectives in total such as CR, PR, RP, n and W have been considered and R_L is considered to be flexible, which can be selected from an acceptable range of values instead of a fixed and pre-specified value. Through exploring the interrelationships of each pair of objectives, the trade-off patterns from several pairs of objectives have been identified such as CR vs PR, W vs n, etc. In order to understand how different design settings may affect the change of trade-off patterns for different objectives, the impact of design parameters such as lower level reliability requirement and maximum allowable failures are evaluated. The Pareto Front of the proposed MO-MSRDT is also compared with the existing single objective MSRDT design with two variations to illustrate the dominance of test plans in Pareto Front and richness of selection. Based on the exploration of multiple objectives trade-offs and Pareto Front, the generation process of optimal test plans is provided to help practitioners with the final decision making. In addition, the impact of prior knowledge on the change of trade-off patterns as well as the selected Pareto Front has been evaluated based on simulated prior distributions with different means and variances.

To conclude, the proposed optimal MO-MSRDT design framework is the first to comprehensively incorporate multiple objectives and multiple states with flexible lower level reliability requirement into RDT design for failure count data. It advocates a more systematic and holistic vision in RDT designs for practitioners in reliability engineering. Even though the scope of this paper is focusing on failure count data, the concept of the frame-

work can be well extended to RDTs with other data types and models, such as failure time data or degradation data. In addition, it is also interesting to consider more objectives from related reliability assurance activities such as reliability growth cost (e.g., when RDT cannot be passed, the product needs to go back to reliability growth at additional cost). With more objectives considered, the selection of optimal test plans can better meet more comprehensive needs of practitioners with different preferences.

Chapter 5

Conclusion

In this dissertation, three advances in optimal RDT design for failure count data are proposed to resolve the data complexity, planning complexity and evaluation criteria complexity. Specifically, the data complexity from conventional BRDT design is that only single testing period or single failure mode is considered, which may not be able to meet the increasing reliability demand of customers. The planning complexity comes from the acceptance decision uncertainty of BRDT when conventional design only focuses on the BRDT itself without considering all the subsequent reliability assurance activities such as reliability growth test and warranty services. The evaluation criteria complexity comes from the single objective consideration of testing cost (e.g., test sample size) in conventional BRDT design, which may not be able to balance the trade-offs among different conflicting evaluation criteria (e.g., consumer's risk, producer's risk, testing cost,etc.) from all related reliability assurance activities.

Chapter 2 focuses on resolving the data complexity with the proposal of MSRDTs (i.e., multi-state RDTs), which includes MSRDTs over multiple time periods and MSRDTs for multiple failure modes. In the MSRDTs over multiple time periods, With every non-overlapping time period of interest treated as a state, a multinomial distribution can be assumed as the joint distribution of failure counts over all the time periods. Two different test strategies are considered to demonstrate the reliability requirements over cumulative time periods, which uses either cumulative failure counts for each cumulative period or separate failure counts over each non-overlapping period, as the passing criteria. In the MSRDTs for multiple failure modes, each failure mode is treated as a state and the reliabil-

ity requirements for each failure model can be customized based on different operating conditions and demonstrated simultaneously. The failure mode with higher reliability requirement tends to dominate the determination of overall minimum test sample size.

Chapter 3 focuses on resolving the planning complexity by considering the uncertainty of BRDT decision and its cost impacts on subsequent reliability assurance activities. A new optimal BRDT design framework is proposed with more global planning horizon to explicitly quantify the acceptance uncertainty of the test using acceptance probability and selects the optimal test plan that minimizes the overall costs of all related activities including BRDT, reliability growth and warranty services. Through the case study, the nonlinear relationships of different design parameters and cost components are evaluated. The changing patterns of overall cost under different cost scenarios and design settings demonstrate the effectiveness of the proposed work and advantages over conventional BRDT design in the overall cost minimization.

Chapter 4 focuses on resolving the evaluation criteria complexity by comprehensively considering multiple objectives of risk criteria and cost components from different related reliability assurance activities in optimal RDT design with multiple time periods and relaxed assumptions of pre-specified lower level reliability requirements. Specifically, five objectives including consumer's risk, producer's risk, rejection probability, RDT cost and warranty services cost are explored to identify their interrelationships and trade-offs. The impact of design parameters on the changing patterns of trade-offs are evaluated. Pareto Front method is used to balance the trade-offs among different objectives and determine the optimal test plans. Compared with existing RDT designs, the proposed MO-MSRDT design can provide more flexibility and richer information in determining/selecting optimal test plans. The decision making strategy for practitioners are also provided based on different objectives, trade-offs and user preferences.

For all the chapters, the impact of prior knowledge incorporation has been explored under the Bayesian framework. For example, in Chapter 2, when the prior knowledge

supports higher reliability, less minimum test sample size will be required for MSRDTs, as compared with non-informative priors. However, if the prior knowledge supports lower reliability than the pre-specified reliability to be demonstrated, more test samples may be needed. In Chapter 3, the impact of different prior distributions are compared through different combinations of prior means and variances. Higher means of reliability with less variance can significantly reduce the test sample size as well as the overall cost. In Chapter 4, the impact of prior knowledge on different changing patterns of objective trade-offs curves has been evaluated. For different pairs of objectives, different patterns such as rotate clockwise, shift/move towards original point, shift/move away from original point can be observed, which correspond to stronger or weaker trade-offs. Prior knowledge is also found to have influence on the richness of Pareto Front with different prior means and variances, which will affect the final decision of optimal test plans.

For future work, there can be three major aspects. First, since all the proposed work has been evaluated under Bayesian framework using Monte Carlo Simulation, it will be ideal to make thorough mathematical justifications through analytical derivations, which can reduce the computational efforts while designing such BRDT tests. Second, even though the scope of the proposed work is focusing on RDT design for failure count data, such as BRDT, MSRDT, etc., the frameworks are readily applicable to RDTs with other data types, such as failure time data or degradation data. Third, considering the related reliability assurance activities such as reliability growth test and warranty services, more detailed modeling can be explored regarding the reliability requirements, iterations of reliability growth and demonstration, costs based on revenue and market conditions and optimal decision making strategies, etc. In addition, it is also practically significant that the three recent advances in design of RDTs discussed in this dissertation can be implemented in a specific application context which focuses on product reliability design and development as well as improvements on customer experience.

References

Allmen, C. and Lu, M.-W. (1993). Sample size for failure detection and reliability demonstration. In *Annual Reliability and Maintainability Symposium 1993 Proceedings*, pages 386–394. IEEE.

Billinton, R. and Wang, P. (1999). Teaching distribution system reliability evaluation using monte carlo simulation. *IEEE Transactions on Power Systems*, 14(2):397–403.

Blischke, W. (1993). *Warranty cost analysis*. CRC Press.

Blischke, W. R. and Murthy, D. (1992). Product warranty management—i: A taxonomy for warranty policies. *European journal of operational research*, 62(2):127–148.

Borzsony, S., Kossmann, D., and Stocker, K. (2001). The skyline operator. In *Proceedings 17th international conference on data engineering*, pages 421–430. IEEE.

Cai, K.-Y., Wen, C.-Y., and Zhang, M.-L. (1991). A critical review on software reliability modeling. *Reliability Engineering & System Safety*, 32(3):357–371.

Chen, S., Lu, L., and Li, M. (2017). Multi-state reliability demonstration tests. *Quality Engineering*, 29(3):431–445.

Coolen, F. and Coolen-Schrijner, P. (2005). Non-parametric predictive reliability demonstration for failure-free periods. *IMA Journal of Management Mathematics*, 16(1):1–11.

Coolen, F. P. and Coolen-Schrijner, P. (2008). Bayesian reliability demonstration. In *Encyclopedia of Statistics in Quality and Reliability*. Wiley.

Crk, V. (2000). Reliability assessment from degradation data. In *Annual Reliability and Maintainability Symposium. 2000 Proceedings. International Symposium on Product Quality and Integrity (Cat. No. 00CH37055)*, pages 155–161. IEEE.

Crow, L. H. (1982). Confidence interval procedures for the weibull process with applications to reliability growth. *Technometrics*, 24(1):67–72.

Crow, L. H. (1984). Methods for assessing reliability growth potential. In *Annual Reliability and Maintainability Symposium, 1984. Proceedings.*, pages 484–489. IEEE.

Crowe, D. and Feinberg, A. (2001). *Design for reliability*. CRC press.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*, volume 16. John Wiley & Sons.

Derringer, G. and Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of quality technology*, 12(4):214–219.

Duane, J. (1964). Learning curve approach to reliability monitoring. *IEEE transactions on Aerospace*, 2(2):563–566.

Duran, B. and Booker, J. (1988). A bayes sensitivity analysis when using the beta distribution as a prior. *IEEE transactions on reliability*, 37(2):239–247.

Ellner, P. M. and Hall, J. B. (2006). An approach to reliability growth planning based on failure mode discovery and correction using amsaa projection methodology. In *RAMS'06. Annual Reliability and Maintainability Symposium, 2006.*, pages 266–272. IEEE.

Fan, T.-H. and Chang, C.-C. (2009). A bayesian zero-failure reliability demonstration test of high quality electro-explosive devices. *Quality and reliability engineering international*, 25(8):913–920.

Fernández, A. J. (2011). Optimal reliability demonstration test plans for k-out-of-n systems of gamma distributed components. *IEEE Transactions on Reliability*, 60(4):833–844.

Fries, A. and Sen, A. (1996). A survey of discrete reliability-growth models. *IEEE Transactions on Reliability*, 45(4):582–604.

Gerokostopoulos, A., Guo, H., and Pohl, E. (2015). Determining the right sample size for your test: Theory and application. In *2015 Annual Reliability and Maintainability Symposium Tutorial Notes*.

Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., and Sun, Y. (2010). A review on degradation models in reliability analysis. In *Engineering Asset Lifecycle Management*, pages 369–384. Springer.

Guo, H., Honecker, S., Mettas, A., and Ogden, D. (2010a). Reliability estimation for one-shot systems with zero component test failures. In *2010 Proceedings-Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–7. IEEE.

Guo, H., Jin, T., and Mettas, A. (2010b). Designing reliability demonstration tests for one-shot systems under zero component failures. *IEEE Transactions on Reliability*, 60(1):286–294.

Guo, H. and Liao, H. (2011). Methods of reliability demonstration testing and their relationships. *IEEE Transactions on Reliability*, 61(1):231–237.

Hamada, M. S., Wilson, A., Reese, C. S., and Martz, H. (2008). *Bayesian reliability*. Springer Science & Business Media.

Hsieh, P. I. and Ling, J. (1999). A framework of integrated reliability demonstration in system development. In *Annual Reliability and Maintainability. Symposium. 1999 Proceedings (Cat. No. 99CH36283)*, pages 258–264. IEEE.

Jensen, W. A. (2015). Binomial reliability demonstration tests with dependent data. *Quality Engineering*, 27(2):253–266.

Jin, G. and Matthews, D. (2014). Reliability demonstration for long-life products based on degradation testing and a wiener process model. *IEEE Transactions on Reliability*, 63(3):781–797.

Jin, T. and Coit, D. W. (2008). Unbiased variance estimates for system reliability estimate using block decompositions. *IEEE Transactions on Reliability*, 57(3):458–464.

Kapur, K., Pecht, M., and Sage, A. (2014). Reliability engineering: Wiley series in systems engineering and management, 20 john wiley & sons. *Inc., Hoboken, New Jersey, USA*.

Kasprzak, E. M. and Lewis, K. E. (2001). Pareto analysis in multiobjective optimization using the collinearity theorem and scaling method. *Structural and Multidisciplinary Optimization*, 22(3):208–218.

Kececioglu, D. (2002). *Reliability engineering handbook*, volume 1. DEStech Publications, Inc.

Kececioglu, D. and Sun, F.-B. (2003). *Environmental Stress Screening: Its Quantification, Optimization and Management*. DEStech Publications, Inc.

Kleyner, A. (2008). Reliability demonstration in product validation testing. In *Handbook of Performability Engineering*, pages 533–542. Springer.

Kleyner, A., Bhagath, S., Gasparini, M., Robinson, J., and Bender, M. (1997). Bayesian techniques to reduce the sample size in automotive electronics attribute testing. *Microelectronics Reliability*, 37(6):879–883.

Kleyner, A., Elmore, D., and Boukai, B. (2015). A bayesian approach to determine test sample size requirements for reliability demonstration retesting after product design change. *Quality Engineering*, 27(3):289–295.

Kleyner, A. and Sandborn, P. (2008). Minimizing life cycle cost by managing product reliability via validation plan and warranty return cost. *International journal of production economics*, 112(2):796–807.

Kleyner, A., Sandborn, P., and Boyle, J. (2004). Minimization of life cycle costs through optimization of the validation program-a test sample size and warranty cost approach. In *Annual Symposium Reliability and Maintainability, 2004-RAMS*, pages 553–558. IEEE.

Krasich, M. (2004). Accelerated testing for demonstration of product lifetime reliability. *Journal of the IEST*, 47(2):48–58.

Krasich, M. (2007). Accelerated reliability growth testing and data analysis method. *Journal of the IEST*, 50(2):98–117.

Kuo, W. (1984). Reliability enhancement through optimal burn-in. *IEEE Transactions on Reliability*, 33(2):145–156.

Kuo, W. and Kuo, Y. (1983). Facing the headaches of early failures: a state-of-the-art review of burn-in decisions. *Proceedings of the IEEE*, 71(11):1257–1266.

Lawless, J. F. (1978). Confidence interval estimation for the weibull and extreme value distributions. *Technometrics*, 20(4):355–364.

Lee, Y.-L., Makam, S., McKelvey, S., and Lu, M.-W. (2015). Durability reliability demonstration test methods. *Procedia Engineering*, 133:31–59.

Leemis, L. M. and Beneke, M. (1990). Burn-in models and methods: a review. *IIE transactions*, 22(2):172–180.

Li, J., Shen, L., and Zhao, L. (2009). Success ratio sequential test plan using development test data. In *2009 8th International Conference on Reliability, Maintainability and Safety*, pages 370–373. IEEE.

Li, M., Zhang, W., Hu, Q., Guo, H., and Liu, J. (2016). Design and risk evaluation of reliability demonstration test for hierarchical systems with multilevel information aggregation. *IEEE Transactions on Reliability*, 66(1):135–147.

Lu, L., Anderson-Cook, C. M., and Robinson, T. J. (2011). Optimization of designed experiments based on multiple criteria utilizing a pareto frontier. *Technometrics*, 53(4):353–365.

Lu, L., Li, M., and Anderson-Cook, C. M. (2016). Multiple objective optimization in reliability demonstration tests. *Journal of Quality Technology*, 48(4):326–342.

Lu, M.-W. and Rudy, R. J. (2001). Reliability demonstration test for a finite population. *Quality and Reliability Engineering International*, 17(1):33–38.

Luo, W., Zhang, C.-h., Chen, X., and Tan, Y.-y. (2015). Accelerated reliability demonstration under competing failure modes. *Reliability Engineering & System Safety*, 136:75–84.

Martz Jr, H. and Waller, R. (1979). A bayesian zero-failure (baze) reliability demonstration testing procedure. *Journal of Quality Technology*, 11(3):128–138.

Mazumdar, M. (1977). An optimum procedure for component testing in the demonstration of series system reliability. *IEEE Transactions on Reliability*, 26(5):342–345.

McKane, S. W., Escobar, L. A., and Meeker, W. Q. (2005). Sample size and number of failure requirements for demonstration tests with log-location-scale distributions and failure censoring. *Technometrics*, 47(2):182–190.

Meeker, W. Q. and Escobar, L. A. (2014). *Statistical methods for reliability data*. John Wiley & Sons.

Meeker, W. Q., Escobar, L. A., and Lu, C. J. (1998). Accelerated degradation tests: modeling and analysis. *Technometrics*, 40(2):89–99.

Meeker, W. Q. and LuValle, M. J. (1995). An accelerated life test model based on reliability kinetics. *Technometrics*, 37(2):133–146.

Meeker, W. Q. and Nelson, W. (1977). Weibull variances and confidence limits by maximum likelihood for singly censored data. *Technometrics*, 19(4):473–476.

Messac, A., Puemi-Sukam, C., and Melachrinoudis, E. (2000). Aggregate objective functions and pareto frontiers: required relationships and practical implications. *Optimization and Engineering*, 1(2):171–188.

Murthy, D. (2006). Product warranty and reliability. *Annals of Operations Research*, 143(1):133–146.

Murthy, D. P. and Blischke, W. R. (2006). *Warranty management and product manufacture*. Springer Science & Business Media.

Nelson, W. (1980). Accelerated life testing-step-stress models and data analyses. *IEEE transactions on reliability*, 29(2):103–108.

Nelson, W. B. (2005). A bibliography of accelerated test plans. *IEEE Transactions on Reliability*, 54(2):194–197.

Nelson, W. B. (2009). *Accelerated testing: statistical models, test plans, and data analysis*, volume 344. John Wiley & Sons.

O'Connor, P. and Kleyner, A. (2012). *Practical reliability engineering*. John Wiley & Sons.

Pham, T. and Turkkan, N. (1992). Bayes binomial sampling by attributes with a general-beta prior distribution. *IEEE transactions on reliability*, 41(2):310–316.

Pintar, A., Lu, L., Anderson-Cook, C. M., and Silver, G. L. (2012). Bayesian estimation of reliability for batches of high reliability single-use parts. *Quality Engineering*, 24(4):473–485.

Pohl, E. A. and Dietrich, D. L. (1995). Environmental stress screening strategies for multi-component systems with weibull failure-times and imperfect failure detection. In *Annual Reliability and Maintainability Symposium 1995 Proceedings*, pages 223–232. IEEE.

Porter, A. (2004). Accelerated reliability qualification in automotive testing. *Quality and Reliability Engineering International*, 20(2):115–120.

Ramírez-Márquez, J. E. and Jiang, W. (2006). On improved confidence bounds for system reliability. *IEEE Transactions on Reliability*, 55(1):26–36.

Reddy, R. K. and Dietrich, D. L. (1994). A 2-level environmental-stress-screening (ess) model: a mixed-distribution approach. *IEEE Transactions on Reliability*, 43(1):85–90.

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Rodríguez-Narciso, S. and Christen, J. A. (2016). Optimal sequential bayesian analysis for degradation tests. *Lifetime data analysis*, 22(3):405–428.

Roocks, P. (2016). Computing pareto frontiers and database preferences with the rpref package. *RJ*, 8(2):393–404.

Sandoh, H. (1991). Reliability demonstration testing for software. *IEEE Transactions on Reliability*, 40(1):117–119.

Sands, A., T. V. (2009). 1 in 3 laptops fail over 3 years: Netbooks fail 20% more than laptops; asus & toshiba the most reliable.

Sarson-Lawrence, J., Sabatini, R., Clothier, R., and Gardi, A. (2014). Experimental determination of low-cost servomotor reliability for small unmanned aircraft applications. In *Applied Mechanics and Materials*, volume 629, pages 202–207. Trans Tech Publ.

Siegmund, D. (2013). *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media.

Smith, D. J. (2017). *Reliability, maintainability and risk: practical methods for engineers*. Butterworth-Heinemann.

Sun, D. and Berger, J. O. (1994). Bayesian sequential reliability for weibull and related distributions. *Annals of the Institute of Statistical Mathematics*, 46(2):221–249.

Sun, Q., Zhang, Z., Feng, J., and Pan, Z. (2012). A zero-failure reliability demonstration approach based on degradation data. In *2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, pages 947–952. IEEE.

Tal, O., McCollin, C., and Bendell, T. (2001). Reliability demonstration for safety-critical systems. *IEEE Transactions on Reliability*, 50(2):194–203.

Tam, S. M. (1995). Demonstrated reliability of plastic-encapsulated microcircuits for missile applications. *IEEE Transactions on Reliability*, 44(1):8–13.

Ten, L. M. and Xie, M. (1998). Bayes reliability demonstration test plan for series-systems with binomial subsystem data. In *Annual Reliability and Maintainability Symposium. 1998 Proceedings. International Symposium on Product Quality and Integrity*, pages 241–246. IEEE.

Tillman, F. A., Kuo, W., Hwang, C., and Grosh, D. L. (1982). Bayesian reliability & availability-a review. *IEEE Transactions on Reliability*, 31(4):362–372.

Tobias, P. A. and Trindade, D. (2011). *Applied reliability*. Chapman and Hall/CRC.

Tsai, C.-C., Tseng, S.-T., and Balakrishnan, N. (2010). Optimal burn-in policy for highly reliable products using gamma degradation process. *IEEE Transactions on Reliability*, 60(1):234–245.

Wang, C. J. (1991). Sample size determination of bogey tests without failures. *Quality and reliability engineering international*, 7(1):35–38.

Wang, P. and Coit, D. W. (2004). Reliability prediction based on degradation modeling for systems with multiple degradation measures. In *Annual Symposium Reliability and Maintainability, 2004-RAMS*, pages 302–307. IEEE.

Wasserman, G. (2002). *Reliability verification, testing, and analysis in engineering design*, volume 153. CRC Press.

Weaver, B. P. and Hamada, M. S. (2008). A bayesian approach to the analysis of industrial experiments: An illustration with binomial count data. *Quality Engineering*, 20(3):269–280.

Wilson, A. G. and Fronczyk, K. M. (2017). Bayesian reliability: Combining information. *Quality Engineering*, 29(1):119–129.

Wong, K. L. (1990). Demonstrating reliability and reliability growth with environmental stress screening data. In *Annual Proceedings on Reliability and Maintainability Symposium*, pages 47–52. IEEE.

Xu, J., Yu, D., Xie, M., and Hu, Q. (2017). An approach for reliability demonstration test based on power-law growth model. *Quality and Reliability Engineering International*, 33(8):1719–1730.

Yang, G. (2002). Environmental-stress-screening using degradation measurements. *IEEE Transactions on Reliability*, 51(3):288–293.

Yang, G. (2007). *Life cycle reliability engineering*. John Wiley & Sons.

Yang, G. (2009). Reliability demonstration through degradation bogey testing. *IEEE transactions on Reliability*, 58(4):604–610.

Yang, G. (2013). Heuristic degradation test plans for reliability demonstration. *IEEE Transactions on Reliability*, 62(1):305–311.

Yang, J. and Sun, F.-B. (1999). A comprehensive review of hard-disk drive reliability. In *Annual Reliability and Maintainability. Symposium. 1999 Proceedings (Cat. No. 99CH36283)*, pages 403–409. IEEE.

Ye, Z.-S., Shen, Y., and Xie, M. (2012). Degradation-based burn-in with preventive maintenance. *European Journal of Operational Research*, 221(2):360–367.

Zhang, C., Lu, X., Tan, Y., and Wang, Y. (2015). Reliability demonstration methodology for products with gamma process by optimal accelerated degradation testing. *Reliability Engineering & System Safety*, 142:369–377.

Appendix A Copyright Permissions

A.1 Permission for Chapter 2

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Sign in | Create Account

Multi-state reliability demonstration tests

Author: Suiyao Chen, , Lu Lu, et al

Publication: Quality Engineering

Publisher: Taylor & Francis

Date: Jul 3, 2017

Rights managed by Taylor & Francis

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

BACK CLOSE

Appendix B Supplemental Materials

B.1 Appendix for Chapter 2

To analytically show the difference between scenarios I and II in the proposed MSRDTs over multiple time periods, let $\Delta H(n, c_1, c_2) = H_{\mathrm{I}}(n, c_1, c_2) - H_{\mathrm{II}}(n, c_1, c_2)$ when $c_1 + c_2$ is fixed, which can be explicitly written as

$$\Delta H(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=c_2+1}^{c_1-y_1} [(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!}) \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n-y_1-y_2}].$$

When $c_1 = 0$, $\Delta H(n, c_1, c_2) = 0$ and both scenarios become equivalent, as shown in Table 2.3-2.5. When $c_1 > 0$, $\Delta H(n, c_1, c_2) > 0$, which indicates that the probability of accepting test plan under scenario II is always smaller than the probability calculated under scenario I. However, this finding does not imply that for a fixed n, one scenario will always give a consistently higher/lower CR than the other. To justify this, let $A = \int_0^{1-R_1} \int_0^{1-R_2-\pi_1} H_{\mathrm{II}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$ and $B = \int_0^1 \int_0^1 H_{\mathrm{II}}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$, $\mathrm{CR}_{\mathrm{II}}$ and CR_{I} can be written as

$$\mathrm{CR}_{\mathrm{II}} = 1 - \frac{A}{B},$$

$$\mathrm{CR}_{\mathrm{I}} = 1 - \frac{A + \Delta A}{B + \Delta B},$$

where we can denote $\Delta A = \int_0^{1-R_1}\int_0^{1-R_2-\pi_1} \Delta H(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$ and $\Delta B = \int_0^1 \int_0^1 \Delta H(n, c_1, c_2) p(\pi_1, \pi_2)\, d\pi_2 d\pi_1$. Then $\mathrm{CR_{II}} - \mathrm{CR_I}$ is given by

$$\mathrm{CR_{II}} - \mathrm{CR_I} = \frac{B\Delta A - A\Delta B}{B(B + \Delta B)}.$$

Although $B > A$, as n, c_1 and c_2 vary, ΔA can be larger/smaller than ΔB. Thus, for a fixed sample size n, neither $\mathrm{CR_{II}} > \mathrm{CR_I}$ nor $\mathrm{CR_{II}} < \mathrm{CR_I}$ will hold consistently. It also explains results in Figure 2.10, Table 2.4 and 2.5 that when controlling CR, one scenario cannot give a consistently larger/smaller minimum sample size than the other scenario.

B.2 Appendix for Chapter 4

Derivations of expected failure probabilities in the corresponding warranty service periods, illustrated by two periods, can be derived as,

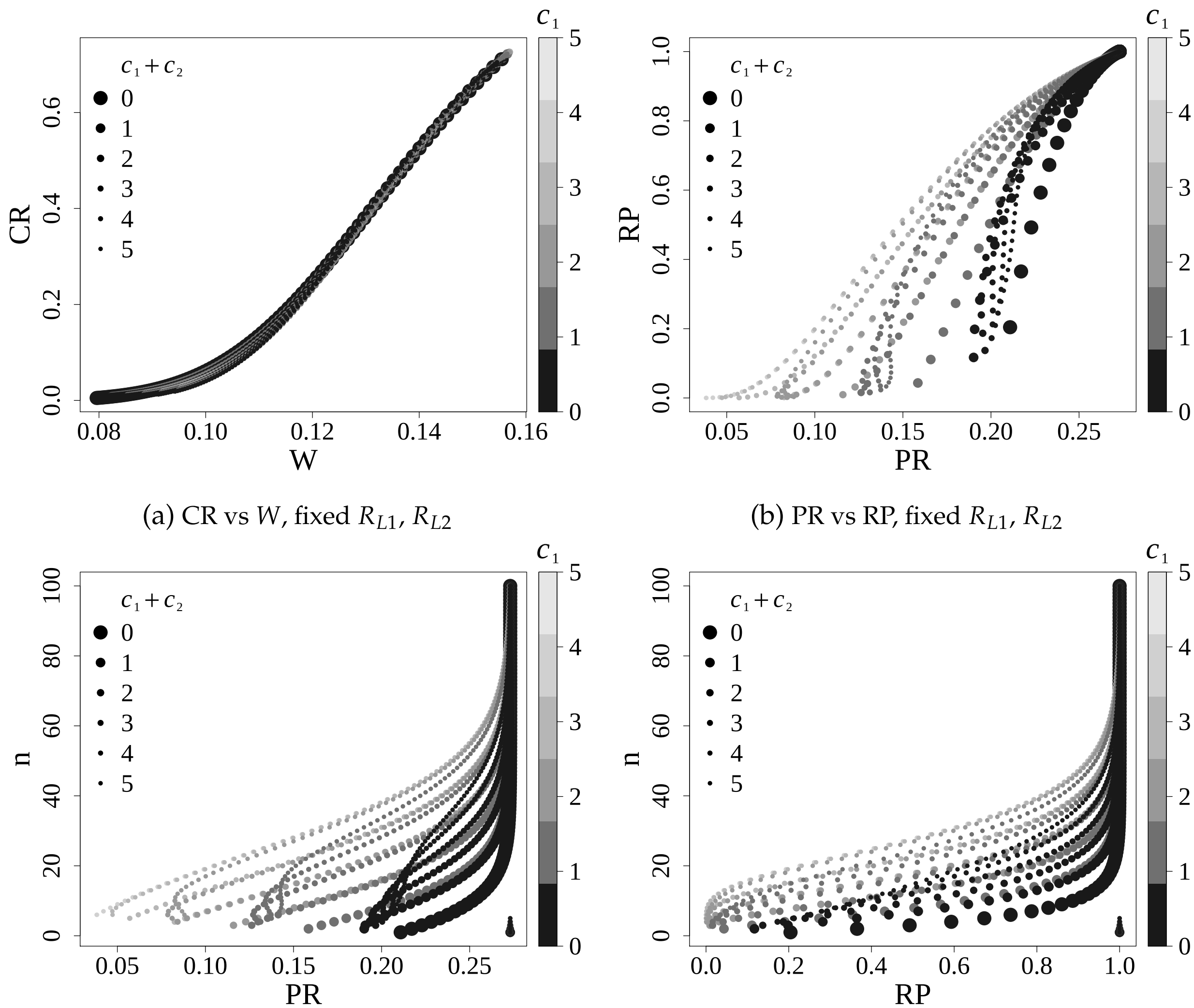
$$\begin{aligned}
&E(\pi_1 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) \\
&= \int_0^1 \pi_1 f(\pi_1 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) d\pi_1 \\
&= \int_0^1 \pi_1 [\int_0^1 f(\pi_1, \pi_2 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) d\pi_2] d\pi_1 \\
&= \int_0^1 \pi_1 [\int_0^1 \frac{p(y_1 \le c_1, y_1 + y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2)}{\int_0^1 \int_0^1 p(y_1 \le c_1, y_1 + y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2} d\pi_1] \\
&= \frac{\int_0^1 \int_0^1 \pi_1 p(y_1 \le c_1, y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}{\int_0^1 \int_0^1 p(y_1 \le c_1, y_1 + y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}; \\
&p(y_1 \le c_1, y_2 \le c_1 + c_2 | \pi_1, \pi_2) = H_{\mathrm{MS}}(n, c_1, c_2)
\end{aligned}$$

$$\begin{aligned}
&E(\pi_2 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) \\
&= \int_0^1 \pi_2 f(\pi_2 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) d\pi_2 \\
&= \int_0^1 \pi_2 [\int_0^1 f(\pi_1, \pi_2 | y_1 \le c_1, y_1 + y_2 \le c_1 + c_2) d\pi_1] d\pi_2
\end{aligned}$$

$$= \int_0^1 \pi_2 [\int_0^1 \frac{p(y_1 \le c_1, y_1 + y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2)}{\int_0^1 \int_0^1 p(y_1 \le c_1, y_1 + y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2} d\pi_2]$$

$$= \frac{\int_0^1 \int_0^1 \pi_2 p(y_1 \le c_1, y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2}{\int_0^1 \int_0^1 p(y_1 \le c_1, y_1 + y_2 \le c_1 + c_2 | \pi_1, \pi_2) p(\pi_1, \pi_2) d\pi_1 d\pi_2};$$

$$p(y_1 \le c_1, y_2 \le c_1 + c_2 | \pi_1, \pi_2) = H_{\mathrm{MS}}(n, c_1, c_2)$$



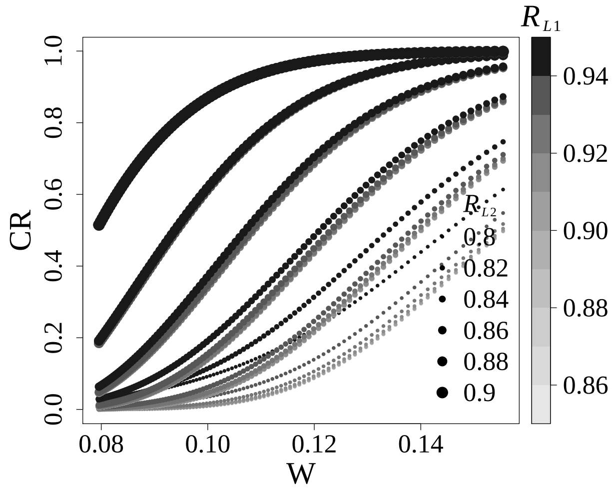
(a) CR vs W, fixed R_{L1}, R_{L2}

(b) PR vs RP, fixed R_{L1}, R_{L2}

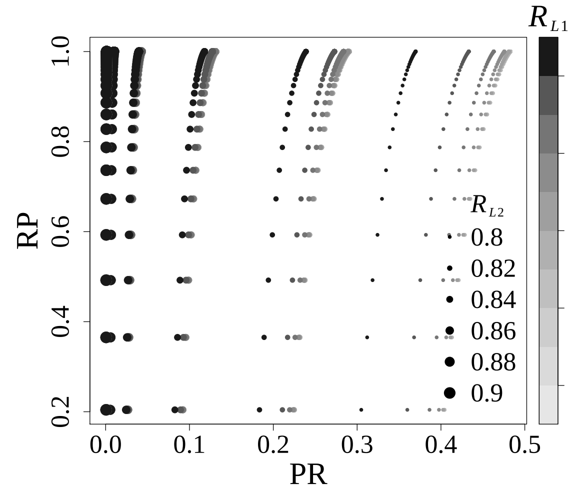
(c) RP vs n, fixed R_{L1}, R_{L2}

(d) RP vs n, fixed R_{L1}, R_{L2}

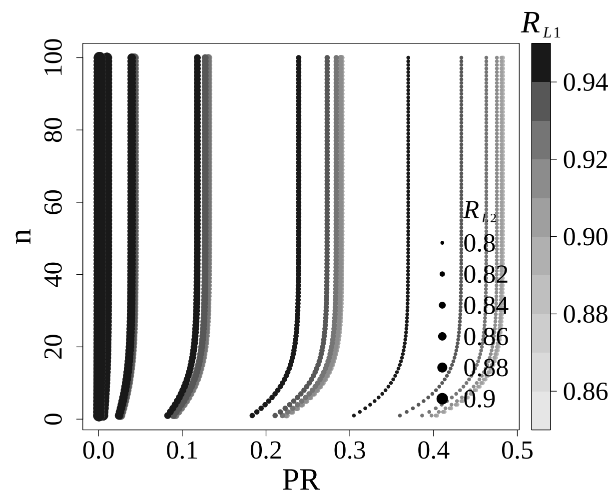
Figure B.1: No obvious trade-off patterns for selected pairs of objectives with fixed lower level reliability requirement



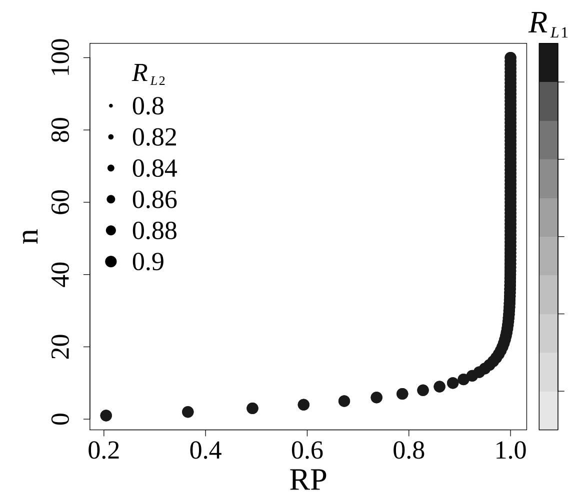
(a) CR vs W, fixed c_1, c_2



(b) PR vs RP, fixed c_1, c_2



(c) RP vs n, fixed c_1, c_2



(d) RP vs n, fixed c_1, c_2

Figure B.2: No obvious trade-off patterns for selected pairs of objectives with fixed maximum allowable failures