ANALYTICS APPROACHES TO IMPROVE STRATEGIC, OPERATIONAL, AND CLINICAL DECISION-MAKING IN HEALTHCARE

A Dissertation
Presented to
The Academic Faculty

By

Caglar Caglayan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology

August 2019

ANALYTICS APPROACHES TO IMPROVE STRATEGIC, OPERATIONAL, AND CLINICAL DECISION-MAKING IN HEALTHCARE

Approved by:

Dr. Turgay Ayer, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Pinar Keskinocak
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paul Griffin
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Siva Theja Maguluri
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Mustafa Y. Sir
Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery
Mayo Clinic

Dr. Yunan Liu
Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University

Date Approved: May 24, 2019

Bilmek, bilmenin ne olduğuna bilmektir,

Bilmek, kendini tanımaktır.

Sen ne olduğunu bilmez isen,

Bu nasıl okumaktır?

To know is to know what knowing is,

To know is to know what one's self is.

If you know do not know who you are,

(Tell me) what kind of knowing this is?

Yunus Emre

I want to dedicate this thesis to my wife Melissa E. Strait, and my parents Gúlúmser Yılmam and Fatih Çağlayan. No words can express my gratitude for their unconditional love and support, which made this thesis and any other accomplishment I have in life possible.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my thesis advisor Dr. Turgay Ayer for his personal and professional support, guidance, and mentorship during my doctoral studies. In addition, I want to acknowledge Dr. Mustafa Sir, Dr. Yunan Liu, and Dr. Christopher Flowers for their contributions to my learning and development. Finally, I want to thank to the members of my thesis committee Dr. Pinar Keskinocak, Dr. Paul Griffin, and Dr. Siva Theja Maguluri.

A manuscript related to Chapter 1 was submitted to the journal *Operations Research* and is currently under revision. The name of the manuscript that is under revision is "Assessing Multi-Modality Breast Cancer Screening Strategies for BRCA 1/2 Gene Mutation Carriers and Other High-Risk Populations". A manuscript related to Chapter 2 was submitted to the journal *Management Science*. The name of the submitted manuscript is "Physician Staffing in Emergency Rooms (ERs): Opening the Black-box of ER Care via a Multi-Class Multi-Stage Network". A manuscript related to Chapter 3 was published by the journal *Cancer*. The name of the published manuscript is "A population-based multistate model for diffuse large B-cell lymphoma-specific mortality in older patients". A manuscript related to Chapter 3 was published by the journal *Clinical Lymphoma, Myeloma & Leukemia*. The name of the published manuscript is "Assessing the Effectiveness of Treatment Sequences for Older Patients With High-risk Follicular Lymphoma With a Multistate Model".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis presents three important and complex medical decision-making problems Çağlar Çağlayan studied during his doctoral studies, describes the analytical methods he utilized and developed, and discusses the methodological and numerical findings and contributions of his work. The works presented in this thesis make contributions to three research topics on clinical decision-making under uncertainty: (i) the development of an optimal multi-modality screening program for women at high-risk for breast cancer, (ii) the determination of optimal physician staffing levels in emergency department under time-varying arrivals, and (iii) the study of the clinical course of follicular and diffuse large B cell lymphomas with the goal of improving treatment outcomes.

In Chapter 1, we study a multi-modality breast cancer screening problem for high-risk population and identify optimal and cost-effective population screening strategies based on the imaging technologies that are in widespread use. Women with certain risk factors such as BRCA 1/2 gene mutations and family history of breast or ovarian cancer are at significantly higher risk for breast cancer. For these high-risk women, the existing guidelines recommend intensified screening starting at an early age, where the use of ultrasound (US) and magnetic resonance imaging (MRI) might address some of the limitations of mammography, the standard screening modality for average-risk women. Yet, the cost and false positive rates of MRI, and the operator dependency of US raise concerns. Currently, there is no consensus on the optimal use of these technologies in conjunction with, or instead of, mammography in high-risk women. To study this problem, we develop a Markov model to capture the disease incidence and progression in high-risk women, and formulate a mixed integer linear program to identify the optimal structured strategies that are practical for implementation. We further study the structure of the optimal strategies, and establish the conditions under which a strategy with more frequent but less sensitive screens yields higher health benefits than a strategy with more sensitive but less frequent screens. Our re-

sults show that (1) for young women, annual screening with ultrasound, is affordable with moderate budgets, and optimal over a wide range of budget levels despite its high operator dependency, (2) for middle-aged women, annual mammography screening is robustly optimal and cost-effective, and (3) the use of MRI, alone or combined with mammogram, leads to outcomes that are not cost-effective.

In Chapter 2, we study a physician staffing and an associated patient routing problem in emergency rooms (ERs) coping with time-varying demand. ERs are complex healthcare delivery systems, characterized by time-varying unscheduled arrivals, medium-to-long service times, high patient volumes, multiple patient classes, and multiple treatment stages. In such a complex system, optimizing the staffing levels of physicians, the most critical resources in ERs, is a major challenge. In this work, we study a staffing problem for ER physicians, and propose a new staffing algorithm that determines the optimum staffing levels stabilizing differentiated tail probability of delay (TPoD) type service targets. Taking a queuing theory approach, we develop a practical and intuitive multi-class multi-stage queuing network describing the ER care delivery as sequences of treatments and order bundles (i.e., groups of diagnostic medical processes). Employing this model, we capture time-varying patient flow in the ER and estimate its load on treatment stations, served by physicians. Treatment queues operate in efficiency-driven regime but experience negligible abandonment as abandonments nearly always occur at earlier stages of the ER care. This observation motivates our proposed new staffing algorithm, which translates the offered load into staffing decisions for efficiency driven queues with perfectly patient customers and TPoD type targets. We analytically show the asymptotic effectiveness of our staffing algorithm for M/M/s queues that operate in efficiency-driven mode. Then, we demonstrate its robustness via realistic and data-driven simulation experiments in various time-varying ER settings, considering non-homogeneous Poisson arrivals, multiple patient classes, multi-stage service, and centralized (pooled) (physicians) under several practical routing rules. Our results show that (1) our proposed staffing approach is effective and

robust for optimizing the ER physician staffing levels in various ER settings, and (2) as the service complexity of an ER increases, the use of dynamic rules, using the current system state for routing decisions, and hybrid policies, combining pre-determined static routing rules with dynamic ones, become necessary to stabilize TPoD targets.

In Chapter 3, we study the clinical course of two types of lymph node cancers, follicular lymphoma (FL) and diffuse large B cell lymphoma (DLBCL). These cancers have different characteristics, where DLBCL is aggressive and FL is recurrent, and have multiple clinical intermediate- or end-points such as the sequence of treatments or cause-specific death. Accordingly, we develop two different continuous-time, multi-state survival analysis models to investigate the clinical course of these diseases following initial treatment with the goal of improving treatment outcomes. We utilize Cox proportional hazards models to specify the impact of prognostic factors on overall survival and cause-specific deaths, and the Aalen-Johansen estimator to project the course of DLBCL over time. In particular, employing the multi-state FL model, we investigate the clinical course of FL under first, second and third line therapies for high-risk patients to assess the effectiveness of various treatment sequences. Our analysis shows that single R-CHOP therapy in any line of treatment improves overall survival for high-risk patients, achieving the most favorable outcome when provided as first-line therapy, but its multiple use for first- and second-line might lead to adverse outcomes. Using the DLBCL model, we examine the role of clinical and socio-demographic factors on DLBCL-associated mortality in the elderly population and identify a cutoff point to stop monitoring DLBCL patients receiving the standard R-CHOP therapy. Utilizing a large population-based dataset, our analysis (1) identifies age, sex, and Charlson comorbidity index as risk factors for DLBCL-specific and other causes of death, and (2) confirms a 5-year cure point for older patients receiving R-CHOP therapy, suggesting to transition survivorship surveillance plans from a focus on lymphoma recurrence-related deaths to non-cancer risks at five years after treatment.

In Chapter 4, we summarize our studies with their contributions to their corresponding

research topics, and conclude the thesis.

CHAPTER 1
OPTIMAL BREAST CANCER SCREENING STRATEGIES FOR WOMEN AT HIGH-RISK FOR BREAST CANCER

1.1 Introduction

Breast cancer is the most common noncutaneous cancer in U.S. women, with an estimated 252,710 new cases of invasive cancer and 63,410 new cases of in situ cancer in 2017. About one in eight U.S. women is projected to ultimately develop breast cancer, leading to more cancer deaths in women than any other cancer after lung cancer, with about 40,000 deaths per year [1].

Some women are at significantly higher risk of developing breast cancer depending on the existence of certain risk factors, including inherited breast cancer gene mutations of BRCA1 or BRCA2, and a personal or family history of breast or ovarian cancer. For example, compared with general population, BRCA 1 mutations carriers are at least four times more likely to develop breast cancer whereas BRCA 2 mutations increase breast cancer risk about three to five times. About 44% to 78% of women with BRCA1 mutations and 31% to 56% of women with BRCA2 mutations are projected to develop breast cancer by age 70, accounting for up to 10% of all breast cancers together [2, 3, 4]. Similarly, breast cancer risk is almost two times higher for women with one-first degree female relative diagnosed with breast cancer, about three times higher for women with two relatives, and nearly four times higher with three or more relatives [5].

Genetic counseling and testing for BRCA mutations can identify such high-risk individuals, and subsequently lead to an improvement in their health outcomes. Although historically such services have been underutilized primarily due to their high-costs [6, 7], with the recent rapid declines in the prices for genetic testing, it is expected that more

women would have access to such services, leading to an increase in the prevalence of detected gene mutation carriers [8]. Indeed, in the past few years, the price of BRCA tests has dropped from several thousands to below $250 [9, 10], which is identified as a cut-off point for a cost-effective population-based testing [11]. Combined with the increasing awareness about gene mutation related cancers, an increasing portion of BRCA gene mutation carrier women are being identified over years, leading to a pressing need for long term management strategies specifically designed for this high-risk subpopulation [12, 13].

Women at high risk of breast cancer may opt for bilateral prophylactic mastectomy, a preventive surgical operation removing both breasts before cancer develops. However, perceived mutilating effects of mastectomy make the decision for surgical prevention difficult [14]. Furthermore, it is not currently recommended by national policy organizations as the standard management strategy and insurance coverage is not guaranteed under federal law [15]. As a less invasive intervention, high-risk women might be offered chemoprevention; however, adherence to these chemoprevention drugs are typically very low [16] and most intervention studies designed to increase the uptake have yielded disappointing results (see [17] and references therein). Instead, intensified screening with an early starting age such as 25 or 30, followed by a subsequent biopsy if an abnormality is detected, is the recommended strategy and the current standard for so-far healthy high-risk women [18, 19, 20].

Among the several screening modalities, mammography is the recommended and most commonly used breast cancer screening technology for average-risk women [21], and most insurers in the U.S. are required to cover it for women at ages 50-74 without a co-pay or deductible. However, the accuracy of mammography in high risk populations, particularly in women with a genetic predisposition and young high-risk women, is not as satisfactory [22, 23, 24]. Accordingly, the use of other screening technologies, such as magnetic resonance imaging (MRI) and breast ultrasound, are encouraged to address the limitations of mammography and enhance cancer detection in these populations (see [25, 26] and references

therein). Breast MRI has been shown to offer additional health benefits for women at high risk due to its higher sensitivity of detecting early breast cancer [23, 27, 28]. Additionally, a number of studies showed that ultrasound usually has higher sensitivity than mammogram in young high-risk women [22, 25, 29]. Based on these findings, the American Cancer Society (ACS) and the American College of Radiology (ACR) have added annual MRI screening to their guidelines for screening women at high risk of breast cancer and the ACR guidelines have also recommended the whole-breast ultrasound as an alternative for high-risk women unable to tolerate MRI [30, 19].

Despite its higher sensitivity and the potential for detecting more cancer cases, MRI is an expensive modality, approximately five times more costly than mammography [31, 32]. Furthermore, it yields substantially higher number of false-positives (i.e. positive screening outcomes when the disease is absent) due to its lower specificity, causing more unnecessary diagnostic follow-ups such as biopsy, and leading to increased screening and diagnostic costs [33, 29, 25]. On the other hand, while ultrasound is much cheaper than MRI, it generally has lower sensitivity than MRI. Further, ultrasound is an operator-dependent technology with substantially higher variation in performance among operators, compared with mammography and MRI [34, 35]. This intra-observer variability is important and is a concern in national guideline development [33, 36], as it affects the number of resulting false-positive readings and hence, the total cost of a screening strategy.

Notwithstanding the strong evidence indicating that the employment of non mammographic technologies may be beneficial for women at high-risk in detecting breast cancer earlier, currently there is no consensus on the optimal use of these technologies in conjunction or instead of mammography [37, 38]. While most guidelines recommend the use of MRI and/or ultrasound in addition to mammograms for use in high-risk women, the current guidelines are neither always specific, nor in agreement with each other about which modality or combination of modalities to use, when and with what frequency. For example, while National Comprehensive Cancer Network (NCCN) recommends starting annual mammog-

raphy and breast MRI as early as at age 25 and 30, respectively, for high-risk women with lifetime breast cancer risk $\geq 20\%$ (estimation of which largely depends on the existence of a family history), the U.S. Preventive Services Task Force (USPSTF) recommends the use of mammography starting at age 40 for high-risk women with family history. On the other hand, the ACS recommends joint annual MRI and mammogram strategy starting at age 30 for high-risk women with a lifetime risk greater $\geq$ 20-25$\%$, with a known BRCA mutation, or with a first-degree relative of BRCA carriers [18]. Furthermore, none of the existing guidelines explicitly specify whether at any age screening frequency should be reduced and ultimately terminated. The lack of uniformity and clarity in the guidelines result in lower adherence to the guideline-specific recommendations among healthcare providers and hence, lead to inconsistency in the practice. Further, comprehensive randomized controlled trials (RCTs), the gold-standard in such decisions, are unlikely to be performed because of the associated large number of patients needed to demonstrate a difference between groups, the long length of follow-up required, and high financial costs incurred [39].

Designing screening strategies that balance the tradeoff between detecting as many cancers as early as possible and minimizing the harms of false-diagnosis, while keeping the overall costs of screening and diagnosis at an affordable range is essential for improving patient outcomes in an cost-effective way. Cost containment has become more critical especially in recent years as the current cost burden of cancer prevention and control is substantial and growing [40, 41], and employed imaging technologies are one of the the main drivers of the increased cost [40, 42]. Accordingly, especially in the midst of an environment of rising health-care costs [40, 41, 43], it is essential to consider the associated medical cost up front, in parallel with efforts to determine improved patient outcomes while developing screening strategies [39].

The complexity and importance of multi-modality breast cancer screening problem cannot be overestimated as the debate and literature around the optimal use of different screening modalities and strategies continue to grow, and the associated cost burden has

become a central concern. Identifying effective multi-modality screening strategies for high-risk women may help decision-makers in developing evidence-based policies, health-care providers in better defining the role of these tests in breast cancer screening, and insurance companies in making coverage decisions. Given the lack of comprehensive RCTs, data-driven mathematical modeling and computational optimization can be promising in conducting comparative effectiveness and cost-effectiveness of multimodality breast cancer screening strategies for women at high risk of breast cancer.

In this study, we study and propose a population-dynamics based modeling approach for the multi-modality breast cancer screening problem in high-risk population to (1) identify the optimal structured screening strategies that are easy for implementation for women at high lifetime risk of developing breast cancer, (2) investigate the impact of resource restrictions and assess the cost-effectiveness of various screening strategies, and (3) gain insight into the underlying structure of the optimal strategies. Adapting a societal perspective, we first develop a discrete-time finite-horizon Markov model to capture age-dependent disease incidence and progression in high-risk population under various screening interventions. We parameterize, calibrate and validate this disease progression model using real data and the best available evidence from the clinical literature. Then, we formulate and solve a mixed integer linear program (MILP) to identify the optimal structured screening strategies, maximizing quality-adjusted life years (QALYs) at various budget levels. We assess the cost-effectiveness of the identified screening strategies with the incremental cost effectiveness ratio (ICER) metric and the role of ultrasound's operator-dependency via a worst-case scenario analysis. Finally, we analyze the structure of optimal screening strategies and establish the sufficiency conditions under which a strategy with more frequent screens yields higher health benefits than the one using a more sensitive modality.

Our main findings shed light on several controversial health policy questions, which can be summarized as follows: (1) Annual ten-year screening with ultrasound alone is affordable with moderate budget, optimal over a wide range of budget levels and cost-effective

in young (aged 25-44 years) high-risk women. Furthermore, these findings continue to hold across a wide range of specificity level for ultrasound, indicating that intra-observer variability in ultrasound reading may not be a serious concern in guideline development for young population groups; (2) for middle-aged (45-74 years) high-risk women, annual mammography screening alone is robustly optimal over a wide budget range and is cost-effective, whereas ultrasound may be an alternative at low to medium budget levels; (3) use of MRI, as recommended by some of the major national guidelines, may lead to cost-ineffective strategies, despite the fact that joint annual MRI and mammogram screening strategy yields the highest health benefits for younger and middle-aged high-risk women.

The rest of the chapter is organized as follows. Section 1.2 summarizes the relevant literature. In Section 1.3, we present a discrete-time Markov model capturing disease progression under various interventions, and propose an optimization model to identify the optimal breast cancer screening strategies for high risk women. In Section 1.4, we present our analytical results. In Section 1.5, we describe model inputs and parameter estimations. In Section 1.6, we present the numerical results and discuss their implications. Finally, in Section 1.7, we discuss our findings and conclude our study.

1.2 Literature Review

There are two main streams of research within the modeling-based disease screening literature. At the one end of the spectrum, there are cost-effectiveness microsimulation models based on clinical trials and at the other hand, there are mathematical modeling based studies. In Table 1.1, we summarize the existing cost-effectiveness studies on breast cancer screening for high-risk population that are relevant to our study. Our key distinction from these simulation-based cost-effectiveness studies is that while they evaluate a limited (and often non-overlapping) predetermined set of strategies, we comprehensively consider a large set of strategies, including but not limited to all of the different strategies proposed and considered by these simulation-based studies.

Table 1.1: Simulation-based Studies on Breast Cancer Screening for High-risk Population

Authors	Year	Examined Policies	Population	Screening Horizon	Screening Interval	Results and Conclusions
[48]	2006	MAM and MAM+MRI	25 year-old BRCA 1/2 mutation carriers	45 years	1 year	MRI is more cost-effective for BRCA 1 mutation carriers (comparing to BRCA 2 mutation carriers)
[49]	2006	MAM, MRI and MAM+MRI	35-49 years old high-risk population	2-7 years	1 year	MRI might be cost-effective for BRCA 1 or 2 mutation carriers.
[50]	2009	MRI, (X-ray) MAM and MRI+(X-ray) MAM	High-risk population	unknown (no access)	unknown (no access)	MRI is cost-effective for BRCA 1 or 2 mutation carriers.
[51]	2009	Annual MAM versus MRI	Young high-risk population	over 25 years	1 year	Annual MRI is not cost-effective even when WTP/QALY = $ 120,000.
[19]	2010	Annual MAM, Annual MRI and Annual MAM+MRI	25 year-old BRCA 1 mutation carriers	over 45 years	1 year	Annual combined screening is cost-effective both when WTP/QALY = $ 50,000 and WTP/QALY = $ 100,000
[32]	2013	MAM, MRI and CBE (alone, alternating, combined)	Women with familial risk	10 years, 15 years, 25 years	6 months - 2 years	MRI may improve survival but is expensive especially in youngest age group.
[52]	2013	digital MAM+MRI (alternating)	25-year old BRCA1/2 mutation carriers	from age 25 until death (over 40 years)	6 months and 1 year	Alternating MAM and MRI screening in every 6 months is the most cost-effective.
[53]	2013	MAM and MAM+MRI	25-year old BRCA 1/2 mutation carriers	40 years (between 25 and 65 years old)	1 year	ICER of MAM+MRI (comparing to annual MAM alone) is ∼$50,000/QALY.
[54]	2013	Dutch, U.K. and U.S. guidelines	BRCA 1/2 mutation carriers	over 50 years	1 year, 2 years and 3 years	Dutch screening is the most cost-effective.

It is noteworthy to emphasize that microsimulation, despite being utilized very commonly in most cancer settings, is not the only methodology proposed in the literature for assessing effectiveness and/or cost-effectiveness of interventions and there are also several other approaches. For instance, as an alternative to the conventional simulation-based approaches, [44] develop a simple and fast one-pass algorithm that is especially useful when there are many scenarios to be simulated or an extensive suite of sensitivity analyses are required to identify the efficient frontier. Based on continuous-time multi-armed bandit model, [45] introduce another framework for adaptive and personalized treatments of certain chronic diseases that can capture the trade-off between the health benefits and costs of medical interventions and hence, can be utilized to generate an entire frontier of cost-effective strategies. [46] and [47] both employ decision-analytic models for cost-effectiveness analysis of a chronic Hepatitis C screening and a more comprehensive newborn screening, respectively.

At the other end of the spectrum, there are studies that focus on different aspects of disease screening problems using analytical approaches. For older studies, we refer the reader

to [55] for an extensive review of modeling based cancer screening studies, and below summarize and discuss the more recent studies. Taking a population-based approach, [56] study mammography screening problem for the average-risk population and employ a partially observable discrete-time finite-horizon Markov chain that captures several age-based dynamics of breast cancer. They evaluate various mammographic screening scenarios by providing an upper bound on the lifetime breast cancer mortality risk for average-risk patients for each considered scenario and report a frontier of efficient strategies as measured by lifetime mortality risk and expected number of mammograms. [57] develop a generic model to capture disease progression and utilize a metaheuristic-based optimization technique to investigate Pareto-optimal breast cancer screening strategies over a ten-year time horizon with annual or total budget constraints. [58] consider individualized mammography screening decisions and use a partially observable Markov decision process (POMDP) that considers some other personal risk characteristics in addition to age and screening history of each patient. They show that personalized screening strategies outperform the existing guidelines with respect to the total expected QALYs, while significantly decreasing the number of mammograms and false-positives. [59] extend [58] and study mammography screening problem under resource constraints. They propose a constrained POMDP maximizing total expected QALYs of patients when they are allowed only a limited number of mammography screenings and find that efficient distribution of available resources between patients with different risk levels leads to significant QALYs gains, especially for the patients with higher breast cancer risk. [60] and [61] study optimal biopsy decision-making after mammography screening for average-risk women. In particular, [60] examine when a woman should undergo a biopsy based on her mammographic features and demographic factors by a finite-horizon MDP formulation, and [61] investigate the impact of budgetary restrictions on breast biopsy decisions.

In addition to breast cancer screening, several other studies consider disease screening optimization in other contexts such as colonoscopy for colorectal cancer, and PSA test for

prostate cancer screening [62, 63]. Considering the effect of age, gender, and personal colorectal cancer (CRC) history on CRC progression, [62] propose a POMDP to optimize colonoscopy screening policies wit the goal of maximizing total QALYs. They show that the optimal colonoscopy screening policies generally need to be more frequent than the current guidelines recommend and in addition to age and CRC history, gender is a critical factor in determining the frequency of screening. [63] formulate a POMDP model for the prostate cancer screening problem, trading off the benefit of early detection by PSA test with the cost of screening and loss of QALYs due to screening and treatment. Maximizing total QALYs and the net benefit, they reflect the patient and societal perspectives, respectively. They show that the traditional guidelines for PSA test are suboptimal and the optimal PSA screening policies differ significantly depending on the perspective of the decision maker.

Another relevant stream of research focuses on allocation of scarce resources and/or assessing cost-effectiveness in chronic disease screening and management. For example, [64] study hepatitis B, a major health population problem both in China and the U.S., via combining decision analysis and Markov models, and assess the cost effectiveness of several interventions. The findings following from their analyses were instrumental in a relevant health policy change on hepatitis B screening affecting millions of people. [65] focus on HIV screening and develop a linear programming model, combined with a Bernoulli process, and improve the allocation of limited Centers for Disease Control and Prevention (CDC)-funded resources to prevent the most new cases of HIV infection. [66] study colorectal cancer screening and diagnosis problem and introduce a dynamic systems model, combining population dynamics and operational control perspective by considering capacity constraints and allocation rules. They show that more resources should be allocated to diagnostic colonoscopy versus screening colonoscopy if the objective is to minimize mortality rather than incidence rate. [67] structurally analyze the competing demand for screening and diagnosis colonoscopy, and build effective policies that dynamically allocate

resources for these two services at the operational-level.

Finally, several studies recently consider other aspects of breast cancer interventions such as preventive surgery to reduce breast cancer risk or genetic testing to identify high-risk population. [68] study the optimal timing and sequence of prophylactic surgeries, such as bilateral mastectomy (BM) or bilateral salpingo-oophorectomy (BSO), to reduce breast and ovarian cancer risks for BRCA mutation carriers. They establish optimal time windows for preventive surgeries BM and BSO, separately for BRCA1 and BRCA2 mutation carriers, and further show the existence of an optimal control limit for the surgery time of a carrier who has already undergone one of the preventive surgeries. [11] employ a decision-analytic model to evaluate cost-effectiveness of population-based testing for a mutation in cancer susceptibility genes BRCA1/2. They conclude that a population-based testing is currently too expensive but can be cost-effective if the cost of genetic testing drops below $250. The authors acknowledge that even when the test is cost-effective, the logistics of "testing more than 100 million women" would be challenging, and argue that funding an aggressive surveillance strategy might offer more health benefits than allocating the fund to the genetic testing of the whole female population over age 30.

A key difference of our study from the existing literature on modeling-based cancer screening/surveillance is that, while the published literature predominantly considers single-modality screening for average-risk population (e.g. mammography screening for average-risk population), we study screening in high-risk population using multi-modality screening technologies. The facts that different screening modalities may be employed with different frequencies over time, and screening modalities can be combined and used in conjunction with each other significantly complicate the problem and make many of the proposed approaches inapplicable. Second, unlike most of the studies in the literature which proposed unstructured policies with frequently changing screening intervals over time, our proposed model enforces structured policies that are easy to implement from a health policy perspective. Third, contrary to the simulation based approaches, our modeling

framework and data-driven findings enable a comprehensive comparison of many potential strategies, and hence may initiate discussions among policy-makers.

1.3 Model Formulation

We take a policy-maker's perspective, where the objective is to maximize population health outcomes by designing effective screening strategies, while considering for budget constraints. Similar to existing guidelines, we consider each age group separately (e.g., 25, 35 or 45-year-old women) and aim to maximize the health outcomes over the entire lifetime of such women, measured in QALYs. In line with reality, we enforce budget constraints over budget planning periods (e.g., 10 years), and systematically vary these budget levels so that our solutions can be adapted in different resource settings. Lastly, we assess the cost-effectiveness of the identified optimal screening strategies at different budget levels using ICER analysis.

Our overall modeling approach is as follows: we first develop and validate a disease progression model to capture incidence and progression of breast cancer under various screening interventions; then, building up on this disease progression model, we formulate an optimization model to identify effective screening strategies for the targeted population, the women at high lifetime risk of developing breast cancer. In the following subsections, we present these two models in detail.

1.3.1 Disease Progression Model

We develop a discrete-state discrete-time finite-horizon Markov model with age-dependent transition probabilities and rewards to capture breast cancer progression in high-risk women. We adapt a societal perspective, and use QALYs to evaluate the health benefits of screening strategies, in compliance with the recommendation of the U.S. Panel on Cost-Effectiveness in Health and Medicine [69, 70]. The disease progression model considers subpopulations of high-risk women with a given initial age (e.g., women at age 25), and simulates disease

incidence and progression over time under different screening actions. The model is developed for the entire screening horizon, covering the time range from the beginning age (e.g., age 25) to the terminal age of screening (e.g., age 100) for the given subpopulation. The model components and the notation used are introduced below.

Time index: Let t denote the time index in years, where $t \in \{1, 2, \ldots, T\}$ with $T < \infty$. We use $T_A = \{1, 2, \ldots, T-1\}$ to denote the time index set excluding the terminal time point T, where no screening action is taken.

Study Population: In line with the existing guidelines [20], we consider age groups separately. Accordingly, the study population corresponds to the cohort of high-risk women with an initial age of a_1 (e.g., age 25) at time $t = 1$, and this population group ages to a_t=$a_1 + t$ - 1 at time $t \in \{1, 2, \ldots, T\}$.

States: Let x_t(s) denote the proportion of the study population in health state s at time t, where $s \in S = \{0, 1, 2, 3, 4, 5\}$ with state **0** representing cancer-free women, states **1** and **2** respectively representing women with so-far undetected in-situ and invasive cancers, states **3** and **4** respectively representing patients initiated in-situ and invasive cancer treatments, and state **5** representing death. We use the notation $S_U = \{0, 1, 2\}$ and $S_T = \{3, 4\}$ to denote the set of unobservable health and observable treatment and post-treatment states, respectively. Previously treated patients need to undergo a more aggressive screening for surveillance, compared with women who have not been diagnosed with cancer [71, 72], and hence, once detected, these patients remain in the treated state and leave the target group for screening.

Actions: An action $a \in A = \{1, 2, 3, 4, 5\}$ represents the screening modality employed, where 1 denotes "No Screening (NO SCREEN)", 2 denotes "Mammography (MAM)", 3 denotes "MRI", 4 denotes "Ultrasound (US)", and 5 denotes "MRI adjunct to mammography (MAM+MRI)". We remark that all of those five actions are permissible for women in states **0**, **1** and **2** whereas the only permissible action for women in a treatment or dead state is "No Screening", and no action is taken at the terminal time period T. We denote the set

of screening actions by $A_S = \{2, 3, 4, 5\}$ by excluding the action a=1, which corresponds to "No Screening".

Transition Probabilities: P_t^a denotes the annual transition probability matrix of the underlying Markov model associated with action a at time t. Figure 1.1 depicts the state transition diagram of disease progression in high-risk population (while state transitions are age-dependent, this dependence is suppressed in the figure for the ease of presentation). In this figure, solid lines represent transitions between health states without any intervention (i.e. when $a = 1$) and the dotted lines represent transitions to the treatment states upon detection, following a "screen" action (i.e. $a \in A_S$), probability of which depends on the sensitivity of the employed modality. Accordingly, $P_t^a(3|1) = P_t^a(4|2) = sens_t(a)$ when $a \in A_S$ and, $t \in T_A$, where $sens_t(a)$ denotes the sensitivity of the corresponding screening technology a at time t, and $sens_t(1)$ is defined as 0 for all $t \in T_A$.

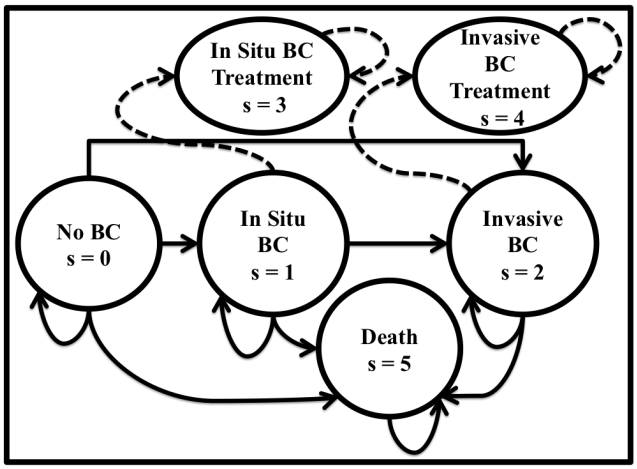


Figure 1.1: Breast Cancer Progression with or without Screening

Solid lines represent transitions between health states without any intervention and the dotted lines represent transitions to the treatment states upon detection, following a "screen" action.

Upon screening, women with positive (i.e. abnormal) screening results undergo a confirmatory diagnostic test such as biopsy, which is assumed to be a perfect follow-up test [73]. Women with positive biopsy results (i.e. women with detected cancer) start cancer treatment, transition to a treatment state, and leave the decision process. On the other hand, women with negative screening results and positive screening but negative biopsy results

remain in the target group for screening in future. Lastly, we remark that specificity of a chosen screening modality affects the overall diagnostic costs, as lower specificity leads to more unnecessary diagnostic exams. However, it does not affect the transition probabilities because in the case of a negative screening test outcome, regardless whether it is a true- or false-negative, no follow-up intervention is made and patients continue to follow their natural disease incidence and progression.

Costs: $c_t(s,a)$ represents the total expected cost of decision a at time t and in state s, including the cost of screening and a follow-up biopsy if performed. Let $c_t^S(s,a)$ and $c_t^F(s,a)$ respectively denote the cost of screening and a follow-up biopsy in state s at time t. Then, $c_t(s,a) = c_t^S(s,a) + p_t^F(s,a)c_t^F(s,a)\ \forall\ s \in S_U$, $a \in A$, $t \in T_A$, where $p_t^F(s,a)$ represents the probability of performing a follow-up test upon a positive screening result in health state s at time t. Accordingly, for $a \in \mathrm{A}_S$, $p_t^F(s,a) = sens_t(a)$ when s $\in \{1,2\}$, $p_t^F(0,a)$ = 1 - $spec_t(a)$, where $spec_t(a)$ represents the specificity of modality a at time t, and $p_t^F(s,1)$ is defined as 0 $\forall\ s \in S$ and $t \in T$. We remark that sensitivity and specificity of some screening modalities change with age (Table 1.3), resulting in varying screening and diagnostics costs at different ages.

Rewards: There are three types of rewards: annual (intermediate), lump-sum (treatment) and terminal rewards. In particular, $r_t(s,a)$ denotes the annual reward defined as the expected QALYs accumulated between time t and $t+1$ when a woman is in health state s and action a is chosen at time t, $R_t(s)$ denotes the lump-sum treatment reward defined as the total expected post cancer treatment QALYs accrued from t until death when the patient is diagnosed with breast cancer and has initiated cancer treatment, and $r_T(s)$ denotes the terminal reward defined as the total expected remaining QALYs at the terminal decision epoch T when the woman's health state is s. More details about reward function estimation can be found in the *Appendix A.6*.

Decision Epochs: Let t denote the time index in years, where $t \in \{1,2,\ldots,T\}$ with $T < \infty$. We use $T_A = \{1,2,\ldots,T-1\}$ to denote the time index set excluding the terminal

time point T, where no screening action is taken.

1.3.2 Optimization Model for Multi-modality Breast Cancer Screening

In this section, we present an optimization model for multi-modality breast cancer screening problem, building upon the disease progression model presented in the previous section. Taking a societal perspective, our objective is to identify structured optimal policies that are practical for implementation, maximizing the total expected QALYs of women at high risk for breast cancer, while considering budget constraints. To be applicable to different resource settings and assess the structure of the resulting policies as budget increases, we solve the optimization models by varying the budget levels.

The overall description of the optimization model is as follows: The model considers a subpopulation of high-risk women with a given initial age (e.g., women at age 25), probabilistically captures their disease incidence and progression over time through stochastic flow balance equations introduced over the disease progression model, and systematically assess the effect of different strategies (i.e. set of screening actions over time) on the total accumulated QALYs through the objective function.

The optimal solution to the base optimization model, introduced in *(1)-(13)* below, may have an irregular structure, where screening modalities and frequencies may change frequently over time (e.g., ultrasound for two years, then MRI for three years, and then mammography for one year etc.), which would prevent the implementation of the resulting strategies in real-life settings. To address this concern, after we introduce the base model, we enforce additional constraints to ensure that the optimal policy would be structured and have limited changes in screening frequencies. More specifically, these constraints guarantee that (i) the chosen screening action remains unchanged over L-year periods and (ii) the frequency of screenings within the L-year periods follows a regular pattern (e.g., annual ultrasound or biennial MRI over the next L=10 years). Further, while the optimization model considers the entire screening horizon similar to the disease progression model, we

consider budget allocation decisions over each of these L-year periods, rather than over the entire planning horizon. This is primarily for practical purposes, because strategic budget plans are typically made over no longer than 10-year periods (while disease burden projections require longer term planning).

There are two types of decision variables in the optimization model. The continuous decision variable $x_t(s,a)$ denotes the proportion of women in health state $s \in S$, who take the screening action $a \in A$ at time $t \in T_A$. The optimization model governs the underlying stochastic process by probabilistically updating the proportion of women in any model states via these continuous decision variables. For the terminal decision period T, there is no decision to take, and $x_T(s)$ denotes the proportion of women in health state s at time T, given the sequence of actions taken in previous periods. The other decision variable $y_t(s,a)$ is a binary variable, indicating the action taken at each decision epoch, and is equal to 1 if action $a \in A$ is taken by the women in health state $s \in S$ at time $t \in T_A$ and is 0 otherwise. In addition to signaling the optimal screening strategy, these binary decision variables are used to introduce several additional constraints that capture the critical features of the multi-modality breast cancer screening problem and ensure that the generated optimal solutions are practical and clinically-meaningful. The base optimization model formulation for our problem is as follows:

$$max \sum_{t\in T_A}\sum_{s\in S_U}\sum_{a\in A} r_t(s,a)x_t(s,a) + \sum_{t\in T_A}\sum_{s\in S_T} R_t(s)(x_{t+1}(s,1) - x_t(s,1)) + \sum_{s\in S_U} r_T(s)x_T(s) \tag{1.1}$$

s.t.

$$\sum_{a\in A} x_1(s,a) = \alpha(s) \quad \forall s \in S \tag{1.2}$$

$$\sum_{a\in A} x_t(s,a) - \sum_{a'\in A}\sum_{j\in S} P_{t-1}^{a'}(s|j)x_{t-1}(j,a') = 0 \quad \forall s \in S, t = 2, ..., T-1 \tag{1.3}$$

$$x_T(s) - \sum_{a' \in A} \sum_{j \in S} P^{a'}_{T-1}(s|j) x_{T-1}(j, a') = 0 \quad \forall s \in S \tag{1.4}$$

$$\sum_{t \in I_k} \sum_{a \in A} \sum_{s \in S_U} c_t(s, a) x_t(s, a) \leq B_k \tag{1.5}$$

$$x_t(s, a) \leq y_t(s, a) \quad \forall s \in S, t \in T_A \tag{1.6}$$

$$\sum_{a \in A} y_t(s, a) = 1 \quad \forall s \in S, t \in T_A \tag{1.7}$$

$$y_t(0, a) = y_t(1, a) \quad \forall t \in T_A \tag{1.8}$$

$$y_t(1, a) = y_t(2, a) \quad \forall t \in T_A \tag{1.9}$$

$$y_t(3, 1) = y_t(4, 1) = y_t(5, 1) = 1 \quad \forall t \in T_A \tag{1.10}$$

$$y_t(s, a) \in \{0, 1\} \quad \forall s \in S, a \in A, t \in T_A \tag{1.11}$$

$$x_t(s, a) \geq 0 \quad \forall s \in S, a \in A, t \in T_A \tag{1.12}$$

$$x_T(s) \geq 0 \qquad \forall s \in S \tag{1.13}$$

The objective value *(1)* represents the total expected QALYs associated with the optimal screening strategy, where the first, second, and third components capture the total QALYs in pre-diagnosis (including disease-free), post-diagnosis (upon detection), and terminal health states, respectively. The difference [x_{t+1}*(s,1)* - x_t*(s,1)*] in the second component accounts for the expected number of women with detected cancer, transitioning to treatment states at time $t \in \mathrm{T}_A$.

Constraint (2) accounts for the initial proportion of the targeted high-risk population with the initial health state distribution α. *Constraints (3)* and *(4)* are stochastic flow-balance equations, capturing and probabilistically updating the proportion of women in health state s over time, depending on chosen sequence of screening actions and the incidence and progression rates over time. *Constraint (5)* introduces a budget constraint for each budget interval of length L, denoted by k, ensuring that the total expected expenditures in each interval *k* (corresponding to L-year periods) do not exceed the allocated budget for

that interval, denoted by B_k.[1]

Constraint (6) links each binary variable $y_t(s,a)$ with a corresponding continuous decision variable $x_t(s,a)$ and prevents illogical solutions such as having a positive $x_t(s,a)$ when $y_t(s,a)$ is zero. *Constraint (7)* guarantees that exactly a single action is chosen for all women in the same health state $s \in S$ at each decision epoch $t \in T_A$, avoiding the possibility of randomized policies, in line with the population-based guidelines. *Constraint (8)* ensures that the same action is chosen for all of the undetected women in the targeted population to be screened. Accordingly, in line with the population-based screening guidelines, all high-risk women of the same age in model states "No BC" (**s=0**), "in-situ BC" (**s=1**), and "invasive BC" (**s=2**) are recommended to take the same action $a \in A$ at time $t \in T_A$.

The output of the above presented base optimization model, formulated by *(1)-(13)*, is a screening strategy that specifies *(i)* whether to screen or not at a given time t and *(ii)* which modality (or joint modalities) to use if screening is performed. Yet, as noted earlier, this strategy may have a dynamic and irregular structure, where screening actions and frequencies change too frequently and hence make the generated strategy impractical for implementation. To address this concern, we enforce additional constraints to ensure that within each L-year period over the entire planning horizon, the strategy is one of the following type: (1) Single screening, (2) Double screening, (3) Biennial screening and (4) Annual screening. Single screening is a strategy with one screening action at the beginning of the corresponding L-year interval, double screening strategy performs two screens, one at the beginning and the other in the middle, biennial screening recommends screening every other year and annual screening strategy recommends screening every year within the L-year interval. As an example, consider the following strategy: MRI screenings between ages 25-34, biennial US screening between ages 35-44, annual mammography screening between ages 45-55 etc. We remark that while such a strategy imposes some regularity

[1] By properly adjusting B_k in *Constraint (5)*, the allocation of the same amount of per person budget can be ensured, after accounting for the women who left the screening process, either due to death or cancer treatment initiation.

into the policy structure and hence is practical, it also offers enough flexibility to capture many different possible combinations of screening frequencies with different modalities over time.

To enforce structured strategies, we utilize four types of binary variables, $\xi(k)$,$\gamma(k)$, $\mu(k)$, and $\theta(k)$, corresponding to the predetermined screening frequencies of single, double, biennial and annual screening for period k, respectively. Based on these binary variables, we introduce "regularity" constraints to ensure that the generated strategies follow a regular pattern. Since the full set of the "regularity" constraints is very long, here we present only a small portion that is sufficient to explain the general concept:

$$\theta(k), \mu(k), \gamma(k), \xi(k) \in \{0, 1\} \tag{1.14}$$

$$\theta(k) + \mu(k) + \gamma(k) + \xi(k) = 1 \tag{1.15}$$

$$\mu(k) + \gamma(k) + \xi(k) = 1 \tag{1.16}$$

$$\gamma(k) - 1 \leq y_{(k-1)*L+1}(0, a) - y_{(k-1)*L+6}(0, a) \leq 1 - \gamma(k) \tag{1.17}$$

$$y_{(k-1)*L+j}(0, 1) \geq \gamma(k) \qquad\qquad j = 2, 3, 4, 5, 7, 8, 9 \tag{1.18}$$

Here, *Constraints (15)* ensures that exactly one of these frequencies is activated for each of the L-year planning intervals. Suppose that the chosen frequency is "double screening" for period k and hence, its corresponding binary variable $\gamma(k) = 1$. Then *Constraint (16)* enforces the same action $a \in A$ to be taken at the two time points t = (k-1)*L+1 and t = (k-1)*L+L/2+1. Further, *Constraint (17)* guarantees that no additional screening is conducted (i.e. a=1) at the other time points (years) of the screening interval k. When the chosen frequency is not "double screening" for period k, i.e. $\gamma(k) = 0$, *Constraints (16)* and *(17)* are not binding in interval k and hence do not pose any limitations. We use similar constraints for other practical frequency structures to capture the full set of the "regularity" constraints that guarantee that the generated strategies are practical for implementation.

The final model is given by *(1)-(13)* and full set of the regularity constraints, the solution of which identifies the optimal practical screening strategies for the specified high-risk group over the entire screening horizon. When we analyze the structure of the optimal strategies and discuss our findings for each subpopulation though, we concentrate on the screening strategies corresponding to the next ten years (rather than the entire lifetime) because: *(i)* based on our interactions with policy makers, we discovered that any discussion of strategies longer than ten year periods is not practical from a strategic planning perspective, especially when budget allocation decisions are taken into account, and *(ii)* it is difficult to gain insights into the structure of the surveillance strategies over the entire screening horizon.

1.4 Analytical Results

In this section, we study the structural properties of regular screening strategies over $L = 10$-year periods (Figure 1.2) and assess the trade-off between the two key parameters of a screening strategy: *sensitivity* of the employed modality and *frequency* of the screenings. In particular, we first separately examine the conditions under which higher sensitivity or frequency is guaranteed to improve health outcomes, measured by expected QALYs, regardless of what the initial cancer prevalence is and which future strategies are implemented after the ten-year period. Then, we investigate the cases where there is a trade-off between the two parameters and establish the conditions under which a strategy with more frequent but less sensitive screens is more favorable than a strategy with more sensitive but less frequent screens.

We start with stating the assumptions we make throughout this section. These assumptions are in line with the evidence in the medical literature and are satisfied by our dataset especially for young and middle-aged high-risk women, for whom the harms associated with screening is negligible comparing to its potential health benefits. The assumptions are stated in main text in an intuitive way, and the formal mathematical conditions correspond-

ing to these assumptions are included in *Appendix A.5*.

Assumption 1: (Detection is better than no detection) For both in situ (i=1) and invasive (i=2) cancer states, detection at any time in the next ten-year period, i.e. $t \in \{1,2,...,10\}$, yields higher expected QALYs than the scenario with no detection by the end of the next ten-year period (followed by a possible detection under any future strategy ϕ implemented after ten years).

Assumption 2: (Early detection is better than late detection) For both in situ (i=1) and invasive (i=2) cancer states, detection at any time t $\in \{1,2,...,9\}$ yields higher expected QALYs than detection at a later time point t+k in the ten-year period, where k $\in \{1,2,...,10\text{-}t\}$.

Assumption 3: Sensitivity of a screening modality $a \in A_S$, denoted by $\text{sens}_t(a)$, is greater than 50% at any time t $\in \{1,2,...,10\}$.

We first present two results, *Propositions 1* and *2*, which show that (i) increased screening frequency or sensitivity leads to better health outcomes, and (ii) these results are robust against the changes in the health state distribution (cancer prevalence) and the future screening actions taken. Unless presented in the main text, all proofs are included in the *Appendix A.5*.

Proposition 1: Between two strategies using the same screening modality with different frequencies, the strategy with higher screening frequency yields higher expected QALYs, regardless of the initial health state distribution and future strategy implemented.

Proposition 2: Between two strategies using different screening modalities with the same frequency, the strategy utilizing a more sensitive modality yields higher expected QALYs, regardless of the initial health state distribution and future strategy implemented.

These two results immediately lead to the following corollary:

Corollary 1: Among affordable ten-year screening strategies, either the one with the highest frequency or the one with the most sensitive modality is the optimal policy over the entire planning horizon for the given budget level. This result holds for any initial health

state distribution and any future strategy that is implemented after the ten-year period.

Propositions 1 and *2* allow a policy-maker to "myopically" focus on the next ten-year strategies at any given time without being concerned about the uncertainty around the future screening strategies that may be implemented beyond the 10-year period. *Corollary 1* narrows the search space of the optimal solutions to the strategies with either the most frequent screens or the most sensitive technology that can be afforded, irrespective of what strategy is implemented beyond the 10-year period. While these results are useful, it is not clear how the optimal policy changes when both screening frequency and sensitivity change simultaneously. Therefore, in our main result below, we analyze the tradeoff between screening frequency and sensitivity. We consider the case where a strategy with more frequent screenings and a strategy with a more sensitive modality are both affordable, and establish the conditions under which the strategy with more frequent but less sensitive screens is more favorable. Before presenting the main theorem, we first introduce some additional notations needed, as follows:

μ_1***(i)*:** Proportion of women in health state i $\in \{0,1,2\}$ in the targeted (sub-)population at time t=1, depending on the initial health state distribution (i.e., prevalence). That is, $\mu_1(i)$ = $\sum_{a \in A} x_1(i,a)$ for i $\in S_U$.

***Pr(s*$_j$*=i|s*$_1$*=0)*:** Probability of being in cancer state i $\in \{1,2\}$ at time $j \in \{2,...,10\}$, given that the patient was healthy at time $t = 1$, i.e. $s_1 = 0$, and no intervention is taken between time 1 and j.

***E*$_\phi$*[s*$_{11}$*=i]*:** Expected future QALYs for a woman with so far undetected stage $i \in \{1,2\}$ cancer at the end of ten-year period (i.e. t=11), gained under a future strategy ϕ implemented after the next ten-year period. The best future clinical scenario is immediate detection right after the ten-year period and the worst one is no detection before death. Accordingly, expected QALYs under the scenarios immediate detection (*ID*) at time $t = 11$ (i.e., right after the next ten-year period) and no detection (*ND*) during the lifetime respectively impose upper and lower bounds on the expected future QALYs.

Rew(t,ϕ\|i): Expected QALYs difference between the scenarios *(1)* detection of a stage $i \in \{1,2\}$ cancer at time $t \in \{1,2,...,10\}$ and *(2)* no detection in the ten-year period possibly followed by a detection at a later time, if any, under future strategy ϕ. We use Rew(t|i)$_{LB}$ and Rew(t|i)$_{UB}$ to denote lower and upper bounds on Rew(t,ϕ|i), respectively.

Rew(t,k\|i): Expected QALYs difference between early detection at time $t \in \{1,2,...,9\}$ and late detection at time t+k, where k $\in \{1,2,...,10\text{-t}\}$, for cancer state $i \in \{1,2\}$.

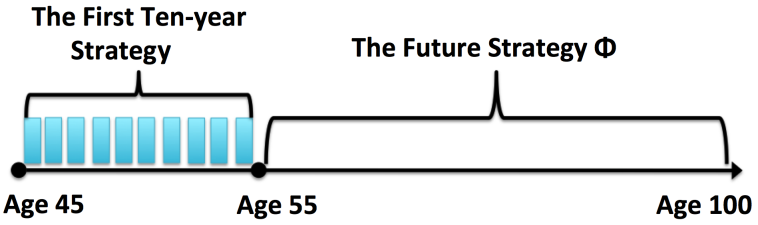


Figure 1.2: The Range of the Screening Strategies for 45-year old High-Risk Women

To investigate the trade-off between sensitivity and frequency, we compare *(A)* "single-screen" with "double-screen", and *(B)* "double-screen" with "biennial" ten-year strategy. In the following result, we characterize the conditions under which a more frequent screening strategy is more effective than a less frequent screening strategy that utilizes a more sensitive technology.

Theorem 1.A: Consider two ten-year screening strategies: single (i.e., every 10-year) screening strategy, call π, and double (i.e., every 5-year) screening strategy, call π', which utilizes a less sensitive modality (i.e., sens$_t(\pi') <$ sens$_t(\pi)$ $\forall$ t = 1,2,...,10). Then, under any future strategy, policy π' yields higher QALYs than policy π, if the following condition holds:

$$\frac{\mu_1(0)}{1-\mu_1(0)} > \frac{1}{Pr(s_6=i|s_1=0)} \frac{sens_1(\pi)-sens_1(\pi')}{sens_6(\pi')} \frac{sens_6(\pi')Rew(1,5|i)+[1-sens_6(\pi')]Rew(1|i)_{UB}}{Rew(6|i)_{LB}} \quad \textit{for i=1,2}$$

Theorem 1.B: Consider two ten-year screening strategies: double screening strategy, call π, and biennial (i.e., every 2-year) screening screening strategy, call π', which utilizes a less sensitive modality (i.e., sens$_t(\pi') <$ sens$_t(\pi)$ $\forall$ t = 1,2,...,10). Then, under a fixed future strategy ϕ, policy π' yields higher QALYs than policy π, if the following conditions and hold:

$$\frac{\mu_1(0)}{\frac{1}{\prod_{t=1}^{5} P_t(0|0)} - \mu_1(0)} > \frac{1}{Pr(s_7=i|s_6=0)+Pr(s_9=i|s_6=0)} \frac{sens_6(\pi)}{min\{sens_7(\pi'),sens_9(\pi')\}} \frac{X(i)}{Rew(9,\phi|i)} \quad \textit{for i=1,2}^2$$

$$\frac{\mu_1(0)}{1-\mu_1(0)} > \frac{1}{Pr(s_3=i|s_1=0)+Pr(s_5=i|s_1=0)} \frac{sens_1(\pi)-sens_1(\pi')}{min\{sens_3(\pi'),sens_5(\pi')\}} \frac{Y(i)}{Z(i)} \quad \textit{for i=1,2}^{34}$$

Theorem 1 establishes the sufficiency conditions under which the strategy with more frequent screenings offsets the disadvantage of using a less sensitive modality and yields greater expected health benefits than the strategy that utilizes a more sensitive technology. The conditions presented in *Theorem 1* can be interpreted based on two critical factors, initial breast cancer prevalence and breast cancer incidence rates, as follows: when the prevalence of the disease in the study population (i.e., 1- $\mu_1(0)$, where $\mu_1(0)$ corresponds to the proportion of disease-free women) is low at the time of initial screenings (i.e., $\mu_1(0) >> 0$), the benefit of using a more sensitive modality is smaller (as compared with the case where initial prevalence is high). On the other hand, if the risk of developing breast cancer is high (i.e, $Pr(s_{1+k}=i|s_1=0) >> 0$ for $i \in \{1,2\}$ and $k > 0$), then, frequent screens might effectively detect the recently developed cancers at earlier stages, leading to more favorable health outcomes as a result of earlier treatment initiation. Accordingly, this result says that if the initial prevalence is (relatively) low and disease progression is (relatively) high such that the conditions in *Theorem 1* are satisfied, then a strategy recommending more frequent screens yields more favorable outcomes than a strategy that employs a more sensitive technology. We remark that in addition to shedding some lights into the overall optimal policy structure, the findings from these structural analyses (i.e., more frequent screens with a less sensitive technology leading to better health outcomes) are also reflected in our real data-driven numerical experiments, as we discuss in the following sections.

[2] X(i)=$sens_7(\pi)$Rew(6,1|i)+[1-$sens_7(\pi)$]$sens_9(\pi)$Rew(6,3|i)+[1-$sens_7(\pi)$][1-$sens_9(\pi)$]Rew(6,ϕ|i)

[3] Y(i)=$sens_3(\pi')$Rew(1,2|i)+[1-$sens_3(\pi')$]$sens_5(\pi')$Rew(1,4|i)+[1-$sens_3(\pi')$][1-$sens_5(\pi')$] $sens_7(\pi')$Rew(1,6|i)+[1-$sens_3(\pi')$][1-$sens_5(\pi')$][1-$sens_7(\pi')$]$sens_9(\pi')$Rew(1,8|i)+[1-$sens_3(\pi')$][1-$sens_5(\pi')$] [1-$sens_7(\pi')$][1-$sens_9(\pi')$]Rew(1,ϕ|i)

[4] Z(i)=$sens_6(\pi)min\{Rew(3,3|i), Rew(5,1|i)\} + [1 - sens_6(\pi)]min\{Rew(3,\phi|i), Rew(5,\phi|i)\}$

1.5 Data and Parameter Estimations

We use multiple sources to estimate our MILP model parameters, which are summarized in Table 1.4. To estimate the age-dependent state transitions between cancer states, we use Wisconsin Breast Cancer Simulation (WBCS), which is a discrete-event simulation model of breast cancer epidemiology developed and validated by a multidisciplinary team in Cancer Intervention and Surveillance Modeling Network (CISNET), a National Cancer Institute (NCI)-sponsored consortium. The model is populated by approximately 3 million women, making up the female population aged 20-100 years of age living in Wisconsin between 1950 and 2000. Details about the WBCS model can be found in [74].

Table 1.2: Cost of Screening Modalities and Biopsy

	Mammography (MAM)	MRI	Ultrasound (US)	MRI adjunct to MAM	Biopsy
Cost	$116	$1,395	$94	$1,511	$1,693

Table 1.3: Sensitivity (Left Table) and Specificity (Right Table) of Modalities

Age	MAM	MRI	US	MRI+MAM
25-30	0.526	0.77	0.593	0.94
31-35	0.526	0.77	0.593	0.94
36-40	0.526	0.77	0.707	0.94
41-45	0.694	0.77	0.786	0.94
46-50	0.761	0.77	0.786	0.94
51-55	0.779	0.91	0.74	0.94
56-60	0.815	0.91	0.74	0.94
61-65	0.809	0.86	0.74	0.94
66-70	0.807	0.86	0.74	0.94
71-75	0.821	0.86	0.74	0.94
76-80	0.839	0.86	0.74	0.94
81-100	0.838	0.86	0.74	0.94

Age	MAM	MRI	US	MRI+MAM
25-30	0.876	0.81	0.892	0.86
31-35	0.876	0.81	0.892	0.86
36-40	0.876	0.81	0.892	0.86
41-45	0.895	0.81	0.968	0.86
46-50	0.894	0.81	0.968	0.86
51-55	0.911	0.898	0.968	0.86
56-60	0.923	0.898	0.968	0.86
61-65	0.927	0.898	0.968	0.86
66-70	0.932	0.898	0.968	0.86
71-75	0.937	0.898	0.968	0.86
76-80	0.939	0.898	0.968	0.86
81-100	0.942	0.898	0.968	0.86

To estimate the incidence of breast cancer among high-risk women, we utilize the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm risk estimation model [75, 76]. BOADICEA is a validated, commonly used risk-estimation tool for high-risk women, calculating the risk of breast and ovarian cancer in women based on key factors such as age, family history and the existence of BRCA1 and BRCA2 gene mu-

tations. BOADICEA is recommended by NCI for use in high-risk women, and has been shown to be well-calibrated for BRCA1 and BRCA2 mutation carriers in the U.S. and other populations [77, 78, 79, 80]. We refer the reader to [81] for a detailed description of the latest version of BOADICEA. In addition to BOADICEA, we use NCI's Breast Cancer Risk Assessment Tool (BCRAT), developed based on the Gail model [82, 83], for women at high-risk for breast cancer due to family history.

We estimate intermediate rewards, representing QALYs, also using the WBCS model, and the lump sum rewards and terminal rewards based on the NCI's Surveillance, Epidemiology, and End Results (SEER) data [84], as described in [85]. The impact of comorbidities is captured by comorbidity-adjusted mortality rates and intolerance to screening and biopsy, as typically experienced by elderly women [86, 51], are captured by a disutility function. SEER data and the incidence ratio as suggested by [87] are used to estimate the initial distribution of the population. The cost of screening modalities and biopsy are drawn from the 2005 Medicare Resource-Based Relative Value Scale [88], and are inflated at a 3% rate to estimate the corresponding 2017 values, which are in line with estimations in reported cost ranges in other sources. To assess the impact of ultrasound's operator-dependency, we conduct a worst-case scenario analysis by using the highest false positive rate reported in literature [25]. Finally, while we present the base-case results in the following section, we conduct extensive sensitivity analyses on many of these parameters, and present the corresponding results in the main text and *Appendix A.4*.

1.6 Numerical Results

We use IBM ILOG CPLEX to solve the MILP formulation of the multi-modality breast cancer screening problem over the lifetime of high-risk women, between ages 25-100 [20]. Our numerical analyses reveal that the structure of the optimal ten-year strategies depends on age groups, which is somewhat expected as the disease progression and the benefits of screening change with age. We characterize these age categories within which the structure

Table 1.4: Breast Cancer Data and Sources

Model Parameter	Source
Breast Cancer Progression Probabilities	WBCS, BOADICEA, NCI (BCRAT)
Intermediate (Annual) rewards	WBCS
Lump-sum (Post-Treatment) rewards	SEER
Terminal Rewards	SEER
Initial population distribution	SEER
Costs of diagnostic tests	Medicare
Sensitivity and specificity of MAM	NCI [89]
Sensitivity and specificity of MRI	[22] , [90], [91], [92]
Sensitivity and specificity of US	[93], [94], [95], [25]
Sensitivity and specificity of MAM and MRI combined	[27]
Disutility of Screening and Biopsy	[96], [97], [98], [99]

of the optimal strategies tend to remain similar. Accordingly, we present and discuss the optimal ten-year strategies based on the following identified three age categories: 25-44 years old (young) women, 45-74 years old (middle-aged) women and over 75 years old (elderly) women. In our base case analysis, we focus on the optimal screening strategies for women with BRCA1 gene mutations (see *Sections 1.6.1-1.6.4*). Later, in *Section 1.6.5*, we extend our analysis to BRCA2 mutation carriers and women with a family history of breast cancer.

In our numerical presentation, we first present the optimal screening strategies for BRCA1+ women, which are identified by solving the MILP model at various budget levels. Then we assess the cost-effectiveness of these different optimal strategies at various budget levels by using incremental cost effectiveness ratio (ICER) analysis [100]. To conduct ICER analysis, we order the identified optimal strategies by the increasing order of required budget and define ICER of strategy i as follows: ICER(i) = [B_{i+1} - B_i]/[Q_{i+1} - Q_i], where B_i is the allocated ten-year budget per person for strategy i and Q_i is the expected QALYs gained under strategy i. In line with the health policy literature [101], we use a willingness-to-pay (WTP) ratio of $ 100,000 and deem that a strategy is cost-effective if its ICER is

lower than this WTP per a unit QALY gain. To identify and eliminate the strategies whose health benefits and costs are outperformed by a mixed strategy of two other alternative strategies, we use the principle of *extended dominance* [102]. For easier interpretation and comparison of the results, we discuss the cost-effectiveness of screening strategies based on budget levels allocated for a single woman, but if preferred, these budget levels could easily be adjusted to compute the required population-level budget by multiplying the per woman budget with the size of the corresponding population.

1.6.1 Results for Young (25-44 Year Old) Women

In this section, we consider the optimal screening strategies for young high-risk women, aged 25-44 years-old. Given the similarity of the policies in the age group of 25-44, we focus on the results for 25 year-old high-risk women in the main text for simplicity of the presentation (For the results for 35 year-old high-risk women, see *Appendix A.1*). Results are presented in Figure 1.3, which depicts the expected QALYs (y-axis) and ten-year budget levels allocated per woman (x-axis). Each point in this QALY/cost curve corresponds to a strategy, and among these strategies, the optimal ten-year policies at various budget levels are depicted to the right of this figure. As an example, biennial ultrasound screening is the optimal ten-year strategy when the budget level is between \$1,430 - \$2,793, and beyond \$2,793 "Annual US" becomes optimal and continues to be optimal until budget is increased to \$16,841. We further compute the corresponding ICER value for each optimal ten-year strategy at various budget levels, and also present them in the sub-table on the right in Figure 1.3.

We make three interesting observations with important policy implications from these results: (1) as the budget increases, the optimal policies tend to typically chance quickly, until annual US becomes optimal, which tend to remain optimal a wide range of budget levels, between \$2,793 - \$16,841. This finding implies that for young high-risk women, "Annual US" is affordable with moderate budget, cost-effective and robustly optimal over

a wide range of budget levels; (2) mammography alone is not optimal at any budget levels since ultrasound screenings achieve higher QALYs with lower costs for this age group; and (3) although recommended by the some of existing guidelines, "Annual MRI+mammography" is not cost-effective with an ICER $100,000, and "Annual MRI" is a dominated strategy (by extended dominance). These results collectively indicate that annual ultrasound should be the preferred ten-year strategy for screening young BRCA1+ women and the use of MRI should be avoided in resource-constrained settings with WTP value of $100,000/QALY.

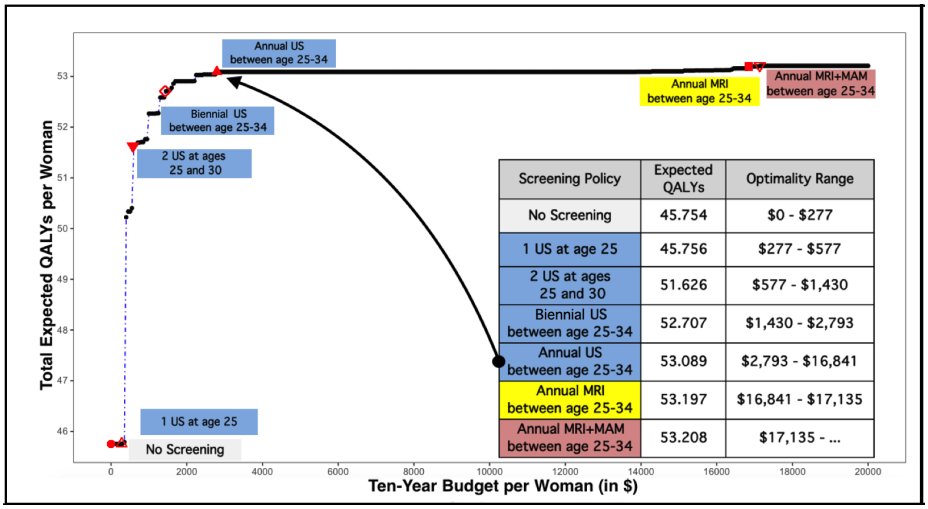


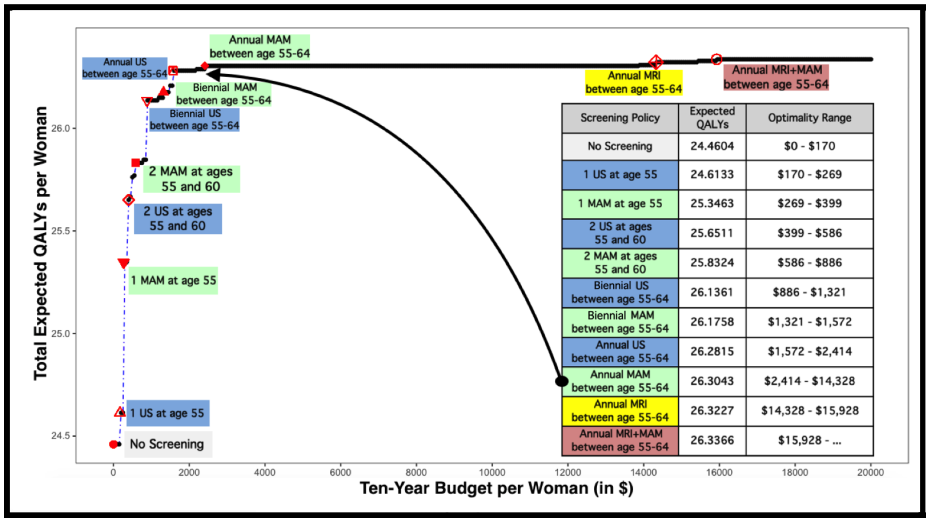
Figure 1.3: Optimal Screening Strategies for 25-Year Old BRCA1+ Women

1.6.2 Results for Middle-Aged (45-74 Year Old) Women

For "middle-aged" high-risk women, we present the results for 55 year-old BRCA 1 mutation carriers (See *Appendix A.1* for the results for 45 and 65 year-old high-risk women). Results are presented in Figure 1.4, which lists the optimal ten-year strategies with their expected QALYs, costs, ICER values, and the budget range over which they continue to be optimal. For example, annual mammography between ages 55-64 yields 26.3043 QALYs, costs $2,414 per woman and remains optimal over the budget range $2,414 - $14,328.

We make the following observations for "middle-aged" BRCA1+ high-risk women from these results: (1) "Annual MAM" is a notably effective strategy: It is affordable with

moderate budget, cost-effective and robustly optimal over a wide range of budget levels; (2) Strategies utilizing ultrasound are optimal for low-to-moderate budget levels and cost-effective and hence, offer alternatives to mammographic screenings when resources are tight; and (3) MRI is not a cost-effective technology as "Annual MRI" is dominated (by extended dominance) and "Annual MRI+MAM" results in an ICER $400,000/QALY, which is way above the WTP=$100,000, despite yielding the highest health benefits. These findings collectively imply that annual mammography screening should be the recommended ten-year strategy in middle-aged high-risk women and the regular use of MRI, alone or jointly with mammography, should be avoided for cost containment.



Screening Policy	Expected QALYs	Optimality Range
No Screening	24.4604	$0 - $170
1 US at age 55	24.6133	$170 - $269
1 MAM at age 55	25.3463	$269 - $399
2 US at ages 55 and 60	25.6511	$399 - $586
2 MAM at ages 55 and 60	25.8324	$586 - $886
Biennial US between age 55-64	26.1361	$886 - $1,321
Biennial MAM between age 55-64	26.1758	$1,321 - $1,572
Annual US between age 55-64	26.2815	$1,572 - $2,414
Annual MAM between age 55-64	26.3043	$2,414 - $14,328
Annual MRI between age 55-64	26.3227	$14,328 - $15,928
Annual MRI+MAM between age 55-64	26.3366	$15,928 - ...

Figure 1.4: Optimal Screening Strategies for 55-Year Old BRCA1+ Women

1.6.3 Results for Elderly (75 Year-Old and Older) Women

For "elderly" high-risk women, we present the results of 75 and 85 year-old BRCA 1+ women together. Table 1.5 lists optimal ten-year screening strategies and expected QALYs, costs, and ICER values corresponding to these strategies. One of the key observations for elderly high-risk women is that annual screening strategies are either not optimal at any budget level or when optimal, they are not cost-effective (e.g., annual mammography, alone or with MRI, is not optimal and annual ultrasound is not cost-effective for 75-year old

women). Further, among the cost-effective strategies, the screening frequency of the most favorable strategy drops from five screens (biennial ultrasound) to two screens (double mammogram) as age of the targeted population increases from 75 to 85. This observed reduction in the frequency of screenings in optimal strategies for elderly population is due to the diminishing benefits of screening with older age as a consequence of increasing comorbidity rates and intensified harms associated with screening and biopsy [103, 51]. The second key observation is that ultrasound alone and mammogram alone both yield affordable strategies that are optimal and cost-effective at certain budget levels. These findings collectively suggest that both mammography alone or ultrasound alone might be used in elderly high-risk populations but the frequency of screenings should be biennial or less and should be decreased with increasing age.

Table 1.5: Optimal First Ten-year Strategies for 75- and 85-Year Old BRCA1+ High-risk Women

75-year Old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	11.643	$0	NA
1 US at age 75	11.743	$197	DOMINATED
1 MAM at age 75	11.785	$272	1,917
2 US at ages 75 and 80	11.817	$357	2,609
2 MAM at ages 75 and 80	11.827	$486	DOMINATED
Biennial US between age 75-84	11.852	$734	10,887
Annual US between age 75-84	11.852	$1,308	1,649,468

85-year Old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	6.503	$0	NA
1 US at age 85	6.548	$205	4,541
1 MAM at age 85	6.551	$276	DOMINATED
2 US at ages 85 and 90	6.555	$307	14,803
2 MAM at ages 85 and 90	6.556	$413	71,116
Biennial US between age 85-94	6.557	$576	141,212

1.6.4 The Role of Ultrasound's Operator Dependency

Compared to mammography and MRI, ultrasound has a higher operator-dependency, causing higher user-variability on the accurate classification of positive (abnormal) image results [34]. Despite its widespread availability, low cost and favorable patient tolerance, the operator-dependency of ultrasound prevents consensus on the consistent use of breast ultrasound as it affects the resulting size of false-positive readings and the total cost of screening strategies and hence, is a major concern in national guideline development [104]. To ad-

dress this concern, we perform a worst-case scenario analysis by using specifically the highest false-positive rates for ultrasound reported in the literature (20.2%) than the values used in the base-case analysis (11.8% and 3.2% below and above age 40, respectively). The corresponding results are presented in Table 1.6, which lists the optimal strategies and the corresponding costs, QALYs and ICERs, for 25-,55- and 75- year old BRCA1+ women, as representatives of young, middle-aged and elderly high-risk populations, respectively.

Table 1.6: Optimal Strategies for 25-, 55- and 75- Year Old BRCA1+ Women with Low US Specificity

25 Year-old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	45.754	$0	NA
1 MAM at age 25	45.756	$326	DOMINATED
1 US at age 25	46.704	$436	DOMINATED
2 MAM at ages 25 and 30	50.554	$670	140
2 US at ages 25 and 30	51.571	$890	DOMINATED
Biennial MAM between age 25-34	52.556	$1,664	496
Biennial US between age 25-34	52.729	$2,204	DOMINATED
Annual MAM between age 25-34	53.041	$3,263	3,300
Annual US between age 25-34	53.089	$4,334	22,443
Annual MRI between age 25-34	53.197	$16,841	DOMINATED
Annual MRI+MAM between age 25-34	53.208	$17,135	106,969

55 Year-old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	24.460	$0	NA
1 MAM at age 55	25.346	$269	304
2 MAM at ages 55 and 60	25.832	$586	652
Biennial MAM between age 55-64	26.176	$1,321	2,141
Annual MAM between age 55-64	26.304	$2,414	8,501
Annual MRI between age 55-64	26.323	$14,328	DOMINATED
Annual MRI+MAM between age 55-64	26.337	$15,928	417,691

75 Year-old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	11.643	$0	NA
1 MAM at age 75	11.785	$272	1,917
2 MAM at ages 75 and 80	11.827	$486	5,017
Biennial MAM between age 75-84	11.848	$1,015	25,700

The key findings of this worst-case analysis, conducted with conservatively high false-positive rates for ultrasound, are as follows: (1) For "young" BRCA1+ women, "Annual US" strategy is still affordable with moderate budget, cost-effective and robustly optimal over a wide-range of budget levels; (2) Mammography alone is the only optimal and cost-effective modality for middle-aged and elderly BRCA1+ women; and (3) the use of MRI is still not cost-effective at a WTP value of $100,000/QALY. Accordingly, with conservatively high false-positive rates, the annual use of ultrasound is still the most favorable cost-effective strategy for "young" high-risk women but ultrasound fails to offer an alternative to mammography in other age groups.

1.6.5 Extension of Results to Other High-Risk Groups

Among high-risk populations, breast cancer incidence and progression rates differ depending on whether the dominant risk factor is family history or inherited gene mutation. In our base numerical analysis, we discuss the results for women at high-risk due to BRCA 1 gene mutations. In this section, we extend the results to two other major risk groups, BRCA 2 mutation carriers and women with a family history of breast cancer. The analyses with base-case parameters reveal that the key results following from BRCA 1 mutations carriers continue to hold for other two high-risk populations: (1) "Annual US" and "Annual MAM" are both affordable with moderate budget, cost-effective and robustly optimal over a wide budget range for young and middle-aged high-risk women, respectively; (2) The use of MRI, alone or with mammography, leads to outcomes that are not cost-effective. Yet, there are some important differences in the optimal strategies between the three different high-risk groups (i.e., BRCA 1+, BRCA 2+, women with family history), which we highlight in the following subsections. The full set of results for BRCA2+ and women with a family history of breast cancer are included in *Appendix A.3*.

Table 1.7: ICER Values of Optimal Screening Strategies for 35-Year Old Women (When The False-Positive Rates of Ultrasound is High)

Screening Policy (High False+ Rates for US)	BRCA 1 Mutation	BRCA 2 Mutation	Family History
No Screening	NA	NA	NA
1 MAM at age 35	DOMINATED	DOMINATED	359
1 US at age 35	129	162	DOMINATED
2 MAM at ages 35 and 40	DOMINATED	DOMINATED	901
2 US at ages 35 and 40	395	473	DOMINATED
Biennial MAM between age 35-44	835	960	3,524
Biennial US between age 35-44	DOMINATED	DOMINATED	DOMINATED
Annual MAM between age 35-44	3,849	4,696	13,942
Annual US between age 35-44	35,500	56,123	208,130
Annual MRI between age 35-44	DOMINATED	DOMINATED	DOMINATED
Annual MRI+MAM between age 35-44	121,175	155,173	454,451

Young High-Risk Women: For "young" women at high-risk due to BRCA 2 gene

mutations or family history of breast cancer, we present the results for 35-year old high-risk women, and compare them against the results for BRCA1+ women. The main difference from the findings for BRCA1+ women is the following: When ultrasound has a low specificity rate, leading to a high number of false-positives, most of the strategies utilizing ultrasound are cost-effective for BRCA1/2 carriers but none of these strategies, including "Annual US", is cost-effective for women with family history (Table 1.7). Relatively lower breast cancer incidence and progression rates experienced by women with family history, compared to BRCA1/2 mutation carriers, might explain why ultrasound fails to offer an alternative to mammography for young women with family history in the clinical scenario with high false-positive rates.

Table 1.8: Screening Strategies for Different High-Risk Groups

Screening Policy (Base-Case)	BRCA 1 Mutation	BRCA 2 Mutation	Family History
2 MAM at ages 45 and 50	DOMINATED	COST-EFFECTIVE	DOMINATED
1 MAM at age 55	COST-EFFECTIVE	DOMINATED	DOMINATED
1 MAM at age 65	COST-EFFECTIVE	DOMINATED	DOMINATED
1 US at age 75	DOMINATED	DOMINATED	COST-EFFECTIVE
1 MAM at age 75	COST-EFFECTIVE	COST-EFFECTIVE	DOMINATED
Annual US between age 75-84	NOT COST-EFFECTIVE	COST-EFFECTIVE	NOT OPTIMAL

Screening Policy (High False+ Rates for US)	BRCA 1 Mutation	BRCA 2 Mutation	Family History
1 MAM at age 45	DOMINATED	DOMINATED	COST-EFFECTIVE
1 US at age 45	COST-EFFECTIVE	COST-EFFECTIVE	DOMINATED
1 MAM at age 55	COST-EFFECTIVE	DOMINATED	DOMINATED
1 MAM at age 65	COST-EFFECTIVE	DOMINATED	DOMINATED

Other Age Groups: For "middle-aged" and "elderly" high-risk women, the differences between optimal strategies of different high-risk groups mostly correspond to low budget scenarios where only a single or at most two screenings are affordable (Table 1.8).The only exception is the "annual" US strategy for 75-year old women. This strategy is not optimal for women with family history of breast cancer, optimal but not cost-effective for BRCA 1 mutation carriers, and optimal and also cost-effective for BRCA 2 mutation carriers. These

results are in line with the differences in five-year breast cancer risk at age 75 for these risk groups: 3.8% for women with family history, 4.9% for BRCA 1 mutation carriers, and 8.0% for BRCA 2 mutation carriers.

1.7 Discussion and Conclusions

In this paper, we study the multi-modality breast cancer screening problem for high-risk population. The performance of mammography, the standard modality for average-risk women, is not satisfactory for high-risk women. MRI and ultrasound are shown to have potential health benefits for high-risk population in detecting breast cancer earlier. Yet, the cost and false positive rates of MRI and operator dependency of US are important concerns in guideline development. Currently, there is no consensus on the optimal and cost-effective design of screening strategies for high-risk women, using MRI and ultrasound either in addition or instead of mammography. Attempting to make an initial step in closing this gap, we identify optimal screening strategies for high risk women, including women with known BRCA 1/2 gene mutations and family history of breast cancer, by considering the technologies that are in widespread use.

Taking a societal perspective, we study optimal screening strategies for high-risk women including BRCA1/2 gene mutation carriers and women with a family history. The key findings of our analysis are as follows: (i) For "young" high-risk women, "annual ultrasound" screening is a very efficient strategy even with high false-positive rates and should be the recommended ten-year policy in resource-constrained settings; (ii) "annual mammography" is optimal strategy for "middle-aged" high-risk women as it is affordable with moderate budget, robustly optimal over a wide budget range with high QALYs, and cost-effective; (iii) The strategies utilizing MRI, either as the primary modality or jointly with mammography as recommended by guidelines, require a large amount of resources due to high cost and high false-positive rates of this modality and lead to outcomes that are not cost-effective.

Both for developed and developing countries, identification of cost-effective surveillance programs that adapt frequent screens such as "Annual US" and "Annual MAM" strategies, is important for young and middle-aged BRCA 1/2 women as timely detection is very critical in these populations due to rapid progression of cancer [105]. Furthermore, the scenarios we identify where ultrasound screenings require low-to-moderate budget and offer a cost-effective alternative to mammogram might address some of the limitations of screening programs in developing countries, where mammography is not widely available [106, 107, 108]. The robustness of "Annual US" in young high-risk populations, for whom the performance of mammography is not satisfactory [109, 110], is especially important, because the lack of proper training to interpret whole-breast ultrasound screenings might cause an increased number of false-positives in developing countries [26]. In older high-risk age groups, the operator-dependency of ultrasound might be a limiting factor for its use as the primary modality, as our results indicate that ultrasound fails to offer cost-effective alternatives to mammogram in the older population when its false-positive rates are high.

Given that assessing all plausible strategies via comprehensive randomized controlled trials is time and cost-prohibitive, our findings can complement clinical studies and be instrumental in designing future trials, developing evidence-based guidelines and informing insurance coverage decisions. Furthermore, this flexible optimization framework can be useful in other healthcare applications, especially when the number of practical interventions is large and a variety of complexities must be captured to generate clinically-meaningful solutions. Finally, our study is also important for the ongoing debate on the cost-effectiveness of individualized or population-based genetic-testing. By enabling targeted high-level surveillance for BRCA1/2+ carriers, genetic-testing has potential to improve health outcomes; however, despite the reduction in prices of these tests, the cost-effective use of screening resources, as shown in this study, is required to justify the widespread clinical adaptation of genetic-testing for BRCA1/2 and other gene mutations.

We should note that our study is not without limitations. First, we primarily utilized

estimations for high risk populations but reliable estimations were not always available for all the modalities. In order to address this limitation, we conducted extensive sensitivity analyses on such parameters. Second, we do not consider emerging technologies, such as molecular breast imaging [111] and breast tomosynthesis [112] due to rarity and lack of sufficient clinical evidence and the combination of ultrasound and mammography together as a single action due to lack of reliable data on sensitivity and specificity rates. Finally, a population-based approach is taken in this study and hence, caution must be exercised if the findings are implemented in clinical practice, where individual patient characteristics such as breast density, and personal values and behaviors, such as adherence to recommended screening actions, play an important role.

CHAPTER 2
PHYSICIAN STAFFING IN EMERGENCY ROOMS VIA A MULTI-CLASS MULTI-STAGE QUEUING NETWORK

2.1 Introduction

We study a staffing problem in the context of a medium-to-large scale emergency room (ER) and determine staffing levels for ER physicians satisfying "tail probability of delay" (i.e., serve $y\%$ in x minutes) type service targets. Conducted in collaboration with Mayo Clinic, our practice-based approach accounts for the key dynamics of the ER care including time-varying demand, multi-stage service and multi-class patients. Keeping practicality as a priority, we develop a queuing network to model the ER care and utilize it to estimate the *offered load* for treatment queues, corresponding the workload on ER physicians. Treatment queues in ER operate in "efficiency-driven" (ED) mode, where servers (physicians) are mostly busy and waiting before service is usually inevitable, but experience no (or, if any, negligible) abandonment. Motivated by these unconventional ED queues, we propose a new staffing algorithm converting *offered load* into staffing for these type of ED queues, and demonstrate its utility in various ER settings. Finally, we tackle an associated routing problem, and discuss several routing rules that are practical for implementation with our solutions.

Below, we first provide a brief background on ER staffing and its impact on the ER care. Next, we motivate the need for a network queuing model and present a list of our contributions.

2.1.1 Background: Emergency Room (ER) and ER Physician Staffing Problem

The importance of an emergency room (ER) for a hospital, and for a healthcare system in general, cannot be overstated. It is one of the major gateways for hospital admissions and plays a pivotal role in acute healthcare delivery by providing critical service for emergency and disaster management. In addition, especially in the U.S., ERs are also used by many non-urgent patients as a means to access general healthcare services [113], as they provide care for more than 140 million patients annually [114].

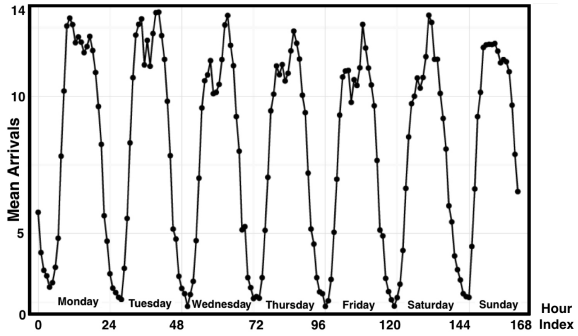


Figure 2.1: Average ER Arrivals - St. Mary Hospital, Mayo Clinic, Rochester, MN (Jan. 2014 - Dec., 2016)

In line with their importance, there is a constant pressure to improve ERs' efficacy and efficiency across the nation [115]. On the one hand, ERs are expected to provide high-quality care in a timely manner despite high patient volumes and crowdedness problem [116, 117]. On the other hand, especially in the midst of rising healthcare costs, eliminating inefficiencies and unnecessary costs in ERs is a central concern [118]. Accordingly, effective and efficient management of ERs, achieving timely high-quality service with efficient resource use, has become an utmost necessity.

An ER's performance is strongly linked with its workforce planning decisions for ER physicians, the most critical human resources in ERs. ER physicians are the leading members of ER care teams, make critical decisions under situations requiring emergency care, and play a decisive role in providing timely and effective medical service. Hence, physician understaffing causes severe consequences for ERs, including increased medical error risk,

prolonged waiting times and higher patient dissatisfaction and abandonment rates [117, 119, 120]. On the other hand, overstaffing can quickly diminish ERs' cost-effectiveness, as ER physician salaries, averaging around $290,000 per year, constitute a significant portion of the overall personnel cost [121]. Finally, contrary to nurses and other ER personnel, it is fairly difficult to make real-time adjustments to the staffing levels of ER physicians, making them a key determinant of ERs' capability in coping with time-varying demand and preventing systematic capacity-demand mismatches.

Table 2.1: Mean Treatment Durations and Subsequent Departure Rates - St. Mary Hospital, Rochester, MN

Patient Class (ESI)	Treatment Stage	Proportion Departed after Treatment	Expected Service Time (hour)
Patient Class 1	Resuscitation	-	1.35
	1	-	0.12
	2	56%	0.45
	3	30%	0.43
	4	14%	0.80
Patient Class 2	1	-	0.27
	2	41%	0.44
	3	33%	0.95
	4	26%	0.59
Patient Class 3	1	-	0.33
	2	50%	0.57
	3	32%	0.90
	4	18%	0.67
Patient Class 4	1	-	0.34
	2	75%	0.62
	3	19%	0.38
	4	6%	0.37
Patient Class 5	1	-	0.34
	2	79%	0.62
	3	16%	0.38
	4	4%	0.37

Optimizing staffing levels is a challenging task for ERs as they exhibit high levels of service complexity. First, ER arrivals are non-deterministic and time-varying (Figure 2.1) with daily and hourly variations [122, 123, 124]. Second, high patient volumes and medium-to-long sojourn times, experienced by many ERs, put a significant amount of workload on ER personnel and cause over-crowding unless managed properly [125]. Third, the ER care delivery is not a single encounter. It involves multiple treatment stages, where a patient is seen multiple times by an ER care team, and is likely to be paused at least a few times for the completion of auxiliary services such as lab work and diagnostic imaging [126, 127]. Fourth, ER patients are not a homogeneous population [128]. Upon arrival,

patients are categorized into different patient classes, where each class has varying degrees of urgency, severity and service requirements. Last but not least, the ER care delivery is not a standard process for all patients, where the average duration of an ongoing treatment and the expected number of additional patient-physicians interactions are both patient type and treatment history dependent, showing variations among patient classes and treatment stages (Figure 2.1).

Canadian Triage and Acuity Scale National Guidelines			
Acuity Level	Time Threshold T (Minutes)	Targeted Service Level (1-α)	Tolerance Target α
Patient Class 1	0	0.98	2%
Patient Class 2	15	0.95	5%
Patient Class 3	30	0.90	10%
Patient Class 4	60	0.85	15%
Patient Class 5	120	0.80	20%
Service Objective:	Achieve $Pr(Wait > T) \leq \alpha$ at all times for each patient class		

Figure 2.2: Canadian Triage and Acuity Scale Guidelines: Patient-Class Specific TPoD Type Service Goals

Reduced waiting time in ERs is shown to be associated with higher patient satisfaction [129] and lower adverse outcome risk [130, 131]. Accordingly, keeping waiting times at acceptable levels is a suitable objective for the ER physician staffing problem. A particularly noteworthy guideline on waiting times is the Canadian Triage and Acuity Scale (CTAS), which is also adopted by many ERs in the U.S. [132]. CTAS categorizes ER patients into five classes (namely, Resuscitation, Emergent, Urgent, Semi-Urgent and Non-Urgent) and specifies two service parameters for each patient category: (1) a threshold value, say T, on waiting before service provided by an *Qualified Medical Professional* (i.e., ER physician), and (2) a tolerance target, say α, serving as an upper bound for the fraction of patients whose waiting time is tolerated to exceed their threshold value (Figure 2.2). To meet CTAS targets, an ER must satisfy that probability of waiting more than threshold T before service is always less than tolerance target α (i.e., Pr(Wait(t)$>T$) $< \alpha$ $\forall$ times t) for all patient classes. These types of service goals are referred to as "tail probability of delay (TPoD)" in queuing theory literature [133] and have been used in many service applications.

2.1.2 A Network Model for ER Care Delivery and A New Staffing Algorithm

Queuing models have been increasingly utilized in healthcare for workforce management and system analysis [134]. Taking a black-box approach, single-station queues (e.g., Erlang-A ,-B and -C models) are among the models that are commonly used to analyze system behavior and performance [135, 136]. However, the ER physician staffing problem demands a more comprehensive framework, as these popular models fail to consider a fair number of the key features of ER care delivery, summarized as follows: **(1)** Physicians are not the only service providers over the course of the ER care delivery as the process also includes various tasks undertaken by other members of the ER care team. **(2)** The medical service provided in ER is not a single event and involves multiple physician-patient interactions [127]. Further, for each patient class, the durations of these physician-patient interactions (i.e., treatments) and the probability of receiving another treatment before departure show variations among treatment stages. **(3)** Some patients leave the ER without receiving treatment. Yet, these abandonments happen mostly at earlier stages of the ER care (i.e., before a patient is placed on his assigned bed and seen by an ER physician). Accordingly, the treatment processes directed by ER physicians experience no (or, if any, negligible) patient abandonment.

Considering these key features of the ER care, we develop a multi-stage queuing network model that (1) captures multiple physician-patient interactions, (2) differentiates the physician-guided works from the services provided by nurses or other ER personnel (e.g., lab work), and (3) separately models the earlier stages of the ER care delivery process to correctly account for abandonments.

Two main building blocks of our multi-stage network are the "Treatment" and "Order Bundle" queues, separating the ER care directed by ER physicians from other services. "Treatment" stations are where ER physicians examine and treat patients over the course of the ER care and are used to calculate the workload on ER physicians. "Order Bundle" stations represent the group of diagnostic medical orders, requested by an ER physician

to better assess a patient's condition before continuing treatment, and are positioned after treatment stations (Figure 2.3). These orders are conducted by nurses or allied health staff and may include lab work (e.g., blood test) or specialty tests (e.g., electrocardiogram and CT scan). Such representation of ER care, describing its central section as a sequence of treatment and order bundle processes, is intuitive and in line with ER physicians' perspective, who view their work as successive examination-decision combinations.

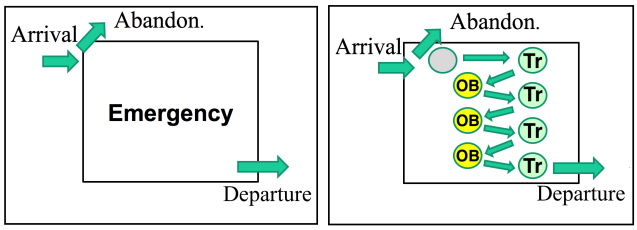


Figure 2.3: Black-box and Multi-Stage Network Models for ER Care Delivery Note: Treatment and order bundle stations are abbreviated with "Tr" and "OB", respectively.

Treatment queues operate with *efficiency-driven* perspective, where servers (physicians) are mostly busy and delays before service are usually inevitable. Yet, contrary to conventional ED regimes, these stations experience no (or, if any, negligible) abandonments, as abandonments almost always occur (and are captured by our network model) at the earlier stages of the ER care (e.g. in the waiting area). This type of an ED regime, especially with TPoD targets and time-varying demand, has not gained much attention since abandonment, a non-negligible feature of the majority of crowded queues, is considered a defining characteristic of an ED model [137, 138] and accordingly, ED queues are commonly studied with abandonment [139, 140]. This motivates us to develop a new staffing formula, translating the load on servers into staffing decisions that meet TPoD targets for ED queues without abandonment (i.e., with perfectly patient customers).

2.1.3 Our Approach and Key Contributions

In collaboration with Mayo Clinic, we study the ER physician staffing problem with time-varying demand. Our objective is to determine optimal staffing levels satisfying patient-

class and treatment-stage specific TPoD targets at all times. To the best of our knowledge, ours is the first study that simultaneously and comprehensively accounts for various real-life features, including time-varying demand, multi-class customers, multi-stage service of discontinuous nature, and differentiated TPoD targets. More specifically, our overall approach and key contributions are as follows:

Multi-Class Multi-Stage Network Model: We develop a multi-stage queuing network to capture the patient flow in ERs and estimate time-varying *offered load* (OL) for treatment queues. This model serves two important purposes: (1) Differentiating patient-physician interactions from services provided by other ER personnel and the earlier stages of the ER care, where abandonments happen, it allows to model and staff the queues that are served by ER physicians more accurately. (2) Opening the black-box of care delivery, it provides more access to and managerial control over the internal components of ER service. Using this modeling framework, a decision-maker can analyze the instantaneous load on physicians at various stages of the ER care, set differentiated TPoD targets for patient classes and treatment stages, and make staffing decisions accordingly.

New Staffing Algorithm: Motivated by the treatment queues, we develop a new and comprehensive algorithm to staff the ED queues that serve under TPoD type service targets and experience no abandonment. We analytically show the asymptotic effectiveness of our staffing rule on stabilizing TPoD for M/M/s queues, operating in ED heavy-traffic mode. Via realistic and data-driven simulation experiments, we demonstrate the robustness of this new algorithm in various ER settings, characterized by time-varying stochastic demand, multi-stage service, multi-class customers (patients), and centralized (*pooled*) servers (ER physicians). Finally, we propose a weighted average technique to avoid overstaffing due to the *pooling effect* (i.e., improved service quality due to the *pooling* of resources), and show its utility in multi-class multi-stage cases.

Practical Routing Rules: The real-time performance of ER physicians also depends on the routing mechanism, determining which patient to be served next. Therefore, we

address the associated *routing* problem for ERs, and assess several practical routing rules, coupled with our staffing solutions, under various clinical scenarios and network setups. We demonstrate that as system complexity (measured by arrival volumes, number of patient classes and the expected number of treatments) increases in ERs, static routing rules fail to meet TPoD targets. We further show that dynamic routing rules, making routing assignments based on the system state, are not always sufficient to cope with the high levels of complexity of medium-to-large scale ERs either. For these cases, we propose *hybrid* routing policies, integrating predetermined priorities (via static rules) and the use of system state (via dynamic rules), and discuss how they, together with "*ED-NoAb*", achieve the desired performance in medium-to-large scale ERs.

The rest of this paper is organized as follows: *Section 2.2* summarizes the relevant literature and highlights our key differences. *Section 2.3* presents our network model. *Section 2.4* discusses the derivation of our staffing method and describes its use in networks with time-varying demand. *Section 2.5* presents our numerical experiments and discusses the implementation of our staffing approach with practical routing rules. *Section 2.6* summarizes our findings, and concludes our study.

2.2 Literature Review

There are two main streams of research within the modeling-based ER care delivery literature. At the one end, there are studies using simulation models to analyze the ER patient flow and evaluate *what-if* scenarios in different ER settings (See [141] for a review). Given the complexity of ER care delivery, a simulation model might offer a (relatively) convenient framework (than an analytical method) to study various operational problems in ERs [142] . Yet, to be accurate and reliable, a simulation model requires meticulous efforts to (1) account for all major system characteristics and interactions, (2) estimate and fine-tune a large set of parameters, and (3) calibrate and validate the model [126]. Further, studying a complex process with a simulation model carries the risk of being *system-of-interest-*

specific, as it requires detailed modeling, and hence, might offer limited insights for the same system in other settings.

At the other end of the spectrum, there are queuing theory based analytical approaches, as ours, focusing on different aspects of health service operations including ER care delivery. Queuing models provide an abstract representation of a service process, and are quite instrumental to study (and gain insights from) system behavior and performance via closed-form formulas. See [134], [135], [136] and [143] for a review the queuing theory applications in ER and other healthcare settings.

2.2.1 A Queuing Theory-based Staffing Approach: Offered-Load Analysis

A central queuing theory-based *staffing* approach for medium-to-large service systems is *offered-load* (OL) analysis [144]. The key idea is to approximate the behavior of the original system with a corresponding infinite-server (IS) queue to compute its OL (i.e., the instantaneous arrival rate per unit service rate) and then use the OL to propose staffing levels for the original system [145]. The foundations of this approach are laid down by [146], who analyze an Erlang-C (i.e., M/M/s) model under heavy-traffic limits (i.e., arrival rate λ and number of servers $s \to \infty$ while service rate μ is kept fixed) with a probability of delay (PoD) target α. They show Pr(Wait > 0) $\to \alpha \in (0,1)$ if and only if $\frac{s-\lambda/\mu}{\sqrt{\lambda/\mu}} \to \beta_\alpha$ for some quality-of-service (QoS) parameter $\beta_\alpha \in (0,\infty)$, leading to the staffing formula "*square-root safety (SRS)*" $s = OL + \beta_\alpha\sqrt{OL}$, OL = λ/μ. Studying Erlang-A (M/M/s+M) and M/M/s+G queues, [137] and [139] extend this approach to many server queues with abandonment. In addition, [137] introduce three staffing regimes, *Efficiency-Driven* (ED), *Quality-Driven* (QD) and *Quality-and-Efficiency-Driven* (QED), to be used depending on whether the emphasis is on (server) efficiency, (service) quality or both. [140] study Erlang-A queue and introduce the "ED+QED" regime (i.e., QED refinement of ED staffing) to stabilize TPoD for stationary queues with abandonment. Also see [147, 148, 149], and [150] for related works on optimal staffing of many-server queues based on asymptotic behavior

analysis.

Experiencing time-varying demand, many service systems do not operate in stationary fashion. Addressing this issue, [151] and [152] generalize OL analysis to time-varying settings, and provide new tools, namely *modified-offered-load* (MOL) and *infinite-server* (IS) approximation, to propose staffing under time-varying demand. Applying a simulation-based iterative algorithm, [153] generalize the SRS to queues with time-varying arrivals and customer abandonment. See [127, 154, 155, 156] for various extensions of the *MOL* method under different performance functions such as PoD, probability of abandonment (PoAb) and mean waiting time. There has also been attempts to address time-varying demand by using traditional models in a non-stationary way. The core idea is to approximate non-stationary system behavior using the exact steady-state formula of stationary queues at each time point. Examples of this approach include *pointwise stationary approximation* and *stationary independent period-by-period* methods, which have been proven useful in a variety of settings, including ER staffing [157, 158, 159].

2.2.2 Repetitive Service and Our Key Differences From The Existing Literature

Beside ours, there exists a few studies accounting for the discontinuous and repetitive nature of health services. [160] study an ER nurse staffing problem with a closed queuing model M/M/s/n, where s nurses serve n patients, attempting to re-enter service when their state switches from "served (stable)" to "in need for service". Based on new many-server asymptotic results, the authors develop heuristics to set nurse staffing levels stabilizing TPoD, or its special case PoD, with a variant of SRS formula. Motivated by healthcare systems, [127] develop a two-station open queuing network with a feedback loop, and using SRS with *MOL* approximation, propose a staffing formula stabilizing PoD in QED systems facing customer re-entries and time-varying demand. In their model, a patient either departs the ER after being served in the first queue or, with a fixed probability, transitions to the second queue, where she waits and then re-visits the first station for service. Both

studies propose staffing levels using a (stationary or time-varying) variant of SRS, consider a homogeneous (single-class) patient population, and assume identical and memoryless "service" and "return to service" processes (i.e., use exponential distributions having the same means for every customer re-entry). Instead, we consider a heterogeneous patient population, model each "treatment" and "return to treatment" process separately, and use "history-dependent" (i.e., treatment stage-specific) service times for the ER care process. The key differences of our study and the existing queuing literature are as follows:

(1) Instead of predominantly used single-station queues, we propose a practical queuing network model that simultaneously considers various complex features including time-varying demand, multi-class customers, and multi-stage service, and differentiated service goals. Specifically, by allowing the use of different (i) general probability distributions for service times, (ii) departure probabilities after service completion, and (iii) TPoD type service targets for each customer type at each service stage, our network model accounts for the differences in the service processes and requirements. This modeling framework can also be applied in other complex service settings and is especially useful for the systems having various service components with different characteristics that cannot be captured by simplistic black-box models.

(2) We propose a new queuing network- and OL analysis-based staffing approach. In particular, we develop and use this approach to meet TPoD targets for ED queues experiencing no abandonment. Yet, our approach can also be extended to other settings, where a different staffing formula (e.g. SRS) is used to stabilize a different performance measure (e.g., PoD) for other types of queuing regimes (e.g., QED).

(3) We tackle some additional challenges that arise from the problem features we jointly consider. In particular, we show how to account for the *pooling effect* of centralized servers while proposing staffing over a multi-class multi-stage network. Further, addressing the associated *routing problem*, we also shed some lights into the relationship between system complexity and staffing-routing policies.

2.3 A Multi-Class Multi-Stage (MCMS) Network Model for ER Care

The ER care delivery process, illustrated in Figure 2.4, can be described as follows: Upon arrival, patients are categorized and prioritized by nurses at triage based on their perceived health risk, and subsequently, begin waiting to be admitted to the ER, during which they might choose to leave. After the waiting period, each patient is assigned to and placed on an empty treatment bed. This process is referred to as "bed placement", during which nurses and other allied health staff interact with the patient and conduct the initial medical assessment. This process is followed by a physician-patient interaction (i.e., treatment process). Once the ER physician completes her initial examination, she might request a set of orders ("order bundle") to assess the patient's medical condition before continuing the treatment. An order bundle may involve lab work (e.g., blood tests), diagnostic imaging (e.g., X-ray) and other procedures, and is completed by nurses and other allied health staff. Upon completion of the orders, the ER physician analyzes the results and continues the patient's treatment with this additional information. If required, she pauses the treatment again and requests more orders. These treatment and order bundle sequences cease when the ER physician arrives at a final disposition decision, ending the patient's ER care. Subsequently, the patient departs the ER either to be admitted to a hospital inpatient unit or to go home. The departure process is supervised by nurses and require no more emergency care.

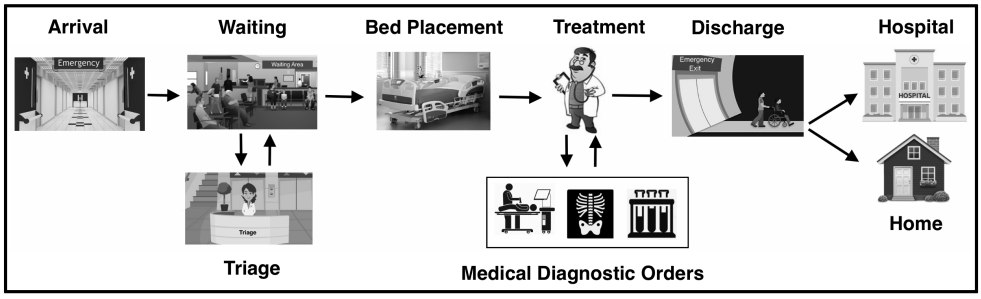


Figure 2.4: Patient Flow in the ER from Arrival to Departure

We develop a queuing network to model the ER care delivery process, illustrated in

Figure 2.5. The model components and the notation used in the manuscript are described below:

Patient Classes: P $= \{1, 2, .., m\}$ denotes the set of patient classes, where m $< \infty$. Using a five-tier triage scale (m=5), we define the patient classes as follows: Level 1 (critical) corresponds to critical conditions, requiring an immediate life-saving medical intervention. Level 2 (emergent) conditions pose a potential threat to life limb or function, requiring rapid medical intervention. Level 3 (urgent) corresponds to a stable health status that causes a certain level of discomfort and carries a risk of progression to a more serious problem over the course of the ER care. Level 4 (semi-urgent) conditions are relatively low-risk but have potential for deterioration if treatment is delayed over an hour. Level 5 (non-urgent) conditions are associated with the lowest risk, allowing for delays in the ER care without significant consequences. In this study, we conduct numerical experiments on Level 2, 3, and 4 patients, constituting 98% of arrivals in our dataset and discuss the impact of not considering Level 1 patients later.

Arrival Process: Arrivals to ER are time-varying and stochastic, and modeled as independent non-homogeneous Poisson processes with rate $\lambda^x(t)$ at time t for patient class x $\in$ P.

Bed Placement, Treatment, and Order Bundle Queues: For patient class $x \in$ P, TR_x := $\{1, 2, ...\}$ denotes the set of treatment queues and μ_i^x denotes the service rate at treatment stage $i \in \text{TR}_x$. Similarly, OB_x denotes the set of order bundle queues for patient class $x \in$ P, where δ_i^x is the service completion rate of order bundle process $i \in \text{OB}_x$. In addition, δ_0^x denotes the completion rate of bed placement process for patient class $x \in$ P.

Abandonment: With varying rates depending on patient class, abandonments, also called *left-without-being-seen* (LWBS), happen in the waiting area while patients wait to be admitted to the ER after triage. No significant abandonment is recorded in our data following the bed placement process. Moreover, this observation also extends to most other ERs as it is unlikely for a patient to prematurely leave the ER care after being placed

on a treatment bed in the ER.

Departure Probabilities: For patient class $x \in$ P, p_i^x denotes the probability of departure after treatment stage $i \in \text{TR}_x$. Then, with probability 1- p_i^x, a patient in class $x \in$ P goes through at least another sequence of order bundle and treatment processes after treatment stage $i \in \text{TR}_x$.

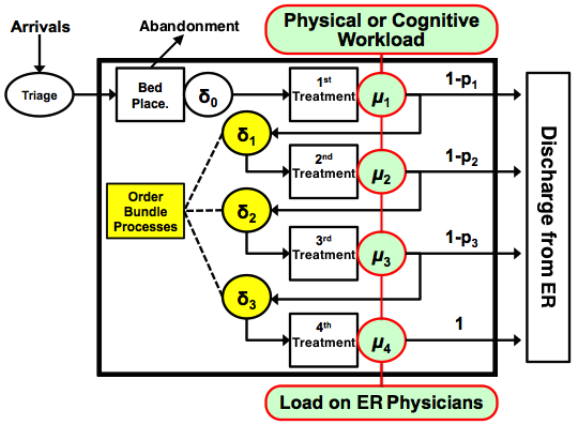


Figure 2.5: A Multi-stage Queuing Network with Four Treatment Stages Note: The patient class index is suppressed for simplicity.

In the next section, we describe how we *(i)* utilize our network model to calculate the instantaneous loads on treatment stations, representing the physical and cognitive workload on ER physicians, and *(ii)* convert them into staffing levels for ER physicians with our staffing algorithm.

2.4 A Staffing Rule for ED Erlang-C Queues and the MCMS Network

Our staffing approach can be decomposed into two steps: *(1)* The calculation of time-varying *offered-load (OL)* for each queue of interest under (a version of) *infinite-server (IS)* approximation, *(2)* the translation of OL functions into staffing levels via an analytic formula-based staffing algorithm. The key idea of our approach to use instantaneous OL functions (of some specific queues) for (overall) staffing decisions, and is applicable to many service systems. Yet, attention must be paid to the characteristics of the queues of interest as they impact both the choice of the OL calculation method (*Step 1*) and the staffing

formula (*Step 2*). Hence, before describing the details of our method and its implementation for the ER physician staffing problem, we first briefly discuss the defining features of treatment queues, capturing the medical services provided by ER physicians.

The operational behavior of treatment queues is characterized by the following factors: (1) Facing high patient volumes, ER physicians (servers) are mostly busy. (2) Delay before each treatment stage (physician-patient interaction) is usually inevitable. (3) At each stage of treatment, keeping the waiting time at an acceptable level is required for patient satisfaction and safety. (4) Abandonments only occur before a patient is placed on an ER bed (i.e., at earlier stages of ER care).

The features of treatment queues define an *efficiency-driven (ED)* regime, aiming reasonable service quality with highly utilized servers, but not a conventional one given the lack of abandonment. For a conventional ED queue, service capacity is set below the arrival rate by a moderate fraction, and abandonments compensate the excess of demand over capacity, preventing asymptotic results from being divergent or degenerate. Yet, there is no (non-negligible) abandonment from treatment queues that can be used to regulate their asymptotic behavior. Such an ED regime, especially with a TPoD target (rather than the commonly used PoAb), has not gained much attention in the literature, given abandonment is usually considered a defining characteristic of ED queues [137, 138]. Accordingly, as we study this special ED regime, we develop a new staffing rule to convert the OL into staffing levels achieving TPoD goals.

We organize the rest of the *Section 2.4* as follows: In *Section 2.4.1*, we formally define an ED regime for Erlang-C (M/M/s) queues, show the asymptotic convergence of TPoD to a non-degenerate limit for this stationary regime, and discuss how this result leads to a new staffing formula (i.e., *ED-NoAb*). In *Section 2.4.2*, we describe how to use *ED-NoAb*, together with *delayed-infinite-server* approximation [154], for a *dedicated* non-stationary $M_t/G/s_t$ queue, serving a single patient class under time-varying (Poisson) demand with TPoD target. In *Section 2.4.3*, we generalize the *ED-NoAb* algorithm to multi-class multi-

stage networks, where *pooled* agents (ER physicians) serve patients from multiple classes under time-varying arrivals and differentiated service goals. We also briefly discuss the *routing* policies we employ with *ED-NoAb* in multi-class multi-stage settings. Finally, in *Section 2.4.4*, we confer about how our staffing approach can be implemented in other queuing systems exhibiting different characteristics.

2.4.1 An Analytic-Formula based Staffing Rule for *Efficiency-Driven* M/M/s Models

Now, we present *Theorem 1*, which *(i)* defines an ED regime for $M/M/s$ queues, and *(ii)* describes the asymptotic (i.e., heavy-traffic limit) behavior of ED $M/M/s$ queues based on tail probability of delay and expected waiting time. The proof of *Theorem 1* is provided in *Appendix B*.

Theorem 1: (An ED Regime for M/M/s) Consider a sequence of $M/M/s$ queues indexed by n, each of which has arrival rate λ_n, service rate $\mu_n = \mu \in (0, \infty)$, and s_n servers. Let the traffic intensity $\rho_n = \frac{\lambda_n}{\mu s_n} < 1$. Consider probability target $0 < \alpha < 1$ and delay threshold $T > 0$. Then, if

$$\lim_{n\to\infty} s_n(1-\rho_n) = \beta \quad as \quad \lambda_n, s_n \to \infty, where \quad \beta = \frac{\ln(1/\alpha)}{\mu T} > 0, \tag{2.1}$$

The probability of delay and the server utilization converge to 1:

$$P(W_n > 0) \to 1 \quad and \quad \rho_n \to 1 \quad as \quad \lambda_n, s_n \to \infty. \tag{2.2}$$

The tail probability of delay and the expected waiting time converge to non-degenerate limits:

$$P(W_n > T) \to \alpha \quad and \quad E[W_n] \to w^* \equiv \frac{T}{\ln(1/\alpha)} \quad as \quad \lambda_n, s_n \to \infty. \tag{2.3}$$

Remark 1: (Heavy-Traffic Regime: *ED Erlang-C*) In *Theorem 1*, the condition 2.1

plays two significant roles regarding the behavior of an $M/M/s$ queue as s_n and λ_n grow to infinity: (1) It keeps the resource capacity s_n and the load on resources λ_n/μ "*matched*" in a way that the system stability is maintained (i.e., $\rho_n < 1$), and (2) it dictates the rate of convergence for server utilization ρ_n (in an order $1/s_n$) (i.e., $\rho_n \to 1$ - β/s_n). Under the condition 2.1, a (heavy-traffic) $M/M/s$ queue operates in an *efficiency-driven* mode, achieving high server utilization at the expense of delays before service (*Theorem 1.i*). We dub this heavy-traffic regime "*ED Erlang-C*", where the TPoD and expected waiting time both converge to non-degenerate limits (*Theorem 1.ii*). Note that the fine-tuning of these metrics (i.e., $P(W > T)$ and $E[W]$) are achieved via controlling the (positive) constant β in 2.1. That is, $E[W]$ is governed by the service parameters α and T, and $P(W > T)$ is stabilized around the probability target α when $\beta = \ln(1/\alpha)/\mu T$.

We explain the key difference of *ED Erlang-C* with the conventional ED regimes, allowing for customer abandonment, as follows: The ED models with impatient customers (e.g., $M/M/s+M$) operate with staffing levels below the OL = λ/μ (i.e., $\rho > 1$) to achieve high server utilization, where $E[W] > 0$, $P(W > 0) \approx 1$ and $P(W > T) > 0$. In such systems, the abandonment rate serves as a regulator to keep the traffic intensity of the served customers < 1, which ensures the system stability [139]. Yet, given the lack of abandonment, all customers are needed to be served in *ED Erlang-C* queues and hence, $\rho < 1$ is required to maintain stability in *ED Erlang-C* regimes. This is satisfied by setting staffing levels above the OL, as we discuss next.

Remark 2: (Asymptotic Staffing Formula: *ED-NoAb*) *Theorem 1* suggests a staffing rule for *large-scale* (i.e., λ, $s >> 0$) $M/M/s$ queues, operating in ED mode with a TPoD target. That is, solving the equation $s(1\text{-}\rho) = s\ (1\text{-}\ \lambda/\mu s) = \beta$ for the staffing level s^*, we obtain that $s^* = \lambda/\mu + \beta$, $\beta > 0$. We refer this staffing formula "*ED-No Abandonment* (*ED-NoAb*)". Notice that *ED-NoAb* sets s^* above the OL = λ/μ, where the excess staffing β is carefully tailored to stabilize the TPoD around the probability target. In particular, with the choice of $\beta = \ln(1/\alpha)/\mu T$ (as in *Theorem 1*), the TPoD $P(W > T) \approx \alpha$ for a *large-scale*

$M/M/s$ queue.

2.4.2 The Use of *ED-NoAb* Staffing with Time-Varying Arrivals

For models with time-varying arrivals, using asymptotic regime staffing formulas together with the time-varying version of the stationary OL has been shown to perform well [151, 152]. The key idea is to compute the time-varying OL for each time t, denoted by $m(t)$, under *IS* approximation, and then use it to replace the stationary OL in the asymptotic staffing formula. A modified version of this method, called *delayed-infinite-server (DIS)* approximation, is proposed by [154]. In this study, we adopt and customize their approach to propose staffing for $M_t/G/s_t$ queues, operating under a TPoD-type service constraint.

[154] study a multi-server queue with general service times, non-homogeneous Poisson arrivals and general abandonments (i.e., M_t/G/s_t + G) and develop a novel two-stage network approach (*DIS* approximation) to stabilize probability of abandonment (PAb). *DIS* approximation decomposes the original M_t/G/s_t + G system into two (artificial) *IS* queues in series, representing the waiting and service areas. In this setting, each arriving customer is required to stay a deterministic amount of time ω in the first station and, unless he abandons the system at the first station, is subsequently admitted to the second station (i.e., service area) without further delay. Then, for any given time $t \geq \omega$, the mean number of busy servers, all operating in the second *IS* queue, is calculated with a time-shifted integral as follows: [*DIS Formula*] $m_\omega(t) = [1-F(\omega)] \int_0^{t-\omega} [1-G_S(t-\omega-u)]\lambda(u)du$, where $G_S(.)$ and $F(.)$ respectively denote the cumulative distribution functions (CDFs) of service and abandonment processes, and λ(t) denotes the arrival rate at time t. Parameter ω is chosen so that PAb is stabilized around $F(\omega)$ (i.e., PAb $\approx F(\omega)$, $\omega \geq 0$) and $m_\omega(t)$ serves as the stochastic OL (to be used for calculating staffing levels).

We make two modifications to customize the *DIS Formula* for $M_t/G/s_t$ queues targeting to achieve P($W(t) > T$) $\approx \alpha$ at all times t: (1) We set the forced delay ω equal to ω^* = $T/ln(1/\alpha)$ based on *Theorem 1* showing that E[W] $\rightarrow$ $T/ln(1/\alpha)$ as $\lambda \rightarrow \infty$ (2.3), (2)

In line with the absence of abandonments, we define a degenerate cdf $F(.)$ for abandonments, where $F(\omega) = 0$ for any $\omega \geq 0$. The modified version of *DIS*-based OL function is calculated as follows for any time $t \geq \frac{T}{ln(1/\alpha)}$:

$$m_{T,\alpha}(t) = \int_0^{t-\frac{T}{ln(1/\alpha)}} [1 - G_S(t - \frac{T}{ln(1/\alpha)} - u)]\lambda(u)du. \tag{2.4}$$

Then, the time-varying staffing levels $s(t)$ is given by $s(t)^* = m_{T,\alpha}(t) + \frac{ln(1/\alpha)}{T}$ E[S], where S is the service time, to ensure P(W(t) $> T$) $\leq \alpha$ at any given time t.

2.4.3 The Use of *ED-NoAb* Staffing for a Network Model with Pooled Servers

To make OL calculations tractable, *DIS* approximation requires the arrival process of a queue to be Poisson [154] and hence, its use over the nodes of the MCMS network are based on the following two features: (1) The Poisson property is preserved for the departure process of a *DIS* queue given the arrival process is also Poisson [161, 154]. (2) Except for the external (original) arrivals, arrivals to any queue of the MCMS network are originated from the departures from its preceding queue (Figure 2.5). Then, when the external arrivals are Poisson, arrivals to any queue of the MCMS network follow a Poisson process under *DIS* approximation [162], and the formula (2.4) can be used for any node of the MCMS network.

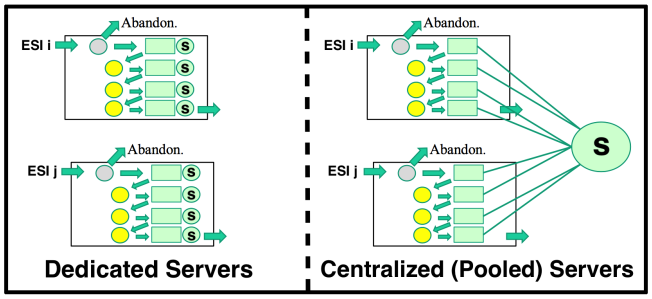


Figure 2.6: Network Models for Medical Services Provided by Dedicated vs. Centralized ER Physicians/Servers(s)

We calculate the OL functions over the MCMS network iteratively by setting the departure rate of a visited queue, or its splitting, as the arrival rate of the next one (See [155] for

a detailed description of departure rate computations). Next, we convert the OL functions into staffing levels that ensure P(W(t) $> T_i^x$) $\leq \alpha_i^x$ holds $\forall\ x \in P$ and $i \in TR_x$ at all times t. During this conversion, it is essential to account for the *pooling effect* (i.e., improved service quality due to the *pooling* of resources) as the ER physicians are *pooled* servers, who serve all patients and treatment queues as opposed to being *dedicated* to a subset of them (Figure 2.6).

For a *pooled*-server system, a *naive* approach for determining the total staffing is to sum the optimal *dedicated* staffing levels without making any adjustments for the *pooling effect*. Neglecting the effect of *pooling*, this approach causes overstaffing in multi-class multi-stage settings, especially for ED queues [163]. Instead, we propose a weighted average technique that addresses the *pooling effect* and avoids overstaffing. This technique normalizes the *excess* staffing, the staffing above the total OL, with arrival rates and assigns higher weights to the queues (and/or patient classes) having higher instantaneous demand. Let S_i^x, λ_i^x(t) and m_i^x(t) respectively denote the service time, arrival rate, *DIS*-based OL function of patient class $x \in P$ in treatment queue $i \in TR_x$ at time t. Then, the total staffing proposed by the *naive* approach and weighted average method at time t, respectively denoted by s_N(t) and s_{WA}(t), are calculated as follows:

The *Naive* Approach: $s_N(t) = \sum_{i,x} s_i^x(t)$, where $s_i^x(t) = m_i^x(t) + \frac{ln(1/\alpha_i^x)}{T_i^x} E[S_i^x]$,

Weighted Average Method: $s_{WA}(t) = \sum_{i,x} m_i^x(t) + \sum_{i,x} \frac{\lambda_i^x(t)}{\sum_{i,x} \lambda_i^x(t)} \frac{ln(1/\alpha_i^x)}{T_i^x} E[S_i^x]$.

Besides staffing levels, the operational performance of an ER depends on *routing* decisions. Hence, to be practical in multi-stage multi-class settings, our solutions must also provide a *routing* mechanism, specifying which patient to be served next [164]. As typically done in the literature [165], we address the *staffing-routing* problem in two separate stages, where we first develop a *staffing* algorithm and then assess its performance under several practical *routing* rules. Table 2.2 lists the static and dynamic routing rules considered in this study, where the dynamic rules use the current system state for *routing* decisions.

Before we conclude this section, we briefly discuss the factors that must be paid atten-

Table 2.2: Brief Descriptions of the Static and Dynamic Routing Rules

Name	Type	Description
Class-Based Routing	Static (Default)	Serve the patient from the class with higher medical urgency first. Within the same patient class, prioritize an earlier treatment stage.
Stage-Based Routing	Static	Serve the head-of-the-line patient waiting at an earlier treatment stage. Within the same treatment stage, prioritize the patient class with higher medical urgency.
Initial-Treatment Routing	Static	Serve the initial treatment stage first with priority given to the class with higher urgency. If no patient is waiting in the initial treatment stage, then apply "class-based" routing.
Low-Volume Based	Static	Serve the patient class with lower volume (i.e., total number of arrivals per day) first. Within the same patient class, prioritize an earlier treatment stage.
Shortest Mean Service Time	Static	Among the waiting patients, serve the one whose service is expected to be shortest. That is, prioritize the patient class and treatment stage with the shortest average service time.
Stage- and Low-Volume-Based	Static	Serve the head-of-the-line patient waiting at an earlier treatment stage first. Within the same treatment stage, prioritize the patient class with lower (daily) volume.
Maximum-Waiting Ratio	Dynamic	Serve the patient with the maximum ratio of so-far waiting at the current treatment stage to pre-determined waiting threshold (of that treatment stage for the class the patient belongs to).
Shortest-or-Violated Deadline	Dynamic	Serve the patient with the minimum difference between so-far waiting at the current treatment stage and waiting threshold (of that treatment stage for the class the patient belongs to).

tion to when our staffing approach is implemented in other settings. Two core elements of our approach are (1) the calculation of the OL functions over the MCMS network, and (2) the translation of the OL functions into staffing levels via an algorithm. Accordingly, when this approach is used in a particular setting, the key decision lies in the choice of the staffing and OL calculation methods that are in line with a system's characteristics including the operational behavior of the queues (e.g., ED, QD, QED), the performance measures to be stabilized (e.g., TPoD, PoD, PAb), and the abandonment behavior of customers.

Modeling the queues having different features separately, the MCMS network enables the use the preferred OL method (and a correspondent staffing algorithm) for each queue. For instance, *DIS* approximation can be used for the queues targeting to control their relatively high PAb whereas *IS* approximation is used for the QED queues with PoD targets. A noteworthy point is that, while converting the OL into staffing decisions, the "*pooling effect*" must be taken into account if the servers are *centralized*. In this study, we propose a weighted average technique, normalizing the excess staffing with instantaneous arrival rates, to determine the overall staffing levels serving all "treatment" queues of the MCMS network. We note that another method (or the use of different weights) might be more appropriate in other settings, as the (magnitude of the) effect of "*pooling*" depends on the operational characteristics of the queues of interest [163].

2.5 Numerical Experiments

We conduct several numerical experiments using a detailed and realistic discrete event simulation, written in MATLAB R2016b, to assess the performance of our proposed staffing approach under time-varying demand over a 24-hour period. The experiments are divided into two main categories. First, we simulate single-class, single-station $M_t/G/s_t$ queues with periodically fluctuating arrivals to assess the effectiveness and robustness of *ED-NoAb* staffing in stabilizing TPoD targets under *challenging* time-varying demand. Second, using real ER data, we simulate various ER scenarios over different network setups to *(i)* examine the performance of *ED-NoAb* staffing on realistic and data-driven ER settings, and *(ii)* analyze the relationship between the staffing-routing decisions and system complexity (set by the number of treatments and patient classes, and arrival volumes).

For each simulation experiment, we run N independent replications. We set N = 1000 for relatively simple cases, and N = 2000 when we simulate a scenario with more than two patient classes and/or treatment stages. At each replication, we send virtual patients to the queues of interest and collect waiting time statistics at fixed times Δt, 2Δt, 3Δt, ..., where Δt = 0.25 hour. Unlike the real patient entities, the virtual patients are terminated right before they start receiving service and hence, do not affect the system performance or behavior in simulation experiments.

For each considered scenario, we calculate the estimated TPoD at time t by the fraction of replications that the virtual patients arriving at time t waited more than the delay threshold T. If the estimated TPoD at time t is above the acceptable level, we refer this outcome as a "*violation*". A "*violation*" indicates that our proposed staffing levels cannot meet the desired TPoD target at all times and hence, lead to *understaffing* for the simulated scenario. If there is no "*violation*", we run the same simulation experiment with *reduced* staffing = FLOOR[$0.90 * s(t)$], where s(t) is the staffing levels proposed by *ED-NoAb*, and check whether our proposed staffing levels cause *overstaffing*.

2.5.1 Performance of *ED-NoAb* in $M_t/G/s_t$ Queues under Challenging Demand

First, we investigate the effectiveness and robustness of *ED-NoAb* staffing in $M_t/G/s_t$ queues under periodically fluctuating demand. In line with our objective, we use a sinusoidal arrival rate function (Figure 2.7) to model the arrivals, following a non-homogeneous Poisson process. We emphasize that we choose a sinusoidal function to generate *challenging* demand rather than to mimic the actual arrival process to an ER, for which we will later use real ER data. We set the arrival rate $\lambda(t) = 10 + 2\sin(t)$ so that the mean daily arrivals roughly matches the number of daily visits to a medium-to-large scale ER ($\approx$ 250 patients/day). The other parameters, determined based on CTAS and our discussions with ER managers, are as follows: Expected service time E[S] = 0.25, 0.50, 0.75 and 1.00 hour, delay threshold T = 0.50, 1.00, 1.50 and 2.00 hour, and tolerance level α = 0.05, 0.10, 0.15 and 0.20.

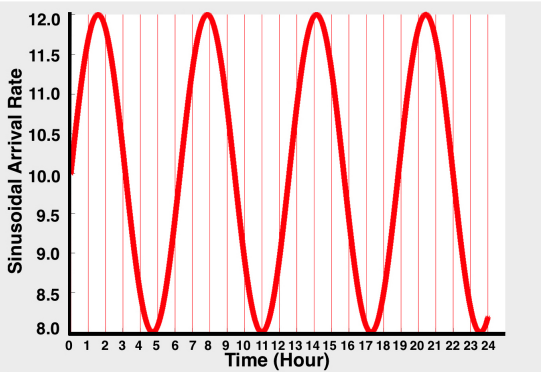


Figure 2.7: Sinusoidal Arrival Rate Function with The Rate $\lambda(t) = 10 + 2\sin(t)$

In total, we conduct 64 simulation experiments. In all 64 cases, each simulated with 1000 independent replications, *ED-NoAb* staffing achieves to meet TPoD targets at all times. On the contrary, when the staffing levels are reduced, TPoD targets are *violated* at certain times during the day in 98.5% (=63/64) of the cases. Table 2.3 summarizes the results. In addition, we provide "TPoD graphs" in *Appendix B*, showing the fraction of patients waited more than threshold T (y-axis) over time (x-axis) and whether the tolerance target α is violated. These results demonstrate *ED-NoAb* staffing effectively stabi-

lizes TPoD targets in $M_t/G/s_t$ queues under *challenging* time-varying demand, and jointly achieves maximum server utilization, as desired in ED regimes, where lower staffing levels fail to meet the desired performance standards.

Table 2.3: Single-Class Single-Stage Simulation Results for 64 Experiments

Experiment Parameters			ED-without-Ab		Max.	[ED-without-Ab*90%]		Experiment Parameters			ED-without-Ab		Max.	[ED-without-Ab*90%]	
E[Service]	Threshold T	Tolerance α	Violations	LoS	Staff	Violations	LoS	E[Service]	Threshold T	Tolerance α	Violations	LoS	Staff	Violations	LoS
0.25	0.50	0.05	NO	0.27	5	YES	0.36	0.75	0.50	0.05	NO	0.74	14	YES	0.78
0.25	0.50	0.10	NO	0.29	5	YES	0.43	0.75	0.50	0.10	NO	0.75	13	YES	0.83
0.25	0.50	0.15	NO	0.31	4	YES	0.60	0.75	0.50	0.15	NO	0.76	12	YES	0.87
0.25	0.50	0.20	NO	0.33	4	YES	0.77	0.75	0.50	0.20	NO	0.79	12	YES	0.93
0.25	1.00	0.05	NO	0.33	4	YES	0.83	0.75	1.00	0.05	NO	0.80	11	YES	0.98
0.25	1.00	0.10	NO	0.36	4	YES	1.06	0.75	1.00	0.10	NO	0.82	11	YES	1.03
0.25	1.00	0.15	NO	0.38	4	YES	1.22	0.75	1.00	0.15	NO	0.87	11	YES	1.14
0.25	1.00	0.20	NO	0.39	4	YES	1.33	0.75	1.00	0.20	NO	0.92	10	YES	1.21
0.25	1.50	0.05	NO	0.37	4	YES	1.20	0.75	1.50	0.05	NO	0.86	11	YES	1.10
0.25	1.50	0.10	NO	0.39	4	YES	1.33	0.75	1.50	0.10	NO	0.93	10	YES	1.23
0.25	1.50	0.15	NO	0.41	4	YES	1.53	0.75	1.50	0.15	NO	0.98	10	YES	1.36
0.25	1.50	0.20	NO	0.43	4	YES	1.71	0.75	1.50	0.20	NO	1.01	10	YES	1.45
0.25	2.00	0.05	NO	0.39	4	YES	1.34	0.75	2.00	0.05	NO	0.93	10	YES	1.25
0.25	2.00	0.10	NO	0.41	4	YES	1.61	0.75	2.00	0.10	NO	1.00	10	YES	1.42
0.25	2.00	0.15	NO	0.44	4	YES	1.91	0.75	2.00	0.15	NO	1.05	10	YES	1.55
0.25	2.00	0.20	NO	0.47	4	YES	2.13	0.75	2.00	0.20	NO	1.16	10	YES	1.82
0.50	0.50	0.05	NO	0.50	9	YES	0.53	1.00	0.50	0.05	NO	0.96	18	NO	0.99
0.50	0.50	0.10	NO	0.52	9	YES	0.58	1.00	0.50	0.10	NO	0.98	17	YES	1.03
0.50	0.50	0.15	NO	0.53	8	YES	0.62	1.00	0.50	0.15	NO	0.99	16	YES	1.09
0.50	0.50	0.20	NO	0.56	8	YES	0.70	1.00	0.50	0.20	NO	1.01	15	YES	1.17
0.50	1.00	0.05	NO	0.57	8	YES	0.72	1.00	1.00	0.05	NO	1.02	15	YES	1.20
0.50	1.00	0.10	NO	0.60	8	YES	0.87	1.00	1.00	0.10	NO	1.06	14	YES	1.39
0.50	1.00	0.15	NO	0.63	7	YES	0.98	1.00	1.00	0.15	NO	1.10	14	YES	1.54
0.50	1.00	0.20	NO	0.67	7	YES	1.10	1.00	1.00	0.20	NO	1.13	14	YES	1.69
0.50	1.50	0.05	NO	0.63	7	YES	0.96	1.00	1.50	0.05	NO	1.09	14	YES	1.51
0.50	1.50	0.10	NO	0.67	7	YES	1.13	1.00	1.50	0.10	NO	1.15	13	YES	1.75
0.50	1.50	0.15	NO	0.73	7	YES	1.33	1.00	1.50	0.15	NO	1.25	13	YES	1.90
0.50	1.50	0.20	NO	0.77	7	YES	1.49	1.00	1.50	0.20	NO	1.26	13	YES	2.04
0.50	2.00	0.05	NO	0.67	7	YES	1.13	1.00	2.00	0.05	NO	1.15	13	YES	1.76
0.50	2.00	0.10	NO	0.74	7	YES	1.39	1.00	2.00	0.10	NO	1.24	13	YES	1.98
0.50	2.00	0.15	NO	0.79	7	YES	1.55	1.00	2.00	0.15	NO	1.33	13	YES	2.20
0.50	2.00	0.20	NO	0.86	7	YES	1.80	1.00	2.00	0.20	NO	1.42	13	YES	2.41

2.5.2 Multi-Class ($m \geq 1$) Multi-Stage ($n \geq 1$) Scenarios with Real ER Data

Using real ER data, we conduct simulation experiments to assess the performance of *ED-NoAb* staffing in realistic ER settings, characterized by time-varying demand, multi-stage service, heterogeneous customer population, and centralized servers (i.e., pooled physician resources). We consider different network setups, consisting of n treatment and n-1 order bundle queues ($1 \leq n \leq 3$), single- (m=1) and multi- ($2 \leq m \leq 3$) class patient population scenarios, and several (static, dynamic and hybrid) routing rules that are practical for implementation. Our main objectives are to (i) demonstrate the effectiveness of our staffing approach in multi-class multi-stage settings and (ii) gain insights into the relationship between the staffing-routing decisions and system complexity (as a function of the number of treatments and patient classes, and arrival volumes).

Data and Parameter Estimation: We estimate the parameters of ER care delivery

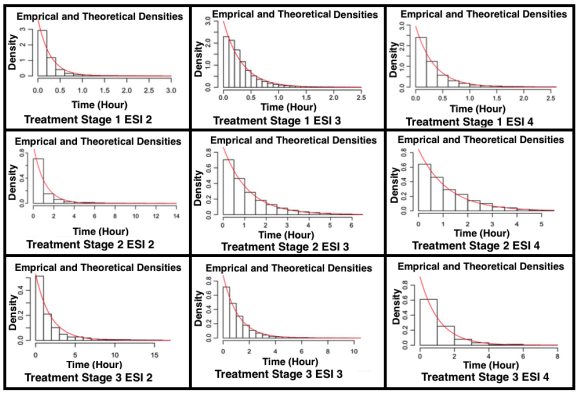


Figure 2.8: (Exponential) Service Time Distributions for Treatment Stages 1-3 and ESI 2-4

process using time-stamp data of approximately 108,000 ER patients, receiving service at the ER of Mayo Clinic Saint Mary's Hospital at Rochester, MN between July, 2012 and September, 2015. In line with the existing healthcare operations literature, we model the ER arrivals as a non-homogeneous Poisson process [166, 167, 168, 127, 169, 170]. Based on the best fit to the observed data, we select exponential distributions (with different means) to model the service times of treatment and order bundle queues (Figure 2.8). We emphasize that exponentially distributed service times are not a requirement, as our approach, relying on the use of $M_t/G/\infty$ queues in MCMS network, allows general service time distributions. Service target parameters (delay threshold T and probability tolerance α) are determined based on CTAS and our discussions with ER managers and physicians.

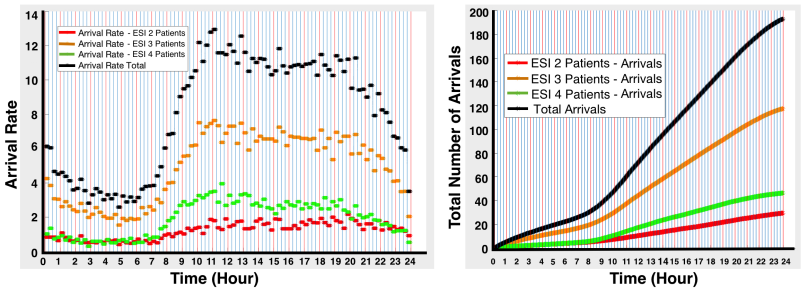


Figure 2.9: Arrival Data of ESI 2, ESI 3, and ESI 4 Patients

Based on Emergency Severity Index (ESI), a five-tier triage scale roughly matching CTAS [132], patients are classified into five categories in the dataset, where a lower number

indicates a higher priority [128]. Similar to [171], we conduct numerical experiments on the patient classes (i.e., ESI 2, 3, and 4) that constitute the majority ($> 98\%$) of ER arrivals, and later, separately discuss the impact of ESI 1 (i.e., the highest priority) patients on staffing decisions and system performance. The arrival volumes of ESI 2, ESI 3 and ESI 4 patients are respectively low, high and medium (Figure 2.9). The differences in arrival volumes and medical urgencies among ESI 2 (high urgency-low volume), ESI 3 (medium urgency-high volume), and ESI 4 (low urgency-medium volume) populations complicate the *routing* decisions. Yet, this variety also provides us with an opportunity to analyze the relationship between *staffing-routing* decisions and system dynamics as we consider various multi-class scenarios in this section.

Table 2.4: Parameters for the Single-Class Cases with ESI 2, ESI 3 and ESI 4 Populations

Patient Class	Population Characteristics	Arrival Volume	Treatment Station 1	Treatment Station 2	Treatment Station 3
ESI 2	Low-Volume-High-Urgency	29.3 arrivals/day	0.25 hours - 5%	0.50 hours - 10%	0.75 hours - 15%
ESI 3	High-Volume-Medium-Urgency	117.3 arrivals/day	0.50 hours - 10%	0.75 hours - 15%	1.00 hours - 20%
ESI 4	Medium-Volume-Low-Urgency	46.2 arrivals/day	0.75 hours - 15%	1.00 hours - 20%	1.25 hours - 25%

Single (n=1) Patient Class: Treating different patient populations separately, we first consider various single-class cases scenarios exhibiting different arrival patterns, service targets, and network complexity (Table 2.4). We use a m-stage network ($1 \leq m \leq 3$), consisting of m treatment and m-1 order bundle stations, with a static routing rule prioritizing an earlier treatment stage.

The *ED-NoAb* staffing meets the TPoD targets for all three single-class cases (ESI 2, ESI 3 and ESI 4) in all of the m-stage network setups considered (m=1, 2, 3). For ESI 3 and ESI 4, the *ED-NoAb* staffing also achieves the maximum server utilization in all scenarios, where the reduced staffing FLOOR[0.90(ED-NoAb)] fails to stabilize the TPoD. For ESI 2, the reduced staffing satisfies the TPoD constraints for single-stage queue (m=1) and two-stage network (m=2) but does not meet the TPoD targets for the three-stage (m=3) network (Figure 2.10).

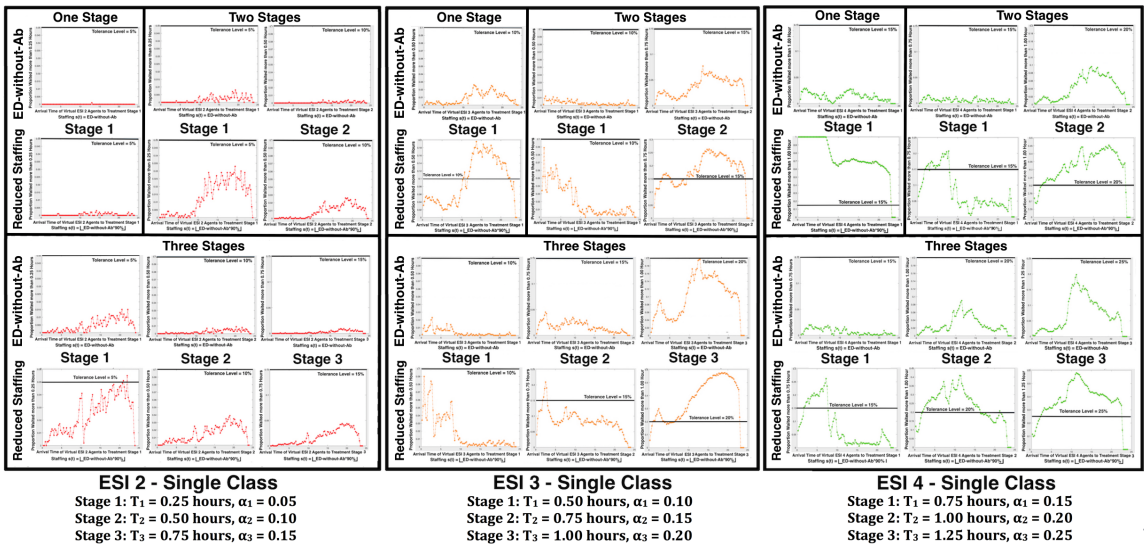


Figure 2.10: TPoD Graphs for Single-Class Cases with ESI 2, ESI 3 and ESI 4 and 1-, 2-, and 3-Stage Networks - Reduced Staffing = 0.90(ED-NoAb)

Based on the results, we make several observations: We derive the *ED-NoAb* staffing formula s_t=m_t+$\beta_{T,\alpha}$ under the heavy-traffic conditions, where the ratio of the increased staffing $\beta_{T,\alpha}$ to OL m_t, R:= $\beta_{T,\alpha}/m_t$, converges to 0. Yet, for the single-station queue with ESI 2 arrivals, low volume and high medical urgency, respectively affecting m_t and $\beta_{T,\alpha}$, cause R to be significantly above zero (R = 9.8 for ESI 2, 0.94 for ESI 3, and 1.62 for ESI 4, assuming a constant rate λ to simplify the calculations). Deviating from the limit conditions, this high ratio of excess staffing to OL leads to overstaffing for the single-station queue with ESI 2 arrivals. The overstaffing disappears when we consider a realistic ER setting, consisting of (possibly) m=3 service stages, or the cases (i.e., ESI 3 or ESI 4) leading to a low R. These findings suggest that (i) the ratio of the increased staffing to OL (rather than just low patient volume), and (ii) system complexity can be used together to predict when *ED-NoAb* might not be the minimum staffing achieving the TPoD targets.

Multi-Class ($n\geq$2) Patients Receiving One (m=1) or Two (m=2) Treatments: Next, we examine single- and two-stage treatment scenarios with a heterogeneous population consisting of two (ESI 2-ESI 3, ESI 2-ESI 4, or ESI 3-ESI 4) or three classes. We use the "class-based routing" rule, prioritizing the class with higher urgency (at an earlier treatment

stage). In all scenarios we consider, *ED-NoAb* succeeds and the *reduced* staffing 0.90(ED-NoAb) fails to meet the TPoD targets (Figure 2.11). Further, under the *reduced* staffing, the TPoD goal with the highest priority, the one for the most urgent class at the initial stage, is violated in all two- and three-class cases with ESI 2 (and almost violated in other cases). This observation is important as it indicates that the performance of the *reduced* staffing cannot be improved by changing the routing policy. Then, for the cases considered, we can conclude that the staffing levels proposed by *ED-NoAb* are indeed the minimum meeting the TPoD targets, and hence, are the maximizer for the server utilization (under TPoD constraints).

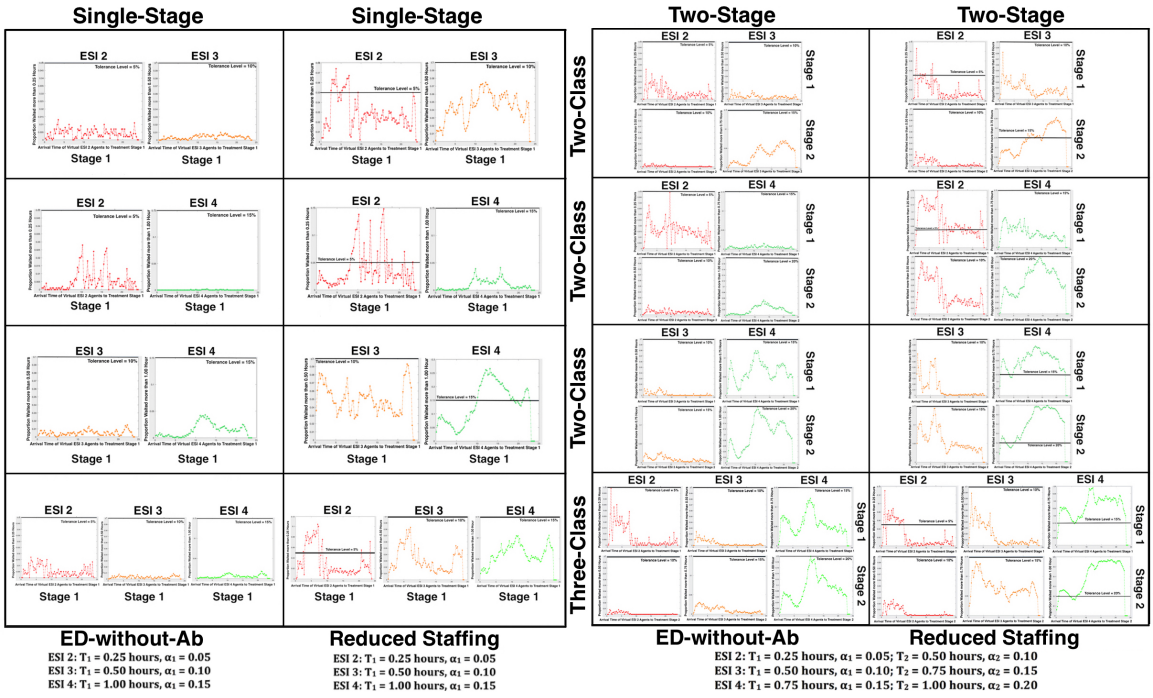


Figure 2.11: TPoD Graphs for Single- and Two-Stage Networks with Multiple ($2 \leq n \leq 3$) Patient Classes - Reduced Staffing = 0.90(ED-NoAb)

ERs with High Complexity- The Need for Dynamic and Hybrid Routing: Finally, we examine multi-class multi-stage cases with high system complexity, where patient population is heterogeneous ($n \geq 2$) and each patient might need, up to, m=3 treatments before departure. In particular, we use a network of m=3 treatment and m-1=2 order bundle queues to model the ER care delivery in these more realistic cases, and consider three two-

class (n=2) and one three-class (n=3) scenarios with the combinations of ESI 2, ESI 3 and ESI 4 populations.

• **Two-Class Three-Stage Cases with ESI 2:** *ED-NoAb* staffing meets TPoD targets under the default "class-based" routing rule for the two-class three-stage scenarios with ESI 2 - ESI 3 and ESI 2 - ESI 4 arrivals. Further, the levels proposed by *ED-NoAb* is the minimum stabilizing all TPoD goals as half of the TPoD constraints, including the one with the highest priority (ESI 2 Stage 1), is violated under the *reduced* staffing in both cases (Figure 2.12).

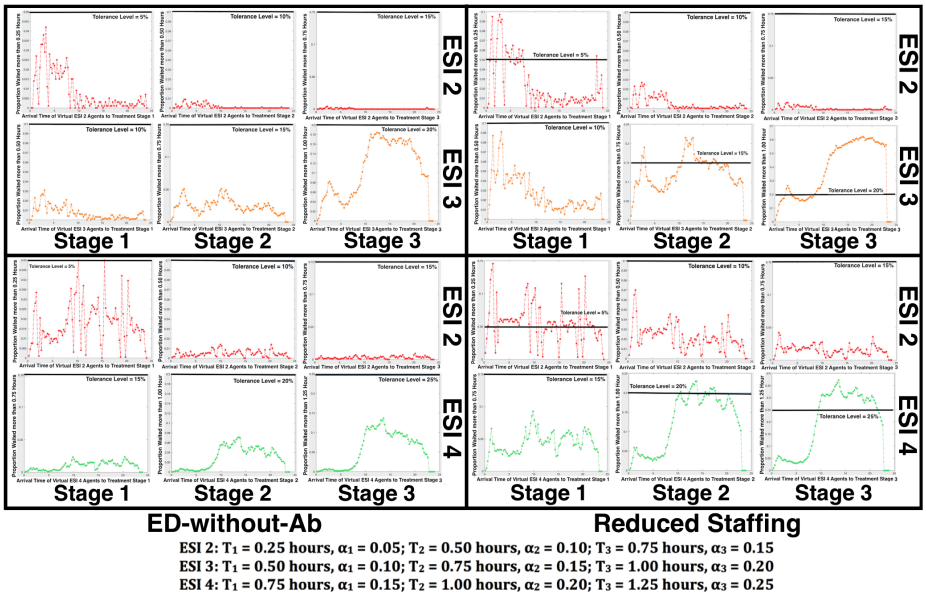


Figure 2.12: TPoD Graphs for Two-Class Three-Stage Cases with ESI 2 - ESI 3 and ESI 2 - ESI 4 Arrivals - Reduced Staffing = 0.90(ED-NoAb)

• **Two-Class Three-Stage Case with ESI 3 - ESI 4:** Under the dynamic routing rules "Shortest-or-Violated-Deadline (SoVD)" or "Maximum-Waiting-Ratio (MWR)", *ED-NoAb* attains all TPoD goals, whereas the *reduced* staffing fails, in the two-class three-stage case with ESI 3 - ESI 4 arrivals (Figure 2.13). On the contrary, none of the static routing rules meets all TPoD targets with the staffing levels of *ED-NoAb* (Figure B.1).

• **Three-Class Three-Stage Case:** For the three-class three-stage case, no dynamic or static rules (that we consider) provides a routing mechanism stabilizing all TPoD goals

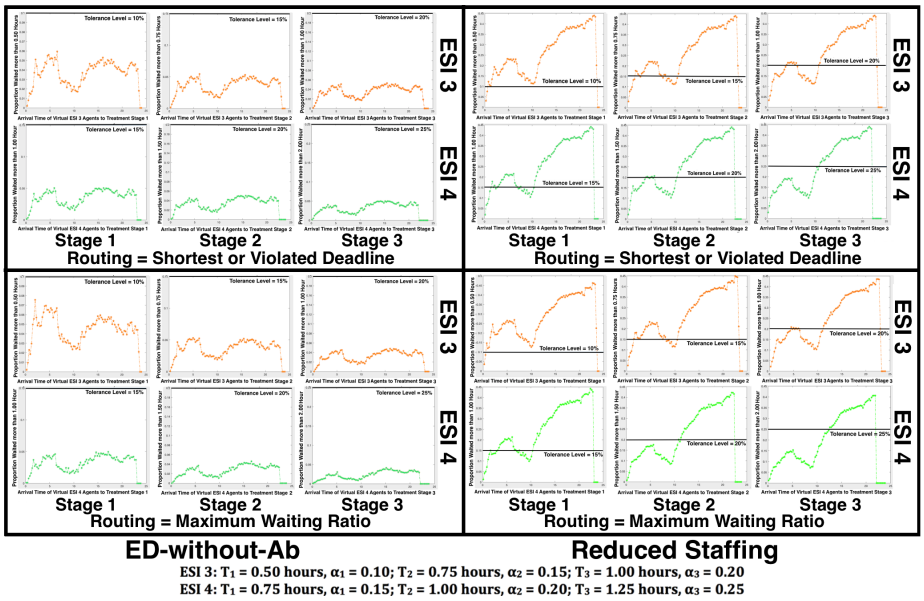


Figure 2.13: TPoD Graphs for Two-Class Three-Stage Case with ESI 3 - ESI 4 Patients and Dynamic Routing - Reduced Staffing = 0.90(ED-NoAb)

with *ED-NoAb* (Figures B.3 and B.4). This motivates us to develop *hybrid* routing rules that combine dynamic and static routing policies. We start with the observation that the only unmet TPoD target under the dynamic rule "SoVD" is the one for ESI 2 patients at the first treatment stage. Based on this observation, we build a *hybrid* routing policy that prioritizes ESI 2 patients at the first treatment and applies "SoVD" for other queues or patient classes. We refer this rule "*SoVD with ESI-2-Stage-1 Priority*". Based on "SoVD", we also design two alternative *hybrid* routing policies, "*SoVD with ESI 2 Priority*" and "*SoVD with Stage 1 Priority*", where the former prioritizes ESI 2 patients (at all treatment stages) and the latter prioritizes the first treatment queue (for all patient classes).

"*SoVD with ESI-2-Stage-1 Priority*" and "*SoVD with ESI 2 Priority*" successfully stabilize all TPoD targets with *ED-NoAb*, where *ED-NoAb* proposes the minimum staffing that meets all TPoD targets for both policies (Figure B.2). On the other hand, "*SoVD with Stage 1 Priority*" cannot satisfy all targets due to the unmet TPoD goal for ESI 2 at Stage 2 (Figure 2.14).

Based the numerical results, we briefly discuss the static, dynamic and hybrid routing

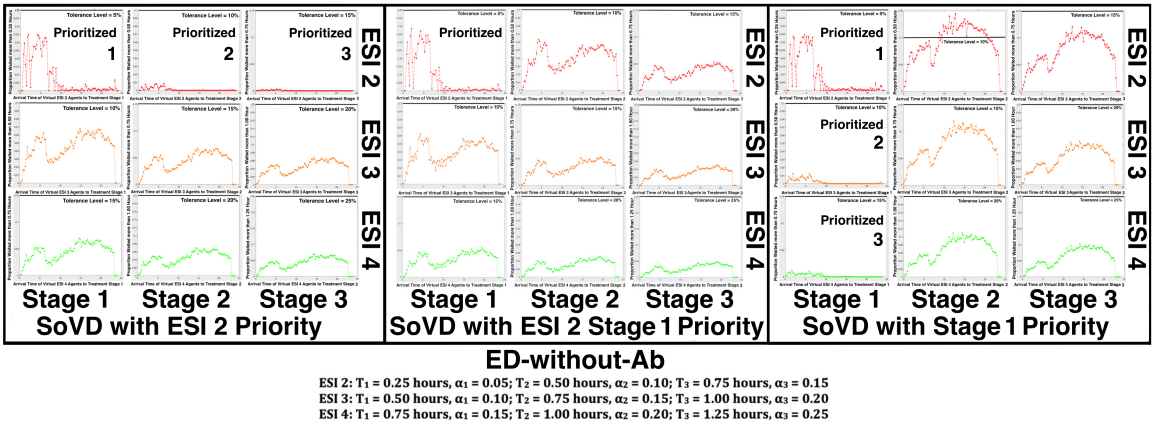


Figure 2.14: TPoD Graphs for Three-Stage Network with Three Patient Classes under Hybrid Routing Rules - Staffing: s(t) = ED-NoAb

rules:

(1) In all scenarios with NoQ<6, where "NoQ"=$n * m$ denotes the number of queues competing for service, the default static rule "class-based" meets all TPoD goals (Table 2.5). Yet, the cases with NoQ≥6, which we refer as the cases with high system complexity, pose challenges for static routing policies. In particular, no static rule meets TPoD targets in three-class three-stage case (NoQ>6) and two-class three-stage case with ESI 3 - ESI 4 arrivals (NoQ=6), and the TPoD targets are barely stabilized in other scenarios with NoQ=6 under the default static rule "class-based".

(2) For the cases with NoQ≥6, dynamic routing rules are needed especially if NoQ>6 or the daily volume of the class with the tightest targets (i.e., highest medical urgency) is large. A notable instance is the two-class three-stage case with ESI 3 - ESI 4 arrivals, where the number of routing decisions to make NoQ=6 and the average daily volume of the most critical class (λ=117.3 arrivals/day for ESI 3) is large. In this scenario, the dynamic rules "SoVD" and "MWR" both meet all TPoD targets under *ED-NoAb* staffing while all static routing rules fail.

(3) As shown by the three-class three-stage case (NoQ=9), the use of dynamic routing rules alone might not always be sufficient to cope with high system complexity of medium-to-large scale ERs. To improve the system performance and address the routing problem in

Table 2.5: Multi-Class Multi-Stage Numerical Experiments Summary

Case	Patient Populations	NoQ	ED-NoAb	Routing Policy
Single-Class Single-Stage	ESI 2	1	Not Minimum	Class-based
Single-Class Single-Stage	ESI 3	1	Minimum	Class-based
Single-Class Single-Stage	ESI 4	1	Minimum	Class-based
Single-Class Two-Stage	ESI 2	2	Minimum	Class-based
Single-Class Two-Stage	ESI 3	2	Minimum	Class-based
Single-Class Two-Stage	ESI 4	2	Minimum	Class-based
Single-Class Three-Stage	ESI 2	3	Minimum	Class-based
Single-Class Three-Stage	ESI 3	3	Minimum	Class-based
Single-Class Three-Stage	ESI 4	3	Minimum	Class-based
Two-Class Single-Stage	ESI 2 - ESI 3	2	Minimum	Class-based
Two-Class Single-Stage	ESI 2 - ESI 4	2	Minimum	Class-based
Two-Class Single-Stage	ESI 3 - ESI 4	2	Minimum	Class-based
Two-Class Two-Stage	ESI 2 - ESI 3	4	Minimum	Class-based
Two-Class Two-Stage	ESI 2 - ESI 4	4	Minimum	Class-based
Two-Class Two-Stage	ESI 3 - ESI 4	4	Minimum	Class-based
Two-Class Three-Stage	ESI 2 - ESI 3	6	Minimum	Class-based
Two-Class Three-Stage	ESI 2 - ESI 4	6	Minimum	Class-based
Two-Class Three-Stage	ESI 3 - ESI 4	6	Minimum	SoVD, MVR
Three-Class Single-Stage	ESI 2 - ESI 3 - ESI 4	3	Minimum	Class-based
Three-Class Two-Stage	ESI 2 - ESI 3 - ESI 4	6	Minimum	Class-based
Three-Class Three-Stage	ESI 2 - ESI 3 - ESI 4	9	Minimum	SoVD with ESI-2-Stage-1 or ESI 2 Priority

real ERs, one alternative is to design hybrid routing rules by integrating managerial insight in the form of static rules with dynamic rules that account for current system state. For example, "*SoVD with ESI-2-Stage-1 Priority*" gives a static priority to the queue with the tightest target (ESI 2 Stage 1), which is violated when SoVD is implemented alone, and stabilizes all TPoD targets in the three-class three-stage case, which is the most realistic case we consider in this study. Yet, the hybrid rules should be tailored carefully since giving a pre-determined (static) priority to a queue improves its performance at the expense of others. Specifically, we construct two other hybrid rules, "*SoVD with ESI 2 Priority*" and "*SoVD with Stage 1 Priority*", by respectively extending the pre-determined priority to all ESI 2 patients or all classes at Stage 1. Yet, the impact of prioritization is not the same as the OLs of ESI 2 patients at Stage 2 and Stage 3, prioritized under "*SoVD with ESI 2 Priority*", are respectively low and medium, whereas the loads of ESI 3 and ESI 4 patients at Stage 1, prioritized under "*SoVD with Stage 1 Priority*", are relatively high and medium, respectively (Figure 2.15). As a result, the former rule causes minor disturbances for non-prioritized queues and meets all TPoD targets but the latter results in delays for non-prioritized queues that are large enough to violate a TPoD target.

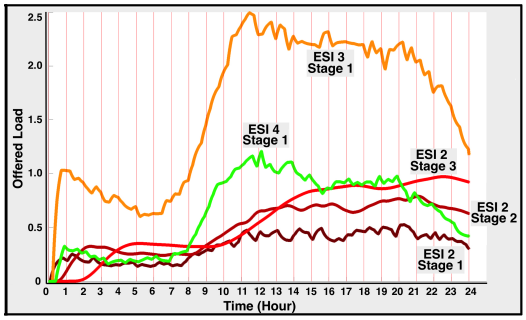


Figure 2.15: Three-Class Three-Stage Case OL Functions for Treatment Stage 1 or ESI 2 Patients

Discussion on High-Urgency ESI 1 Patients: ESI 1 patients, representing patients with the highest urgency, make up a very small portion ($\approx 1\%$ in our dataset) of the total arrivals. However, they might have a significant impact on the utilization of ER resources as their life-threatening conditions require immediate medical intervention and might take longer to be stabilized. The OL and staffing calculations for ESI 1 patients, requiring minor changes in our approach, are as follows: First, we make two modifications in the network model by removing triage (for ambulance arrivals), and replacing bed placement queue with "treatment stage 0" (operating room) queue, served by ER physicians. Second, we use the SRS staffing formula [152] for treatment stages 0 and 1, as these are QED queues with PoD type service targets. The rest of the ER care delivery process and staffing calculations are the same as other patient classes.

Despite the importance of ESI 1 patients, an ER manager might choose to exclude them in the calculation of staffing levels due to two reasons: (1) ESI 1 arrivals are rare events without a clear predictable pattern, where the historical data is not necessarily representative of the underlying stochastic process. Accordingly, it is unclear at what time of the day the number of ER physicians should systematically be increased so that an ER copes with an ESI 1 arrival without affecting the treatment of other patients. (2) Given the high salary cost of ER physicians, increasing the staffing levels of ER physicians to account for rare and unpredictable ESI 1 arrival(s) is not an cost-effective approach. Adjusting the staffing levels with SRS formula even for limited portions of the day could still be costly since PoD

type service targets, reflecting the urgency of ESI 1 cases, require higher staffing levels so that earlier treatment stages operate in QED mode for ESI 1 patients.

Based on our discussions with ER administrators, we conclude that a practical and cost-effective solution that can be implemented in ERs is to make staffing decisions excluding rare ESI 1 cases, and then allocate all resources necessary when an ESI 1 patient arrives at the expense of violating some of the TPoD targets for other patients temporarily. This solution is based on the fact that TPoD targets are "*soft (non-binding)*" constraints, determined by a managerial board as ideal goals rather than enforced by a federal agency, and is in line with the current practice.

2.6 Discussions and Conclusion

ERs are complex healthcare delivery systems, expected to provide service in a timely and cost-effective manner. Workforce planning of ER physicians, the most critical and expensive human resources in ERs, plays a pivotal role in determining ERs' operational efficacy and efficiency. In this paper, we study a *staffing* and an associated *routing* problem for ER physicians, who serve a multi-class patient population in a repetitive service setting under time-varying demand. Our objective is to determine the minimum number of ER physicians that are needed to meet TPoD type service targets for all patient classes and at all treatment stages at any given time.

2.6.1 The MCMS Network and A New Staffing Algorithm

We propose a new staffing approach based on a practical and intuitive queuing network model. Representing the patient flow in the ER via tandem queues, the network model opens the black-box of ER care and captures the key aspects of ER service. Utilizing this model with *DIS* assumption, we compute the OLs for treatment queues of the network, which are served by ER physicians and do not experience abandonments despite operating in ED regime. Motivated by these unconventional ED queues, we develop a new algo-

rithm, "*ED-NoAb*", that translates OL into staffing for ED queues with perfectly patient customers and TPoD type service targets. We analytically show that *ED-NoAb* asymptotically stabilizes TPoD for stationary M/M/s queues operating in ED regime and we numerically demonstrate its robustness in time-varying settings with $M_t/G/s_t$ queues coping with periodically fluctuating demand. Finally, we generalize our approach to multi-class multi-stage ER settings, where we also capture the *pooling* effect, and perform a comprehensive analysis with real ER data by considering various network setups with different patient populations under practical *routing* rules. Our analysis shows that *ED-NoAb* algorithm is effective in meeting differentiated TPoD targets while ensuring high resource utilization for various ER settings.

Our approach addresses the some of the major limitations of black-box models. In particular, the network model we propose is practical and comprehensive enough to capture the key features of the ER care and account for many real life challenges ERs face. Further, by opening the black-box of ER service, it provides more access to and managerial control over the internal components of ER care delivery process. Utilizing the network model, an ER manager can set differentiated service goals at each treatment stage for each patient class, estimate the load on the ER physicians at various treatment stages for any given time, and make staffing decisions based on these factors accordingly. In addition, despite the highly complex nature of ER care delivery process, the network model is intuitive enough to gain insights and implement in practice. In this study, we focus on queues operate with an *efficiency-driven* philosophy, aiming to maximize server utilization while keeping waiting times at acceptable levels, and experience no significant abandonments but our approach can also be used for other complex service systems.

Despite the practicality our approach, we should note that an ER manager might choose to use our staffing solution as a reference rather than directly implementing them. This is because TPoD targets are *non-binding* service goals and the managerial board might have other incentives and concerns. For instance, an ER manager might prefer lower staffing lev-

els at the expense of violating certain TPoD goals sometimes, as strictly satisfying all TPoD targets might not be cost-effective. Another implementation-related issue is the length of intervals over which the staffing levels remain constant. In our analyses, we allow changes in the number of ER physicians in every 15 minutes to be able to accurately monitor the performance of our staffing solutions. Yet, in practice, ER physician shifts may start only at pre-determined times such as 30-minute or 1-hour intervals. In such cases, the staffing levels should be set the maximum number of ER physicians over the no-change interval to avoid any TPoD violation.

2.6.2 Practical Routing Rules and The Role of the System Complexity

Besides the *staffing* problem, we also study an associated *routing* problem and analyze the relationship between *staffing-routing* decisions and system complexity. Our numerical results show that dynamic and, eventually, hybrid routing rules become necessary as the system complexity increases with increasing number of treatment stages and patient classes, and daily arrival volumes. The dynamic routing rule "SoVD", directing a physician to the patient with the (most or) almost violated delay threshold, is particularly important for our solutions since it serves as the backbone for the proposed hybrid *routing* policies. A challenge for "SoVD" or any other dynamic routing policy, in terms of implementation, is the requirement of keeping track of waiting times, possibly using a real-time location services (RTLS) [172]. Yet, our discussions with ER managers reveal that the "SoVD" rule can be still implemented in ERs (to a certain extent) without the use the RTLS technology. Specifically, the waiting time of a patient for the next treatment can roughly be estimated based on the arrival time of the last medical order in the electronic medical records and then this estimate can then be used to implement an approximate version of "SoVD".

Another interesting result of our study is the failure of "*SoVD with Stage 1 Priority*" rule to meet TPoD. This is important as routing rules similar to "*SoVD with Stage 1 Priority*", prioritizing the initial treatment stage after triage for all patient classes, are quite

common in practice. The routing rules prioritizing the initial treatment are commonly used in many ERs partly because of the critical metric "time to initial service", affecting both performance and financial evaluations. "Time to initial service" (*Door to Diagnostic Evaluation by a Qualified Medical Professional*) is one of the key metrics, based on which an ER's performance is assessed, and further, is required to be reported the Centers for Medicare and Medicaid Services (CMS) for the receipt of reimbursements. Accordingly, the improvements efforts in ERs are primarily focused on the first patient-physician interaction such as implementing the routing mechanisms that reduce the waiting time before the first treatment. In this study, we propose a more comprehensive approach, where service goals are determined for each treatment stage and all stages are accounted for when staffing decisions are made, and show the sub-optimal of "*SoVD with Stage 1 Priority*" routing policy. Yet, higher level interventions (e.g. at federal policy, reimbursement and metric-reporting levels) might be required to create incentives for ERs focus on all treatment stages (and record and report related statistics) as they make their staffing and routing decisions.

2.6.3 Conclusion

As stated by [171], it is challenging to capture all prevalent features of the ER care delivery process with a single analytic model given the complexity of ER operations. Our study is no exception and has the following limitations: First, we assume a constant service rate for the queues of the network model whereas, in reality, the service rate might server specific [173] and depend on the current load on the system [170]. We limit the scope of this study with the performance of the ER under regular conditions and do not consider the rare but very critical conditions such as a mass casualty event. We do not explicitly account for the interruption of service due to equipment malfunction, lost document, or other causes [174]. We assume the same skill level among servers (ER physicians) in this study. Other approaches (e.g., see [175] and [176]) might be more appropriate for the service systems where the skills of the servers differentiate significantly. We assume there are sufficient

physical resources (e.g., total number of beds) available and hence, access to the ER might be delayed, causing some abandonments, but never blocked. This assumption may not hold for small-sized ERs. The misclassification of patient classes, analyzed by [177], is not considered in this study. Finally, we do not impose the "*same patient-same staff*" rule, where ideally each patient is treated by the same ER team throughout her ER stay.

To summarize, in collaboration with Mayo Clinic, we study a staffing problem for ER physicians. The problem features we consider are time-varying demand, multi-class patient population, differentiated service targets, and discontinuous, treatment-history dependent, and multi-stage service provided by centralized servers. We develop a queuing network model that captures the patient flow in the ER, utilize it to compute the workload on ER physicians, and propose a new staffing algorithm that translates the load on ER physicians into staffing decisions. Using real ER data, we show the effectiveness of our staffing approach in various ER settings, and demonstrate the need for dynamic and, eventually, hybrid routing rules as the system complexity increases.

CHAPTER 3
MULTI-STATE SURVIVAL ANALYSIS FOR CLINICAL DECISION-MAKING: APPLICATIONS TO FOLLICULAR LYMPHOMA AND DIFFUSE LARGE B CELL LYMPHOMA

3.1 Background

Diffuse large B-cell lymphoma (DLBCL) is the most common type of non-Hodgkin lymphoma (NHL) in the United States, accounting for about one out of every four lymphomas. DLBCL tends to be a fast-growing (aggressive) lymphoma, requiring rapid initiation of a medical therapy, but it often responds well to treatment and might be cured. About one out of five lymphomas in the United States is a follicular lymphoma (FL). Contrary to DLBCL, FL is usually a slow-growing (indolent) lymphoma, and also often responds well to treatment. However, follicular lymphomas are hard to cure, where patients with FL usually experience multiple remission-relapse cycles over the course of this disease.

In this chapter, we present our survival analysis studies, investigating the clinical courses of DLBCL and FL after initial treatment over time as a function of socio-demographic and clinical factors. Our main objectives are (1) to investigate optimal treatment sequences and regimens for older high-risk FL patients, (2) to identify the predictors affecting the course of FL for older high-risk population, (3) to identify factors influencing the course of DLBCL and to quantify their impact on DLBCL-specific mortality in patients older than age 65, (4) to determine the optimal stopping time to end the surveillance of older DLBCL patients for their lymphoma recurrence-related death risk.

3.2 Assessing the Effectiveness of Treatment Sequences for Follicular Lymphoma Patients with a Multi-state Model

3.2.1 Introduction

Follicular lymphoma (FL) is the most common indolent lymphoma and the second most common non-Hodgkin lymphoma accounting for about 20-25% of all lymphomas in Western countries [178] [179]. For most patients, FL is an incurable disease characterized by an indolent behavior with an initial period of observation followed by favorable response to initial therapy. Most newly diagnosed FL patients treated with rituximab (R) alone or R-chemotherapy enjoy prolonged progression free survival (PFS) and overall survival (OS) but some patients with FL undergo transformation to a more aggressive histology [180] and approximately 20% experience progression of disease (PD) within two years of first-line chemoimmunotherapy irrespective of treatment choice [181].

Large randomized studies have consistently shown FL patients with early PD have poorer OS compared to those patients who did not relapse within two years [182] [183] [184]. Consistent with these trials, an analysis of data from the National LymphoCare cohort study (NLCS) in the United States involving 588 patients with stage II-IV FL treated with first-line R-chemotherapy, showed that 19% relapsed within two years of diagnosis [181]. Importantly, OS was markedly reduced in the early progression group, with a 5-year survival rate of 50% from the 2-year risk-defining progression event compared with 90% for patients without early progression. Validation of these data in an independent cohort of FL patients for Iowa/Mayo confirmed poor 5-year OS in early relapsing patients [181]. Along with early PD, advanced age also has been identified as a high-risk factor for poor clinical course of FL and worse survival. In particular, age $>$ 60 years has been shown to be a key adverse prognostic factor associated with poorer OS and PFS and is a component of the follicular lymphoma international prognostic index [185] [186].

Limited data exist regarding the outcomes associated with the sequencing of first, sec-

ond, and third line therapy in FL, which remains important since most patients relapse and require sequential treatment. Moreover, no study exists regarding the impact of sequential therapies among the most vulnerable FL patients who are > 60 years and require a second treatment within two years of initial therapy. We developed a continuous-time multi-state model to capture the clinical course of FL for older high-risk patients, and conducted a multi-state survival analysis (1) to examine the outcomes associated with first-, second-, and third-line therapies and (2) to assess the effects of socio-demographic and clinical factors on outcomes at each line and OS.

3.2.2 Methods

Our multi-state model consists of three treatment states, alive after first- (TX1), second- (TX2), and third-line treatment (TX3); and an absorbing health state"Dead" (Figure 3.1). FL patients enter a treatment state with the initiation of their corresponding treatment and depart either when they die or when their next treatment is initiated. We used the Aalen-Johansen estimator [187], a generalization of Kaplan-Meier (KM) estimator [188], to assess the likelihood of being in one of the four clinical states at a given time. The AJ estimator is a convenient and reliable (nearly unbiased) nonparametric estimator for multi-state models, does not assume any form on probability distributions, and can cope with censored observations that exist in clinical data. It was mathematically shown that the Aalen-Johansen estimator is the maximum likelihood estimator of multi-state models [189] [190], and provides consistent estimates both for Markov and non-Markov models [191].

3.2.3 Patients, Data Source and Variables

Data from the Surveillance, Epidemiology, and End Results (SEER) 2000-2009 registry, linked with Medicare claims data through 2011, were used to identify patients with a histologically confirmed first primary diagnosis of FL based on the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) histology codes [192]. The SEER

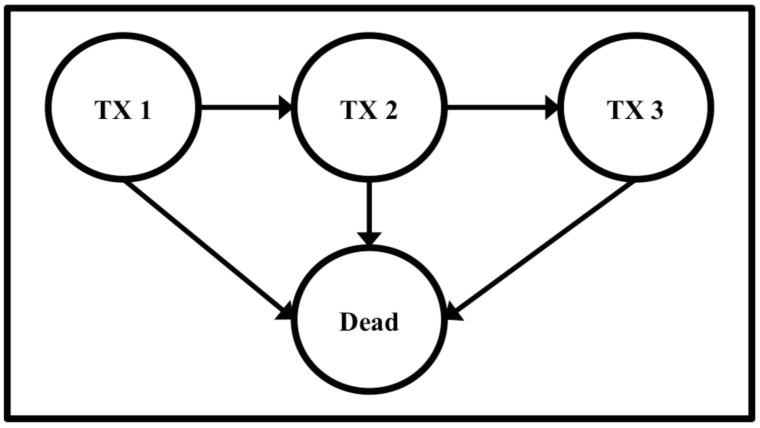


Figure 3.1: A Multistate Model for the Clinical Course of Follicular Lymphoma

program is a National Cancer Institute sponsored epidemiologic surveillance system of population-based tumor registries that routinely seek to collect demographic and clinical information on all incident cases occurring in SEER areas [193]. Medicare is the primary health insurer for 93% of the United States population aged 65 years and older. Medicare claims data contain information collected to cover health care services provided to Medicare beneficiaries. Since the median age at diagnosis for patients with FL is $>$ 65 years, the SEER-Medicare data offer a valuable source to examine FL patterns of care and outcomes.

The SEER-Medicare dataset included a total of 8411 FL patients $\geq$ 65 years. After our examination for eligibility, 3177 patients were excluded for the following reasons: insufficient data (n = 676), diagnosis before year 2000 (n = 1342), no received treatment after diagnosis (n = 821), and progression more than 2 years after initial treatment (n = 338) (Figure 3.2). All 5234 patients included in this analysis (1) had an advanced age $>$ 65 years, a high-risk factor for poor PFS and OS, (2) received at least a first-line therapy after being diagnosed in the era of routine first-line rituximab and R-chemotherapy ($>$ year 2000) and (3) either had an early PD $<$ 2 years after first-line therapy (71%), a risk factor for poorer OS, or experience median OS $<$ 3 years (29%).

Starting with their diagnosis, patients were observed until death or end of follow-up (12/31/2009 for SEER and 12/31/2011 for Medicare data). We used Medicare as the primary source for time of death information and adjusted the missing entries via SEER data,

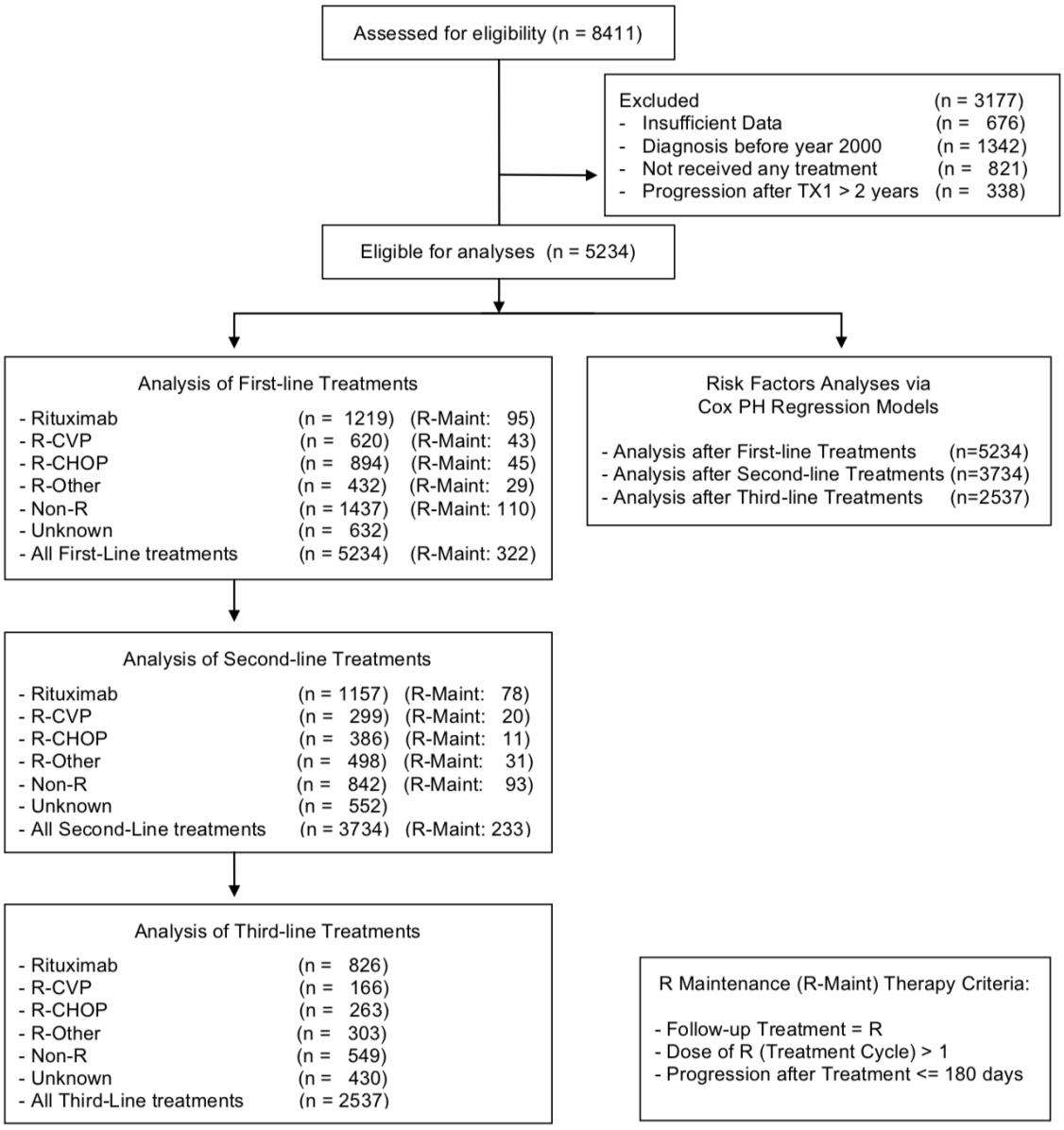


Figure 3.2: Consort Flow Diagram Reporting the Number of Patients in Each Analysis

assuming the 15th day of the month as the time of death. The primary variables of interest were the receipt of first-, second- and third-line treatment, and OS under these treatment conditions. We identified FL-directed treatment strategies using ICD, 9th Revision, Clinical Modification (ICD-9-CM) Diagnosis; ICD-9-CM Procedural; Current Procedural Terminology (CPT); HCPCS; and revenue centers codes [193] on inpatient, outpatient and physician claims. We grouped treatments into five main categories as follows: (1) R with or without radiation, (2) R-cyclophosphamide and vincristine (R-CVP), (3) R- cyclophosphamide, hydroxydaunorubicin, and vincristine (R-CHOP), (4) R with other chemotherapy combinations (R-Other) and (5) treatments that did not contain R (Non-R). Data also included patients with unknown treatments. When we investigated the optimal first-, second- and third-line therapies, we presented the results for the treatment groups R, R-CVP, R-CHOP and R-Other. When we examined the impact of clinical and socio-demographic risk factors on the clinical course of FL, we considered all patients, including patients with Non-R and unknown treatments. To adjust for the use of R maintenance therapy following induction, we classified a treatment as involving R-maintenance if R-alone was received within 180 days of the previous treatment and more than one dose of R treatment was given (Figure 3.2).

3.2.4 Results

We assessed 8411 patients in the SEER-Medicare data for eligibility and included 5234 patients, diagnosed between 2000 and 2009, in this study (Figure 3.2): 71% of these patients received a second-line therapy within two years, and 29% received no further therapy after their initial therapy and experienced median OS less than three years. All patients were above age 65, the mean age was 76 years and median duration of time from diagnosis to first-line treatment was 1.37 months (95% Confidence Intervals (CI): 1.33-1.43). Table C.1 (see Appendix C) displays the baseline patient characteristics by treatment type.

In this FL population enriched for patients with early progression, 1219 received R

alone, 620 R-CVP, 894 R-CHOP, and 432 R-Other as first line therapy and 322 patients received R maintenance therapy after first-line treatment. R-CHOP achieved the highest survival rates at two and five years among the first-line treatments, which were: 76% and 46% for R, 82% and 55% for RCVP, 84% and 71% for RCHOP, and 74% and 54% for R-Other, respectively (Figure 3.3). RCHOP in second- and third-line therapies also were associated with the most favorable five-year OS rates after the initiation of second- and third-line treatments, which were: 50% and 49% for R, 47% and 38% for R-CVP, 55% and 61% for R-CHOP, and 47% and 46% for R-Other as second- and third-line therapies, respectively (see Figure C.1 and C.2 in Appendices). Consistent with these findings, median OS from the initiation of corresponding treatment was the highest for RCHOP patients at any treatment line and was 36.3 months (95% CI: 33.7, 38.7) for first-line R-CHOP, 34.1 months (95% CI: 28.6, 37.9) for second-line R-CHOP, and 40.4 months (95% CI: 33.5, 45.7) for third-line R-CHOP therapy (see Table 3.1).

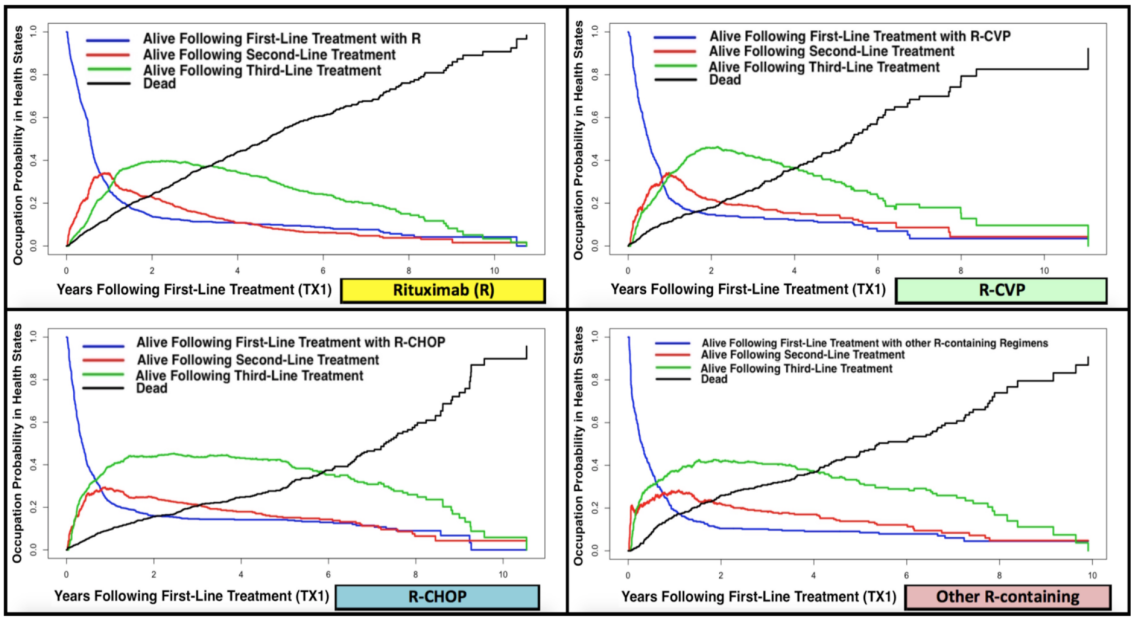


Figure 3.3: Graph of Occupation Probabilities over Time by First-Line Treatments

RCHOP had the greatest impact on OS when provided as the first-line therapy. For patients receiving first-line R-CHOP and then two other therapies (R-CHOP + X + X), the mortality rate at two years after the corresponding line of therapy was 10.9%, 13.8%,

Table 3.1: Median Overall Survival (OS) Duration from the Initiation of Treatment

Overall Survival from Tx1			Overall Survival from Tx2			Overall Survival from Tx3		
First-Line Treatment	Number of Patients	Months (95% CI)	Second-Line Treatment	Number of Patients	Months (95% CI)	Third-Line Treatment	Number of Patients	Months (95% CI)
Rituximab (R)	1219	33.3 (31.1,35.1)	Rituximab (R)	1157	29.6 (27.9,31.7)	Rituximab (R)	826	25.2 (23.3,27.2)
R-CVP	620	30.7 (28.4,32.1)	R-CVP	299	27.4 (24.0,30.7)	R-CVP	166	26.0 (23.4,30.5)
R-CHOP	894	36.3 (33.7,38.7)	R-CHOP	386	34.1 (28.6,37.9)	R-CHOP	263	40.4 (33.3,45.7)
R-Other	432	34.8 (30.2,39.8)	R-Other	498	28.9 (26.5,33.3)	R-Other	303	28.9 (24.7,35.2)
All First-line	5234	33.3 (32.2,34.3)	All Second-line	3734	29.6 (28.5,30.8)	All Third-line	2537	25,8 (24.8,26.8)
Overall Survival: Time to Last Follow-Up or Death			Tx: Treatment					

and 17.8% when this statistics collected after the initiation of first-, second- and third-line treatments, respectively (Table 3.2). The corresponding mortality rates for second-line R-CHOP (X + R-CHOP + X) were 18.0%, 20.8% and 23.7%, and for third-line R-CHOP (X + X + R-CHOP) were 25.5%, 31.1% and 34.0% when R-CHOP was combined with any other two regimens. Over the course of sequential treatments, first-line R-CVP followed by R-CHOP also was a particularly favorable combination with no deaths within 24 months of initiating first line therapy (Table 3.3). Patients who received R-Other as first-line therapy in this cohort had poorer outcomes with all second line regimens if they received a (subsequent) second-line therapy (Table 3.3).

Table 3.2: The Impact of R-CHOP on Mortality as a First-, Second- or Third-line Therapy

Treatment Sequences where R-CHOP is Provided as First-, Second- or Third-line Therapy			Number of Patients	Percentage (%) of Deaths within 24 months			Total % of Deaths
TX1	TX2	TX3		After TX1	After TX2	After TX3	
R-CHOP	Any Treatment except R-CHOP	Any Treatment except R-CHOP	304	10.9%	13.8%	17.8%	30.9%
Any Treatment except R-CHOP	R-CHOP	Any Treatment except R-CHOP	245	18.0%	20.8%	23.7%	43.7%
Any Treatment except R-CHOP	Any Treatment except R-CHOP	R-CHOP	106	25.5%	31.1%	34.0%	51.9%

Compared to other regimens, first-line R-CHOP was effective across all age subgroups. Among first-line R-CHOP patients, the five-year OS rates were 0.70 for age 66-70, 0.71 for age 71-75, 0.75 for age 76-80, and 0.66 for age $\geq$ 81. On the contrary, the five-year OS rates were lower for patients receiving other R-based regimens including R, R-CVP

Table 3.3: Percentage of Deaths within 24 months of First-line Treatment

First-line Therapy (TX 1)	Rituximab (n= 1219)	R-CVP (n= 620)	R-CHOP (n= 894)	R-Other (n= 432)	All First Line (n= 5234)
Second-line Therapy (TX 2)					
Rituximab (R) (n=1157)	9.7%	5.7%	9.7%	17.2%	12.8%
R-CVP (n=299)	23.4%	21.1%	8.0%	29.0%	23.1%
R-CHOP (n=386)	18.9%	0.0%	17.3%	21.2%	21.2%
R-Other (n=498)	28.1%	14.0%	12.3%	25.9%	22.1%
All Second Line (n=3734)	17.7%	12.1%	11.7%	23.4%	20.4%

or R-Other first-line therapies. For these groups of patients, the five-year OS rates were 0.52 for age 66-70, 0.52 for age 71-75, 0.49 for age 76-80, 0.48 for age ≥ 81. In terms of comorbid conditions, first-line R-CHOP was slightly less effective for patients with CCI $>$ 1, where the five-year OS rates were 0.71 for CCI ≤ 1 and 0.66 for CCI > 1. Not surprisingly, FL stage affected the performance of the first-line R-CHOP, where the five-year OS rate dropped from 0.75 (for stage ≤ 2) to 0.66 when stage > 2. Though moderately, the histology was also influential in the effectiveness of R-CHOP. The five-year OS rates were 0.74 for FL grade 1/2 and 0.70 for grade 3, respectively. The factors affecting outcomes for R-CHOP were not as influential in patients receiving front-line R treatment. The effect of age, comorbidity, stage, grade, and performance status on five-year OS rates among patients receiving R-Only for their first-line therapy were as follows: 0.48 for age 66-70, 0.50 for age 71-75, 0.47 for age 76-80, 0.51 for age ≥ 81; 0.49 for CCI > 1, 0.49 for CCI ≤ 1; 0.48 for FL stage ≤ 2, 0.51 for FL stage > 2; 0.50 for grade 1/2, 0.49 for grade 3; and 0.50 for favorable performance status and 0.48 for unfavorable performance status.

3,734 patients received a second-line therapy and 65% of these patients received a subsequent third-line therapy within two years of their second-line therapy including: 750 R, 198 R-CVP, 286 R-CHOP and 356 R-Other. For these patients, empirical two- and five-year OS rates after second-line therapies were: 83% and 55% for R, 76% and 45% for R-CVP, 77% and 55% for R-CHOP, 75% and 45% for R-Other respectively.

Multivariable Cox regression models revealed that the presence of B-symptoms, being married, and grade 1/2 FL histology were associated with earlier initiation of the second-line therapy and age $>$ 85 years was associated with elevated risk of death after first-line treatment (Table C.2). The Cox models also substantiated that R-CHOP in any-line improved OS by reducing the rate of death after first-, second- and third-line treatments. Compared to the baseline treatment group R, first-line R-CHOP reduced the risk of death after first-line treatment, with hazard ratio (HR) 0.60 (95% CI: 0.47-0.77), and improved OS after any subsequent second- and third-line treatments, with HR 0.40 (95% CI: 0.29-0.53) and 0.63 (95% CI: 0.53-0.76), respectively. Second- and third-line R-CHOP treatments had similar positive impacts on OS: The mortality risk was reduced by second-line R-CHOP with HR 0.61 (95% CI: 0.44-0.84) after second-line and 0.80 (95% CI: 0.66-0.96) after any third-line treatment, and by third-line R-CHOP with HR 0.81 (95% CI: 0.66-1.00).

As might be expected, RCHOP followed by second-line RCHOP yielded unfavorable clinical outcomes. Among all first-line R-CHOP patients, who also received a second-line therapy, two-year mortality rates were: RCHOP followed by R, 9.7%; RCHOP followed by RCVP, 8%; RCHOP followed by RCHOP, 17.3%; and RCHOP followed by R-Other, 12.3% (Table 3.3). Similarly, the repeated use of R-CVP (i.e., R-CVP + R-CVP) yielded adverse outcomes with the highest two-year mortality rate 21.1%, compared to other clinical scenarios, where first-line R-CVP was followed by another second-line regimen: R-CVP + R = 5.7%, R-CVP + R-CHOP = 0%, and R-CVP + R-Other = 14% (Table 3.3). On the contrary, R-CHOP followed by R-CVP and R-CVP followed by R-CHOP both resulted in the low two-year mortality rates, 0% and 8%, respectively (Table 3.3).

3.2.5 Discussion

Currently there is no standard of care for the initial treatment following diagnosis of FL in the United States, where the management options include watchful waiting, single agent therapy, combination chemotherapy with immunotherapy, and radiation [25]. For instance,

the initial first-line strategies among 2728 patients reported by NLCS included: watchful waiting in 17.7%, R alone in 13.9%, a clinical trial in 6.1%, radiation therapy in 5.6%, chemotherapy in 3.2%, and R with chemotherapy in 51.9% [194]. Since that publication, R-bendamustine and maintenance R following chemoimmunotherapy have emerged as other common first line approaches [182] [183] [184], yet no treatment strategy has demonstrated superior outcomes over all alternatives. The impacts of clinical and biological data or functional imaging on FL outcomes have been analyzed in a number of studies [183] [195] [196] [197], and early relapsing patients are significantly more likely than patients without early PD to have high FL International Prognostic Index scores (p=0.007) [181] and to have worse OS. At present, identifying the subset of patients at greatest risk for early PD and the optimal treatment strategies for these patients remain unmet clinical needs [180]. However, limited data exist examining the patterns of care, impact of certain regimens across lines of therapy, or the optimal therapy sequence for patients with FL who experience early PD.

In this population of 5234 patients, enriched for older FL patients with early initiation of second-line therapy, our analysis revealed that R-CHOP at any-line of treatment was associated with the highest OS, and the most favorable impact was achieved when R-CHOP was provided as a first-line therapy (Table C.3). Yet, as might be expected, RCHOP followed by second-line RCHOP yielded unfavorable clinical outcomes. One possible explanation for this is anthracycline toxicity due to doxorubicin. Anthracycline-containing regimens are associated with toxicities including myelosuppression, mucositis, febrile neutropenia, and cardiomyopathy particularly in older patients, like the patients included in our dataset. Older individuals with lymphoma have been shown to be particularly vulnerable to cardiac and hematologic toxicity with anthracycline-based therapies [198] [199]. [29, 30]. A second explanation includes the development of drug resistance upon chemotherapy re-challenge [200]. Studies on the efficacy of anthracycline re-challenge have focused on breast cancer and remains inconclusive [201]. Drug resistance also would explain the unfavorable outcomes seen with repeat use of R-CVP (Table 3.3). The issue of sub-optimal

outcomes with the repetitive use of the same regimen, R-CHOP or R-CVP in particular, might be addressed by using R-CHOP or R-CVP as first- and the other as second-line treatment since such treatment sequences resulted in low two-year mortality rates in our analysis of this SEER-Medicare dataset.

Our study has some limitations. First, we analyzed the impact of first-, second- and third-line FL treatments and other factors on OS; a more in depth analysis could assess the effect of these factors specifically on FL-associated mortality by using the same multi-state framework but distinguishing ?death due to FL? and ?death from other causes? if the data permitted. Second, the population examined was older FL patients and was enriched for high-risk patients. Accordingly, the findings may not be directly generalizable to younger or average-risk patients. Third, novel therapies have been developed for relapsed FL including bendamustine [202], obinotuzumab [203], lenalidomide [204] [205], ibrutinib [206], and idelalisib [207] [208] [209] that were not directly assessed in this analysis of patients diagnosed with FL from 2000 through 2009. While some of these agents were included in the category of "R-Other" treated patients (see Table C.4 in Appendix), these interventions were poorly represented in this dataset, and will likely impact future outcomes.

Other data-related limitations were as follows: 632 (12%) first-line, 552 second-line (15%), and 430 (17%) third-line treatments were unidentified due to data limitations. Further, despite our efforts to adjust for the use of R maintenance therapy, there might be some patients for whom our method failed to correctly categorize the sequence of treatments due to the limitations of SEER-Medicare dataset. As with any observational study, the lack of selection criteria (for the choice of R-CHOP and other therapies) is another limitation of the SEER-Medicare dataset. Although we conducted several multivariable Cox regression analyses to assess the effectiveness of R-CHOP and found that findings were robust and could not be attributed to another covariate, we could only account for the factors that were available in the dataset. The Cox models accounted for age, gender, comorbidities,

FL prognostic factors and other demographic covariates known to influence outcomes in FL. Yet, additional unmeasured factors could also influence treatment selection and confound assessment of the relationship between treatment and outcomes. Although this was a SEER population-based study, linkage with Medicare claims yielded a cohort with a mean age of 76 years. Thus, caution must be exercised when interpreting these findings for clinical practice, where individual patient characteristics - such as existence of co-morbidities - influence the selection and performance of treatments strategies especially for older individuals. In these analyses, we adjusted for comorbidities using the CCI. However, other unmeasured comorbidities may influence practice patterns and outcomes. Moreover, the lack of tumor size measures or tumor size related indications for treatment, such as Groupe d'Etude des Lymphomes Folliculaires (GELF) criteria [210], in the SEER registry or Medicare claims datasets prevented us from analyzing the impact of tumor burden on outcomes. The inclusion of such information in disease-specific clinical data collection in cancer registries will enable further analyses and improve the clarity of findings from such studies.

Prior predictive models have focused on effects of a single line of therapy on a single outcome of interest such as PFS or OS, while our model allows us to untangle the effects of first-line and additional therapies on subsequent treatment progression and from factors pertaining to all-cause death. Knowledge regarding predictive factors for earlier transition to next line treatment provides clinicians with more specific information to use in the decision-making process, and is particularly valuable in a high-risk, older population. Age has consistently been an important risk factor for worse FL outcomes [185] [186], and age above 85 years was associated with elevated risk of death after initial (first-line) treatment for this high-risk population. In addition, age above 80 years was associated with earlier initiation of third-line treatment, indicating more rapid disease progression after the second-line therapy for older patients. B symptoms, as expected, were associated with worse OS [186], as well as with earlier transitions between therapies. Contrary to prior studies that identified being married as a significant predictor of improved outcomes in FL

[211], we found that married status predicted earlier initiation of the second-line therapy.

While other analyses have documented the poor outcomes associated with early PD following first-line chemo-immunotherapy [181], this is one of the first to indicate in a large dataset that early progression from second- to third-line therapy is also associated with poor 2-year and 5-year OS. However, we acknowledge that our study selected for patients experiencing early PD following all first-line therapies as opposed to relapse following chemo-immunotherapy only [181], and the prognosis for these patient groups is not necessarily the same. National clinical trials are now underway for patients with early progression after first-line therapy and their findings might shed more light on this important issue.

In this study, we modeled the clinical course of FL with a multi-state model, employed the Aalen-Johansen estimator to estimate the clinical course of FL over time under various treatments, and utilized Cox regression models to quantify the impact of socio-demographic and clinical factors. Multi-state models offer an ideal framework to accurately analyze survival data with multiple intermediate states and/or multiple endpoints [212]. By capturing different types of events and the relationships between these events separately within the same model, the multi-state modeling framework distinguishes the differential effects of baseline clinical factors and subsequent treatment events across lines of care, and allows us to evaluate their distinct influences on outcomes in more detail. This comprehensive modeling framework was especially important for the accuracy of our study since examining inter-dependent clinical events within the same model, such as competing risks of OS (i.e., death) and different phases of progression free survival (i.e., across multiple lines of subsequent treatment), address biases in estimates that have been demonstrated to occur when events are examined in isolation [213] [214] [215] [216]. Our use of the large SEER-Medicare claims database was also critical for our method as the availability of a large dataset enabled us to utilize a nonparametric method and hence, avoid potentially unrealistic parametric assumptions on the clinical course of FL over time.

Our study provides the first attempt to examine multiple clinical states beyond diagnosis for patients with FL using the most contemporary data available in the type of large dataset required to perform this type of analysis. Yet, despite its strengths in conceptual design, analysis, and modeling, our study would significantly benefit from a dataset that includes more modern front-line chemo-immunotherapy agents such as bendamustine. For instance, data from the BRIGHT study [217] suggested that bendamustine combined with rituximab produced equivalent outcomes to R-CHOP or R-CVP for patients with previously untreated advanced stage FL. In addition, data from the STiL study [182] suggested that bendamustine and rituximab improved progression free survival (PFS) when compared to R-CHOP for patients with advanced stage FL receiving first-line therapy. Unfortunately, since bendamustine was underrepresented in this SEER-Medicare dataset, we were not able to conduct further analysis on the effectiveness of treatments that include this agent. To our knowledge, currently, no existing resource provides sufficient detailed information on subsequent lines of therapy under first-line bendamustine + rituximab treatment and sufficient follow-up to describe outcomes. Yet, as large clinical datasets with more contemporary treatment strategies, such as the Follicular Lymphoma Analysis of Surrogate Hypothesis (FLASH) collaboration [218], are now becoming available, we are optimistic that similar multi-state survival analyses will be conducted on these new and richer datasets to produce more insightful results.

3.2.6 Conclusion

We conducted a multi-state survival analysis to study the clinical course of FL under first-, second- and third-line treatments, and assess the impact of sequential therapies on immediate and subsequent line outcomes. Our analysis revealed that R-CHOP at any-line of treatment improved the OS, achieving the most favorable impact when provided as the first-line therapy. Together these findings suggest that R-CHOP remains an appropriate comparator for clinical trials involving patients with high-risk FL at any line of therapy

when compared to R-monotherapy, R-CVP and R-Other. Yet, caution must be exercised to the generalization of the effectiveness of first-line R-CHOP as the study population derived from the SEER-Medicare data, consisting of all first-line patients, was enriched for older high-risk patients with poor FL outcomes (early PD or death).

The key clinical practice points of this study are as follows: (i) No standard treatment exists for high-risk follicular lymphoma (FL) patients and the effectiveness of sequential therapies remains unclear, (ii) Utilizing a large dataset and a multi-state model, this survival analysis study examined the clinical course of FL under first-, second- and third-line treatments, (iii) In this FL population enriched for patients with early progression from first- to second-line therapy, B-symptoms, being married, and grade 1/2 histology were associated with earlier initiation of second-line therapies. Early progression from second- to third-line therapy was shown to be associated with poor OS, (v) R-CHOP at any-line of treatment was demonstrated to improve the OS rates, achieving the most favorable impact when provided as the first-line therapy (iv) Our findings suggest that R-CHOP remains an appropriate comparator for clinical trials involving patients with high-risk FL at any line of therapy when compared to R-monotherapy, R-CVP and other R-containing treatments (i.e., R-Other), and (vi) Using the same regimen in first- and second-line treatments, in particular R-CHOP + R-CHOP and R-CVP + RCVP, yielded adverse clinical outcomes with high two-year mortality rates. The multi-state model approach allowed a more detailed analysis of the impact of clinical covariates on the phases of care and ensured the model accuracy for distinguishing endpoints. The utility of this approach is not limited to FL and it can be applied to other clinical situations to inform decision-making on sequences of therapy when longitudinal time-to-event data are available for patient populations.

3.3 A Population-Based Multi-state Model for Diffuse Large B Cell Lymphoma-Specific Mortality in Older Patients

3.3.1 Introduction

Diffuse large B-cell lymphoma (DLCBL) is the most common lymphoid malignancy in the United States, accounting for about one-fourth of all non-Hodgkin lymphomas [219]. DLBCL is an aggressive disease with untreated survival less than 12 months and a peak incidence after age 60 years [220]. The majority of patients who receive the standard regimen of rituximab with cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) survive more than five years and are considered cured [221] [222] [223] [224]. Yet, a sizeable fraction of DLBCL patients do not respond to initial therapy or experience relapse [221] [225] [226] [227] while others die from comorbid conditions or treatment-related complications. At present, autologous stem transplant is the only known curative approach in the setting of relapse and older individuals rarely are candidates for this treatment.

Identifying predictors of poor outcomes have been a central focus of recent research in DLBCL, where a variety of biological, clinical and socio-demographic factors have been identified as contributors to the disparity in DLBCL outcomes [224] [228] [229] [230] [231]. Most recently, utilizing a large dataset, Howlader et al. conducted a population-based study on patients $\geq$ age 20 to investigate predictors for DLBCL-associated death, and assess cure outcomes with risk stratification in the R-CHOP era (i.e., from 2002 onward) [229]. The authors identified advanced stage, older age, black race, and Hispanic ethnicity as predictors of worse DLBCL-specific survival and found that being married and female sex were associated with a lower risk of DLBCL-specific death. The authors also concluded that there was no clear (so-called "cure") point, where DLBCL-specific mortality risk reaches a plateau over time and patients can be considered cured [225] [232] [233]. Despite its valuable findings, this study was limited by missing key prognostic variables, such as treatment data [234], and by the modeling approach examining cancer- and other

cause-specific deaths separately as this approach has been shown lead to biased estimates [213] [215] [216], and potentially incorrect conclusions [235], despite its popularity.

Compared to younger patients, patients older than 60 years experience significantly worse DLBCL outcomes [236] that have been shown to be influenced by geriatric syndromes, frailty, comorbid diseases and variations in disease biology [237] but these factors have not commonly been addressed in large cohort studies. The International Prognostic Index (IPI), which includes advanced stage, elevated lactate dehydrogenase, age over 60 years, an Eastern Cooperative Oncology Group performance status ≥ 2, and the involvement of more than one extranodal site, has been the primary clinical tool used for risk stratification [236] [238], but requires adjustments for older individuals. A study specifically focusing on DLBCL patients $\geq$ age 60 can aid in better characterizing risk factors for elderly population as well as identifying older patients in need of novel therapies at diagnosis or relapse.

In this study, we examined the clinical course of DLBCL in elderly patients in the R-CHOP era by conducting a multi-state survival analysis with a large population-based dataset that includes patient-level demographic, clinical and treatment information. Our objectives were to investigate the prognostic factors affecting overall and cause-specific survivals in the general and R-CHOP (sub-) populations within a comprehensive model.

3.3.2 Methods

Given that older individuals are at risk for multiple causes of death due to comorbidities, in our study, we used a multi-state framework that captures DLBCL- and other causes-associated deaths separately within a single model, so that we can account for competing risks and accurately identify factors affecting cause-specific mortalities.

We developed two multi-state models. Utilizing the first model, we analyzed the course of DLBCL in the general elderly population with four model states: Alive after Diagnosis, Alive after Treatment (TX), Death due to DLBCL and Death from Other Causes (Figure

3.4). We built a separate second multi-state model for the sub-population of patients receiving first-line RCHOP therapy. The health states examined by the second model are as follows: Alive after First-Line Treatment (TX1) with R-CHOP, Alive after Subsequent Therapy, Death due to DLBCL and Death due to Other Causes (Figure 3.4).

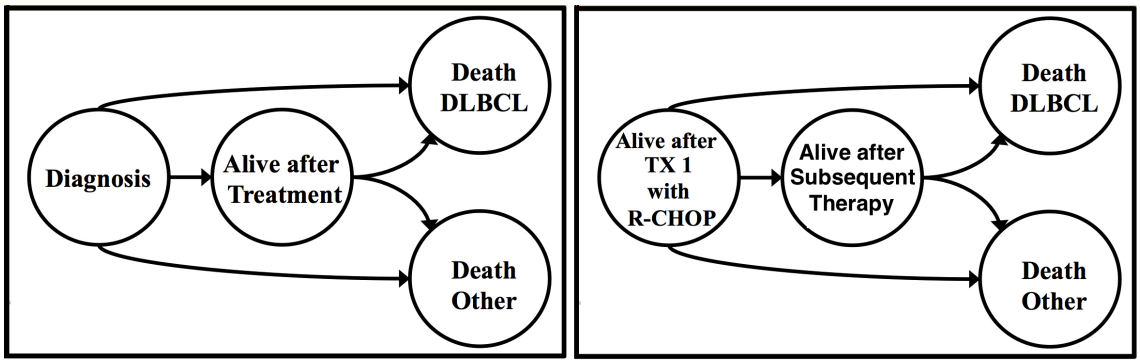


Figure 3.4: Multi-state Models for DLBCL Patients and First-Line R-CHOP Patients

We employed the Aalen-Johansen estimator [187], a generalization of Kaplan-Meier (KM) estimator [188], to assess the probability of being one of the model states at any given time. The Aalen-Johansen estimator can cope with censored observations, is the nonparametric maximum likelihood estimator of multi-state models [189] [190] [239], and provides consistent and reliable estimates both for Markov and non-Markov models [191]. We used multivariable Cox proportional hazard models [240] [241] to examine the association between clinical and socio-demographic factors and the time of the clinical events of interest. To do so, first, we constructed a multivariable Cox regression model for the conventional overall survival (OS) model and analyzed the impact of predictors on OS. Then, we fit a multivariable Cox model to each transition between the model states of the multi-state models and identified the risk factors having a (statistically) significant impact on the timing and rate of each clinical event. This is one of the major strengths of a multi-state approach as the use of Cox models over each transition (e.g., early deaths before treatment) allows the identification of the covariates that define different populations (having different prognosis and characteristics), which can lead to the design of tailored interventions for specific poor-risk subgroups based on data for the cause of death.

3.3.3 Data Source and Description

We used Surveillance, Epidemiology, and End Results (SEER) data registry (2002-2009) and Medicare claims dataset (2002-2011), to identify individuals with a histologically confirmed initial primary DLBCL diagnosis. Sponsored by the National Cancer Institute (NCI), the SEER program operates (high-quality) population-based cancer registries by systematically collecting clinical and socio-demographic information on cancer cases in all SEER designated areas [193]. Medicare is a public health organization, and the primary source of healthcare insurance for more than 90% of the U.S. citizens ≥ 65 years old. Since most DLBCL patients are diagnosed after age 60, the SEER-Medicare data offer a valuable source to analyze DLBCL treatment patterns and outcomes.

Patients were observed from diagnosis until death or end of follow-up (12/31/2011 for all-cause and 12/31/2009 for cause-specific mortality). All patients were older than 65 years at diagnosis. Patient age, sex, marital status, race, geographical region (Northeast, Midwest, South, or West), residency demographics (metropolitan, urban, or rural area), education level (as a census tract-level characteristics), and cause of death were collected from the SEER dataset. Other potential predictors included in the data were Charlson comorbidity index, DLBCL stage, extranodal primary site of involvement, and the receipts of treatments. We categorized treatments into five main categories: (i) rituximab (R), (ii) Rcyclophosphamide and vincristine (RCVP), (iii) Rcyclophosphamide, hydroxydaunorubicin, and vincristine (RCHOP), (iv) other Rcontaining regimens (ROther), and (v) regimens that do not contain R (Non-R). We also included the patients with unknown treatments in general analyses. DLBCL subtype was not included in our analysis due to data limitation in SEER.

3.3.4 Results - The General Elderly Population

We examined 11,780 patients with DLBCL older than 65 years. The median age at diagnosis was 76 years. Patients were diagnosed between 2002 and 2009, 9,508 patients received a

first-line therapy, 5,732 died within two years of diagnosis (69% due to DLBCL), and 7,599 died before the last day of follow-up (58% due to DLBCL). Table 3.4 displays the baseline characteristics for all patients (n=11,780), patients receiving a first-line therapy (n=9,508), and patients with R-CHOP first-line therapy (n=3,610). Median time to treatment was short for all first-line regimens: 1.07 months for R, 1.07 months RCVP, 1.00 month for RCHOP, 1.03 months for R-Other, and 0.87 months for Non-R. Two and five-year OS rates after diagnosis were the highest among patients who received R-CHOP as first-line treatment at 76.9% and 62.1%, and were 53.6% and 34.3% for R, 57.6% and 38.6% for R-CVP, 66.2 % and 49.4% for R-Other, 45.9% and 31.7% for Non-R treatments, respectively (Figure C.3).

Table 3.4: Patient Characteristics at Diagnosis

Patient Characteristics		All Patients (n = 11,780)	Patients with Treatment (n = 9,508)	Patients with First-Line R-CHOP (n = 3,610)
Age (years)	66 - 70	19.1%	20.9%	27.0%
	71 - 75	22.5%	24.1%	28.8%
	76 - 80	23.8%	24.4%	24.7%
	81 - 85	20.6%	19.4%	15.2%
	≥ 86	14.1%	11.2%	4.4%
Stage	Stage I	29.5%	30.2%	30.8%
	Stage II	18.6%	19.3%	21.1%
	Stage III	15.5%	16.0%	16.8%
	Stage IV	30.4%	29.3%	25.7%
	Unknown	6.1%	5.3%	5.5%
Primary Site	Nodal	62.8%	63.6%	66.2%
	Extranodal	37.2%	36.4%	31.0%
Charlson Index	Charslon Index = 0	53.1%	54.9%	60.5%
	Charslon Index = 1	25.8%	26.0%	25.4%
	Charslon Index ≥ 2	21.1%	19.1%	14.1%
Gender	Male	46.2%	45.8%	47.4%
	Female	53.8%	54.2%	52.6%
Marital Status	Married	53.1%	55.7%	61.8%
	Other Status	42.8%	40.4%	33.3%
	Unknown	4.2%	3.9%	4.9%
High School Diploma Only	Less than 25%	42.2%	42.6%	44.5%
	More than 25%	57.4%	57.1%	55.1%
Race	Caucasian	87.9%	88.3%	89.4%
	African American	3.8%	3.5%	3.0%
	Asian American	3.4%	3.4%	3.0%
	Other	4.9%	4.8%	4.6%
Region Demographics	West	41.6%	41.2%	41.5%
	Northeast	20.7%	2.0%	20.1%
	Midwest	13.0%	13.2%	14.4%
	South	24.6%	24.7%	24.0%
Residency Demographics	Metro Area	83.1%	82.9%	83.2%
	Urban or Rural	16.9%	17.1%	16.8%
First-Line Treatment Group	Rituximab (R)	6.2%	7.7%	NA
	R-CVP	3.6%	4.4%	NA
	R-CHOP	30.7%	38.0%	100.0%
	Other R-containing	6.2%	7.7%	NA
	Non R-containing	14.4%	17.8%	NA
	Unknown	19.7%	24.4%	NA
	No First-Line Treatment	19.3%	NA	NA

The analyses based on multivariable Cox regression models revealed that age $>$ 70

years, advanced stage, Charlson comorbidity index (CCI) score ≥ 1, and not being married were associated with an increased risk of DLBCL-caused mortality for patients with or without treatment (Table 3.5). In particular, comparing to the reference age group 66-70, the hazard ratios (HRs) for DLBCL-specific mortality after first-line treatment were 1.25, 1.46, 1.88, and 2.26 for age group 71-75, 76-80, 81-85, and ≥ 86, respectively. For patients receiving treatment, being female (HR = 0.91) and having higher socioeconomic status (HR = 0.91) were associated with a lower risk of DLBCL-related death following treatment.

Table 3.5: Multivariable Cox Regression Models for DLBCL Patients

Multivariable Analyses Cox PH Regression Model			Overall Survival	Multi-state Model: Diagnosis -> Alive after Treatment -> Death DLBCL, Death Other				
FACTOR	CLASSIFICATION	REFERENCE GROUP	Diagnosis --> Death All	Diagnosis --> Treatment	Diagnosis --> Death DLBCL	Diagnosis --> Death Other	Treatment --> Death DLBCL	Treatment --> Death Other
AGE AT DIAGNOSIS (Years)	66 - 70	X						
	71 - 75		**1.28** (1.18, 1.39)		**1.50** (1.19, 1.89)		**1.25** (1.11, 1.41)	**1.29** (1.13, 1.47)
	76 - 80		**1.61** (1.49, 1.74)		**2.36** (1.86, 2.86)	**1.50** (1.14, 1.98)	**1.46** (1.30, 1.64)	**1.58** (1.38, 1.79)
	81 - 85		**2.05** (1.90, 2.22)	**0.89** (0.83, 0.94)	**2.66** (2.15, 3.30)	**2.26** (1.75, 2.93)	**1.88** (1.67, 2.11)	**2.36** (2.07, 2.69)
	≥ 86		**2.53** (2.33, 2.76)	**0.68** (0.63, 0.73)	**3.61** (2.92, 4.47)	**3.01** (2.32, 3.92)	**2.26** (1.98, 2.59)	**3.72** (3.21, 4.32)
STAGE	STAGE I	X						
	STAGE II		**1.25** (1.17, 1.34)	**1.11** (1.04, 1.17)	**1.54** (1.29, 1.83)		**1.28** (1.14, 1.43)	
	STAGE III		**1.43** (1.33, 1.54)	**1.23** (1.16, 1.32)	**2.05** (1.71, 2.46)	**1.35** (1.03, 1.78)	**1.54** (1.37, 1.72)	
	STAGE IV		**1.78** (1.67, 1.88)	**1.15** (1.09, 1.22)	**2.54** (2.20, 2.94)	**1.81** (1.49, 2.20)	**1.95** (1.77, 2.14)	**1.32** (1.19, 1.47)
SITE	NODAL	X						
	EXTRANODAL			**0.90** (0.86, 0.94)		**1.24** (1.06, 1.45)		**1.11** (1.02, 1.21)
CHARLSON INDEX	CHARLSON INDEX = 0	X						
	CHARLSON INDEX = 1		**1.30** (1.23, 1.38)		**1.17** (1.03, 1.34)	**1.51** (1.26, 1.82)	**1.15** (1.06, 1.26)	**1.42** (1.29, 1.56)
	CHARLSON INDEX ≥ 2		**1.75** (1.65, 1.85)		**1.85** (1.63, 2.10)	**2.57** (2.16, 3.07)	**1.37** (1.25, 1.50)	**2.02** (1.83, 2.23)
GENDER	MALE	X						
	FEMALE		**0.91** (0.87, 0.96)	**1.12** (1.08, 1.17)		**0.76** (0.65, 0.89)	**0.91** (0.85, 0.99)	**0.75** (0.69, 0.82)
MARITAL STATUS	MARRIED	X						
	ALL OTHER STATUSES		**1.11** (1.05, 1.16)	**0.85** (0.82, 0.89)	**1.24** (1.11, 1.39)	**1.40** (1.18, 1.65)	**1.12** (1.03, 1.21)	**1.13** (1.04, 1.24)
HIGH SCH. ONLY	LESS THAN %25	X						
	MORE THAN %25		**1.12** (1.06, 1.18)	**0.97** (0.92, 1.00)			**1.10** (1.02, 1.20)	**1.18** (1.09, 1.28)
RACE	CAUCASIAN	X						
	ASIAN AMERICAN					**0.54** (0.32, 0.92)		
REGION STATUS	WEST	X						
	SOUTH		**1.10** (1.03, 1.17)				**1.12** (1.01, 1.23)	
TX1 GROUP	Rituximab (R)	X						
	R-CVP		**0.84** (0.72, 0.97)				**0.75** (0.61, 0.93)	
	R-CHOP		**0.62** (0.56, 0.68)				**0.51** (0.44, 0.59)	**0.69** (0.60, 0.80)
	Other R-containing		**0.79** (0.69, 0.90)				**0.70** (0.59, 0.84)	
	Non R-containing		**1.27** (1.14, 1.40)				**1.34** (1.16, 1.54)	
	No Treatment		**3.69** (3.35, 4.07)				N/A	N/A

Age, stage, (male) gender, CCI, and (the lack of) marital status were also predictors of poor OS and non-DLBCL mortality whereas extranodal site of disease, and race were found to influence non-DLBCL mortality rate (Table 3.5). As expected, treatment was shown to reduce the risk of DLBCL-specific death, where R-CHOP achieved the greatest impact among first-line therapies (HR: 0.51, 95% confidence interval (CI) 0.4-0.6).

3.3.5 Results - The Subpopulation of Patients Receiving First-Line R-CHOP Therapy

Among 11,780 DLBCL patients examined, 3,610 received a first-line R-CHOP therapy: 937 (26%) of these patients received no further treatment and 2,673 (74%) received a subsequent therapy after R-CHOP within a median of 2.90 months (Figure C.4). The receipt of R-CHOP as a first-line therapy decreased with increasing age. Out of 9,508 patients receiving a first-line treatment, the percentages of first-line R-CHOP were 43%, 39%, 32%, 23% and 9% within the age subgroups 66-70, 71-75, 76-80, 81-85, and $\geq$ 86, respectively. On the contrary, the adoption of first-line R-CHOP increased slightly with the (later) year of diagnosis. The rates of R-CHOP, out of all patients diagnosed in the same period, increased from 28% in the period 2002-2005 to 34% in the period 2006-2009.

Table 3.6: Multivariable Cox Regression Models for First-Line R-CHOP Patients

Multivariable Cox PH Regression Model			Overall Survival	Multi-state: Alive after TX1 -> Alive After Subsequent TX -> Death DLBCL, Death Other				
FACTOR	CLASSIFICATION	REFERENCE GROUP	TX1 R-CHOP --> Death All Cause	TX1 R-CHOP --> Subsequent TX	TX1 R-CHOP --> Death DLBCL	TX1 R-CHOP --> Death Other	Subsequent TX --> Death DLBCL	Subsequent TX --> Death Other
AGE AT DIAGNOSIS	66 - 70	X						
	71 - 75		**1.41** (1.21-1.64)		**1.66** (1.06-2.60)	**2.41** (1.45-4.01)	**1.29** (1.01-1.64)	**1.32** (1.04-1.68)
	76 - 80		**1.54** (1.32-1.79)		**1.92** (1.23-3.02)	**3.13** (1.89-5.19)		**1.59** (1.25-2.02)
	81 - 85		**2.41** (2.05-2.83)		**3.03** (1.92-4.78)	**5.52** (3.30-9.24)	**1.61** (1.22-2.12)	**2.60** (2.02-3.35)
	≥ 86		**3.10** (2.43-3.86)		**3.34** (1.77-6.30)	**8.21** (4.33-15.6)		**3.89** (2.75-5.51)
STAGE	STAGE I	X						
	STAGE II		**1.19** (1.02-1.38)	**0.83** (0.75-0.93)			**1.39** (1.06-1.83)	
	STAGE III		**1.48** (1.27-1.74)	**0.73** (0.65-0.82)	**1.72** (1.11-2.69)		**2.37** (1.83-3.07)	**1.40** (1.09-1.80)
	STAGE IV		**1.70** (1.48-1.95)	**0.89** (0.81-0.99)	**2.17** (1.46-3.23)		**2.41** (1.92-3.04)	**1.46** (1.18-1.81)
SITE	NODAL	X						
	EXTRANODAL					**1.34** (1.01-1.80)		
CHARLSON INDEX	CHARLSON INDEX = 0	X						
	CHARLSON INDEX = 1		**1.39** (1.23-1.56)		**1.91** (1.39-2.62)	**1.42** (1.02-1.97)	**1.29** (1.05-1.58)	**1.33** (1.10-1.61)
	CHARLSON INDEX ≥ 2		**1.71** (1.49-1.97)		**2.25** (1.56-3.24)	**1.86** (1.27-2.73)		**2.06** (1.67-2.54)
GENDER	MALE	X						
	FEMALE		**0.75** (0.68-0.83)		**0.65** (0.48-0.87)	**0.54** (0.40-0.73)	**0.82** (0.69-0.98)	**0.79** (0.67-0.93)
MARITAL STATUS	MARRIED	X						
	ALL OTHER STATUSES					**1.35** (1.00-1.85)		
HIGH SCH. ONLY	LESS THAN %25	X						
	MORE THAN %25		**1.14** (1.03-1.27)		**1.38** (1.03-1.85)			
RACE	CAUCASIAN	X						
	AFRICAN AMERICAN		**1.39** (1.06-1.83)			**2.72** (1.53-4.85)		
	ASIAN AMERICAN							
	OTHER							
TX1 GROUP	TX1 = R-CHOP	X	X	X	X	X	X	X
SUBSEQUENT THERAPY AFTER TX1 = R-CHOP	Rituximab (R), R-CVP or R-CHOP	X		NA	NA	NA	X	X
	R-Other or Non-R Containing			NA	NA	NA		
	Unknown			NA	NA	NA		
	No Second-Line Therapy		**1.25** (1.10-1.42)	NA	NA	NA	NA	NA

Similar to the general population, increased age, advanced stage, and CCI $\geq$1 were associated with worse OS and DLBCL-specific mortality whereas being female improved OS and reduced cause-specific mortality risk both due to DLBCL and from other causes (Table 3.6). The cumulative mortality rate associated with DLBCL was 14.0% at two years and 18.6% at five years under the front-line R-CHOP, after which it plateaued and was

exceeded by the risk of other causes mortality thereafter (Figure 3.5). Among the patients receiving no subsequent (i.e., continuation of the first-line or a new second-line) therapy after R-CHOP, lower socio-economic status and being African American were respectively associated with higher risk of death from DLBCL (HR: 1.38) and other causes (HR: 2.72).

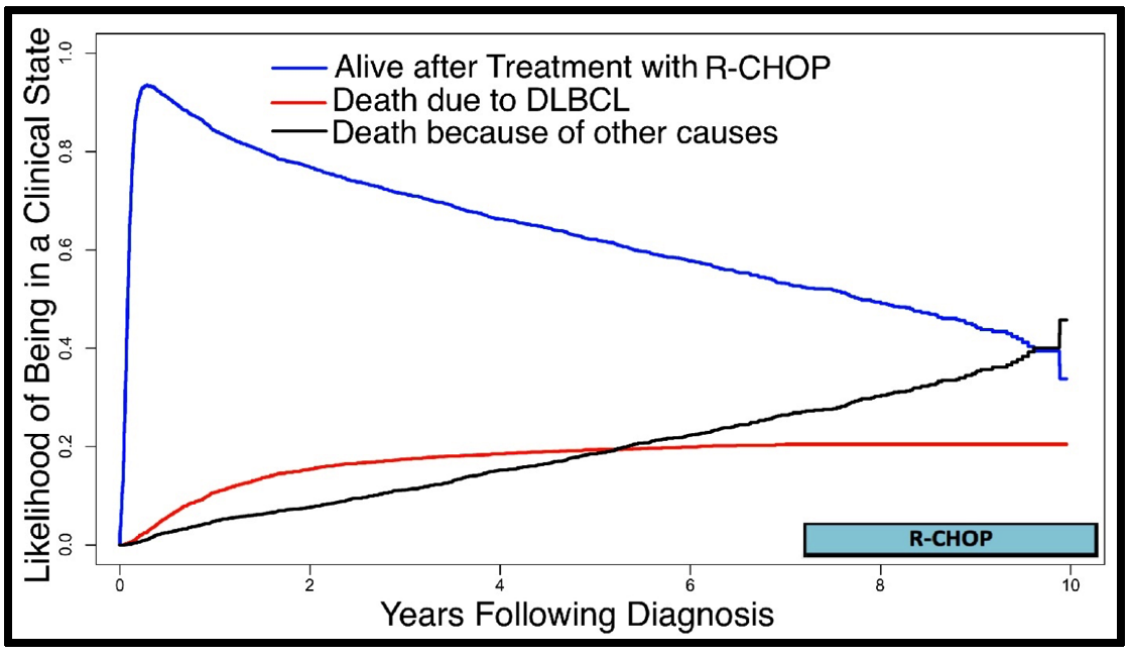


Figure 3.5: Cause-Specific Death and Survival Probabilities for R-CHOP Patients

3.3.6 Discussion

Traditionally, the decisions for initial management of older (and recently diagnosed) DLBCL patients have been dependent predominantly on clinician judgment [242] [243] [244]. The current standard treatment for DLBCL patients $>$ 60 years was based on the Groupe d? Etude des Lymphomes de l?Adulte (GELA) NHL trial (GELA LNH- 98-5). The GELA NHL trial demonstrated that the long-term survival outcomes were improved for these patients when R was administered together with cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) therapy [221] [245], and this result was later confirmed by the RICOVER-60 trial [246]. Approaches using dose-reduced CHOP with an anti-CD20 antibody also demonstrated favorable progression free survival with tolerable toxicity in pa-

tients $\geq$ 80 years, but both appear to produce inferior outcomes compared to patients of 80 years treated with R-CHOP at standard doses on GELA LNH- 98-5 [247] [248]. Comorbidities and changes in functional status might complicate anthracycline-based chemotherapy treatments, as shown in a large epidemiological study that analyzed treatment patterns in the U.S. for older patients with DLBCL [225]. The toxicities related to R-CHOP therapy exacerbate with increasing age, functional disability and comorbidity [249] [250]. Previous studies and our current analysis suggested that even "unfit" elderly patients may still benefit from anthracycline-based chemotherapy, thus making it imperative to undertake a careful assessment of fitness for R-CHOP before considering less toxic and potentially less effective alternatives [251] [252] [253].

Two large database-based analyses used the SEER-Medicare dataset to characterize treatment and survival outcomes for older individuals with DLBCL. Williams et al. examined 1,156 patients >80 years diagnosed with DLBCL from 2002-2009 and found that R-CHOP, as compared to non-anthracycline containing regimens, was associated with better lymphoma-related (HR = 0.58) and overall (HR = 0.45) survival outcomes [254]. Tien et al. examined 8,262 Medicare patients diagnosed with DLBCL from 2000-2006 and found that OS was highest in patients treated with an anthracycline-containing regimen plus R [255]. A recent systematic review conducted on the available evidence related to the inclusion of anthracycline found that the initial management strategies of DLBCL for older patients might benefit from the potential use of a comprehensive assessment, taking comorbidities and frailty/fitness into account (instead of age alone), as the available evidence shows that an extensive geriatric evaluation before treatment correctly anticipates treatment toxicities and (the course of) OS [256].

In this nationwide retrospective analysis of a large cohort of older individuals with DLBCL (n=11,780), we confirmed increased age and stage as predictors of worse DLBCL outcomes [254]. We found that older age and advanced stages are associated with worse OS, and increase the likelihood of cause-specific deaths, from DLBCL and other causes,

both before and after treatment. In particular, age > 70 and stage $\geq$ II both increased the risk of all-cause and DLBCL-related mortality. The CCI has previously been shown to predict worse OS for DLBCL adults age 18-64 years [230]; and our results affirmed that it is a prognostic factor in this elderly population ≥ 65 years old as well. Increased comorbities (CCI ≥ 1) did not influence time to initial treatment but were associated with worse OS and increased risk of death due to DLBCL, suggesting an important interplay between DLBCL and comorbidities. Higher CCI was also associated with increased risk of death from other causes both for patients receiving and those not receiving a treatment. Being married were shown to improve OS and cause-specific survivals before/after treatment, indicating a positive impact of marital support on DLBCL outcomes. Married patients were also associated with a shorter time to treatment. Our analysis confirmed the association between female gender and improved DLBCL outcomes [229] and demonstrated an increased likelihood of an earlier treatment among females, implying that gender traits may impact the time to initiate treatment and treatment outcomes. Surprisingly, nearly all socio-demographic and clinical factors associated with DLBCL-specific mortality were also factors associated with all other causes of death. The data and our findings suggest that these mitigating factors need to be considered in optimizing therapy both to increase efficacy and reduce adverse events.

Our analysis also demonstrated that higher socio-economic status was associated with a shorter time to treatment, better OS and lower DLBCL-specific (and other cause) mortality risk after therapy. Yet, there was no identified effect of socio-economic status on cause-specific survivals for patients receiving no treatment. African Americans more commonly present with DLBCL with a younger age at diagnosis and have been shown to have worse outcomes in previous studies [257] [258]. In this population of older patients with DLBCL, African American race was not associated with differences in OS or DLBCL-specific survival in the general population. Yet, in the R-CHOP subpopulation, African American race was found to be associated with worse DLBCL-specific survival among the (first-line

R-CHOP) patients receiving no subsequent therapy (after R-CHOP).

The utility of surveillance imaging and the optimal time to stop monitoring DLBCL patients have been analyzed by a couple of recent studies. A meta-analysis examining PET/CT surveillance in DLBCL patients achieving clinical remission recommended against routine surveillance imaging but suggested it is reasonable to consider in high-risk patients [232]. Cohen et al. [259] concluded that, for high-risk DLBCL patients, routine imaging, if performed, should be limited to the first two years after induction therapy as the probability of relapse, a significant risk factor for salvage therapy failure and early mortality [232], declines significantly after two years. Maurer et al. [225] also suggested a two-year cut-off point to stop observing DLBCL patients receiving standard immune-chemotherapy for relapse as their analysis revealed that the patients with event-free two-year survival after diagnosis had OS comparable to that of the age- and sex-matched general population. In their analysis of 18,047 DLBCL patients, diagnosed in the R-CHOP era (i.e., from 2002 onward), Howlader et. al [229] concluded that there was no clear point where DLBCL-specific mortality plateaued over time. Yet, this study was limited as the authors failed to determine the patients in SEER receiving R-CHOP due to data-related restrictions. In our study, we identified 3,610 R-CHOP patients and found that DLBCL-specific mortality was exceeded by the cumulative other causes-specific mortality risk 5.25 years after diagnosis (Figure 3.5) and remained nearly constant afterwards. This result suggests that, for elderly DLBCL patients, five years after front-line R-CHOP can be considered as a natural time point to transition the focus of survivorship surveillance plans from DLBCL-specific to other mortality risks. This finding provides additional credence for the common approach in clinical practice.

In the SEER-Medicare dataset, 38% (=3,610/9,508) of the first-line treatments were R-CHOP, which is lower than expected given that our study corresponds to the R-CHOP era. Utilizing a subset of the same dataset, Williams et al. [254] examined 3,176 (2,581 between age 66-80, and 595 above age 80) first-line R-CHOP patients diagnosed during

the period 2002-2009, and suggested that for some individuals co-morbidities or age were the main factors precluding the use of CHOP/R-CHOP therapy. Age was influential in the receipt of first-line R-CHOP in our study as well, where the proportion of patients receiving R-CHOP within age subgroups decreased with increasing age. Various other reasons, such as race, were shown to influence the slow adoption of chemo-immunotherapy [260], which also affects the adoption of R-CHOP despite the demonstrated effectiveness of R-CHOP in 2002 by the French GELA group [221]. We observed the year of diagnosis was also a factor in the adoption of R-CHOP, where the annual rate of patients receiving R-CHOP noticeably increased between 2002 and 2009.

We should note the limitations of our study. The study population was ≥ 65 years old at the time of diagnosis. Accordingly, further analyses are needed to generalize the findings of this study to younger age groups. As with many other observational studies, the lack of selection criteria for the choice of treatments was another dataset-related limitation. Due to the nature of SEER-Medicare dataset in which the determination of the chemotherapy regimen being administered is derived from claims data, we were unable to determine chemotherapy dosing. This issue is an important one to consider in other very large clinical datasets with direct ascertainment of the treatment regimen, dosing, timing of dose delivery, and outcomes. Large clinical datasets such as the Surrogate Endpoints for Aggressive Lymphoma (SEAL) collaboration are now becoming available to address such question [261].

In terms of methodology, even though the Aalen-Johansen estimator is the canonical nonparametric maximum likelihood estimator for multi-state models with guaranteed asymptotical convergence [214] [189] [262] [215] caution must be exercised as the size of the available data gets smaller. This is unlikely to pose an issue for subpopulation analyses (e.g., patients using R-CHOP) [213] but might be a concern for individualized predictions as each covariate defining an individual patient also shrinks the size of the available data that the estimator utilizes. In such cases, either Cox regression models can be appended

to make personalized inferences or a computational tool, such as micro-simulation, might be preferred. Finally, despite including the registry of subsequent therapies after first-line R-CHOP, the SEER-Medicare dataset does not specify whether a subsequent therapy is a continuation of the first-line or a new (i.e., second-line) treatment. This limitation prevented us from conducting further analyses on the subsequent therapies following first-line R-CHOP (in the R-CHOP subpopulation).

Despite these limitations, it is worth mentioning some of the key strengths our multi-state approach. First, the availability of a large dataset enabled us to utilize a nonparametric method and hence, avoid potentially unrealistic parametric assumptions on the clinical course of DLBCL (over time) following diagnosis or treatment [189] [263] [190]. Second, with the use of multi-state model, we captured the competing risks of DLBCL-specific and other cause mortality within the same framework, whereas studying them separately was shown to yield biased estimates [214] [213] [215] [216] [264]. Third, by using separate model states for diagnosis and treatment, we were able to distinguish patients with and without treatment (and differentiate early deaths before treatment from others), and define the subgroups having different prognosis, characteristics [220], and clinical implications for improving outcomes. Finally, as we were able to identify the patients receiving R-CHOP in the SEER-Medicare dataset, we could conduct an in-depth survival analysis over the R-CHOP sub-population with a tailored multi-state model for older patients who are administered the current, optimal first-line therapy.

3.3.7 Conclusion

We conducted a population-based multi-state survival analysis on the clinical course of DLBCL in R-CHOP era for older individuals. Utilizing a large population-based dataset, we confirmed and showed that increased age, advanced DLBCL stage, higher Charlson comorbidity index and the lack of marital support as significant predictors for poor OS and increased risk of DLBCL-associated death in the elderly population. Our findings

accentuated the importance of comorbidity and functional status assessments in the evaluation of older patients with lymphoma prior to therapy initiation and the potential role of gender characteristics, social support and socio-economic status on treatment outcomes. Our multi-state analysis revealed that DLBCL-associated mortality plateaued and was exceeded by other-cause specific mortality risk five years after R-CHOP treatment. This result provides strong support for transitioning survivorship surveillance plans from a focus on lymphoma relapse-related deaths to non-cancer risks at five years following treatment. Additional research using the multi-state approach with datasets including younger patients and detailed information on later lines of treatment is required to assess the impact of sequential therapies following R-CHOP.

CHAPTER 4
CONCLUSION

In Chapter 1, I studied a multi-modality breast cancer screening problem. Women with certain risk factors are at higher risk of developing breast cancer. Some of these factors are gene mutations, family history of breast or ovarian cancer, or elevated breast density. For these high-risk women, intensified screening with non-mammographic technologies such as ultrasound (US) and magnetic resonance imaging (MRI) have been suggested to address some of the limitations of mammography, the standard screening modality for average-risk women. However, MRI is significantly more expensive and yields substantially higher number of false-positives than mammograms, while US is less sensitive than MRI and is more operator dependent, despite being cheaper. Currently there is no consensus on the optimal use of US and MRI in conjunction with, or instead of, mammography in high-risk women. I studied and proposed a population-dynamics based optimization approach for the multi-modality BC screening problem for the high-risk population. I developed a Markov model to capture the disease incidence and progression in high-risk women, parameterize and calibrate it using the best available evidence and formulate a mixed integer linear program to identify optimal structured strategies that are practical for implementation. I assessed the cost-effectiveness of the identified strategies and the role of ultrasound's operator-dependency. I further shed some light on the structure of the optimal screening strategies by establishing the sufficiency conditions under which a strategy with more frequent screens yields higher health benefits than the one using a more sensitive modality. The main findings are the following: (1) Annual screening with ultrasound alone, even when its false-positive rates are high, and annual mammogram screening are respectively optimal with moderate budgets and robustly cost-effective in young (aged 25-44 years) and middle-aged (45-74 years) high-risk women, and (2) the annual use of MRI, alone or com-

bined with mammogram, leads to outcomes that are not cost-effective. These findings can be helpful in designing future trials, developing evidence-based guidelines and informing insurance coverage decisions.

In Chapter 2, I studied a physician staffing and an associated patient routing problem in the emergency room (ER) with time-varying stochastic arrivals. The ER is a highly complex service system. Patient visits are unscheduled and time-varying in terms of arrival rates and acuity levels. Designing, modeling and managing such a complex service system is a challenge and determining appropriate staffing levels while balancing costs and targeted service levels lies at the heart of this challenge. In this study, to design a model consistent with the perspective of ER physicians, I developed order bundle notion, which views the ER care delivery process as a sequence of treatments and diagnostic test orders placed by a physician. Based on this notion, I introduced an intuitive, realistic and tractable model of ER composed of a multi-class multi-stage queuing network with multiple targeted service levels. Utilizing infinite-server approximation and offered load analysis, I first approximated the expected stochastic load on the physicians in the ER. Then I converted the calculated workload into staffing decisions, by using a new staffing formula I developed, with the goal of satisfying the targeted tail probability of delay constraints. I analytically showed that our new staffing rule asymptotically satisfied the desired convergence for the M/M/s queues that operate efficiency-driven mode. Then, I demonstrated the robustness of our approach via realistic and data-driven simulation experiments in various (time-varying) ER settings, considering multi-stage service, multi-class patients and pooled servers (physicians). I also studied the associated patient routing problem and showed that as the service complexity of an ER increases, the use of dynamic rules, using the current system state for routing decisions, and hybrid policies, combining pre-determined static routing rules with dynamic ones, become necessary to stabilize TPoD targets. Overcoming the major limitations of black-box models, our study offered a new queuing theory-based staffing approach that is detailed enough to capture the key dynamics of complex service systems,

and practical enough to implement in practice.

In Chapter 3, I conducted multi-state survival analyses for clinical decision-making, with applications to follicular lymphoma (FL) and diffuse large B-cell lymphoma (DLBCL). DLBCL is a fast-growing lymphoma with high rates of treatment under correct therapies whereas FL is a slow-growing disease with lower rates of full recovery. Despite having different characteristics, the courses of both of these diseases include multiple clinical intermediate- or end-points of interest. In the works presented in Chapter 3, I developed two different continuous-time, multi-state models to investigate the clinical course of these diseases. Based on the SEER-Medicare data, one of the largest available datasets on FL and DLBCL, I took a data-driven solution approach and utilized nonparametric methods such empirical transition matrix estimates as well as other tools of survival analysis. Using the DLBCL model, I examined the role of clinical and socio-demographic factors on DLBCL-associated mortality in the elderly population, showed the effect of age, sex, and Charlson comorbidity index on cause-specific mortalities, and identified a five-year cutoff point to stop monitoring DLBCL patients receiving the standard R- CHOP therapy. This approach, coupled with the findings, enriched the current literature as the existing studies based their conclusions mostly on overall survival rather than cause-specific mortality. Employing the multi-state FL model, I studied the clinical course of FL under first, second and third line therapies for high-risk patients to assess the effectiveness of various treatment sequences. I showed that single R-CHOP therapy in any line improved overall survival for high-risk patients, achieving the most favorable outcome when provided as first-line therapy. This is important as currently there is no standard for treatment of these patients and hence, our findings might influence the current medical practice.

Appendices

APPENDIX A

APPENDIX FOR CHAPTER 1

A.1 Base-Case Analysis: Optimal Age-Specific Screening Strategies

Table A.1: Optimal Ten-year Strategies for 35 Year-old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	ICER ($/QALY)
No Screening	37.611	$0	NA
1 US at age 35	37.641	$279	DOMINATED
2 US at ages 35 and 40	42.189	$508	111
Biennial US between age 35-44	43.170	$1,258	764
Annual US between age 35-44	43.535	$2,232	2,672
Annual MRI between age 35-44	43.613	$16,451	DOMINATED
Annual MRI+MAM between age 35-44	43.646	$16,728	131,040

Table A.2: Optimal Ten-year Strategies for 45 Year-old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	ICER ($/QALY)
No Screening	30.778	$0	NA
1 US at age 45	30.904	$161	DOMINATED
2 US at ages 45 and 50	33.235	$407	166
2 MAM at ages 45 and 50	33.625	$665	DOMINATED
Biennial US between age 45-54	34.257	$916	498
Biennial MAM between age 45-54	34.320	$1,540	DOMINATED
Annual US between age 45-54	34.532	$1,618	2,542
Annual MAM between age 45-54	34.552	$2,818	61,127
Annual MRI between age 45-54	34.585	$15,311	DOMINATED
Annual MRI+MAM between age 45-54	34.625	$16,213	184,616

Table A.3: Optimal Ten-year Strategies for 65 Year-old BRCA1+ Women

Screening Policy	Expected QALYs	Expected Ten-year Cost	ICER ($/QALY)
No Screening	17.849	$0	NA
1 US at age 65	18.017	$188	DOMINATED
1 MAM at age 65	18.232	$272	711
2 US at ages 65 and 70	18.388	$388	747
2 MAM at ages 65 and 70	18.421	$541	DOMINATED
Biennial US between age 65-74	18.542	$832	2,873
Biennial MAM between age 65-74	18.559	$1,182	DOMINATED
Annual US between age 65-74	18.600	$1,479	11,234
Annual MAM between age 65-74	18.610	$2,150	64,385
Annual MRI between age 65-74	18.616	$13,698	DOMINATED
Annual MRI+MAM between age 65-74	18.625	$15,226	846,729

A.2 Optimal Screening Strategies with Low US Specificity

Table A.4: Optimal Strategies for 35 Year-old BRCA1+ Women with Low US Specificity

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	37.611	$0	NA
1 MAM at age 35	37.633	$327	DOMINATED
1 US at age 35	41.010	$438	129
2 MAM at ages 35 and 40	41.557	$682	DOMINATED
2 US at ages 35 and 40	42.266	$934	395
Biennial MAM between age 35-44	43.129	$1,654	835
Biennial US between age 35-44	43.219	$2,237	DOMINATED
Annual MAM between age 35-44	43.510	$3,119	3,849
Annual US between age 35-44	43.543	$4,294	35,500
Annual MRI between age 35-44	43.613	$16,451	DOMINATED
Annual MRI+MAM between age 35-44	43.646	$16,728	121,175

Table A.5: Optimal Strategies for 45 Year-old BRCA1+ Women with Low US Specificity

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	30.778	$0	NA
1 MAM at age 45	30.900	$306	DOMINATED
1 US at age 45	32.788	$446	222
2 MAM at ages 45 and 50	33.625	$665	262
Biennial MAM between age 45-54	34.320	$1,540	1,259
Annual MAM between age 45-54	34.552	$2,818	5,503
Annual MRI between age 45-54	34.585	$15,311	DOMINATED
Annual MRI+MAM between age 45-54	34.625	$16,213	184,616

Table A.6: Optimal Strategies for 65 Year-old BRCA1+ Women with Low US Specificity

Screening Policy	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	17.849	$0	NA
1 MAM at age 65	18.232	$272	711
2 MAM at ages 65 and 70	18.421	$541	1,425
Biennial MAM between age 65-74	18.559	$1,182	4,637
Annual MAM between age 65-74	18.610	$2,150	18,941
Annual MRI between age 65-74	18.616	$13,698	DOMINATED
Annual MRI+MAM between age 65-74	18.625	$15,226	846,729

A.3 Numerical Results for BRCA2+ Carriers and Women with Family History

Table A.7: Optimal Strategies for 35 Year-old Women with BRCA2+ and Family History

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	38.868	$0	NA
1 US at age 35	40.799	$287	DOMINATED
2 US at ages 35 and 40	42.460	$493	137
Biennial US between age 35-44	43.450	$1,248	763
Annual US between age 35-44	43.786	$2,260	3,013
Annual MRI between age 35-44	43.861	$17,170	DOMINATED
Annual MRI+MAM between age 35-44	43.883	$17,467	156,232

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	43.274	$0	NA
1 US at age 35	44.197	$287	311
2 US at ages 35 and 40	44.533	$451	485
Biennial US between age 35-44	44.833	$1,179	2,427
Annual US between age 35-44	44.946	$2,200	9,040
Annual MRI between age 35-44	44.973	$17,484	DOMINATED
Annual MRI+MAM between age 35-44	44.985	$17,795	400,798

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	38.868	$0	NA
1 MAM at age 35	38.890	$337	DOMINATED
1 US at age 35	41.657	$451	162
2 MAM at ages 35 and 40	41.963	$679	DOMINATED
2 US at ages 35 and 40	42.687	$939	473
Biennial MAM between age 35-44	43.448	$1,670	960
Biennial US between age 35-44	43.539	$2,281	DOMINATED
Annual MAM between age 35-44	43.774	$3,200	4,696
Annual US between age 35-44	43.796	$4,440	56,123
Annual MRI between age 35-44	43.858	$17,170	DOMINATED
Annual MRI+MAM between age 35-44	43.880	$17,467	155,173

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	43.274	$0	NA
1 MAM at age 35	44.212	$337	359
1 US at age 35	44.226	$451	DOMINATED
2 MAM at ages 35 and 40	44.556	$646	901
2 US at ages 35 and 40	44.566	$906	DOMINATED
Biennial MAM between age 35-44	44.832	$1,619	3,524
Biennial US between age 35-44	44.865	$2,246	DOMINATED
Annual MAM between age 35-44	44.943	$3,178	13,942
Annual US between age 35-44	44.950	$4,461	208,130
Annual MRI between age 35-44	44.967	$17,484	DOMINATED
Annual MRI+MAM between age 35-44	44.979	$17,795	454,451

Table A.8: Optimal Strategies for 45 Year-old Women with BRCA2+ and Family History

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	31.091	$0	NA
1 US at age 45	31.218	$165	DOMINATED
2 US at ages 45 and 50	33.214	$385	181
2 MAM at ages 45 and 50	33.749	$651	498
Biennial US between age 45-54	34.033	$890	616
Biennial MAM between age 45-54	34.385	$1,546	DOMINATED
Annual US between age 45-54	34.605	$1,624	1,285
Annual MAM between age 45-54	34.630	$2,894	51,818
Annual MRI between age 45-54	34.669	$16,052	DOMINATED
Annual MRI+MAM between age 45-54	34.697	$17,023	210,955

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	34.137	$0	NA
1 US at age 45	34.534	$165	DOMINATED
2 US at ages 45 and 50	35.013	$337	385
2 MAM at ages 45 and 50	35.202	$605	DOMINATED
Biennial US between age 45-54	35.425	$805	1,135
Biennial MAM between age 45-54	35.436	$1,480	DOMINATED
Annual US between age 45-54	35.520	$1,549	7,808
Annual MAM between age 45-54	35.534	$2,864	98,933
Annual MRI between age 45-54	35.549	$16,430	DOMINATED
Annual MRI+MAM between age 45-54	35.560	$17,459	546,028

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	31.091	$0	NA
1 MAM at age 45	31.214	$315	DOMINATED
1 US at age 45	32.992	$459	242
2 MAM at ages 45 and 50	33.749	$651	254
Biennial MAM between age 45-54	34.379	$1,546	1,419
Annual MAM between age 45-54	34.626	$2,894	5,462
Annual MRI between age 45-54	34.659	$16,052	DOMINATED
Annual MRI+MAM between age 45-54	34.693	$17,023	210,936

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	34.137	$0	NA
1 MAM at age 45	34.916	$315	404
1 US at age 45	34.921	$459	DOMINATED
2 MAM at ages 45 and 50	35.202	$605	1,015
Biennial MAM between age 45-54	35.436	$1,480	3,732
Annual MAM between age 45-54	35.528	$2,864	15,165
Annual MRI between age 45-54	35.542	$16,430	DOMINATED
Annual MRI+MAM between age 45-54	35.554	$17,459	547,873

Table A.9: Optimal Strategies for 55 Year-old Women with BRCA2+ and Family History

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	23.880	$0	NA
1 US at age 55	24.032	$175	DOMINATED
1 MAM at age 55	24.048	$277	DOMINATED
2 US at ages 55 and 60	25.247	$409	192
2 MAM at ages 55 and 60	25.499	$601	DOMINATED
Biennial US between age 55-64	25.777	$922	969
Biennial MAM between age 55-64	25.965	$1,371	DOMINATED
Annual US between age 55-64	26.096	$1,638	2,242
Annual MAM between age 55-64	26.124	$2,505	31,014
Annual MRI between age 55-64	26.145	$14,787	DOMINATED
Annual MRI+MAM between age 55-64	26.164	$16,435	349,593

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	25.725	$0	NA
1 US at age 55	25.878	$175	DOMINATED
1 MAM at age 55	26.236	$277	DOMINATED
2 US at ages 55 and 60	26.423	$354	327
2 MAM at ages 55 and 60	26.461	$545	DOMINATED
Biennial US between age 55-64	26.604	$817	2,557
Biennial MAM between age 55-64	26.625	$1,279	DOMINATED
Annual US between age 55-64	26.674	$1,539	10,319
Annual MAM between age 55-64	26.687	$2,445	74,082
Annual MRI between age 55-64	26.700	$15,209	DOMINATED
Annual MRI+MAM between age 55-64	26.704	$16,934	842,072

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	23.880	$0	NA
1 MAM at age 55	24.048	$277	DOMINATED
2 MAM at ages 55 and 60	25.310	$601	420
Biennial MAM between age 55-64	25.961	$1,371	1,184
Annual MAM between age 55-64	26.120	$2,505	7,126
Annual MRI between age 55-64	26.141	$14,787	DOMINATED
Annual MRI+MAM between age 55-64	26.159	$16,435	349,540

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	25.725	$0	NA
1 MAM at age 55	25.893	$277	DOMINATED
2 MAM at ages 55 and 60	26.461	$545	740
Biennial MAM between age 55-64	26.619	$1,279	4,630
Annual MAM between age 55-64	26.680	$2,445	19,251
Annual MRI between age 55-64	26.690	$15,209	DOMINATED
Annual MRI+MAM between age 55-64	26.697	$16,934	846,846

Table A.10: Optimal Strategies for 65 Year-old Women with BRCA2+ and Family History

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	17.288	$0	NA
1 US at age 65	17.456	$193	DOMINATED
1 MAM at age 65	17.471	$280	DOMINATED
2 US at ages 65 and 70	18.047	$446	589
2 MAM at ages 65 and 70	18.122	$607	DOMINATED
Biennial US between age 65-74	18.323	$939	1,780
Biennial MAM between age 65-74	18.348	$1,292	DOMINATED
Annual US between age 65-74	18.420	$1,598	6,825
Annual MAM between age 65-74	18.435	$2,264	43,008
Annual MRI between age 65-74	18.444	$14,450	DOMINATED
Annual MRI+MAM between age 65-74	18.458	$15,271	576,419

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	18.135	$0	NA
1 US at age 65	18.302	$193	DOMINATED
1 MAM at age 65	18.452	$280	DOMINATED
2 US at ages 65 and 70	18.550	$372	895
2 MAM at ages 65 and 70	18.574	$528	DOMINATED
Biennial US between age 65-74	18.653	$814	4,294
Biennial MAM between age 65-74	18.666	$1,180	DOMINATED
Annual US between age 65-74	18.691	$1,489	17,757
Annual MAM between age 65-74	18.699	$2,196	91,526
Annual MRI between age 65-74	18.704	$15,043	DOMINATED
Annual MRI+MAM between age 65-74	18.711	$15,945	1,190,374

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	17.288	$0	NA
1 MAM at age 65	17.471	$280	DOMINATED
2 MAM at ages 65 and 70	18.122	$607	728
Biennial MAM between age 65-74	18.348	$1,292	3,026
Annual MAM between age 65-74	18.435	$2,264	11,180
Annual MRI between age 65-74	18.444	$14,450	DOMINATED
Annual MRI+MAM between age 65-74	18.458	$15,271	576,419

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	18.135	$0	NA
1 MAM at age 65	18.452	$280	DOMINATED
2 MAM at ages 65 and 70	18.574	$528	1,201
Biennial MAM between age 65-74	18.666	$1,180	7,114
Annual MAM between age 65-74	18.699	$2,196	30,583
Annual MRI between age 65-74	18.704	$15,043	DOMINATED
Annual MRI+MAM between age 65-74	18.711	$15,945	1,190,374

Table A.11: Optimal Strategies for 75 Year-old Women with BRCA2+ and Family History

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	11.581	$0	NA
1 US at age 75	11.716	$203	DOMINATED
1 MAM at age 75	11.726	$281	1,940
2 US at ages 75 and 80	11.774	$380	2,077
2 MAM at ages 75 and 80	11.785	$513	DOMINATED
Biennial US between age 75-84	11.827	$773	7,413
Annual US between age 75-84	11.833	$1,357	93,114

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	11.660	$0	NA
1 US at age 75	11.789	$203	1,567
1 MAM at age 75	11.801	$281	DOMINATED
2 US at ages 75 and 80	11.828	$363	4,087
2 MAM at ages 75 and 80	11.837	$494	DOMINATED
Biennial US between age 75-84	11.859	$752	12,591
Annual US between age 75-84	NOT OPTIMAL	NOT OPTIMAL	NOT OPTIMAL

Screening Policy BRCA 2 Mutation	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	11.581	$0	NA
1 MAM at age 75	11.726	$281	1,940
2 MAM at ages 75 and 80	11.785	$513	3,931
Biennial MAM between age 75-84	11.824	$1,059	13,834

Screening Policy Family History	Expected QALYs	Expected Ten-year Cost	Updated ICER ($/QALY)
No Screening	11.660	$0	NA
1 MAM at age 75	11.801	$281	1,981
2 MAM at ages 75 and 80	11.837	$494	5,949
Biennial MAM between age 75-84	11.855	$1,042	31,423

A.4 Sensitivity Analysis

In this section, we perform a series of univariate sensitivity analyses on the cost of screening technologies and biopsy, and the disutility associated with screening and biopsy, and discuss our findings.

A.4.1 Sensitivity Analysis - Cost Function

We conduct sensitivity analysis on cost function by considering all possible combinations among low (minimum), medium (average), and high (maximum) cost scenarios separately (Table A.12), where medium (average) values are employed in the base case numerical study. Change in the cost of one of the technologies might result in a change in optimal screening strategies, comparing to the base study results. We name a change as "robust" if it occurs under all possible scenarios that are separately examined. The robust changes in optimal screening strategies, resulting from the use of alternative cost values for screening modalities or biopsy, are presented in Table A.13.

Table A.12: Cost Range of Screening Modalities and Biopsy

Imaging Technology/Cost	Minimum (Low)	Average (Medium)	Maximum (High)
Mammography (MAM)	$94	$116	$161
MRI	$1,075	$1,395	$2,016
Ultrasound (US)	$94	$94	$94
MRI adjunct to MAM	$1,169	$1,511	$2,177
Biopsy	$1,379	$1,693	$2,363

Table A.13: Robust Strategy Modifications due to Changes in Cost

Revised Policy	MAM Low	MAM High	MRI Low
Annual MRI+MAM between age 25-34	-	-	ICER COST-EFFECTIVE
Biannual MAM between age 45-54	-	NOT OPTIMAL	-
1 MAM at age 55	-	DOMINATED	-
1 MAM at age 65	-	DOMINATED	-
1 US at age 75	-	ICER COST-EFFECTIVE	-
1 MAM at age 75	-	DOMINATED	-
2 MAM at ages 75 and 80	ICER COST-EFFECTIVE	-	-
1 MAM at age 85	-	NOT OPTIMAL	-
2 MAM at ages 85 and 90	-	DOMINATED	-

Our study reveals two important findings regarding the impact of screening costs on optimal strategies: (1) The most critical cost factor, affecting both the optimality and cost-effectiveness of identified strategies, is the cost of mammography. Under high (maximum) cost scenario of mammography, "single MAM" loses its cost-effectiveness and/or optimality at all age groups (also making "single US" cost-effective) and "double MAM" becomes dominated for all "elderly" women. On the other hand, a decrease in the cost of mammography affects only "double MAM", by making it cost-effective for all "elderly" women. (2) "Annual MRI+MAM" becomes cost-effective for 25-year-old high-risk women under all scenarios of low MRI cost. This result suggests that lowering the cost of MRI can make "Annual MRI+MAM" a cost-effective strategy for "young" high-risk women and that the cost of MRI plays a more critical role than cost of biopsy and mammography on cost-effectiveness of this strategy.

A.4.2 Sensitivity Analysis - Disutility Function

We utilize a disutility function, capturing the harms associated with a screening action, for elderly high-risk populations, for whom tolerance to aggressive screening is a significant concern and the benefits of screenings diminish due to increased comorbidities and complications [103, 51]. In our base case study, we consider a linear disutility function increasing with age both for screening technologies and biopsy for women over age 75. In this section, we investigate the impact of an alternative (i) structure (i.e., constant rather than linearly increasing) and (ii) initiation age (e.g., age 55 or 65). For "young" age groups, these changes in disutility function have no impact on their ten-year strategies. For "middle-aged" high-risk women, sensitivity analysis reveals that disutility initiation age affects cost-effectiveness or optimality of "annual MAM" screening only. Activating disutility at age 65 causes "annual MAM" strategy to be sub-optimal for women above 65 years old and activating it at age 55 makes "annual MAM" either sub-optimal or dominated (hence not cost-effective) for all "middle-aged" women. These results suggest that when

affordable, "annual US", instead of annual "MAM", can be used for "middle-aged" high-risk individuals who have a significant intolerance to biopsy procedure or for whom biopsy has a significantly elevated complication risk. Finally, for "elderly" high-risk women, the structure of the disutility function has an impact on the cost-effectiveness of the strategies with the highest health benefits (namely, "annual US" for 75- and "biennial US" for 85-year old women). With a constant structure, "annual US" and "biennial US" both become cost-effective for 75-and 85-year old women, respectively.

A.5 Proofs of Analytical Results

In this section, we provide the proofs of the analytical results we presented in the main text. We use the following notational conventions about the summation and product signs: $\sum_{i=S}^{M} ... = 0$ and $\prod_{i=S}^{M} ...$ =1 when M $<$ S. We start with introducing additional notations that will be needed for the proofs (in addition to the ones defined in the main text):

μ_t**(i):** Proportion of women in health state i $\in \{0,1,2\}$ in the targeted (sub-)population at time t $\in \{1,2,...,10\}$, which changes probabilistically over time as a function of disease prevalence, incidence and progression with respect to the natural history of breast cancer. Accordingly, $\mu_1(i)$ corresponds to the the initial health state distribution for $i \in S_U$ and is equal to $\sum_{a \in A} x_1(i,a)$.

$\hat{\mu_t}$**(i**$|\pi,\pi'$**):** Proportion of women with so far undetected stage $i \in \{1,2\}$ cancer at time t $\in \{1,2,...,10\}$, when strategy π or π' is implemented. That is, $\hat{\mu_t}$(i$|\pi,\pi'$) corresponds the fraction of stage i cancers that would not be detected by either one of the strategies π or π', regardless of which one of them has been implemented by time t.

Next, we restate Assumptions (1)-(3) and present the resulting mathematical statements imposed by these assumptions.

Assumption 1: *(Detection is better than no detection) For both in situ (i=1) and invasive (i=2) cancer states, detection at any time in the next ten-year period, i.e.* $t \in \{1,2,...,10\}$*, yields higher expected QALYs than the scenario with no detection by the end*

of the next ten-year period (followed by a possible detection under any future strategy ϕ implemented after ten years). Accordingly, the conditions (1.1) and (1.2), expressed below, hold for all t $\in \{1,2,...,10\}$ under any future strategy ϕ:

(1.1) Rew(t,ϕ|1):= $R_t(1) - [r_t(1)+\sum_{j=t}^{9}[\prod_{s=t}^{j} P_s(1|1)r_{j+1}(1)]+ \prod_{j=t}^{10} P_j(1|1)E_\phi[s_{11}=1]) +$
$\sum_{s=1}^{10-t}[\prod_{j=t}^{t+s-2} P_j(1|1)P_{t+s-1}(2|1)[r_{s+t}(2)+\sum_{k=t+s}^{9}[\prod_{j=s+t}^{k} P_j(2|2)r_{k+1}(2)]+ \prod_{j=s+t}^{10} P_j(2|2)E_\phi[s_{11}=2]]+$
$\prod_{k=t}^{9} P_k(1|1)P_{10}(2|1)E_\phi[s_{11}=2]) > 0$

(1.2) Rew(t,ϕ|2):= $R_t(2) - [r_t(2)+\sum_{j=t}^{9}[\prod_{s=t}^{j} P_s(2|2)r_{j+1}(2)]+ \prod_{j=t}^{10} P_j(2|2)E_\phi[s_{11}=2]) > 0$

Assumption 2: *(Early detection is better than late detection) For both in situ (i=1) and invasive (i=2) cancer states, detection at any time t $\in \{1,2,...,9\}$ yields higher expected QALYs than detection at a later time point t+k in the ten-year period, where k $\in \{1,2,...,10-t\}$. Accordingly, the conditions (2.1) and (2.2) hold for all t $\in \{1,2,...,9\}$, k $\in \{1,2,...,10-t\}$:*

(2.1) Rew(t,k|1)=: $R_t(1) - r_t(1)+\sum_{j=t}^{t+k-2}[\prod_{s=t}^{j} P_s(1|1)r_{j+1}(1)]+ \prod_{j=t}^{t+k-1} P_j(1|1)R_{t+k}(1) +$
$\sum_{s=1}^{k-1}[\prod_{j=t}^{t+s-2} P_j(1|1)P_{t+s-1}(2|1)[r_{s+t}(2)+\sum_{m=t+s}^{t+k-2}[\prod_{j=s+t}^{m} P_j(2|2)r_{m+1}(2)]+ \prod_{j=s+t}^{t+k-1} P_j(2|2)R_{t+k}(2)]]+$
$\prod_{j=t}^{t+k-2} P_j(1|1)P_{t+k-1}(2|1)R_{t+k}(2) > 0$

(2.2) Rew(t,k|2)=: $R_t(2) - r_t(2)+\sum_{j=t}^{t+k-2}[\prod_{s=t}^{j} P_s(2|2)r_{j+1}(2)]+ \prod_{j=t}^{t+k-1} P_j(2|2)R_{t+k}(2) > 0$

Assumption 3: *Sensitivity of a screening modality $a \in A_S$, denoted by $sens_t(a)$, is greater than 50% at any time t $\in \{1,2,...,10\}$. That is, $sens_t(\pi) \geq 0.50$ for all t $\in \{1,2,...,10\}$ and for any screening strategy π.*

We use "$\prec$" to denote the order between the strategies in terms of expected total QALYs. That is, $\pi \prec \pi'$ means that strategy π' yields more expected QALYs than strategy π. We construct proofs by going backward in time and visiting each time point the compared strategies differ. For future reference, we name the time points, where the compared strategies (and hence, the clinical courses of women with breast cancer following these strategies) differ, as *critical time points*.

A.5.1 Proof of Proposition 1

Proposition 1: *Between two strategies using the same screening modality with different frequencies, the strategy with higher screening frequency yields higher expected QALYs, regardless of the initial health state distribution and future strategy implemented.*

To prove *Proposition (1)*, we fix the screening modality being used, call ψ, and systematically compare the strategies utilizing this modality with different frequencies. In this context, the *critical time points*, where different strategies lead to different outcomes, are the times of extra screenings implemented by the more frequent strategy. At each *critical time point*, a portion of so far undetected cancers is detected with the extra screening of the more frequent ten-year strategy, denoted as strategy π'. These cancers either are detected later or remain undetected by the end of ten-year screening period when the less frequent ten-year strategy, denoted as strategy π, is implemented instead. Accordingly, to prove *Proposition (1)*, we show that for all critical time points, the detections under the extra screenings of the more frequent strategy π' yield more expected QALYs than the QALYs obtained with the alternative clinical course and outcomes observed under the less frequent strategy π (in the ten-year period and afterwards). We complete this proof in four steps: *Propositions (1.1)*, *(1.2)*, *(1.3)*, and *(1.4)*.

Proposition (1.1): No Intervention Strategy $\pi \prec$ Single-Screen Strategy π'

The only difference between "no intervention strategy" π and "single-screen strategy" π' is the clinical course of women whose cancers are detected at time t=1 under strategy π'. These cancer patients leave the screening process and initiate their cancer treatments under strategy π'. However, if "no intervention strategy" π is implemented, these cancers remain undetected by the end of the ten-year interval. Accordingly, for a fixed future strategy ϕ, implemented after ten years, the total expected QALYs difference between the strategies π' and π is equal to μ_1**(1)sens**$_1(\psi)$**Rew(1,**ϕ**|1)+** μ_1**(2)sens**$_1(\psi)$**Rew(1,**ϕ**|2)**. In this expression, the term μ_1**(i)sens**$_1(\psi)$ corresponds the proportion of cancers detected in stage $i \in \{1,2\}$

at time t=1 by strategy π'. **Rew(1,ϕ|i)** is the total expected QALYs difference between strategies π' and π for a woman with stage $i \in \{1,2\}$ cancer at time t=1. This difference in the expected QALYs arises from the difference in the course of women with breast cancer under strategies π' and π. Figure A.1 depicts the expected QALYs difference between the strategies π' and π for women with invasive cancer (i=2) at time t=1. Since $\mu_1(1) > 0$ and $\mu_2(1) > 0$ for any initial health state distribution, $\text{sens}_1(\psi) > 0$ and $\text{Rew}(1,\phi|i) > 0$ for $i \in \{1,2\}$ by *Assumption (1)*, we conclude that "Single-Screen" strategy π' yields more total expected QALYs than "No Intervention" strategy π for any future strategy ϕ.

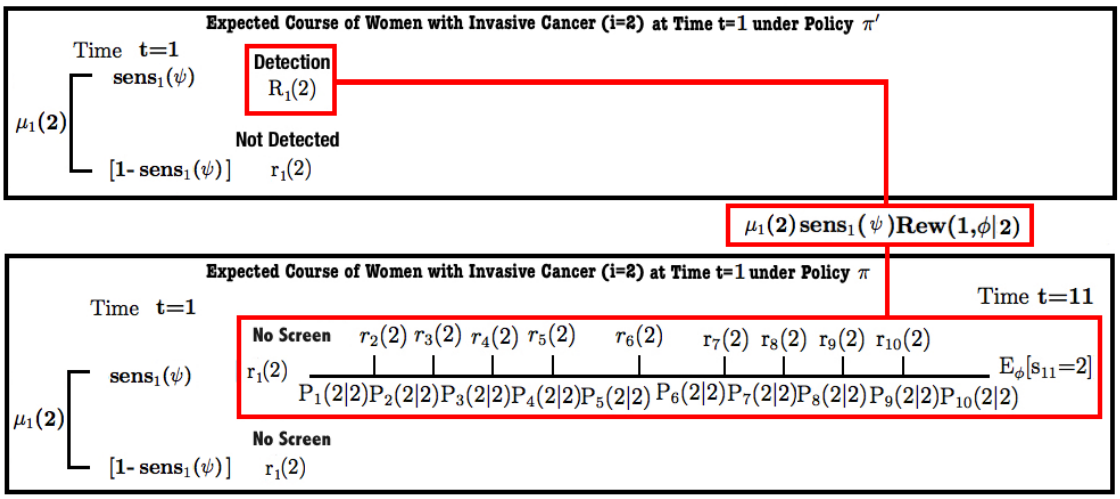


Figure A.1: The Difference Between the Expected Courses of Women with Invasive Cancer under No Intervention Strategy π and Single-Screen Strategy π'

Proposition (1.2): Single-Screen strategy $\pi \prec$ Double-Screen strategy π'

Under the single-screen and double-screen strategies π and π' the expected clinical course of healthy women and women with cancer does not differentiate until time t=6. At time t=6, an additional screening is conducted by strategy π', detecting $\hat{\mu}_6$**(i|π,π')sens**$_6(\psi)$ cancers in stage $i \in \{1,2\}$. These patients with detected breast cancer are assigned a one-time lump sum reward R_6(i), corresponding to expected lifetime QALYs after stage i cancer treatment, and depart the screening process. Under single-screen strategy π, the cancers of these patients remain undetected by the end of the ten-year interval, and the patients continue to receive annual rewards associated with their health states. Given

they survive, they receive further rewards after the ten-year period, depending on the implemented future strategy ϕ. **Rew(6,ϕ|i)** captures the total expected QALYs difference between these two clinical scenarios observed under the strategies π and π' for a woman with stage i $\in \{1,2\}$ cancer at time t=6. Accordingly, for a fixed future strategy ϕ, the total expected QALYs difference between the strategies π' and π is equal to $\hat{\mu_6}$**(1**$|\pi,\pi'$**)sens**$_6(\psi)$**Rew(6,**ϕ**|1)+** $\hat{\mu_6}$**(2**$|\pi,\pi'$**)sens**$_6(\psi)$**Rew(6,**ϕ**|2)** (Figure A.2). As $\hat{\mu_6}(1|\pi,\pi') > 0$ and $\hat{\mu_6}(2|\pi,\pi') > 0$ for any initial health state distribution, sens$_6(\psi)>0$ and Rew(6,ϕ|i)>0 for $i \in \{1,2\}$ by *Assumption (1)*, we conclude that "Double-Screen" strategy π' yields more total QALYs than "Single-Screen strategy" π for any future strategy ϕ.

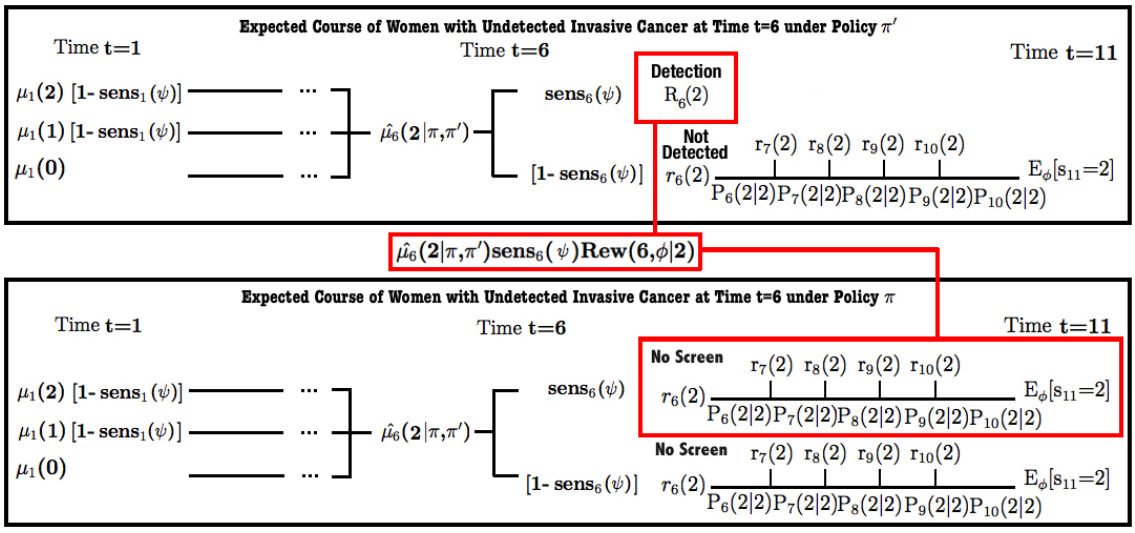


Figure A.2: The Difference Between the Expected Courses of Women with Invasive Cancer under Single-Screen strategy π and Double-Screen strategy π'

Proposition (1.3): Double-Screen strategy $\pi \prec$ Biennial-Screen strategy π'

Biennial screening strategy π' has screenings at time t=1, 3, 5, 7, and 9 whereas double-screen strategy π recommends screenings at time t=1 and 6. We show the superiority of strategy π' step-by-step by visiting all critical points backward in time and stop at the time point t=3, where the strategies differ first. In order to reduce algebraic computations, we shift the screening at time t=5 (under strategy π') to time t=6 and only focus on time points t=3, 7 and 9 (rather than time t=5 and 6 as well).[1] Figure A.3 demonstrates the total QALYs

[1]This shift is justified because *Assumption (2)* states that early detection yields higher QALYs in terms of

differences at each critical time point $t \in \{3, 7, 9\}$.

We start the proof with the latest critical point t=9, where biennial strategy π' recommends an additional screening. $\hat{\mu_9}(1|\pi,\pi')$ and $\hat{\mu_9}(2|\pi,\pi')$ respectively corresponds to the proportions of women with undetected in situ (i=1) and invasive (i=2) cancers at time t=9, regardless of whether strategy π' or π has been implemented by this time point. [100*sens$_9(\psi)$]% of these cancers are expected to be detected by the screening recommended by strategy π' at time t=9. Then, for a fixed future strategy ϕ, the total expected QALYs difference of the strategies π' and π for so-far undetected cancer cases at time t=9 is equal to **[$\hat{\mu_9}$(1|π,π')sens$_9(\psi)$Rew(9,ϕ|1)+ $\hat{\mu_9}$(2|π,π')sens$_9(\psi)$Rew(9,ϕ|2)]**. This term is positive for any initial health state distribution and future strategy ϕ since Rew(9,ϕ|i)$>$ 0 holds for $i \in \{1,2\}$ by *Assumption (1)*.

- Time **t=9**: $\hat{\mu_9}(1|\pi,\pi')\text{sens}_9(\psi)\text{Rew}(9,\phi|1) + \hat{\mu_9}(2|\pi,\pi')\text{sens}_9(\psi)\text{Rew}(9,\phi|2)$
- Time **t=7**: $\hat{\mu_7}(1|\pi,\pi')\text{sens}_7(\psi)\text{Rew}(7,\phi|1) + \hat{\mu_7}(2|\pi,\pi')\text{sens}_7(\psi)\text{Rew}(7,\phi|2)$
- Time **t=3**: $\hat{\mu_3}(1|\pi,\pi')\text{sens}_3(\psi)[\text{sens}_6(\psi)\text{Rew}(3,3|1)+[1-\text{sens}_6(\psi)]\text{Rew}(3,\phi|1)]+$
  $\hat{\mu_3}(2|\pi,\pi')\text{sens}_3(\psi)[\text{sens}_6(\psi)\text{Rew}(3,3|2)+[1-\text{sens}_6(\psi)]\text{Rew}(3,\phi|2)]$

Figure A.3: The QALYs Differences between Biennial-Screen strategy π' and Double-Screen strategy π due to Screenings at time t=3, 7, and 9

Now, we proceed to the preceding critical time point t=7, where biennial screening strategy π' recommends a screening. $\hat{\mu_7}$(i|π,π') corresponds to the proportions of women with so-far undetected stage i,$i \in \{1,2\}$, cancers at time t=7, regardless of whether strategy π' or π has been implemented by this time point. We will show that strategy π', conducting a screening at time t=7 and another one at time t=9, yields more total expected QALYs for these so-far undetected cancer cases. Among these cancer patients, we can discard the cancer cases that will be detected later at time t=9, since we have already shown the superiority of biennial strategy π' for them. Hence, to prove the superiority of π' for women with so-far

time and hence, by shifting a screening to a later time, we cause strategy π' to lose some of its advantage on strategy π.

undetected cancers at time t=7, it suffices to focus on cancer cases that are detected at time t=7 by strategy π'. For a fixed future strategy ϕ, the total expected QALYs difference between strategies π' and π for these cancer cases is equal to **[**$\hat{\mu_7}$**(1**$|\pi$**,**π'**)sens**$_7(\psi)$**Rew(7,**ϕ**|1)+** $\hat{\mu_7}$**(2**$|\pi$**,**π'**)sens**$_7(\psi)$**Rew(7,**ϕ**|2)]**, which is positive for any initial health state distribution and future strategy ϕ.

Finally, we proceed the first critical time point t=3. At time t=3, strategy π' recommends a screening, and detects a portion of so far undetected cancers. The expected course of cancers that are not detected at time t=3 by the strategy π' is the same under both strategies π' or π by time t=7, where the strategies differ again. We can discard these cases, as we have already shown the superiority of π' at time t=7 and afterwards, and can turn our attention to the cancers that are detected under strategy π' at time t=3. Under double-screen strategy π, these cancers remain undetected at time t=3, as no screening is conducted, a fraction of them are expected to be detected later at time t=6 and given they survive, the rest remains undetected at least by the end of ten-year period with their future clinical course depending on the future strategy ϕ. By taking all these possible scenarios into account, the total expected QALYs difference between strategies π' and π for these cancer cases can be expressed by $\hat{\mu_3}$**(1**$|\pi$**,**π'**)sens**$_3(\psi)$**[sens**$_6(\psi)$**Rew(3,3|1)+[1- sens**$_6(\psi)$**]Rew(3,**ϕ**|1)]+** $\hat{\mu_3}$**(2**$|\pi$**,**π'**)sens**$_3(\psi)$**[sens**$_6(\psi)$**Rew(3,3|2)+[1- sens**$_6(\psi)$**]Rew(3,**ϕ**|2)]**. The term Rew(3,3|i) captures the expected QALYs difference between detecting a stage i cancer at time t=3 or later at time t=6 (3+3). The term Rew(3,ϕ|i) captures the expected QALYs difference between detecting a stage i cancer at time t=3 and not detecting it by time t=10, with possible detection after the ten-year period under the future strategy ϕ. By *Assumptions (1)* and *(2)*, the terms Rew(3,3|i) and Rew(3,ϕ|i) are both positive for both cancer states i=1,2. As a result, the overall term $\hat{\mu_3}$**(1**$|\pi$**,**π'**)sens**$_3(\psi)$**[sens**$_6(\psi)$**Rew(3,3|1)+[1- sens**$_6(\psi)$**]Rew(3,**ϕ**|1)]+** $\hat{\mu_3}$**(2**$|\pi$**,**π'**) sens**$_3(\psi)$ **[sens**$_6(\psi)$**Rew(3,3|2)+[1- sens**$_6(\psi)$**]Rew(3,**ϕ**|2)]** is > 0 for any health state distribution μ and future strategy ϕ.

Combining the results for the critical time points t=3, 7 and 9, we conclude that biennial

screening strategy π' yields more expected QALYs than double-screen strategy π over the entire ten-year period under any health state distribution μ and future strategy ϕ.

Proposition (1.4): Biennial-Screen strategy $\pi \prec$ Annual-Screen strategy π'

The biennial and annual strategies π and π' differ at times t=2, 4, 6, 8 and 10, where only the annual-screen strategy recommends screenings. Again, we construct our proof iteratively by analyzing these critical time points backward in time and show that the annual-screen strategy π' yields higher QALYs at each critical time point $t \in \{2, 4, 6, 8, 10\}$ for any health state distribution μ and future strategy ϕ. Figure A.4 demonstrates the total QALYs differences between the strategies π' and π for so-far undetected cancer cases at each critical time point.

- Time **t=10**: $\hat{\mu_{10}}(\mathbf{1}|\pi,\pi')\mathbf{sens}_{10}(\psi)\mathbf{Rew(10,}\phi|\mathbf{1)} + \hat{\mu_{10}}(\mathbf{2}|\pi,\pi')\mathbf{(2)}\ \mathbf{sens}_{10}(\psi)\mathbf{Rew(10,}\phi|\mathbf{2)}$
- Time **t=8**: $\hat{\mu_8}(\mathbf{1}|\pi,\pi')\mathbf{sens}_8(\psi)[\mathbf{sens}_9(\psi)\mathbf{Rew(8,1|1)}+[\mathbf{1-}\ \mathbf{sens}_9(\psi)]\mathbf{Rew(8,}\phi|\mathbf{1)}]+$
  $\hat{\mu_8}(\mathbf{2}|\pi,\pi')\mathbf{sens}_8(\psi)[\mathbf{sens}_9(\psi)\mathbf{Rew(8,1|2)}+[\mathbf{1-}\ \mathbf{sens}_9(\psi)]\mathbf{Rew(8,}\phi|\mathbf{2)}]$
- Time **t=6**: $\sum_{i=1,2} \hat{\mu_6}(\mathbf{i}|\pi,\pi')\ \mathbf{sens}_6(\psi)\ [\mathbf{sens}_7(\psi)\mathbf{Rew(6,1|i)}+\ [\mathbf{1-}\ \mathbf{sens}_7(\psi)]\mathbf{sens}_9(\psi)$
  $\mathbf{Rew(6,3|i)}+[\mathbf{1-sens}_7(\psi)][\mathbf{1-sens}_9(\psi)]\mathbf{Rew(6,}\phi|\mathbf{i)}]$
- Time **t=4**: $\sum_{i=1,2} \hat{\mu_4}(\mathbf{i}|\pi,\pi')\mathbf{sens}_4(\psi)\ [\mathbf{sens}_5(\psi)\mathbf{Rew(4,1|i)}+\ [\mathbf{1-}\ \mathbf{sens}_5(\psi)]\mathbf{sens}_7(\psi)\mathbf{Rew(4,3|i)}$
  $+\ [\mathbf{1-sens}_5(\psi)][\mathbf{1-sens}_7(\psi)]\mathbf{sens}_9(\psi)\mathbf{Rew(4,5|i)}\ +\ [\mathbf{1-sens}_5(\psi)][\mathbf{1-sens}_7(\psi)]$
  $[\mathbf{1-sens}_9(\psi)]\ \mathbf{Rew(4,}\phi|\mathbf{i)}]$
- Time **t=2**: $\sum_{i=1,2} \hat{\mu_2}(\mathbf{i}|\pi,\pi')\ \mathbf{sens}_2(\psi)\ [\mathbf{sens}_3(\psi)\mathbf{Rew(2,1|i)}+\ [\mathbf{1-}\ \mathbf{sens}_3(\psi)]\mathbf{sens}_5(\psi)\mathbf{Rew(2,3|i)}$
  $+\ [\mathbf{1-sens}_3(\psi)][\mathbf{1-sens}_5(\psi)]\mathbf{sens}_7(\psi)\mathbf{Rew(2,5|i)}\ +\ [\mathbf{1-sens}_3(\psi)][\mathbf{1-sens}_5(\psi)]$
  $[\mathbf{1-sens}_7(\psi)]\mathbf{sens}_9(\psi)\mathbf{Rew(2,7|i)} + [\mathbf{1-sens}_3(\psi)][\mathbf{1-sens}_5(\psi)][\mathbf{1-sens}_7(\psi)]$
  $[\mathbf{1-sens}_9(\psi)]\mathbf{Rew(2,}\phi|\mathbf{i)}]$

Figure A.4: The QALYs Differences between Annual-Screen strategy π' and Biennial-Screen strategy π due to Screenings at time t=2, 4, 6, 8, and 10

Time t=2 is the first time point the strategies π and π' differ in the ten-year screening period. At time t=2, a screening is performed under annual strategy π' and detects a portion of so-far undetected cancer cases. These cancers remain undetected at time t=2 if biennial strategy π is implemented instead but they might be detected later by one of the screenings performed by strategy π at times t= 3, 5, 7, 9 or after the first ten-year period by future strategy ϕ. The total expected QALYs between the strategies π and π' for so-far

undetected cancers at time t=2 is computed by taking all of these scenarios into account (Figure A.4). The component Rew(2,k|i) captures the expected QALYs between detecting a stage i cancer at time t=2 or later at time $t+k$. The term Rew(2,ϕ|i) captures the expected QALYs between detecting a stage i cancer at time t=2 or not detecting it by time t=10, with possible detection later with the future strategy ϕ. These terms (i.e., Rew(2,k|i) for i= 1, 2 and k= 1, 3, 5, 7, and Rew(2,ϕ|i) for i= 1, 2 and any future strategy ϕ) are positive by *Assumptions (1)* and *(2)*, and $\hat{\mu_2}(i|\pi,\pi') \geq 0$ for $i \in \{1,2\}$ for any health state distribution μ. Accordingly, the whole term capturing the total QALYs difference between π and π' at time t=2 is positive. This result shows the superiority of annual strategy π' over biennial strategy π for the cancers detected under strategy π' at time t=2.

We also need to show the superiority of strategy π' for the cancer cases that are not detected by the additional screening of strategy π' at time t=2. These undetected cancers might be detected at times t= 3, 5, 7, 9, which is a clinical scenario that doesn't impact the QALYs difference between the strategies π and π' as both recommend screenings at these points. Alternatively, if they survive, they join the proportion of women with so-far undetected cancers at time t= 4, 6, 8, 10. As a result, we need to show the superiority of strategy π' at times $t \in \{4, 6, 8, 10\}$ to demonstrate the superiority of strategy π' at time t=2 over the entire targeted population. That's why, we follow a backward-style proof since the complete result for any critical time point, where the strategies differ, relies on the results shown for the critical time points that will come later.

The proofs for the time points $t \in \{4, 6, 8, 10\}$ follow the same structure as of the proof for the time point t=2 presented above. That is, the objective is the show that the total expected QALYs differences between strategies π' and π at each one of these critical time points are positive, and the results follow from the fact that the terms Rew(t,k|i) and Rew(t,ϕ|i) are positive by *Assumptions (1)* and *(2)*, and $\hat{\mu_t}(i|\pi,\pi') \geq 0$. Accordingly, the overall terms (Figure A.4), each capturing the total expected QALYs difference at one of the critical time points, are positive. Combined, these results show that annual-screening

strategy π' yields higher total expected QALYs than biennial screening strategy π over the entire ten-year screening period for any future strategy ϕ and any health distribution μ.

A.5.2 Proof of Proposition 2

Proposition 2: *Between two strategies using different screening modalities with the same frequency, the strategy utilizing a more sensitive modality yields higher expected QALYs, regardless of the initial health state distribution and future strategy implemented.*

To prove *Proposition (2)*, we compare the strategies that utilize different modalities with different sensitivity rates under the same fixed screening schedule. The critical time points, where the difference in compared strategies causes a difference in the clinical course of women with cancers, correspond to the time points where screenings are conducted. For any two strategies that are compared, the expected result of screenings with different modalities can be categorized into three groups: (i) so far undetected cancer cases that remain undetected under both strategies, (ii) so far undetected cancer cases that are detected under both strategies and (iii) so far undetected cancer cases that are detected under the strategy using the more sensitive modality but remains undetected at that time point under the other strategy. Since the implementation of one strategy, instead of the other, affects the clinical course of the women in group (iii) only, it suffices to focus on these patients to prove *Proposition (2)*.

We evaluate the four practical schedule cases (i.e., 1, 2, 5 and 10 screenings in the ten-year interval) separately and complete the proof in four steps: *Propositions (2.1)*, *(2.2)*, *(2.3)*, and *(2.4)*. For each fixed screening schedule, we begin with the latest critical time point, where a screening is recommended, and then visit all critical points step-by-step backward in time because the proof at any critical time point relies on the proof of a later point in time. We use $\text{sens}_t(\pi)$ and $\text{sens}_t(\pi')$ to respectively denote the sensitivity of the modality used at time $t \in \{1, ..., 10\}$ by strategies π' and π, where strategy π' always uses a more sensitive modality (i.e $\text{sens}_t(\pi') > \text{sens}_t(\pi) \; \forall \; t \in \{1, ..., 10\}$).

Proposition (2.1): Single Screening Strategy (Frequency=1)

Under single screening strategy, a screening is performed at the beginning (i.e., at time t=1) and no other intervention is taken during the ten-year interval. Due to utilizing a more sensitive technology, strategy π' is expected to detect μ_1**(i)[sens**$_1(\pi')$**-sens**$_1(\pi)$**]** more stage i cancers at time t=1. These cancers remain undetected under strategy π by the end of the ten-year period and might be detected later by the future strategy ϕ. Since the clinical course of cancer cases (Figure A.5) constitute the only difference between the impacts of strategies π' and π, the total expected QALYs difference between these two policies is equal to μ_1**(1)[sens**$_1(\pi')$**-sens**$_1(\pi)$**]Rew(1,**ϕ**|1)+**μ_1**(2)[sens**$_1(\pi')$**-sens**$_1(\pi)$**]Rew(1,**ϕ**|2)** for a future strategy ϕ. The term Rew(1,ϕ|i) > 0 for $i \in \{1,2\}$ under any future strategy ϕ by *Assumption (1)*, $\mu_1(1) > 0$ and $\mu_1(2) > 0$ for any health state distribution μ, and sens$_1(\pi')>$ sens$_1(\pi)$ as the strategy π' uses a more sensitive modality. Hence, the overall term, capturing the total QALYs difference of strategies π' and π, is positive, which shows the superiority of strategy π' that utilizes a more sensitive modality than strategy π.

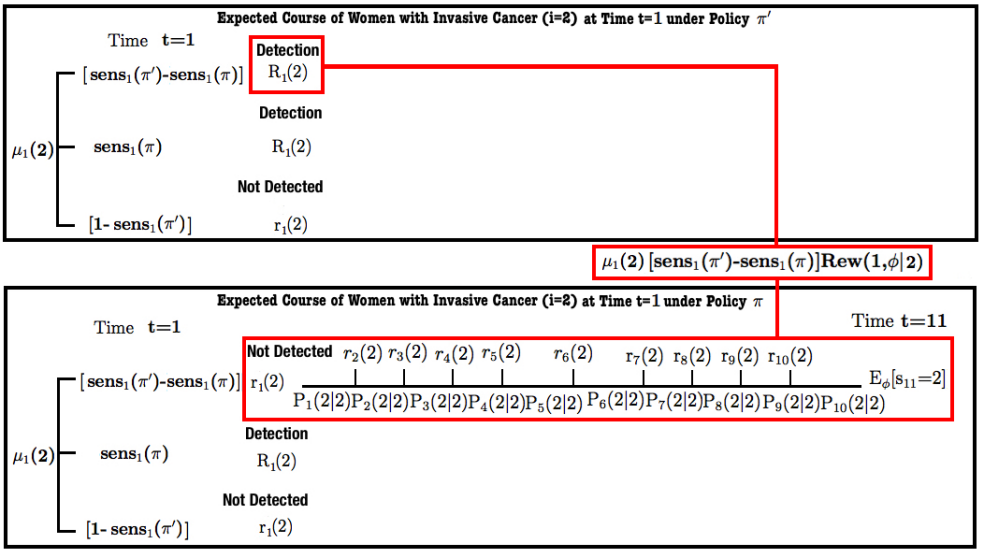


Figure A.5: The Difference Between the Expected Courses of Women with Invasive Cancer under Strategies π and π'

Proposition (2.2): Double Screening Strategy (Frequency=2)

Double screening strategies π' and π conduct two screenings in the ten-year interval,

one at time t=1 and the other at time t=6. We analyze the critical time point t=6 first and then t=1. Regardless of whether strategy π' or π has been implemented by time t=6, there are women with so-far undetected cancers at time t=6. Among these women, strategy π' is expected to detect $\hat{\mu_6}$**(i**$|\pi,\pi'$**)[sens**$_6(\pi')$**-sens**$_6(\pi)$**]** more stage i cancers at time t=6 due to utilizing a more sensitive modality. These cancers remain undetected under strategy π by the end of ten-year period and might be detected later by the future strategy ϕ. Then, the total expected QALYs difference between strategies π' and π for these so-far undetected cancer cases at time t=6 is equal to $\sum_{i=1,2} \hat{\mu_6}$**(i**$|\pi,\pi'$**)[sens**$_6(\pi')$**-sens**$_6(\pi)$**]Rew(6,**ϕ**|i)** for a future strategy ϕ. This term is positive since Rew(6,ϕ|i) > 0 for i$\in \{1,2\}$ under any future strategy ϕ by *Assumption (1)*, sens$_6(\pi')>$ sens$_6(\pi)$, and $\hat{\mu_6}$(i$|\pi,\pi') > 0$ for i$\in \{1,2\}$ with any health state distribution μ. As a result, strategy π', the strategy with more sensitive modality, yields more expected QALYs for women with so far undetected cancers at time t=6.

At time t=1, both strategies schedule a screening. Due to its more sensitive modality, strategy π' is expected to detect μ_1**(i)[sens**$_1(\pi')$**-sens**$_1(\pi)$**]** more stage i cancers at time t=1. These cancers remain undetected at time t=1 under strategy π but might be detected later either at time t=6, or after the ten-year period by the future strategy ϕ. Taking all these possible scenarios into account, the total expected QALYs difference between strategies π' and π at time t=1 is equal to $\sum_{i=1,2} \mu_1$**(i)[sens**$_1(\pi')$**-sens**$_1(\pi)$**][sens**$_6(\pi)$**Rew(1,5|i)+[1-sens**$_6(\pi)$**]Rew(1,**ϕ**|i)]**. Given Rew(1,5|i)[2]> 0 for i$\in \{1,2\}$ by *Assumption (2)*, Rew(1,ϕ|i)[3] > 0 for $i \in \{1,2\}$ under any future strategy ϕ by *Assumption (1)*, sens$_1(\pi')>$ sens$_1(\pi)$ and μ_1(i) > 0 for i$\in \{1,2\}$ with any health state distribution μ, the overall term for the total QALYs difference is positive, indicating that strategy π' yields more expected QALYs at time t=1. Combining the results for both critical time points t=1 and t=6, we conclude that for any future strategy ϕ and health state distribution μ, strategy π', the strategy with a more

[2]Rew(1,5|i) is the QALYs difference between detecting a stage i cancer at time t=1 or later at time t=6.

[3]Rew(1,ϕ|i) is the QALYs difference between the detection of a stage i cancer at time t=1 and no detection by the end of the ten-year period.

sensitive modality, yields higher expected QALYs when two screenings are scheduled in the ten-year screening interval.

Proposition (2.3): Biennial Screening Strategy (Frequency=5)

Biennial screening strategies π' and π both schedule five screenings at times t=1, 3, 5, 7 and 9. Since strategy π' uses a more sensitive modality at each $t \in \{1, 3, 5, 7, 9\}$, it is expected to detect $\hat{\mu}_t$**(i**$|\pi,\pi'$**)[sens**$_t(\pi')$**-sens**$_t(\pi)$**]** more stage i cancers among the women with so-far undetected cancers. If strategy π is implemented, these cancer cases, which are detected at time t only by strategy π', remain undetected at this time point but might be detected later by either a subsequent screening of strategy π or future strategy ϕ. The expected QALYs difference between strategies π' and π at $t \in \{1, 3, 5, 7, 9\}$ is calculated by probabilistically taking all of these clinical scenarios into account (Figure A.6). At each critical time point, the expected QALYs difference is positive as Rew(t,ϕ|i) > 0 by *Assumption (1)*, Rew(t,k|i) > 0 by *Assumption (2)*, sens$_t(\pi')$-sens$_t(\pi)$] > 0 due to the higher sensitivity of strategy π' modality, and $\hat{\mu}_t$(i|π,π') > 0 as μ is a probability distribution. Accordingly, the summation of the results for $t \in \{1, 3, 5, 7, 9\}$, each corresponding to the expected QALYs difference at one of critical time points, is positive, which shows that strategy π', the strategy using a more sensitive modality, yields more total expected QALYs over the entire ten-year screening interval.

Proposition (2.4): Annual Screening Strategy (Frequency=10)

Annual screening strategies π' and π both schedule yearly screenings all time points t=1,2,...,10 within ten-year period. Due to employing a more sensitive modality, strategy π' is expected to detect $\hat{\mu}_t$**(i**$|\pi,\pi'$**)[sens**$_t(\pi')$**-sens**$_t(\pi)$**]** more stage i cancers at each $t \in \{1, 2, ..., 10\}$ from the population with so-far undetected cancers. If strategy π is implemented, these cancer cases remain undetected at this time point but might be detected later by either a subsequent screening of strategy π or future strategy ϕ. The difference in the clinical course of these cancer cases leads to the expected QALYs difference between strategies π' and π at each time $t \in \{1, 2, ..., 10\}$, which is calculated by probabilistically taking

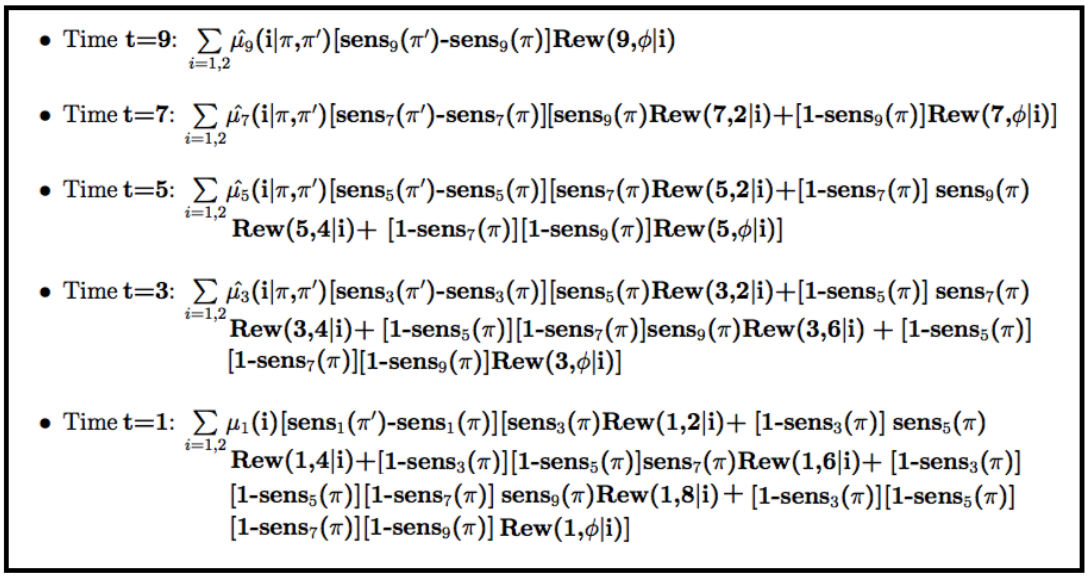


Figure A.6: The QALYs Differences between Biennial-Screen Strategies π' and π due to Screenings at time t=1, 3, 5, 7, and 9

all of the clinical scenarios into account (Figure A.7). At each time point $t \in \{1, 2, ..., 10\}$, the expected QALYs difference is positive since Rew(t,ϕ|i)[4] > 0 by *Assumption (1)*, Rew(t,k|i)[5] > 0 by *Assumption (2)*, sens$_t(\pi')$-sens$_t(\pi)$] > 0 due to the higher sensitivity of strategy π' modality, and $\hat{\mu}_t$(i|π,π') > 0 as μ is a probability distribution. Accordingly, the total QALYs difference, the summation of the results for $t \in \{1, 2, ..., 10\}$, is positive, which indicates that the strategy using a more sensitive modality (i.e., strategy π') yields more total expected QALYs under annual screening schedule.

A.5.3 Proof of Corollary 1

Corollary 1: *Among affordable ten-year screening strategies, either the one with the highest frequency or the one with the most sensitive modality is the optimal policy over the entire planning horizon for the given budget level. This result holds for any initial health state distribution and any future strategy that is implemented after the ten-year period.*

[4]Rew(t,ϕ|i) captures the expected QALYs difference resulting from the detection of a stage i cancer at time t and no detection by the end of the ten-year interval.

[5]Rew(t,k|i) captures the expected QALYs difference resulting from the detection of a stage i cancer at time t, by strategy π', or later at time t+k by strategy π.

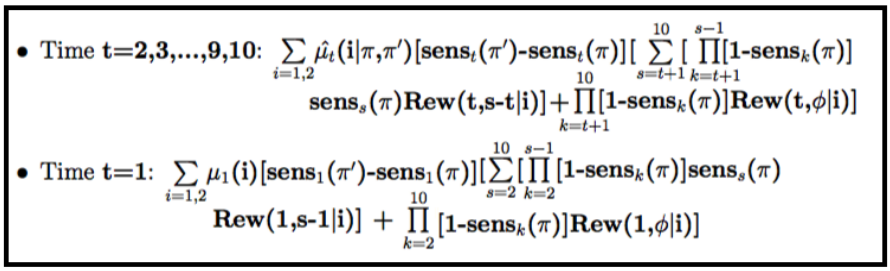


Figure A.7: The QALYs Differences between Annual-Screen Strategies π' and π due to Screenings at time t=1, 2, 3, ... and 10

The proof of *Corollary (1)* immediately follows from *Propositions (1)* and *(2)*. We show that *Corollary (1)* holds by "proof by contradiction". Assume that *Corollary (1)* is not true. Then there is an optimal ten-year screening strategy that achieves the highest expected QALYs, let's say strategy δ, without utilizing neither (i) the highest screening frequency nor (ii) the most sensitive modality that is affordable. Alternative (i) cannot be true since *Proposition (1)* states that the strategy using the same modality with strategy δ and having higher frequency yields more expected QALYs than strategy δ. Similarly, alternative (ii) cannot be true either since *Proposition (2)* states that the strategy having the same frequency with strategy δ and utilizing a more sensitive modality yields more expected QALYs than strategy δ. As a result, a strategy cannot be optimal without utilizing either (i) the highest frequency or (ii) the most sensitive modality options.

A.5.4 Proof Lemma 1

For a fixed given future strategy ϕ, *Lemma 1* establishes the conditions under which single screening strategy π is outperformed by "double" screening strategy π', which utilizes a less sensitive modality than strategy π. After proving *Lemma 1*, we generalize this result to all future strategies by proving *Theorem 1 (A)*.

Lemma 1: *Consider two ten-year screening strategies: single (i.e., every 10-year) screening strategy, call π, and double (i.e., every 5-year) screening strategy, call π', which utilizes a less sensitive modality (i.e., $sens_t(\pi') < sens_t(\pi)\ \forall\ t$ = 1,2,...,10). Then, under*

a fixed future strategy ϕ, strategy π' yields higher QALYs than strategy π, if the following condition (E.1) holds for both cancer states i=1 and i=2:

$$\frac{\mu_1(0)}{1-\mu_1(0)} > \frac{1}{Pr(s_6=i|s_1=0)} \frac{sens_1(\pi)-sens_1(\pi')}{sens_6(\pi')} \frac{[sens_6(\pi')Rew(1,5|i)+[1-sens_6(\pi')]Rew(1,\phi|i)]}{Rew(6,\phi|i)}$$

Proof (Lemma 1): Let either double-screen strategy π' or single-screen strategy π be implemented in the ten-year screening interval and a fixed future strategy ϕ be implemented afterwards. Both strategies π' and π conduct a screening at time t=1, and strategy π' conducts an additional screening at time t=6. Due to utilizing a more sensitive screening modality, strategy π is expected to detect μ_1(i)[sens$_1(\pi)$-sens$_1(\pi')$] more cancers in health state i $\in \{1,2\}$ at time t=1. These cancer cases are not detected at time t=1 if strategy π' is implemented but, if survive, they might be detected later by strategy π' at time t=6 or after the ten-year period by future strategy ϕ. Figure A.8 depicts the expected clinical course of these cancer cases under strategies π' and π, based on which the total expected QALYs surplus of strategy π over strategy π' at time t=1 is calculated. The total expected QALYs surplus of strategy π, due to utilizing a more sensitive modality at time t=1, is equal to $\sum_{i=1,2} \mu_1$**(i)[sens$_1(\pi)$-sens$_1(\pi')$][sens$_6(\pi')$Rew(1,5|i)+[1-sens$_6(\pi')$]Rew(1,ϕ|i)]**.

At time t=6, strategy π' conducts an additional screening and is expected to detect $\hat{\mu_6}$**(i**$|\pi,\pi'$**)sens**$_6(\pi')$ stage i cancers, which would remain undetected regardless whether strategy π' or π has been implemented by time t=6. Since single-screen strategy π does not conduct a screening at time t=6, these cancers remain undetected by the end of the ten-year interval. Figure A.9 depicts the expected clinical course of these cancer cases under strategies π' and π. The total expected QALYs surplus of strategy π', due to utilizing an additional screening at time t=6, is equal to $\sum_{i=1,2} \hat{\mu_6}$**(i**$|\pi,\pi'$**)sens**$_6(\pi')$**Rew(6,**ϕ|**i)** .

Accordingly, double-screen strategy π' generates a QALYs surplus at time t=6, due to utilizing an additional screening, and single-screen strategy π generates a QALYs surplus at time t=1, due to utilizing a more sensitive modality. Now, we will show that if the inequality **(E.1)** holds, then the QALYs surplus of double-screen strategy π' is higher and hence, strategy π' yields more expected QALYs. Assume that the inequality **(E.1)** holds for cancer

states i= 1,2. Then $\frac{\hat{\mu_6}(i|\pi,\pi')}{\mu_1(i)} > \frac{\mu_1(0)Pr(s_6=i|s_1=0)}{\mu_1(i)} > \frac{\mu_1(0)Pr(s_6=i|s_1=0)}{1-\mu_1(0)}$ holds for cancer states i= 1,2. This implies that $\frac{\hat{\mu_6}(i|\pi,\pi')}{\mu_1(i)} > \frac{sens_1(\pi)-sens_1(\pi')}{sens_6(\pi')} \frac{[sens_6(\pi')Rew(1,5|i)+[1-sens_6(\pi')]Rew(1,\phi|i)]}{Rew(6,\phi|i)}$ holds for i= 1,2. Accordingly, under the fixed future strategy ϕ, $\frac{\hat{\mu_6}(i|\pi,\pi')}{\mu_1(i)} \frac{sens_6(\pi')}{sens_1(\pi)-sens_1(\pi')} \frac{Rew(6,\phi|i)}{[sens_6(\pi')Rew(1,5|i)+[1-sens_6(\pi')]Rew(1,\phi|i)]} > 1$ holds for each cancer state i=1 and i=2, leading to $\sum_{i=1,2} \hat{\mu_6}(i|\pi,\pi')sens_6(\pi')Rew(6,\phi|i) > \sum_{i=1,2} \mu_1(i)[sens_1(\pi)\text{-}sens_1(\pi')][sens_6(\pi')\, Rew(1,5|i)+ [1\text{-}sens_6(\pi')]Rew(1,\phi|i)]$. This shows the superiority of double-screen strategy π' and concludes the proof.

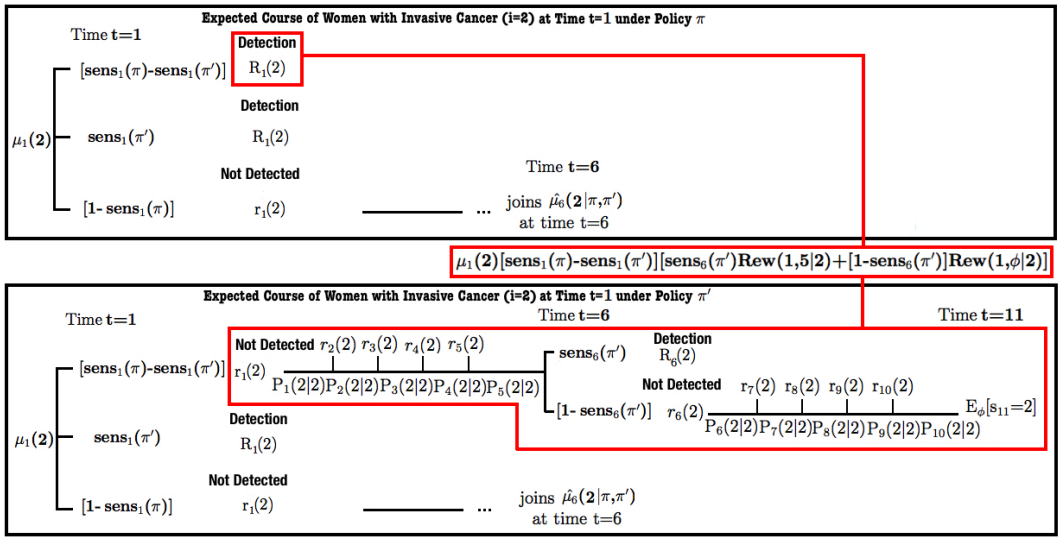


Figure A.8: The Clinical Course of Women with Invasive Cancer at time t=1 under Strategies π and π'

A.5.5 Proof for Theorem 1.A

Theorem 1.A: *Consider two ten-year screening strategies: single (i.e., every 10-year) screening strategy, call π, and double (i.e., every 5-year) screening strategy, call π', which utilizes a less sensitive modality (i.e., $sens_t(\pi') < sens_t(\pi)\ \forall\ t$ = 1,2,...,10). Then, under any future strategy, policy π' yields higher QALYs than policy π, if the following condition (E.2) holds:*

$$\frac{\mu_1(0)}{1-\mu_1(0)} > \frac{1}{Pr(s_6=i|s_1=0)} \frac{sens_1(\pi)-sens_1(\pi')}{sens_6(\pi')} \frac{sens_6(\pi')Rew(1,5|i)+[1-sens_6(\pi')]Rew(1|i)_{UB}}{Rew(6|i)_{LB}} \quad \textit{for i=1,2}$$

Proof (Theorem 1.A): When the inequality *(E.2)* holds the inequality **(E.1)** holds for

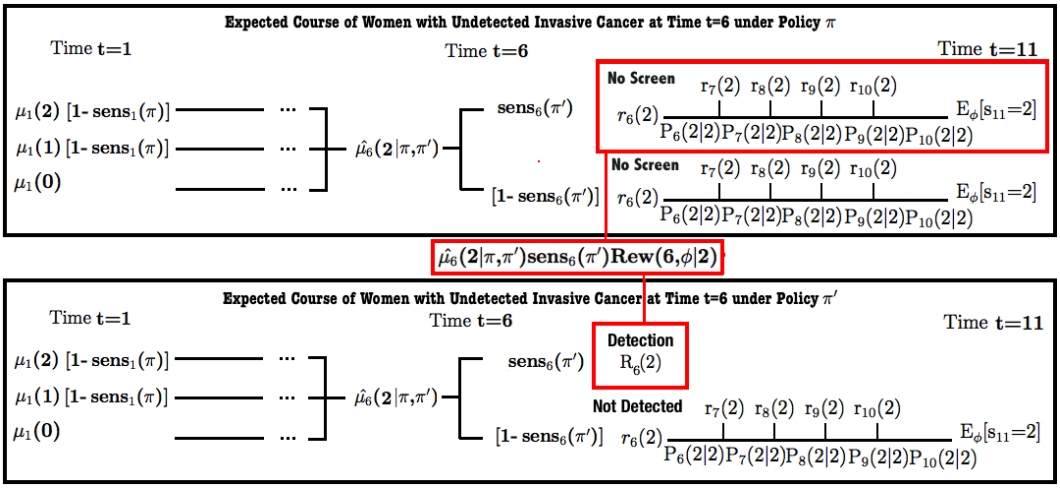


Figure A.9: The Clinical Course of Women with Invasive Cancer at time t=6 under Strategies π and π'

any future strategy ϕ since by definition, $Rew(1|i)_{UB} > Rew(1,\phi|i)$ and $Rew(6|i)_{LB} > Rew(6,\phi|i)$ for both in situ (i=1) and invasive (i=2) cancer states. Then, the result follows immediately from *Lemma 1*.

A.5.6 Proof for Theorem 1.B

Theorem 1.B: *Consider two ten-year screening strategies: double screening strategy, call π, and biennial (i.e., every 2-year) screening screening strategy, call π', which utilizes a less sensitive modality (i.e., $sens_t(\pi') < sens_t(\pi)\ \forall\ t$ = 1,2,...,10). Then, under a fixed future strategy ϕ, policy π' yields higher QALYs than policy π, if the following conditions (E.3)*[6] *and (E.4)*[7][8] *hold:*

$$\frac{\mu_1(0)}{\frac{1}{\prod_{t=1}^{5} P_t(0|0)} - \mu_1(0)} > \frac{1}{Pr(s_7=i|s_6=0)+Pr(s_9=i|s_6=0)} \frac{sens_6(\pi)}{min\{sens_7(\pi'),sens_9(\pi')\}} \frac{X(i)}{Rew(9,\phi|i)} \quad \textit{for i=1,2}$$

$$\frac{\mu_1(0)}{1-\mu_1(0)} > \frac{1}{Pr(s_3=i|s_1=0)+Pr(s_5=i|s_1=0)} \frac{sens_1(\pi)-sens_1(\pi')}{min\{sens_3(\pi'),sens_5(\pi')\}} \frac{Y(i)}{Z(i)} \quad \textit{for i=1,2.}$$

Proof (Theorem 1.B): Let either biennial-screen strategy π' or double-screen strategy

[6] X(i)=sens$_7(\pi)$Rew(6,1|i)+[1-sens$_7(\pi)$]sens$_9(\pi)$Rew(6,3|i)+[1-sens$_7(\pi)$][1-sens$_9(\pi)$]Rew(6,ϕ|i)

[7] Y(i)=sens$_3(\pi')$Rew(1,2|i)+[1-sens$_3(\pi')$]sens$_5(\pi')$Rew(1,4|i)+[1-sens$_3(\pi')$][1-sens$_5(\pi')$] sens$_7(\pi')$Rew(1,6|i)+[1-sens$_3(\pi')$][1-sens$_5(\pi')$][1-sens$_7(\pi')$]sens$_9(\pi')$Rew(1,8|i)+[1-sens$_3(\pi')$][1-sens$_5(\pi')$] [1-sens$_7(\pi')$][1-sens$_9(\pi')$]Rew(1,ϕ|i)

[8] Z(i)=$sens_6(\pi)min\{Rew(3,3|i), Rew(5,1|i)\} + [1 - sens_6(\pi)]min\{Rew(3,\phi|i), Rew(5,\phi|i)\}$

π be implemented in the ten-year screening interval and a fixed future strategy ϕ be implemented afterwards. Biennial-screen strategy π' conducts five screenings at time t= 1, 3, 5, 7 and 9. Double-screen strategy π utilizes a more sensitive modality and conducts two screenings at times t= 1 and 6. Accordingly, strategy π generates more QALYs at times t= 1, 6, whereas strategy π' yields QALYs surplus at times t= 3, 5, 7 and 9. We will show that the additional QALYs obtained under strategy π' offsets the additional benefits of strategy π when the inequalities *(E.3)* and *(E.4)* hold.

If the inequality *(E.3)* holds for both cancer states i=1,2, then the left hand side component $\frac{\mu_1(0)\prod_{t=1}^{5} P_t(0|0)}{1-\mu_1(0)\prod_{t=1}^{5} P_t(0|0)}$ [Pr(s_7=i|s_6=0)+Pr(s_9=i|s_6=0)] is greater than the right hand side $\frac{sens_6(\pi)}{min\{sens_7(\pi'),sens_9(\pi')\}}\ \frac{X(i)}{Rew(9,\phi|i)}$, which is equivalent to $\mu_6(0)$[Pr(s_7=i|s_6=0)+Pr(s_9=i|s_6=0)] min{$sens_7(\pi')$,$sens_9(\pi')$} Rew(9,ϕ|i) > [1-$\mu_6(0)$] $sens_6(\pi)$ X(i), $i \in \{1, 2\}$. Using the definition of *X(i)* and summing up both sides of the inequality over i=1,2, we derive the inequality $\sum_{i=1,2} \hat{\mu_7}(i|\pi,\pi')sens_7(\pi')Rew(7,\phi|i)+\hat{\mu_9}(i|\pi,\pi')sens_9(\pi')Rew(9,\phi|i) > \sum_{i=1,2} \hat{\mu_6}(i|\pi,\pi')$ $sens_6(\pi')[sens_7(\pi)\ Rew(6,1|i) + [1-sens_7(\pi)]sens_9(\pi)Rew(6,3|i) + [1-sens_7(\pi)][1-sens_9(\pi)]$ Rew(6,ϕ|i)].[9] This shows that the total expected QALYs surplus obtained under strategy π' due to the screenings at time points t=7, and 9 is higher than the QALYs surplus obtained under strategy π due to the screening conducted at time point t=6.

If the inequality *(E.4)* holds for both cancer states i=1, 2, then the inequality $\mu_1(0)$ [Pr(s_3=i|s_1=0)+Pr(s_5=i|s_1=0)] min{$sens_3(\pi')$,$sens_5(\pi')$} Z(i) > [1-$\mu_1(0)$] [$sens_1(\pi)$-$sens_1(\pi')$] Y(i) is satisfied. This implies that $\sum_{i=1,2} \hat{\mu_3}(i|\pi,\pi')sens_3(\pi')$ [$sens_6(\pi)$Rew(3,3|i)+[1-$sens_6(\pi)$] Rew(3,ϕ|i)]+$\hat{\mu_5}$(i|π,π')$sens_5(\pi')$[$sens_6(\pi)$Rew(5,1|i)+[1-$sens_6(\pi)$]Rew(5,ϕ|i)] is greater than the term $\sum_{i=1,2} \mu_1(i)$[$sens_1(\pi)$-$sens_1(\pi')$] Y(i) holds for both cancer states i=1,2. The summa-

[9] The inequality **(E.3)** leads to this result since the following inequalities hold both for i=1 and i=2:

- $\hat{\mu_7}(i|\pi,\pi')sens_7(\pi')Rew(7,\phi|i)+\hat{\mu_9}(i|\pi,\pi')sens_9(\pi')Rew(9,\phi|i) >$ $\mu_6(0)$[Pr(s_7=i|s_6=0)+Pr(s_9=i|s_6=0)] min{$sens_7(\pi')$,$sens_9(\pi')$} Rew(9,ϕ|i) ,
- [1-$\mu_6(0)$] > μ_6(i) ,
- X(i)=$sens_7(\pi)$Rew(6,1|i)+[1-$sens_7(\pi)$]$sens_9(\pi)$Rew(6,3|i)+[1-$sens_7(\pi)$][1-$sens_9(\pi)$]Rew(6,ϕ|i).

tion of both sides of the inequality over i=1,2 shows that the total expected QALYs surplus obtained under strategy π' due to the screenings at time points t=3, and 5 is higher than the QALYs surplus obtained under strategy π due to the screening conducted at time point t=1.

As shown, the inequality *(E.3)* implies that the screenings of strategy π' at time points t=7, and 9 yields more additional QALYs than the screening conducted at time point t=6 by strategy π. Similarly, the inequality *(E.4)* implies that the screenings of strategy π' at time points t=3, and 5 yields more additional QALYs than the QALYs surplus of screening strategy π, due to its more sensitive screening at time point t=1. Combined, these results show that when the inequalities *(E.3)* and *(E.4)* both hold for cancer states i=1, 2, biennial screening strategy π' is expected to generate more QALYs surplus with its screenings at t=1, 3, 5, 7 and 9 than double screening strategy π under a fixed future strategy ϕ. This concludes the proof.

A.6 Intermediate Rewards and Disutility Function

We use $r_t(s, a)$ to denote the intermediate expected QALYs accrued between time t and $t + 1$ when a woman's current health state is $s \in S_U$ and the screening action is $a \in A$. It consists of two main components: reward function, denoted by $\hat{r}_t(s, a)$, and the disutility function, denoted by $u_t(s, a)$.

We employ the "half-cycle correction method" to calculate the first component of intermediate rewards. Accordingly, we assign the full decision interval length to the intermediate reward of each woman if she remains alive and the half decision interval length in the case of death. The underlying assumption is that if death occurs, it happens, on average, in the middle of the decision interval. Then, the general form of the reward function corresponding to health state $s \in S_U$ and action $a \in A$ at time $t \in T_A$ is the following: $\hat{r}_t(s, a) = CP_t^a(Alive|s) + \frac{C}{2}P_t^a(Dead|s)$

C is a constant that denotes the length of time interval between consecutive decision epochs, $P_t^a(Dead|s)$ denotes the probability of death and $P_t^a(Alive|s)$ denotes the prob-

ability of survival given the health state is $s \in S_U$ and action is equal to $a \in A$ at time $t \in T_A$. In our study, C =1 since decisions are made every year, $P_t^a(Dead|s)$ = $P_t^1(5|s)$, i.e. probability of death under natural history of disease progression, and $P_t^a(Alive|s)$ = 1 -$P_t^a(Dead|s)$ $\forall$ s $\in$ S_U, a $\in$ A and t $\in$ T_A.

We account for quality of life reductions due to the harms associated with diagnostic actions (i.e. screening and biopsy) by subtracting a certain amount from $\hat{r}_t(s,a)$. This reduction is captured by disutility function $u_t(s,a)$. In our study, we activate this disutility function starting from a certain age, where the harm associated with diagnostic actions is not negligible and hence, cannot be ignored. We employ an indicator function $1_{\{t \geq DIT\}}$, which is equal to 1 if time t is no less than the disutility initiation time DIT and 0 otherwise. Accordingly, the general form of the intermediate reward corresponding to health state $s \in S_U$ and action $a \in A$ at time $t \in T_A$ is the following: $r_t(s,a) = \hat{r}_t(s,a) - 1_{\{t \geq DIT\}} u_t(s,a)$

In our base case analysis, we choose "DIT" in a way that it corresponds to age 75 and conduct sensitivity analysis to measure its impact on optimal strategies (e.g. changing initiation age to age 65 or 55). This concludes the general description of intermediate rewards.We now proceed with the detailed description of disutility function. Similar to intermediate rewards, disutility function consists of two main components: Disutility associated with screening, denoted by $u_t^{scr}(s,a)$, and disutility associated with (follow-up) biopsy, denoted by u_t^{bio}(s). The general form of the disutility function, corresponding to harms associated with action $a \in A$ for health state $s \in S_U$ at time $t \in T_A$, is the following: $u_t(s,a) = u_t^{scr}(s,a) + q_t^{bio}(s,a) u_t^{bio}(s)$

where the probability of performing a biopsy $q_t^{bio}(s,a)$ = 1 - $spec_t(a)$ when s=0 (healthy) and $q_t^{bio}(s,a)$ = $sens_t(a)$ when s=1, 2 (undetected -asymptomatic- cancer states) for all a $\in$ A_{SCR}. There is no disutility when the chosen action is "No Screening" (i.e. u_t(s,a) = 0 when a=1).

The literature reports that false positive screening has higher disutility than a true positive one [99] and disutility (both associated with screening and biopsy) increases as women

get older ([86] and the references therein). Accordingly, we penalize false positive screening results higher than true positive ones and use a linear function both for screening and biopsy disutility functions (accounting for increasing disutility with age) in our base case (numerical) study. There are three parameters, allowing us to adjust for these features of the disutility function: (i) false-positive penalty constant α_{FPP}, accounting the ratio of disutility of a false positive screening to the one of a true positive screening, (ii) screening linearity factor α_{SLF}, setting the ratio of screening disutility at age 100 when compared to the one at age 40 and (iii) biopsy linearity factor α_{BLF}, setting the ratio of biopsy disutility at age 100 when compared to the one at age 40.

We continue discussing the disutility function, by analyzing the mathematical expression of its components in detail. We start with providing the relation between time t, the current age of patient a_t and screening initial age a_1: $t = a_t - a_1 + 1 \quad \forall \quad t \in \{1, 2, ...T - a_1 + 1\}$

We use this relationship below when we define the components of disutility function. Terminal age T is set to age 100 in our study.

A.6.1 Disutility Associated with Screening $u_t^{scr}(s, a)$

The disutility of negative screening is set to 0.5 days at age 40 (and at younger ages) for any single modality [96] and 1 day when two modalities (i.e. MRI adjunct to mammography) is utilized together. We designate $n(a)$ to denote the number of modalities employed by screening action a ∈ A, which is equal to 0 when a=1, 1 when a ∈ {2,3,4} and 2 when a=5. Then general form of the disutility associated with screening is the following:

$u_t^{scr}(s, a) = \frac{0.5}{365} * n(a)$ when $a_t \in \{25, 26, ..., 40\}, a \in A \; and \; s \in S_U$

$u_t^{scr}(s, a) = \frac{0.5}{365} + \beta_S(a_t - 40)] * n(a)$ when $a_t \in \{41, .., 99\}, a \in A \; and \; s \in S_U$

$u_t^{scr}(s, a) = \alpha_{SLF} * \frac{0.5}{365} * n(a)$ when $a_t = 100, a \in A \; and \; s \in S_U$

Screening disutility linearity slope β_S = $\frac{0.5}{365} \frac{\alpha_{SLF}-1}{100-40}$. Screening linearity factor α_{SLF}=1 and =2 when a constant or an increasing disutility function is employed, respectively. Ac-

cordingly, β_S disappears when harm associated with screening is assumed to remain constant with age (i.e. when α_{SLF} is set to 1).

A.6.2 Disutility Associated with Biopsy u_t^{bio}(s)

The disutility of true positive screening at age 40 (and at younger ages) is set equal to two weeks [97, 98].

$u_t^{bio}(s) = \frac{2}{52}$ when $a_t \in \{25, 26, ..., 40\}\ and\ s \in \{1, 2\}$

$u_t^{bio}(s) = \frac{2}{52} + \beta_B(a_t - 40)$ when $a_t \in \{41, 42, ..., 99\}\ and\ s \in \{1, 2\}$

$u_t^{bio}(s) = \alpha_{BLF} * \frac{2}{52}$ when $a_t = 100\ and\ s \in \{1, 2\}$

$u_t^{bio}(s) = \alpha_{FPP} * u_t^{bio}(s + 1)$ when $a_t \in \{25, 26, ..., 99\}\ and\ s = 0$

Biopsy disutility linearity slope $\beta_B = \frac{2}{52}\frac{\alpha_{BLF}-1}{100-40}$. Biopsy linearity factor α_{BLF}=1 and =2 when a constant or an increasing disutility function is employed, respectively. Accordingly, β_B disappears when harm associated with biopsy is assumed to remain constant with age (i.e. when α_{BLF} is set to 1). False-positive penalty constant α_{FPP} is set to 2.

APPENDIX B

APPENDIX FOR CHAPTER 2

B.1 Overview of Existing Asymptotic Staffing Rules

In the literature, three main modes of operation, based on which analytic-formula based staffing rules with asymptotic convergence developed, has been identified for queuing regimes: Quality-Driven (QD), Efficiency-Driven (ED), and Quality-and-Efficiency Driven (QED) [139]. QD regimes target timely service (i.e., service quality) with the goal of minimal delays and abandonments, ED regimes emphasize server efficiency with the goal of maximum server utilization and QED regimes balance these two perspectives and target high (service) quality and (server) efficiency simultaneously rather than letting one perspective (goal) dominating the other. In particular, the main staffing results for these three asymptotic queuing regimes are developed by studying M/M/s (+M or +G) queue, characterized by Poisson arrivals (M/./.) with rate λ, exponential service (./M/.) with rate μ, s identical servers (././s), and if abandonment is included, exponentially (+M) or generally (+G) distributed patience time with pdf g(.). Summary of the our and other main staffing rules with their corresponding staffing level formulas and targeted performance measures, asymptotically achieved as λ and s increase indefinitely and μ held fixed, are provided in Table B.1. See [146], [137], [139], and [140] for the details and [165] for an overview.

ER treatment queues, where servers are ER physicians and the service is the medical treatment provided, are characterized as *efficiency-driven (ED)*(busy-server) regimes. In an ED regime, servers are generally busy and some delay in service is inevitable, and under heavy traffic limit conditions, expected waiting time $E[W] > 0$, probability of delay $P(W > 0) \approx 1$ and tail probability of delay $P(W > T) > 0$ for any T>0. The key difference between ER treatment queues, operating as *no-abandonment-ED*, and the other

Table B.1: Asymptotic Staffing Rules with Associated Staffing Levels and Performance Goals

Asymptotic Staffing Rule	Staffing Level Formula	Key Performance Measure(s)	Other Measures
Efficiency-Driven (ED)	s = (1-γ)*$\frac{\lambda}{\mu}$, $\gamma > 0$	P(Ab) $\to \gamma$	P(W > 0) $\to 1$, $\rho_e \to 1$
Quality-Driven (QD)	s = (1+γ)*$\frac{\lambda}{\mu}$, $\gamma > 0$	P(Ab) $\to 0$, E[W] $\to 0$	P(W > 0) $\to 0$, $\frac{P(Ab)}{E[W]}$=g(0)
Quality-and-Efficiency Driven (QED)	s = $\frac{\lambda}{\mu} + \beta_\alpha \sqrt{\frac{\lambda}{\mu}}$	P(W > 0) $\to \alpha$	P(Ab) $\to 0$, E[W] $\to 0$, $\rho_e \to 1$
QED tuning of ED (ED+QED)	s = (1-γ_T)*$\frac{\lambda}{\mu} + \delta_{T,\alpha} \sqrt{\frac{\lambda}{\mu}}$	P(W $> T$) $\to \alpha$	P(Ab) $\to \gamma_T$
ED-No Abandonment (ED-NoAb)	s = $\frac{\lambda}{\mu} + \frac{\ln(1/\alpha)}{\mu T}$	P(W $> T$) $\to \alpha$, E[W] $\to \frac{T}{\ln(1/\alpha)}$	P(W > 0) $\to 1$, $\rho_e \to 1$

Note: "Ab" and "W" are abbreviations for abandonment and waiting time, respectively.
Effective server utilization (load per server) ρ_e:= ρ*(1-P(Ab)), where $\rho = \frac{\lambda}{\mu s}$.

efficiency-driven regimes is the (non)existence of abandonment. As noted in the main text, abandonments from the ER happen at the earlier stages (queue stations) of the ER care delivery process (multi-stage network) and ER treatment queues experience no abandonment. When the ED (or ED+QED) formula is used without abandonments by setting the abandonment target γ (or γ_T) = 0, the resulting staffing level s is equal to the offered load $\frac{\lambda}{\mu}$, and cannot maintain stability (as ρ = 1). Hence, we develop a new staffing rule, *ED-NoAb*, to stabilize the tail probability of delay P(W $> T$) around α for ED regimes with negligible, if any, abandonment events (i.e., P(Ab) ≈ 0).

B.2 Proof of Theorem 1

We first present some steady-state results for M/M/s (Erlang-C) queue that we use to prove *Theorem 1*.

Relevant Steady-State Results for M/M/s (Erlang-C) *Consider a M/M/s (Erlang-C) queue model with (Poisson) arrival rate λ, (exponentially distributed) service rate μ, and s identical servers. The steady-state results for this queue, when the expected traffic intensity (server utilization) $\rho := \frac{\lambda}{\mu s} < 1$ holds, are as follows [265]:*

- Probability of delay P(W > 0) is equal to

$$P(W > 0) = \frac{1}{1 + A/B}, \quad \textit{where} \quad A = \sum_{k=0}^{s-1} \frac{(s\rho)^k}{k!} \quad \textit{and} \quad B = \frac{(s\rho)^s}{s![1-\rho]}, \quad \text{(B.1)}$$

- Conditional tail probability of delay P(W > T|W > 0) is equal to

$$P(W > T|W > 0) = e^{-s[1-\rho]\mu T} \quad for \quad T > 0, \tag{B.2}$$

- Conditional expected waiting time, given service is delayed, is equal to

$$E[W|W > 0] = \frac{1}{s(1-\rho)\mu}. \tag{B.3}$$

Now, we provide three results that will help us prove *Theorem 1*. From now on, assume $\rho_n \in (0,1)\ \forall\ n$.

Lemma 1: *Let* $\gamma_n = \sum_{k=0}^{n-1} \frac{(n\rho_n)^k}{k!} e^{-n\rho_n}$. *Then* $\gamma_n \to 1/2$ *as* $n \to \infty$ *and* $n(1-\rho_n) \to \beta > 0$.

Proof (Lemma 1): Let S_n be the sum of n independent and identically distributed Poisson random variables, each with rate ρ_n. Then, S_n is also a Poisson random variable [266] with rate $n\rho_n$, mean $E[S_n] = n\rho_n$ and variance $V[S_n] = n\rho_n$ [267].

It is easy to recognize that $\gamma_n = P(S_n \leq n-1) = P(\frac{S_n - E[S_n]}{\sqrt{V[S_n]}} \leq \frac{(n-1)-n\rho_n}{\sqrt{n\rho_n}})$. Then, by *Central Limit theorem* [268], $\gamma_n \to P(N(0,1) \leq K)$, where N(0,1) is the standard normal distribution and $K = \lim_{n\to\infty} \frac{(n-1)-n\rho_n}{\sqrt{n\rho_n}}$ as $n(1-\rho_n) \to \beta$.

Since $\frac{(n-1)-n\rho_n}{\sqrt{n\rho_n}} = \frac{n(1-\rho_n)-1}{n^{1/2}\sqrt{\rho_n}}$, it is easy to see that K=0 as $n(1-\rho_n) \to \beta$ and $n \to \infty$. Accordingly, the result $\gamma_n \to 1/2$ immediately follows as $\gamma_n \to P(N(0,1) \leq 0) = 1/2$.

Lemma 2: *Let* $\xi_n = \frac{(n\rho_n)^n}{n![1-\rho_n]} e^{-n\rho_n}$. *Then* $\xi_n \to \infty$ *as* $n \to \infty$ *and* $n(1-\rho_n) \to \beta > 0$.

Proof (Lemma 2): Let's apply Stirling's formula [269] to ξ_n in order to approximate the factorial $n!$ term by $(2\pi n)^{1/2} n^n e^{-n}$. Then, we obtain $\xi_n \sim \frac{(n\rho_n)^n}{(2n\pi)^{1/2} n^n e^{-n}[1-\rho_n]} e^{-n\rho_n} = \frac{n^{1/2} e^{n(1-\rho_n)}}{\sqrt{2\pi} n[1-\rho_n]} (\rho_n)^n$. Since $\rho_n = 1-(1-\rho_n) = 1 - \frac{n(1-\rho_n)}{n}$, we get $\xi_n \sim \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{e^{n(1-\rho_n)}}{n[1-\rho_n]} [1 - \frac{n(1-\rho_n)}{n}]^n$. As $n \to \infty$ and $n(1-\rho_n) \to \beta > 0$, the components $\frac{\sqrt{n}}{\sqrt{2\pi}} \to \infty$, $\frac{e^{n(1-\rho_n)}}{n[1-\rho_n]} \to \frac{e^\beta}{\beta}$, and $[1 - \frac{n(1-\rho_n)}{n}]^n \to 1$. Accordingly, $\xi_n \to \infty$.

Lemma 3: *Let* γ_n *and* ξ_n *be as defined in Lemma 1 and Lemma 2. Then, the terms* $\gamma_n/\xi_n \to 0$ *and* $\frac{1}{1+\gamma_n/\xi_n} \to 1$ *as* $n \to \infty$ *and* $n(1-\rho_n) \to \beta > 0$.

Proof (Lemma 3): *Lemma 3* directly follows from *Lemma 1* and *Lemma 2*: Since γ_n $\rightarrow$ 1/2, by *Lemma 1*, and $\xi_n \rightarrow \infty$, by *Lemma 2*, as n $\rightarrow \infty$ and n(1-ρ_n) $\rightarrow \beta > 0$, the terms $\gamma_n/\xi_n \rightarrow 0$ and $\frac{1}{1+\gamma_n/\xi_n} \rightarrow 1$.

Theorem 1: *Consider a sequence of $M/M/s$ queues, indexed by n, each of which has arrival rate λ_n, service rate $\mu_n = \mu \in (0, \infty)$, and s_n servers. Let the traffic intensity $\rho_n = \frac{\lambda_n}{\mu s_n} < 1$. Consider probability target $0 < \alpha < 1$ and delay threshold $T > 0$. If*

$$\lim_{n\rightarrow\infty} s_n(1-\rho_n) = \beta \quad \textit{as} \quad \lambda_n, s_n \rightarrow \infty, \qquad \textit{where } \beta = \frac{\ln(1/\alpha)}{\mu T} > 0, \tag{B.4}$$

(i) The probability of delay and the server utilization converge to 1:

$$P(W_n > 0) \rightarrow 1 \quad \text{and} \quad \rho_n \rightarrow 1 \quad \text{as} \quad \lambda_n, s_n \rightarrow \infty.$$

(ii) The tail probability of delay and the expected waiting time converge to non-degenerate limits:

$$P(W_n > T) \rightarrow \alpha \quad \text{and} \quad E[W_n] \rightarrow w^* \equiv \frac{T}{\ln(1/\alpha)} \quad \text{as} \quad \lambda_n, s_n \rightarrow \infty.$$

Proof (Theorem 1): The proofs of *Theorem 1 (i)* and *(ii)* are as follows:

(i) Let the condition B.4 holds. Further, let $A_n = \sum_{k=0}^{s_n-1} \frac{(s_n\rho_n)^k}{k!}$, $B_n = \frac{(s_n\rho_n)^{s_n}}{s_n![1-\rho_n]}$, $\gamma_{s_n} = A_n * e^{-s_n\rho_n}$ and $\xi_{sn} = B_n * e^{-s_n\rho_n}$. Then by B.1, $P(W_n > 0) = \frac{1}{1+A_n/B_n} = \frac{1}{1+\gamma_{s_n}/\xi_{s_n}}$, when $\rho_n < 1$. By *Lemma 3*, $P(W_n > 0) = \frac{1}{1+\gamma_{s_n}/\xi_{s_n}} \rightarrow 1$ as λ_n,$s_n \rightarrow \infty$ in a way that $s_n(1-\rho_n) \rightarrow \beta$. Further, since $\frac{\beta}{s_n} \rightarrow 0$ as $s_n \rightarrow \infty$ for $\beta > 0$, the server utilization $\rho_n \rightarrow (1 - \frac{\beta}{s_n}) \rightarrow 1$ when the condition B.4 holds.

(ii) Under B.4, $P(W_n > T | W_n > 0) \rightarrow e^{-\beta\mu T} = \alpha$ and $E[W_n | W_n > 0] \rightarrow \frac{1}{\beta\mu} = \frac{T}{\ln(1/\alpha)}$ by B.2, B.3, and the choice of $\beta = \frac{ln(1/\alpha)}{\mu T}$. As $P(W_n > 0) \rightarrow 1$ when $\lambda_n, s_n \rightarrow \infty$, as shown in *Proof (Theorem 1) (i)*, the results $P(W_n > T) \rightarrow \alpha$ and $E[W_n] \rightarrow \frac{T}{\ln(1/\alpha)}$ immediately follow from the definitions of conditional probability and conditional expectation.

B.3 Numerical Results: TPoD Graphs

As described in the manuscript (*Section 2.5 Numerical Experiments*), we conduct simulation experiments to assess the performance of *ED-NoAb* staffing under time-varying demand. Each simulation experiment corresponds to a scenario with fixed parameters. For each experiment, we run N independent replications and at each replication, we send virtual patients to the queue(s) of interest at fixed times $t_1, t_2, ...$ to collect statistics. Accordingly, in total, there are N virtual patients that are (generated and) dispatched to the queues of interest at the fixed time point t_i, where $i \in \{1,2,...,\}$. Then, for each t_i, the fraction of patients waiting more than delay threshold T is given by $\frac{N(t_i)}{N}$, where $N(t_i)$ is the total number of virtual patients with arrival time t_i and waiting time over T. The result of each experiment is summarized by (what we call) a "TPoD graph", depicting the fractions of patients waiting more than threshold T (y-axis) over time (x-axis),and demonstrating whether the tolerance α is violated. If the TPoD targets are attained with *ED-NoAb*, the same experiment is run with the *reduced* staffing levels $0.90 * $ *ED-NoAb*.

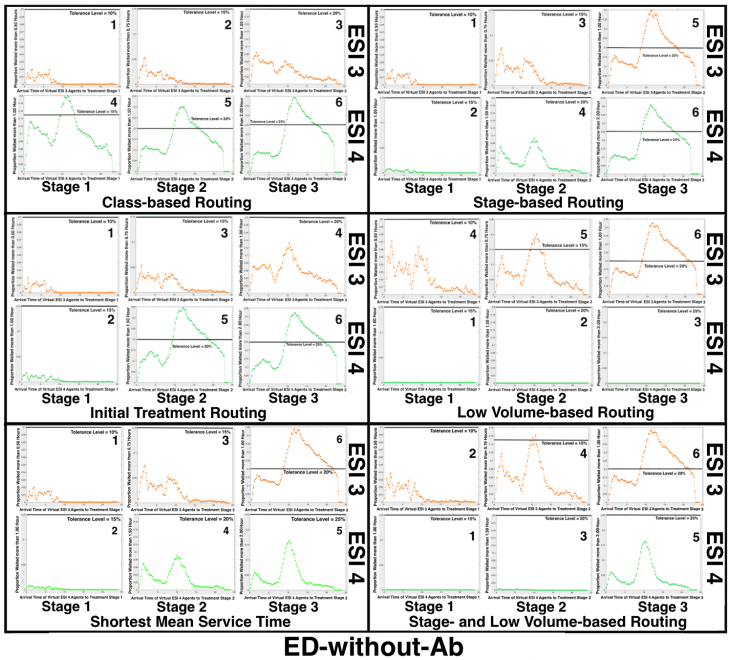


Figure B.1: TPoD Graphs for Two-Class Three-Stage Case with ESI 3 - ESI 4 Arrivals and Static Routing Rules

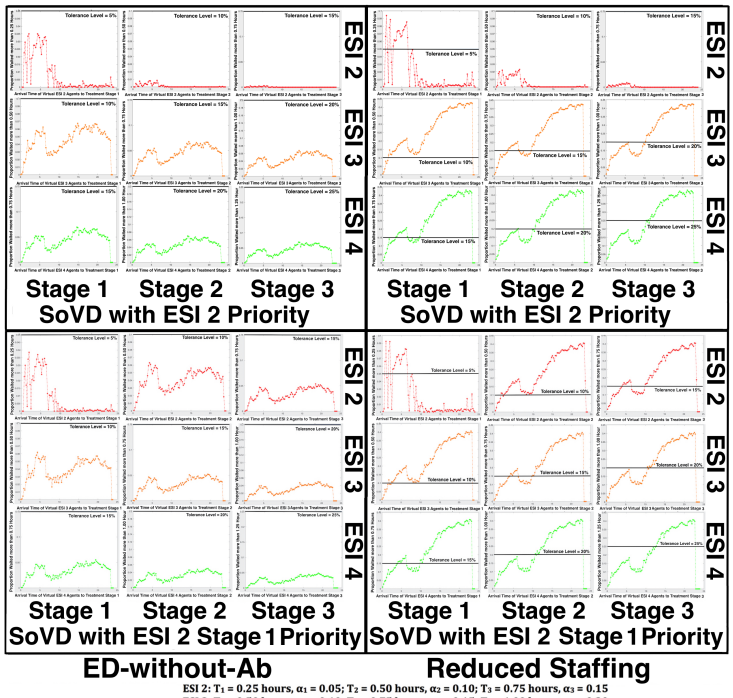
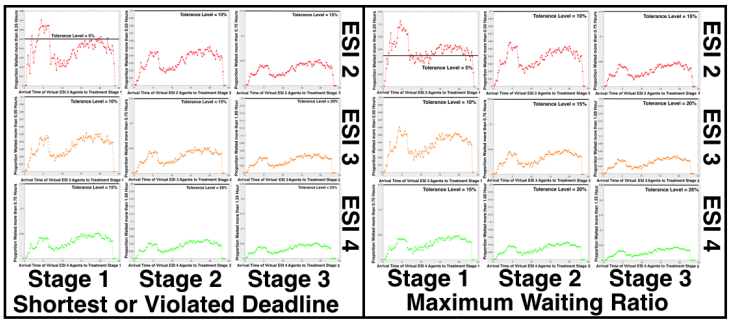


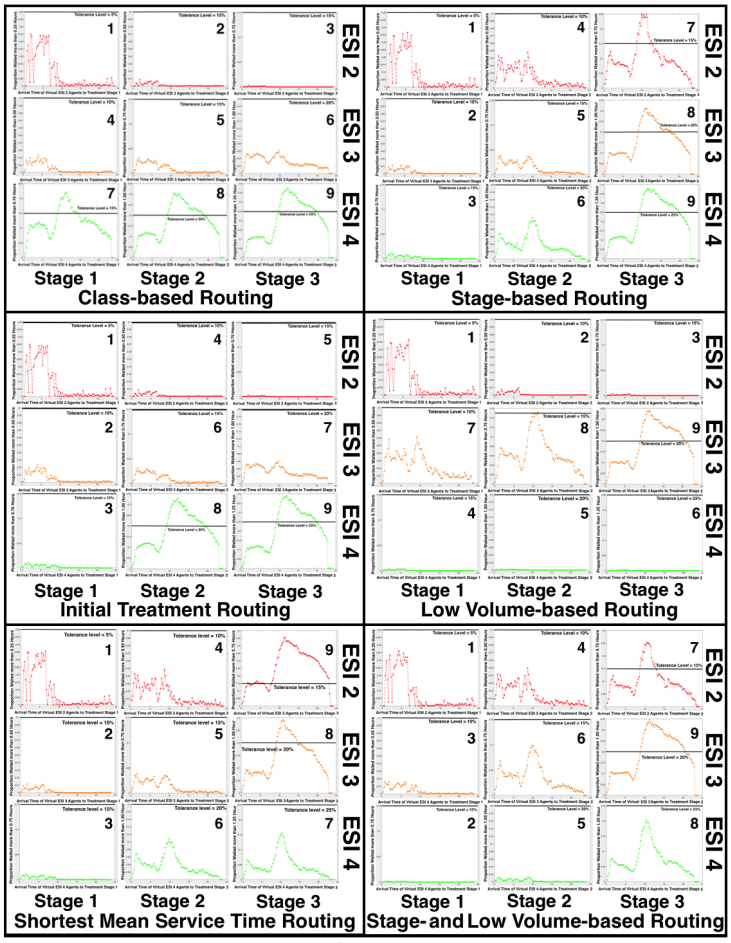
Figure B.2: TPoD Graphs for Three-Class Three-Stage Case with Hybrid Routing Rules - Reduced Staffing = 0.90(ED-NoAb)



ED-without-Ab

ESI 2: $T_1 = 0.25$ hours, $\alpha_1 = 0.05$; $T_2 = 0.50$ hours, $\alpha_2 = 0.10$; $T_3 = 0.75$ hours, $\alpha_3 = 0.15$
ESI 3: $T_1 = 0.50$ hours, $\alpha_1 = 0.10$; $T_2 = 0.75$ hours, $\alpha_2 = 0.15$; $T_3 = 1.00$ hours, $\alpha_3 = 0.20$
ESI 4: $T_1 = 0.75$ hours, $\alpha_1 = 0.15$; $T_2 = 1.00$ hours, $\alpha_2 = 0.20$; $T_3 = 1.25$ hours, $\alpha_3 = 0.25$

Figure B.3: TPoD Graphs for Three-Class Three-Stage Case with Dynamic Routing Rules



ED-without-Ab

ESI 2: $T_1 = 0.25$ hours, $\alpha_1 = 0.05$; $T_2 = 0.50$ hours, $\alpha_2 = 0.10$; $T_3 = 0.75$ hours, $\alpha_3 = 0.15$
ESI 3: $T_1 = 0.50$ hours, $\alpha_1 = 0.10$; $T_2 = 0.75$ hours, $\alpha_2 = 0.15$; $T_3 = 1.00$ hours, $\alpha_3 = 0.20$
ESI 4: $T_1 = 0.75$ hours, $\alpha_1 = 0.15$; $T_2 = 1.00$ hours, $\alpha_2 = 0.20$; $T_3 = 1.25$ hours, $\alpha_3 = 0.25$

Figure B.4: TPoD Graphs for Three-Class Three-Stage Case with Static Routing Rules

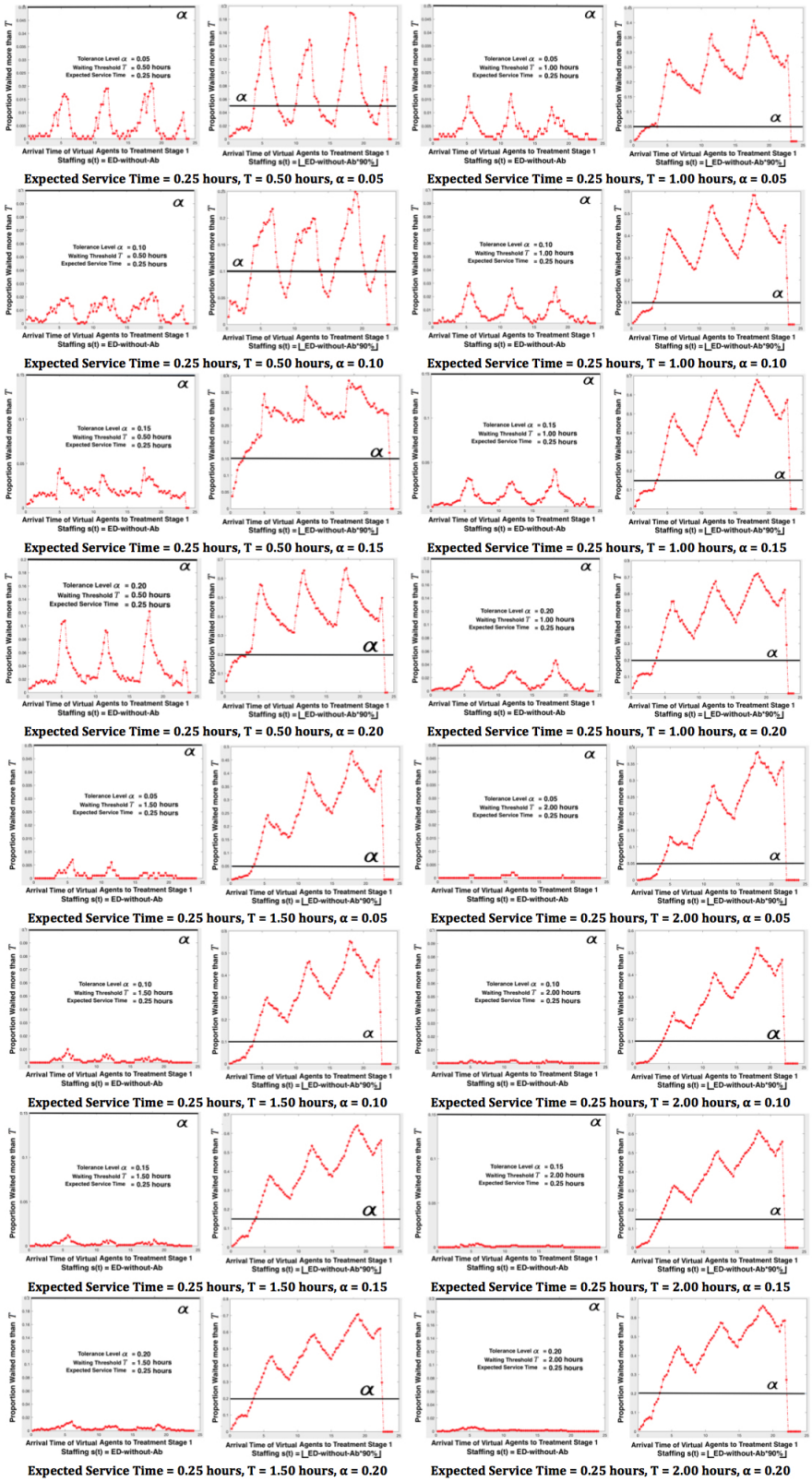


Figure B.5: $M_t/G/s_t$ Queue Experiments with Mean Service Time = 0.25 Hour

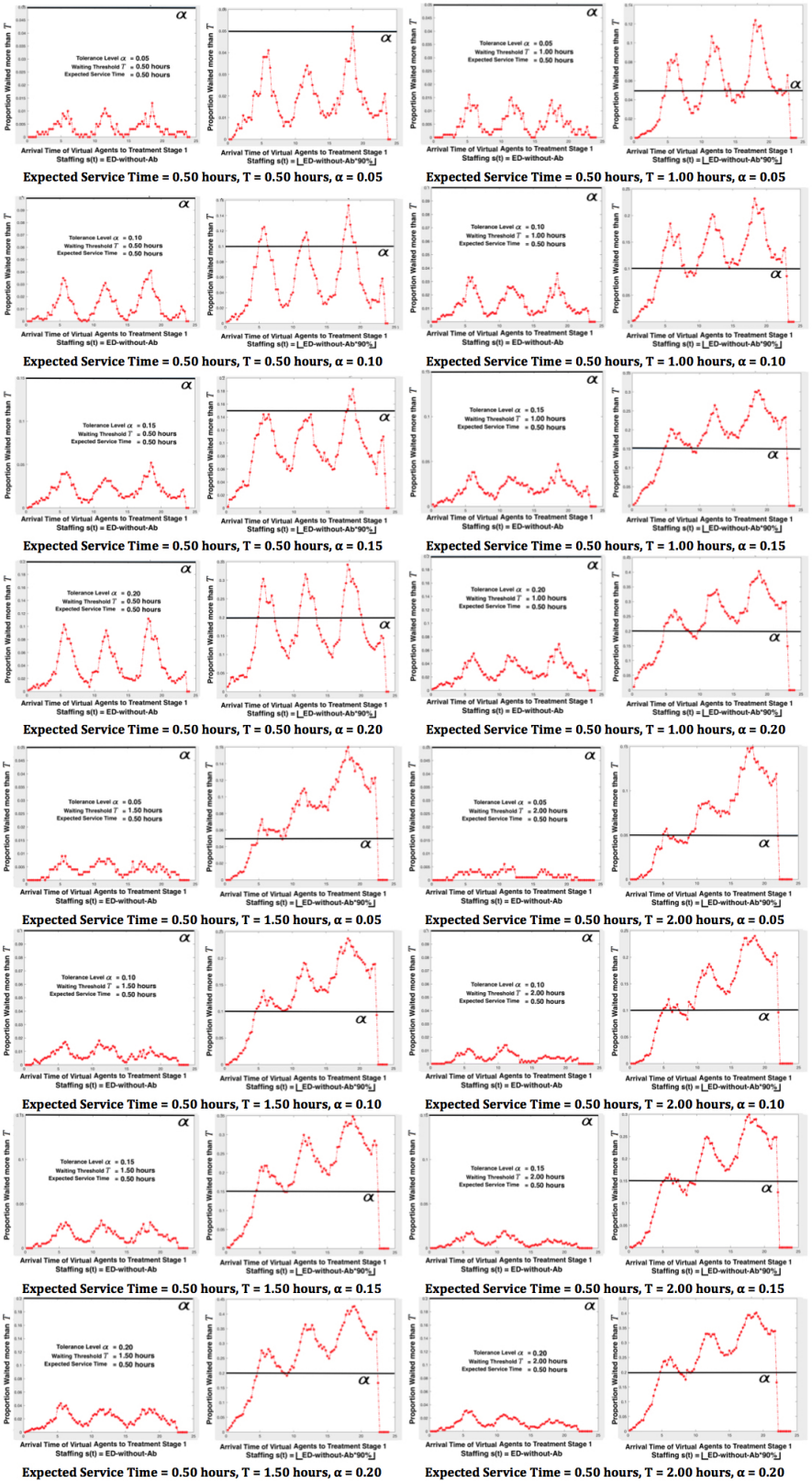


Figure B.6: $M_t/G/s_t$ Queue Experiments with Mean Service Time = 0.50 Hour

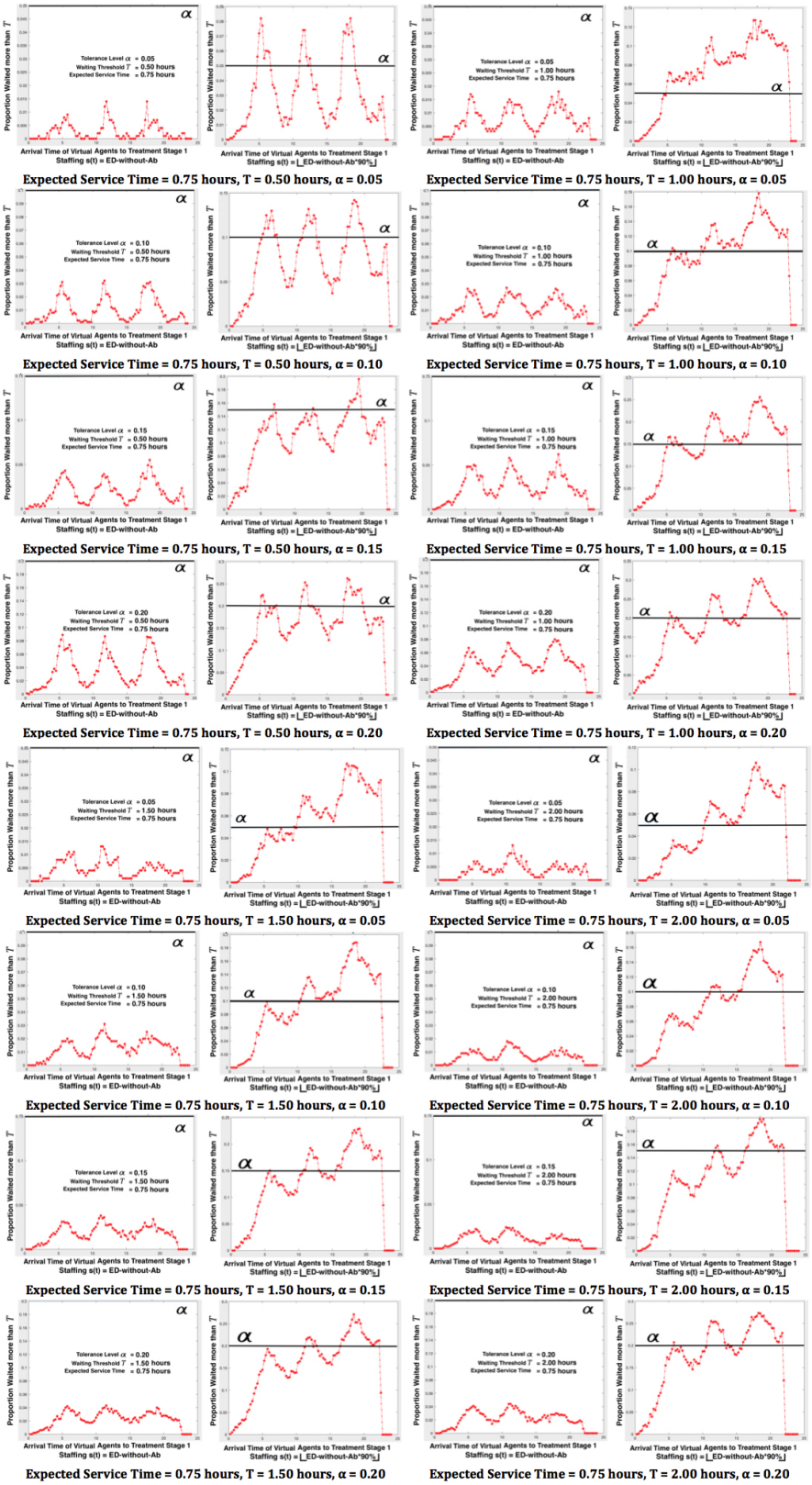


Figure B.7: $M_t/G/s_t$ Queue Experiments with Mean Service Time = 0.75 Hour

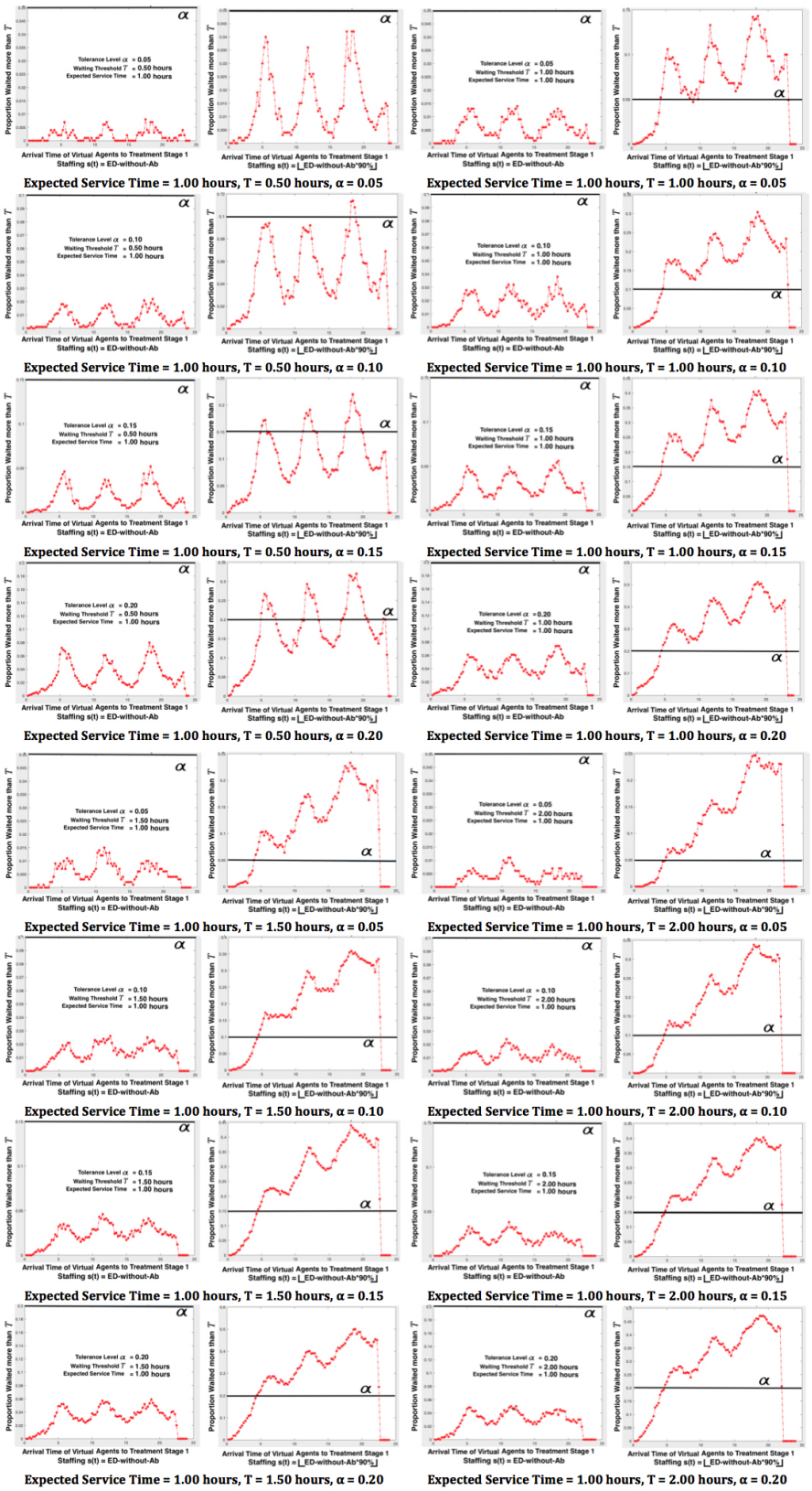


Figure B.8: $M_t/G/s_t$ Queue Experiments with Mean Service Time = 1.00 Hour

APPENDIX C

APPENDIX FOR CHAPTER 3

C.1 Assessing the Effectiveness of Treatment Sequences for Follicular Lymphoma Patients with a Multi-state Model

Table C.1: Patient Characteristics for First-Line Treatments

Patient Characteristics Clinical and Demographic Factors		Rituximab (R) (n = 1219)	R-CVP (n = 620)	R-CHOP (n = 894)	R-Other (n = 432)	All First-Line (n = 5234)
Age at TX1 Initiation (Years)	Mean	75.9	76.2	75.9	76.6	76.0
	Standard Dev.	6.5	6.8	6.6	6.7	6.7
Age Group at TX1 Initiation (Years)	66 - 70	24.2%	26.0%	24.8%	23.6%	25.3%
	71 - 75	27.0%	22.1%	28.0%	21.8%	25.4%
	76 - 80	22.3%	27.9%	21.6%	25.2%	22.9%
	≥81	26.5%	24.0%	25.6%	29.4%	26.5%
Race	Caucasian	93.8%	93.9%	93.1%	92.8%	93.9%
	African American	3.8%	3.1%	2.9%	4.6%	3.5%
	Other	2.5%	3.1%	4.0%	2.6%	2.6%
Sex	Male	43.2%	42.3%	42.2%	44.4%	42.8%
	Female	56.9%	57.7%	57.8%	55.6%	57.2%
Marital Status	Married	57.7%	57.3%	59.2%	55.8%	57.2%
	Other	35.9%	38.1%	34.3%	40.5%	37.1%
	Unknown	6.4%	4.7%	6.5%	3.7%	5.7%
% High School Education Only	< 25%	29.6%	29.4%	30.5%	27.3%	30.0%
	≥ 25%	66.2%	66.1%	66.0%	69.9%	66.1%
	Unknown	4.2%	4.5%	3.5%	2.8%	3.9%
FL Stage	I/II	48.7%	47.6%	44.0%	47.7%	47.8%
	III/IV	44.9%	47.3%	49.2%	46.8%	46.1%
	Unknown	6.5%	5.2%	6.8%	5.6%	6.1%
Primary Site of Involvement	Nodal	83.6%	84.5%	83.1%	81.9%	83.6%
	Extranodal	16.4%	15.5%	16.9%	18.1%	16.5%
Charlson Comorbidity Index	Charlson Index = 0	64.6%	62.9%	63.5%	62.0%	63.5%
	Charlson Index = 1	22.9%	23.2%	22.9%	26.9%	23.2%
	Charlson Index ≥ 2	12.6%	13.9%	13.5%	11.1%	13.3%
Performance Status	Fully Active	84.7%	84.0%	83.0%	81.7%	83.3%
	Restricted	15.3%	16.0%	17.0%	18.3%	16.7%
History of Anemia	Absent	88.2%	85.3%	85.8%	88.9%	87.1%
	Present	11.8%	14.7%	14.2%	11.1%	12.9%
B Symptoms	Present	12.9%	7.7%	11.2%	12.7%	14.3%
	Absent	3.5%	2.3%	3.5%	3.7%	4.5%
	Unrecorded	83.7%	90.0%	85.4%	83.6%	81.2%
FL Histology	Grade 1 or 2	52.9%	52.9%	53.0%	47.9%	53.1%
	Grade 3	18.0%	16.9%	15.0%	19.0%	16.4%
	Not Specified	29.1%	30.2%	32.0%	33.1%	30.5%
Residency Demographics	Metro Area	78.0%	79.8%	74.8%	77.6%	77.5%
	Urban or Rural	22.0%	20.0%	25.2%	22.5%	22.5%
Region Demographics	West	21.7%	20.8%	21.4%	22.0%	21.5%
	Northeast	25.8%	27.7%	22.4%	26.4%	25.1%
	Midwest	18.7%	23.4%	23.0%	22.9%	22.1%
	South	33.8%	28.1%	33.2%	28.7%	31.4%

Table C.2: Multivariable Analyses via Cox PH Regression Models

Model		Conventional OS Model	Multi-state Model with the Clinical States Alive after TX1, TX2, TX3 and Death				
Factor / Transition		Treatment → Death (OS)	Tx1 → Tx2	Tx1 → Death	Tx2 → Tx3	Tx2 → Death	Tx3 → Death
Histology	Grade 1/2 - Reference	--	--	--	--	--	--
	Grade 3		0.91 (0.83-0.99)				
B Symptoms present		1.38 (1.16-1.65)	1.42 (1.23-1.65)				
Performance Status = No limitation					0.87 (0.79-0.97)		
Region	West - Reference	--	--	--	--	--	--
	Northeast			0.83 (0.70-0.99)			
Marital Status	Married– Reference	--	--	--	--	--	--
	Other Marital Status	0.91 (0.83-0.98)	0.92 (0.87-0.99)				0.86 (0.76-0.97)
Residency	Rural/Urban Reference	--	--	--	--	--	--
	Metropolitan Area						1.16 (1.01-1.34)
Age (years)	Age 65-70 - reference	--	--	--	--	--	--
	Age > 80				1.09 (1.00-1.19)		
	Age > 85			1.27 (1.02-1.60)			
First line of therapy							
	R - Reference	--	--	--	--	--	--
	R-CVP	0.78 (0.68-0.90)	1.16 (1.06-1.29)		1.23 (1.09-1.39)	0.50 (0.36-0.71)	
	R-CHOP	0.57 (0.50-0.64)	1.26 (1.16-1.37)	0.60 (0.47-0.77)	1.21 (0.82-1.35)	0.40 (0.29-0.53)	0.63 (0.53-0.76)
	R-Other		1.58 (1.41-1.77)	0.57 (0.39-0.84)			
Second line of therapy							
	R - Reference	--	--	--	--	--	--
	R-CVP				0.72 (0.62-0.83)		
	R-CHOP	0.82 (0.70-0.96)			0.75 (0.66-0.85)	0.61 (0.44-0.84)	0.80 (0.66-0.96)
	R-Other				0.87 (0.77-0.97)		
Third line of therapy							
	R - Reference	--	--	--	--	--	--
	R-CHOP						0.81 (0.66-1.00)

Table C.3: State Occupation Probabilities over Time for First Line Treatments

First-Line	Rituximab (R)				R-CVP				R-CHOP				Other R-containing			
Year	TX1	TX2	TX3	Death	TX1	TX2	TX3	Death	TX1	TX2	TX3	Death	TX1	TX2	TX3	Death
Year 0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
Year 0.25	0.75	0.16	0.06	**0.04**	0.69	0.16	0.12	**0.03**	0.61	0.16	0.20	**0.03**	0.53	0.20	0.23	**0.03**
Year 0.5	0.56	0.24	0.12	**0.07**	0.49	0.25	0.19	**0.07**	0.39	0.27	0.29	**0.05**	0.38	0.24	0.31	**0.08**
Year 0.75	0.35	0.33	0.21	**0.10**	0.36	0.28	0.26	**0.10**	0.30	0.27	0.34	**0.08**	0.27	0.27	0.33	**0.13**
Year 1	0.26	0.33	0.27	0.13	0.22	0.33	0.34	0.11	0.22	0.29	0.39	0.10	0.19	0.27	0.37	0.17
Year 1.25	0.21	0.27	0.35	**0.17**	0.19	0.31	0.37	**0.13**	0.20	0.27	0.42	**0.11**	0.15	0.27	0.39	**0.18**
Year 1.5	0.18	0.26	0.37	**0.20**	0.16	0.26	0.42	**0.15**	0.19	0.24	0.44	**0.13**	0.14	0.23	0.42	**0.21**
Year 1.75	0.16	0.23	0.39	**0.22**	0.15	0.23	0.45	**0.16**	0.17	0.25	0.44	**0.14**	0.12	0.24	0.42	**0.23**
Year 2	0.14	0.23	0.39	0.24	0.15	0.21	0.46	0.18	0.16	0.25	0.44	0.16	0.10	0.22	0.42	0.26
Year 2.25	0.13	0.21	0.40	**0.27**	0.14	0.20	0.45	**0.21**	0.16	0.23	0.45	**0.16**	0.10	0.20	0.42	**0.27**
Year 2.5	0.12	0.19	0.39	**0.29**	0.14	0.19	0.44	**0.22**	0.15	0.22	0.45	**0.17**	0.10	0.19	0.42	**0.29**
Year 2.75	0.12	0.17	0.39	**0.32**	0.13	0.19	0.43	**0.24**	0.15	0.22	0.44	**0.19**	0.10	0.19	0.41	**0.30**
Year 3	0.11	0.16	0.38	0.35	0.13	0.18	0.42	0.26	0.15	0.21	0.44	0.21	0.10	0.18	0.41	0.31
Year 3.25	0.11	0.15	0.37	**0.37**	0.13	0.17	0.41	**0.29**	0.14	0.20	0.44	**0.22**	0.10	0.17	0.41	**0.32**
Year 3.5	0.11	0.13	0.37	**0.39**	0.13	0.17	0.39	**0.31**	0.14	0.19	0.44	**0.22**	0.09	0.17	0.40	**0.33**
Year 3.75	0.11	0.12	0.35	**0.42**	0.13	0.15	0.37	**0.35**	0.14	0.18	0.44	**0.23**	0.09	0.17	0.38	**0.36**
Year 4	0.11	0.11	0.34	0.44	0.12	0.15	0.36	0.37	0.14	0.18	0.43	0.25	0.09	0.17	0.37	0.37
Year 4.25	0.10	0.11	0.33	**0.46**	0.12	0.15	0.35	**0.38**	0.14	0.18	0.43	**0.25**	0.09	0.15	0.36	**0.40**
Year 4.5	0.10	0.10	0.33	**0.48**	0.11	0.14	0.33	**0.41**	0.14	0.16	0.43	**0.27**	0.09	0.14	0.34	**0.42**
Year 4.75	0.10	0.09	0.31	**0.51**	0.11	0.14	0.31	**0.43**	0.14	0.16	0.42	**0.28**	0.09	0.14	0.33	**0.44**
Year 5	0.10	0.08	0.29	0.54	0.11	0.14	0.30	0.45	0.14	0.15	0.41	0.29	0.09	0.14	0.31	0.46
Year 6	0.09	0.06	0.24	0.61	0.07	0.11	0.22	0.60	0.13	0.14	0.35	0.37	0.08	0.11	0.29	0.52
Year 7	0.08	0.05	0.20	0.68	0.03	0.09	0.18	0.70	0.11	0.11	0.31	0.47	0.06	0.08	0.25	0.61
Year 8	0.05	0.04	0.14	0.77	0.03	0.04	0.13	0.79	0.09	0.06	0.26	0.59	0.05	0.05	0.17	0.74
Year 9	0.04	0.02	0.08	0.86	0.03	0.04	0.10	0.83	0.07	0.04	0.15	0.74	0.05	0.05	0.07	0.83
Year 10	0.04	0.02	0.03	0.91	0.03	0.04	0.00	0.92	0.00	0.04	0.06	0.90	0.05	0.05	0.00	0.91

Table C.4: The List of R-Other Therapies in the SEER-Medicare Dataset

Line of Treatment	Therapies that are Grouped into R-Other Category
First-Line	CFHOR, CFMOR, CFMR, CFR, CHMR, CHR, CMOR, CMORX, CMR, CR, CRX, FMR, FR, HMR, HOR, HR, MR, OR, RXZ, RX
Second-Line	BRXZ, CFMR, CFOR, CFR, CHMR, CHR, CHRX, CMOR, CMR, CR, FMR, FR, FRX, HMOR, HOR, HORX, HR, MOR, MR, MRX, OR, ORX, RXZ, RZ
Third-Line	CFMOR, CFMR, CFR, CHMR, CHR, CHRX, CMOR, CMORX, CMR, CR, CRX, FMR, FMRZ, FR, HMOR, HOR, HR, MOR, MR, MRX, MRXZ, OR, RXZ, RZ

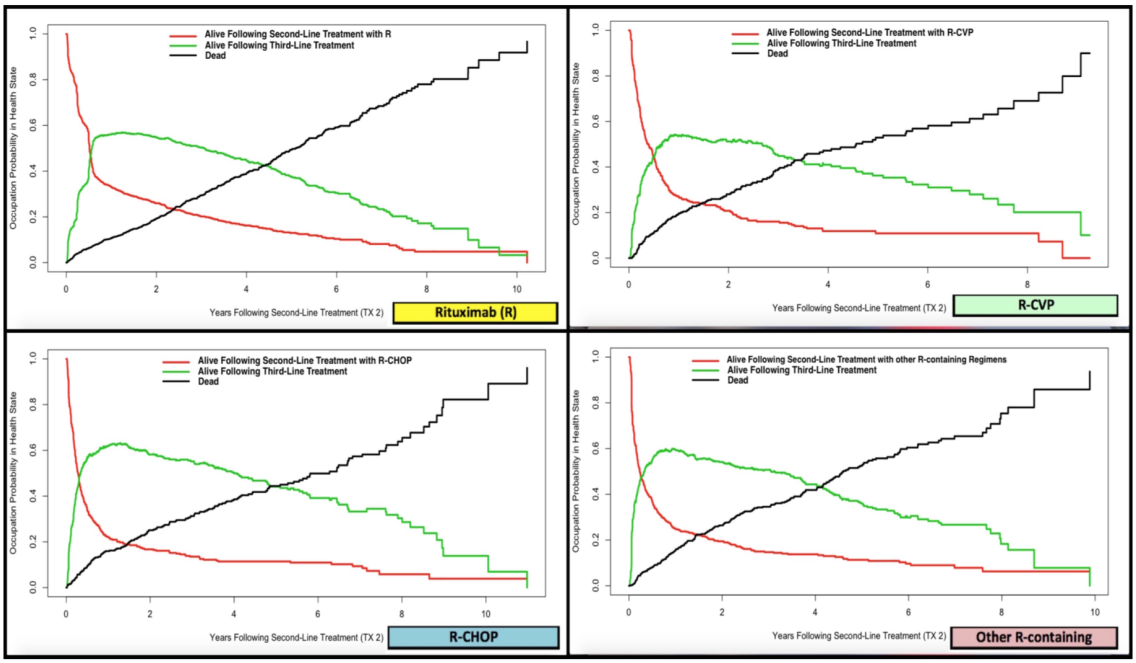


Figure C.1: Graph of Occupation Probabilities over Time for Second-Line Treatments

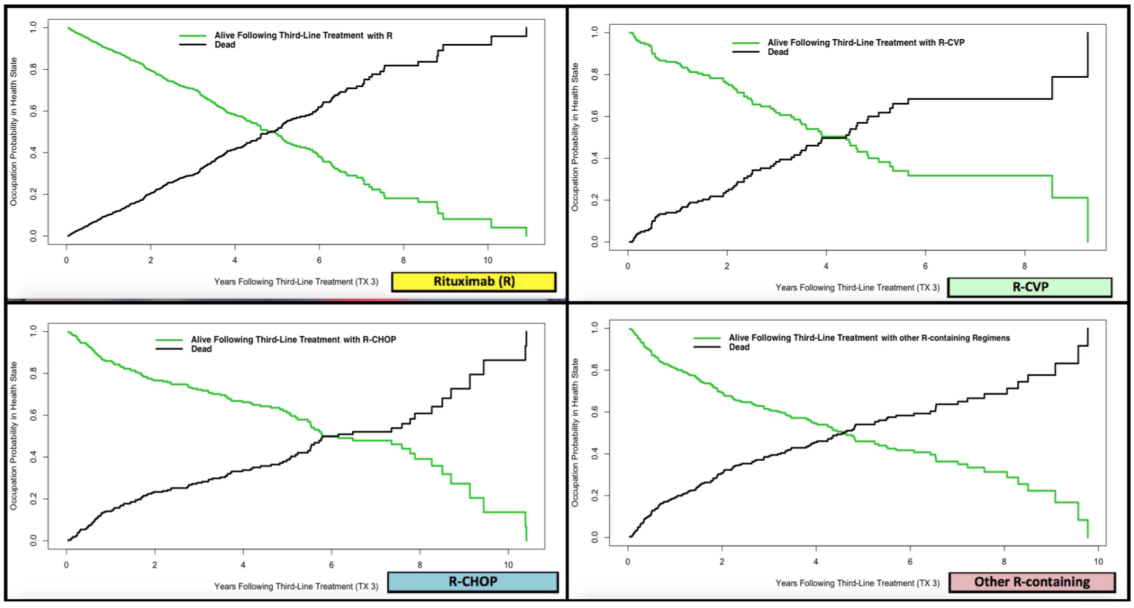


Figure C.2: Graph of Occupation Probabilities over Time for Third-Line Treatments

C.2 A Population-Based Multi-state Model for Diffuse Large B Cell Lymphoma-Specific Mortality in Older Patients

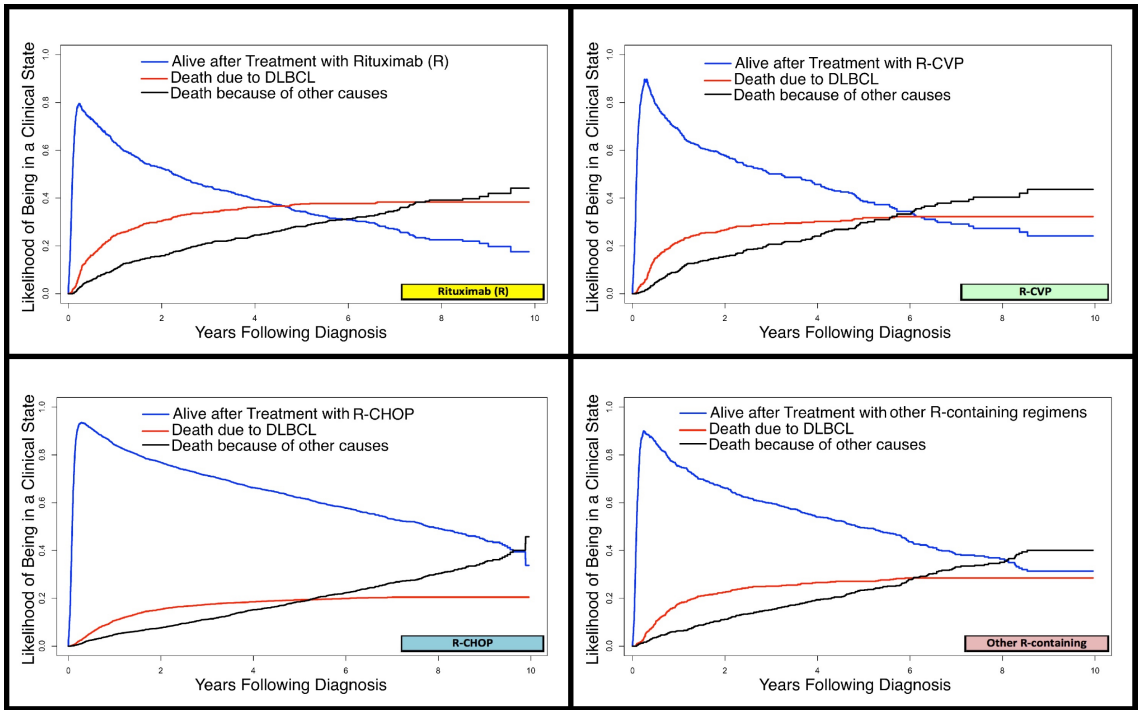


Figure C.3: Survival and Cause-Specific Death Probabilities for Initial Treatments

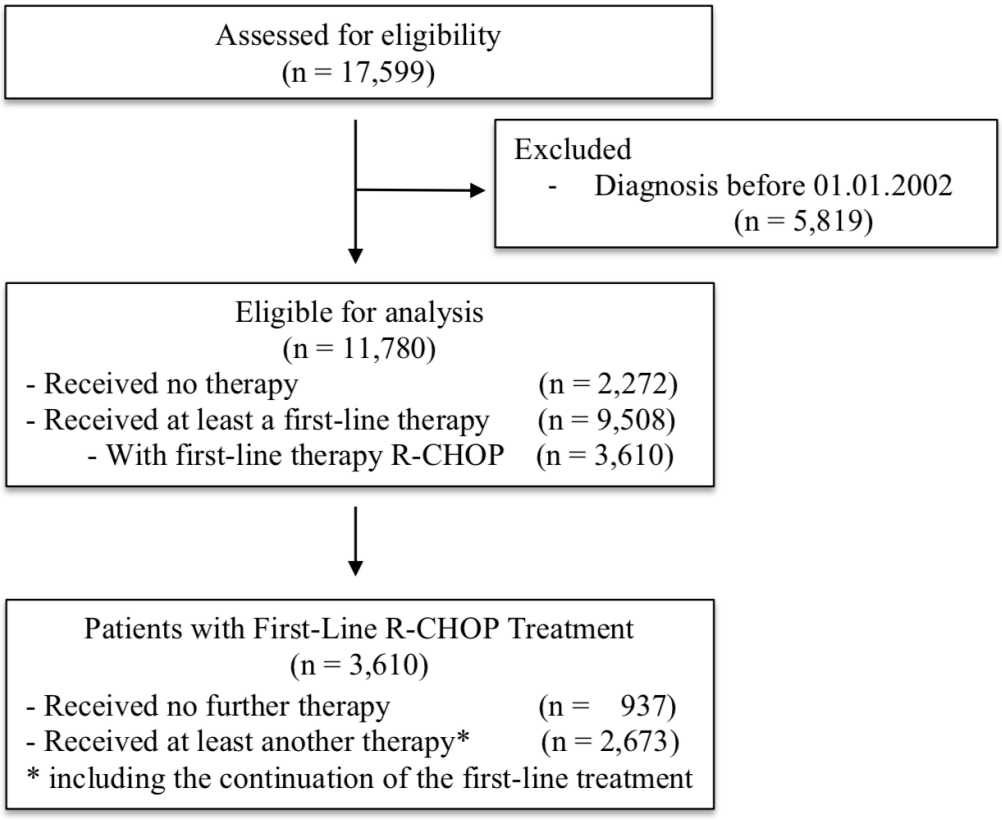


Figure C.4: CONSORT Flow Diagram Reporting the Number of Patients

REFERENCES

[1] American, Cancer, and Society, *Cancer facts & figures 2017, atlanta,ga*, Atlanta,GA, 2017.

[2] A. Antoniou, P. D. P. Pharoah, S. Narod, H. A. Risch, J. E. Eyfjord, J. L. Hopper, N. Loman, H. Olsson, O. Johannsson, A. Borg, and B. Pasini, "Average risks of breast and ovarian cancer associated with brca1 or brca2 mutations detected in case series unselected for family history: A combined analysis of 22 studies," *The american journal of human genetics*, vol. 72, no. 5, pp. 1117–1130, 2003.

[3] S. Chen and G. Parmigiani, "Meta-analysis of brca1 and brca2 penetrance," *Journal of clinical oncology*, vol. 25, no. 11, pp. 1329–1333, 2007.

[4] Breastcancer.org, *U.s. breast cancer statistics*, Last Modified on January 9, 2018 at 10:49 AM, 2015.

[5] C. G. on Hormonal Factors in Breast Cancer (CGHFBC), "Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including women with breast cancer and 101,986 women without the disease.," *The lancet*, vol. 358, no. 209, pp. 1389–1399, 2001.

[6] M. Wood, B. Flynn, and A. Stockdale, "Primary care physician management, referral, and relations with specialists concerning patients at risk for cancer due to family history," *Public health genomics*, vol. 16, no. 3, pp. 75–82, 2013.

[7] M. Wood, P. Kadlubek, T. Pham, D. Wollins, K. Lu, J. Weitzel, M. Neuss, and K. Hughes, "Quality of cancer family history and referral for genetic counseling and testing among oncology practices: A pilot test of quality measures as part of the american society of clinical oncology quality oncology practice initiative," *Journal of clinical oncology*, vol. 32, no. 8, pp. 824–829, 2014.

[8] B. Drohan, C. Roche, J. Cusack, and K. Hughes, "Hereditary breast and ovarian cancer and other hereditary syndromes: Using technology to identify carriers," *Annals of surgical oncology*, vol. 19, no. 6, pp. 1732–1737, 2012.

[9] C. Genomics, *Color's brca test - get the brca test for* $*99*, Accessed on January 30, 2018, 2018.

[10] E. Lopatto, "Genetic testing for breast cancer gets more affordable - welcome to the post-patent world," *The verge*, 2015.

[11] E. Long and P. Ganz, "Cost-effectiveness of universal brca1/2 screening: Evidence-based decision making," *Jama oncology*, vol. 1, no. 9, pp. 1217–1218, 2015.

[12] F. Guo, J. Hirth, Y. Lin, G. Richardson, L. Levine, A. Berenson, and Y. Kuo, "Use of brca mutation test in the us, 2004?2014," *American journal of preventive medicine*, vol. 52, no. 6, pp. 702–709, 2017.

[13] Z. Chen, K. Kolor, S. Grosse, J. Rodriguez, J. Lynch, R. Green, W. Dotson, M. Bowen, and M. Khoury, "Trends in utilization and costs of brca testing among women aged 18?64 years in the united states, 2003?2014," *Genetics in medicine*, 2017.

[14] K. A. Metcalfe, D. Birenbaum-Carmeli, J. Lubinski, J. Gronwald, H. Lynch, P. Moller, P. Ghadirian, W. D. Foulkes, J. Klijn, E. Friedman, and C. Kim Sing, "International variation in rates of uptake of preventive options in brca1 and brca2 mutation carriers," *International journal of cancer*, vol. 122, no. 9, pp. 2017–2022, 2008.

[15] S. G. Komen, *Preventive surgery prophylactic mastectomy*, 2016.

[16] W. Owens, T. Gallagher, M. Kincheloe, and V. Ruetten, "Implementation in a large health system of a program to identify women at high risk for breast cancer," *Journal of oncology practice*, vol. 7, no. 2, pp. 85–88, 2011.

[17] L. Reimers and K. Crew, "Tamoxifen versus raloxifene versus exemestane for chemoprevention," *Current breast cancer reports*, vol. 4, no. 3, pp. 207–215, 2012.

[18] A. C. S. (ACS), *Breast cancer facts & figures 2015-2016, atlanta,ga*, Atlanta,GA, 2015.

[19] C. H. Lee, D. D. Dershaw, D. Kopans, P. Evans, B. Monsees, D. Monticciolo, R. J. Brenner, L. Bassett, W. Berg, S. Feig, and E. Hendrick, "Breast cancer screening with imaging: Recommendations from the society of breast imaging and the acr on the use of mammography, breast mri, breast ultrasound, and other technologies for the detection of clinically occult breast cancer," *Journal of the american college of radiology*, vol. 7, no. 1, pp. 18–27, 2010.

[20] L. C. Farrington, "2013 breast cancer screening guidelines and recommendations," *Home health care management & practice*, vol. 26, no. 1, pp. 45–48, 2014.

[21] S. O.S.S.E. G. (SOSSEG), "Reduction in breast cancer mortality from organized service screening with mammography: 1. further confirmation with extended data," *Cancer epidemiology biomarkers & prevention*, vol. 15, no. 1, p. 45, 2006.

[22] C. K. Kuhl, S. Schrading, C. C. Leutner, N. Morakkabati-Spitz, E. Wardelmann, R. Fimmers, W. Kuhn, and H. H. Schild, "Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer," *Journal of clinical oncology*, vol. 23, no. 33, pp. 8469–8476, 2005.

[23] E. Warner, H. Messersmith, P. Causer, A. Eisen, R. Shumak, and D. Plewes, "Systematic review: Using magnetic resonance imaging to screen women at high risk for breast cancer," *Annals of internal medicine*, vol. 148, no. 9, pp. 671–679, 2008.

[24] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe, "Mammographic density and the risk and detection of breast cancer," *New england journal of medicine*, vol. 356, no. 3, pp. 227–236, 2007.

[25] W. A. Berg, Z. Zhang, D. Lehrer, R. A. Jong, E. D. Pisano, R. G. Barr, M. Bohm-Velez, M. C. Mahoney, W. P. Evans, L. H. Larsen, and M. J. Morton, "Detection of breast cancer with addition of annual screening ultrasound or a single screening mri to mammography in women with elevated breast cancer risk," *Jama*, vol. 307, no. 13, pp. 1394–1404, 2012.

[26] W. Berg, A. Bandos, E. Mendelson, D. Lehrer, R. Jong, and E. Pisano, "Ultrasound as the primary screening test for breast cancer: Analysis from acrin 6666," *Jnci: Journal of the national cancer institute*, vol. 108, no. 4, 2016.

[27] E. J. Granader, B. Dwamena, and R. C. Carlos, "Mri and mammography surveillance of women at increased risk for breast cancer: Recommendations using an evidence-based approach," *Academic radiology*, vol. 15, no. 12, pp. 1590–1595, 2008.

[28] A. Narayan, K. Visvanathan, and S. Harvey, "Comparative effectiveness of breast mri and mammography in screening young women with elevated risk of developing breast cancer: A retrospective cohort study," *Breast cancer research and treatment*, vol. 158, no. 3, pp. 583–589, 2016.

[29] J. R. Scheel, J. M. Lee, B. L. Sprague, C. I. Lee, and C. D. Lehman, "Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts," *American journal of obstetrics and gynecology*, vol. 212, no. 1, pp. 9–17, 2015.

[30] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, and R. A. Smith, "American cancer society guidelines for breast screening with mri as an adjunct to mammography," *Ca: A cancer journal for clinicians*, vol. 57, no. 2, pp. 75–89, 2007.

[31] A. N. Tosteson, N. K. Stout, D. G. Fryback, S. Acharyya, B. A. Herman, L. G. Hannah, and E. D. Pisano, "Cost-effectiveness of digital mammography breast cancer screening," *Annals of internal medicine*, vol. 148, no. 1, pp. 1–10, 2008.

[32] S. Saadatmand, M. M. Tilanus-Linthorst, E. J. Rutgers, N. Hoogerbrugge, J. C. Oosterwijk, R. A. Tollenaar, M. Hooning, C. E. Loo, I. M. Obdeijn, E. A. Heijnsdijk, and H. J. de Koning, "Cost-effectiveness of screening women with familial risk for breast cancer with magnetic resonance imaging," *Journal of the national cancer institute*, vol. 105, no. 17, pp. 1314–1321, 2013.

[33] W. A. Berg, J. D. Blume, J. B. Cormack, E. B. Mendelson, D. Lehrer, M. Behm-Velez, E. D. Pisano, R. A. Jong, W. P. Evans, M. J. Morton, and M. C. Mahoney, "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer," *Jama*, vol. 299, no. 18, pp. 2151–2163, 2008.

[34] W. Berg, J. Blume, J. Cormack, and E. Mendelson, "Operator dependence of physician-performed whole-breast us: Lesion detection and characterization," *Radiology*, vol. 241, pp. 355–365, 2006.

[35] T. Greene, C. Cocilovo, A. Estabrook, L. Chinitz, C. Giuliano, S. Smith, and P. Tartter, "A single institution review of new breast malignancies identified solely by sonography," *Journal of the american college of surgeons*, vol. 203, no. 6, pp. 894–898, 2006.

[36] N. Houssami and. Miglioretti D., *Breast cancer screening: Making sense of complex and evolving evidence. chapter 10: Screening for breast cancer*. Academic Press, 2016.

[37] S. Zervoudis, G. Iatrakis, E. Tomara, A. Bothou, G. Papadopoulos, and G. Tsakiris, "Main controversies in breast cancer," *World journal of clinical oncology*, vol. 5, no. 3, p. 359, 2014.

[38] F. Sardanelli, F. Podo, F. Santoro, S. Manoukian, S. Bergonzi, G. Trecate, D. Vergnaghi, M. Federico, L. Cortesi, S. Corcione, and S. Morassut, "Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk italian 1 study): Final results," *Investigative radiology*, vol. 46, no. 2, pp. 94–105, 2011.

[39] C. I. Lee, L. W. Bassett, and C. D. Lehman, "Breast density legislation and opportunities for patient-centered outcomes research," *Radiology*, vol. 264, no. 3, pp. 632–636, 2012.

[40] R. Sullivan, J. Peppercorn, K. Sikora, J. Zalcberg, N. J. Meropol, E. Amir, D. Khayat, P. Boyle, P. Autier, I. F. Tannock, and T. Fojo, "Delivering affordable cancer care in high-income countries," *The lancet oncology*, vol. 12, no. 10, pp. 933–980, 2011.

[41] A. B. Mariotto, K. R. Yabroff, Y. Shao, E. J. Feuer, and M. L. Brown, *Projections of the cost of cancer care in the united states: 2010-2020*. Journal of the National Cancer Institute, 2011.

[42] M. A. Dinan, L. H. Curtis, B. G. Hammill, E. F. Patz, A. P. Abernethy, A. M. Shea, and K. A. Schulman, "Changes in the use and costs of diagnostic imaging among medicare beneficiaries with cancer, 1999-2006," *Jama*, vol. 303, no. 16, pp. 1625–1631, 2010.

[43] K. R. Yabroff, E. B. Lamont, A. Mariotto, J. L. Warren, M. Topor, A. Meekins, and M. L. Brown, "Cost of care for elderly cancer patients in the united states," *Journal of the national cancer institute*, vol. 100, no. 9, pp. 630–641, 2008.

[44] S. Suen and J. D. Goldhaber-Fiebert, "An efficient, noniterative method of identifying the cost-effectiveness frontier," *Medical decision making*, vol. 36, no. 1, pp. 132–136, 2016.

[45] D. M. Negoescu, K. Bimpikis, M. L. Brandeau, and D. A. Iancu, "Dynamic learning of patient response types: An application to treating chronic diseases," *Management science*, 2017.

[46] S. Liu, L. C. LE, M. H. M, and J. Goldhaber-Fiebert, "Cost-effectiveness analysis of risk-factor guided and birth-cohort screening for chronic hepatitis c infection in the united states," *Plos one*, vol. 8, no. 3, 2013.

[47] L. Cipriano, C. Rupar, and G. Zaric, "The cost-effectiveness of expanding newborn screening for up to 21 inherited metabolic disorders using tandem mass spectrometry: Results from a decision-analytic model," *Value in health*, vol. 10, no. 2, pp. 83–97, 2007.

[48] S. K. Plevritis, A. W. Kurian, B. M. Sigal, B. L. Daniel, D. M. Ikeda, F. E. Stockdale, and A. M. Garber, "Cost-effectiveness of screening brca1/2 mutation carriers with breast magnetic resonance imaging," *Jama*, vol. 295, no. 20, pp. 2374–2384, 2006.

[49] I. Griebsch, J. Brown, C. Boggis, A. Dixon, M. Dixon, D. Easton, R. Eeles, D. G. Evans, F. J. Gilbert, J. Hawnaur, and P. Kessar, "Cost-effectiveness of screening with contrast enhanced magnetic resonance imaging vs x-ray mammography of women at a high familial risk of breast cancer," *British journal of cancer*, vol. 95, no. 7, pp. 801–810, 2006.

[50] C. Taneja, J. Edelsberg, D. Weycker, A. Guo, G. Oster, and J. Weinreb, "Cost effectiveness of breast cancer screening with contrast-enhanced mri in high-risk women," *Journal of the american college of radiology*, vol. 6, no. 3, pp. 171–179, 2009.

[51] S. G. Moore, P. J. Shenoy, L. Fanucchi, J. W. Tumeh, and C. R. Flowers, "Cost-effectiveness of mri compared to mammography for breast cancer screening in a high risk population," *Bmc health services research*, vol. 9, p. 1, 2009.

[52] C. Chubiz, L. J. E., G. J. M., K. M. E., L. C. Y., H. K. P., M. E. F., R. P. M., and G. Gazelle, "Cost-effectiveness of alternating magnetic resonance imaging and digital mammography screening in brca1 and brca2 gene mutation carriers," *Cancer*, vol. 119, no. 6, pp. 1266–1276, 2013.

[53] R. Pataky, L. Armstrong, S. Chia, A. J. Coldman, C. Kim-Sing, B. McGillivray, J. Scott, C. M. Wilson, and S. Peacock, "Cost-effectiveness of mri for breast cancer screening in brca1/2 mutation carriers," *Bmc cancer*, vol. 13, p. 1, 2013.

[54] D. Bock, V. G. H., J. K. M., O. L., S. J. C., D. S., F. M. D., H. N. T., and M. W. Greuter, "Which screening strategy should be offered to women with brca 1 or brca2 mutations & quest; a simulation of comparative cost-effectiveness.," *British journal of cancer*, vol. 108, no. 8, pp. 1579–1586, 2013.

[55] O. Alagoz, T. Ayer, and F. S. Erenay, *Operations research models for cancer screening*. Encyclopedia of Operations Research and Management Science: Wiley, 2011.

[56] L. M. Maillart, J. S. Ivy, S. Ransom, and K. Diehl, "Assessing dynamic breast cancer screening policies," *Operations research*, vol. 56, no. 6, pp. 1411–1427, 2008.

[57] M. S. Rauner, W. J. Gutjahr, K. Heidenberger, J. Wagner, and J. Pasia, "Dynamic policy modeling for chronic diseases: Metaheuristic-based identification of pareto-optimal screening strategies," *Operations research*, vol. 58, no. 5, pp. 1269–1286, 2010.

[58] T. Ayer, O. Alagoz, and N. K. Stout, "Or forum-a pomdp approach to personalize mammography screening decisions," *Operations research*, vol. 60, no. 5, pp. 1019–1034, 2012.

[59] M. Cevik, T. Ayer, O. Alagoz, and B. Sprague, "Analysis of mammography screening policies under resource constraints," *Production and operations management*, vol. 27, no. 5, pp. 949–972, 2018.

[60] J. Chhatwal, O. Alagoz, and E. S. Burnside, "Optimal breast biopsy decision-making based on mammographic features and demographic factors," *Operations research*, vol. 58, no. 6, pp. 1577–1591, 2010.

[61] M. U. Ayvaci, O. Alagoz, and E. S. Burnside, "The effect of budgetary restrictions on breast cancer diagnostic decisions," *Manufacturing & service operations management*, vol. 14, no. 4, pp. 600–617, 2012.

[62] F. Erenay, O. Alagoz, and A. Said, "Optimizing colonoscopy screening for colorectal cancer prevention and surveillance," *Manufacturing & service operations management*, vol. 3, no. 16, pp. 381–400, 2014.

[63] J. Zhang, B. Denton, H. Balasubramanian, N. Shah, and B. Inman, "Optimization of psa screening policies: A comparison of the patient and societal perspectives," *Medical decision making*, vol. 2, no. 32, pp. 337–349, 2012.

[64] D. Hutton, M. Brandeau, and S. So, "Doing good with good or: Supporting cost-effective hepatitis b interventions," *Interfaces*, vol. 41, no. 3, pp. 289–300, 2011.

[65] E. Yaylali, P. Farnham, K. Schneider, S. Landers, O. Kouzouian, A. Lasry, D. Purcell, T. Green, and S. Sansom, "From theory to practice: Implementation of a resource allocation model in health departments," *Journal of public health management and practice (jphmp)*, vol. 22, no. 6, p. 567, 2016.

[66] E. Güneş, E. Örmeci, and D. Kunduzcu, "Preventing and diagnosing colorectal cancer with a limited colonoscopy resource," *Production and operations management*, vol. 24, no. 1, pp. 1–20, 2015.

[67] E. Örmeci, E. Güneş, and D. Kunduzcu, "A modeling framework for control of preventive services," *Manufacturing & service operations management*, vol. 18, no. 2, pp. 227–244, 2015.

[68] E. Nohdurft, E. Long, and S. Spinler, "Was angelina jolie right? optimizing cancer prevention strategies among brca mutation carriers," *Decision analysis*, vol. 14, no. 3, pp. 139–169, 2017.

[69] M. C. Weinstein, J. E. Siegel, M. R. Gold, M. S. Kamlet, and L. B. Russell, "Recommendations of the panel on cost-effectiveness in health and medicine," *Jama*, vol. 276, no. 15, pp. 1253–1258, 1996.

[70] P. Neumann, G. Sanders, L. Russell, J. Siegel, and T. Ganiats, *Cost-effectiveness in health and medicine*. Oxford University Press, 2016.

[71] J. Yoon, M. Kim, E. Kim, and H. Moon, "Imaging surveillance of patients with breast cancer after primary treatment: Current recommendations," *Korean journal of radiology*, vol. 16, no. 2, pp. 219–228, 2015.

[72] D. Lam, N. Houssami, and J. Lee, "Imaging surveillance after primary breast cancer treatment," *American journal of roentgenology*, vol. 208, no. 3, pp. 676–686, 2017.

[73] S. H. Parker, F. Burbank, R. J. Jackman, C. J. Aucreman, G. Cardenosa, T. M. Cink, C. Jr, E. J. L., W. G. W. Evans 3rd, and P. Garver, "Percutaneous large-core breast biopsy: A multi-institutional study.," *Radiology*, vol. 193, no. 2, pp. 359–364, 1994.

[74] D. G. Fryback, N. K. Stout, M. A. Rosenberg, A. Trentham-Dietz, V. Kuruchittham, and P. L. Remington, "The wisconsin breast cancer epidemiology simulation model," *Monographs-national cancer institute*, vol. 36, p. 37, 2006.

[75] A. C. Antoniou, P. P. D. Pharoah, P. Smith, and D. F. Easton, "The boadicea model of genetic susceptibility to breast and ovarian cancer," *British journal of cancer*, vol. 91, no. 8, pp. 1580–1590, 2004.

[76] A. C. Antoniou, A. P. Cunningham, J. Peto, D. G. Evans, F. Lalloo, S. A. Narod, H. A. Risch, J. E. Eyfjord, J. L. Hopper, M. C. Southey, and H. Olsson, "The boadicea model of genetic susceptibility to breast and ovarian cancers: Updates and extensions," *British journal of cancer*, vol. 98, no. 8, pp. 1457–1466, 2008.

[77] R. J. MacInnis, A. Bickerstaffe, C. Apicella, G. S. Dite, J. G. Dowty, K. Aujard, K. A. Phillips, P. Weideman, A. Lee, M. B. Terry, and G. G. Giles, "Prospective validation of the breast cancer risk prediction model boadicea and a batch-mode version boadiceacentre," *British journal of cancer*, vol. 109, no. 5, pp. 1296–1301, 2013.

[78] A. C. Antoniou, R. Hardy, L. Walker, D. G. Evans, A. Shenton, R. Eeles, S. Shanley, G. Pichert, L. Izatt, S. Rose, and F. Douglas, "Predicting the likelihood of carrying a brca1 or brca2 mutation: Validation of boadicea, brcapro, ibis, myriad and the manchester scoring system using data from uk genetics clinics," *Journal of medical genetics*, vol. 45, no. 7, pp. 425–431, 2008.

[79] A. C. Antoniou, F. Durocher, P. Smith, J. Simard, and D. F. Easton, "Brca1 and brca2 mutation predictions using the boadicea and brcapro models and penetrance estimation in high-risk french-canadian families," *Breast cancer research*, vol. 8, p. 1, 2005.

[80] C. H. Barcenas, G. M. Hosain, B. Arun, J. Zong, X. Zhou, J. Chen, J. M. Cortada, G. B. Mills, G. E. Tomlinson, A. R. Miller, and L. C. Strong, "Assessing brca carrier probabilities in extended families," *Journal of clinical oncology*, vol. 24, no. 3, pp. 354–360, 2006.

[81] A. J. Lee, A. P. Cunningham, K. B. Kuchenbaecker, N. Mavaddat, D. F. Easton, and A. C. Antoniou, "Boadicea breast cancer risk prediction model: Updates to cancer incidences, tumour pathology and web interface," *British journal of cancer*, vol. 110, no. 2, pp. 535–545, 2014.

[82] J. Costantino, M. Gail, D. Pee, S. Anderson, C. Redmond, J. Benichou, and H. Wieand, "Validation studies for models projecting the risk of invasive and total breast cancer incidence," *Journal of the national cancer institute*, vol. 91, no. 18, pp. 1541–1548, 1999.

[83] M. Gail and P. Mai, "Comparing breast cancer risk assessment models," *Jnci: Journal of the national cancer institute*, vol. 102, no. 10, pp. 665–668, 2010.

[84] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun, "Cancer statistics," *Ca: A cancer j. clinicians*, vol. 59, no. 4, pp. 225–249, 2009.

[85] E. Arias, "United states life tables: 2004," *National vital statistics reports*, vol. 56, no. 9, pp. 1–40, 2007.

[86] L. Walter and M. Schonberg, "Screening mammography in older women: A review.," *Jama*, vol. 311, no. 13, pp. 1336–1347, 2014.

[87] M. H. Gail, J. P. Costantino, J. Bryant, R. Croyle, L. Freedman, K. Heizlsouer, and V. Vogel, "Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer," *Journal of the national cancer institute*, vol. 91, no. 21, pp. 1829–1846, 1999.

[88] A. M. A. (AMA), *Medicare resource-based relative value scale 2005. chicago,il 3: American medical association: 2005*, 2005.

[89] N. funded Breast Cancer Surveillance Consortium (HHSN261201100031C), "Downloaded on november 30, 2014 from the breast cancer surveillance consortium web site - `http://breastscreening.cancer.gov/`.,"

[90] M. O. Leach, C. R. Boggis, A. K. Dixon, D. F. Easton, R. A. Eeles, D. G. Evans, F. J. Gilbert, I. Griebsch, R. J. Hoff, P. Kessar, S. R. Lakhani, S. M. Moss, A. Nerurkar, A. R. Padhani, L. J. Pointon, D. Thompson, and W. R.M.M. S. Group, "Screening with magnetic resonance imaging and mammography of a uk population at high familial risk of breast cancer: A prospective multicentre cohort study (maribs)," *The lancet*, vol. 365, no. 9473, pp. 1769–1778, 2005.

[91] M. Kriege, C. T. Brekelmans, C. Boetes, P. E. Besnard, H. M. Zonderland, I. M. Obdeijn, R. A. Manoliu, T. Kok, H. Peterse, M. M. Tilanus-Linthorst, and S. H. Muller, "Efficacy of mri and mammography for breast-cancer screening in women

with a familial or genetic predisposition," *New england journal of medicine*, vol. 351, no. 5, pp. 427–437, 2004.

[92] A. I. Hagen, K. A. Kvistad, L. Maehle, M. M. Holmen, H. Aase, B. Styr, A. Vab, J. Apold, P. Skaane, and P. Miller, "Sensitivity of mri versus conventional screening in the diagnosis of brca-associated breast cancer in a national prospective series," *The breast*, vol. 16, no. 4, pp. 367–374, 2007.

[93] T. M. Kolb, J. Lichy, and J. H. Newhouse, "Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: An analysis of 27,825 patient evaluations 1," *Radiology*, vol. 225, no. 1, pp. 165–175, 2002.

[94] N. Houssami, S. Ciatto, L. Irwig, J. M. Simpson, and P. Macaskill, "The comparative sensitivity of mammography and ultrasound in women with breast symptoms: An age-specific analysis," *The breast*, vol. 11, no. 2, pp. 125–130, 2002.

[95] C. D. Lehman, C. I. Lee, V. A. Loving, M. S. Portillo, S. Peacock, and W. B. DeMartini, "Accuracy and value of breast ultrasound for primary imaging evaluation of symptomatic women 30-39 years of age," *American journal of roentgenology*, vol. 199, no. 5, pp. 1169–1177, 2012.

[96] J. S. Mandelblatt, M. E. Wheat, M. Monane, R. D. Moshief, J. P. Hollenberg, and J. Tang, "Breast cancer screening for elderly women with and without comorbid conditions: A decision analysis model," *Annals of internal medicine*, vol. 116, no. 9, pp. 722–730, 1992.

[97] I. T. Gram, E. Lund, and S. E. Slenker, "Quality of life following a false positive mammogram," *British journal of cancer*, vol. 62, no. 6, p. 1018, 1990.

[98] V. Velanovich, "Immediate biopsy versus observation for abnormal findings on mammograms: An analysis of potential outcomes and costs," *The american journal of surgery*, vol. 170, no. 4, pp. 327–332, 1995.

[99] C. C. Earle, R. H. Chapman, C. S. Baker, C. M. Bell, P. W. Stone, E. A. Sandberg, and P. J. Neumann, "Systematic overview of cost-utility assessments in oncology," *Journal of clinical oncology*, vol. 18, no. 18, pp. 3302–3317, 2000.

[100] A. Detsky and I. Naglie, "A clinician's guide to cost-effectiveness analysis," *Annals of internal medicine*, vol. 113, no. 2, pp. 147–154, 1990.

[101] P. Neumann, J. Cohen, and M. Weinstein, "Updating cost-effectiveness-the curious resilience of the $50,000-per-qaly threshold," *New england journal of medicine*, vol. 371, no. 9, pp. 796–797, 2014.

[102] S. B. Cantor, "Cost-effectiveness analysis, extended dominance, and ethics a quantitative assessment," *Medical decision making*, vol. 14.3, pp. 259–265, 1994.

[103] E. S. Burnside, J. Chhatwal, and O. Alagoz, "What is the optimal threshold at which to recommend breast biopsy?" *Plos one*, vol. 7, no. 11, e48820, 2012.

[104] H. Shin, H. Kim, and J. Cha, "Current status of automated breast ultrasonography," *Ultrasonography*, vol. 34, no. 3, p. 165, 2015.

[105] L. Liberman, "Breast cancer screening with mri - what are the data for patients at high risk?" *The new england journal of medicine*, vol. 351, no. 5, p. 497, 2004.

[106] O. Omidiji, P. Campbell, N. Irurhe, O. Atalabi, and O. Toyobo, "Breast cancer screening in a resource poor country: Ultrasound versus mammography," *Ghana medical journal*, vol. 51, no. 1, pp. 6–12, 2017.

[107] P. Rajaraman, B. Anderson, P. Basu, J. Belinson, A. D'Cruz, P. Dhillon, P. Gupta, T. Jawahar, N. Joshi, U. Kailash, and S. Kapambwe, "Recommendations for screening and early detection of common cancers in india," *The lancet oncology*, vol. 16, no. 7, e352–e361, 2015.

[108] M. Galukande and E. Kiguli-Malwadde, "Rethinking breast cancer screening strategies in resource-limited settings," *African health sciences*, vol. 10, no. 1, 2010.

[109] C. Brekelmans, C. Seynaeve, C. Bartels, M. Tilanus-Linthorst, E. Meijers-Heijboer, C. Crepin, A. van Geel, M. Menke, L. Verhoog, A. van Den Ouweland, and I. Obdeijn, "Effectiveness of breast cancer surveillance in brca1/2 gene mutation carriers and women with high familial risk," *Journal of clinical oncology*, vol. 19, no. 4, pp. 924–930, 2001.

[110] L. Scheuer, N. Kauff, M. Robson, B. Kelly, R. Barakat, J. Satagopan, N. Ellis, M. Hensley, J. Boyd, P. Borgen, and L. Norton, "Outcome of preventive surgery and screening for breast and ovarian cancer in brca mutation carriers," *Journal of clinical oncology*, vol. 20, no. 5, pp. 1260–1268, 2002.

[111] M. O'Connor, "Molecular breast imaging: An emerging modality for breast cancer screening," *Jbreast cancer management*, vol. 4, no. 1, pp. 33–40, 2015.

[112] S. Patterson and M. Roubidoux, "Update on new technologies in digital mammography," *International journal of women's health*, vol. 6, p. 781, 2014.

[113] W. Northington, J. Brice, and B. Zou, "Use of an emergency department by nonurgent patients," *The american journal of emergency medicine*, vol. 23, no. 2, pp. 131–137, 2005.

[114] N. C. for Health Statistics (NCHS), "National hospital ambulatory medical care survey: 2014 emergency department summary tables," *Centers for disease control and prevention*, 2014.

[115] J. Adams, "Emergency department overuse: Perceptions and solutions," *Jama*, vol. 309, no. 11, pp. 1173–1174, 2013.

[116] S. Bernstein, D. Aronsky, R. Duseja, S. Epstein, D. Handel, U. Hwang, M. McCarthy, K. John McConnell, J. Pines, N. Rathlev, and R. Schafermeyer, "The effect of emergency department crowding on clinically oriented outcomes," *Academic emergency medicine*, vol. 16, no. 1, pp. 1–10, 2009.

[117] K. Johnson and C. Winkelman, "The effect of emergency department crowding on patient outcomes: A literature review," *Advanced emergency nursing journal*, vol. 33, no. 1, pp. 39–54, 2011.

[118] P. Smulowitz, L. Honigman, and B. Landon, "A novel approach to identifying targets for cost reduction in the emergency department," *Annals of emergency medicine*, vol. 61, no. 3, pp. 293–300, 2013.

[119] E. Berger, "Physician burnout: Emergency physicians see triple risk of career affliction," *Annals of emergency medicine*, vol. 61, no. 3, A17–A19, 2013.

[120] R. Batt and C. Terwiesch, "Waiting patiently: An empirical study of queue abandonment in an emergency department," *Management science*, vol. 61, no. 1, pp. 39–59, 2015.

[121] J. Lam, "The emn salary survey: Now and then how the average ep changed in two years," *Emergency medicine news*, vol. 40, no. 5, p. 28, 2018.

[122] M. Wargon, B. Guidet, T. Hoang, and G. Hejblum, "A systematic review of models for forecasting the number of emergency department visits," *Emergency medicine journal*, vol. 26, no. 6, pp. 395–399, 2009.

[123] S. Pitts, J. Pines, M. Handrigan, and A. Kellermann, "National trends in emergency department occupancy, 2001 to 2008: Effect of inpatient admissions versus emergency department practice intensity," *Annals of emergency medicine*, vol. 60, no. 6, pp. 679–686, 2012.

[124] J. Boyle, M. Jessup, J. Crilly, D. Green, J. Lind, M. Wallis, P. Miller, and G. Fitzgerald, "Predicting emergency department admissions," *Emergency medicine journal*, vol. 29, no. 5, pp. 358–365, 2012.

[125] J. C. Moskop, D. P. Sklar, J. M. Geiderman, R. M. Schears, and K. J. Bookman, "Emergency department crowding, part 1 - concept, causes, and moral consequences," *Annals of emergency medicine*, vol. 53, no. 5, pp. 605–611, 2009.

[126] D. Sinreich and Y. Marmor, "Emergency department operations: The basis for developing a simulation tool," *Iie transactions*, vol. 37, no. 3, pp. 233–245, 2005.

[127] G. Yom-Tov and A. Mandelbaum, "Erlang-r: A time-varying queue with reentrant customers, in support of healthcare staffing," *Manufacturing & service operations management*, vol. 16, no. 2, pp. 283–299, 2014.

[128] N. Gilboy, T. Tanabe, D. Travers, and A. Rosenau, "Chapter 3. esi level 2, chapter 4. esi levels 3-5 and expected resource need," in *Emergency severity index (esi): A triage tool for emergency department care, implementation handbook*, Agency for Healthcare Research and Quality (AHRQ) Publication No. 11(12)-0014, 2011.

[129] A. Sayah, L. Rogers, K. Devarajan, L. Kingsley-Rocker, and L. Lobon, "Minimizing ed waiting times and improving patient flow and experience of care," *Emergency medicine international*, 2014.

[130] A. Guttmann, M. Schull, M. Vermeulen, and T. Stukel, "Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from ontario, canada," *The bmj*, p. 2983, 2011.

[131] E. Carter, S. Pouch, and E. Larson, "The relationship between emergency department crowding and patient outcomes: A systematic review," *Journal of nursing scholarship*, vol. 46, no. 2, pp. 106–115, 2014.

[132] C. Fernandes, P. Tanabe, N. Gilboy, L. Johnson, R. McNair, A. Rosenau, P. Sawchuk, D. Thompson, D. Travers, N. Bonalumi, and R. Suter, "Five level triage: A report from the acep/ena five level triage task force," *Journal of emergency nursing*, vol. 31, no. 1, pp. 39–50, 2005.

[133] Y. Liu, "Staffing to stabilize the tail probability of delay in service systems with time-varying demand," *Operations research*, 2018.

[134] C. Lakshmi and S. Iyer, "Application of queueing theory in health care: A literature review," *Operations research for health care*, vol. 2, no. 1, pp. 25–39, 2013.

[135] S. Saghafian, G. Austin, and S. Traub, "Operations research/management contributions to emergency department patient flow optimization: Review and research prospects," *Iie transactions on healthcare systems engineering*, vol. 5, no. 2, pp. 101–123, 2015.

[136] M. Defraeye and I. Van Nieuwenhuyse, "Staffing and scheduling under nonstationary demand for service: A literature review," *Omega*, vol. 58, pp. 4–25, 2016.

[137] O. Garnett, A. Mandelbaum, and M. Reiman, "Designing a call center with impatient customers," *Manufacturing & service operations management*, vol. 4, no. 3, pp. 208–227, 2002.

[138] W. Whitt, "Efficiency-driven heavy-traffic approximations for many-server queues with abandonments," *Management science*, vol. 50, no. 10, pp. 1449–1461, 2004.

[139] S. Zeltyn and A. Mandelbaum, "Call centers with impatient customers: Many-server asymptotics of the m/m/n+ g queue," *Queueing systems*, vol. 51, no. 3, pp. 361–402, 2005.

[140] A. Mandelbaum and S. Zeltyn, "Staffing many-server queues with impatient customers: Constraint satisfaction in call centers," *Operations research*, vol. 57, no. 5, pp. 1189–1205, 2009.

[141] L. Laker, E. Torabi, D. France, C. Froehle, E. Goldlust, N. Hoot, P. Kasaie, M. Lyons, L. Barg-Walkow, M. Ward, and R. Wears, "Understanding emergency care delivery through computer simulation modeling," *Academic emergency medicine*, 2017.

[142] S. Zeltyn, Y. Marmor, A. Mandelbaum, B. Carmeli, O. Greenshpan, Y. Mesika, S. Wasserkrug, P. Vortman, A. Shtub, T. Lauterman, and D. Schwartz, "Simulation-based models of emergency departments: Operational, tactical, and strategic staffing," *Acm transactions on modeling and computer simulation (tomacs)*, vol. 21, no. 4, p. 24, 2011.

[143] X. Hu, S. Barnes, and B. Golden, "Applying queueing theory to the study of emergency department operations: A survey and a discussion of comparable simulation studies," *International transactions in operational research*, vol. 25, no. 1, pp. 7–49, 2018.

[144] W. Whitt, "What you should know about queueing models to set staffing requirements in service systems," *Naval research logistics*, vol. 54, pp. 476–484, 2007.

[145] ——, "Om forum-offered load analysis for staffing," *Manufacturing & service operations management*, vol. 15, no. 2, pp. 166–169, 2013.

[146] S. Halfin and W. Whitt, "Heavy-traffic limits for queues with many exponential servers," *Operations research*, vol. 29, no. 3, pp. 567–588, 1981.

[147] S. Borst, A. Mandelbaum, and M. Reiman, "Dimensioning large call centers," *Operations research*, vol. 52, no. 1, pp. 17–34, 2004.

[148] A. Janssen, J. Van Leeuwaarden, and B. Zwart, "Refining square-root safety staffing by expanding erlang c," *Operations research*, vol. 59, no. 6, pp. 1512–1522, 2011.

[149] B. Zhang, J. Van Leeuwaarden, and B. Zwart, "Staffing call centers with impatient customers: Refinements to many-server asymptotics," *Operations research*, vol. 60, no. 2, pp. 461–474, 2012.

[150] A. Mandelbaum and P. Momcilovic, "Queues with many servers and impatient customers," *Mathematics of operations research*, vol. 37, no. 1, pp. 41–65, 2012.

[151] W. Massey and W. Whitt, "An analysis of the modified offered-load approximation for the nonstationary erlang loss model," *The annals of applied probability*, pp. 1145–1160, 1994.

[152] O. Jennings, A. Mandelbaum, W. Massey, and W. Whitt, "Server staffing to meet time-varying demand," *Management science*, vol. 42, no. 10, pp. 1383–1394, 1996.

[153] Z. Feldman, A. Mandelbaum, M. W. A., and W. Whitt, "Staffing of time-varying queues to achieve time-stable performance," *Management science*, vol. 54, no. 2, 324?338, 2008.

[154] Y. Liu and W. Whitt, "Stabilizing customer abandonment in many-server queues with time-varying arrivals," *Operations research*, vol. 60, no. 6, pp. 1551–1564, 2012.

[155] ——, "Stabilizing performance in networks of queues with time-varying arrival rates," *Probability in the engineering and informational sciences*, vol. 28, no. 4, pp. 419–449, 2014.

[156] ——, "Stabilizing performance in a service system with time-varying arrivals and customer feedback," *European journal of operational research*, vol. 256, no. 2, pp. 473–486, 2017.

[157] L. V. Green and P. J. Kolesar, "The pointwise stationary approximation for queues with nonstationary arrivals," *Management science*, vol. 37, no. 1, pp. 84–97, 1991.

[158] L. Green, J. Soares, J. Giglio, and R. Green, "Using queueing theory to increase the effectiveness of emergency department provider staffing," *Academic emergency medicine*, vol. 13, no. 1, pp. 61–68, 2006.

[159] L. V. Green, P. J. Kolesar, and W. Whitt, "Coping with time-varying demand when setting staffing requirements for a service system," *Production and operations management*, vol. 16, no. 1, pp. 13–39, 2007.

[160] F. Vericourt and O. Jennings, "Nurse staffing in medical units: A queueing perspective," *Operations research*, vol. 59, no. 6, pp. 1320–1331, 2011.

[161] S. Eick, W. Massey, and W. Whitt, "The physics of the mt/g/∞ queue," *Operations research*, vol. 41, no. 4, pp. 731–742, 1993.

[162] W. Massey and W. Whitt, "Networks of infinite-server queues with nonstationary poisson input," *Queueing systems*, vol. 13, no. 1-3, pp. 183–250, 1993.

[163] J. Tsitsiklis and K. Xu, "On the power of (even a little) resource pooling," *Stochastic systems*, vol. 2, no. 1, pp. 1–66, 2012.

[164] Z. Aksin, M. Armony, and V. Mehrotra, "The modern call center: A multi-disciplinary perspective on operations management research," *Production and operations management*, vol. 16, no. 6, pp. 665–688, 2007.

[165] M. Armony and A. Mandelbaum, "Routing and staffing in large-scale service systems: The case of homogeneous impatient customers and heterogeneous servers," *Operations research*, vol. 59, no. 1, pp. 50–65, 2011.

[166] M. McCarthy, S. Zeger, R. Ding, D. Aronsky, N. Hoot, and G. Kelen, "The challenge of predicting demand for emergency department services," *Academic emergency medicine*, vol. 15, no. 4, pp. 337–346, 2008.

[167] S Maman, *Uncertainty in the demand for service: The case of call centers and emergency departments*, Doctoral dissertation, Technion-Israel Institute of Technology, Faculty of Industrial and Management Engineering), 2009.

[168] G. Allon, S. Deo, and W. Lin, "The impact of size and occupancy of hospital on the extent of ambulance diversion: Theory and evidence," *Operations research*, vol. 61, no. 3, pp. 544–562, 2013.

[169] P. Shi, M. Chou, J. Dai, D. Ding, and J. Sim, "Models and insights for hospital inpatient operations: Time-dependent ed boarding time," *Management science*, vol. 62, no. 1, pp. 1–28, 2015.

[170] M. Armony, S. Israelit, A. Mandelbaum, Y. Marmor, Y. Tseytlin, and G. Yom-Tov, "On patient flow in hospitals: A data-based queueing-science perspective," *Stochastic systems*, vol. 5, no. 1, pp. 146–194, 2015.

[171] S. Saghafian, W. Hopp, M. Van Oyen, J. Desmond, and S. Kronick, "Patient streaming as a mechanism for improving responsiveness in emergency departments," *Operations research*, vol. 60, no. 5, pp. 1080–1097, 2012.

[172] D. Claudio, M. Velzquez, W. Bravo-Llerena, G. Okudan, and A. Freivalds, "Perceived usefulness and ease of use of wearable sensor-based systems in emergency departments," *Iie transactions on occupational ergonomics and human factors*, vol. 3, no. 3-4, pp. 177–187, 2015.

[173] S. Levin, D. Aronsky, R. Hemphill, J. Han, J. Slagle, and D. France, "Shifting toward balance: Measuring the distribution of workload among emergency physician teams," *Annals of emergency medicine*, vol. 50, no. 4, pp. 419–423, 2007.

[174] S. Levin, D. France, R. Hemphill, I. Jones, K. Chen, D. Rickard, R. Makowski, and D. Aronsky, "Tracking workload in the emergency department," *Human factors*, vol. 48, no. 3, pp. 526–539, 2006.

[175] P. Chevalier and J. Schrieck, "Optimizing the staffing and routing of small?size hierarchical call centers," *Production and operations management*, vol. 17, no. 3, pp. 306–319, 2008.

[176] O. Akşin, N. Çakan, F. Karaesmen, and E. Ormeci, "Flexibility structure and capacity design with human resource considerations," *Production and operations management*, vol. 24, no. 7, pp. 1086–1100, 2015.

[177] N. Argon and S. Ziya, "Priority assignment under imperfect information on customer type identities," *Manufacturing & service operations management*, vol. 11, no. 4, pp. 674–693, 2009.

[178] D. Tan and S. Horning, "Follicular lymphoma: Clinical features and treatment," *Hematology/oncology clinics of north america*, vol. 22, no. 5, pp. 863–882, 2008.

[179] L. R. Teras, C. E. DeSantis, J. R. Cerhan, L. M. Morton, A. Jemal, and C. R. Flowers, "2016 us lymphoid malignancy statistics by world health organization subtypes," *Ca cancer j clin*, 2016.

[180] C. Casulo, L. Nastoupil, N. Fowler, J. Friedberg, and C. Flowers, "Unmet needs in the first-line treatment of follicular lymphoma," *Annals of oncology*, p. mdx189, 2017.

[181] C. Casulo, M. Byrtek, K. Dawson, X. Zhou, C. Farber, C. Flowers, J. Hainsworth, M. Maurer, J. Cerhan, B. Link, and A. Zelenetz, "Early relapse of follicular lymphoma after rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone defines patients at high risk for death: An analysis from the national lymphocare study," *Journal of clinical oncology*, vol. 33, no. 23, pp. 2516–2522, 2015.

[182] M. Rummel, N. Niederle, G. Maschmeyer, G. Banat, U. von Grnhagen, C. Losem, D. Kofahl-Krause, G. Heil, M. Welslau, C. Balser, and U. Kaiser, "Bendamustine plus rituximab versus chop plus rituximab as first-line treatment for patients

with indolent and mantle-cell lymphomas: An open-label, multicentre, randomised, phase 3 non-inferiority trial.," *The lancet*, vol. 381, no. 9873, pp. 1203–1210, 2013.

[183] O. Press, J. Unger, L. Rimsza, J. Friedberg, M. LeBlanc, M. Czuczman, M. Kaminski, R. Braziel, C. Spier, A. Gopal, and D. Maloney, "A comparative analysis of prognostic factor models for follicular lymphoma based on a phase iii trial of chop-rituximab vs chop+ 131iodine-tositumomab," *Clinical cancer research*, clincanres–1120, 2013.

[184] G. Salles, J. Seymour, F. Offner, A. Lpez-Guillermo, D. Belada, L. Xerri, P. Feugier, R. Bouabdallah, J. Catalano, P. Brice, and D. Caballero, "Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (prima): A phase 3, randomised controlled trial," *The lancet*, vol. 377, no. 9759, pp. 42–51, 2011.

[185] M. Federico, M. Bellei, L. Marcheselli, S. Luminari, A. Lopez-Guillermo, U. Vitolo, B. Pro, S. Pileri, A. Pulsoni, and S. Soubeyran P. andCortelazzo, "Follicular lymphoma international prognostic index 2: A new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project," *Journal of clinical oncology*, vol. 27, no. 27, pp. 4555–4562, 2009.

[186] P. Solal-Celigny, P. Roy, P. Colombat, J. White, J. Armitage, R. Arranz-Saez, W. Au, M. Bellei, P. Brice, D. Caballero, and B. Coiffier, "Follicular lymphoma international prognostic index," *Blood*, 2004.

[187] O. Aalen and S. Johansen, "An empirical transition matrix for non-homogeneous markov chains based on censored observations," *Scandinavian journal of statistics*, pp. 141–150, 1978.

[188] E. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the american statisticial association*, vol. 53, no. 282, pp. 457–481, 1958.

[189] P. Andersen, O. Borgan, R. Gill, and N. Keiding, *Statistical methods based on counting processes*. New York, NY: Springer-Verlag., 1993.

[190] Borgan, "Aalen-johansen estimator," in *Encyclopedia of biostatistics*. Published online: 15 July 2005 http://dx.doi.org/10.1002/0470011815.b2a11001 ed.: John Wiley and Sons, Inc., 2005.

[191] S. Datta and G. Satten, "Validity of the aalen-johansen estimators of stage occupation probabilities and nelson-aalen estimators of integrated transition hazards for non-markov models," *Statistics and probability letters*, vol. 55, no. 4, pp. 403–411, 2001.

[192] World-Health-Organization, *International classification of diseases for oncology*, Third Edition, First Revision. Geneva: World Health Organization (WHO), 2013.

[193] J. Warren, C. Klabunde, D. Schrag, P. Bach, and G. Riley, "Overview of the seer-medicare data: Content, research applications, and generalizability to the united states elderly population," *Medical care*, vol. 40, no. 8, pp. IV–3, 2002.

[194] J. Friedberg, M. Taylor, J. Cerhan, C. Flowers, H. Dillon, C. Farber, E. Rogers, J. Hainsworth, E. Wong, J. Vose, and A. Zelenetz, "Follicular lymphoma in the united states: First report of the national lymphocare study," *Journal of clinical oncology*, vol. 27, no. 8, pp. 1202–1208, 2009.

[195] L. Arcaini, M. Merli, F. Passamonti, S. Rizzi, V. Ferretti, S. Rattotti, C. Pascutto, M. Paulli, and M. Lazzarino, "Validation of follicular lymphoma international prognostic index 2 (flipi2) score in an independent series of follicular lymphoma patients," *British journal of haematology*, vol. 149, no. 3, pp. 455–457, 2010.

[196] M. Meignan, A. Cottereau, A. Versari, L. Chartier, J. Dupuis, S. Boussetta, I. Grassi, R. Casasnovas, C. Haioun, H. Tilly, and V. Tarantino, "Baseline metabolic tumor volume predicts outcome in high-tumor-burden follicular lymphoma: A pooled analysis of three multicenter studies," *Journal of clinical oncology*, vol. 34, no. 30, pp. 3618–3626, 2016.

[197] A. Pastore, V. Jurinovic, R. Kridel, E. Hoster, A. Staiger, M. Szczepanowski, C. Pott, N. Kopp, M. Murakami, H. Horn, and E. Leich, "Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: A retrospective analysis of a prospective clinical trial and validation in a population-based registry," *The lancet oncology*, vol. 16, no. 9, pp. 1111–1122, 2015.

[198] C. Davis, J. Cohen, K. Shah, D. Hutcherson, M. Surati, K. Valla, E. Panjic, C. Handler, J. Switchenko, and C. Flowers, "Efficacy and tolerability of anthracycline-based therapy in elderly patients with diffuse large b-cell lymphoma," *Clinical lymphoma myeloma and leukemia*, vol. 15, no. 5, pp. 270–277, 2015.

[199] H. Muss, D. Berry, C. Cirrincione, D. Budman, I. Henderson, M. Citron, L. Norton, E. Winer, and H. C.A., "Toxicity of older and younger patients treated with adjuvant chemotherapy for node-positive breast cancer: The cancer and leukemia group b experience," *Journal of clinical oncology*, vol. 25, no. 24, pp. 3699–3704, 2007.

[200] E. Kuczynski, D. Sargent, A. Grothey, and R. Kerbel, "Drug rechallenge and treatment beyond progression [mdash] implications for drug resistance," *Nature reviews clinical oncology*, vol. 10, no. 10, pp. 571–587, 2013.

[201] C. Palmieri, J. Krell, C. James, C. Harper-Wynne, V. Misra, S. Cleator, and D. Miles, "Rechallenging with anthracyclines and taxanes in metastatic breast cancer.," *Nature reviews clinical oncology*, vol. 7, no. 10, pp. 561–574, 2010.

[202] J. Friedberg, P. Cohen, L. Chen, K. Robinson, A. Forero-Torres, A. La Casce, L. Fayad, A. Bessudo, E. Camacho, M. Williams, and R. Van der Jagt, "Bendamustine in patients with rituximab-refractory indolent and transformed non-hodgkin's lymphoma: Results from a phase ii multicenter, single-agent study," *Journal of clinical oncology*, vol. 26, no. 2, pp. 204–210, 2008.

[203] L. Sehn, N. Chua, J. Mayer, G. Dueck, M. Trneny, K. Bouabdallah, N. Fowler, V. Delwail, O. Press, G. Salles, and J. Gribben, "Obinutuzumab plus bendamustine versus bendamustine monotherapy in patients with rituximab-refractory indolent non-hodgkin lymphoma (gadolin): A randomised, controlled, open-label, multicentre, phase 3 trial," *The lancet oncology*, vol. 17, no. 8, pp. 1081–1093, 2016.

[204] J. Leonard, S. Jung, J. Johnson, B. Pitcher, N. Bartlett, K. Blum, M. Czuczman, J. Giguere, and B. Cheson, "Randomized trial of lenalidomide alone versus lenalidomide plus rituximab in patients with recurrent follicular lymphoma: Calgb 50401 (alliance)," *Journal of clinical oncology*, vol. 33, no. 31, pp. 3635–3640, 2015.

[205] M. Wang, N. Fowler, W.-B. N., L. Feng, J. Romaguera, S. Neelapu, F. Hagemeister, M. Fanale, Y. Oki, B. Pro, and J. Shah, "Oral lenalidomide with rituximab in relapsed or refractory diffuse large cell, follicular and transformed lymphoma: A phase ii clinical trial," *Leukemia*, vol. 27, no. 9, pp. 1902–1909, 2013.

[206] C. Ujjani, S. Jung, B. Pitcher, P. Martin, S. Park, K. Blum, S. Smith, M. Czuczman, M. Davids, E. Levine, and L. L.D., "Phase 1 trial of rituximab, lenalidomide, and ibrutinib in previously untreated follicular lymphoma: Alliance a051103," *Blood*, vol. 128, no. 21, pp. 2510–2516, 2016.

[207] A. Gopal, B. Kahl, S. de Vos, N. Wagner-Johnston, S. Schuster, W. Jurczak, I. Flinn, C. Flowers, P. Martin, A. Viardot, and K. Blum, "Pi3kdelta inhibition by idelalisib in patients with relapsed indolent lymphoma," *New england journal of medicine*, vol. 370, no. 11, pp. 1008–1018, 2014.

[208] A. Gopal, B. Kahl, C. Flowers, P. Martin, S. Ansell, E. Abella-Dominicis, B. Koh, W. Ye, P. Barr, G. Salles, and J. Friedberg, "Idelalisib is effective in patients with high-risk follicular lymphoma and early relapse after initial chemoimmunotherapy," *Blood*, vol. 129, no. 22, pp. 3037–3039, 2017.

[209] G. Salles, S. Schuster, S. de Vos, N. Wagner-Johnston, A. Viardot, K. Blum, C. Flowers, W. Jurczak, I. Flinn, B. Kahl, and M. P., "Efficacy and safety of idelalisib in patients with relapsed, rituximab-and alkylating agent-refractory follicular lym-

phoma: A subgroup analysis of a phase 2 study," *Haematologica*, vol. 102, no. 4, e156–e159, 2017.

[210] P. Brice, Y. Bastion, E. Lepage, N. Brousse, C. Haioun, P. Moreau, N. Straetmans, H. Tilly, I. Tabah, and P. Solal-Celigny, "Comparison in low-tumor-burden follicular lymphomas between an initial no-treatment policy, prednimustine, or interferon alfa: A randomized study from the groupe d'etude des lymphomes folliculaires," *Journal of clinical oncology*, vol. 15, no. 3, pp. 1110–1117, 1997.

[211] M. Wieduwilt, L. Tao, C. Clarke, T. Keegan, S. Glaser, S. Gomez, and M. Martinez, "Impact of marital status on the survival of patients with hematologic malignancies reported to the california cancer registry," *Journal of clinical oncology*, vol. 34, no. 15, p. 6555, 2016.

[212] P. K. Andersen and N. Keiding, "Multi-state models for event history analysis," *Stat methods med res*, vol. 11, no. 2, pp. 91–115, 2002.

[213] P. K. Andersen, R. B. Geskus, T. de Witte, and H. Putter, "Competing risks in epidemiology: Possibilities and pitfalls," *Int j epidemiol*, vol. 41, no. 3, pp. 861–70, 2012.

[214] P. C. Austin, D. S. Lee, and J. P. Fine, "Introduction to the analysis of survival data in the presence of competing risks," *Circulation*, vol. 133, no. 6, pp. 601–9, 2016.

[215] T. A. Gooley, W. Leisenring, J. Crowley, and B. E. Storer, "Estimation of failure probabilities in the presence of competing risks: New representations of old estimators," *Stat med*, vol. 18, no. 6, pp. 695–706, 1999.

[216] S. Basu and R. Tiwari, "Breast cancer survival, competing risks and mixture cure model: A bayesian analysis," *Journal of the royal statistical society: Series a (statistics in society)*, vol. 173, no. 2, pp. 307–329, 2010.

[217] I. Flinn, R. van der Jagt, B. Kahl, P. Wood, T. Hawkins, D. MacDonald, M. Hertzberg, Y. Kwan, D. Simpson, M. Craig, and K. Kolibaba, "Open-label, randomized, non-inferiority study of bendamustine-rituximab or r-chop/r-cvp in first-line treatment of advanced indolent nhl or mcl: The bright study," *Blood*, pp.blood–2013, 2014.

[218] Q. Shi, C. Flowers, W. Hiddemann, R. Marcus, M. Herold, A. Hagenbeek, E. Kimby, H. Hochster, U. Vitolo, B. Peterson, and E. Gyan, "Thirty-month complete response as a surrogate end point in first-line follicular lymphoma therapy: An individual patient-level analysis of multiple randomized trials," *Journal of clinical oncology*, vol. 35, no. 4, pp. 552–560, 2016.

[219] L. R. Teras, C. E. DeSantis, J. R. Cerhan, L. M. Morton, A. Jemal, and C. R. Flowers, "2016 us lymphoid malignancy statistics by world health organization subtypes," *Ca cancer j clin*, 2016.

[220] C. R. Flowers, R. Sinha, and J. M. Vose, "Improving outcomes for patients with diffuse large b-cell lymphoma," *Ca cancer j clin*, vol. 60, no. 6, pp. 393–408, 2010.

[221] B. Coiffier, E. Lepage, J. Briere, R. Herbrecht, H. Tilly, R. Bouabdallah, P. Morel, E. Van Den Neste, G. Salles, P. Gaulard, F. Reyes, P. Lederlin, and C. Gisselbrecht, "Chop chemotherapy plus rituximab compared with chop alone in elderly patients with diffuse large-b-cell lymphoma," *N engl j med*, vol. 346, no. 4, pp. 235–42, 2002.

[222] M. Pfreundschuh, L. Trumper, A. Osterborg, R. Pettengell, M. Trneny, K. Imrie, D. Ma, D. Gill, J. Walewski, P. L. Zinzani, R. Stahel, S. Kvaloy, O. Shpilberg, U. Jaeger, M. Hansen, T. Lehtinen, A. Lopez-Guillermo, C. Corrado, A. Scheliga, N. Milpied, M. Mendila, M. Rashford, E. Kuhnt, and M. Loeffler, "Chop-like chemotherapy plus rituximab versus chop-like chemotherapy alone in young patients with good-prognosis diffuse large-b-cell lymphoma: A randomised controlled trial by the mabthera international trial (mint) group," *Lancet oncol*, vol. 7, no. 5, pp. 379–91, 2006.

[223] T. M. Habermann, E. A. Weller, V. A. Morrison, R. D. Gascoyne, P. A. Cassileth, J. B. Cohn, S. R. Dakhil, B. Woda, R. I. Fisher, B. A. Peterson, and S. J. Horning, "Rituximab-chop versus chop alone or with maintenance rituximab in older patients with diffuse large b-cell lymphoma," *J clin oncol*, vol. 24, no. 19, pp. 3121–7, 2006.

[224] L. H. Sehn, J. Donaldson, M. Chhanabhai, C. Fitzgerald, K. Gill, R. Klasa, N. MacPherson, S. O'Reilly, J. J. Spinelli, J. Sutherland, K. S. Wilson, R. D. Gascoyne, and J. M. Connors, "Introduction of combined chop plus rituximab therapy dramatically improved outcome of diffuse large b-cell lymphoma in british columbia," *J clin oncol*, vol. 23, no. 22, pp. 5027–33, 2005.

[225] M. J. Maurer, H. Ghesquieres, J. P. Jais, T. E. Witzig, C. Haioun, C. A. Thompson, R. Delarue, I. N. Micallef, F. Peyrade, W. R. Macon, T. Jo Molina, N. Ketterer, S. I. Syrbu, O. Fitoussi, P. J. Kurtin, C. Allmer, E. Nicolas-Virelizier, S. L. Slager, T. M. Habermann, B. K. Link, G. Salles, H. Tilly, and J. R. Cerhan, "Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large b-cell lymphoma treated with immunochemotherapy," *J clin oncol*, vol. 32, no. 10, pp. 1066–73, 2014.

[226] J. F. Larouche, F. Berger, C. Chassagne-Clement, M. Ffrench, E. Callet-Bauchu, C. Sebban, H. Ghesquieres, F. Broussais-Guillaumot, G. Salles, and B. Coiffier, "Lymphoma recurrence 5 years or later following diffuse large b-cell lymphoma:

Clinical characteristics and outcome," *J clin oncol*, vol. 28, no. 12, pp. 2094–100, 2010.

[227] J. W. Friedberg and R. I. Fisher, "Diffuse large b-cell lymphoma," *Hematol oncol clin north am*, vol. 22, no. 5, pp. 941–52, ix, 2008.

[228] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson J., L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, "Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–11, 2000.

[229] N. Howlader, A. B. Mariotto, C. Besson, G. Suneja, K. Robien, N. Younes, and E. A. Engels, "Cancer-specific mortality, cure fraction, and noncancer causes of death among diffuse large b-cell lymphoma patients in the immunochemotherapy era," *Cancer*, vol. 123, no. 17, pp. 3326–3334, 2017.

[230] X. Han, A. Jemal, C. R. Flowers, H. Sineshaw, L. J. Nastoupil, and E. Ward, "Insurance status is related to diffuse large b-cell lymphoma survival," *Cancer*, vol. 120, no. 8, pp. 1220–7, 2014.

[231] Q. Chen, T. Ayer, L. J. Nastoupil, J. L. Koff, A. D. Staton, J. Chhatwal, and C. R. Flowers, "Population-specific prognostic models are needed to stratify outcomes for african-americans with diffuse large b-cell lymphoma," *Leuk lymphoma*, vol. 57, no. 4, pp. 842–51, 2016.

[232] J. B. Cohen, D. M. Kurtz, A. D. Staton, and C. R. Flowers, "Next-generation surveillance strategies for patients with lymphoma," *Future oncol*, vol. 11, no. 13, pp. 1977–91, 2015.

[233] N. Howlader, L. Morton, E. Feuer, C. Besson, and E. Engels, "Contributions of subtypes of non-hodgkin lymphoma to mortality trends," *Cancer epidemiology and prevention biomarkers*, vol. 25, no. 1, pp. 174–179, 2016.

[234] J. L. Koff and C. R. Flowers, "Prognostic modeling in diffuse large b-cell lymphoma in the era of immunochemotherapy: Where do we go from here?" *Cancer*, vol. 123, no. 17, pp. 3222–3225, 2017.

[235] F. Campigotto, D. Neuberg, and J. Zwicker, "Biased estimation of thrombosis rates in cancer studies using the method of kaplan and meier," *Journal of thrombosis and haemostasis*, vol. 10, no. 7, pp. 1449–1451, 2012.

[236] M. Shipp, "International non-hodgkin's lymphoma prognostic factors, project: A predictive model for aggressive non-hodgkin's lymphoma," *N engl j med*, vol. 329, no. 14, pp. 987–94, 1993.

[237] P. Fields and D. Linch, "Treatment of the elderly patient with diffuse large b cell lymphoma," *British journal of haematology*, vol. 157, no. 2, pp. 159–170, 2012.

[238] L. H. Sehn, B. Berry, M. Chhanabhai, C. Fitzgerald, K. Gill, P. Hoskins, R. Klasa, K. J. Savage, T. Shenkier, J. Sutherland, R. D. Gascoyne, and J. M. Connors, "The revised international prognostic index (r-ipi) is a better predictor of outcome than the standard ipi for patients with diffuse large b-cell lymphoma treated with r-chop," *Blood*, vol. 109, no. 5, pp. 1857–61, 2007.

[239] O. Aalen, O. Borgan, and H. Gjessing, *Survival and event history analysis: A process point of view*. New York, NY: Springer Science and Business Media, 2008.

[240] D. Cox, "Regression models and life-tables.," *Journal of the royal statistical society, b*, vol. 34, no. 2, pp. 187–220, 1972.

[241] ——, "Partial likelihood," *Biometrika*, pp. 269–276, 1975.

[242] V. Morrison, P. Hamlin, P. Soubeyran, R. Stauder, P. Wadhwa, M. Aapro, and S. Lichtman, "Approach to therapy of diffuse large b-cell lymphoma in the elderly: The international society of geriatric oncology (siog) expert position commentary.," *Annals of oncology*, vol. 26, no. 6, pp. 1058–1068, 2015.

[243] V. Morrison, P. Hamlin, P. Soubeyran, R. Stauder, P. Wadhwa, M. Aapro, and S. Lichtman, "Diffuse large b-cell lymphoma in the elderly: Impact of prognosis, comorbidities, geriatric assessment, and supportive care on clinical practice," *An international society of geriatric oncology (siog) expert position paper. journal of geriatric oncology*, vol. 6, no. 2, pp. 141–152, 2015.

[244] S. Chaganti, T. Illidge, S. Barrington, P. Mckay, K. Linton, K. Cwynarski, A. McMillan, A. Davies, S. Stern, and K. Peggs, "Guidelines for the management of diffuse large b?cell lymphoma," *British journal of haematology*, vol. 174, no. 1, pp. 43–56, 2016.

[245] P. Feugier, A. Van Hoof, C. Sebban, P. Solal-Celigny, R. Bouabdallah, C. Ferme, B. Christian, E. Lepage, H. Tilly, F. Morschhauser, and P. Gaulard, "Long-term results of the r-chop study in the treatment of elderly patients with diffuse large b-cell lymphoma: A study by the groupe d?etude des lymphomes de l?adulte," *J clin oncol*, vol. 23, no. 18, pp. 4117–4126, 2005.

[246] M. Pfreundschuh, J. Schubert, M. Ziepert, R. Schmits, M. Mohren, E. Lengfelder, M. Reiser, C. Nickenig, M. Clemens, N. Peter, and C. Bokemeyer, "Six versus

eight cycles of bi-weekly chop-14 with or without rituximab in elderly patients with aggressive cd20+ b-cell lymphomas: A randomised controlled trial (ricover-60)," *The lancet oncology*, vol. 9, no. 2, pp. 105–116, 2008.

[247] P. Hamlin, S. Satram-Hoang, C. Reyes, K. Hoang, S. Guduru, and S. Skettino, "Treatment patterns and comparative effectiveness in elderly diffuse large b-cell lymphoma patients: A surveillance, epidemiology, and end results-medicare analysis," *The oncologist*, vol. 19, no. 12, pp. 1249–1257, 2014.

[248] F. Peyrade, S. Bologna, V. Delwail, J. Emile, L. Pascal, C. Ferm, J. Schiano, B. Coiffier, B. Corront, H. Farhat, and C. Fruchart, "Combination of ofatumumab and reduced-dose chop for diffuse large b-cell lymphomas in patients aged 80 years or older: An open-label, multicentre, single-arm, phase 2 trial from the lysa group," *The lancet haematology,*, vol. 4, no. 1, e46–e55, 2017.

[249] F. Merli, S. Luminari, G. Rossi, C. Mammi, L. Marcheselli, A. Ferrari, M. Spina, A. Tucci, C. Stelitano, I. Capodanno, and A. Fragasso, "Outcome of frail elderly patients with diffuse large b-cell lymphoma prospectively identified by comprehensive geriatric assessment: Results from a study of the fondazione italiana linfomi," *Leukemia & lymphoma*, vol. 55, no. 1, pp. 38–43, 2014.

[250] A. Wieringa, K. Boslooper, M. Hoogendoorn, P. Joosten, T. Beerden, H. Storm, R. Kibbelaar, G. Veldhuis, H. van Kamp, B. van Rees, and H. Kluin?Nelemans, "Comorbidity is an independent prognostic factor in patients with advanced?stage diffuse large b?cell lymphoma treated with r?chop: A population?based cohort study," *British journal of haematology*, vol. 165, no. 4, pp. 489–496, 2014.

[251] Y. Bastion, J. Blay, M. Divine, P. Brice, D. Bordessoule, C. Sebban, M. Blanc, H. Tilly, P. Lederlin, E. Deconinck, and B. Salles, "Elderly patients with aggressive non-hodgkin's lymphoma: Disease presentation, response to treatment, and survival–a groupe d'etude des lymphomes de l'adulte study on 453 patients older than 69 years," *Journal of clinical oncology*, vol. 15, no. 8, pp. 2945–2953, 1997.

[252] A. Tucci, M. Martelli, L. Rigacci, P. Riccomagno, M. Cabras, F. Salvi, C. Stelitano, A. Fabbri, S. Storti, S. Fogazzi, and S. Mancuso, "Comprehensive geriatric assessment is an essential tool to support treatment decisions in elderly patients with diffuse large b-cell lymphoma: A prospective multicenter evaluation in 173 patients by the lymphoma italian foundation (fil)," *Leukemia & lymphoma*, vol. 56, no. 4, pp. 921–926, 2015.

[253] M. Yoshida, T. Nakao, M. Horiuchi, H. Ueda, K. Hagihara, H. Kanashima, T. Inoue, E. Sakamoto, M. Hirai, H. Koh, and T. Nakane, "Analysis of elderly patients with diffuse large b?cell lymphoma: Aggressive therapy is a reasonable approach for 'unfit' patients classified by comprehensive geriatric assessment," *European journal of haematology*, vol. 96, no. 4, pp. 409–416, 2016.

[254] J. Williams, A. Rai, J. Lipscomb, J. Koff, L. Nastoupil, and C. Flowers, "Disease characteristics, patterns of care, and survival in very elderly patients with diffuse large b-cell lymphoma," *Cancer*, vol. 121, no. 11, pp. 1800–1808, 2015.

[255] Y. Tien, B. Link, J. Brooks, K. Wright, and E. Chrischilles, "Treatment of diffuse large b-cell lymphoma in the elderly: Regimens without anthracyclines are common and not futile," *Leukemia & lymphoma*, vol. 56, no. 1, pp. 65–71, 2015.

[256] R. Lin, M. Behera, C. Diefenbach, and C. Flowers, "Role of anthracycline and comprehensive geriatric assessment for elderly patients with diffuse large b-cell lymphoma," *Blood*, vol. 130, no. 20, pp. 2180–2185, 2017.

[257] P. Shenoy, N. Malik, A. Nooka, R. Sinha, K. Ward, O. Brawley, J. Lipscomb, and C. Flowers, "Racial differences in the presentation and outcomes of diffuse large b?cell lymphoma in the united states," *Cancer*, vol. 117, no. 11, pp. 2530–2540, 2011.

[258] C. Flowers, P. Shenoy, U. Borate, K. Bumpers, T. Douglas-Holland, N. King, O. Brawley, J. Lipscomb, M. Lechowicz, R. Sinha, and R. Grover, "Examining racial differences in diffuse large b-cell lymphoma presentation and survival," *Leukemia & lymphoma*, vol. 54, no. 2, pp. 268–276, 2013.

[259] J. B. Cohen, M. Behera, C. A. Thompson, and C. R. Flowers, "Evaluating surveillance imaging for diffuse large b-cell lymphoma and hodgkin lymphoma," *Blood*, vol. 129, no. 5, pp. 561–564, 2017.

[260] C. Flowers, S. Fedewa, A. Chen, L. Nastoupil, J. Lipscomb, O. Brawley, and E. Ward, "Disparities in the early adoption of chemoimmunotherapy for diffuse large b-cell lymphoma in the united states," *Cancer epidemiology and prevention biomarkers*, vol. 21, no. 9, pp. 1520–1530, 2012.

[261] Q. Shi, N. Schmitz, F. Ou, J. Dixon, D. Cunningham, M. Pfreundschuh, J. Seymour, U. Jaeger, T. Habermann, C. Haioun, and H. Tilly, "Progression-free survival as a surrogate end point for overall survival in first-line diffuse large b-cell lymphoma: An individual patient?level analysis of multiple randomized trials (seal)," *Journal of clinical oncology*, vol. 36, no. 25, pp. 2593–2602, 2018.

[262] A. Suzukawa, "Asymptotic properties of aalen-johansen integrals for competing risks data," *Journal of the japan statistical society*, vol. 32, no. 1, pp. 77–93, 2002.

[263] O. Aalen, O. Borgan, and H. Gjessing, *Survival and event history analysis: A process point of view*. New York, NY: Springer Science and Business Media, 2008.

[264] F. Campigotto, D. Neuberg, and J. Zwicker, "Biased estimation of thrombosis rates in cancer studies using the method of kaplan and meier," *Journal of thrombosis and haemostasis*, vol. 10, no. 7, pp. 1449–1451, 2012.

[265] R. B. Cooper, *Introduction to queueing theory*. 2nd Edition, Elsevier North-Holland, New York, 1981.

[266] D. Evans and L. Leemis, "Algorithms for computing the distributions of sums of discrete random variables," *Mathematical and computer modelling*, vol. 40, no. 13, pp. 1429–1452, 2004.

[267] M. D. Ward, *Probability, stat/ma 416, fall 2014 course notes, purdue university*, 2014.

[268] W. Feller, *An introduction to probability theory and its applications (volume 2) second edition. chapter 8 basic limit theorems*. John Wiley & Sons Incorporated, New York, 1971.

[269] ——, *An introduction to probability theory and its applications: Volume 1. chapter 2.9 elements of combinatorial analysis - stirling's formula*. John Wiley & Sons Incorporated, New York, 1968.

VITA

Çağlar Çağlayan was born in Istanbul, Turkey. He is a researcher, working on topics of clinical decision-making under uncertainty. His research mainly focuses on disease management and health service operations problems such as breast cancer screening, physician staffing in emergency departments, and lymphoma treatment. He utilizes a variety of analytical methods for his research including deterministic and stochastic optimization, queuing theory, discrete event simulation, survival analysis, and machine learning. He received his Bachelor of Science degree in Industrial Engineering from Boğaziçi University in 2011. He received a Master of Industrial Engineering degree from North Carolina State University in 2013, and a Master of Science degree in Operations Research from Georgia Institute of Technology in 2015. He plans to receive his doctoral degree in Operations Research from Georgia Institute of Technology in 2019.

Caglar Caglayan - August 2013