ABSTRACT

TEJADA, JEREMY JOHN. An Integrated Discrete-Event/System Dynamics Simulation Model of Breast Cancer Screening for Older US Women. (Under the direction of Julie S. Ivy and James R. Wilson).

The objective of this research is to develop, validate, and exploit a simulation modeling framework for evaluating the effectiveness of breast cancer screening policies in the near future (that is, over the period 2012–2020) for US women who are at least 65 years old. This work includes an examination of key components in the breast cancer screening process for older women, and an approach to defining and modeling those components using simulation. In the near future, it is expected that half of newly diagnosed breast cancer cases will be in women 65 and older. This development, along with the aging US population, is evidence that older women will become the prevalent patient cohort in the breast cancer population of the United States. This research utilizes a two-phase simulation modeling approach. The first simulation is a natural history model of breast cancer incidence and progression in randomly sampled individuals from the designated population of older US women. The second simulation is an integrated screening-and-treatment model that uses knowledge about the genesis of breast cancer within the same population gained from the natural history model to estimate the benefits of different policies for screening the designated population and treating the relevant individuals. Both simulation models are composed of interacting submodels that represent key aspects of the incidence, progression, screening, treatment, survival, and cost of breast cancer in the population of older US women as well as the overall structure of the system for detecting and treating this disease. We discuss the rationale for combining the discrete-event and system-dynamics modeling techniques for the analysis of this problem, with an underlying goal of identifying the benefit of using this integrated approach. Our methodology is “individualized” in the sense that we simulate the lives of individual women who are representative of the designated population, and each woman’s risk of being diagnosed with breast cancer is based on her individual risk factors. Other problem areas are explored in this research, including the development of techniques for input modeling, general systems modeling, and output analysis that are

specifically adapted to address the special needs of simulation-based health care decision making.

An Integrated Discrete-Event/System Dynamics Simulation Model of Breast Cancer Screening for Older US Women

by
Jeremy John Tejada

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina

2012

APPROVED BY:

_____	_____
Dr. Julie Ivy Committee Co-Chair	Dr. James R. Wilson Committee Co-Chair
_____	_____
Dr. Russell E. King	Dr. Peter Bloomfield

DEDICATION

To my family, friends, professors and staff who helped me achieve my goal of completing my Ph.D.

BIOGRAPHY

Jeremy Tejada is a Ph.D. student in the Edward P. Fitts Department of Industrial and Systems Engineering at North Carolina State University. He was born on February 24, 1987, in Tucson, AZ, and raised in Garland, TX. He is the son of John Tejada and Andrea Tejada, and brother of Kristen Tejada. In May 2008, he graduated from Texas Tech University with a Bachelor of Science degree in industrial engineering with a minor in mathematics. His undergraduate academic achievements include memberships in Alpha Pi Mu and Tau Beta Pi National Honor Societies. In August 2008, he came to North Carolina State University to pursue graduate studies. He received his Master of Industrial Engineering degree in December 2009. His research interests include simulation modeling, health care applications, breast cancer, statistics, and operations research. His hobbies include playing rugby, watching sports, cooking, and eating. While working on his Ph.D., he served as a research assistant, and as a teaching assistant for graduate and undergraduate simulation and production classes.

ACKNOWLEDGEMENTS

First and foremost, I would like thank my advisor, Dr. James R. Wilson, for his insightful perspective, his ability to work with my personality, his willingness to go out of his way to help me prepare for my qualifying exam, and his guidance throughout my entire time at NC State. Special thanks go to Dr. Julie Ivy for serving as my committee co-chair, developing my interest in breast cancer research, and the advice she has given me throughout my graduate studies. I thank Dr. Bonnie Yankaskas and Dr. Kathleen Diehl for providing their guidance on breast cancer related matters. I thank Dr. Russell King for recruiting me to come to NC State, serving on my committee, and helping me maintain a long-term perspective. I thank Dr. Peter Bloomfield for serving on my committee, and for his guidance in statistics. I thank Dr. Thom Hodgson for his input on teaching, and his general guidance throughout my graduate studies. Finally, I especially thank my parents and my sister for their support during both my undergraduate and graduate studies. Their advice and patience was a great help while I was working on this dissertation. I would also like to thank my friends who kept me smiling throughout my time in graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Problem Statement and Research Objectives

The objective of this research is to develop, validate, and exploit a simulation modeling framework for evaluating the effectiveness of breast cancer screening policies in the near future (that is, over the period 2012-2020) for US women who are at least 65 years old in the near future, that is, over the period 2012–2020. This work includes an examination of key components in the breast cancer screening process for older women, and an approach to defining and modeling those components using simulation. In the near future, it is expected that half of newly diagnosed breast cancer cases will be in women 65 and older. This development, along with the aging US population, is evidence that older women will become the prevalent patient cohort in the breast cancer population of the United States. This research utilizes a two-phase simulation modeling approach. The first simulation is a natural history model of breast cancer incidence and progression in random samples of women from the designated population of older US women. The second simulation is an integrated screening-and-treatment model that uses knowledge about the genesis of breast cancer within the same population gained from the natural history model to estimate the benefits of different policies for screening the designated population and treating the relevant individuals. Both simulation models are composed of interacting submodels that represent key aspects of the incidence, progression, screening, treatment, survival, and cost of breast cancer in older US women as well as the overall structure of the system for detecting and treating this disease. Both models have the "warm-up" period 2001–2011, and then they simulate the future years 2012–2020.

The natural history model is a purely discrete event simulation (DES) model that contains a population growth submodel as well as the previously mentioned submodels. The outputs of the natural history simulation model are databases of older women whose breast cancer histories are entirely known; and these histories are critical inputs to the integrated screening simulation model. Additionally, we provide point and confidence interval (CI)

estimates of several system-wide performance measures on an annual basis during the period 2001–2020, including: cancer incidence rates and stage distribution at diagnosis with perfect annual screening; cancer deaths without treatment; cancer prevalence; and the percentages of different invasive-cancer outcomes (breast cancer death, non–breast cancer death, or survival past 2020). We also infer the distribution of several patient-specific performance measures, such as the time for a cancer to reach lethal size when untreated.

The screening-and-treatment simulation model integrates DES and system dynamics (SD) modeling techniques into a single model where both stochastic details and population-level state variables are accounted for and allowed to interact. In addition to the previously mentioned submodels, the DES/SD model contains a population-level SD submodel that represents overall access to (and satisfaction with) screening as well as public awareness of the need for screening. The integrated DES/SD model provides a flexible tool for evaluating the effectiveness of a wide range of population-level screening policies that can be compared directly on individual women who are representative of the designated population. The effectiveness measures for alternative screening policies are compared using statistically designed experiments and optimization routines, and conclusions are presented about the most effective screening policies in terms of several key performance measures. We discuss the rationale for combining the discrete-event and system-dynamics modeling techniques for the analysis of this problem, with an underlying goal of identifying the benefit of using this integrated approach. Our methodology is “individualized” in the sense that we simulate the lives of individual women who are representative of the designated population, and each woman’s risk of being diagnosed with breast cancer is based on her individual risk factors. Other problem areas are explored in this research, including the development of techniques for input modeling, general systems modeling, and output analysis that are specifically adapted to address the special needs of simulation-based health care decision making. Figure 1-1 graphically depicts the overall structure of the simulation models, information about the major inputs, and information about the two simulation models and their submodels.

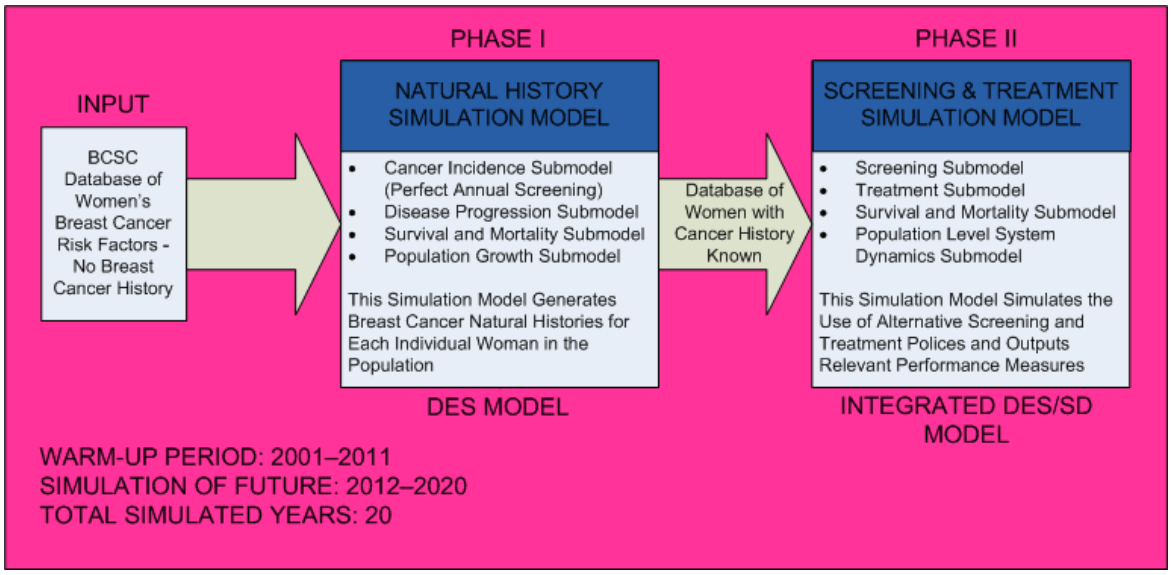


Figure 1-1. Overall Structure of the Simulation Models

1.2 Overview of the Dissertation

The delivery of health care is currently failing to meet the expectations of patients, physicians, administrators and government entities [8]. The methods of operations research have been applied to the health care domain for over sixty years. Simulation is one of the most widely used modeling approaches in operations research, and many would consider it the method of choice for modeling health care systems [8]. We focus on two different approaches to simulation modeling, DES and SD. These are two of the most widely used approaches by modelers in academia and in industry [8, 9]. To this point, the two modeling approaches have coexisted, but have almost exclusively remained separate modeling techniques that are generally used independently.

The integration of these two modeling techniques will help inform guidelines for breast cancer screening, focusing on women at least 65 years of age. As a result of the baby boom generation, the US population is aging. In 1980, 11.3% of the population was at least 65 years old. In 2030, it is projected that over 20% of the population will be at least 65 years old [28, 42], a dramatic increase. Medical advances have also made it possible for people to live longer; life-tables from the Social Security Administration [87] reveal the life

expectancy of 65-year-old women is 19.5 years, which suggests women 65 and over still have considerable life expectancy. With the aging of the US population, larger numbers of older patients will present complex scenarios; and effective screening guidelines for these age groups will become very important [1, 28, 42, 58, 77]. Breast cancer is one of the most common cancers among North American women [42], with over 200,000 new cases of breast cancer in 2009 [105]. The benefits of mammography for middle-aged women are commonly accepted, and much work has been done in evaluating the costs and benefits of screening women in this age group for breast cancer [1, 42, 45, 59]. On the other hand, there are no well-established screening guidelines for women at least 65 years old [42, 45, 77, 86]. Furthermore, clinical trials for breast cancer screening have generally not included women at least 65 years old, and clinicians do not anticipate any clinical trials specific to breast cancer screening in the near future [1, 21, 45]. This research addresses the aforementioned gap in breast cancer screening guidelines using simulation. We provide a detailed discussion of breast cancer in Chapter 2.

DES models progress by scheduling events at discrete points in time, advancing the simulation clock to the time of the event soonest to occur, executing the relevant logic at that event epoch, and then continuing to repeat these steps until reaching the end of the simulation's time horizon. More simply, a DES model is a structured network of queues through which entities flow and services are rendered at discrete points in time. The DES approach is typically used for health care systems, where patients join some form of a waiting list for appointments, examination, and treatments [10]. This approach is also used for disease/biological process models, where it can be used to model diseases within individuals, and to determine optimal screening and treatment policies. DES is explored in further detail in Chapter 2.

In contrast to DES models, SD models do not focus on individual entities; rather, the entities are represented collectively as an amount of “fluid” that flows through different parts of the system to simulate entity movements at an aggregate level. An SD model uses state

variables, which are often called "stocks," to represent fluid levels in different parts of the system, with inflows and outflows from these stocks being governed by "flow rates" and other system status variables. The fundamental concept behind SD modeling is that the structure interconnecting different parts of the system determines overall system behavior. SD models typically have both continuous and discrete state variables, whose evolution over time is governed by the following: (a) differential and/or difference equations for the continuous state variables; and (b) piecewise constant functions of simulated time with jumps at event times for the discrete state variables. Typical SD models represent the evolution of these state variables over time, and the output analysis is focused on particular state variables of interest. For example, Murray [65] is an SD model of colorectal cancer, where the state variables of interest are the colorectal cancer screening rates. This model is revisited in Chapter 3. The integrated screening model has an SD submodel and several DES submodels that interact with one another. The population-level state variables evolving over time in the SD submodel are linked to stochastic details of individual women in the DES submodel using "hybrid levels" and a logistic regression model for adherence to breast cancer screening policies. Hybrid levels are SD levels (global variables) that are also used in the DES submodel to alter logic and the probabilities of certain events occurring. The logistic regression model uses a combination of attributes (local variables) that are assigned to each individual woman and primary SD levels to compute the probability of adherence to screening appointments for use in the DES screening submodel. The adherence or non-adherence of each woman alters both future discrete-events and SD levels. The development of the SD submodel required a study of the structure of the breast cancer screening system. The term “system” refers to the interdependent group of components that make up the US national infrastructure for breast cancer screening such as physicians, Medicare, patients, and doctors. SD modeling is explored in further detail in Chapter 2.

The idea of an integrated DES/SD model applied to health care is discussed briefly in the literature [11, 19], but few models have been developed; and these papers are simply a call for further research on this topic. In this research, we integrate these two modeling

approaches into a unified approach that has the power to give insights into specific operational details as well as the dynamics of the system's response to changes in its state variables. SD is used to represent the time-dependent response of the overall system for breast cancer screening and treatment, while DES will be used to represent screening events, treatment events, and other relevant discrete-time events for each individual woman in the designated population served by this system. For example, screening compliance of the population is a state variable of interest in the SD submodel that is linked to individual compliance in the DES submodel. Brailsford [9] states "if it were possible to develop a methodology with the advantages of both – the detailed, stochastic, individual patient level approach of DES combined with the whole-systems, strategic view of SD – this approach would have benefits far beyond health care." This research is a step towards that goal. Although the applications of this type of modeling approach could be easily extended beyond the domain of health care, breast cancer screening provides a particularly instructive example of the advantages of this approach.

Arena 13.9 produced by Rockwell Automation [48] is the software package chosen to construct both simulation models and to integrate the different modeling approaches. It has been chosen because of its versatility in that it can model both discrete and continuous variables, and it can also approximate the solution of a system of ordinary differential equations governing the behavior of smooth (continuous) system status variables over time. An arbitrary number of attributes can be assigned to entities, and these attributes can contain logical statements of arbitrary complexity using Arena's "Expression Builder." Arena is typically described as DES software, but it has the continuous modeling capabilities needed for developing SD models and has been used to create SD simulation models [65].

The health care system, and in particular breast cancer screening, will benefit from a simulation methodology that takes a holistic system view and still captures stochastic details. For health care in general and breast cancer screening in particular, there is a structure guiding the actions of the people and components that make up the system. That structure is

imposed by a number of sources, including government policy and insurance requirements. Although this overall structure defines the general behavior of the population, each individual has a different experience defined by a number of detailed stochastic components such as the growth rate of an individual's tumor or the reliability of the screening test for their particular cancer. A combined DES/SD model addresses the structure of the system with SD and the stochastic details with DES.

Cramp [20] has examined several models for assessing health policy strategies which can be applied to assessing the performance of breast cancer screening. He cites the fact that the forces of the external structure play a large role in the health care system. He calls for the use of tools that enable various complex scenarios to be analyzed and results of those scenarios to be accurately predicted. A scenario is defined as a hypothetical set of events formulated with the purpose of focusing attention on causal processes and decisions. In creating these scenarios, it is important to combine quantitative data with the qualitative and subjective knowledge of health care experts. This simulation modeling methodology creates an environment where complex breast cancer screening scenarios are played out in the future period 2012–2020 and performance measures of interest are estimated.

In this dissertation the development of the relevant simulation models has been divided into two distinct phases. The first phase concerns a natural history model of breast cancer incidence and progression in random samples of women from the designated population of older US women. The second phase concerns an integrated screening-and-treatment model that uses the knowledge about breast cancer within that same population that was gained from the natural history model to estimate the effects of different policies for screening and treatment of older US women. The natural history model is a purely discrete event simulation model. It establishes a baseline for the benefits of screening by (a) determining the age that cancer is detected for each afflicted woman in the population given perfect annual screening; and then (b) allowing that cancer to remain untreated. We accomplish this using a risk model that assumes screening is performed each year and

follow-up is done to ensure accurate diagnosis. Once diagnosis occurs, the cancer is allowed to continue growing according to a Gompertzian tumor growth model until death from breast cancer occurs, death from natural causes occurs, or the simulation reaches the end of the time horizon, noting that we compute the age at breast cancer death for all women with breast cancer even if the death would occur after the year 2020. Upon completion, the natural history model generates a database that contains each woman's individual non–breast cancer and cancer history, such as cancer onset age and age at non–breast cancer death. This database is input to the integrated screening model, where imperfect screening policies at different frequencies, treatment, survival, and costs are considered.

The rest of this dissertation is organized as follows. In Chapter 2, we provide essential background information on simulation in health care, breast cancer, DES and SD modeling, and the integration of these modeling techniques. In Chapter 3 we explain the development of the natural history model and all its submodels, which are used to simulate the risk and onset of cancer in several randomly sampled populations. In this chapter we also summarize some results from the natural history model. In Chapter 4 we explain the development and structure of the integrated DES/SD screening-and-treatment simulation model and all of its submodels, which are used to determine the effectiveness and amount of life saved when using selected policies for breast cancer screening and treatment. Chapter 4 also includes the following: (a) an analysis of the results of the integrated screening simulation model using designed statistical experiments and optimization routines; (b) point and CI estimates for important performance measures; and (c) a discussion of all these results, including identification of the statistically best screening policy in terms the most important performance measures. In Chapter 5, we discuss the main findings of this research, we examine the limitations of the simulation models, and we make recommendations for future work.

CHAPTER 2: LITERATURE REVIEW AND RESEARCH MOTIVATION

This chapter begins with a discussion of health care simulation models, including a method of classifying such models. The next section focuses on breast cancer, with special emphasis on the following topics: a description of the disease; the screening and treatment for the disease and; and past modeling approaches. This chapter also includes descriptions of the two simulation modeling techniques that will be integrated in the Tejada model and how they have been used in general applications to health care as well as in specific applications to breast cancer screening. Finally we elaborate the problems and research questions that motivated this research.

2.1 Simulation in Health Care

Simulation models have been developed to improve health care systems for over sixty years. In a survey of simulation modeling in health and health care delivery, Fone [33] attempted to account for and classify the papers identified as simulation models applied to problems in healthcare. Papers are separated by their quality, originating country, and decade of publication. Table 2-1 presents the results of this survey. A quick glance at the last line in Table 2-1 reveals that in total, 182 papers were evaluated in the survey.

Table 2-1. Summary Data on Papers Appraised [33]

Paper Area	Total	Grade				Country of Origin			Year of Publication				
		A	B	C	D	UK	USA	Other	Pre 1980	1980–1984	1985–1989	1990–1994	1995–1999
Hospital Planning	94	16	36	26	16	8	73	13	26	11	12	19	26
Percentage	51.6	17.0	38.3	27.7	17.0	8.5	77.7	13.8	27.7	11.7	12.8	20.2	27.7
Communicable Disease	7	2	2	2	1	1	2	4	0	1	0	2	4
Percentage	3.8	28.6	28.6	28.6	14.3	14.3	28.6	57.1	0.0	14.3	0.0	28.6	57.1
Economic Evaluation	17	5	7	2	3	2	9	6	0	1	0	1	15
Percentage	9.3	29.4	41.2	11.8	17.6	11.8	52.9	35.3	0.0	5.9	0.0	5.9	88.2
Screening	44	11	26	7	0	4	14	26	0	2	5	17	20
Percentage	24.2	25.0	59.1	15.9	0.0	9.1	31.8	59.1	0.0	4.5	11.4	38.6	45.5
Miscellaneous	20	3	9	5	3	6	11	3	1	0	3	4	12
Percentage	11.0	15.0	45.0	25.0	15.0	30.0	55.0	15.0	5.0	0.0	15.0	20.0	60.0
All Papers	182	37	80	42	23	21	109	52	27	15	20	43	77
Percentage	100.0	20.3	44.0	23.1	12.6	11.5	59.9	28.6	14.8	8.2	11.0	23.6	42.3

Simulation has only recently begun to be accepted in the health care community despite its introduction over thirty years ago [30]. It is important that this work take into account the fact that health care professionals should at the very least, believe that the model is a valid representation of the real world, and the model results are of value to policy makers. Sanchez et al. [83] address the emerging issues in health care simulation. Some of the important issues identified that are relevant to this research include the following:

1. combining models of information technology with traditional process models;
2. improving capabilities for verifying and validating simulation models;
3. developing models without data – using expert opinion;
4. developing new and innovative tools for process analysis in health care;
5. conflicting objectives—for example, between stakeholders such as patients interested in living as long and comfortably as possible while keeping costs down, and insurance companies who want to pay only as much as they have to.

Standridge [89] explains the applications and issues in regards to simulation in health care. He states that the application of simulation in health care can be divided into four categories: public policy; patient treatment processes; capital expenditure requirements; and operating policies. The importance of this paper lies in the convincing case made by the author that all these issues need to be addressed concurrently. Eldabi et al. [30] argue that the immense potential for simulation in health care has not been realized; and they provide a particularly insightful survey of professionals using simulation in health care which helps reinforce the need for this dissertation research. The responses that are directly applicable are cited in Table 2-2. To the best of our knowledge, a single model which integrates DES and SD has not yet been constructed. The subjects surveyed consisted of 12 UK academics, 4 North American academics, and 6 UK industry personnel.

Table 2-2. Responses: Current Trends of Simulation in Health Care [30]

Respondent	Response
1	Taking a "whole system" view – it is necessary to consider the links between component parts of the health care system rather than view each in isolation
5	Research on combining DES and SD approaches to incorporate the benefits of both techniques and model large complex interconnected systems (such as the use of SD to more large populations and trends, while DES captures lower level planning detail such as Emergency beds and resources, etc.)
6	Development of hybrid techniques and methodologies is important. Trend in the health care domain appears to be focused on metrics, improved performance, cost effectiveness analysis, and more accurate predictions of future activities for better resource utilization
8	There is a need to be much more ambitious to develop whole system approaches to this delivery chain. This requires a rate mixture of hard and soft approaches. The technical work could be done in system dynamics,

It is clear from the survey results reported in [30] that the majority of researchers favor pursuing the idea of mixing methodologies rather than expanding a single methodology in order to address the complexity and diversity of problems presented by health care systems. The authors assert that new methods for simulation modeling and for interacting

with the user must be created. The community needs a starting point from which to work towards integrating several methodologies, and a framework within which the combined methodologies might operate. User interaction and the integration of methodologies are needs that are addressed in this research. This literature provides evidence that an integrated simulation model that can capture the overall system dynamics as well as stochastic details could be useful in solving problems presented by breast cancer screening and in other applications.

2.1.1 Taxonomy of Health Care Simulation Models

Simulation models in health care can be broadly classified into three groups that are called level-1, level-2, and level-3 and that are explored in detail by Brailsford [9, 10]. Level-1 models, represent the human body, and they are usually called “disease” models; but it is possible to use them to represent biological processes among healthy individuals. These models can be at the body-system level, the organ level, or the cellular/microbiological level. The most frequent application of these models is for the evaluation of the cost effectiveness of some intervention, such as screening. One example of a level-1 DES model is the simulation developed by Davies and Brailsford [23, 24], where the authors sought to identify the best screening procedures for diabetic retinopathy, a disease that can lead to blindness if it is not detected early enough. In addition, Roberts et al. [78] and Tafazzoli [95] developed an object-oriented DES model to evaluate the effectiveness of screening for colorectal cancer. In addition to DES models, level-1 SD models also exist. For example, a level-1 SD model created by Evenden [31] examines screening and intervention techniques (education, pamphlets, etc.) for preventing the spread of chlamydia.

Level 2 models are operational or tactical models at the health care unit level. The unit level refers to a clinic, or a hospital department such as the operating room or emergency room. These models are usually concerned with modeling behavior and movement of individual patients (entities), but do not focus on biological processes. They tend to be more similar to conventional process-oriented simulation models, because they attempt to identify

and address bottlenecks, which is similar to the traditional method for using simulation models in the design and analysis of manufacturing and production systems. The academic literature contains thousands of level 2 models. DES is the modeling approach of choice for level 2 models; and to the best of our knowledge, there are no SD models that would be classified as level 2 models. An example of a level 2 DES model can be found in Harper and Shahani [40], where a simulation model of an 800-bed hospital in the United Kingdom was created in order to determine the hospital's bed capacity structure for departments as a function of the level of demand. Shifting beds from a low-demand department to another department experiencing peak demand at a given time can increase hospital efficiency. Another capacity-planning model of an intensive care unit (ICU) was developed by Griffiths [38], and it was used to determine staffing and bed requirements for the ICU. There are a number of other level 2 models, and many different health care units have been modeled by numerous researchers. As a basis for research, it should be noted that the system boundaries of these models are the physical walls of the unit under consideration, and currently, the effects of policy decisions on the dynamic complexity of the overall organization are not considered.

Level 3 models are at the strategic level and are system-wide models which usually do not model individual patients. There are comparatively very few of these types of models in the literature. Level 3 models are almost always created using the SD approach, because it lends itself to considering long-term and broad effects in a relatively low level of detail. These types of models are becoming increasingly popular in the United Kingdom. One of the best known examples of a level 3 SD model was developed by Lane et al. [55], and it explores the relationships between waiting times in the emergency room and hospital bed closures. One of the interesting findings of this study was the fact that the immediate impact of bed shortages was not observed in the ER, and that it did not make sense to look at any single measure in isolation because none of them individually were the cause of the problem. The system was too complex, and so a holistic system view was required as future work. Another example of a level 3 SD model was developed by Brailsford et al. [12], and it looked

at the entire UK health care system to determine the cause of extreme spikes in demand in recent years. Finally, Murray [65] developed an SD model to address colorectal cancer care. As discussed previously, this model uses the SD approach and observes trends in screening rates over time. Review of this thesis provided insight regarding many SD modeling concepts relevant to cancer screening.

2.2 Breast Cancer: The Disease, Screening, Treatment, and Modeling Approaches

The Centers for Disease Control and Prevention (CDC) estimate that one-in-eight US women (12.5%) will develop breast cancer at some point in their lives [18]. In addition, it is expected that half of the newly diagnosed cases of breast cancer will be in women who are at least 65 years old [28, 42]. The aging of the US population will cause the rate of occurrence of breast cancer in older women to increase faster than in the overall population [67]. Therefore older women will become the prevalent patient cohort in the breast cancer population. Breast cancer is believed to behave less aggresively in older women [21, 27, 28, 60, 73], and these women have a significantly greater chance of death from other medical conditions. Although benefits of screening these women may be limited, a 65-year-old woman has an expected lifetime of approximatley 19 years [86, 87], which suggests screening could still provide significant benefits for older women. The complexity of screening-and-treatment decision making for older women highlights a critical need to understand the implications and quantify the the impact of population-based policies on individuals.

The *Cancer Trends Progress Report* estimates the annual cost of breast cancer to be at least $12.2 billion per year, and these costs are expected to increase with advances in medical technology and the associated treatments [67]. The majority of older women are eligible for Medicare/Medcaid, which is paid for in part by US taxpayers. Thus, patients, health care providers, and the general population all have a stake in breast cancer screening. It is well established that screening techniques, particularly mammography, can detect breast cancer before it becomes symptomatic, improving the chances of survival as well as quality

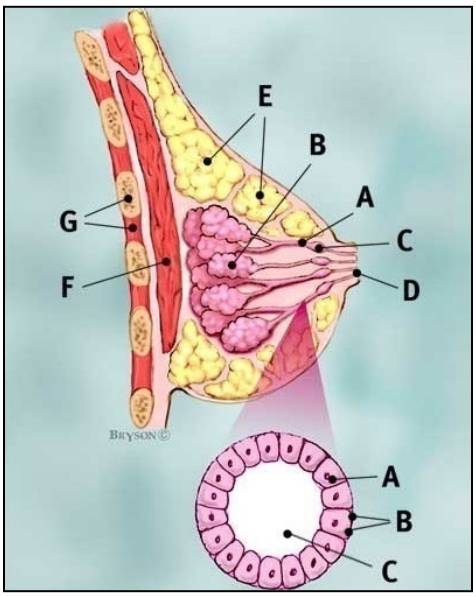
of life after treatment [32]. While considerable research has been done regarding screening policies for middle-aged women, well-defined screening policies for older women do not currently exist [45, 77, 86]. In the rest of this section, we review the background on breast cancer biology, screening and treatment for breast cancer, and existing approaches to mathematically modeling these phenomena.

2.2.1 The Disease

Breast cancer can be defined as the uncontrolled growth of cells in the breast [105]. Cancer occurs as a result of changes or mutations in the genes that are responsible for the growth rate of cells and keeping cells healthy. Usually cells in the body regenerate via an organized process of cell growth, wherein new healthy cells are created as old ones die. However over time, mutations can alter cells by turning certain genes on and other genes off. The altered cell has the ability to continue to divide without control or order, creating more cells exactly like it and eventually forming a tumor. A tumor can be either benign or malignant. Benign tumors are those that are not considered to be a health threat in the sense that the cells of the tumor are similar to normal cells, they grow very slowly, and they do not spread to other parts of the body. By contrast, malignant tumors are cancerous; and cells from these tumors will eventually spread to other parts of the body and cause death. The term *invasive breast cancer* is used to describe a malignant tumor that originated in the cells of the breast [105].

Over time the malignant cells will invade healthy breast tissue and eventually the lymph nodes under the arms, which act as one of the body's filters of toxins. If the cancerous cells spread to the lymph nodes, then the cancerous cells have access to other parts of the body and will eventually spread to major organs. Generally, the stages of breast cancer are based upon the size of the tumor, whether the lymph nodes have been affected, and the distance that the cancerous cells have spread from the breast to other parts of the body. Even though cancer is caused by genetic mutations, only 5%–10% of genetic cell mutations causing breast cancer are inherited from family members, while the other 90%–95% of cases

are due to cell mutations caused by normal aging of the human body [105]. Most breast cancers originate in the cells of the lobules, the milk-producing glands, or the ducts, the passageways used to drain milk from the lobules to the nipple. Less commonly, breast cancer originates in the stromal tissue, which contains fatty and fibrous connective tissues of the breast. Figure 2-1 shows a diagram of a typical breast, including the surrounding areas and areas mentioned above where cancer usually begins to form.



Breast profile:
A Ducts
B Lobules
C Dilated section of duct to hold milk
D Nipple
E Fat
F Pectoralis major muscle
G Chest wall/rib cage

Enlargement
A Normal duct cells
B Basement membrane
C Lumen (center of duct)

Figure 2-1. Diagram of a Typical Breast [105]

An understanding of breast cancer biology within the body is critical to the development not only of the tumor model but also of the model for the spread of the disease (metastasis). These disease models will be used to evaluate different screening and treatment policies. The most important anatomical variables are tumor size and the degree of metastasis to other parts of the body. A Gompertzian tumor growth equation [69] will be used to model the size of the initial tumor within the breast. The level of metastasis within the body will also be modeled using a stage progression model that determines the probability of the cancer being in the local, regional, or distant stage given the tumor size. Breast cancer is divided into stages, stages 0, I, II and some cancers in stage III are considered early stage breast

cancers. Stage III and stage IV are considered advanced stages of the disease that usually involve the spread of cancer to other parts of the body. Table 2-3 and Table 2-4 show the different stages of breast cancer. Within these stages, there are various assignments for values of T (tumor), N (lymph nodes), and M (distant metastasis). For example, T0 means that there is no evidence of a primary tumor, and TX means that the primary tumor could not be diagnosed. This structured type of system is used to keep track of the state of the disease as it progresses within the human body [105].

Table 2-3. Detailed Stages of Breast Cancer [105]

Stage	Definition
Stage 0	Cancer cells remain inside the breast duct, without invasion into normal adjacent breast tissue.
Stage I	Cancer is 2 centimeters or less and is confined to the breast (lymph nodes are clear).
Stage IIA	No tumor can be found in the breast, but cancer cells are found in the auxiliary lymph nodes (the lymph nodes under the arm) OR the tumor measures 2 cm or smaller and has spread to the auxiliary lymph nodes OR the tumor is larger than 2 cm but no larger than 5 cm and has not spread to the auxiliary lymph nodes.
Stage IIB	The tumor is larger than 2 cm but no larger than 5 cm and has spread to the auxiliary lymph nodes OR the tumor is larger than 5 cm but has not spread to the auxiliary lymph nodes.
Stage IIIA	No tumor is found in the breast. Cancer is found in auxiliary lymph nodes that are sticking together or to other structures, or cancer may be found in lymph nodes near the breastbone OR the tumor is any size. Cancer has spread to the auxiliary lymph nodes, which are sticking together or to other structures, or cancer may be found in lymph nodes near the breastbone.
Stage IIIB	The tumor may be any size and has spread to the chest wall and/or skin of the breast AND may have spread to auxiliary lymph nodes that are clumped together or sticking to other structures or cancer may have spread to lymph nodes near the breastbone. Inflammatory breast cancer is considered at least stage IIIB
Stage IIIC	There may either be no sign of cancer in the breast or a tumor may be any size and may have spread to the chest wall and/or the skin of the breast AND the cancer has spread to lymph nodes either above or below the collarbone AND the cancer may have spread to auxiliary lymph nodes or to lymph nodes near the breastbone.
Stage IV	The cancer has spread — or metastasized — to other parts of the body.

Table 2-4. Local, Regional, and Distant Stages of Breast Cancer [105]

Stage	Definition
Local	The cancer is confined within the breast.
Regional	The lymph nodes, primarily those in the armpit, and possibly those near the collarbone, are involved.
Distant	The cancer is found in other parts of the body and other organs are involved.

There are several risk factors which can affect the possibility of getting breast cancer as well as the outcome of the treatment process. Other than gender, age is the most important risk factor for breast cancer. As a woman ages, the probability of developing some form of breast cancer increases significantly; but it is believed that tumors grow more slowly in older women [21, 27, 28, 59, 73]. Combining this information with the fact that the US population is expected to increase in average age consistently over the next twenty years, the appropriate conclusion is that breast cancer among women in the US will become an increasingly complex problem. In the near future, it will become more important to have screening guidelines for women over the age of 65.

Among the risk factors that have been identified, some are controllable by the person at risk and some are not. Controllable factors include: weight, diet, exercise, smoking, drinking alcohol, exposure to estrogen, recent use of oral contraceptives (birth control pill), stress, and anxiety. Gender, age, family history, personal history, race, exposure to radiation therapy, exposure to estrogen, breast cellular changes, pregnancy and breastfeeding, and exposure to outdated medicine to prevent miscarriages are among the risk factors that cannot be controlled by the person at risk [105]. Some other risk factors that have recently been recognized and are incorporated in this research are body mass index (BMI) and breast density. For the older population, there is an idea that other diseases may actually begin to take precedence over breast cancer. These are referred to as “comorbidities.” Some of the specific diseases in this category include, but are not limited to, the following: hypertension, diabetes, coronary artery disease, congestive heart failure, dementia, and other cancers such as lung or ovarian cancer. Comorbidities for people over a certain age (65, 70, or 75) may substantially reduce the risk of death as a result of breast cancer, thus complicating the decision about how frequently to screen this population.

2.2.1.1 Ductal Carcinoma In-Situ (DCIS)

DCIS is the most commonly occurring noninvasive breast cancer. As the name indicates, DCIS begins in the milk ducts of the breast. The term in-situ means “in its original place.” Hence, this type of cancer has not spread to any other parts of the breast other than the milk ducts. DCIS is not considered to be life-threatening like invasive cancer; but it needs to be treated, and it increases the risk of being diagnosed with invasive breast cancer later in life. The majority of recurrences happen within 5–10 years of the original diagnosis, and the recurrence rate is about 30% [69]. There are approximately 60,000 cases of DCIS diagnosed each year, and they make up about one in every five new breast cancer cases. This number may seem large, but people are living longer and the risk of cancer increases with increasing age. The increase in screening technology also enables easier and earlier detection this type of cancer, which contributes to the increasing number of DCIS cases [105]. While DCIS is considered an additional risk factor for developing invasive breast cancer later in life, very little is known about the progression of DCIS to invasive cancer.

2.2.2 Screening and Treatment

The earlier breast cancer is diagnosed, the greater is the probability of surviving and recovering. Screening is a formal method of examining certain characteristics of the breast, either physically by touch, or by digital medical imaging, in order to determine whether breast cancer or any signs or symptoms related to breast cancer are present at the time of screening. Treatment refers to the process that is begun once it is known that cancer is present, and the stage to which the cancer has progressed within an individual [105]. The treatment process may include surgery, radiation therapy, chemotherapy, hormonal therapy, targeted therapies, and holistic medicines, or some combination of these. In general, there are three types of tests for breast cancer: screening, diagnostic, and monitoring. Screening tests (such as a mammogram) are used for early detection of breast cancer among healthy individuals; diagnostic tests (such as a biopsy) are given to those who are suspected of having breast cancer through symptoms or screening results; and monitoring tests (such as a

blood chemistry test) are used to monitor the progress of treatments and therapies as well as to monitor recurring cancers [105]. Table 2-5 outlines the different types of tests that exist.

It is outside the scope of this dissertation to present the details of every test for every category, but the basics of screening and treatment are reviewed. Berg [5] describes tailored supplemental screening for breast cancer, and explains the current state of breast cancer screening as well as the potential future directions.

Table 2-5. Breast Cancer Screening, Diagnostic, and Monitoring Tests

Screening Tests	**Diagnostic Tests**	**Monitoring Tests**
Physical Exam (Self or Physician)	Biopsy	Blood Cell Counts
Mammography	Breast MRI	Blood Chemistries
Digital Tomosynthesis	Bone Scan	Blood Marker Tests
Ductal Lavage	CAT Scan	
Thermograhpy	Chest X-Ray	
Molecular Breast Imaging	Ultrasound	
Breast MRI	SPoT-Light HER2 CISH	
Sonography	Oncotype DX	
	IHC Tests (ImmunoHistoChemistry)	
	FISH Test (Fluorescence In Situ Hybridization)	

Physical breast exams and mammography are the most common forms of screening for breast cancer, and mammography is the only screening technique that is currently recommended for population-based screening in the United States [86]. Therefore, mammography and clinical detection (detection of a lump in the breast by physical exam) will be the focus of the modeling efforts in our research. Table 2-6 summarizes current

mammography screening guidelines from a variety of sources, including the American Cancer Society (ACS) and American Medical Association (AMA). This table shows that well-defined screening guidelines have not been developed for older woman by any of the public medical authorities.

Table 2-6. Mammography Screening Guidelines [77]

Agency	Middle-Aged Recommendation	Older Recommendation
Medicare/Medicaid Reimbursement	Annual screening mammography for women >40 yrs (if eligible)	For all women after 70 years old, risk factors associated with breast cancer should be explored every 1 – 2 years to discuss breast cancer screening and to address the risks and benefits associated with screening, as well as the individual's comorbidities, life expectancy, health status and quality of life.
American Academy of Family Physicians	Every 1–2 yrs for women 50–69 yrs	
American College of Obstetricians and Gynecologists	Every 1–2 yrs for women >50 yrs	
American Cancer Society	Yearly for women >40 yrs	
American Medical Association	Yearly after 50 yrs	
Canadian Task Force on Preventive Health Care	Every 1–2 yrs for women 50–69 yrs	
US Preventive Services Task Force	Every 1–2 yrs for women 50–74 yrs	

Nine randomized trials have shown that the benefits of mammography screening for breast cancer include a 20% reduction in the mortality rate; however, eight of those trials did not consider the screening of older patients [1, 21]. Of course, this is complicated by the fact that breast cancer in older patients is believed to be less aggressive [21, 27, 28, 73]; and the benefits of screening may be offset by risks from comorbidities that could lead to other causes of death. Upon examination of Table 2-6, it is evident that clear screening guidelines for women over 65 have not yet been developed. These issues form the need for considerable research regarding effective screening strategies for older women, and these strategies could be dependent upon risk factors, chronological age, and a number of other variables.

Biopsy is the most common form of diagnostic test. Once the results of the screening test indicate that the woman may have breast cancer, sample tissue from her breast is taken and analyzed by a specialist in order to determine if cancer is present; and if it is, then the characteristics of the cancer are determined. Several other diagnostic tests are listed in Table

2-5, and are generally used in certain instances where a particular type of test has the ability to provide information that cannot be obtained from other diagnostic tests. Blood tests are the standard for monitoring tests; however if the physician believes the cancer has changed stage or spread to other parts of the body including the lymph nodes, then he may order some diagnostic tests in order to confirm or refute his hypothesis about the spread of the cancer [105].

2.2.4 Past Modeling Approaches

In this section, we present an overview of some of the previous modeling approaches to provide the context for the proposed work in the breast cancer screening research area and within the science of simulation. Previous models have made simplifying assumptions that either limit the applicability or validity of the model, or do not focus on the same factors that this research addresses. Previous modeling approaches to breast cancer screening are important because they provide a framework for the research completed in this dissertation.

2.2.4.1 Analytical Models

Kirch and Klein [49, 50] develop one of the earlier analytical models for examination schedules of age-dependent diseases, where the objective is to minimize the delay in detection of the disease. The problem is to decide how many examinations should be given in each of the eleven five-year age periods between the ages 25 and 79. In [50], they consider two basic cases, one where the delay until detection is a constant, and one where the delay until detection is a random variable. They also provide a procedure for solving this scheduling optimization problem. In the case of breast cancer, results indicate that nonperiodic schedules will yield an average savings of 2%–3% in the number of exams (delay) prior to detection. The results also showed that a significant improvement in detection time could be achieved if screening were semiannually as opposed to annually. They also suggested that for mass-screening programs, each individual's screening interval should be dependent on age instead of fixed throughout life. This is an early analytical model that contains many simplifying assumptions, and is not particularly flexible or robust.

Schwartz [84] develops a Markov model of breast cancer that incorporates the different states of the disease, tumor size progression, and progression to the lymph nodes in order to predict the benefits of screening strategies in terms of life expectancy gain, and probability of detection prior to nodal involvement. The screening strategies compared screening every 10, 5, 3, 2, and 1 year(s). This model assumes test efficiency and rate of disease spread are stationary, does not account for false positives and negatives in the screening process, and assumes the decision maker (the physician) is omniscient and has every piece of necessary information. It also looks at screening from a purely individual perspective, as opposed to allowing some interaction with the surrounding environment. The results of this model suggest that life expectancy and the probability of detection prior to nodal involvement can be improved by screening annually as opposed to less frequent screening. It is also suggested that if the goal is to maximize life expectancy, then screening should be concentrated in the earlier years (40–70 years old). However, if the goal is to minimize the probability of recurrence of the disease, then screening should be concentrated in the later years. Understanding this approach to modeling was beneficial in that it provided a structure for important aspects of breast cancer screening that need to be taken into account in order to have an accurate model. It also provided a basic idea of performance measures used to assess the performance of different screening policies. This model is restricted by Markovian assumptions about progression of the disease.

Ozekici and Pliska's [70] delayed Markov scheduling model is formulated as a generic scheduling problem that can be solved by dynamic programming, and it is later applied to breast cancer screening. This is one of the first scheduling models to take into account false positives and negatives, and it also allows the time until deterioration begins (i.e., the time until onset of the disease) to be any nonnegative random variable. This model determines the optimal inspection schedule of a system that starts in a "good" state, deteriorates gradually over time with any number of "defective" states, and ends in a "failed" state. The objective is to minimize the sum of the costs associated with the disease, which are subjectively assigned; moreover, the model is sensitive to these costs. Applying this

procedure to breast cancer screening yields an optimal policy that involves a sequence of physical and mammographic examinations at ages 50, 55.5, 60, 65, and 75 with some caveats. Some inaccuracies of this model include the assumption of stationary transition probabilities (i.e., disease progression with age is constant), and the fact that once a person has the disease, there is no further diagnostic process. This research is an interesting application of an inspection scheduling problem with false positives and false negatives applied to breast cancer. However, it is a very simple model which a small portion of the problem in breast cancer screening and it also contains a number of simplifying assumptions that this dissertation addresses, such as treatment and survival.

Parmigiani [71] uses Bayesian decision techniques to develop a model that predicts the effect of age on the minimum-cost screening schedule. The model used a non-Markovian stochastic process to represent the progression of the disease; however the model fails to distinguish deaths due to breast cancer from deaths not due to breast cancer. Each factor considered in this model is handled in isolation. An approach which considers the interaction of selected screening factors would clearly be more representative of the real-world system.

Mandelblatt et al. [60] develop a model for breast cancer screening in older women that takes into account two comorbid conditions, hypertension and congestive heart failure. Moreover, this model only considers whether or not screening should be offered to older women, and does not explore the proper screening policies for women in this age group.

Baker [2] develops a model that uses data from a study in the United Kingdom for parameterization; and she uses a continuous-time, non-Markovian stochastic process to model disease progression and screening. There are several limitations of this model, including limited sample data, poor fitting of the data to model parameters, and quite subjective cost values.

Lee and Zelen [57] develop a stochastic model that is able to predict breast cancer death and survival rates based on a number of factors such as examination schedules, population age distribution, follow-up time, and survival condition. Maillart and Ivy [59] describe a Markov model of dynamic breast cancer screening policies. They hypothesize that the screening interval at younger ages should be different than the screening interval at older ages and evaluate a large range of policies to develop a set of "efficient policies" measured by a lifetime breast cancer mortality risk metric and by the expected mammogram count. In this scenario, a patient could potentially select a policy based on individual circumstances. One of the authors' main conclusions is that screening should begin early in life and should continue until late in life regardless of the screening intervals adopted, which implies screening older women is beneficial.

The models that have been reviewed to this point are analytical in nature and attempt to provide a closed-form solution to complex problems. One drawback common to all these types of models is their inability to individualize the results. Another drawback is their reliance on many simplifying assumptions. Substantially better screening-and-treatment policies could be achieved by exploiting a methodology that allows results to be applied on an individual level and that allows these results to be a function of the individual risk factors. Simulation has the ability to remedy this problem because of its versatility and the fact that it does not require a closed-form solution to the problem at hand; instead it simulates the system and provides estimates of performance measures of interest over an experimental region containing all feasible solutions to the given problem. Generally, simulation provides a vast amount of insight, as opposed to solving for certain parameters and variables in a very specific mathematical model; moreover, simulation provides a framework of answering a variety of "what if" questions about changes in screening policies. Simulation is ideal for comparing the results of systems when changes are made that could affect the output measures of interest.

2.2.4.2 Simulation Models

There have been relatively fewer attempts to model the breast cancer screening process using simulation. Michaelson et al. [62] and Fryback et al. [34] both use simulation modeling to answer breast cancer screening questions.

Michaelson et al. [62] develops a simulation methodology to test the ability of different screening intervals to detect breast cancer prior to the spread of the disease, using data on rates of breast cancer growth and metastasis. According to the taxonomy of Section 2.1.1, this is a level-1 model. This is a basic simulation model and does not consider anything that happens after the cancer is diagnosed. The work of Michaelson et al. [62] is a good starting point for modeling the disease in a simulation environment, but it needs to be extended in a number of ways to account for a vast number of risk factors and population variables that are not taken into consideration. The results of the model are fairly straightforward, suggesting that more frequent screening intervals could increase the number of breast cancers identified in the early stage of biological development.

Fryback et al. [34] develop a level-1 DES model of breast cancer in the Wisconsin population. Four interacting processes involved in breast cancer care are modeled: the natural history of the disease, detection of the disease, treatment of the disease, and competing cause mortality. The disease submodel is parameterized using data on breast cancer in the state of Wisconsin, and it provides a good framework for a breast cancer disease model. The model of Frybeck et al. [34] is somewhat simplistic in that it uses several critical factors to come up with a total onset proportion, or proportion of women in the population who get cancer regardless of their individual risk factors; and then it simulates the disease according to the indicated onset. The screening techniques considered are “clinical surfacing,” where the tumor is discovered physically, and mammography. A simple function was used to estimate technological improvements over time. The treatment process is modeled as a process where the women is either cured of the cancer and life proceeds normally, or she dies of breast cancer in the course of treatment. These are the only two possible outcomes of the treatment

process in this model. This model is retrospective, whereas our research looks into future and attempts to quantify the potential benefit of screening over the next 25 years. The results of the Wisconsin model indicate that breast cancer mortality for the period 1975–2000 would have been reduced by 38.3% with mammography screening and postcancer therapy compared with the option of not using these techniques during this time period.

2.3 DES Modeling

DES is one of the most widely used modeling techniques in operations research. It is important to understand the basic concepts surrounding DES and how it has been used to model health care systems [8, 9]. We begin by providing a brief description of the fundamental concepts behind DES.

DES models progress by scheduling events at discrete points in time, advancing the simulation clock to the time of the event to occur soonest, executing the relevant logic at that event epoch, and then continuing to repeat these steps until reaching the end of the simulation's time horizon. Times between events (such as arrivals or services) are usually stochastic and are represented as random variables. In order to schedule events in the proper chronological order, samples are taken from the appropriate distributions and assigned to the relevant variables in the simulation model logic. The entities that advance through parts of a health care model are typically patients; but this is not always true and other creative modeling techniques can be utilized (for example, the basic entities moving through the various parts of the model might be health care providers or pieces of equipment). Entities (patients) each carry their own patient-specific attributes with them as they move through the model. These attributes can be used to determine the routing of the patient through different wards of a hospital, as well as to keep track of relevant statistics such as waiting time in queues, service times, and many other measurements of interest.

The DES approach is typically used for health care systems in which patients join some form of a waiting list for appointments, examination, and treatments [10]. This methodology allows the user to determine the current state of the overall system (in terms of operational efficiency, as well as variable parameters), and to ask "what if" questions to improve the design of the current system, or to design a new system from scratch. Jun et al. [47] provide an extensive survey of applications of discrete event simulation in health care clinics. The authors state that DES can be used to forecast the impact of changes in patient flow, to examine resource needs, or to investigate the complex relationships among the different model variables. These options provide managers with the ability to select and configure system components in order to reconfigure the simulation system and evaluate the simulated performance of the reconfigured system without actually altering the present system in the real world. This mode of using discrete-event simulation is of vital importance in the realm of health care, where it would be highly unethical, for example, to let a breast cancer patient's cancerous tissue or tumor progress untreated solely for the purpose of gaining insight into its biological behavior. Instead, we define and mathematically model what we know about the disease, and predict the impact of changes to screening practices, medication with certain drugs, and changes in operational procedures and public policy.

2.4 SD Modeling

System dynamics modeling is an analytical modeling approach which originated from the general systems theory approach of von Bertalanffy, and the foundations of SD were laid by Jay Forrester at MIT [8]. As stated in the introduction of this dissertation, the fundamental principle of SD is that the structure of the system will determine its behavior. The way the separate components of any system are related to each other and affect each other will determine the long-run behavior of the system as a whole. The reason SD modeling is important is because this approach can reveal highly counterintuitive types of emergent behavior by the system at hand, and it is necessary to understand the reasons for such unexpected behaviors [8]. SD provides a framework for solving these types of problems.

There are two different sides to SD modeling, the quantitative side and the qualitative side. The qualitative portion involves the construction of a causal loop diagram (CLD), which is a graphic depiction of the manner in which the different components of the system are related. The arrows that relate two components can be signed as either positive or negative. A positive sign indicates that when the level of the state variable for the originating component increases, so does the level of the state variable for the destination component. A negative sign indicates the opposite: when the value of the originating variable increases, the value of the destination component decreases. In essence the signs denote the direction of influence, but do not quantify the magnitude of the influence. The goal of constructing this diagram is to gain a better understanding of the problem through exploring relationships between different variables and the overall structure of the system [51, 91]. Murray [65] defines an SD model that predicts the demand for screening for colorectal cancer. The causal loop diagram from Murray [65] is given in Figure 2-2. The colorectal cancer screening process contains many of the same elements as the breast cancer screening process. That is, the structure of the two systems is made up of very similar components: The role of the patient, the role of the physician, and the role of the health care delivery system.

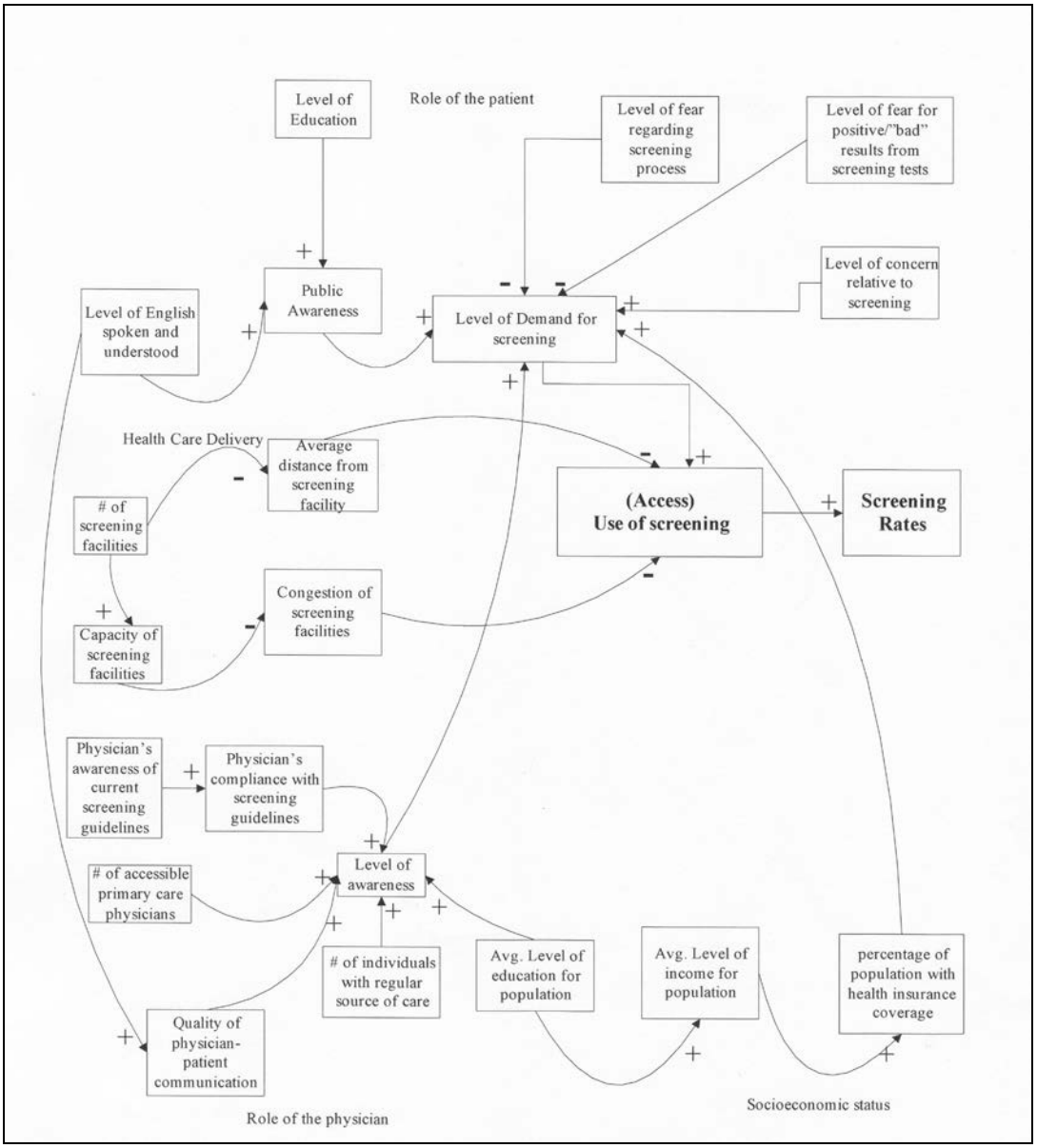


Figure 2-2. Casual Loop Diagram of Colon Cancer Care System [65]

For causal loop diagrams, balancing loops are defined as loops that retain the status quo of the system and keep it operating in steady state. Note carefully that in a CLD, the term loop is synonymous with the graph-theoretic terms circuit and cycle because a CLD is a directed graph [26]. In a CLD, balancing loops can be identified by the use of an odd number of negative signs in given loop. On the other hand, reinforcing loops, or vicious circles,

contain an even number of negative signs and are loops where the system can effectively spiral out of control. The identification of these two types of loops is in itself a valuable learning process. A major benefit of this type of approach is that it often identifies unintended consequences of changes in system variables [8].

The quantitative portion of the SD model development process requires the modeler to convert the causal loop diagram from the qualitative part of the modeling process to a stock-flow diagram. The best way to think of this type of diagram is as a system of water reservoirs connected by pipes and governing valves. The water that flows around the system is analogous to entities in a DES model, except that the aggregation of all relevant entities is represented as a continuous quantity (state variable or stock); thus such a quantity can represent an amount of money, the size of a population or subpopulation, the mass of an inventory of raw material, or the number of items in a finished-goods inventory [8]. The influences in the causal loop diagram need to be converted to mathematical relationships that relate component X to component Y. Determining the form of these mathematical relationships is one of the goals of this research. Figure 2-3 is from a model of hospital admissions [8], and it provides a graphical depiction of a casual loop diagram and the stock-flow diagram that was generated from it.

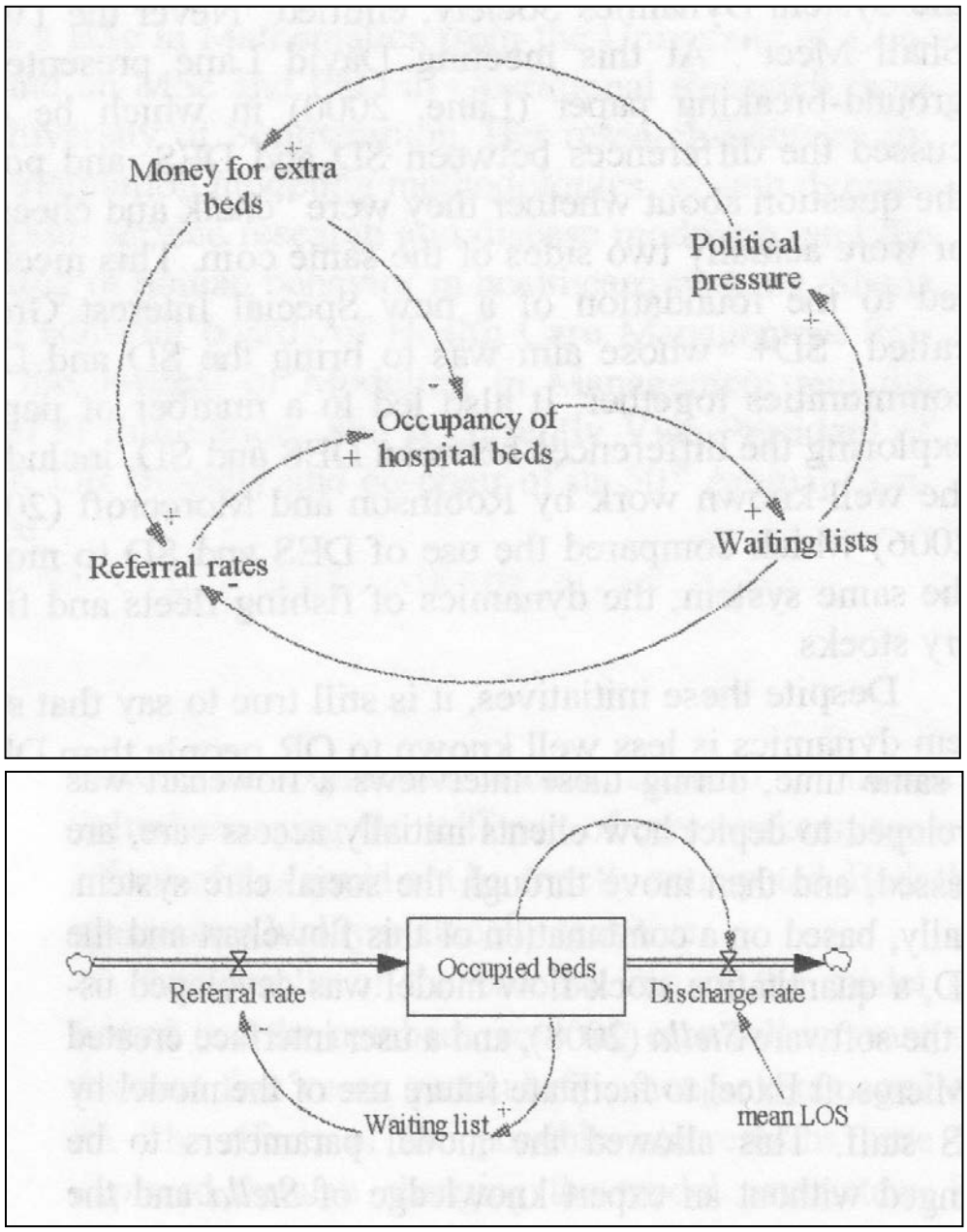


Figure 2-3. Casual Loop and Stock-Flow Diagrams from Hospital Admissions Model [8]

Models are parameterized by defining the initial values of the governing rates, and initial levels of the stocks. After these mathematical relationships are defined, the model can be run and the values of variables over time and under different conditions can be analyzed. It should be noted that SD models are typically deterministic in nature, and do not involve any stochastic components or samples of random variables. Time is handled by a discretization process where the time-step, Δt, is chosen so that all rates are constant over the short period of time Δt.

SD modeling is suited for health care systems because it has the ability to model complex systems whose boundaries extend into other organizations. In health care, it is very rare that a system of interest involves well-defined boundaries and does not interact with the surrounding environment. This is compounded by the fact that health care has multiple stakeholders who have different perspectives regarding different aspects of system operation. In DES, an overwhelming amount of expert opinion is needed to define model variables and parameters and the relationships between these quantities, because the necessary data is often limited or not available at all. Data requirements for SD models are much less since the view of the model is more aggregated and at a higher level, but the required data are harder to get. Another key advantage of SD models is that they typically run very fast, while on the other hand, execution time is a major concern with of DES models.

Similarly, Dangerfield [22] describes the emergence of the use of SD modeling in the European health care arena. The paper describes the origins and methodology of system dynamics and relates it to health care, similar to Brailsford [8]. One of the most important conclusions drawn by the authors of both these papers is that there are a relatively few papers that utilize SD modeling, and that the health care arena has an immense potential to benefit from this type of modeling methodology. Brailsford [8] states “there is a clear potential for system dynamics modeling to be employed in support of health care policy.”

2.5 Integration of DES and SD

Several studies have investigated the differences between DES and SD [8, 11, 19, 63, 89, 93, 96, 97, 103], some on a theoretical and methodological level, and some from the user's perspective. It is important to understand all the underlying differences from a variety of perspectives in order to assess the possible benefit of combining the two methodologies. The DES methodology is based on the premise that the randomness inherent in the system as well as the system structure will determine its long-run behavior, and DES is usually used to model systems in a high level of detail. SD modeling on the other hand, is based on the idea that the structure of system will determine its long-run behavior. Based on an extensive review of the literature and lengthy discussions with faculty specializing in the fields of health care and simulation, we have great reason to expect that combining the DES and SD techniques could be greatly beneficial in gaining insight into complex systems, such as breast cancer screening. The idea being that overall system dynamics can be modeled using SD modeling techniques, and other aspects which occur at discrete points in time can be modeled using DES. Structure and randomness both play a vital role in the long-run behavior of the system, and the integrated screening model combines these two methodologies into a single approach with the ability to capture the dynamics of high-level system strategy (health care delivery and government involvement), as well as low-level randomness and structure involved in the screening process for individual women. We assert that there is the potential for significant benefits to combining these two modeling approaches, and relevant literature on the subject largely supports this conclusion.

Morecraft and Robinson [63] develop an SD model and a DES model of the same system, a simple fishery. The process of building both the SD and DES models is described in detail, although the models themselves are actually very simple. The importance of this work lies in the conclusions reached after developing both types of models, which include the following:

1. “Overall it is apparent that while SD illuminates ‘deterministic complexity,' DES illuminates ‘constrained randomness.' Either or both may be important in understanding and explaining puzzling dynamics. SD and DES should therefore not be seen as opposing modeling approaches, but as complementary.”

2. “The tragedy is that too few researchers realize that both deterministic [SD] and stochastic [DES] models have important roles to play in the analysis of any particular system. Slavish approach to one particular approach can lead to disaster. ... So pursuing both approaches simultaneously ensures that we do not become trapped by either deterministic fantasy or unnecessary mathematical detail.”

Table 2-7 and Table 2-8 [63] provide insights into technical and conceptual differences between the two approaches. These tables reinforce the idea of combining both approaches. These conclusions are in direct support of the concepts behind this research.

Table 2-7. Technical Differences between DES and SD [63]

Issue	Discrete Event Simulation	System Dynamics
1	Systems (such as health care) can be viewed as networks of queues and activities.	Systems (such as health care) can be viewed as a series of stocks and flows.
2	Objects in a system are distinct individuals (such as patients in a hospital), each possessing characteristics that determine what happens to that individual.	Entities (such as patients) are treated like molecules of a fluid, flowing through reservoirs or tanks connected by pipes.
3	Activity durations are sampled for each individual from probability distributions and the modeler has almost unlimited flexibility in the choice of these functions and can easily specify nonexponential dwelling times.	The time spent in each reservoir is modeled as a delay with limited flexibility to specify a dwelling time other than exponential.
4	State changes occur at discrete points of time.	State changes are continuous.
5	Models are by definition stochastic in nature.	Models are deterministic.
6	Models are simulated in unequal times steps, when “something happens.”	Models are simulated in finely sliced time steps of equal duration.

Table 2-8. Conceptual Differences between DES and SD [63]

Issue	Discrete Event Simulation	System Dynamics
Perspective	Analytic, emphasis on detail complexity.	Holistic; emphasis on dynamic complexity.
Resolution of Models	Individual entities, attributes, decisions, and events.	Homogenized entities, continuous policy pressures and emergent behavior.
Data Sources	Primarily numerical with some judgmental elements.	Broadly drawn.
Problems Studies	Operational.	Strategic.
Model Elements	Physical, tangible, and some informational.	Physical, tangible, judgmental and information links.
Agents Represented in Models As…	Decision Makers.	Boundedly rational policy makers.
Clients Find the Model…	Opaque/dark grey box, nevertheless convincing.	Transparent/fuzzy glass box, nevertheless compelling.
Model Outputs	Point predictions and detailed performance measures across a range of parameters, decision rules, and scenarios.	Understanding of structural source of behavior modes, location of key performance measures and effective policy levers.

Tako and Robinson [96] provide an empirical study comparing the DES and SD simulation approaches, and the results are summarized in Table 2-9. The uniqueness in this study is that it allows one group of users to become familiar with and analyze the output of a DES model and an SD model of the same system, the prison system in the United Kingdom, where the populations of the prisons are growing rapidly and a major problem seems to be people who reoffend. Before conducting the experiment, the authors review the relevant literature and compare DES and SD in a number of categories, including: model understanding, complexity, validity, usefulness, and results. The authors conclude that users would find significant differences in the two modeling approaches. The results of the study are actually quite interesting. Due to the limited amount of time available, users could not be exposed to both types of models.

Table 2-9. Comparison of DES and SD Model Literature [96]

Methodology	DES	SD
Model Understanding		
Understanding (parts of) the Model	The client does not understand the underlying mechanics.	Models (links and flows) are transparent to the client.
Animation	Animation and graphic tools help model understanding.	No animation. Visual display of model aids model understanding.
Complexity		
Level of Detail	Emphasis on detail complexity.	Emphasis on dynamic complexity.
Feedback	Feedback is not explicit.	Feedback effects are clear to the client.
Model Validity		
Credibility	Both models are perceived as representative, provide realistic outputs and create confidence in decision-making.	
Model Usefulness		
Learning Tool	Used less as learning tools.	"Learning laboratories" that enhance users learning.
Strategic Thinking	Mostly used in solving operational/tactical issues.	Aid strategic thinking.
Communication Tool	Both models are seen as good communication tools and facilitate communication with the client.	
Model Results		
Nature of Results	Provides statistically valid estimates of systems performance. Results aid in instrumental learning.	Results provide a full picture of the system. Results aid in conceptual learning.
Interpretation of Results	More difficult, requires users to have a statistical background.	Outputs are easily interpreted, little or no statistical analysis is required.
Results Observation	Randomness/variation of results is explicit.	Generally deterministic results, which convey casual relationships between variables.

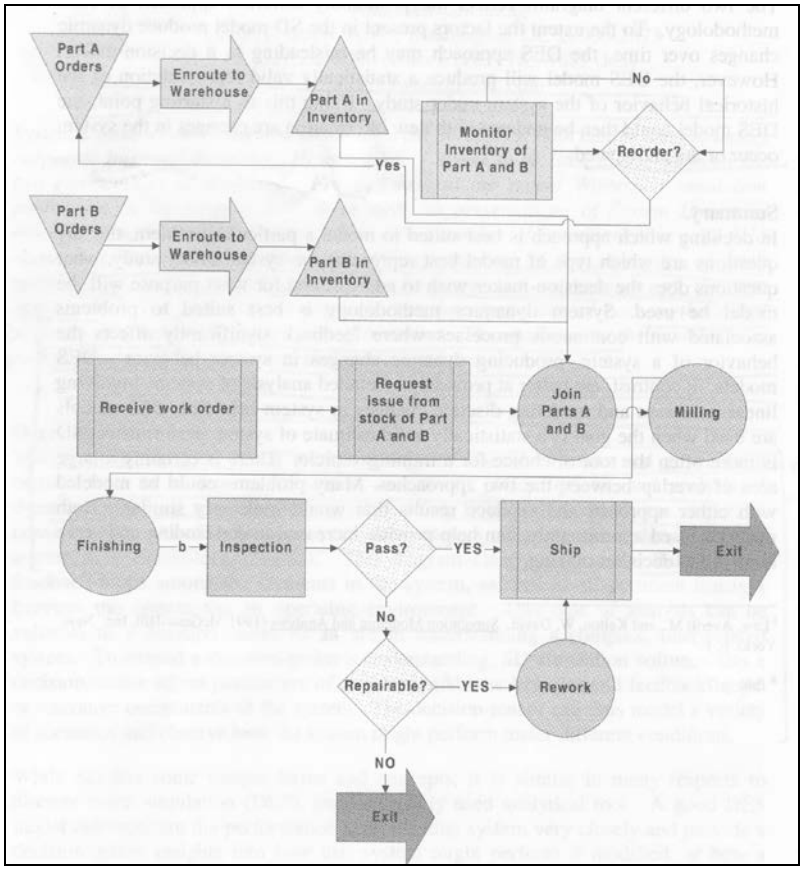
Each group of users was only subjected to one of the two types of modeling, took for granted the features of the model they were given to work with, and were not really able to provide deep insights into the differences in the way users perceive the two simulation modeling approaches. The results of the comparison experiment are given in Table 2-10. Slight differences were found in terms of interpretation and importance of animation, but no real significant differences were identified in the study. This suggests that users are comfortable with both modeling approaches; therefore, the integration of the two approaches might easily

be accepted by typical users. A combined model has the potential to bring the strengths of the two separate concepts together, but this will surely create some unexpected dilemmas about which methodology should prevail when modeling certain system components.

Table 2-10. Comparison Between DES and SD Model Experiments and Results [96]

Model Use	DES	SD
Model Understanding		
Understanding (parts of) the Model	No differences in user's opinions.	
Factors That Helped Understanding	Animation the most important factor.	Accompanying model descriptions most important.
Complexity		
Level of Detail	Similar in level of perceived detail.	
Feedback	Feedback effects more explicit to DES model users..	—
Model Validity		
Representative of Real Problem	—	SD model just more representative as compared to DES model.
Realistic Outputs	Outputs are perceived to be similarly realistic.	
Confidence in Outputs	Similar level of confidence in model outputs.	
Model Usefulness		
Learning	Similar level of learning achieved from using the models.	
Strategic Thinking	Same level of perceived strategic thinking involved.	
Communication of Ideas	Same level of communication perceived to have taken place.	
Model Results		
Instrumental/Conceptual Learning	Both DES and SD aid instrumental learning. SD aids in conceptual learning to a higher extent.	
Interpretation of Results	DES models were more difficult to interpret.	-
Attitude When Interpreting Results	No differences in user's attitude.	

Sweester [93] compares the two modeling approaches from the perspective of evaluating businesses' decision making, and the purpose is to identify the strengths and weaknesses of both approaches. Another visual representation of the differences between the two approaches is shown in Figure 2-4, which presents a typical diagram of production process from both the SD and DES perspectives.



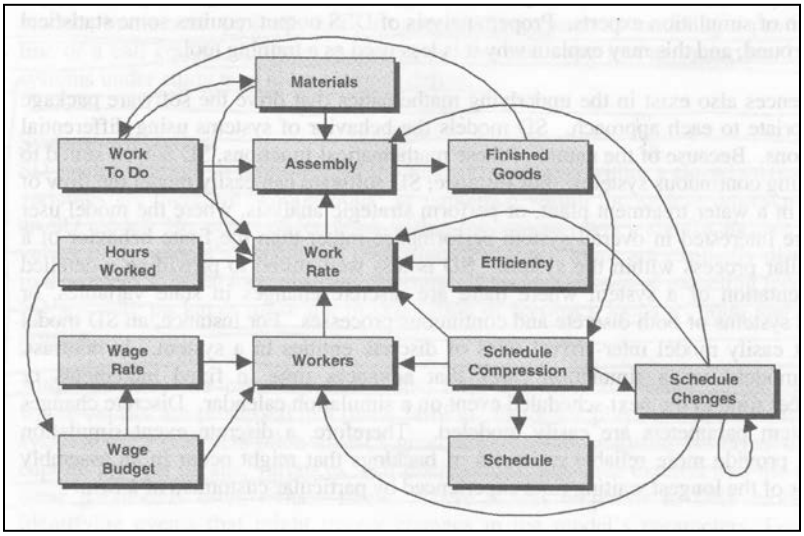


Figure 2-4. DES Model (Top) and SD Model (Bottom) of Same Production Process [93]

The study concludes SD is suited to modeling problems with continuous variables whose interactions significantly affect the behavior of the system, while DES is used to develop statistically valid estimates of system performance. What the paper fails to explore is the idea that the system could contain continuous variables whose feedback significantly affects the behavior of the system, and there could be a need to develop a statistically valid estimate of system performance. The relevant literature clearly reveals that many aspects of breast cancer screening present both of these characteristics and that the system is suitable for an integrated model.

Tako and Robinson [96] fulfill one of their goals for future research from [97] by observing users as they build SD and DES models of the same UK prison system, and noting the differences in model development from a user standpoint. Seven specific topics were observed: problem structuring, conceptual modeling, data inputs, model coding, validation and verification, results and experimentation, and implementation. The results of the study suggest that all modelers switch continuously between the seven topics, but DES modelers seem to follow a more linear progression than SD modelers. Model coding is the central topic for DES modelers, while conceptual modeling is primarily followed by model coding for SD modelers. However, the major finding was that neither approach revealed users that followed these seven topics in any particular order. Other findings contribute to the framework for integrating the two approaches, such as “most modelers did not consider alternative modeling approaches before modeling. This conveys that modelers use the simulation approach they are more familiar with rather than the approach that is most appropriate to the situation.” It is reasonable to believe certain systems could benefit from a combined approach.

Brailsford et al. [11] compares the two approaches in the modeling of an emergency department (ED) and suggests that the relative strengths and limitations of both approaches are in fact complementary. “We argue that the integrated simulation models can assist in dealing with the efficiency issues within the ED walls, but can also provide a better understanding of system-wide factors beyond the immediate control of ED managers.” They

describe three potential benefits of the integrated approach: inward and outward looking, analytical insights at all levels of the system, and manager-friendly models. The conclusion of the paper states: “We believe there is much to be gained by integrating these two approaches into a unified framework for both conceptual and implementation purposes. ... We recommend future research into conceptual and implementation issues of this area.” This is further support for the integration of these two approaches.

Chahal and Eldabi [19] explore reasons that DES and SD could potentially be combined and predict some potential benefits of this type of model. This paper provides a remarkably clear and compelling argument for why DES and SD should be combined. The benefits and reasons are nearly identical to those discussed in [11]. The authors suggest using the outputs of a DES model as the inputs of an SD model, and vice versa; and they also admit that they have not developed a hierarchical structure to specify the way the two models would interact. There may be some merit to this type of approach, but there seems to be much greater value in integrating these two approaches so that elements of both types of simulation models can be represented simultaneously. This provides evidence of the fact that a gap exists in simulation methodology that has not yet been addressed by researchers.

Ventkatesawaran [103] develops an SD model of a production planning system and uses the output from that model as the input to a DES model of the shop floor. This model is a first-step to combining these two approaches, but these two models are built in different software packages and are not integrated together in a single model. This dissertation is a direct extension of this work and seeks to integrate these two approaches into a single model where both models run concurrently and are linked.

To the best of our knowledge, a simulation methodology which integrates DES and SD modeling approaches into a single model has not yet been developed. The problem statement and research questions addressed are presented in the final sections of this chapter.

2.6 Research Motivation

In this section we elaborate the motivation for this research beyond the brief discussion given in Chapter 1. By enumerating the main problems in breast cancer screening and in simulations of the breast cancer screening-and-treatment process, we also identify the main research questions to be addressed in the dissertation.

2.6.1 The Problems in Breast Cancer Screening

The problem in breast cancer screening has been introduced earlier sections, but will be explored in more detail in this section. The aging U.S. population, the lack of screening guidelines for older women, and the prominence of breast cancer among women in the United States are three of the major factors that contribute to the need for research in breast cancer screening [1, 28, 42, 58, 77]. With regard to breast cancer screening, there is little guidance for older patients [42, 45, 77, 86].

Currently, there are no well-established guidelines for women 70 and older (Table 2-6), nor are there established clinical guidelines for the age at which screening for breast cancer should be discontinued. The vagueness of the current guidelines leaves much to be desired in terms of an effective decision making process for the older patient and her physician. The major goals of the simulation model are to determine the most effective screening policy for women over 65, which could be dependent on age or specific individual risk factors. The appropriate age at which to stop screening is also of interest. The Carolina Mammography Registry (CMR), the Surveillance, Epidemiology and End Results (SEER) Program (www.seer.cancer.gov), the Breast Cancer Surveillance Consortium (BCSC), and relevant literature provide reliable sources of information that are used for model parameterization and validation. Medicare claims data assist in determining cost-of-care estimates, since most women over 65 have Medicare as their primary insurance.

The majority of the studies in this area have ignored factors that will be considered in this research, do not focus on older patients, or contain assumptions that may have become outdated with technological advances. Previous studies have not focused on policy evaluation and development, whereas that will be the primary objective of this research. In addition, the American Cancer Society has stated there is no reason to believe there will be more randomized clinical trials for breast cancer screening in the near future [1, 21, 45, 86]. Upcoming clinical trials will be pharmaceutical clinical trials as opposed to clinical trials for breast cancer screening Considerable controversy remains regarding how frequently to recommend screening and at what age to stop screening older women. There is considerable evidence that physicians have difficulty differentiating the highly heterogeneous older patient population. Given that women over the age of 65 will become the most prevalent patient cohort in the breast cancer population, it is imperative that research into the optimal screening strategies for this patient population be performed. Next, we state the problems from a simulation modeling perspective.

2.6.2 The Problems in Simulation Modeling

Jun et al. [47] describes specifically how discrete event simulation has been used to optimize patient flow and allocation of scarce resources with health care systems. The results of the survey show that there is a lack of papers that address integrated systems. Another conclusion is that this type of modeling has been made possible by recent advances in software. This is important, because the tools to proceed with this integrated type of model have only recently become available. Arena 13.9 [48] is an extremely versatile simulation program that has the ability to model discrete and continuous variables simultaneously, as well as use arbitrarily complex expressions in routing entities through the model. This software will be used in the development of both the natural history simulation model and the integrated screening model. The fact that simulation software that has the ability to handle

this type of integrated system is already available is a result of advancing technology. Arena also has the ability to interface with Visual Basic and C++ programming languages.

Several problems arise when the concept of integrating these two approaches is pursued, including:

1. What components of the overall breast cancer screening-and-treatment system should be modeled using the DES approach and which should be modeled using SD?
2. Why and how does combining these two approaches result in a model which more accurately depicts the overall system, and how can this benefit be quantified?
3. What specific links between the components of the DES submodel and the SD submodel should be present for the most realistic representation of the system?
4. What information will be shared by the two approaches as the simulation runs both simultaneously?
5. What are the inputs for the SD components and for the DES components of the system?
    a. Are they related?
    b. Will choosing whether a component is to be modeled using SD or DES be based upon the available data to model that system component?
6. What will the outputs of the system be that provide the analyst with better insights compared with developing and analyzing these two types of models separately?
7. What types of "prefabricated" modeling tools and components should be developed to facilitate rapid building of simulation models for design and evaluation of breast cancer screening policies?
    a. Submodels of the process-interaction part of the overall model representing operations or flows of people, information, etc., that occur in essentially similar form at numerous points in the overall model; and

b. New output-analysis procedures that have been adapted to the specific requirements of simulation-based evaluation of breast cancer screening policies.

Our research addresses these questions. However, capturing the big picture and the small details of a system at the same time is intrinsically difficult, and can pose unanticipated challenges during the development of the integrated screening model.

2.6.3 Research Questions Addressed

Along with the problems and research questions already stated, this research seeks to address the following questions:

1. How does using a combined DES/SD model help to capture the actual behavior of the breast cancer screening process?
2. What aspects of the breast cancer screening process should be modeled with DES and what aspects should be modeled with SD?
3. How will the SD and DES submodels be linked?
4. How can we measure the risk of having a cancer detected based on specific breast cancer risk factors?
5. How can we attain a database of women with appropriate breast cancer risk factors to sample from so that we can determine incidence on an individual basis?
6. How can we model the disease within the body?
7. How can we determine the stage of breast cancer at diagnosis?
8. What performance measures typically define the effectiveness of a breast cancer screening policy?
9. How can the effect of policy changes on the performance of breast cancer screening best be measured?

10. Which screening policy is most effective for older US women in terms of the performance measures referenced in Question 8?
11. How can the Tejada model be validated using historical data?
12. How can the best screening policy be identified from the (large) set of feasible screening policies?
13. How can we honestly represent the randomness and intrinsic uncertainty in the simulation-generated estimates of expected system performance?

These questions characterize the key goals of our research. This dissertation will be the first step of many in combining these two simulation methodologies in the health care modeling arena. To this point, we have provided background on simulation modeling methodology, health care, and specifically breast cancer. The following three chapters review the models used to address the problems and research questions we have just discussed. We begin with our natural history simulation model.

CHAPTER 3: THE NATURAL HISTORY SIMULATION MODEL

This chapter focuses on the detailed development and results of the natural history simulation model. As stated previously, this model is made up of a number of interacting submodels, namely: (a) the cancer incidence submodel; (b) the disease progression submodel; (c) the survival and mortality submodel; and (d) the population growth submodel. The most important input to this model is the sampling database that contains information about breast cancer risk factors from actual women in the designated population of US women 65 and older. All of the submodels and their components are explained in detail in this chapter.

This model establishes a baseline on the benefits of screening and treatment by determining the earliest a cancer could be detected for each woman in the population with perfect annual screening for all women, and it assumes all cancers remain untreated. The time-step for this model is one year, i.e., we simulate all the events for a given year, then we move to the next year and simulate all the events for that year, and this process repeats until we reach the end of the time horizon. This approach to the operation of the natural history model makes the most effective use of the available literature, because the Barlow risk model [3] predicts a woman's risk each year of being diagnosed with breast cancer as a function of her risk factors given that she received a mammogram; and life-tables with breast cancer deaths removed [79] predict the probability of surviving one more year as a function of a woman's current age. To establish a baseline that will later enable us to investigate the results of using the best-case scenario, we assume in the natural history DES model that every woman is screened every year with perfect compliance until one of the following occurs: (a) detection for a woman who develops breast cancer; (b) death from a cause other than breast cancer; or (c) the end of the simulation is reached at the year 2020 for a woman who survives to that time and does not develop breast cancer. However, when a cancer is detected, it is not treated. The cancer is allowed to progress, and it can eventually cause a breast cancer–related death; or a natural death from other causes may occur before death from breast cancer. It is also possible that the woman lives beyond the year 2020. The integrated screening model

considers screening at less frequent intervals and as a function of individual risk factors, postdiagnosis (adjuvant) treatment, costs, and survival after diagnosis.

First, the natural history DES model creates the initial population representative of US women at least 65 using a random sampling procedure. The model assigns each woman the attributes (see Table 3-1 for a list of attributes) of a randomly sampled woman in the data set provided by the Breast Cancer Surveillance Consortium (BCSC) [15], and then that woman enters into the disease submodel. The age distribution of the initial simulated population represents that of US women who are at least 65, and was derived from 2000 US census data [43]. Following the creation of the initial population and the assignment of risk-factor attributes for each individual in that population, the risks of breast-cancer detection and natural death are simulated in one-year time-steps until the occurrence of non–breast cancer related death, breast cancer–related death, or the end of the simulation in the year 2020. Each woman in the model can only have breast cancer detected once, and that event is regarded as the first occurrence of breast cancer in her lifetime. The population of US women at least 65 is currently growing rapidly, so it was essential to include a submodel for population growth that adds women who are turning 65 to the current population each year. The initial simulated population size in the initial model year 2001 is 20,582, which is 0.1% of the number of women over 65 in the United States in 2001. This population grows to over 30,000 by 2020. For each of the 10 simulated populations used to initialize an independent replication (run) of the natural history model, 50,000 samples are taken with replacement from the database, which has about 250,000 entries. The natural history DES model was created using Arena 13.9 by Rockwell Software [48].

Figure 3-1 shows the possible transitions among health states for each individual woman in the simulation model. All women who are diagnosed with invasive cancer are followed past 2020 until their death from breast cancer or natural causes. Women who are diagnosed with DCIS are not "cancer free", but they cannot have a death from breast cancer since DCIS cannot progress, the only possible transition from DCIS is to other cause death.

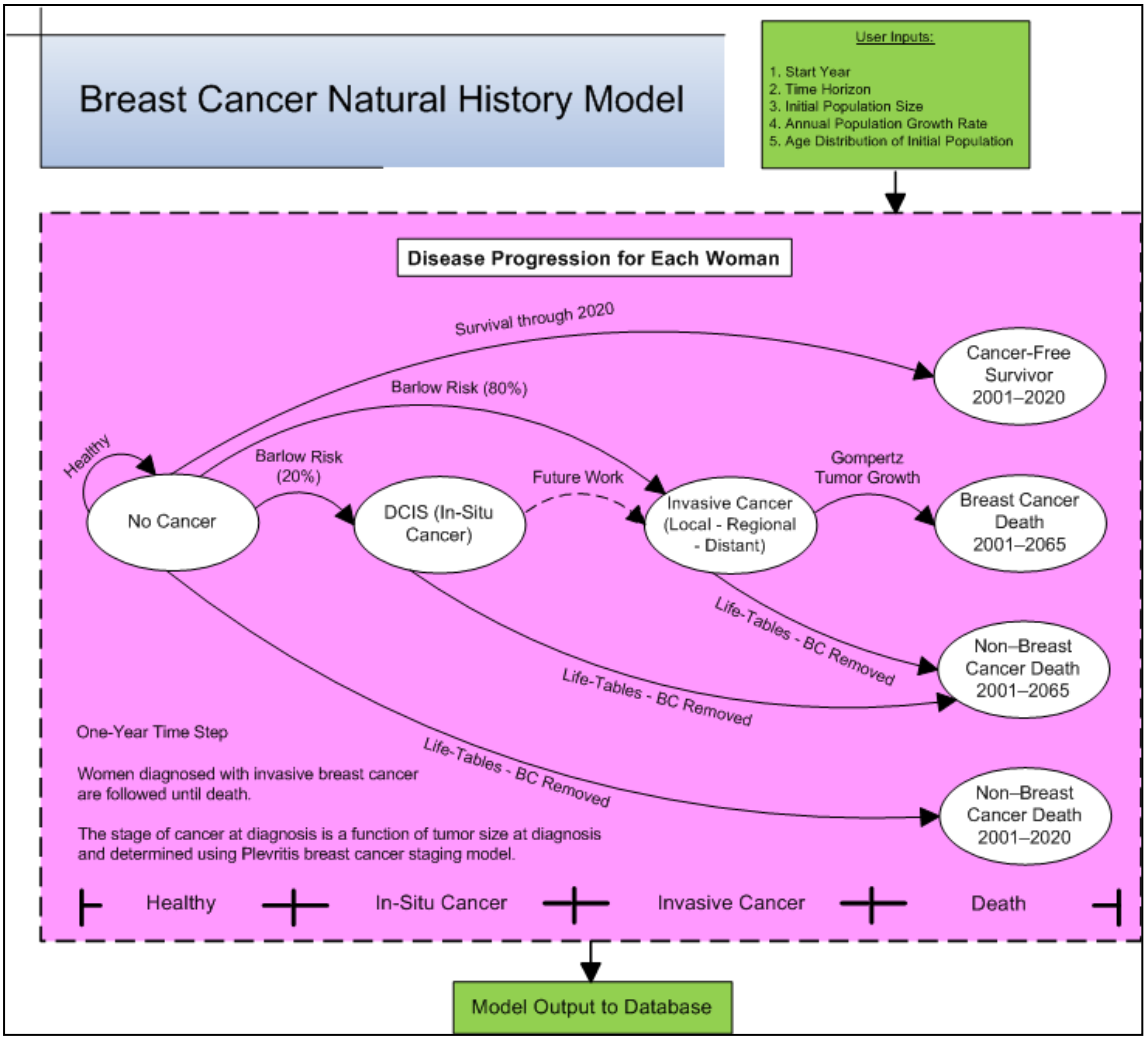


Figure 3-1. Health-State Transitions in the Natural History Model

3.1 Cancer Incidence Submodel

The cancer incidence submodel consists of the Barlow breast cancer risk model being used in conjunction with each woman's risk factors. Risk factors for each woman were sampled from the BCSC data set at the very beginning of the model. This allows the simulation to estimate the probability of being diagnosed with breast cancer each year given that a mammogram was received in that year (this is the condition imposed by the Barlow risk model). Each year this probability is computed for each woman, and this is the method used for determining which women in the model will get cancer and the year they were

diagnosed. Cancer incidence can be defined as the number of women who get cancer over a given time period, typically one year. These raw numbers are usually converted into rates per 100,000 women, which is the format SEER uses to present incidence data that we use for validation. The remainder of this section presents the details of the BCSC data and the Barlow risk model.

3.1.1 BCSC Data

A key component of this research is locating a data set containing risk factors similar to those of women in the US population. For example, age, race, body mass index (BMI), breast density, and many other factors are important in determining a woman's risk of developing breast cancer. We were required to submit a formal research proposal and request for data to the BCSC in order to obtain the individualized data set required for this model, and the BCSC provided a data set containing risk factors for slightly over one million women that has been de-identified to protect the confidentiality of the women, the radiology facilities, and the mammography registries—i.e., the data do not include any dates, origin of the data, patient identifiers, or the assessment (outcome) of the index screening mammogram [15]. Since the data set consists of information from seven mammography registries in different locations across the United States, it is believed to be representative of the women in the US population. Based on the attributes in this data set, the Barlow risk model [3] can be used to calculate an individual woman's risk of developing invasive breast cancer or DCIS within the next year, given that she received a mammogram. The risk factors used to calculate risk are assumed to stay constant throughout each woman's lifetime, with the exception of age and menopausal status.

The following is a description of the BCSC that has been taken from their web site:

> The Breast Cancer Surveillance Consortium (BCSC) is a research resource for studies designed to assess the delivery and quality of breast cancer screening and related patient outcomes in the United States. The BCSC is a collaborative network of seven mammography registries with linkages to tumor and/or pathology registries. The network is supported by a central Statistical Coordinating Center. Currently, the Consortium's database contains information on 7,521,000 mammographic examinations, 2,017,869 women, and 86,700 cancer cases (72,800 invasive cancers and 13,800 In Situ). A full description of the history and early work of the BCSC can be found at: http://breastscreening.cancer.gov/espp_report.html.

None of the women who participated had a previous diagnosis of breast cancer, but all of them had undergone breast mammography at some point in the previous five years. This limits the application of this research to women who have no prior history of breast cancer. Cancer registry and pathology data were linked by the BCSC to the data on mammography and incidence of invasive breast cancer or DCIS. Table 3-1 contains a list of the breast cancer risk factors for which data was collected during the BCSC study, and the appropriate predictor (independent) variables in the logistic regression equations of the Barlow risk model.

Table 3-1. Risk Factor Attributes of Individuals in the BCSC Risk Model Data set

Factor #	Factor Symbol	Factor Name	Columns	Factor Coding
0	None	menopaus	1	0 = premenopausal; 1 = postmenopausal or age>=55 ; 9 = unknown
1	x_1	agegrp	3–4	1 = 35–39; 2 = 40–44; 3 = 45–49; 4 = 50–54; 5 = 55–59; 6 = 60–64; 7 = 65–69; 8 = 70–74; 9 = 75–79; 10 = 80–84
2	x_2	density	6	BI-RADS breast density codes 1 = Almost entirely fat; 2 = Scattered fibroglandular densities; 3 = Heterogeneously dense; 4 = Extremely dense; 9 = Unknown or different measurement system
3	x_3	race	8	1 = white; 2 = Asian/Pacific Islander; 3 = black; 4 = Native American; 5 = other/mixed; 9 = unknown
4	x_4	Hispanic	10	0 = no; 1 = yes; 9 = unknown
5	x_5	bmi	12	Body mass index: 1 = 10–24.99; 2 = 25–29.99; 3 = 30–34.99; 4 = 35 or more; 9 = unknown
6	x_6	agefirst	14	Age at first birth: 0 = Age < 30; 1 = Age 30 or greater; 2 = Nulliparous; 9 = unknown
7	x_7	nrelbc	16	Number of first-degree relatives with breast cancer: 0 = zero; 1= one; 2 = 2 or more; 9 = unknown
8	x_8	brstproc	18	Previous breast procedure: 0 = no; 1 = yes; 9 = unknown
9	x_9	lastmamm	20	Result of last mammogram before the index mammogram: 0 = negative; 1 = false positive; 9 = unknown
10	x_{10}	surgmeno	22	Surgical menopause: 0 = natural; 1 = surgical; 9 = unknown or not menopausal (menopaus=0 or menopaus=9)
11	x_{11}	hrt	24	Current hormone therapy: 0 = no; 1 = yes; 9 = unknown or not menopausal (menopaus=0 or menopaus=9)

3.1.2 The Barlow Risk Model

The Barlow breast cancer risk model [3] was developed using the data collected by the BCSC during their six-year study. The risk model consists of logistic regression equations that predict the probability of being diagnosed with breast cancer (invasive or DCIS) in the next year for premenopausal and postmenopausal women, respectively. SAS statistical software was used to conduct the analysis. We verified this model by using SAS code made publicly available by the authors (Appendix A) to reproduce the statistical model and the corresponding results (Appendix B).

The dependent (response) variable is a binary variable for which the values 0 and 1 denote the absence and presence of breast cancer, respectively; and the predictor variables (factors or "covariates") are the categorical independent variables presented in Table 3-1. Note that the BCSC provided data for women ages 35 and older, and we incorporated the possibility of including younger women in the logic of many of the submodels. For premenopausal and postmenopausal women, Barlow [3] formulated separate logistic regression equations to predict the risk of being diagnosed with breast cancer within one year.

Logistic regression [13] is typically used to predict the probability of occurrence of a specific event using the logistic function. There are several reasons for using the logistic function and logistic regression to predict probabilities. The logistic function is given in Equation (3.1). It is easy to see that the function's domain consists of all real numbers, while its range is restricted between zero and one. This form makes logistic regression appealing for predicting probabilities. The logistic function is of the form:

$$f(L) = \frac{e^L}{1+e^L} \quad \text{for } -\infty < L < \infty. \tag{3.1}$$

In logistic regression, the probability of interest, *p*, is related to the predictor variables using the natural log function. Typically, Equation (3.2) is referred to as the "logit" function in the regression equation:

$$L = \text{Logit(p)} = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n. \qquad (3.2)$$

where x_1 , ... , x_n are the independent variables whose values (coded factor levels) are used to estimate the relevant probability *p*. In order to calculate the predicted probability of the event of interest based on the risk factors, Equation (3.2) must be solved for *p*. Equation (3.3) shows the results of such algebraic manipulation. Rearranging (3.2), we see:

$$e^L - p - pe^L = 0 \Rightarrow p = \frac{e^L}{1+e^L} = f\left(L\right). \qquad (3.3)$$

Barlow [3] created two separate logistic regression models to predict the risk of being diagnosed with breast cancer within one year. One model is used for premenopausal women, and the other is used for postmenopausal women. Cancer rates suggest that risk patterns are not the same for premenopausal and postmenopausal women, and the risk factors which are significant for developing cancer are quite different for these two groups of women. All eleven of the risk factors in Table 3-1 were tested for significance in both regression models. There are no interaction terms included in order to keep the model simple enough to be used practically, and the authors state that the interaction terms were not significant. In order to verify the models, 75% of the random sample (the training data set) was used to estimate the statistical models; and the prediction was tested in the remaining 25% of the random sample (the validation data set). The concordance value (*c*-statistic) is used to determine the quality of the prediction for each individual woman. The *c*-statistic is a value that varies from 0.5 (discriminating power not better than chance) to 1.0 (perfect discriminating power). It can be interpreted as the percent of all possible pairs of cases in which the model assigns a higher probability to a correct case than to an incorrect case. The entire output from the analysis is

available in Appendix B. The *c*-statistic for the Barlow risk model is 0.624, which the authors suggest is acceptable in practice.

It is important to note that a very strict *p*-value criterion of $p < 0.0001$ is required for a risk factor to be statistically significant and included in the model. The reason the *P*-value criterion is so strict is because of the large size of the data set. Small differences could appear significant at the .05 level due to the size of the data set, so a more stringent criterion was chosen by Barlow [3]. For premenopausal women there were only four statistically significant risk factors: age, family history, breast density, and history of breast procedures. The following risk factors were not statistically significant: race ($p = 0.40$), ethnicity ($p = 0.41$), age at first birth ($p = 0.48$), BMI ($p = 0.89$), and the result of previous mammogram ($p= 0.0018$). Equation (3.4) is the regression equation for premenopausal women, which was formed using the factors given in Table 3-1. Equation (3.5) is the explicit equation that can be used to calculate the one-year risk when given the values of the regression coefficients and the values of the categorical variables for a specific woman. The logit for premenopausal women can be calculated according to Equation (3.6), noting that the double equals sign is a true/false statement that evaluates to either zero or one.

$$L_{\text{PRE}} = \text{Logit}(p_{\text{PRE}}) = \ln\left(\frac{p_{\text{PRE}}}{1-p_{\text{PRE}}}\right) = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_8 x_8 + \beta_9 x_9, \tag{3.4}$$

where the independent variables x_2, x_3, x_8, and x_9 are defined in Table 3-1,

$$p_{\text{PRE}} = f(L_{\text{PRE}}) = \frac{\exp(L_{\text{PRE}})}{1+\exp(L_{\text{PRE}})} = \frac{1}{1+\exp(-L_{\text{PRE}})}, \tag{3.5}$$

$$
\begin{aligned}
L_{\text{PRE}} &= -7.4979 + (\text{AgeGroup} == 1)*(0) + (\text{AgeGroup} == 2)*(0.2046) \\
&+ (\text{AgeGroup} == 3)*(0.6388) + (\text{AgeGroup} == 4)*(0.8574) \\
&+ (\text{Previous Procedure} == 0)*(0) + (\text{Previous Procedure} == 1)*(0.3873) \\
&+ (\text{Previous Procedure} == 9)*(-0.1431) + (\text{First Degree} == 0)*(0) \\
&+ (\text{First Degree} == 1)*(0.4306) + (\text{First Degree} == 2)*(0.7455) \\
&+ (\text{First Degree} == 9)*(-0.1134) + (\text{Breast Density} == 1)*(0) \\
&+ (\text{Breast Density} == 2)*(0.6818) + (\text{Breast Density} == 3)*(1.2067) \\
&+ (\text{Breast Density} == 4)*(1.3694) + (\text{Breast Density} == 9)*(1.1966).
\end{aligned}
\tag{3.6}
$$

For example, consider a premenopausal woman who is age 43 (category 2), has one relative with cancer (category 1), whose breasts are heterogeneously dense (category 3), and who has had no previous breast procedures (category 0). We can calculate her risk of being diagnosed with cancer in the next year. First we must calculate the value of the logit, shown in Equation (3.7), then we can calculate the desired probability, shown in Equation (3.8),

$$
L_{\text{PRE}} = -7.4979 + (1)*(0.2046) + (1)*(0) + (1)*(0.4306) + (1)*(1.2067) = -5.656, \tag{3.7}
$$

$$
p_{\text{PRE}} = \frac{1}{1+\exp(-L_{\text{PRE}})} = \frac{1}{1+e^{5.656}} = 0.00348. \tag{3.8}
$$

For postmenopausal women there were eleven statistically significant risk factors from Table 3.1: age x_1 , breast density x_2, race x_3, ethnicity x_4, BMI x_5, age at first birth x_6, family history x_7, history of breast procedures x_8, result of last mammogram x_9, surgical menopause x_{10}, and hormone replacement therapy x_{11}. Equation (3.9) is the regression equation for postmenopausal women, which was formed using the factors given in Table 3-1. Equation (3.10) is the explicit equation that can be used to calculate the one-year risk given the values of the regression coefficients and the values of the categorical variables for a specific woman. The logit for postmenopausal women can be calculated according to Equation (3.11):

$$L_{\text{POST}} = \text{Logit}(p_{\text{POST}}) = \ln\left(\frac{p_{\text{POST}}}{1-p_{\text{POST}}}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11}, \tag{3.9}$$

$$p_{\text{POST}} = f(L_{\text{POST}}) = \frac{\exp(L_{\text{POST}})}{1+\exp(L_{\text{POST}})} = \frac{1}{1+\exp(-L_{\text{POST}})}, \tag{3.10}$$

$$
\begin{aligned}
L_{\text{POST}} &= -7.0713 + (\text{AgeGroup} == 3)*(0) + (\text{AgeGroup} == 4)*(0.283) \\
&+ (\text{AgeGroup} == 5)*(0.6743) + (\text{AgeGroup} == 6)*(0.8196) \\
&+ (\text{AgeGroup} == 7)*(0.9034) + (\text{AgeGroup} == 8)*(1.026) \\
&+ (\text{AgeGroup} == 9)*(1.1086) + (\text{AgeGroup} == 10)*(1.2033) \\
&+ (\text{Ethnicity} == 0)*(0) + (\text{Ethnicity} == 1)*(-0.3072) \\
&+ (\text{Ethnicity} == 9)*(-0.0641) + (\text{Race} == 1)*(0) \\
&+ (\text{Race} == 2)*(-0.2251) + (\text{Race} == 3)*(0.0935) \\
&+ (\text{Race} == 4)*(-0.6108) + (\text{Race} == 5)*(-0.0269) \\
&+ (\text{Race} == 9)*(-0.00037) + (\text{BMI} == 1)*(0) + (\text{BMI} == 2)*(0.1337) \\
&+ (\text{BMI} == 3)*(0.2463) + (\text{BMI} == 4)*(0.3827) + (\text{BMI} == 9)*(0.0295) \\
&+ (\text{AAFB} == 0)*(0) + (\text{AAFB} == 1)*(0.1871) + (\text{AAFB} == 9)*(0.0203) \\
&+ (\text{Previous Procedure} == 0)*(0) + (\text{Previous Procedure} == 1)*(0.262) \\
&+ (\text{Previous Procedure} == 9)*(0.0585) + (\text{First Degree} == 0)*(0) \\
&+ (\text{First Degree} == 1)*(0.2722) + (\text{First Degree} == 2)*(0.5080) \\
&+ (\text{First Degree} == 9)*(-0.0488) + (\text{Hormone Therapy} == 0)*(0) \\
&+ (\text{Hormone Therapy} == 1)*(0.171) + (\text{Hormone Therapy} == 9)*(0.124) \\
&+ (\text{Surg Meno} == 0)*(0) + (\text{Surg Meno} == 1)*(-0.1701) \\
&+ (\text{Surg Meno} == 9)*(-0.0603) + (\text{Last Mamm} == 0)*(0) \\
&+ (\text{Last Mamm} == 1)*(0.5241) + (\text{Last Mamm} == 9)*(0.1840) \\
&+ (\text{Breast Density} == 1)*(0) + (\text{Breast Density} == 2)*(0.7377) \\
&+ (\text{Breast Density} == 3)*(1.0799) + (\text{Breast Density} == 4)*(1.1482) \\
&+ (\text{Breast Density} == 9)*(1.0449).
\end{aligned}
\tag{3.11}
$$

The example given for premenopausal women can easily be extended to postmenopausal women; the probability of being diagnosed with cancer at some point within the next year is calculated in the exact same way using the postmenopausal model equations.

The Gail model [35] is another popular breast cancer risk model that is prevalent in breast cancer literature and is used by the National Cancer Institute to educate the public. More information can be found in [35] and at http://www.cancer.gov/bcrisktool/about-

tool.aspx#gail. The Gail Model predicts the probability that a woman will be diagnosed with cancer within the next five years. This is different from the Barlow model, which has a one-year time span. The risk factors used in the Gail model are also different than those used in the Barlow model. The Gail model uses the following risk factors: age, personal history, age at the start of menstruation, age at the first live birth, and the family history of breast cancer. Unfortunately, we were unable to find an up-to-date, readily available data set that contains these attributes for a group of older US women. The Gail also model uses risk factors that were known in 1989. By contrast, the Barlow risk model was published in 2006; and its associated study was started in 1996, when recent information about new relevant risk factors had been discovered. For the reasons discussed, the Barlow model was chosen for use in the natural history simulation model, but it is important to explore and acknowledge other valid models for calculating risk also exist in the literature.

3.1.2.1 DCIS Risk

The Barlow risk model provides us with the ability to determine women in the population who would be diagnosed with DCIS given that they received a mammogram. However, it does not provide us with a method of modeling the time-dependent risk of DCIS progressing to invasive breast cancer. According to the data collected during the BCSC study, 20% of breast cancer diagnoses were DCIS and the other 80% were invasive breast cancer. Thus, when the risk model indicates cancer will be diagnosed for a specific woman in the next calendar year, we assume 20% of the cancers are DCIS. After consulting with breast cancer experts such as my cochair Dr. Julie Ivy [59], Stanford professor Dr. Sylvia Plevritis [75], and University of Michigan physician Dr. Kathleen Diehl, we determined that the behavior of DCIS and its progression to invasive cancer are not well understood at this time; and we have chosen to leave that area for future research in hopes that more knowledge about DCIS will be available in the future. Despite not being able to track DCIS progression to invasive cancer, we can estimate incidence in both the natural history model and the integrated screening model. A woman who is diagnosed with DCIS in the natural history model either survives until the end of the simulation or dies from some other cause; but her

DCIS cannot progress to invasive cancer in the natural history model so that there is no possibility of breast cancer being the cause of death.

3.2 Disease Progression Submodel

The disease progression submodel consists of a Gompertzian tumor growth model [68] that tracks the size of the primary tumor from onset to death, and the Plevritis stage progression model [74] that is used to determine the stage of invasive breast cancer at diagnosis as a function of primary tumor size. The Gompertzian tumor growth model was found in the literature and was validated against three data sets for untreated tumor growth; however, we extended the basic model in several ways that we felt made the model more realistic and better suited to our purposes. The Plevritis model allows us to determine whether the invasive cancer is in the local, regional, or distant stage at the time of diagnosis. The stage of invasive breast cancer at diagnosis has a dramatic effect on the type of treatment and on the patient's quality of life for the rest of her life. Different screening policies will lead to different stage distributions at diagnosis, and this is an important measure of screening effectiveness that we are able to capture in the Tejada model. We are also able to validate the natural history model by comparing the stage distribution resulting from perfect screening to SEER's reported 65-and-older stage distribution, which reflects the actual level of screening in the population. The remainder of this section presents the details of the Gompertzian tumor growth model and the Plevritis stage progression model.

3.2.1 Gompertzian Tumor Growth Model

One of the most important pieces of the natural history model is the tumor growth model, which describes the growth of cancer over time within the body. A Gompertz model proposed by Norton [69] of human breast cancer growth is used to describe the size of a single breast cancer tumor at time t by $N(t)$, the number of malignant cells in the tumor at time t. Other tumor growth models [85, 88] were considered for use in the natural history model, but Norton's approach was chosen because there are abundant data that suggest breast

cancer growth in an individual woman can be accurately represented by a general Gompertz function [61, 69, 72, 76, 92]. Equation (3.12) was derived from Norton [69] and describes tumor growth from some initial tumor size $N(t_1)$ at time t_1 to the size $N(t_2)$ at some future time t_2,

$$N(t_2) = N(t_1)\exp\left(k\left(1-e^{-b(t_2-t_1)}\right)\right), \quad \text{for } (t_2 \geq t_1), \tag{3.12}$$

where:

$N(t_1) =$ Individual's tumor size (cells) at the time t_1,

$N(t_2) =$ Individual's tumor size (cells) at the time t_2,

$N(\infty) =$ Maximum size of tumor (cells),

$N_{\mathrm{L}} =$ Tumor size at which breast cancer is considered lethal,

$b =$ Tumor growth constant,

$t = t_2 - t_1 =$ Time delay between tumor sizes $N(t_1)$ and $N(t_2)$ (months),

$$k = \ln\left(\frac{N(\infty)}{N(t_1)}\right).$$

Norton assumes $N(t_1) = N(t_{\mathrm{cd}})$, the size at clinical (or symptomatic) detection, and $N(t_2) = N_{\mathrm{L}}$. He was interested in determining the best values of $N(t_{\mathrm{cd}})$, N_{L}, and $N(\infty)$; and he was also interested in determining the distribution of b, the tumor growth rate. Norton uses a data set provided by Bloom [6] that consists of data on 250 women who had untreated breast cancers and were followed in Middlesex Hospital of London between 1805 and 1933. Norton uses an iterative scheme to solve for the following parameter values and distributions when fitting the Gompertzian model to the Bloom data:

$N(t_{\mathrm{cd}}) = 4.8 \text{ x } 10^9$ cells,

$N_{\mathrm{L}} = 10^{12}$ cells,

$$N(\infty) = 3.1 \text{ x } 10^{12} \text{ cells,}$$

$$b \sim \text{Lognormal(E[b] = 0.07079, SD[b] = 0.05731).}$$

Norton acknowledges that in reality, the values of $N(t_{cd})$, N_L, and $N(\infty)$ are random variables and not constants. This analysis did not provide information about the size of breast cancer at mammographic detection, which is a parameter that is necessary for our simulation model. Thus, we extended some of the assumptions of this rather simplistic model in a number of ways in order to account for newly acquired information about breast cancer, such that the growth rate is dependent upon age. In the following section, we describe how we extended and updated this tumor growth model for use in the natural history simulation model. It should also be noted that in reality, tumors reach a terminal size; and breast cancer death is caused by metastasis to other organs, which prevents those organs from functioning properly. In this research, we refer to N_L as the size at which the tumor becomes lethal because in reality, the cancer will never reach that size, but it represents the size the cancer would be when it becomes lethal if it kept growing and the body could support an ambiguously large tumor.

3.2.1.1 Extensions and Additions to Tumor Growth Model

In the relevant literature, some data are given in terms of tumor diameter, and other data is given in terms of number of cells in the tumor. Thus, it is important to be able to relate these two measures. Because most tumor cells are roughly $2.0x10^{-2}$ mm in diameter [7, 74, 102], it is reasonable to assume that the tumor density, D, has a value of 2.3873 x 10^8 (cells/cm^3) [62]. Therefore at time *t*, the tumor volume (cm^3) is given by Equation (3.13):

$$V(t) = N(t)/D \text{ cm}^3. \qquad (3.13)$$

And since each tumor is assumed to be spherical in shape, we see that the tumor diameter (cm) in terms of the number of cells is given by Equation (3.14), and the cell count $N(t)$ is given in terms of the tumor diameter $d(t)$ by Equation (3.15):

$$d(t) = 2\left[3N(t)/(4\pi D)\right]^{1/3}, \tag{3.14}$$

$$N(t) = D\left[(4/3)\pi(d(t)/2)^3\right]. \tag{3.15}$$

We will exploit the relationships between the tumor-size measures $N(t), V(t)$, and $d(t)$ in dealing with estimates of tumor sizes at different phases of growth, and in determining the stage of the cancer at diagnosis.

Another extension to the tumor growth model involved adding variability to the parameters N_{L} and $N(\infty)$ since it is known they are not constants. We make the very basic assumption that N_{L} and $N(\infty)$ are normally distributed random variables, and that the mean of these normal distributions is the best-fit value provided by Norton (see above). We assume a conservative coefficient of variation of 0.05, which determines the standard deviations of the corresponding normal distributions. Figure 3-2 and Figure 3-3 are graphs of those probability distributions. These assumptions are not great departures from the original model, but we believe they make the tumor growth model a more accurate representation of reality.

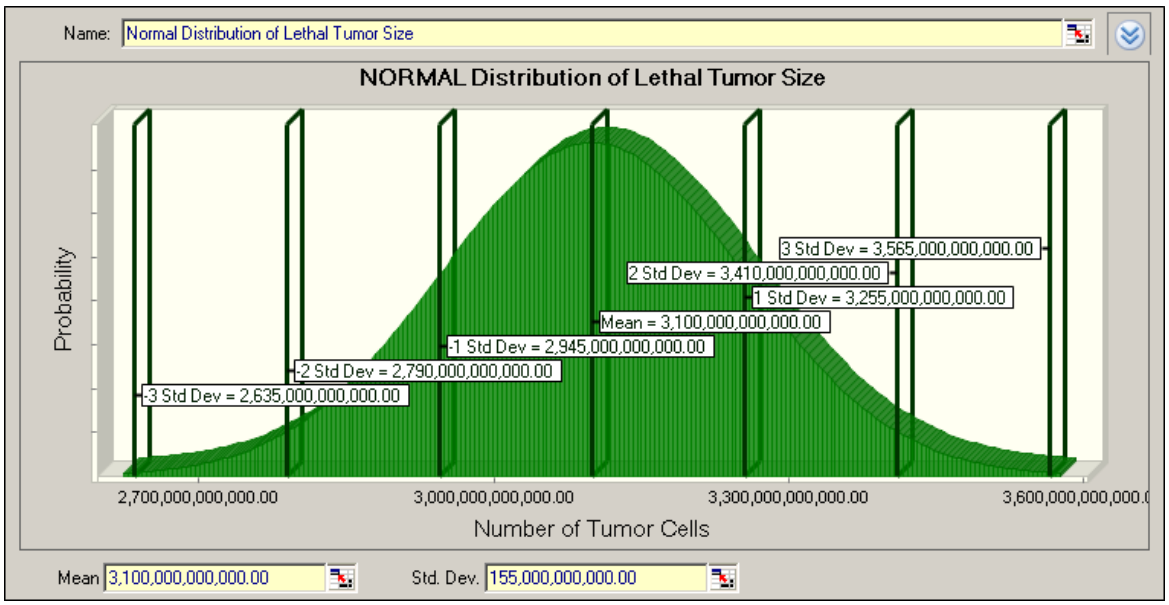


Figure 3-2. Distribution of Lethal Tumor Size, N_{L}

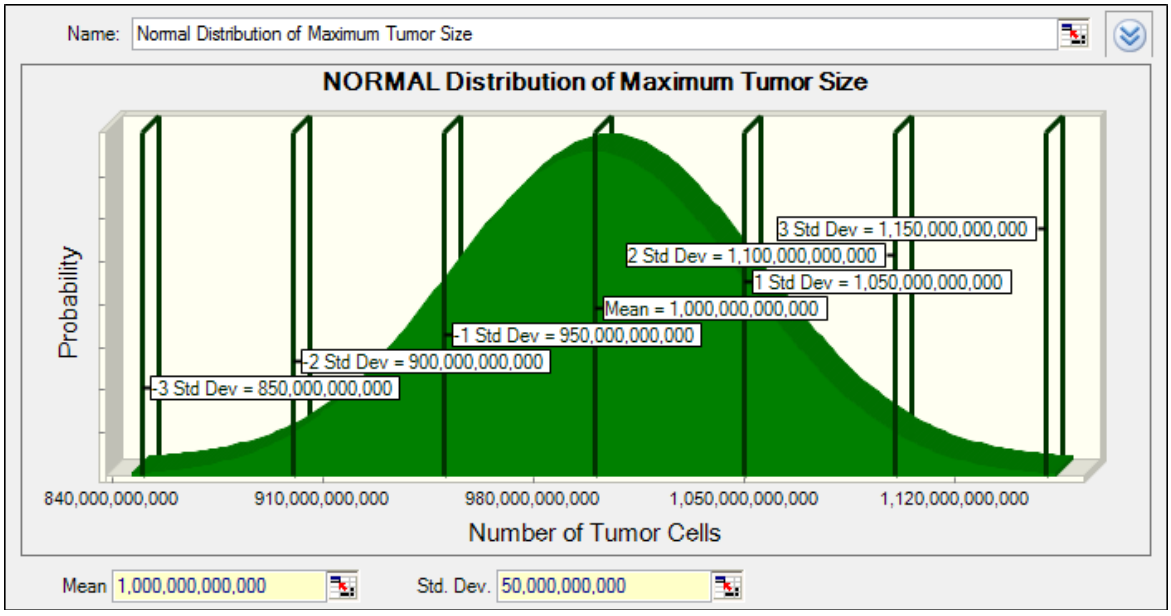


Figure 3-3. Distribution of Maximum Tumor Size, $N(\infty)$

As previously mentioned, breast cancer is believed to behave less aggresively in older women [21, 27, 28, 60, 73]. Thus, to accurately capture the behavior of tumor growth in older women, the tumor growth rate needs to be age-dependent. Clearly for younger women, the tumor growth rate will generally be larger, leading to faster-growing tumors; and for older women the tumor growth rate will be smaller, leading to slower-growing tumors. In order to make the tumor growth rate a function of age, we vary the lognormal mean and standard deviation based on the original distribution of b given by Norton. We do this as follows: we assign the 25^{th} percentile of the original distribution of b as the expected value of the lognormal distribution of b for a 25-year-old woman; we assign the 75^{th} percentile of the original distribution of b as the expected value of the lognormal distribution of b for a 75-year-old woman; and we let the mean of b vary linearly for all ages between 25 and 75. The distribution for women over 75 is the same as for a 75-year-old woman. The coefficient of variation of b is held constant, so the standard deviation of b for each age is directly determined by the mean of b for that age; and the shape of this distribution is the same for each age. Figure 3-4 is a graph of Norton's original lognormal distribution for b (x 1,000 for clarity) and Table 3-2 shows the percentiles of this distribution. Figure 3-5 and Figure 3-6 are graphs of the probability distributions of tumor growth constant b for a 65-and a 75-year-old woman after using the method described above to determine the age dependency. Other approaches for modeling tumor growth as a function of age may be considered in the future.

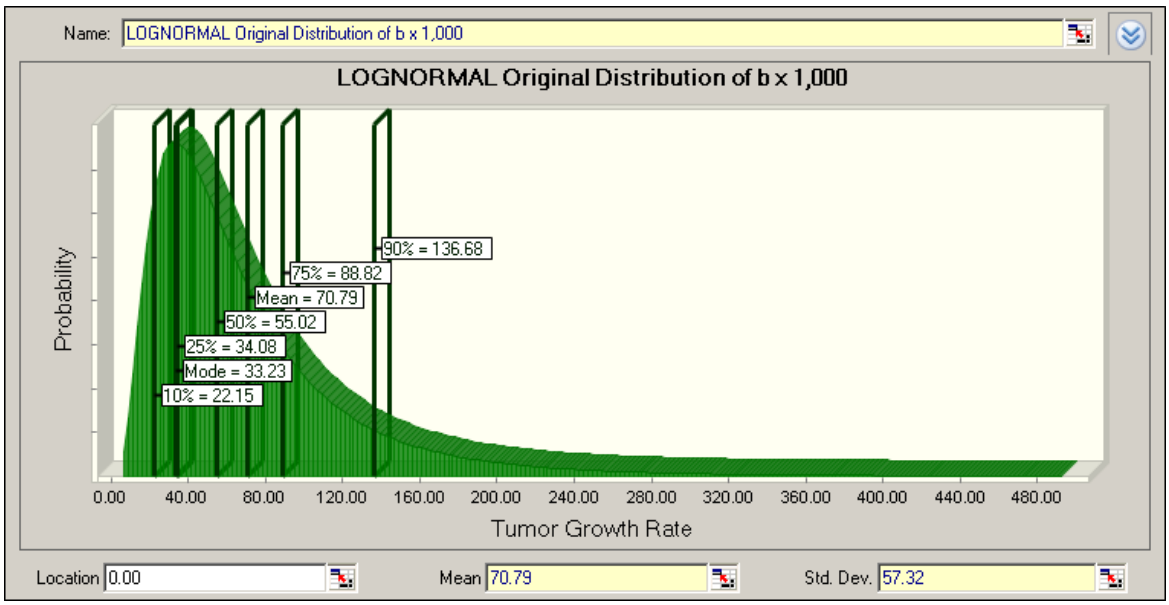


Figure 3-4. Original Distribution of Tumor Growth Constant *b* x 1,000

Table 3-2. Percentiles for Original Distribution of Tumor Growth Constant *b*

Original Distribution on Tumor Growth Rate b	
Percentile	**Value**
5	0.01711
10	0.02215
15	0.02336
20	0.03027
25	0.03408
30	0.03791
35	0.04185
40	0.04596
45	0.05032
50	0.05502
55	0.06015
60	0.06586
65	0.07233
70	0.07984
75	0.08882
80	0.10001
85	0.11484
90	0.13668
95	0.17690

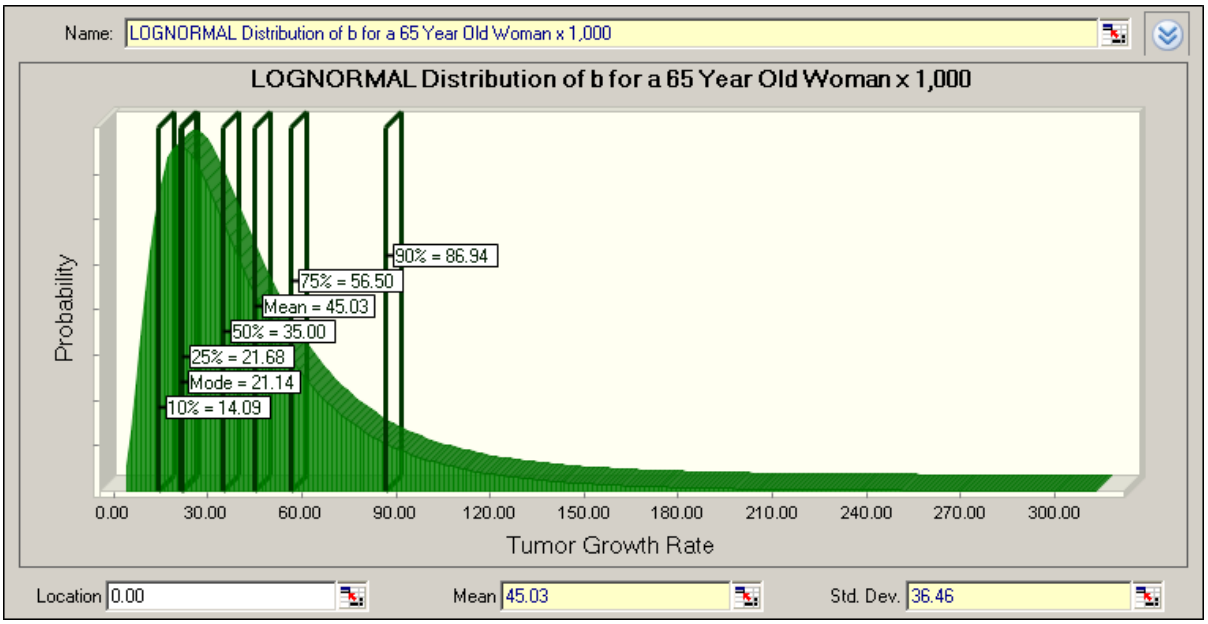


Figure 3-5. Distribution of Tumor Growth Constant *b* for a 65-Year-Old Woman x 1,000

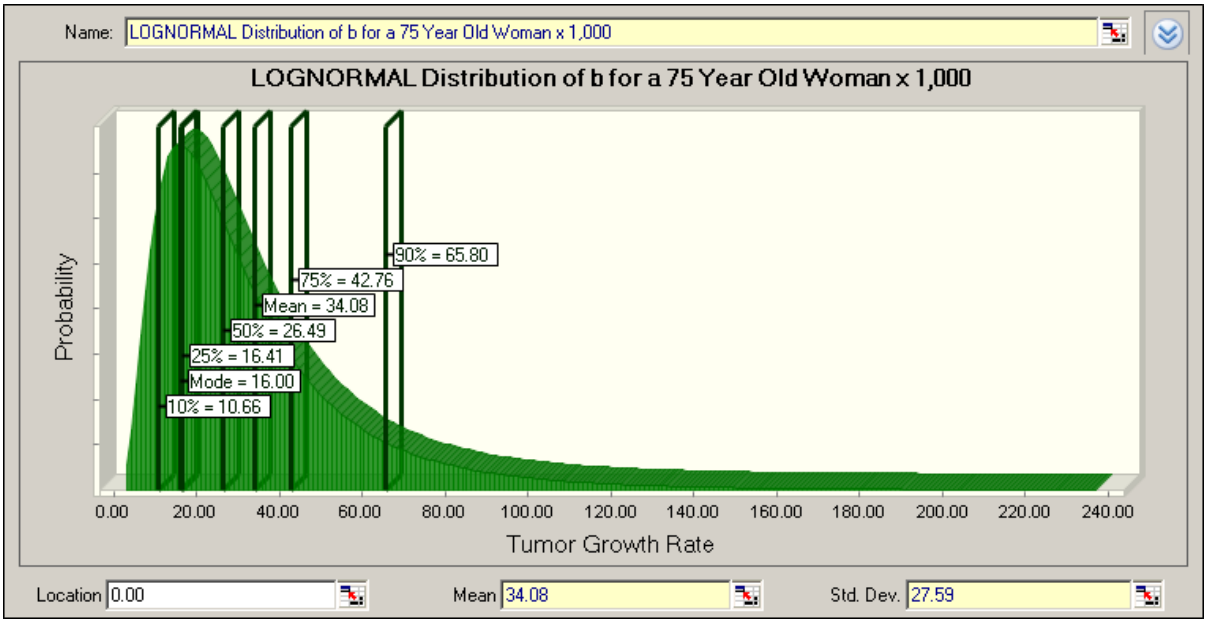


Figure 3-6. Distribution of Tumor Growth Constant *b* for a 75-Year-Old Woman x 1,000

The natural history model assumes women are being screened for cancer by mammography every year. However, at some point most women with breast cancer will experience symptoms of the disease or they will detect a lump in the breast; and it is possible that this may happen before or after cancer is detected by mammography. Thus, we need to estimate the size of each cancer at mammographic detection, $N(t_{\text{md}})$, and at clinical detection, $N(t_{\text{cd}})$.

The size of the tumor at mammographic detection $N(t_{\text{md}})$ in a population being screened yearly is a direct function of the tumor growth rate for each woman with cancer. If the tumor is growing fast, then it will likely be larger at detection than a tumor that is growing slowly. Duffy [29] provides information on the size of tumors at mammographic detection, and identifies six different categories for tumor diameter (mm) at mammographic and clinical detection: 1–10 mm, 10–15 mm, 15–20 mm, 20–30 mm, 30–50 mm, and 50–75 mm. Thus, we define six different classifications of tumor growth: very fast, fast, slightly fast, slightly slow, slow, and very slow. When a woman's tumor is detected mammographically, her tumor growth constant b is sampled from the appropriate lognormal distribution for an individual of her age as detailed on pp. 66–67. In order to determine which tumors fell in each category, we use selected percentiles of the original distribution for b that was posed by Norton for b. At first glance it might appear that we should use selected percentiles of the age-dependent distribution from which the woman's growth constant b was sampled; however the tumor-growth classifications {very fast, fast, slightly fast, slightly slow, slow, and very slow} apply to the entire population of women with untreated invasive breast cancer instead of a subpopulation of women in a certain age range—and Norton's original lognormal distribution was specifically formulated to describe the former population of all women with untreated invasive breast cancer.

If $F_0(u) \equiv \Pr\{b \le u\}$ denotes Norton's original cumulative distribution function (CDF) of the growth constant b for all nonnegative values of u, then, for example, the 10th

percentile of Norton's original distribution is $F_0^{-1}(0.10)$. In terms of this notation, we used the 10th, 25th, 50th, 75th, and 90th percentiles of Norton's original distribution to define the six growth classifications for each woman's tumor:

$$\text{Class} = \begin{cases} \text{Very slow}, & \text{if } b \le F_0^{-1}(0.10), \\ \text{Slow}, & \text{if } F_0^{-1}(0.10) < b \le F_0^{-1}(0.25), \\ \text{Slightly slow}, & \text{if } F_0^{-1}(0.25) < b \le F_0^{-1}(0.50), \\ \text{Slightly fast}, & \text{if } F_0^{-1}(0.5) < b \le F_0^{-1}(0.75), \\ \text{Fast}, & \text{if } F_0^{-1}(0.75) < b \le F_0^{-1}(0.90), \\ \text{Very fast}, & \text{if } F_0^{-1}(0.9) < b. \end{cases} \tag{3.16}$$

Table 3-3 summarizes this information. In addition, we assume that the tumor size is uniformly distributed with minimum and maximum parameter values defined in Table 3-3. Using Equation (3.15) and converting each sampled diameter $d(t)$ to cm, we can easily relate $d(t)$ to the corresponding number of cells $N(t)$ for use in the tumor growth model.

Table 3-3. Relationship between Tumor Size at Mammographic Detection and Tumor Growth Constant b

Tumor Growth Class	Tumor Growth Constant b		Diameter $d(t)$ at Detection (mm)	
	Min	Max	Min	Max
Very Slow	0	0.02215	1	10
Slow	0.02215	0.03408	10	15
Slightly Slow	0.03408	0.05502	15	20
Slightly Fast	0.05502	0.08882	20	30
Fast	0.08882	0.13668	30	50
Very Fast	0.13668	Infinity	50	75

Tumor size at clinical detection $N(t_{cd})$ is not a function of the tumor growth rate. Regardless of how fast the tumor is growing, at some point the woman with cancer will become symptomatic and seek care. Norton provides the information we need to infer a

distribution of the size of tumors at clinical detection. He assumes the minimum size for clinical recognition is 10^9 cells (20 mm) and the maximum is 5 x 10^9 cells (34 mm). He also provided a best-fit estimate of 4.8 x 10^9 cells (33.7 mm). As in many stochastic simulation applications, we chose to fit a generalized beta distribution to these estimated characteristics of $N(t_{cd})$; and we assumed that the standard deviation of $N(t_{cd})$ is one-sixth of its range so that we have

$$\sqrt{\text{Var}[N(t_{cd})} = \frac{1}{6}[5\times10^{10} - 10^9] = 8.17\times10^9 \text{ cells;} \qquad (3.17)$$

see Section 2 of Kuhl et al. [52]. Using this information, we can define the distribution of the size at clinical detection (cells) as a Beta distribution with minimum 10^9, a mode of 4.8 x 10^9, and maximum of 5 x 10^9, which leads to the first shape parameter having the value of 4.149 and the second shape parameter having the value of 1.166. Clearly if $t_{md} < t_{cd}$, then the cancer was detected by mammography before clinical symptoms were present. On the other hand if $t_{md} > t_{cd}$, the cancer was detected because clinical symptoms caused the patient to seek care. All women are assumed to seek care when clinical symptoms surface from invasive cancer in the natural history model. DCIS is assumed to be asymptomatic.

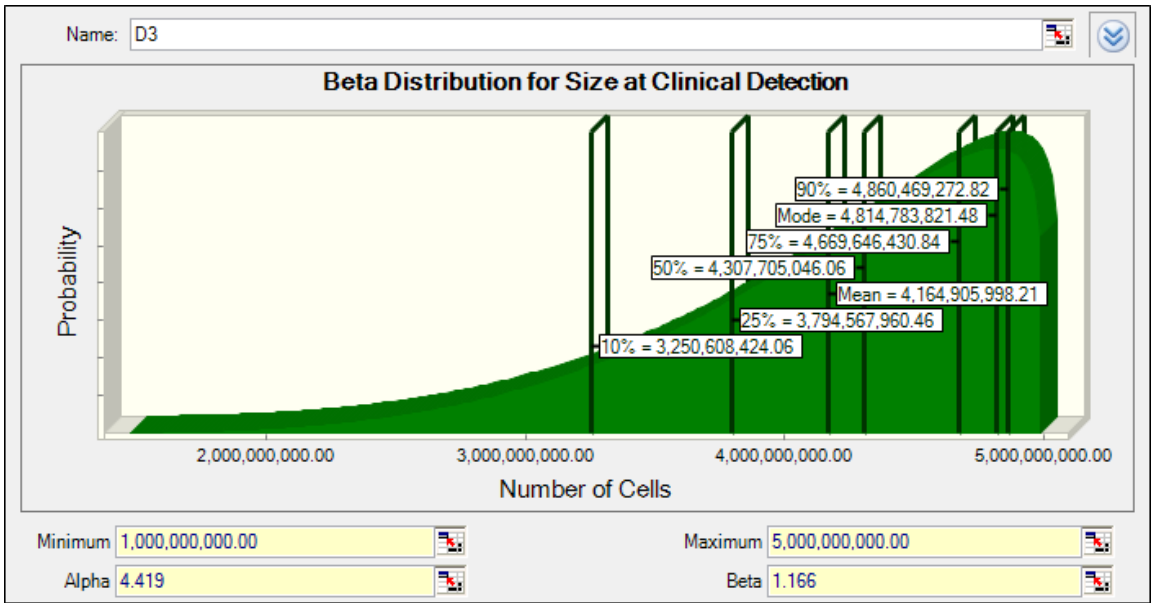


Figure 3-7. Distribution of Tumor Size at Clinical Detection, $N(t_{cd})$

Onset ages are of interest, and our method for calculating them is as follows. In back-calculating a woman's age at the onset of the tumor, we define the following quantities:

$N(t_1) = N(t_{on})$ is the tumor size (cells) at the time of cancer onset $t_1 = t_{on}$.

For a woman whose cancer is detected by mammography ($t_{md} < t_{cd}$):

$N(t_2) = N(t_{md})$ is the tumor size (cells) at the time $t_2 = t_{md}$ of mammographic detection.

For a woman whose cancer is detected clinically ($t_{md} > t_{cd}$):

$N(t_2) = N(t_{cd})$ is the tumor size (cells) at the time $t_2 = t_{cd}$ of clinical detection.

We described previously how the values of $N(t_{cd})$, b, and $N(t_{md})$ are determined. We make the following assumptions based on estimates of parameter values and distributions that were provided by Norton [69], other cited references, and our own assumptions:

$N(t_1) = 2000$ cells [62],

$b \sim$ Lognormal(age-dependent mean and standard deviation) as described on p. 66 and in Table 3-2,

$t_2 = \min\{t_{md}, t_{cd}\}$

$N(t_{md}) \sim$ Uniform (lower, upper bound dependent on b) cells as described by Equation (3.16) and Table 3-3,

$N(t_{cd}) \sim 10^9 + (4 \times 10^9) * \text{Beta}(4.149, 1.166)$ cells [69],

$N(t_2) = N(t_{on}) = \min\{N(t_{md}), N(t_{cd})\}$ cells

$N(\infty) \sim$ Normal(1 x 10^{12}, 5 x 10^{10}) cells

Given values of $N(t_1)$, $N(t_2)$, $N(\infty)$, and b, we can calculate the time delay $t_D = t_2 - t_1$ (in months) required for the cancer to grow from its initial size to the size at detection:

$$t_D = t_2 - t_1 = (-b)^{-1} \ln\{1 - k^{-1} \ln[N(t_2)/N(t_1)]\} \text{ months.} \quad (3.18)$$

We can then subtract this time delay from the woman's current age, and thus find AGE_{on}, the theoretical onset age of the cancer:

$$\text{AGE}_{on} = \text{CurrentAge} - (t_D/12) \text{ years.} \quad (3.19)$$

Similarly, by letting $N(t_1) = N(t_{on})$ but now letting $N(t_2) = N_L$, we calculate the time delay for the cancer to grow from its initial size to lethal size, and the age at which death from breast cancer would occur if the cancer remains untreated. In the natural history model, all cancers are assumed to be left untreated, and all cancers which reach lethal size are assumed to cause breast cancer death. However, after computation of the breast cancer death age, the model determines whether or not a death from other causes would have occurred prior to breast cancer death; and if so, then the death is recorded as a death not caused by breast cancer.

In summary, the tumor growth model works in the following manner in the natural history simulation model: If it is determined that a woman has been diagnosed with invasive breast cancer (not DCIS), then her age is used to determine the appropriate distribution for b and then to sample from that distribution a value of b, her tumor growth constant. The size at mammographic detection, $N(t_{\text{md}})$, is assigned in conjunction with the tumor growth constant b as previously described on p. 66 and in Table 3-2. The size at clinical detection, $N(t_{\text{cd}})$, is then sampled from the Beta distribution previously described on p. 71. If $N(t_{\text{md}}) > N(t_{\text{cd}})$ (which is equivalent to the condition $t_{\text{md}} > t_{\text{cd}}$), then it is assumed that clinical symptoms would have occurred before mammographic detection; and the woman would have acted on those symptoms at an earlier time, leading to an earlier diagnosis and smaller tumor size at detection. The tumor size at detection, whether clinical or mammographic, is important because it is used to determine the stage of cancer at diagnosis. The next section discusses our model of metastasis, which is how we determine the stage of cancer progression within the body at the time of diagnosis. Next, we calculate the onset age, the age at mammographic detection, the age at onset of clinical symptoms, and the breast cancer death age. If the projected age of death caused by breast cancer is greater than the projected age of death from other causes, then in the natural history model the woman's time of death is determined by the latter projected age and a non-BC cause of death is recorded; otherwise the woman's time of death is determined by the former projected age and breast cancer is recorded as the cause of death. We also record an observation for each important variable for each woman diagnosed with breast cancer, enabling us to infer the distribution of unknown random variables, such as time to reach clinical symptoms.

Finally, it is important to note that other tumor growth models were considered, and that this model was chosen because it offered advantages over the others. Other tumor growth models which were considered included the Speer model [88] and a model proposed

by Shumate and El-Shenawee [85]. The Speer model [88] is a slightly more complicated version of the Gompertz model. Essentially, it assumes the same underlying Gompertzian kinetics (small tumors grow quickly while large tumors grow slowly) but allows for the tumor to grow in "spurts" instead of steadily throughout the course of a woman's life. These spurts occur stochastically throughout the life of the tumor. The equation which governs the Speer model is more complicated than the equation that governs the Gompertz model, and it provides the same goodness-of-fit. Lastly, a model proposed by Shumate and El-Shenawee [85] was evaluated. This is the most complicated model reviewed, and the model went into a level of detail beyond what was required for the simulation model. The model has three dimensions, and tracks the volume and shape of the tumor as well as the number of cells in the tumor at any given time. This specific model was not tailored specifically to breast cancer tumors, but is a representation of general tumor growth that was applied to breast cancer. This model carries much more information than needed for the purposes of our simulation model and was not developed specifically for breast cancer.

3.2.2 Model of Metastasis: Cancer Stage at Diagnosis

Invasive breast cancer is a progressive disease, and the stage at diagnosis plays a significant role in determining not only the type of treatment used but also the patient's prospects for survival. As we discussed in Section 2, breast cancer is typically defined in terms of three stages: local, regional, and distant. Refer to Table 2-4 for a definition of these stages. In-situ cancers such as DCIS have extremely high survival rates; and according to SEER data [66], the relative five-year relative survival rate for women of any age who are diagnosed with DCIS is 100%. That is, women who are diagnosed with DCIS live just as long as other women in the population without DCIS. Women who are diagnosed with local, regional, and distant cancer have five-year relative survival rates of 98.5%, 83.6%, and 23.2%, respectively. The ultimate cause of breast cancer death is the spread of malignant cells to other parts of the body and the resulting destruction of other organs such as the brain and liver [62]. Clearly, both the natural history simulation model and integrated screening model need a method for determining the stage of breast cancer at diagnosis since diagnosis

could be at a later time in the integrated screening model. The stage of breast cancer at diagnosis is dependent upon the size of the tumor at diagnosis. Smaller tumors are often associated with breast cancers that have not spread to other parts of the body. The larger the tumor, the greater the chance that the cancer has spread to the lymph nodes or other major organs. We evaluated two models in the literature for staging breast cancer according to tumor size, the Plevritis model [75] and the Michaelson model [62]. After carefully evaluating the merits of both models, we chose to use the Plevritis model because it is a simpler and more tractable model that can be readily translated into the simulation software language.

Plevritis [75] uses SEER data [66] to construct at stochastic model of the natural history describing the progression in the stages of breast cancer. The important result is a model that allows us to estimate the probability of breast cancer being in the local, regional, and distant stages, as a function of tumor diameter. Figure 3-8 is a graphical depiction of the model. The model fits clinical data reasonably well, and was easy to incorporate into the simulation model since we have a method for determining the diameter of breast cancer tumors at diagnosis. In the integrated screening model, the stage at diagnosis is used to determine survival after treatment. Recall that at time t, $N(t)$, $d(t)$, and $V(t)$ respectively denote the number of cells in the tumor (in mm), the diameter of the tumor, and the volume of the tumor (in mm^3). We now provide some important details of the Plevritis model.

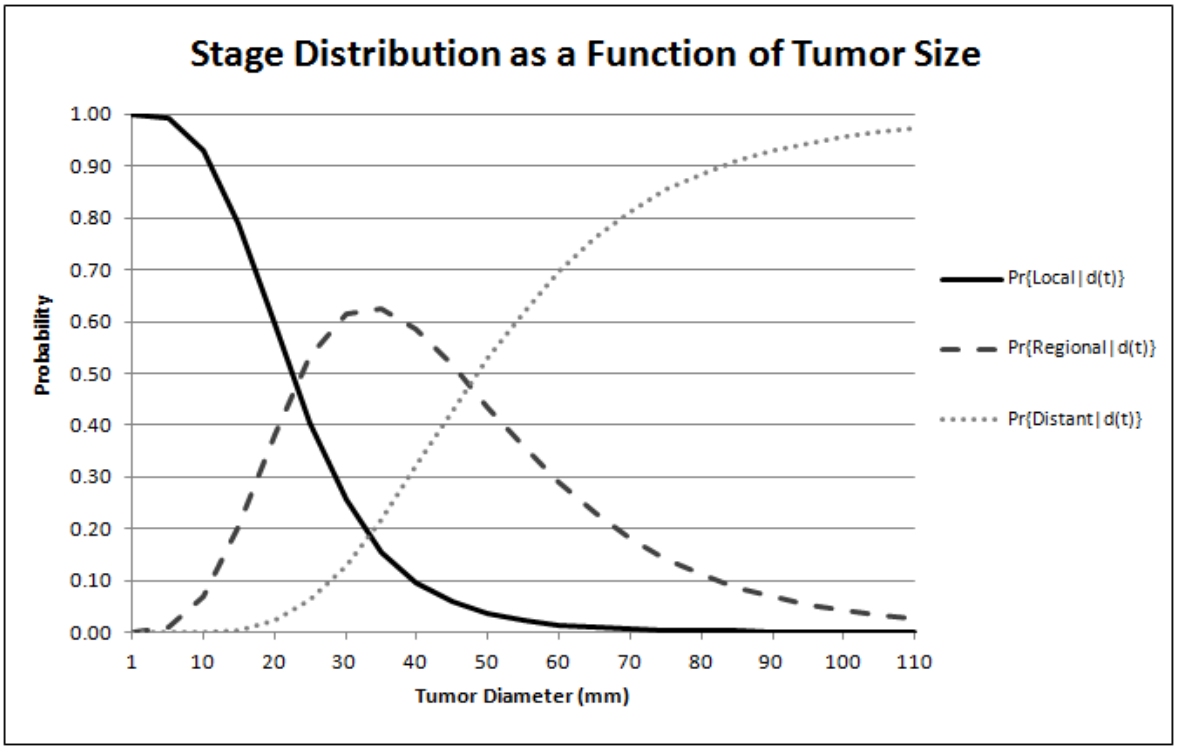


Figure 3-8. Stage Distributions for Untreated Breast Cancers [75]

Given $N(t_{cd}) = n$, Plevritis computes the following:

$$v = n/\left(2.3873\times10^{5}\ \text{cells/mm}^3\right) = \text{tumor volume in mm}^3, \tag{3.20}$$

$$v_0 = \frac{4}{3}\pi\left[\frac{d(0)}{2}\right]^3 = \frac{4}{3}\pi\left[\frac{2\text{mm}}{2}\right]^3 = 4.188\ \text{mm}^3 = \text{initial tumor volume}, \tag{3.21}$$

$$\begin{aligned}\Pr\{\text{Local} \mid N(t_{\text{cd}}) = n\} = \Pr\{\text{Local} \mid N(t_{\text{cd}}) = v\} &= \frac{\alpha\beta^{\alpha}\gamma\left[\beta + (\eta+\gamma)(v-v_0)\right]^{-(\alpha+1)}}{\alpha\beta^{\alpha}\gamma\left[\beta+\gamma(v-v_0)\right]^{-(\alpha+1)}} \\ &= \left[\frac{\beta+\gamma(v-v_0)}{\beta+(\eta+\gamma)(v-v_0)}\right]^{\alpha+1},\end{aligned} \tag{3.22}$$

$$\Pr\{\text{Regional} \mid N(t_{cd}) = n\} = \Pr\{\text{Regional} \mid N(t_{cd}) = v\}$$
$$= \frac{\alpha\beta^{\alpha}\dfrac{\eta\gamma}{(\eta-\omega)}\left\{\left[\beta+(\gamma+\omega)(v-v_0)\right]^{-(\alpha+1)} - \left[\beta+(\eta+\gamma)(v-v_0)\right]^{-(\alpha+1)}\right\}}{\alpha\beta^{\alpha}\gamma\left[\beta+\gamma(v-v_0)\right]^{-(\alpha+1)}} \tag{3.23}$$
$$= \left(\frac{\eta}{\eta-\omega}\right)\times\left\{\left[\frac{\beta+\gamma(v-v_0)}{\beta+(\gamma+\omega)(v-v_0)}\right]^{\alpha+1} - \Pr\{\text{Local} \mid N(t_{cd}) = n\}\right\},$$

$$\Pr\{\text{Distant} \mid N(t_{cd}) = n\} = 1 - \Pr\{\text{Local} \mid N(t_{cd}) = n\} - \Pr\{\text{Regional} \mid N(t_{cd}) = n\}. \tag{3.24}$$

Table 3-4 shows the maximum likelihood estimates for the parameters γ, η, ω, β, and α with asymptotic confidence intervals conditioned on $\alpha = \beta$. Using the parameter values given in Table 3-4 with the exception of setting $\gamma = 0$ in Equations (3.22), (3.23), and (3.24) produces a set of equations that estimate breast cancer stage distribution only conditioned on the tumor size.

Table 3-4. Maximum Likelihood Estimates of Parameters, with Asymptotic Confidence Intervals Conditioned on $\alpha = \beta$ [75]

Parameter	Estimates of the Natural Log of Parameter Values	95% CI Endpoints of Natural Log of Parameter Values	Estimates of Parameter Values
$\hat{\gamma}$	–9.602	[–9.624, –9.580]	6.759×10^{-5}
$\hat{\eta}$	–9.636	[–9.661, –9.610]	6.533×10^{-5}
$\hat{\omega}$	–11.765	[–11.816, –11.713]	7.771×10^{-6}
$\hat{\beta}$	–0.165	[–0.187, –0.143]	0.8478
$\hat{\alpha}$	$\hat{\beta}$	—	—

After these probabilities have been determined as a function of the tumor size at detection for each woman who has a clinical or mammographic detection, we use those probabilities to determine the stage of each woman's cancer at diagnosis in both the natural history model and the integrated screening model.

Michaelson et al. [62] estimate the distribution of the probability $1/P$ that a single cell will leave the primary tumor and form a distant metastasis. On each day t we know the number of cells in the tumor, $N(t)$, and the probability $1/P$ that each of those cells will independently form a distant metastasis at that time. Michaelson estimates that 25% of women have a $1/P$ value of 10^{-11}, 50% of women have a $1/P$ value of 10^{-12}, and 25% of women have a $1/P$ value of 10^{-13}. On each day t given the current values of $N(t)$ and $1/P$, the number of cells that metastasize, $X(t)$, is a random variable whose conditional distribution is binomial with the number of trials $N(t)$ and the success probability $1/P$ so that $X(t)$ has the following conditional probability mass function (PMF) given $N(t)$ and $1/P$,

$$\Pr\{X(t)=x \mid N(t), 1/P\} = \left\{[N(t)]!\Big/\left[x![N(t)-x]!\right]\right\}(1/P)^x\left[1-(1/P)\right]^{N(t)-x} \quad \text{for } X(t)=0,1,\ldots,N(t). \tag{3.25}$$

To sample the number of cells that metastasize according to the conditional PMF (3.25), we would need to iteratively compute successive points on that conditional PMF, accumulate the corresponding points on the cumulative distribution function, and then exploit the inverse transform method [48]. We considered using this approach to compute the time at which the first distant metastasis occurred and the size of the tumor at that time. However, this method requires computation of the probability of metastasis on each day from initial size to lethal size, which would drastically increase the run time of the natural history simulation model. In addition, the model requires our simulation software to realize extremely small probabilities (on the order of 10^{-13}) an extremely large number of times, and Arena does not handle these situations well. Note that in a simulation, we could exploit the Poisson approximation to the binomial PMF,

$$\Pr\{X(t)=x \mid N(t),1/P\} \approx \frac{[N(t)/P]^x \exp[-N(t)/P]}{x!} \quad \text{for } x=0,1,\ldots,N(t); \tag{3.26}$$

see Section 2.5.1 of Hoel [41]. It follows immediately from (3.26) that the random variable $X(t)$ could be sampled from a Poisson distribution with mean $N(t)/P$, but the execution time for this approximation for Michelson's approach is still prohibitively large. Lastly, this model considers regional metastasis only and provides no way to measure the point at which regional metastasis occurs. For these reasons, we use Plevritis' model of cancer staging.

3.2.2.1 Regional Cancer, Lymph Node Involvement, and Detailed Staging

Carter et al. [17] describe the relationship of tumor size to lymph node status and overall survival by analyzing SEER data. Their article contains data regarding the distribution of breast cancer cases by size and lymph node status; and this information enables us to estimate the number of positive lymph nodes given a certain tumor size. Lymph nodes are a medium for the cancer cells to reach other parts of the body, and the more lymph nodes that are positive for cancer the more likely it is that the cancer has or will spread to other parts of the body. The authors divide the number of positive lymph nodes into three categories: 0 nodes positive, 1–3 nodes positive, and 4+ nodes positive.

Table 3-5 provides a summary of this data, where we have factored out unknown cases because they do not provide useful information. Table 3-6 shows the probabilities of lymph node involvement as a function of tumor size that are used in the natural history model.

Table 3-5. Distribution of Breast Cancer Cases by Size and Lymph Node Status [17]

Tumor Size (mm)	Cases	0 Nodes	1–3 Nodes	4+ Nodes	Unknown	Total	Total–Unknown
0–5	21,530	9,721	563	197	11,049	21,530	10,481
5–10	37,075	23,816	2,261	541	10,457	37,075	26,618
10–20	93,875	58,654	14,035	4,475	16,711	93,875	77,164
20–30	54,610	27,139	12,194	6,440	8,837	54,610	45,773
30–40	23,880	9,497	5,704	4,412	4,267	23,880	19,613
40–50	11,786	3,866	2,692	2,753	2,475	11,786	9,311
50–100	17,015	4,120	3,135	5,054	4,706	17,015	12,309
100+	2,580	382	290	670	1,238	2,580	1,342

Table 3-6. Probabilities of Lymph Node Involvement Used in the Natural History Model

Tumor Size (mm)	0 Nodes	1–3 Nodes	4+ Nodes	Total 1+ Nodes	Number of Nodes Given Node–Positive	
					1–3 Nodes	4+ Nodes
0–5	0.9275	0.0537	0.0188	0.0725	0.74079	0.25921
5–10	0.8947	0.0849	0.0203	0.1053	0.80692	0.19308
10–20	0.7601	0.1819	0.0580	0.2399	0.75824	0.24176
20–30	0.5929	0.2664	0.1407	0.4071	0.65440	0.34560
30–40	0.4842	0.2908	0.2250	0.5158	0.56386	0.43614
40–50	0.4152	0.2891	0.2957	0.5848	0.49440	0.50560
50–100	0.3347	0.2547	0.4106	0.6653	0.38283	0.61717
100+	0.2846	0.2161	0.4993	0.7154	0.30208	0.69792

The Plevritis model allows us to determine the probability that cancer is in the regional stage (i.e., the stage in which only lymph nodes are involved and all other organs are free of the malignant cells) as a function of tumor size at diagnosis. For the case in which the cancer is in the regional stage, Carter et al. [17] provide data that are used to estimate the number of positive lymph nodes. In Table 3-6, the two columns on the far right represent the conditional probability of having 1–3, or 4+ positive nodes as a function of tumor size, given that there are positive lymph nodes. If it is determined there are 1–3 positive nodes, then we assume that the number of positive nodes is uniformly distributed on the integers 1–3. If more than three nodes are positive, then we simply record that fact that there are four or more

nodes positive. We use this information and the size of the tumor at diagnosis to estimate the stage at diagnosis at the more detailed level (refer to Table 2-3). Table 3-7 shows our method for determining the detailed stage at diagnosis as a function of the tumor size (i.e., diameter) and lymph node involvement.

Table 3-7. Method for Determining Detailed Stage as a Function of Tumor Size and Lymph Node Involvement

Stage (L,R,D)	**Tumor Size S (mm)**	**Number of Positive Nodes**	**Stage (Detailed)**
Local	$S < 20$	0	1
	$20 \leq S \leq 50$	0	2A
	$S > 50$	0	2B
Regional	$S < 20$	Any	2A
	$20 \leq S \leq 50$	Any	2B
	$S > 50$	1–3 Nodes Positive	3A
		4+ Nodes Positive	3B
Distant	Any	Any	4

One of the concerns about using the Michelson model for metastasis was the fact that it did not provide a method for determining whether the cancer was in the local or regional stage at diagnosis. If the cancer is not in the distant stage, then the "Total 1+ Nodes" column in Table 3-6 is an estimate of the probability that lymph nodes are involved as a function of tumor size at diagnosis. If lymph nodes are involved, then the cancer is in the regional stage; otherwise the cancer is in the local stage. If it is determined that the cancer is in the regional stage, the number of nodes that are positive can be assigned as previously described, and the detailed stage of cancer can again be determined from Table 3-7. However, the computational inefficiency of the Michaelson model still makes its use infeasible.

3.3 Survival and Mortality Submodel

Nearly all of the women who participated in the BCSC study are still alive, so there is no information about death included with the BCSC data. In order to assign a death age for women who do not die of cancer before 2020, we use life-tables provided by Rosenberg [79] in which breast cancer has been removed as a cause of death. An abridged version of the life-

tables used in the natural history model is provided in Table 3-8. It should be noted that there is a table like this for every birth year in the period 1900–2000; and although the current version of the natural history model only requires birth years for the period 1907–1955, all birth years were included to facilitate future implementation of an "all-ages model." In the natural history model, each year a woman has a probability of death occurring from causes other than breast cancer, and this probability is assigned according to the aforementioned life-tables. Thus for each woman who does not die from breast cancer, we compute her age at death from other causes; and we store this quantity for use in statistics calculations and for use when the same population is resimulated in the integrated screening model. This approach allows the natural history model to differentiate between breast cancer–related deaths and deaths from other causes, making it possible to calculate life-years saved by using different screening policies, and the number of cancer deaths averted in any given year.

Table 3-8. Abridged Life–Tables with Breast Cancer Removed as a Cause of Death [79]

Birth Year	Age	Probability of Death from Causes Other Than Breast Cancer
1940	65	0.01331
1940	66	0.01475
1940	67	0.01611
1940	68	0.01732
1940	69	0.01844
1940	70	0.01964
1940	71	0.02106
1940	72	0.02266
1940	73	0.02449
1940	74	0.02657
1940	75	0.02897
1940	76	0.03162
1940	77	0.03439
1940	78	0.03725
1940	79	0.04031
1940	80	0.04389
1940	81	0.04806
1940	82	0.05269
1940	83	0.05783
1940	84	0.06358
1940	85	0.07010
…	…	…
1940	110	1.00

* Life Table data collection stops at age 119. The simulation model uses the assumption that after age 119, the probability of death in the next year remains the same as it was at age 119.

3.4 Population Growth Submodel

It is important to account for changes in the size of the screening population over time. To determine the current rate of growth of the population of US women over 65, we obtained US Census data and determined the average growth of the US population for the last ten years [99]. We fitted a linear regression model to the time series of percentage increases in the population of older US women since the year 2000, shown in Figure 3-9. We go back to the year 2000 because we want to capture the recently increasing trend in this

population of women, and the end of the year 2000 is the beginning of the simulation "warm-up" period. We simulated population growth in the following manner: at the end of the first year, we know the number of women who died in the first year, the number of women still alive, and the expected size of the target population at the end of the next year based on the fitted regression equation. Therefore, we can determine the expected growth of the target population over the next year; and we repeat this procedure every year. The number of women who actually come into the designated population each year is a Poisson distributed random variable with its mean equal to the expected population growth over that year. Figure 3-10 shows 0.1% of the actual size of the 65 and older female population during the period 2001–2010. The graph is presented in this form and with extended axes because later in the chapter to determine if our model accurately represents the data for the period 2001–2010, we plot the same data superimposed on a plot of the sample mean population size for the years 2001–2020 computed across the 10 simulated populations. This is an example of why a "warm-up" period consisting of the years 2001–2010 was chosen, as this greatly assists with model validation.

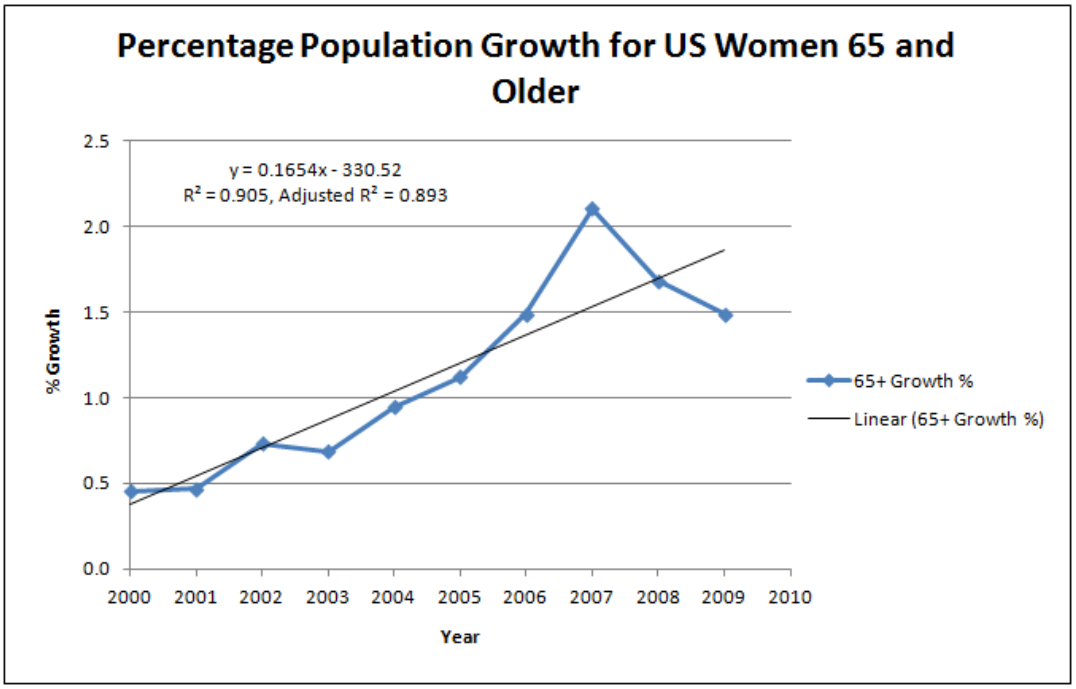


Figure 3-9. Population Growth for Older US Women Since 2000

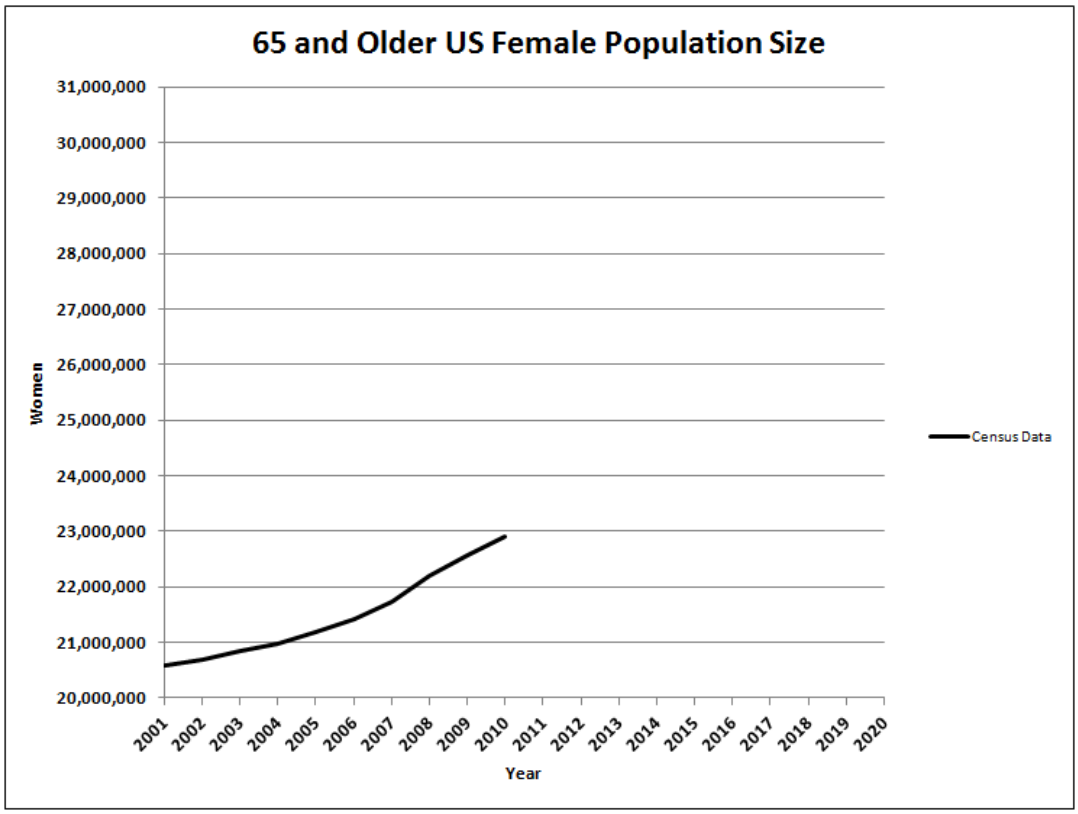


Figure 3-10. Actual 65 and Older US Female Population Size for the period 2001–2010

3.5 Summary of Model Parameters, Assumptions, and Logic

This section summarizes the key assumptions underlying the natural history model as well as the model's main inputs (run parameters). This section also contains a detailed discussion of model logic, including an explanation of how all the individual submodels are used together. We begin by listing the model's run parameters.

3.5.1 Run Parameters

1. We replicate the natural history model for 10 different randomly sampled populations.
2. The initial population size is 20,582 women, which is 0.1% of the actual size of the 65 and older female population at the end of the year 2000.

3. The simulation begins in the year 2001 and runs for 20 years, ending in 2020.
4. With the following hardware and software, the average runtime is 1 minute per population, or 10 minutes total:
    a. 32-bit Arena using 4 GB of RAM; and
    b. 2.93 GHz Intel Core i7 processor.
5. Each year the model evaluates the risks of death caused by breast cancer and by other causes.
6. Women diagnosed with invasive cancer during the period 2001–2020 are followed until death in "Model Shutdown" to allow computation of QALYs saved in the Phase II model.
    a. Youngest age at 2020 = 65, Oldest Age = 110, and
    b. Latest Conceivable Year Woman Could be in “Model Shutdown” = 2065.

Most of the statistical and breast cancer–related assumptions were described in detail in previous sections. However, it is also beneficial to list the key assumptions in one place to get an idea of how they all fit together to form the natural history model.

3.5.2 Key Assumptions

1. None of the women in the study population have had a previous diagnosis of breast cancer. Their risk factor attributes (such as family history of breast cancer) are assigned by random sampling from a BCSC data set.
2. None of the cancers in the population are treated. They are allowed to progress until they become lethal, death occurs from other causes, or the model run ends.
3. Annual screening is the most frequent screening interval considered for this group of women. We simulate the best-case detection scenario, which is assumed to occur under annual screening with perfect compliance.

4. All women who have not yet been diagnosed with cancer receive a mammogram every year, and no appointments are missed. This best-case scenario theoretically allows us to determine the true history of cancer in the designated population.
5. The designated population grows annually according to a linear model fit to the annual growth percentages observed for that population during the period 2000–2009.
6. There are two ways cancer can be detected: mammography and clinical presentation of symptoms.
7. Out of all older women for whom breast cancer is detectable, 20% are diagnosed with DCIS, and the other 80% are diagnosed with invasive cancer. These percentages are based on BCSC data.
8. If a woman will be diagnosed with breast cancer during the next year and she is scheduled to have a mammogram that year, then the time of diagnosis is uniformly distributed over that year.
9. Every woman diagnosed with invasive cancer is assumed to have a tumor that grows according to the Gompertzian tumor growth model presented in Section 3.3. This tumor growth model is used to determine each woman's age at mammographic detection, her age at clinical detection, the stage of the cancer at diagnosis, and the age at which the tumor becomes lethal. Section 3.3 includes all detailed assumptions about tumor growth.
10. The Plevritis natural history model of breast cancer staging is used to determine the stage at diagnosis directly from the tumor size at diagnosis.
11. Life-tables from the Social Security Administration are used to determine the projected time at which death would occur from other causes. For women who are afflicted with breast cancer, this determines their death age, unless they survive to the end of the simulation's time horizon. For women who are afflicted with breast cancer, death from other causes could occur before the projected time of death from breast cancer, especially because breast cancer is less aggressive in older women.
12. Women who are not afflicted with cancer will either die from other causes or survive to the end of the simulation. Women who are diagnosed with invasive cancer can

either have a breast cancer death from a lethal tumor, or death from other causes before that tumor becomes lethal.

3.5.3 Detailed Discussion of Model Logic

All of the aforementioned submodels were created in Arena, and incorporated into a simulation model that we have referred to as the natural history simulation model. The following discussion is a detailed description of the logic contained in the model. First, the model initializes some global variables and arrays used for determining its initial parameter values. A list and description of important global variables including their names, dimension (if array), type, and their initial values is given in Table 3-9. All these initial variable values are written to an external plain-text file so that they can be accessed by the integrated screening model.

Table 3-9 shows that we have accounted for the possibility that in the future this model may be used for younger women. The majority of the model logic would not need to be changed should we decide to investigate the impact of screening and treatment on women of other age groups. The final year of the simulation is 2020, as we felt it would be difficult to predict what breast cancer screening might be like more than ten years from now, and we balanced that with desire to be able to look at breast cancer screening in the future. The initial simulated population size of 20,582 women was chosen for several reasons. Each woman is represented as a separate entity in the simulation model. The number of entities is limited by the memory and speed of the computer; and as the number of entities becomes very large, the simulation run time increases rapidly. The BCSC data set only contains information about 250,000 different combinations of risk factors associated with women in the study who were at least 65, and we wanted to ensure random samples were taken. The different simulated populations of older women can be thought of as a blocking factor in a statistically designed experiment, and we test the screening policies on several different simulated populations of older women so that we are able to factor out the differences between simulated populations

and we can more accurately estimate the true mean differences in performance between candidate screening policies. Also, each simulated population grows by about 33% by the final year of the simulation, and then the next logical larger value to choose for an initial population size would be 205,820, 1% of the total population size. All our computational experience indicates that having over 100,000 entities in a simulation model can begin to have a great effect on run time, and it actually is not necessary to sample this many women to get accurate point and CI estimations of mean performance.

Table 3-9. Global Variables Initialized at the Beginning of the Natural History Model

Variable Name	Type	Description	Initial Value(s)
65OlderModel	Binary(0,1)	0 if ages 35 and older are considered 1 if ages 65 and older are considered	1
IncludeUSPopAge	Binary(0,1)	0 if BCSC database is used to determine age of initial population 1 if 2000 US Census is used to determine age of initial population	1
PopDist65[5]	Real	Given we are only considering women over 65, these conditional percentages represent the percentage of the population that is 65–70, 70–75,75–80, 80–85, and 85+	0
PopDistAllAges[11]	Real	Given we are considering all women over 35, these conditional percentages represent the percent of the population that is 35-40, 40–45,45–50, …, and 85+	0
AllAgesSampleSize	Integer	Total number of women in the BCSC data set	1,007,660
65OlderSampleSize	Integer	Number of women in the BCSC data set who are at least 65 years old	250,509
WriteData	Binary(0,1)	0 if not writing data to files 1 if writing data to files	1
IncludedDistantEvents	Integer(0,1,2)	0 if not considering stage at diagnosis 1 if using the Michaelson metastasis / regional node model 2 if using the Plevritis stage at diagnosis model	2
SimLength	Integer	Length of the simulation in years	25
SimStartYear	Integer	Year the simulation begins	2012
InitialPopSize	Integer	Number of women in the initial population	20,582
CurrentPopSize	Integer	Number of women in currently in the simulation model	InitialPopSize
DeathsThisYear	Integer	Number of deaths that occurred in current model year	0
TotalDeaths	Integer	Total number of deaths at the end of the current model year	0
GrowthIndex	Integer	Value of 1 for the first year and increased by 1 each year; used as *x*-value in the linear model (*y=mx+b)* for the population growth percentage (PopGrowthRatePercent)	1
PopGrowthRatePercent	Real	The population growth percentage calculated according to the linear model: PopGrowthRatePercent = (0.165)*(GrowthIndex) – 330.52	0
NeededPopSize	Integer	The expected number of women that would needed to be added to the population for it to grow according to the linear model	0
GrowthThisYear	Integer	The actual population growth for the current year ~ Poisson(NeededPopSize)	0
NumOfMamms	Integer	The number of mammograms given in the 25-year time horizon	0
SignalCouter	Integer	Specifies the signal number for the 25 signals that are used to ensure the data is written to the output file in the proper order	2012
TumorDensityB	Real	The density of breast cancer tumors (Cells/Cm^3)	238,732,414.6

Following global variable initialization, the initial simulated population is created (still at time 0), and the women in the initial simulated population proceed through their first year of life in the model. After the first year's events occur for each woman in the initial simulated population, time is advanced by one year; women turning 65 during the next year are added to the simulated population; and then we perform the years events for all women in the simulated population. Some of the attributes for each woman are initialized when the corresponding entity is created and inserted into the simulation. We use the term "attribute" to describe each characteristic that is uniquely and specifically assigned to an individual entity; and almost every entity in this model corresponds to a woman in the designated population.

Table 3-10. Attributes Initialized for the Women When First Created

Attribute Name	Type	Description/Initial Values
YearEnteredSim	Integer	Year the woman entered the simulation model.
InitialAge	Integer	The woman's initial age in the simulation model, initialized according to US census data (Table 3-13) for initial population, assumed to be 65 for women who enter the population in later years.
AgeExact	Integer	The woman's current age in the simulation model, initialized to InitialAge.
PostMenoAge	Real	The age at which each woman enters menopause (needed only if considering younger women) - sampled from a discrete uniform distribution on the range 43–59 [68].

The age distribution of the initial population is based on 2000 US census data [36], which is presented in different forms in Table 3-11, Table 3-12, and Table 3-13. Table 3-11 shows the US Census numbers from 2000 and 2010 by age and gender, and it also shows the percentage change over that ten-year period. Table 3-12 shows the details of the population data from the 2000 census, including the percentage of the population falling into each age-gender category. Table 3-13 was derived from the data in Table 3-12, and the two highlighted columns represent the age distribution of the initial population considering only women at least 65, and all women 35 and older. This is another example of an area where we accounted for the fact that this model may be used for younger women in the future. All

women who enter the model after the first year are assumed to do so because they are women turning 65 years old; thus, their initial age in the model is 65. The age when menopause occurs is needed only if younger women are considered and is necessary because the risk model used depends on menopausal status. The NIH [54] estimates that 51 is the average age at which menopause occurs, but it could occur at any age in the range from the mid-forties to late-fifties.

Table 3-11. US Population by Age and Sex: 2000 and 2010 [43]

Age	2000		2010		Percent Change, 2000 to 2010	
	Male	Female	Male	Female	Male	Female
0–4	9,810,733	9,365,065	10,319,427	9,881,935	5.185%	5.519%
5–9	10,523,277	10,026,228	10,389,638	9,959,019	-1.270%	-0.670%
10–14	10,520,197	10,007,875	10,579,862	10,097,332	0.567%	0.894%
15–19	10,391,004	9,828,886	11,303,666	10,736,677	8.783%	9.236%
20–24	9,687,814	9,276,187	11,014,176	10,571,823	13.691%	13.967%
25–29	9,798,760	9,582,576	10,635,591	10,466,258	8.540%	9.222%
30–34	10,321,769	10,188,619	9,996,500	9,965,599	-3.151%	-2.189%
35–39	11,318,696	11,387,968	10,042,022	10,137,620	-11.279%	-10.980%
40–44	11,129,102	11,312,761	10,393,977	10,496,987	-6.605%	-7.211%
45–49	9,889,506	10,202,898	11,209,085	11,499,506	13.343%	12.708%
50–54	8,607,724	8,977,824	10,933,274	11,364,851	27.017%	26.588%
55–59	6,508,729	6,960,508	9,523,648	10,141,157	46.321%	45.696%
60–64	5,136,627	5,668,820	8,077,500	8,740,424	57.253%	54.184%
65–69	4,400,362	5,133,183	5,852,547	6,582,716	33.001%	28.238%
70–74	3,902,912	4,954,529	4,243,972	5,034,194	8.739%	1.608%
75–79	3,044,456	4,371,357	3,182,388	4,135,407	4.531%	-5.398%
80–84	1,834,897	3,110,470	2,294,374	3,448,953	25.041%	10.882%
85+	1,226,998	3,012,589	1,789,679	3,703,754	45.858%	22.943%

Table 3-12. US Population by Age and Sex: 2000 Details [43]

Age	Male	Percent	Female	Percent
0–4	9,810,733	3.49%	9,365,065	3.33%
5–9	10,523,277	3.74%	10,026,228	3.56%
10–14	10,520,197	3.74%	10,007,875	3.56%
15–19	10,391,004	3.69%	9,828,886	3.49%
20–24	9,687,814	3.44%	9,276,187	3.30%
25–29	9,798,760	3.48%	9,582,576	3.41%
30–34	10,321,769	3.67%	10,188,619	3.62%
35–39	11,318,696	4.02%	11,387,968	4.05%
40–44	11,129,102	3.95%	11,312,761	4.02%
45–49	9,889,506	3.51%	10,202,898	3.63%
50–54	8,607,724	3.06%	8,977,824	3.19%
55–59	6,508,729	2.31%	6,960,508	2.47%
60–64	5,136,627	1.83%	5,668,820	2.01%
65–69	4,400,362	1.56%	5,133,183	1.82%
70–74	3,902,912	1.39%	4,954,529	1.76%
75–79	3,044,456	1.08%	4,371,357	1.55%
80–84	1,834,897	0.65%	3,110,470	1.11%
85+	1,226,998	0.44%	3,012,589	1.07%
Total	138,053,563	49.05%	143,368,343	50.95%

Table 3-13. Initial Age Distribution for Women 35 and Older to Women 65 and Older Conditioned on Minimum Population Age

Age	35–39						
35–39	0.15180	**40–44**					
40–44	0.15067	0.17764	**45–49**				
45–49	0.13606	0.16041	0.19506	**50–54**			
50–54	0.11957	0.14096	0.17141	0.21295	**55–59**		
55–59	0.09258	0.10915	0.13272	0.16489	0.20950	**60–64**	
60–64	0.07534	0.08882	0.10801	0.13418	0.17048	0.21567	**65–69**
65–69	0.06822	0.08042	0.09780	0.12150	0.15437	0.19528	0.24897
70–74	0.06597	0.07777	0.09457	0.11749	0.14928	0.18884	0.24077
75–79	0.05810	0.06849	0.08329	0.10347	0.13147	0.16631	0.21204
80–84	0.04160	0.04905	0.05965	0.07410	0.09415	0.11910	0.15185
85+	0.04010	0.04728	0.05750	0.07143	0.09075	0.11481	0.14637
Sum Check	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
Conditional Sum	0.266800	0.2263	0.1861	0.1498	0.1179	0.0932	0.0731

After the initial age is assigned, risk factor attributes from women in the BCSC are randomly chosen for each woman in the simulation model. We divided the BCSC data set according to age such that each five-year interval has its own set of risk factors. Each woman in the simulation model is assigned attributes from a randomly sampled woman within her five-year age interval in the corresponding BCSC data set. At this point, each woman in the model proceeds into the submodels for cancer incidence and for survival and mortality, where we determine the risk of being diagnosed with cancer and the risk of death from other causes on a yearly basis. There are four possible outcomes for each woman in any given year: death from other causes, cancer detection, death from breast cancer (postdetection), or no change in health status. The Barlow risk model computes the risk of being diagnosed with cancer given that a mammogram was performed; thus it is assumed that all women are screened every year until they die, a cancer is detected, or the simulation run ends. The life-tables are used to determine the risk of death from other causes each year, which increases as a function of age. If it is determined that a death from other causes would occur, then the corresponding woman is assumed to have died on a random day during the next year, and her death age is recorded, along with the fact that a death from other causes occurred. The woman then departs the simulation model, and all of her attributes are recorded upon her departure. If there is neither a cancer detection nor death during a given year for a particular woman, then she will experience no other events during that year. When all the women in the model have had their risks of cancer detection and death estimated and evaluated, the simulation can advance to the next year. When the simulation moves to the next year, the number of women who died in the previous year is known, and we can estimate the amount by which the population will grow during that year using our linear model of population percentage growth, so we can directly determine the expected number of women who will turn 65 that year and enter the simulated population. To simulate the number of women that will actually turn 65 that year and enter the population, we generate a Poisson random variable with a mean equal to the expected population growth for that year. The specified number of women enter the model, are assigned initial attributes and risk factor attributes, and then join the rest of the simulated population in the cancer incidence and survival and

mortality submodels. The next year is simulated in the same manner as the previous year, and this logic is repeated until the end of the simulation time-horizon is reached. We now explain the events that take place when cancer detection occurs, and the simulation model logic that governs the disease progression submodel.

If the Barlow risk model determines that during the current year a woman in the simulation will have a mammogram revealing the presence of cancer (invasive or DCIS), then the time of cancer detection is assumed to be a random day during that year. A brief analysis of the BCSC data shows that 19.58% of cancer diagnoses were DCIS, and all others were invasive breast cancers. Thus, once it is determined that a woman will be diagnosed with cancer during the current year, with probability 0.1958 she will be diagnosed with DCIS; otherwise she will be diagnosed with invasive breast cancer.

As we mentioned previously, little is known about the probability or time it takes for DCIS to progress to invasive breast cancer; thus our model has no method for estimating whether or not a specific case of DCIS will progress to invasive cancer, and a breast cancer death from DCIS is assumed never to occur. Thus, once a woman is diagnosed with DCIS, she continues in the simulation model until she dies from other causes or until the simulation ends. Her annual risk of developing cancer is no longer estimated, because the risk model assumes that women have had no previous cancer diagnoses. We choose to include DCIS incidence, although we could not estimate its progression, because DCIS is becoming more prevalent in the population; and we believe it will be valuable to have an estimate of DCIS incidence as well as invasive cancer incidence. In the integrated screening model we account for DCIS treatment (cost, QALYs, etc.), so DCIS still has an effect on the overall cost of screening and treatment for the designated population. Once physicians have more reliable information on DCIS progression, future work may include adding a submodel that estimates DCIS progression to invasive cancer and incorporating that progression into the natural history simulation model. The remaining 80.42% of women for whom cancer is detected are

diagnosed with invasive breast cancer, and we now explain the logic behind the disease progression submodel that governs the progression of invasive cancer.

When a woman is diagnosed with invasive breast cancer, we assume that a primary breast cancer tumor is present; and that this tumor grows from onset to its final size (i.e., the size at which it becomes lethal or its size when the woman dies from other causes) according to the individualized Gompertzian tumor growth model presented in Section 3.3. The first tumor characteristic assigned to a woman with invasive cancer is the tumor growth rate b, which is assigned from a lognormal distribution whose mean and standard deviation are a functions of the woman's age at diagnosis as explained in previous sections. Once the tumor growth rate for the woman is known, we classify her rate of tumor growth into one of six classes in accordance with the percentiles of the original distribution as summarized in Table 3-3. Next, we assigned the size at mammographic detection, which is a random variable whose distribution depends on the growth rate of the tumor (slow-growing tumors have smaller detection sizes, and faster-growing tumors have larger detection sizes). The size at clinical detection is also a random variable that is assigned according to the specified Beta distribution shown in Figure 3-7. If the size at clinical detection is smaller than the size at mammographic detection, then we assume that particular woman would have had her cancer detected before the most recent mammography based on physical symptoms or self-exam, and we count this as a clinical detection instead of a mammographic detection. In these cases, the size and stage of the tumor at diagnosis as well as the age at diagnosis are based on clinical detection and not mammographic detection. We also assign the size at which the tumor becomes lethal according to the aforementioned normal distribution. After assigning the sizes at clinical and mammographic detection, determining the method of detection, and assigning the lethal size, we calculate the following important cancer attributes for the associated woman using the tumor growth equation:

1. The cancer onset age (age that the cancer began to grow);
2. The time delay (months) from onset to mammographic detection;
3. The time delay (months) from onset to clinical detection; and
4. The time delay (months) from onset to lethal size.

Other individual cancer characteristics are known for each woman with cancer detected at this point in time; recall that the values from the natural history model are based on perfect annual screening, our assumed baseline (most frequent screening) scenario. Some of the characteristics of interest are:

1. The method of cancer detection (clinical or mammographic);
2. The age of the woman when breast cancer is first detected;
3. The size of the tumor at detection (diameter, volume, and number of cells);
4. The size of the tumor when it becomes clinically symptomatic (diameter, volume, and number of cells);
5. The size that the tumor must reach to have a lethal effect; and
6. The age at which breast cancer will have a lethal effect.

Next, we determine the stage of breast cancer at diagnosis using the Plevritis natural history model of breast cancer stage progression, which estimates the conditional probability that breast cancer is in the local, regional, and distant stages as a function of tumor size, given that breast cancer symptoms have been detected. The units for this model are diameter of the tumor (in mm), and the reason we constructed an effective method for converting between the number of cells in a tumor and the diameter of that tumor now becomes apparent. At this point in the model logic, we know the size of each woman's tumor at detection (mammographic or clinical), and we use the curves from the Plevritis stage progression model presented in Section 3.4 to estimate the conditional probabilities that the cancer is in the local, regional, and distant stages, respectively. We use this information to

determine 95% confidence intervals (CIs) on the stage distribution at diagnosis under perfect annual screening.

Following the determination of the stage at diagnosis, each woman is returned to the survival and mortality submodel in order to determine if she would die of natural causes or the simulation model would end before she would experience death from breast cancer. If a woman with invasive cancer is still alive at the end of the year 2020, then she is deemed a survivor of invasive cancer, although we do compute her cause of death and age at death after 2020. We compute 95% confidence intervals on the outcome of invasive cancer, that is, on the percentage women with invasive cancers surviving through 2020, the percentage of women with invasive cancers that have a breast cancer death before 2020, and the percentage of women with invasive cancer leading that have a death from other causes before 2020. This concludes the discussion of the natural history simulation model logic. The next section presents the results of the natural history model, which include graphs of 95% CIs on incidence rates, prevalence rates, and death rates as well as 95% CIs on some relevant population characteristics.

3.6 Important Results, Discussion, and Validation Considerations

There are several important results from the natural history model presented in this section. There are a few different classes of results, including annual 95% CIs on population cancer statistics, 95% CIs on cancer statistics for the population as a whole over the entire model time horizon, 95% CIs on stage distribution, 95% CIs on invasive cancer outcome percentages, and distributions that were inferred from individual observations of certain individual cancer attributes. Many of the results are presented graphically in this section, but the raw data for the confidence intervals and SEER data are available in Appendix C: Raw Data for Graphical Results from the Natural History Model.

3.6.1 Annual 95% CI's on Population Cancer Statistics

The following is a list of important results for which 95% confidence intervals were computed for each of the years 2001–2020, and their definitions.

1. Cancer Incidence - the expected value of the number of women that are diagnosed with cancer in a population over a certain finite period of time, typically chosen to be one year. Usually presented in the form of incidence rates per 100,000 women.
2. Cancer Prevalence - the expected value of the total number of women in a population that have breast cancer regardless of the time of diagnosis. Usually presented in the form of incidence rates per 100,000 women or percentages of the designated population.
3. Deaths - the expected value of the number of cancer deaths, deaths from causes other than breast cancer, and the total number of deaths that occur each year.
4. Population Size - the expected values of the size of the population each year

3.6.1.1 Incidence

We begin with cancer incidence rates per 100,000 women, which we divide into invasive cancer incidence rates, DCIS incidence rates, and total incidence rates. Each graph contains the mean, upper and lower limits of a 95% CI about the mean, and SEER data (age adjusted) for comparison and validation purposes. Figure 3-11 is a graph of annual invasive cancer incidence rates, Figure 3-12 is a graph of annual DCIS incidence rates, and Figure 3-13 is a graph of annual total breast cancer incidence rates.

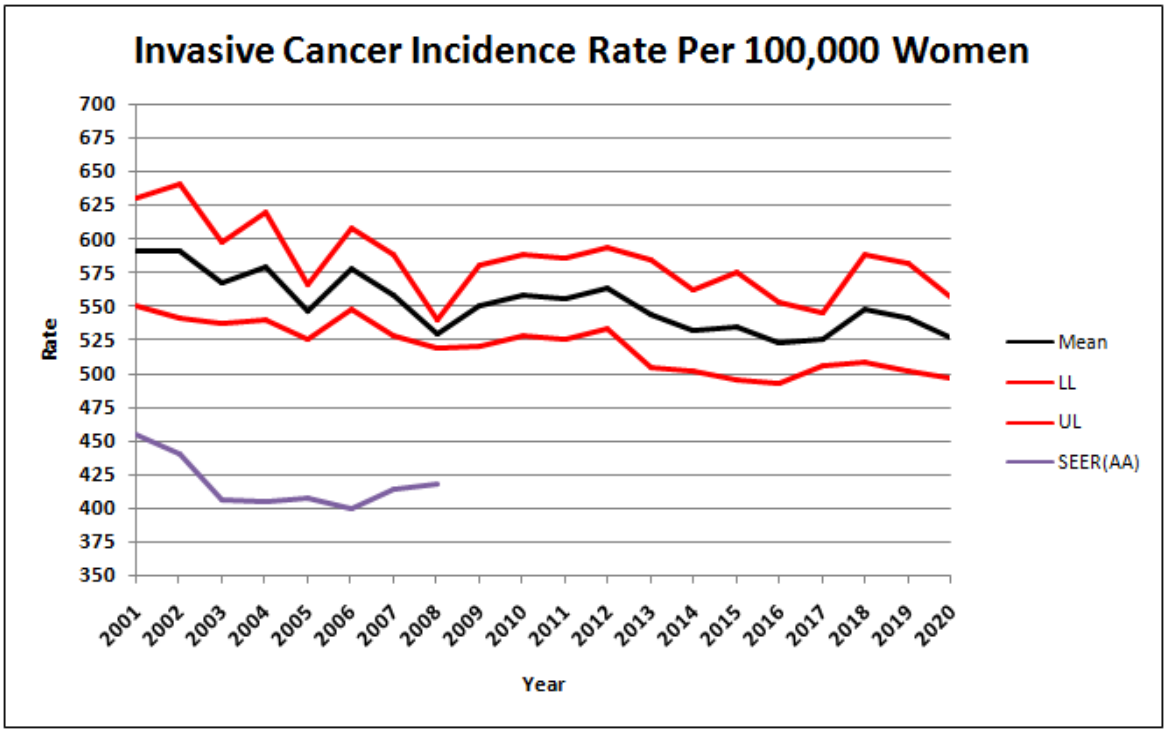


Figure 3-11. Annual Invasive Cancer Incidence Rates Per 100,000 Women

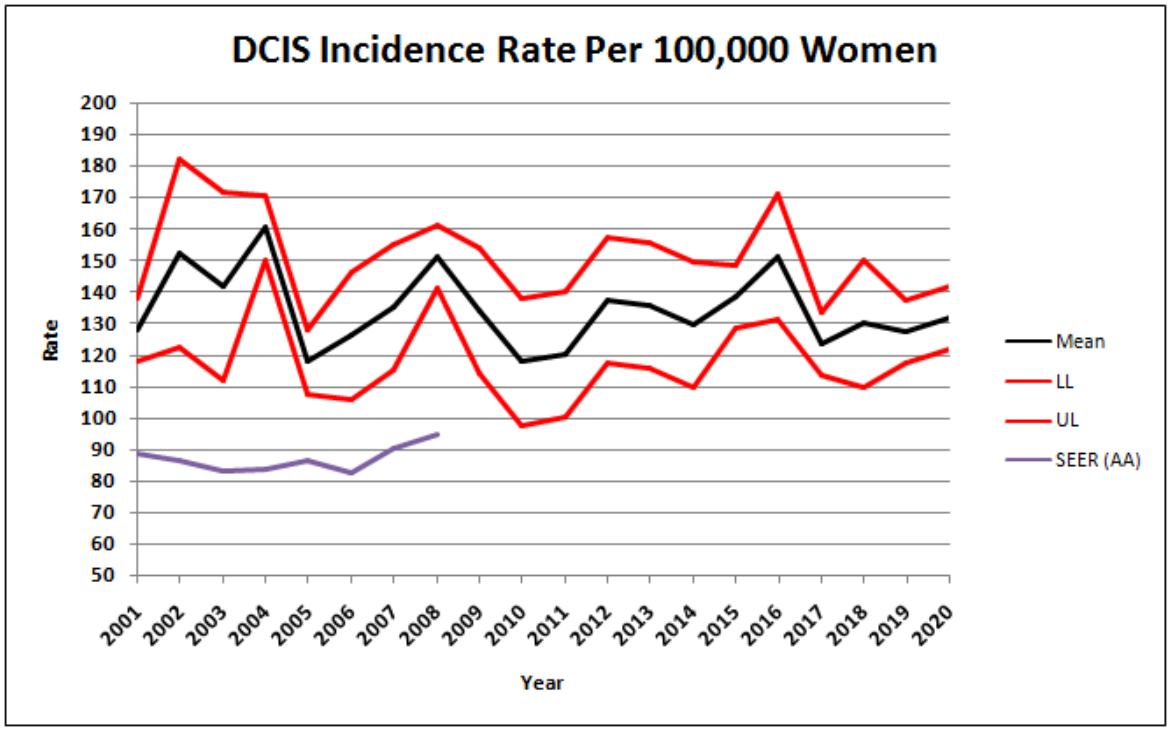


Figure 3-12. Annual DICS Incidence Rates Per 100,000 Women

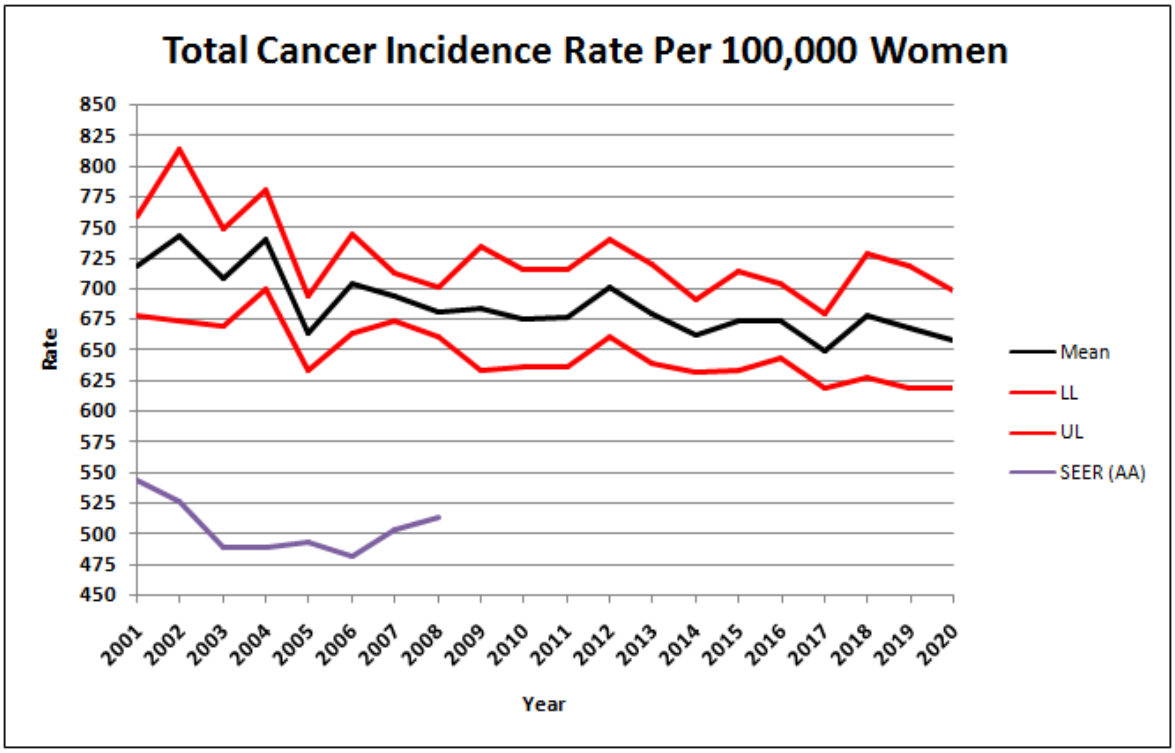


Figure 3-13. Annual Total Cancer Incidence Rates Per 100,000 Women

These three graphs show that the cancer incidence rate in the natural history model is considerably higher than age-adjusted SEER data over the period 2001–2008 when SEER data were available and the model was in the "warm-up" period. Incidence rates do not require a "warm-up" period because unlike prevalence rates, they are not cumulative statistics. SEER data is representative of the actual level of screening and diagnosis in the population during the period 2001–2008, which is certainly not perfect annual screening. Thus, we would expect that with perfect annual screening, the incidence rates would be considerably higher than what we actually see in the population, which is happening in each of the three graphs. How much higher the incidence rates should be cannot be known, but what is known is that the simulation results should be greater by a significant margin; and our model captures this expected behavior. We claim these are reasonable estimates of mean incidence rates under perfect annual screening during the "warm-up" period and most importantly, in the future. At a minimum, this result does not invalidate the model. We

calibrate the SD model, which governs adherence to screening for the population, using the SEER data presented here.

3.6.1.2 Prevalence

We move now to cancer prevalence rates per 100,000 women, which we again divide into invasive cancer prevalence rates, DCIS prevalence rates, and total prevalence rates. Each graph again contains the sample mean as well as upper and lower limits of a 95% CI for the true prevalence rates; however, SEER data for prevalence is not presented on an annual basis. SEER prevalence data from a few different studies simply present the prevalence percentages for certain age groups, and we use a recent US study for comparison with our results. Figure 3-14 and Figure 3-15 are graphs of annual invasive cancer prevalence rates and percentages respectively. Figure 3-16 and Figure 3-17 are graphs of DCIS prevalence rates and percentages, respectively. Finally, Figure 3-18 and Figure 3-19 are graphs of total breast cancer prevalence rates and percentages, respectively.

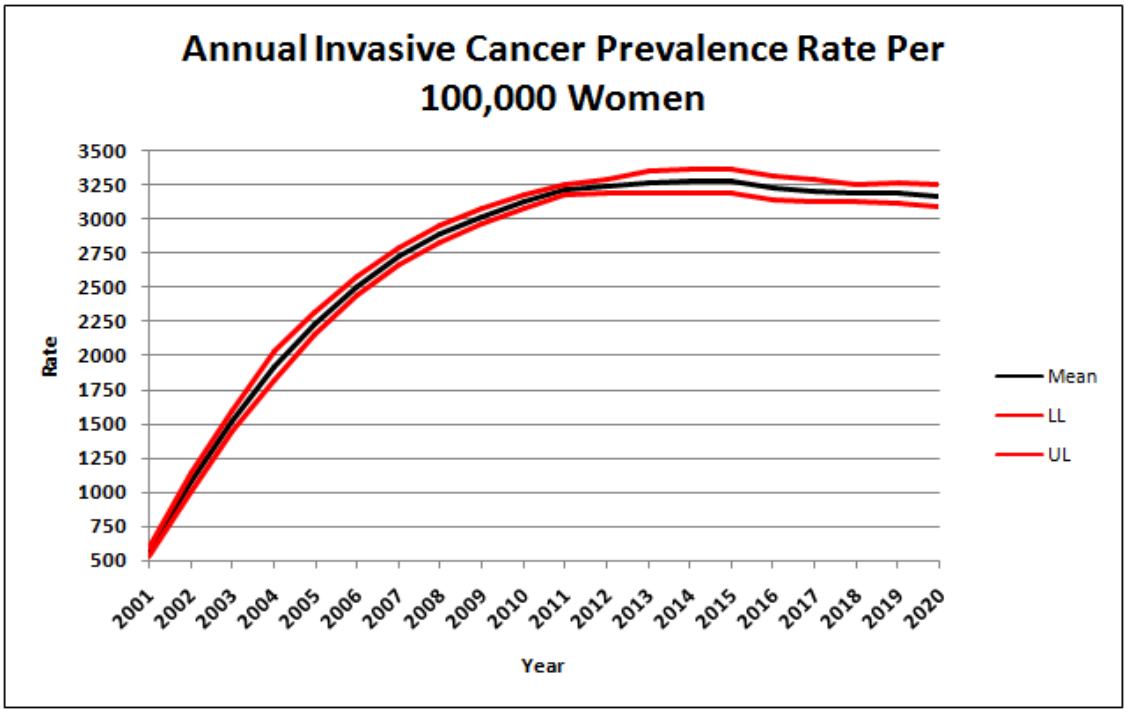


Figure 3-14. Phase I Annual Invasive Cancer Prevalence Rates Per 100,000 Women

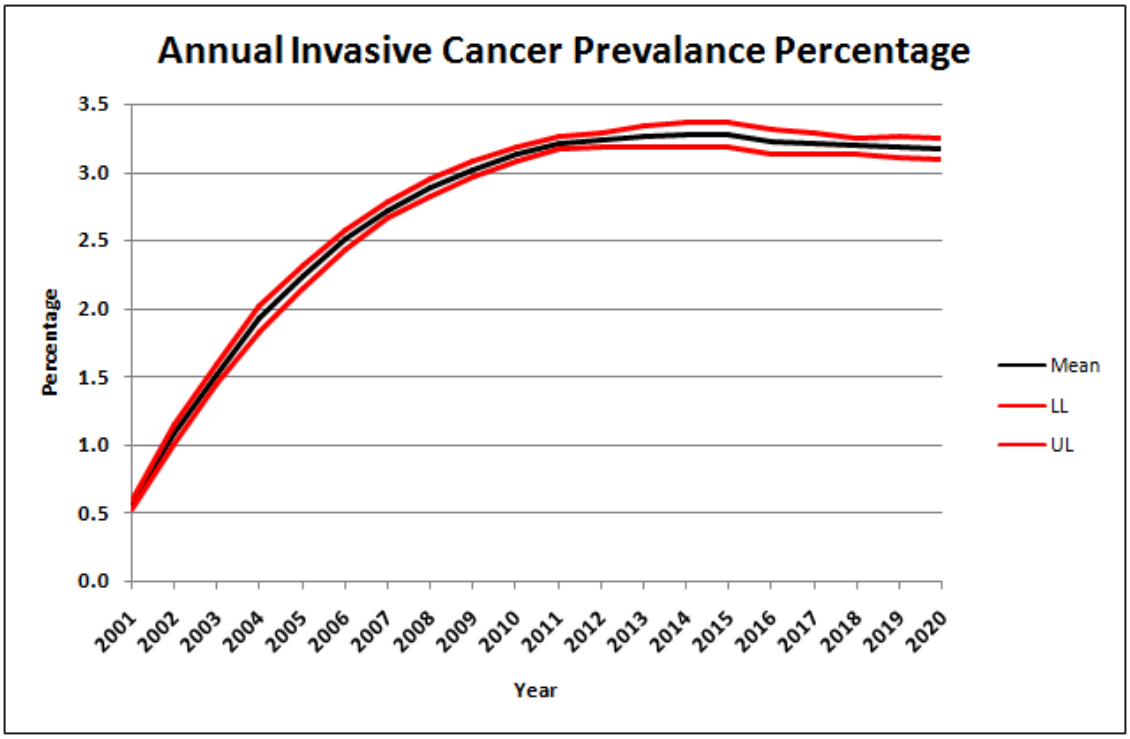


Figure 3-15. Phase I Annual Invasive Cancer Prevalence Percentages

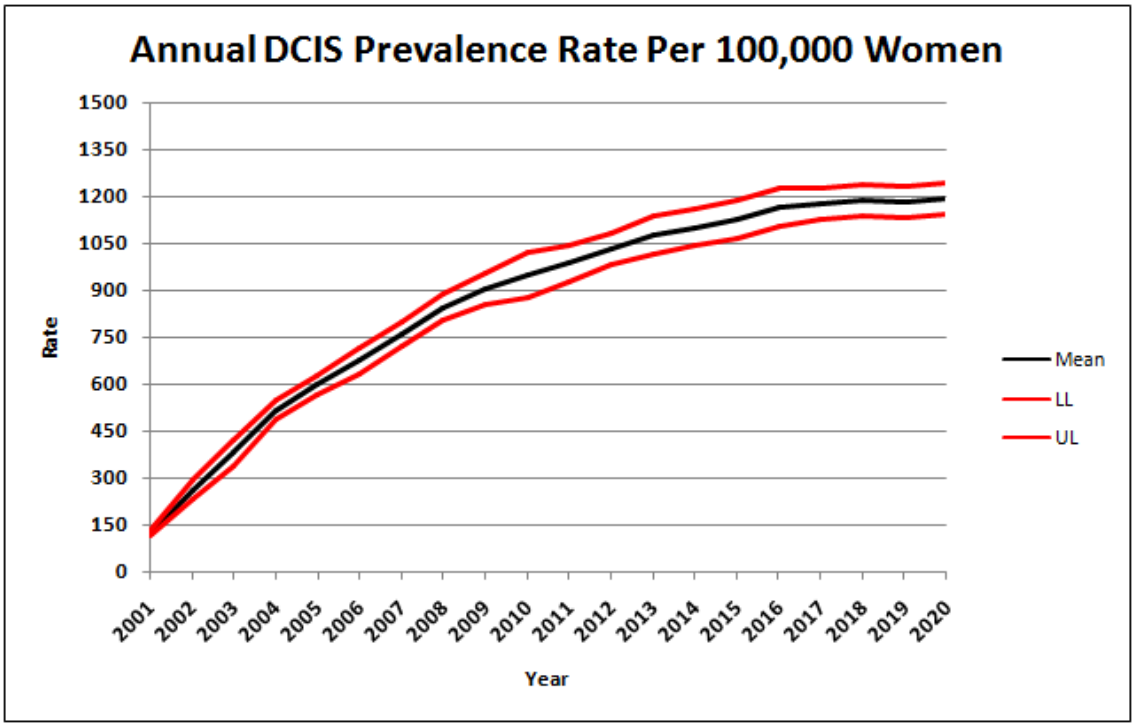


Figure 3-16. Phase I Annual DCIS Prevalence Rates Per 100,000 Women

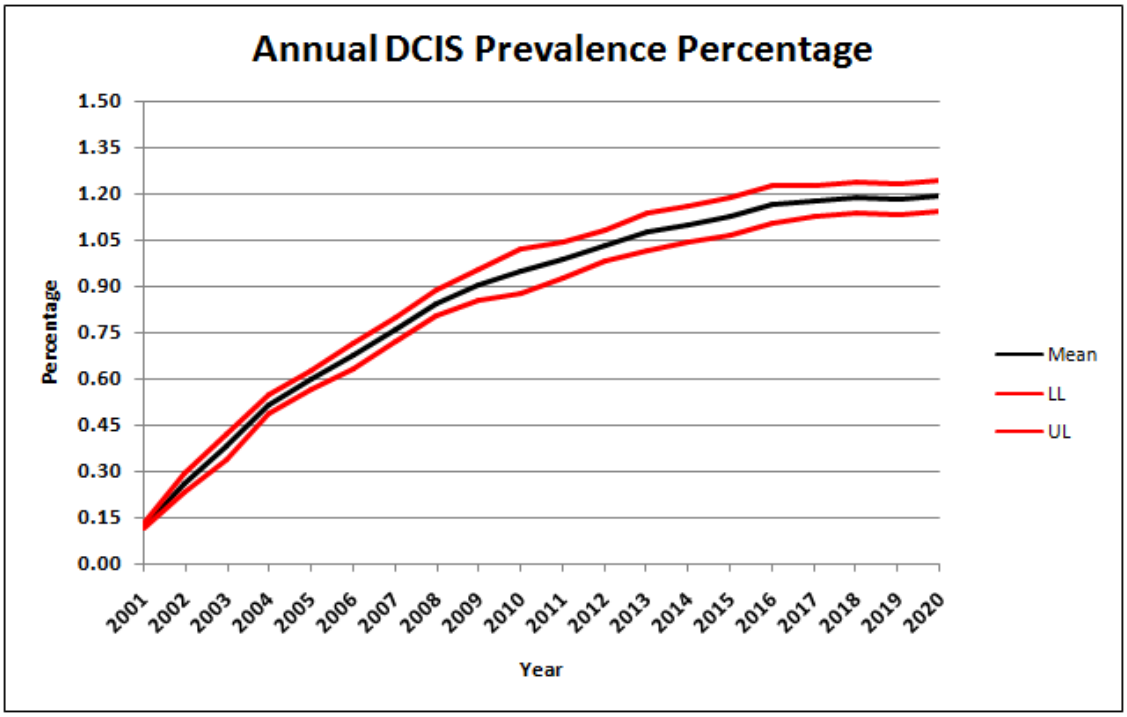


Figure 3-17. Phase I Annual DCIS Prevalence Percentages

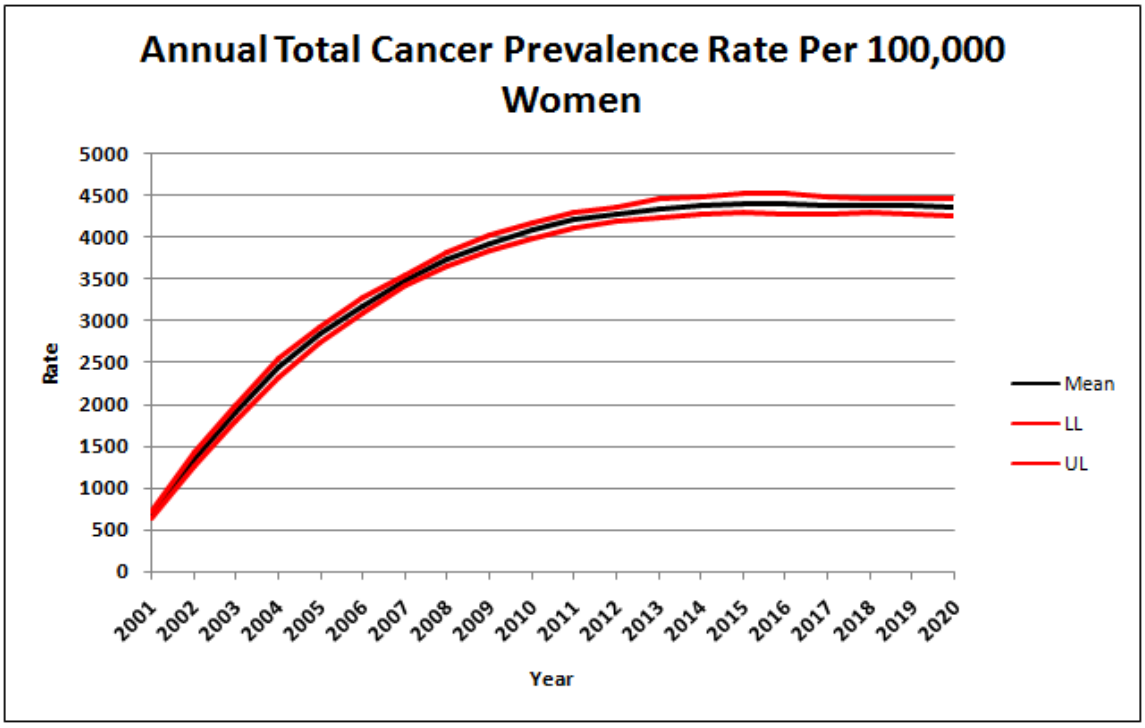


Figure 3-18. Phase I Annual Total Prevalence Rates Per 100,000 Women

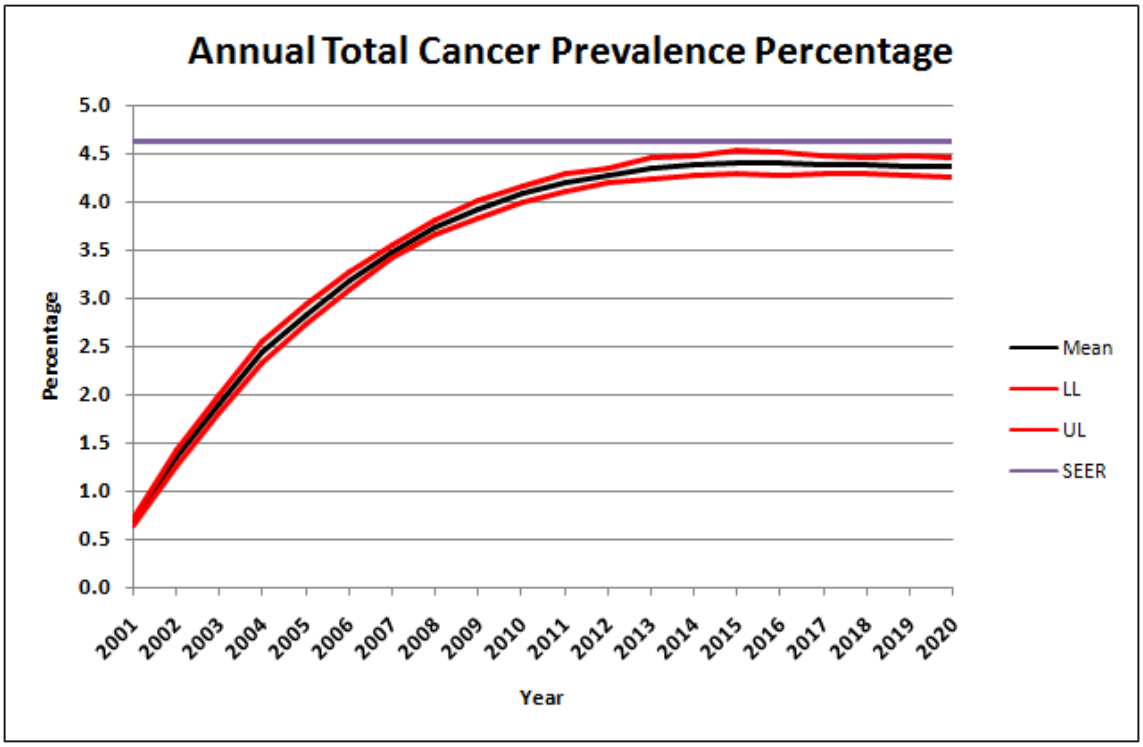


Figure 3-19. Phase I Annual Total Prevalence Percentages

Table 3-14. SEER Prevalence Data by Age from 2000 Study of US Population

Age-Adjusted Prevalence Percent (US 2000)	
Age	**Prevalence Rate**
65–69	4.62%
70–74	4.83%
75–79	5.08%
80–84	5.34%
85+	4.97%

Prevalence rates require a "warm-up" period before achieving a "representative" behavior pattern in about the year 2012 (the beginning of the future). The reason for this is that we began with a population that was cancer free, so the initial prevalence was zero. However, after several years the prevalence percentage in the simulated population begins to mimic the real population; and it is the "warmed-up" simulation results that are comparable with SEER data. The prevalence percentage of 4.62% for ages 65–69 is plotted in Figure 3-14 and is the primary value we use for comparison because by the end of our model, the majority of the women are between the ages of 65 and 74. Additionally, the prevalence percentage only slightly increases to 4.83% for ages 70–74. We would expect "representative" prevalence from our model to be less than the prevalence given by SEER for a population of women 65 and older for one major reason—we are not treating any of the cancers that are detected. When cancers are treated, the women with those cancers remain alive much longer than if they were left untreated, leading to a greater build up (or prevalence) of breast cancer within the population. Since our natural history model assumes no treatment, women with cancer are dying off at a much faster rate, which in turn leads to lower prevalence rates. Beyond the warm-up period, the total cancer prevalence rate fluctuates between 4.3% and 4.5%, which is slightly less than the values reported by SEER as we would expect. Again, at the very minimum, this does not invalidate our study; and in fact our model results are quite close to SEER data with the slight departure induced by of lack of treatment.

3.6.1.3 Deaths

Death rates per 100,000 women are divided into breast cancer death rates, death rates from other causes, and total death rates. Each graph contains the sample mean, upper and lower limits of a 95% CI for the theoretical death rate about the mean, and SEER data for comparison and validation purposes. Figure 3-11 is a graph of annual invasive cancer incidence rates, Figure 3-12 is a graph of annual DCIS incidence rates, and Figure 3-13 is a graph of annual total breast cancer incidence rates.

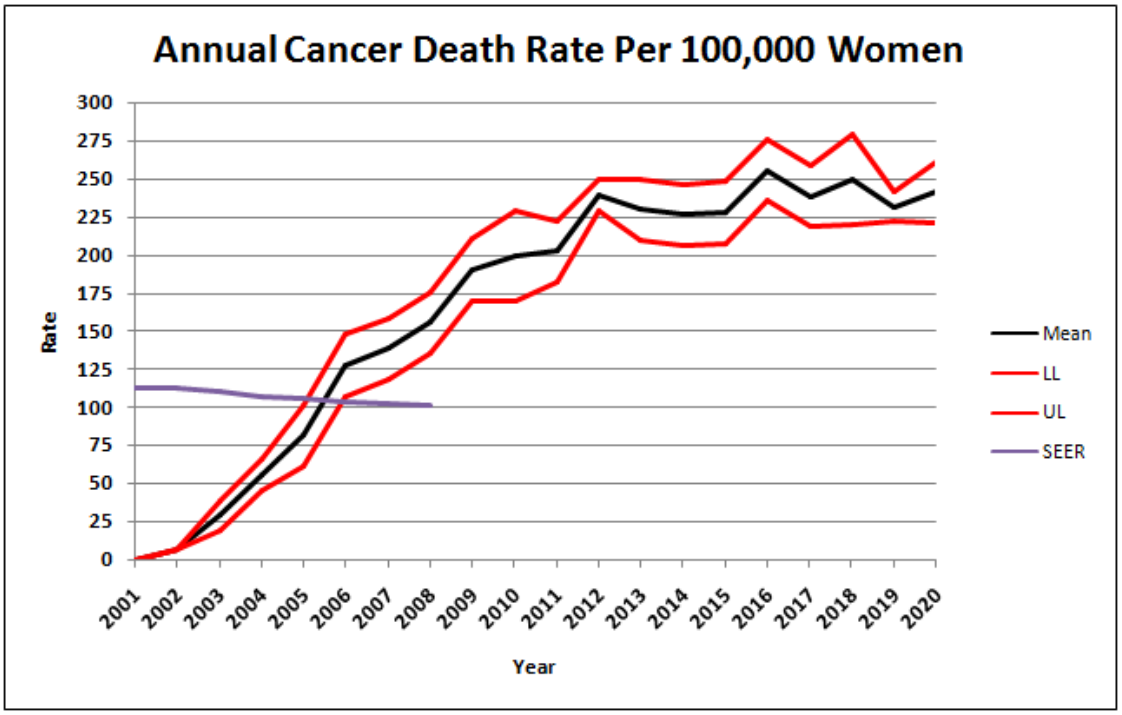


Figure 3-20. Phase I Annual Cancer Death Rates Per 100,000 Women with SEER Data for Comparison

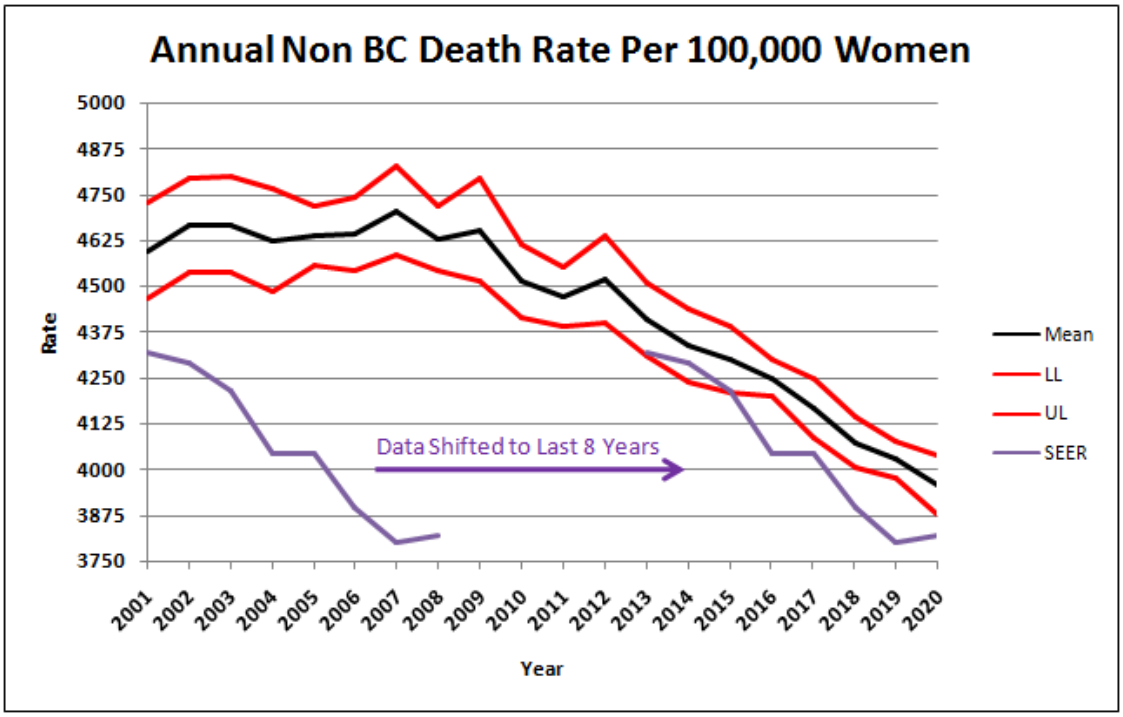


Figure 3-21. Phase I Annual Non Breast Cancer Death Rates Per 100,000 Women with SEER Data for Comparison

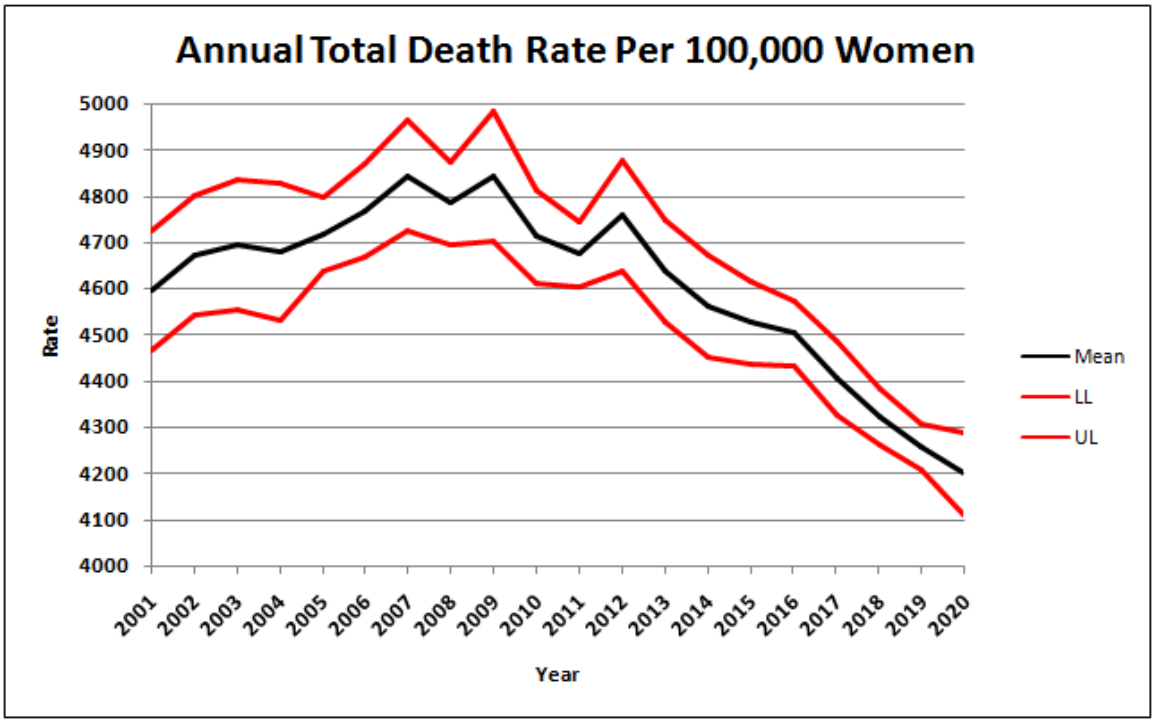


Figure 3-22. Phase I Annual Total Death Rates Per 100,000 Women

Cancer death rates require a "warm-up" period before achieving a "representative" behavior pattern in about the year 2012 similar to prevalence. The reason for this is that we began with a population that was cancer free, so it took several years for cancer to start having lethal effects on the women in the model. Again, after several years the cancer death rate in the simulated population begins to mimic the real population; and it is the "warmed-up" simulation results that are useful to compare with SEER age-adjusted data. Since we have perfect annual screening and in addition we are not treating any cancers, we would expect the cancer death rate from the natural history model to be significantly higher than the breast cancer death rate actually occurring in the population. In the first place, we are detecting a greater number of cancers. Secondly and most importantly, we are not treating cancers at all, whereas in the actual US population the majority of breast cancer deaths are treated. Beyond the "warm-up" period, the cancer death rate from the model is about 2.5 times as large as the real cancer death rate. Again, we know the death rate should be greater, but it is impossible to know how much greater it should be. Thus, we feel these are accurate estimates of the mean cancer death rate for the period 2012–2020 under the assumption of perfect annual screening and without treating cancer.

Deaths from causes other than breast cancer also require a "warm-up" period. We would expect the "warmed-up" death rate from causes other than breast cancer to be similar to the SEER age-adjusted death rate. Of course, we only have SEER data for the period 2001–2008, so we shifted this 8 years of data to the last 8 years of the model after it has had a chance to warm up; and then our model tracks the value and the trend remarkably well. This is further evidence that the natural history model is an accurate representation of breast cancer screening amongst older US women.

3.6.1.4 Population Size and Growth

As previously stated, the initial population size in 2001 was chosen to be 0.10% of the actual size of the US population of women who were at least 65 at the end of the year 2000, which was obtained from US census data [99]. We chose this value because it allows

us to easily compare output from our model with actual data from the population, by either multiplying the output of our simulation model by 1,000 or dividing the actual data from the population by 1,000. In Chapter 1, we asserted that women over 65 would soon become the prevalent breast cancer patient cohort, partially because the older female population is expected to grow rapidly in the near future, and additionally because of their increased risk of developing breast cancer. The age distribution of the initial simulated population was derived from 2000 US Census data, and population growth occurs because women turning 65 enter the simulation model on a yearly basis. Our simulated population size can be compared with the actual 65 and older female population size for the period 2001–2010. Figure 3-23 shows the estimated mean population size from the simulation model alongside the actual population data. The graph shows that the simulated population is tracking the actual population extremely well, which leads us to conclude that our population growth model is a valid model for population growth for the future years 2012–2020. This graph also shows that the population size grows by about 50% over the twenty-year simulation, which is representative of the rapidly growing older population in the United States.

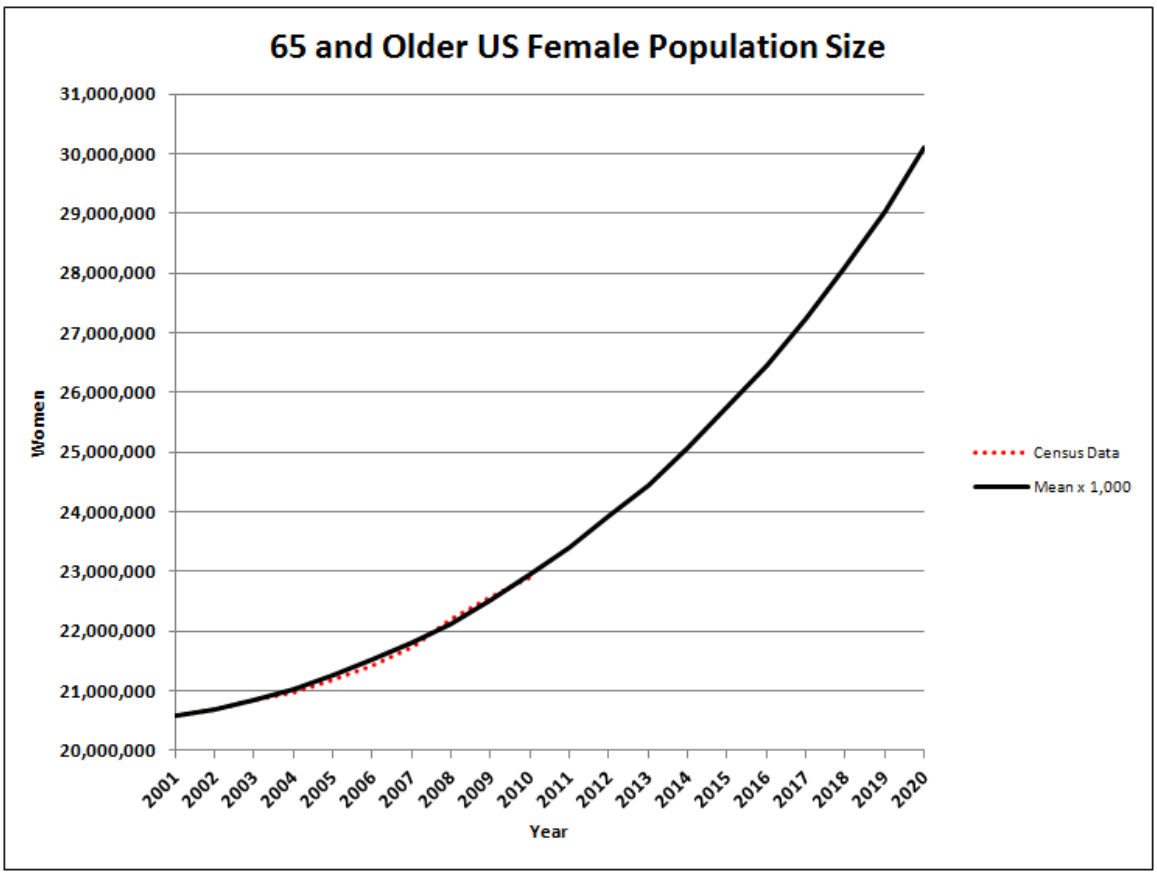


Figure 3-23. Phase I Annual Mean Simulated Population Size with Census Data for Comparison

3.6.2 95% CIs on Theoretical Cancer Performance Measures for the Entire Population Over Twenty Years

This section provides 95% confidence intervals on the expected values of the main cancer performance measures and other population characteristics. The goal of this research is to make recommendations for future breast cancer screening policies, so Table 3-15 contains performance measures for the full twenty years from 2001–2020, for the 11 year "warm-up" period 2001–2011, and for the future period 2012–2020. There are some interesting results in this table that merit a brief discussion.

We reported the number of cancers that were detected by mammography as opposed to clinical detection; and as expected with perfect annual screening, nearly all cancers are detected by mammography before symptoms become present. However, there were still some fast-growing cancers that caused symptoms to become present before mammographic detection. Perfect annual screening over twenty years leads to 470,000 mammograms being administered over this time period; and since the population size is 0.1% of the actual US population, this number can be multiplied by 1,000 to estimate the actual number of mammograms that would be given in the United States if there were perfect annual screening. Equivalent information to that given in Table 3-15 is presented for scenarios in the integrated screening model at the end of Chapter 4.

Table 3-15. 95% CIs on Population Characteristics Based on Natural History Model

Performance Measure	Years	Mean	HW	LL	UL
Number of Invasive Cancers Detected	2001–2020	2,633.90	43.41	2,590.49	2,677.31
	2001–2011	1,344.60	25.99	1,318.61	1,370.59
	2012–2020	1,289.30	23.09	1,266.21	1,312.39
Number of DCIS Cancers Detected	2001–2020	641.70	18.78	622.92	660.48
	2001–2011	321.30	14.36	306.94	335.66
	2012–2020	320.40	14.84	305.56	335.24
Method of Detection - Mammography	2001–2020	2,590.20	42.85	2,547.35	2,633.05
	2001–2011	1,324.50	25.63	1,298.87	1,350.13
	2012–2020	1,265.70	23.28	1,242.42	1,288.98
Method of Detection - Clinical Presentation	2001–2020	43.70	3.44	40.26	47.14
	2001–2011	20.10	2.69	17.41	22.79
	2012–2020	23.60	3.11	20.49	26.71
Number of Cancer Deaths	2001–2065	1,348.00	27.00	1,310.00	1,407.00
	2001–2011	264.00	11.62	244.00	294.00
	2012–2020	570.30	11.42	546.00	593.00
	2020–2065	513.70	17.76	482.00	551.00
Number of Non-BC Deaths with Invasive Cancer	2001–2065	1,285.90	33.40	1,193.00	1,353.00
	2001–2011	327.90	15.29	289.00	352.00
	2012–2020	519.70	15.11	490.00	543.00
	2020–2065	438.30	12.96	414.00	468.00
Number of Mammograms Given	2001–2020	468,833.20	1,439.76	467,393.44	470,272.96
	2001–2011	235,216.60	606.86	234,609.74	235,823.46
	2012–2020	233,616.60	922.14	232,694.46	234,538.74
Number of Mammograms per Invasive Detection	2001–2020	178.09	3.09	175.00	181.18
Number of Mammograms per DCIS Detection	2001–2020	731.71	21.38	710.33	753.09
Number of Mammograms per Cancer Detection	2001–2020	143.21	2.61	140.60	145.82
Number of Non-BC Deaths	2001–2020	21,137.00	78.04	21,058.96	21,215.04
	2001–2011	11,022.70	86.32	10,936.38	11,109.02
	2012–2020	10,114.30	73.18	10,041.12	10,187.48
Number of Survivors	2020	28,827.50	147.23	28,680.27	28,974.73
Initial Population Size	2001–2020	20,582.00	-	20,582.00	20,582.00
Average Population Size	2001–2020	22,529.99	72.07	22,457.92	22,602.06
Final Population Size	2001–2020	28,827.50	147.23	28,680.27	28,974.73
Total Number of Women Simulated	2001–2020	50,799.30	156.58	50,642.72	50,955.88

3.6.3 95% CIs on Stage Distribution at Diagnosis

The stage distribution at diagnosis reveals the percentage of cancers that were detected in the local, regional, and distant stages, respectively. This is a very important statistic, because breast cancer mortality is largely dependent on the stage at diagnosis. According to SEER data, if invasive breast cancer is diagnosed in the local stage, then there is a 20% chance of dying from breast cancer. This increases to 45% if invasive cancer is diagnosed in the regional stage, and to 80% if invasive cancer is diagnosed in the distant stage. Table 3-16 shows 95% CIs on the number of invasive breast cancers diagnosed in each stage and the percentages of invasive breast cancers diagnosed in each stage from the natural history model. Table 3-15 also includes SEER data on the percentages of invasive breast cancer diagnosed in each stage for comparison purposes. Figure 3-25 and Figure 3-27 are pie charts of the stage distribution at diagnosis from the natural history model and from SEER data, respectively. The percentage of cancers diagnosed in the local stage is 20% higher in the natural history model than in SEER data. The comparable percentage of cancers diagnosed in the regional stage was less by 13%, and the comparable percentage diagnosed in the distant stage was less by 4.5%. Clearly cancers are being diagnosed much earlier in the natural history model than they are in the actual population, which is expected with perfect annual screening. As would be expected, most cancers were diagnosed in the local stage, and very few were diagnosed in the regional and distant stages. However, cancers remain untreated in the natural history model; and thus the benefits of these earlier stages at detection cannot be determined until the execution of the integrated screening model, where treatment is considered. We assert that this estimate of the stage distribution at diagnosis from the natural history model is a reasonable estimate of the stage distribution at diagnosis with perfect annual screening. It is impossible to actually know the stage distribution at diagnosis with perfect annual screening, but cancers should be detected much earlier than in the actual population, and the natural history model captures this expected behavior.

Table 3-16. Stage Counts and Distribution at Diagnosis: Results of Natural History Model vs SEER Data for Comparison

COUNTS	Mean	HW	LL	UL
Local	2191.2	29.1	2162.1	2220.3
Regional	410.5	6.3	404.2	416.8
Distant	45.3	6.5	38.8	51.8

DIST	Mean	HW	LL	UL
Local	82.78	0.38	82.40	83.16
Regional	15.51	0.32	15.19	15.83
Distant	1.71	0.25	1.46	1.96

SEER DIST	Mean	Sim Mean	Diff
Local	29.5	85.027	55.5
Regional	44.0	13.457	-30.5
Distant	16.3	1.516	-14.8
Unstaged	10.2	-	-

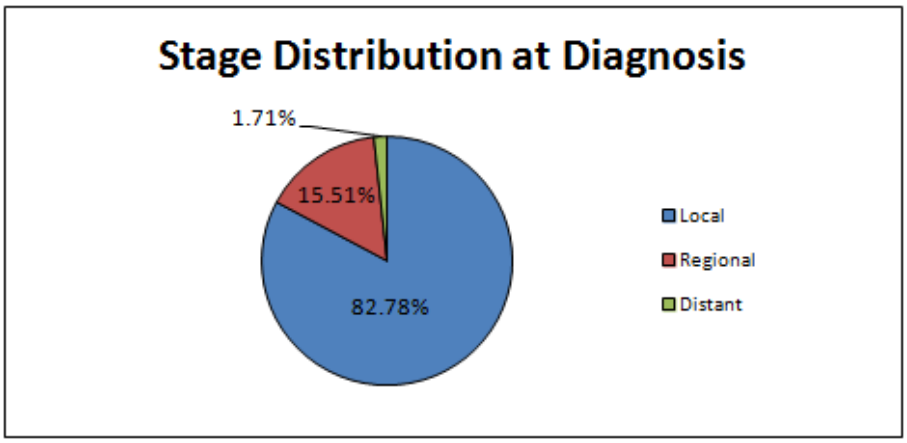


Figure 3-24. Simulated Stage Distribution at Diagnosis Pie Chart

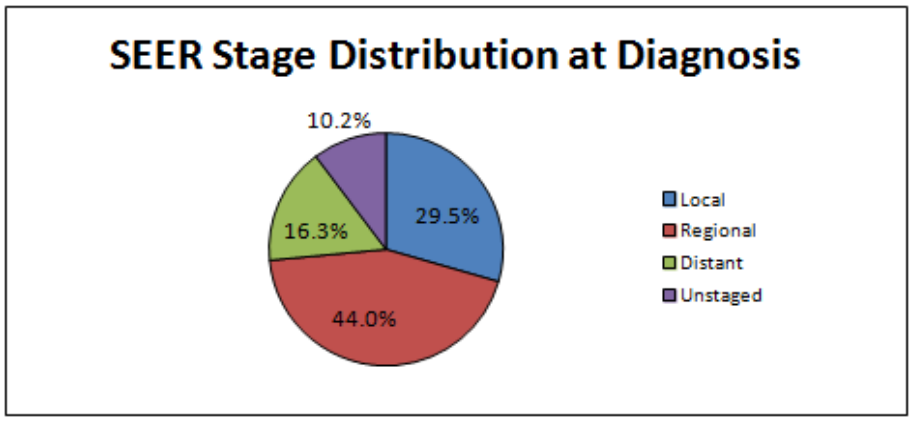


Figure 3-25. SEER Stage Distribution at Diagnosis Pie Chart

3.6.4 95% CIs on Invasive Cancer Outcomes

Another result of interest is the distribution of invasive cancer outcomes. There are three potential outcomes for a woman who is diagnosed with breast cancer in the natural history model: breast cancer death, death from a cause other than breast cancer, or survival through the year 2020. Of particular interest is the percentage of women with invasive cancers who die from a cause other than breast cancer. Since there is no treatment in the natural history model, women who are diagnosed with breast cancer and die from a cause other than breast cancer would not benefit from treatment. Ideally, we would like to maximize the percentage of women with invasive cancers who survive through 2020, and minimize the percentage of breast cancers resulting in death. These outcomes from the natural history model are later compared to outcomes from the integrated screening model. Figure 3-26 is pie chart of the distribution of invasive cancer outcomes, and Table 3-17 shows the 95% CI's on those distribution percentages.

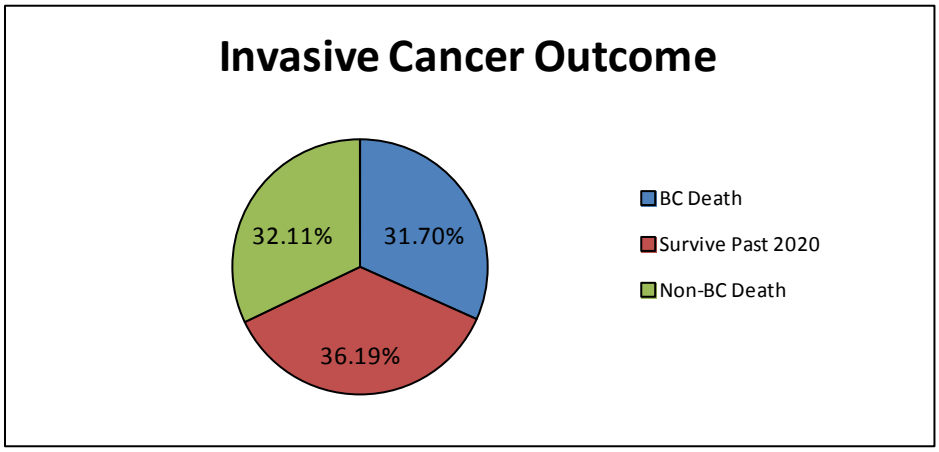


Figure 3-26. Pie Chart of Distribution of Invasive Cancer Outcomes from Natural History Model

Table 3-17. 95% CIs on Distribution of Invasive Cancer Outcomes from Natural History Model

Performance Measure	Mean	HW	LL	UL
Percentage of Invasive Cancers Leading to BC Death	31.7	0.69	31.0	32.4
Percentage of Women with Invasive Cancers Still Surviving (2020)	36.2	0.66	35.5	36.9
Percentage of Invasive Cancers Leading to Non BC Death	32.1	0.57	31.5	32.7

3.6.5 95% CIs on Tumor Growth-Rate Classifications

As previously stated, there is evidence that suggests breast cancers grow more slowly in older women than they do in younger women. To account for this, the tumor growth-rate distribution has a mean that decreases with age, leading to slower tumor growth rates than were implied by the original distribution specified by Norton [69]. Tumor growth rates were divided into six classifications: very slow, slow, slightly slow, slightly fast, fast, very fast. Table 3-18 contains 95% CIs on the expected number of tumors in each class and the percentage of tumors in each class. Figure 3-27 is a pie chart for the percentage of tumors in each class. These results show that the majority of tumors are slow growing, and this is expected with an older population. This also implies that the majority of tumors are detected at small sizes with perfect annual screening, and again this is something that is expected. In the next section we use each woman's tumor growth rate to infer the new distribution of tumor growth rates, which is still lognormal but has a smaller mean.

Table 3-18. Phase I Tumor Growth Rate Counts and Distribution CI Data After Age Adjustment

COUNTS	Mean	HW	LL	UL
Very Slow	1376.5	34.19	1342.31	1410.69
Slow	557.2	12.58	544.62	569.78
Slightly Slow	419.3	16.26	403.04	435.56
Slightly Fast	198.2	5.48	192.72	203.68
Fast	62.7	7.50	55.20	70.20
Very Fast	20.0	3.92	16.08	23.92
DISTRIBUTION	**Mean**	**HW**	**LL**	**UL**
Very Slow	52.25	0.75	51.5038	53.0038
Slow	21.16	0.52	20.6418	21.6818
Slightly Slow	15.92	0.54	15.3781	16.4581
Slightly Fast	7.53	0.23	7.2978	7.7578
Fast	2.38	0.27	2.1076	2.6476
Very Fast	0.76	0.15	0.6108	0.9108

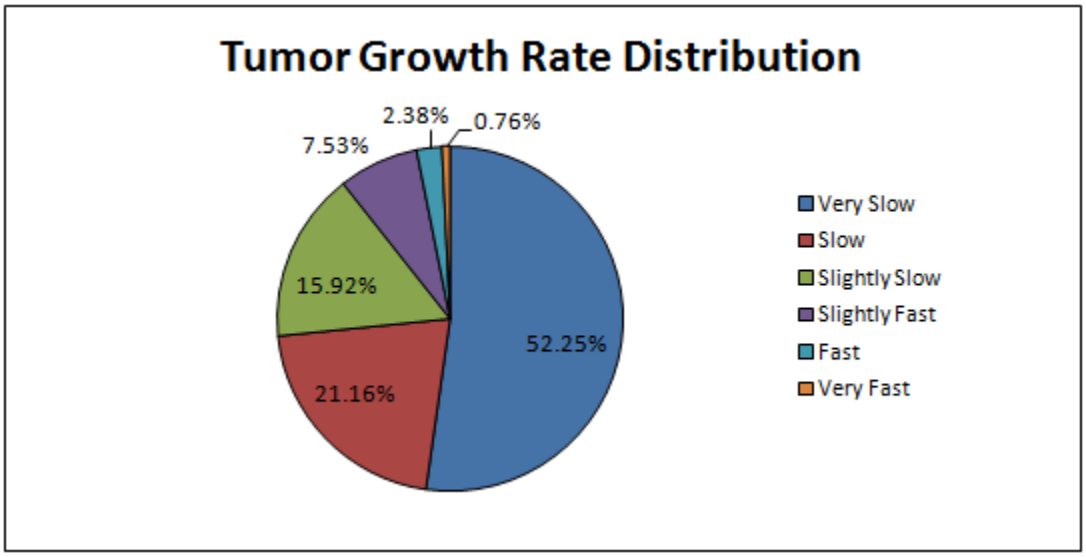


Figure 3-27. Phase I Categorical Distribution of Tumor Growth Rate for Women over 65

3.6.6 Distributions Inferred from Individual Observations of Cancer Attributes

The StatFit 2.0 statistical distribution fitting software [36] was used to fit probability distributions to cancer attributes that may be of interest to clinicians and other researchers. Each woman from all ten simulated populations who is diagnosed with cancer produces a cancer history; and in addition, certain cancer attributes of interest were written to a separate text file that can easily be loaded into StatFit for analysis. StatFit does have a limit of 8,000 observations, but 8,000 samples is certainly adequate for distribution fitting. Large sample sizes also render *p*-values meaningless in most cases. The following criteria were used to evaluate the fit of alternative distributions to data: visual inspection of the fitted CDF, visual inspection of the fitted PDF, a P-P plot of the fit, chi-squared and Kolmogrov-Smirinov *p*-values, and the ability of the fitted distribution to capture the first four sample moments.

In this section we simply present Table 3-19, which contains the list of cancer attributes we fitted distributions to, the distribution that was the best fit, and the parameters of the best-fitting distribution. For more detailed information about the fits, including graphs of the fitted CDFs and PDFs, the P-P plots, the moments of the sample data and fitted

distributions, and the associated goodness-of-fit *p*-values, see Appendix D: Inferred Fitted Distributions from Natural History Model.

StatFit has the capability to fit the following continuous probability distributions:

1. Beta
2. Chi Squared
3. Erlang
4. Exponential
5. Gamma
6. Lognormal
7. Pearson 5
8. Pearson 6
9. Power Function
10. Rayleigh
11. Triangular
12. Uniform
13. Weibull

Table 3-19. Phase I Inferred Distributions and Their Parameters

Performance Measure	Distribution	Parameters
Time for Tumors to Grow to Minimum Size Detectable by Mammography (months)	Lognormal	μ = 38.6, σ = 29.4, *Min (Offset)* = 4.0
Time for Tumors to Become Clinically Detectable or Symptomatic (months)	Pearson 6	β = 106.4, p = 2.8, q = 4.7, *Min (Offset)* = 3.0
Time Until Tumors to Become Lethal (months)	Pearson 6	β = 235.4, p = 2.82, q = 4.81, *Min (Offset)* = 7.0
Size When Tumor Becomes Detectable by Mammography (mm)	Lognormal	μ = 2.27, σ = 0.639, *Min (Offset)* = 1.0
Size When Tumor Becomes Clinically Detectable (mm)	BETA	α_1 = 2.27, α_2 = 0.639, *Min (Offset)* = 1.0
Cancer Onset Age for Women 65 and Older	Pearson 6	β = 383.7, p = 23.8, q = 248.6, *Min (Offset)* = 36.0
Tumor Growth Rate for Women 65 and Older	Lognormal	μ = -3.84, σ = 0.75, *Min (Offset)* = 0.0

3.7 Relationship with Integrated Screening Simulation Model

As indicated in Chapter 1, the output of the natural history model is used as input to the integrated screening model, and the results from the natural history model can be compared with the results of different screening scenarios simulated in the integrated screening model. The results presented earlier in this section are those we felt might be of most interest to researchers or clinicians, but they are not vital inputs to the integrated screening model. The initial database of women's attributes used in the natural history model contains only women who have not previously been diagnosed with breast cancer and contains no information about each individual's cancer history or future. The natural history model essentially creates cancer histories for each woman who enters the model, and those women who get cancer have their cancer attributes recorded along with their original risk-factor attributes at the end of the simulation. This database is used as the input to the integrated screening model, essentially providing a trace or realization of each woman's entire cancer history. For example, when a woman who has cancer is screened with mammography in the integrated screening model, the size of her primary tumor at the time of the screening is known, and so is probability of detection as a function of tumor size. Thus, we can compute the probability of detection accurately because each woman's tumor size at any point in the model is known. The integrated screening model provides an environment in which the effects of alternative screening policies on individuals can be measured and compared. Chapter 4 details the development of the integrated screening simulation model, with a detailed analysis of the results and screening policy recommendations for older women.

CHAPTER 4: INTEGRATED DES/SD SCREENING SIMULATION MODEL

This chapter focuses on the detailed development and results of the integrated DES/SD screening simulation model. This model is made up of a number of interacting submodels, namely: a) the screening submodel; b) the treatment submodel; c) the survival and mortality submodel; d) the costing submodel; and e) the population level SD submodel. The most important input to this model is the modified sampling database that is created by the natural history and that now contains information about breast cancer risk factors and entire breast cancer histories from all women in the simulated population. Stochastic details such as screening events, diagnostic procedure decisions, and treatment decisions are modeled using DES; however these stochastic details are subject to the influence of the population-level SD model, which focuses on adherence to screening. Like the natural history model, the time-step for the integrated screening model is one year.

The integrated screening model is unique in that it has elements of all three levels of models discussed in Chapter 2. The level-1 element is the natural history model of the onset and evolution of breast cancer within the sampled population, modeled using DES. The level-2 elements include screening, treatment, and survival; these elements are modeled using DES and represent the stochastic discrete events that take place in the integrated screening model. The level-3 elements are modeled using SD and will include structural elements of the breast cancer screening system that describe the role of the physician, the government, and the overall health care delivery system. Models in the literature typically focus on elements from one of these three levels. The hybrid DES/SD model developed in this research captures the structural behavior of the system through the operation of the SD submodel; and the DES submodel generates the stochastic details that provide additional insight, as opposed to considering the modeling problem from a single perspective.

This integrated simulation model is used to determine the cost-effectiveness of alternative screening policies for older US women, as well as to provide numerous breast cancer related population statistics for each alternative policy. A screening policy consists of

a screening interval and an age at which screening should stop. The screening interval and the stopping age can be a function of a woman's individual breast cancer risk factors, her annual risk, or any other individual characteristics captured in the simulation. Each of the 10 randomly sampled populations from the natural history model are simulated with a screening policy in place, and the results of using that screening policy are delivered via in Microsoft Excel. It is important to note that the alternative screening policies tested in this simulation were only in effect for the future years 2012–2020. As we have previously indicated, for the past years 2001–2011, there was no well-defined screening policy for women of this age group. However, population-level breast cancer data from SEER and the BCSC for the majority of those years is available. Thus, we evaluated multiple theoretical screening policies and selected the one that produced results which most closely resembled actual data. Our ability to produce results which resembled actual data with reasonable theoretical screening policies is one piece of evidence that our model is a valid representation of breast cancer screening in the US. With the cancer history of the every woman in each population now known, each woman's tumor size at any given time can be computed along with the probability of detection given that tumor size. Thus, both false positives and false negatives are considered for screening mammograms. If a screening mammogram is positive, whether it is a true positive or false positive, the woman is sent for a diagnostic procedure. False positives are considered for diagnostic procedures but false negatives are not; and if a woman with cancer is sent for a diagnostic procedure (additional imaging), then she will also be sent for additional work-up (biopsy). Additional work-up procedures include open biopsy, core needle biopsies, and FNA; and we assume these exams are perfect. It is then decided whether a woman with cancer will be treated, and her survival is then determined according to whether or not she receives treatment.

If treatment is received, then the number of years until death from cancer given treatment is sampled from a conditional distribution that is a function of the cancer stage (local, regional, or distant) at diagnosis. The cancer's stage at diagnosis is determined by the Plevritis stage progression model [75] just as it is in the natural history model. It is possible

for women who are treated to experience death from other causes during or after treatment, and this probability is computed on a yearly basis according to the breast cancer adjusted life-tables. If a woman with cancer who is treated experiences death from other causes in the natural history model, then she is assumed to die at the same time from other causes in the integrated screening model. Women who are not treated are assumed die of cancer or other causes exactly as they did in the natural history model. For women who are treated, it likely that they will live longer than they did in the natural history model, which did not consider treatment; and the difference in the number of years lived are called "life-years saved".

We compute life-years saved and "quality-adjusted life-years (QALYs) saved," which are life years saved discounted by a utility factor that adjusts for the person's quality of life for that year. For example, a woman living with distant cancer will have a lower health utility for that year than a woman living cancer-free or a woman in the local stage of cancer. Nevertheless, screening and treatment costs money; and in the case of women of age 65+ in the United States, it costs society money, as these women are reimbursed for these costs by Medicare, which society pays for in Medicare taxes. The costs can be divided into four categories: screening costs, diagnostic imaging costs, additional work-up costs, and treatment costs. All costs are discounted forwards and backwards to 2012 dollars so that costs from the future and the past are compared on the same basis. The recognized measure of cost-effectiveness for population screening policies is cost per QALY saved, that is, how much does society has to spend to save a QALY. While still debated, one method for determining the cost effectiveness of screening policies is by comparing the cost per QALY saved to the Gross Domestic Product (GDP) per capita for the country in which the screening policy is being used [107]. If the cost per QALY saved is less than three times the GDP per capita, then the screening policy is considered cost-effective; and if the cost per QALY saved less than the GDP per capita, then the screening policy is considered very cost-effective [107]. The 2010 GDP per capita in the US is $47,184.47 [106]. Incremental cost per QALY saved is the difference between cost per QALY saved for two policies, and this measure is used to determine if one screening policy is significantly more cost-effective than another. Thus, the

major outcome from this model is a screening policy or set of screening policies that minimize some screening-policy performance measure such as cost per QALY saved. In particular, our simulation optimization may seek to minimize the cost of screening, diagnosis, and treatment per QALY saved for cancers detected by mammography in the period 2012–2020. Breast cancer deaths averted by screening is another important performance measure that we may choose to maximize subject to some restrictions on cost.

The remainder of this chapter describes the development of the integrated screening simulation model, including a detailed discussion of all the model assumptions and a description of each of the interacting submodels. We begin with the stochastic details modeled using DES, we describe the population level SD model, and finally we describe how the two models exchange information and affect one another while running simultaneously. Lastly, we present the results of using this model the determine cost-effective screening or life saving policies, provide a detailed discussion of these results, and also discuss verification and validation.

4.1 DES Submodel

Detailed individualized components of this simulation are modeled using discrete-event logic, and other population-level components are modeled using system dynamics equations. Individual characteristics, screening appointments, treatment decisions, survival after treatment, costs, and stochastic elements are all modeled using discrete-event logic.

When the integrated screening model is executed, first a user interface is displayed that allows the user to select values for variables using option buttons and drop-down menus. As detailed in Section 4.4, the user interface provides an intuitive method for rapidly assigning values to key parameters the control the operation of the model. The following parameters are initialized each time the model is executed:

1. Simulation length = 20 years;
2. Simulation start year = 2001;
3. Initial population size = 20,582 women;
4. Screening interval for 2001–2011 (default value = 1 year);
5. Stopping age of screening for 2001–2011 (default value = Beta distributed with a mode of 75);
6. Screening policy type for 2012–2020
    a. Screening interval for all women for 2012–2020;
    b. Screening intervals for high risk women and low-risk women separately for 2012–2020 (requires user to specify percentage of women that are "high risk")
    c. Screening Intervals for each woman based on their individual risk factors for 2012–2020;
7. Stopping age of screening for 2012–2020 (default vale = deterministically specified);
8. The level of breast cancer research for 2001–2011 and 2012–2020 separately (default value = 2 (medium));
9. The level of breast cancer public advertising for 2001–2011 and 2012–2020 separately (default value = 2 (medium));
10. The trend in the number of breast cancer screening facilities for 2012–2020 (default value = 0 (constant)); and
11. The trend in the capacity of breast cancer screening facilities for 2012–2020 (default value = 1 (slightly increasing)).

Women (or entities) enter the integrated screening model exactly as they entered the natural history model. At the end of the natural history model, each woman's individual attributes, including those attributes representing her cancer history, were written to a database; and the entities for different women are inserted into the database in the same order that women entered the natural history model (as opposed to the order that they left the

model). We sequentially sample women from this database and assign their individual attributes, allowing the correct number of women to enter the population on a yearly basis, exactly as they did in Phase I. Thus, we are resimulating (re-creating) the *same* 10 randomly sampled populations that were used in the natural history model; but now we have knowledge of each woman's cancer history (or lack thereof), so that we can use the integrated screening model to perform 10 independent replications of the sample path (trajectory) of the selected population over the given time horizon. This approach enables us to compute means and confidence intervals (CIs) for the expected values of all selected performance measures.

Upon entering the integrated screening model, each woman has her individual attributes assigned from the database created by the natural history model, and then she enters the screening portion of the DES submodel. The time-step for discrete events is one year, the same as in the natural history model. The following events occur each and every year for each woman:

1. Age is increased by one.
2. If no comorbid diseases are present, then the probability of developing a comorbid condition within the last year is computed and used to determine health status.
3. If a woman will have a clinical detection in the current year and it will be prior to a mammographic detection, then she has a clinical detection and proceeds to diagnostic exams.
4. The screening submodel determines whether or not the woman should enter by determining if she should have an appointment based on the screening interval and stopping age specified by the screening policy. If she meets the criteria for screening that year (i.e., her age is evenly divisible by the screening interval and is less than the stopping age), then she is screened for breast cancer with mammography, provided she adheres to the screening policy that year. If she is

not eligible for screening, then she has no other events in the current year, and she waits for the start of the next year to repeat the process.

The screening submodel enforces the screening policy on an individual basis, samples the probability of adherence to each screening appointment on an individual basis, and determines the type of screening, diagnostic, and work-up exams on an individual basis. False positives and negatives can occur for screening exams, diagnostic exams can have false positive results but not false negative results, and work-up exams such as biopsies are assumed to be perfect. If cancer is present, then the stage at diagnosis is determined according to the Plevritis breast cancer staging model just as in the natural history model.

The treatment submodel is rather simple, as the details of treatment are not currently the focus of this research. Only women who have a breast cancer detected enter the treatment submodel. Through communications with our breast cancer experts, we determined the probability a woman over 65 would be treated for breast cancer given her age and whether or not she has other comorbid diseases. If a woman does not receive treatment, then her death age will be exactly the same as it was in the natural history model, where treatment was not considered; and her cause of death will be the same as it was in the natural history model. If a woman does receive treatment, then the time until death from breast cancer is computed in the survival submodel; and there are different curves for the three different stages of breast cancer, since survival largely depends on the stage of breast cancer at diagnosis. Each year the breast cancer adjusted life-tables are used to compute the probability that death is caused by something other than breast cancer (i.e., it is a non–breast cancer death).

The survival submodel is only entered by a woman who has breast cancer detected and who receives treatment according to the treatment submodel. The survival submodel calculates the ages at which death would be caused by breast cancer or some other cause, and the lower age governs the death age and cause of death. Then utilities from breast cancer

literature are used to compute the number of life-years saved as well as the number of QALYs saved.

The costing submodel keeps track of all of the costs within the integrated screening model. There are four major types of costs: the cost of screening exams, the cost of diagnostic exams, the cost of work-up exams, and the cost of treatment. These costs are added up and used to compute the cost-effectiveness of each alternative screening policy.

4.1.1 Screening Submodel

If it has been determined that a woman is eligible for screening during a given year according to the prescribed screening policy and that she will adhere to the screening policy during that year, then she enters the screening submodel. Since we have knowledge of each woman's entire cancer history, we know whether or not cancer is present at the time of screening and the size of all tumors at all times. If the woman being screened does not have cancer, then there will not be a cancer detection: although there could be a false positive screening exam and even a false positive diagnostic exam, the work-up biopsy will reveal that she does not have cancer. A false positive screening exam will not necessarily lead to a false positive diagnostic exam, but such an outcome is possible. If the woman being screened does have cancer, then her tumor size at the time of screening is known and is used to determine the probability of detection by screening. If the cancer is detected, then the woman will go onto diagnostic imaging and biopsy, which will confirm the positive diagnosis. If the cancer is not detected (e.g., it is too small, it is missed by radiologists, etc), then no diagnostic exam will be performed, the overall result is counted as a false negative, and the woman will continue to be screened according to the screening policy. With the help of our breast cancer experts, we have designed three types of screening policies that can be used for the period 2012–2020: interval screening, risk-based screening, and factor-based screening. For the period 2001–2011, there was no well-defined screening policy, so we know that a risk-based or factor-based policy was not in effect during that period, and it would be incorrect to use these types of policies for that time period. Thus, when selecting the

screening policy that most closely matches SEER data, we only test basic interval screening policies. However, we are mostly interested in how alternative screening policies can impact the future; and using this model, we have the ability to test over 10^{30} different screening policies, but not all at once.

4.1.1.1 Screening Policies

We have constructed three types of screening policies that can be used for the period 2012–2020: interval screening, risk-based screening, and factor-based screening. Interval screening assigns the same screening interval and stopping age to every woman in the population. Risk-based screening allows for different screening intervals for both high-risk and low-risk women. Factor-based screening allows for different screening intervals based on one individual breast cancer risk factor, such as breast density, the number of first degree relatives with breast cancer, or some combination of these factors. These individualized risk-based and factor-based policies may be more cost-effective than simply basing the screening interval on age, a practice for which there is no good basis. For the "model calibration" period 2001–2011, the screening policy is set to the interval screening policy that yields results most closely matching SEER data from that time period. The screening policy changes in 2012 to the prescribed policy, which lasts through 2020.

These three types of screening policies are essentially three different methods of assigning a screening interval to each woman in the population. However, a screening policy is made up of both a screening interval and a stopping age. For our purposes, it is really necessary to consider the different situations presented by screening in the past and screening in the future. In the past, there was no well-defined stopping age for screening, so women stopped screening at different ages. In order to account for this, we have a set of beta distributions for stopping age, all having a minimum of 65 and a maximum of 100, but with modes varying from 70 to 100 in steps of 5. This gives us seven different stopping ages that we can use in combination with different screening intervals and policies for the past. In the future, we actually want to assign a fixed (constant, deterministic) stopping age at which

screening is stopped for everyone in the population. For future policies, possible values of the stopping ages range from 70 to 100 in steps of 5. Stopping ages are for everyone in the population, and they are not dependent on risk classification or risk factors. Future work could include allowing stopping ages to be based on individual factors, but this creates another dimension to the optimization and would slow it down considerably.

4.1.1.1.1 Interval Screening

Interval screening is the simplest of the types of screening policies. The screening intervals considered are every one, two, three, four, and five years. Whether using deterministic or stochastic stopping ages, there are seven potential values. This leaves us with 35 combinations of screening interval and stopping age, or in other words 35 possible screening policies of this type.

4.1.1.1.2 Risk-Based Screening

Risk-based screening was suggested by our breast cancer experts, and we use the Barlow one-year risk to determine each woman's overall risk of being diagnosed with breast cancer. The only risk factor that changes over time for each woman in the model is her age, while all other risk factors remain constant. Thus, we recorded each woman's one-year risk upon entering the natural history model, determined the percentiles of the data given in Table 4-1, and created a histogram of the data given in Figure 4-1. We allowed the definition of high risk to vary, so that it can correspond to the top 5%, 10%, 15%, or 20% of women of age 65+ in terms of one-year risk.

Table 4-1. Percentiles for Barlow 1-Year Risk Distribution From Women in the Simulation Model

Percentile	Data Values
0%	0.00105
5%	0.00276
10%	0.00367
15%	0.00442
20%	0.00495
30%	0.00565
40%	0.00626
50%	0.00689
60%	0.00751
70%	0.00822
80%	0.00909
85%	0.00974
90%	0.01054
95%	0.01187
100%	0.03026

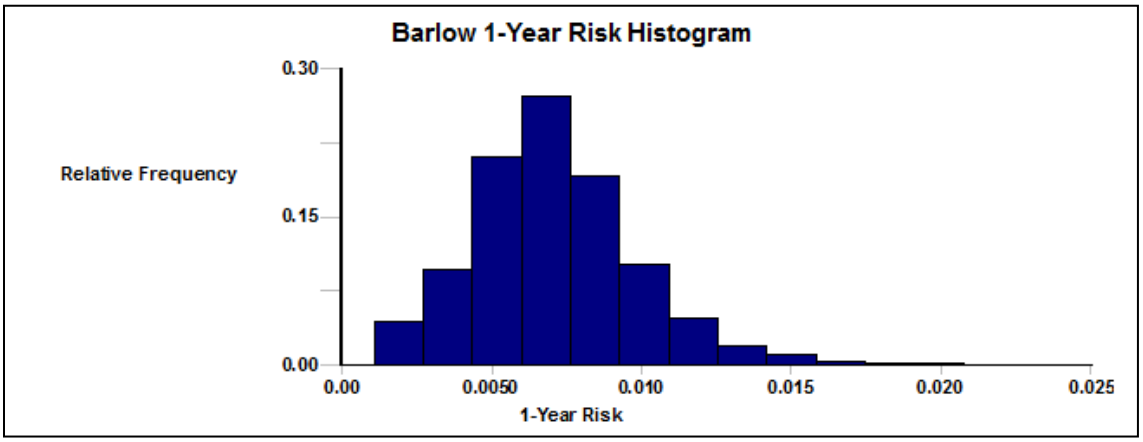


Figure 4-1. Histogram of Barlow 1-Year Risk for Women Entering the Simulation Model

Once a definition of high risk is chosen, different screening intervals can be selected for high-risk and low-risk women, with the same options of every one, two, three, four, or five years for each risk classification. In creating optimizations that search through all risk-based policies, we enforce the constraint that the screening interval for low-risk women be greater than or equal to the screening interval for high-risk women, leaving us with 15 combinations of screening intervals. Given that there are 15 combinations of screening intervals, four different definitions of screening, and seven different potential stopping ages, we have 420 risk-based screening policies. An example of a risk-based screening policy is that high-risk women are screened every year, and low-risk women are screened every two years, where by definition high risk corresponds to the top 5% of women of age 65+ in terms of one-year risk; and screening is stopped at the age of 85.

4.1.1.1.3 Factor-Based Screening

Factor-based screening is the most versatile and complex type of screening policy that the integrated screening model can handle. There are eleven risk factors used in the Barlow risk model: five of those factors have three levels, three factors have four levels, two factors have five levels, and one factor has six levels, as shown in Table 4-2.

Table 4-2. Table of Risk-Factors, Their Levels, and Their Categorical Variables

Risk Factor	**Categories and Categorical Variables (43 Total)**					
Age	65-69 FreqAge7	70-74 FreqAge8	75-79 FreqAge9	80+ FreqAge10		
Ethnicity	Non-Hispanic FreqEth0	Hispanic FreqEth1	Unknown FreqEth9			
Race	White FreqRace1	Asian FreqRace2	Black FreqRace3	Native American FreqRace4	Other FreqRace5	Unknown FreqRace9
BMI (kg/m^2)	<25 FreqBMI1	25-29.99 FreqBMI2	30-34.99 FreqBMI3	>35 FreqBMI4	Unknown FreqBMI9	
Age at First Birth	<30 FreqAAFB1	≥30 FreqAAFB2	No Children FreqAAFB3	Unknown FreqAAFB9		
Previous Breast Procedure	No FreqPrevProc1	Yes FreqPrevProc2	Unknown FreqPrevProc9			
1st Deg Relatives with BC	None FreqFirstDeg1	1 FreqFirstDeg2	≥2 FreqFirstDeg3	Unknown FreqFirstDeg9		
Hormone Therapy Use	No FreqHRT1	Yes FreqHRT2	Unknown FreqHRT9			
Surgical Menopause	No FreqSurgMeno1	Yes FreqSurgMeno2	Unknown FreqSurgMeno9			
Result of Last Mammogram	Negative FreqLastMamm1	False Positive FreqLastMamm2	Unknown FreqLastMamm9			
Breast Density (BI-RADS)	1 FreqDensity1	2 FreqDensity2	3 FreqDensity3	4 FreqDensity4	Unknown FreqDensity9	

Factor-based screening works in the following way:

1. The user selects the factors on which the screening interval should be based. The number of selected factors can range from only one factor to all eleven risk factors.
2. For those factors that are selected, a screening interval is specified for each level of that factor. For example, age is divided into four levels, so an interval must be specified for each age group.

a. If only one factor is selected, then each woman's screening interval is determined directly by her individual level for that factor.
b. If more than one factor is selected, then there will be a screening interval assigned to each woman for each factor selected; and the smallest such interval (taken over all selected factors) is used as the screening interval for that woman. Table 4-3 is a visual aid to this process.

Table 4-3. Method for Assigning Factor-Based Screening Intervals to Each Woman

Risk Factor	**Screening Frequency Attribute**	**Possible Attribute Values for an Individual {Variable Values}**
Age	FreqAge	{FreqAge7, FreqAge8, FreqAge9, FreqAge10}
Ethnicity	FreqEthnicity	{FreqEth0, FreqEth1, FreqEth9}
Race	FreqRace	{FreqRace1, FreqRace2, FreqRace3, FreqRace4, FreqRace5, FreqRace9}
BMI (kg/m^2)	FreqBMI	{FreqBMI1, FreqBMI2, FreqBMI3, FreqBMI4, FreqBMI9}
Age at First Birth	FreqAAFB	{FreqAAFB1, FreqAAFB2, FreqAAFB3, FreqAAFB9}
Previous Breast Procedure	FreqPrevProc	{FreqPrevProc1, FreqPrevProc2, FreqPrevProc9}
1st Deg Relatives with BC	FreqFirstDegree	{FreqFirstDegree1, FreqFirstDegree2, FreqFirstDegree3, FreqFirstDegree9}
Hormone Therapy Use	FreqHRT	{FreqHRT1, FreqHRT2, FreqHRT9}
Surgical Menopause	FreqSurgMeno	{FreqSurgMeno1, FreqSurgMeno2, FreqSurgMeno9}
Result of Last Mammogram	FreqLastMamm	{FreqLastMamm1, FreqLastMamm2, FreqLastMamm9}
Breast Density (BI-RADS)	FreqDensity	{FreqDensity1, FreqDensity2, FreqDensity3, FreqDensity4, FreqDensity9}

The actual screening frequency for an individual woman can be computed using Equation (4.1).

$$\min\left\{\text{Freq}\left\{\begin{array}{c}\text{Age, Ethnicity, Race, BMI, AAFB, PrevProc,}\\ \text{FirstDegree, HRT, SurgMeno, LastMamm, Density}\end{array}\right\}\right\} \tag{4.1}$$

As with the other two types of screening policies, the user must specify a stopping age for screening from amongst the seven values. With a little mathematical insight, one can see that if the screening policy is based on all eleven factors and a stopping age, there are 16,329,600 screening policies. If we then allow for the possibility of using any subset of the eleven risk factors, the number of screening policies is so large as to preclude any attempt to optimize system performance by total enumeration. Thus, we must have an intelligent method for searching through these screening policies to find the policy or group of policies that is the most cost-effective.

4.1.1.2 Screening Exams

For the purposes of our model, screening exams are assumed to use either film or digital mammograms. Both false positive and false negative screening exams are possible; and although these results are not dependent on whether film or digital mammography is used (data limitation), the probabilities of these events are affected by the level of screening technology (see the SD submodel in Section 4.2 for details). The only difference between a digital and film mammogram in our model is the cost of each exam. For the period 2001–2009, the BCSC has data on the total number of mammograms performed and on the number of digital mammograms performed; see Table 4-4. We use this data set to determine the percentage of digital mammography. Upon examining the trend, we set a maximum percentage of digital mammograms at 95%, and we anticipate that this value be reached by 2015. We then fit a fourth-degree polynomial model to the percentage of digital mammograms during the period 2001–2009, and used this curve (unadjusted R^2 = 0.991) to

predict the percentage of digital mammograms during the period 2010–2014. Figure 4-2 shows a graph of this percentage over the entire model time horizon from 2001 to 2020.

Table 4-4. BCSC Digital Mammography Data

Year	Total Number of Mammograms	Number of Digital Mammograms	% Digital
2001	716,432	7,070	0.987%
2002	762,526	23,252	3.049%
2003	744,179	40,171	5.398%
2004	720,718	59,465	8.251%
2005	595,505	80,581	13.532%
2006	490,465	134,981	27.521%
2007	482,256	198,981	41.260%
2008	461,427	255,692	55.413%
2009	383,536	254,257	66.293%
2010	-	-	74.294%
2011	-	-	82.484%
2012	-	-	88.776%
2013	-	-	93.108%
2014	-	-	95.000%
2015	-	-	**95%**
2016	-	-	**95%**
2017	-	-	**95%**
2018	-	-	**95%**
2019	-	-	**95%**
2020	-	-	**95%**

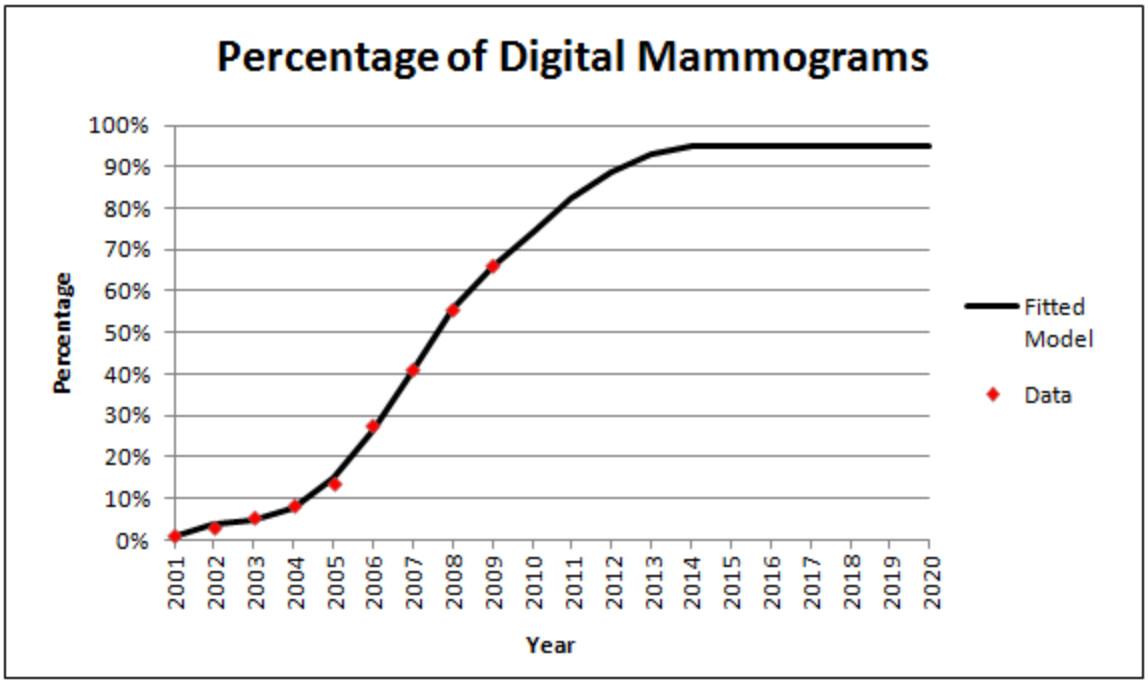


Figure 4-2. Percentage of Digital Mammograms for the period 2001–2020

The probability of a false positive screening exam is from BCSC data and is dependent of each woman's age and the time since her last mammogram. These values are given in Table 4-5. If a positive screening exam occurs (whether the result is a true positive or a false positive), then the woman is given a diagnostic exam as a second step in the detection process. The probability of a false negative screening exam is a function of the size of the tumor at the time of the screening exam, and the data was provided by the CISNET breast cancer modeling group [34]. Table 4-6 gives these values, and Figure 4-3 shows a graph of the probability of detection as a function of tumor size. If a screening exam is negative (whether the result is a true negative or a false negative), then the woman proceeds to the next year in the model; and the usual events for that year are executed at the appropriate times. Both false positives and false negatives for screening exams are affected by the level of breast cancer screening technology at the time of the exam. A detailed analysis of this effect is given in Section 4.2.

Table 4-5. Probability of False Positive Screening Exam by Age and Time Since Last Mammogram [14]

Screening Performance Data		
Age/Time Since Last Mammogram	**Specificity**	**Pr{False Positive}**
65-69		
9-15 months	92.9%	7.1%
16-20 months	92.5%	7.5%
21-27 months	92.0%	8.0%
28+ months	90.5%	9.5%
No previous mammography	83.5%	16.5%
70-74		
9-15 months	93.5%	6.5%
16-20 months	93.3%	6.7%
21-27 months	92.0%	8.0%
28+ months	91.3%	8.7%
No previous mammography	83.5%	16.5%
75+		
9-15 months	93.9%	6.1%
16-20 months	93.7%	6.3%
21-27 months	92.7%	7.3%
28+ months	92.3%	7.7%
No previous mammography	85.2%	14.8%

Table 4-6. Probability of Mammographic Detection as a Function of Tumor Size [34]

Mammogram Sensitivity			
Size (mm)		**Pr{Detection}**	
Min	**Max**	**Min**	**Max**
0	2	0	0.3
2	5	0.3	0.3
5	7.5	0.3	0.65
7.5	15	0.65	0.8
15	20	0.8	0.9
20	50	0.9	0.99
50	80	0.99	1.00
80	∞	1.00	1.00

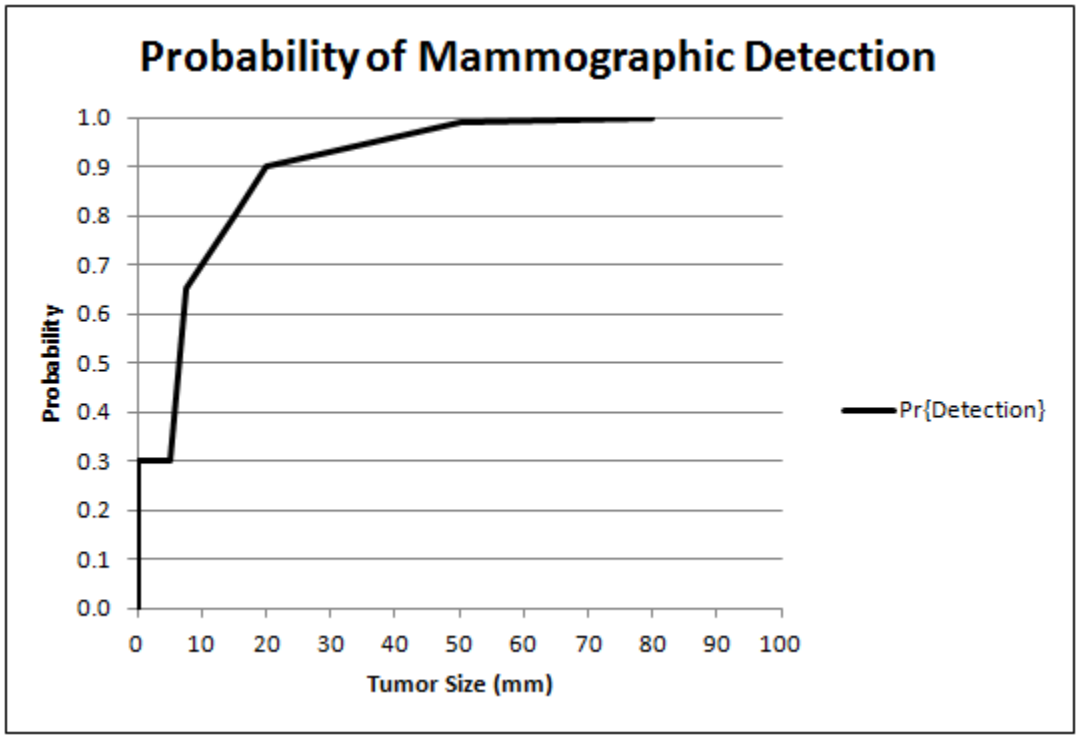


Figure 4-3. Graph of the Probability of Mammographic Detection as a Function of Tumor Size [34]

4.1.1.3 Diagnostic Exams

Diagnostic exams are given to women who have a positive screening exam. There are no false negatives for diagnostic exams: if a woman with cancer makes it to a diagnostic exam, then she will be sent for a work-up biopsy; and at that point where the presence of cancer will be confirmed. However, it is possible to have false-positive diagnostic exams, which will eventually lead to a benign work-up biopsy. The probability of a false positive diagnostic exam is based on BCSC data and is dependent of each woman's age and the time since her last mammogram. These values are given in Table 4-7. If a positive diagnostic exam occurs (whether the result is a true positive or a false positive), then the woman is given a work-up exam as a third and final step in the detection process. For costing purposes, three types of diagnostic exams are considered: mammography, ultrasound, and MRI. For a given patient, one, two, or all three of these diagnostic exams could be performed. Dr. Yankaskas of the Carolina Mammography Registry provided data on the percentages of

patients that receive one, two, and three exams along with what type of exams were performed for the years 2001–2009. The data are given in Table 4-8. We used the data from 2009 for the years 2010–2020.

Table 4-7. Probability of False Positive Diagnostic Exam by Age and Time Since Last Mammogram [14]

Diagnostic Performance Data		
Age/Time Since Last Mammogram	**Specificity**	$\Pr\{\text{False Positive}\}$
65–69		
9–15 months	34.8%	65.2%
16–20 months	34.9%	65.1%
21–27 months	51.6%	48.4%
28+ months	65.6%	34.4%
No previous mammography	72.3%	27.7%
70–74		
9–15 months	41.6%	58.4%
16–20 months	48.7%	51.3%
21–27 months	56.0%	44.0%
28+ months	66.4%	33.6%
No previous mammography	75.4%	24.6%
75+		
9–15 months	43.3%	56.7%
16–20 months	55.9%	44.1%
21–27 months	54.7%	45.3%
28+ months	69.4%	30.6%
No previous mammography	71.4%	28.6%

Table 4-8. Diagnostic Exams Distribution Percentages by Year for the Period 2000–2007

Year	Diagnostic Exams							
	None Only	Mamm. Only	Ultrasound Only	MRI Only	Mamm.+ Ultrasound	Mamm.+ MRI	Ultrasound+ MRI	All 3
2000	33.22	37.85	17.17	0.04	11.72	0.00	0.00	0.00
2001	30.27	38.98	17.53	0.00	13.23	0.00	0.00	0.00
2002	29.59	37.55	20.24	0.00	12.62	0.00	0.00	0.00
2003	30.04	37.73	16.93	0.04	15.26	0.00	0.00	0.00
2004	33.44	38.96	14.29	0.00	13.31	0.00	0.00	0.00
2005	32.76	37.64	13.50	0.00	16.09	0.00	0.00	0.00
2006	29.05	37.31	13.96	0.00	19.26	0.13	0.09	0.18
2007	19.35	42.96	14.42	0.43	21.84	0.19	0.34	0.48

4.1.1.4 Work-Up Procedures

Work-up procedures consist of open biopsies, core needle biopsies (CNBs), and (on a very small scale) fine needle aspirations (FNAs). Work-up exams are perfect – that is, if cancer is present, then the work-up exam will return a positive result; and if cancer is not present, then the work-up exam will return a negative result. We assume that only one work-up procedure is performed on each woman, and Table 4-9 shows the percentages of women who get the different types of work-up exams. Our breast cancer experts consulted with their radiologists and used their own experience and judgment to come up with the values in Table 4-9, and we are confident of their validity for our modeling purposes.

Table 4-9. Work-Up Procedure Distribution

Work Up Procedure	Percentage
FNA	**2%**
Core Needle Biopsy	**58%**
CNB (Ultrasound)	25%
CNB (Mammography)	25%
CNB (MRI)	25%
CNB (Palpation)	25%
Open Biopsy	**40%**
Open (Needle Loc)	50%
Open (Palpation)	50%

4.1.2 Treatment Submodel

Every woman who is diagnosed with breast cancer will have some probability of being treated that is a function of her age and health status at the time of diagnosis, and this probability is determined by the treatment submodel. Table 4-10 shows the treatment percentages by age and health status at diagnosis. Our breast cancer experts consulted with their radiologists and used their own experience and judgment to come up with the values in Table 4-10, and we are confident of their validity for our modeling purposes.

Table 4-10. Treatment Percentages by Age and Health Status

Age	Comorbidities Not Present	Comorbidities Present
65–69	99.5%	95%
70–79	99%	95%
80–89	98%	80%
90–99	95%	60%
100–110	90%	50%

Currently, treatment decisions are not the focus of the model, and we do not consider the exact type of treatment, just whether or not treatment occurs. In the future, it is possible to expand on the treatment submodel so that it incorporates a detailed representation of the type of treatment administered and the outcome. If it is determined that treatment will not be received, then the woman will have a breast cancer or non–breast cancer cause of death exactly as she did in the natural history model, where treatment was also ignored. If it is determined that treatment will be received, then the woman enters the survival submodel, where a new survival curve generates a (possibly) new cause of death and age at death.

4.1.3 Survival Submodel

Only women who receive treatment enter the survival submodel. This submodel alters the woman's survival after being diagnosed with breast cancer from what it was in the natural history model, where no treatment was given. SEER provides data for the time until breast cancer death by stage at diagnosis for women treated for breast cancer [66]. This only takes into account death from breast cancer or breast cancer complications, however the woman

could also have a death from other causes, and this probability is evaluated on a yearly basis based on the breast cancer–adjusted life tables. Thus each woman is assigned a death age from breast cancer and a death age from natural causes, and the minimum of these two values determines her actual death age and cause of death. SEER gives the probability that a woman would still be alive after *x* years, depending on whether she was diagnosed with cancer in either the local, regional, or distant stage. These probabilities can actually be used as the CDFs for the time until death given treatment as a function of stage at diagnosis, and we can use the method of inversion to sample from these CDFs as we have described earlier in this dissertation. Figure 4-4 shows graphs of the survival by stage. If treatment is given, there is a good possibility each woman will live longer than she did in the natural history model. Life years saved is defined as the number of years lived in the integrated screening model minus the number of years lived in the natural history model. When a woman receives treatment, it is likely that she will live longer than she would have without receiving treatment, especially if diagnosed in the early stages. The number of life-years lived beyond what would have occurred without screening and treatment is an important measure of the effectiveness of a screening policy. It is essentially the benefit in a cost-effectiveness analysis based on the cost-benefit ratio. However, each life-year saved is typically multiplied by a utility that represents the quality of the woman's life for that year relative to a woman who is completely healthy, and the result accumulated over all life-years saved is termed the woman's quality-adjusted life years saved.

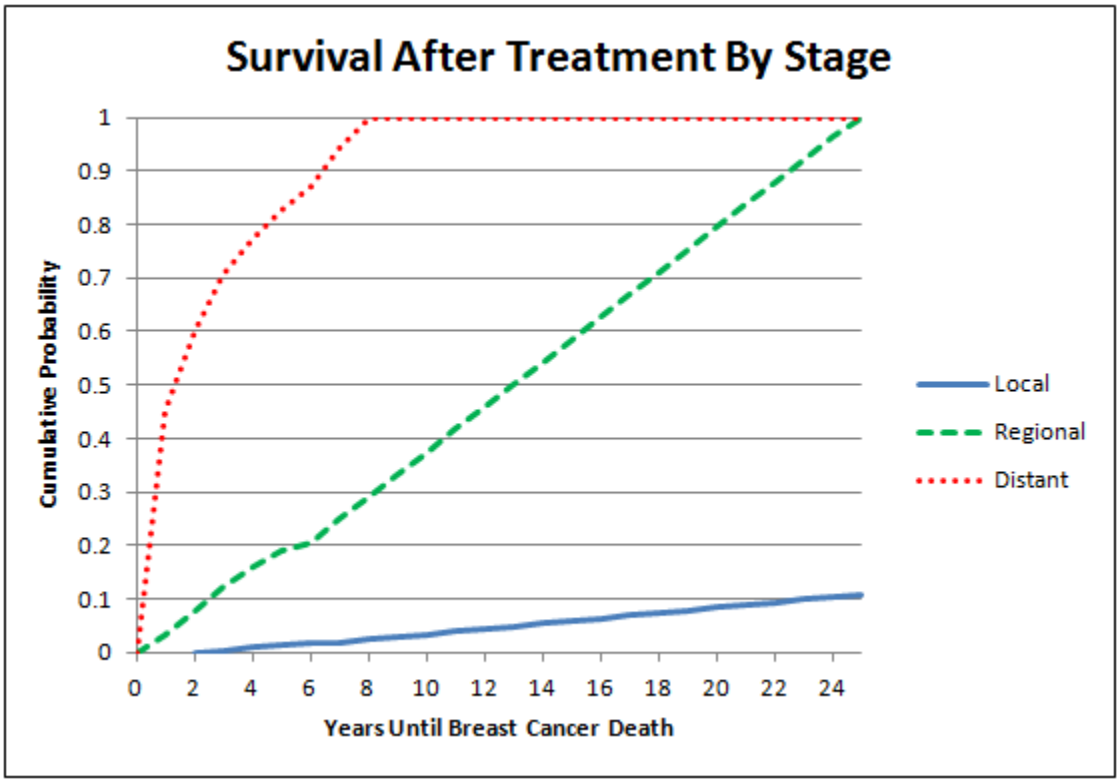


Figure 4-4. CDFs for Survival After Treatment by Stage at Diagnosis [66]

4.1.3.1 Quality Adjusted Life Years and Utility

As mentioned in the previous section, life-years saved are typically multiplied by a utility and converted to quality-adjusted life-years saved for the purposes of cost-benefit analysis. This requires utilities that represent the quality of a woman's life at different ages and preferably after being diagnosed at different stages of breast cancer. A woman who is diagnosed with distant cancer with have a lower quality of life than a healthy woman or a woman diagnosed with local cancer that can be easily treated before it spreads to other areas of the body. Along the same lines, a younger woman has a higher quality of life than women in general, because a younger woman has a longer life expectancy and is generally healthier than an older woman. The cost-effectiveness of breast cancer screening for middle-aged women had been addressed by many papers in the literature, and Tosteson et al. [98] provide utilities used to compute QALY for breast cancer screening in women of all ages. Table 4-11 gives these utilities as a function of age and stage at diagnosis. Each year a woman lives

beyond what she did in the natural history model, a life-year saved is recorded; and the corresponding number of quality-adjusted life-years saved is also computed and recorded.

Table 4-11. Health State Utilities Used to Compute QALYs [98]

Health State Utilities	Health State				
	Healthy	In-Situ	Localized	Regional	Distant
Age 60–69	0.81	0.73	0.73	0.61	0.49
Age 70–79	0.77	0.69	0.69	0.58	0.46
Age 80+	0.72	0.65	0.65	0.54	0.43

4.1.4 Costing Submodel

Because we ultimately intend to determine the most cost-effective screening policy or policies from the perspective of society, we use Medicare reimbursement rates data in order to determine the costs of exams and treatment. There are four types of costs tracked in the integrated screening model:

1. Cost of screening exams,
2. Cost of diagnostic exams,
3. Cost of work-up exams, and
4. Cost of treatment.

Each of these four types of costs is also broken down into the following categories:

1. Costs separately incurred during the periods 2001–2011 and 2012–2020; and
2. Costs incurred from clinically detected breast cancer and costs incurred from mammographic detections.

The total costs for the periods 2001–2011 and 2012–2020 are computed and then combined to yield the overall cost in a cost-effectiveness analysis. All costs are discounted forwards and backwards to 2012 dollars using the Equation (4.2) :

$$\text{DiscountedCost} = \text{Cost} * (1+i)^{2012\text{-CurrentYear}}. \quad (4.2)$$

We assume the interest rate or time value of money is 0.05 or 5% for these calculations.

We obtained costing data from two sources. The primary source was the Medicare reimbursement database [100], which gives the costs of many procedures according to geographic location. The costs from the different geographic locations were inputted into **Stat::Fit** distribution fitting software and appropriate fits were found for each data set, representing each type of procedure. the details of this distribution fitting including goodness-of-fit statistics are presented in Appendix E. Some procedures did not have data available, and in those cases we used the costs listed in [98], which is another paper on cost-effectiveness of breast cancer screening. Table 4-12 presents the distributions and constants used for the costs of screening, diagnostic, and additional work-up procedures. The Medicare reimbursement rate has been fairly constant for these exams over the last five years, so we did not inflate these costs as we did treatment costs.

Table 4-12. Exam Costing Information [98, 100]

Screening Costs ($)	**Distribution**	**Parameters**
Film	Lognormal	μ = 83.6, σ = 8.61, *Min (Offset)* = 71
Digital	Lognormal	μ = 146, σ = 18.8, *Min (Offset)* = 121
Diagnostic Imaging Costs ($)	**Distribution**	**Parameters**
Film	Lognormal	μ = 89, σ = 8.24, *Min (Offset)* = 75
Digital	Lognormal	μ = 137, σ = 14.6, *Min (Offset)* = 114
Ultrasound	Lognormal	μ = 99.6, σ = 11.4, *Min (Offset)* = 83
MRI	Constant	$787.13
Additional Work-Up Costs ($)	**Distribution**	**Parameters**
Fine Needle Aspiration	Pearson 6	β = 739.5, *p* = 3.03, *q* = 29.2, *Min (Offset)* = 457
Core Needle Biopsy (Ultrasound Guided)	Pearson 6	β = 4872.7, *p* = 3.04, *q* = 115.4, *Min (Offset)* = 706
Core Needle Biopsy (Mammography Guided)	Constant	$946.51
Core Needle Biopsy (MRI Guided)	Constant	$1044.54
Core Needle Biopsy (Palpation Guided)	Constant	$351.26
Open Biopsy (Needle Localization)	Constant	$2061.01
Open Biopsy (Palpation Guided)	Constant	$1699.06

Note: μ is the overall mean cost of each procedure, *Min* is the minimum value.

For our modeling purposes, it is convenient to define treatment costs by stage. Treatment costs by stage are also given in [98], a paper which uses costs from 2005. These costs by stage of treatment and stage at diagnosis are given in Table 4-13.

Table 4-13. Treatment Costing Information By Health State and Treatment Phase [98]

Treatment Costs in 2005 ($)	Health State				
	Healthy	In-Situ	Localized	Regional	Distant
Initial Treatment	-	$14,510	$18,470	$20,920	$0
Ongoing Treatment	-	$1,510	$1,630	$2,430	$4,980
Terminal treatment	-	$15,400	$20,530	$27,880	$25,560
Total	-	$31,420	$40,630	$51,230	$30,540

However, treatment costs and reimbursement rates can be expected to increase as new and improved treatments become available. In order to account for this increase in the costs of treatment, these costs are multiplied by the ratio between the medical CPI for the year in which treatment is given, and the medical CPI for 2005 (the year these costs were obtained),

$$\text{CPI(Treatment Year)}/\text{CPI(2005)}. \tag{4.3}$$

Equation (4.4) shows how treatment costs are computed,

$$\text{TreatmentCost(CurrentYear)} = \text{TreatmentCost(2005)} * \left[\frac{\text{CPI(CurrentYear)}}{\text{CPI(2005)}} \right]. \tag{4.4}$$

Medical CPI data was obtained from the U.S. Bureau of Labor Statistics online database [101]. Data is available as far back as 1935, and we fit a second-degree polynomial model (R^2 = 0.9913) to this data in order to project the medical CPI for the period 2012–2020. The data and projected values are given in Table 4-14, and a graph of the information provided in this table is given in Figure 4-5.

Table 4-14. Medical CPI Data for 2001–2011 [101] and Model Results for 2012–2020

Year	Medical CPI	Ratio (2005)
2001	272.8	0.8441
2002	285.6	0.8837
2003	297.1	0.9192
2004	310.1	0.9595
2005	323.2	1.0000
2006	336.2	1.0402
2007	351.054	1.0862
2008	364.065	1.1264
2009	375.613	1.1622
2010	388.436	1.2018
2011	399.2603	1.2353
2012	410.096	1.2689
2013	423.864	1.3115
2014	437.861	1.3548
2015	452.085	1.3988
2016	466.538	1.4435
2017	481.219	1.4889
2018	496.129	1.5351
2019	511.266	1.5819
2020	526.632	1.6294

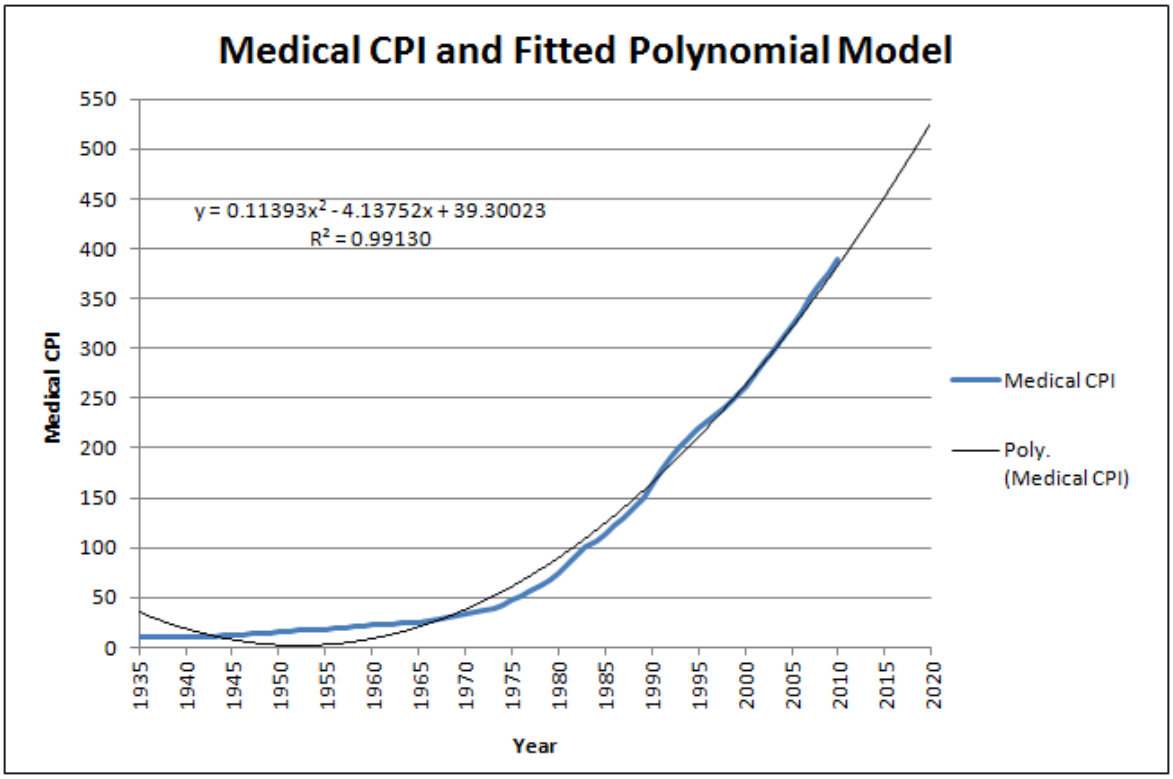


Figure 4-5. Graph of Medical CPI from Past Data [101] and Predicted Future Values

Having information about both cost and QALYs saved from both mammographic and clinical detections for the periods 2001–2011 and 2012–2020, we can now form the cost-effectiveness ratio cost per QALY saved for each of these time periods and types of detection. The most important of these is the cost per QALY saved from mammographic detections for the period 2012–2020. This completes the description of the SD submodels and their functions. The SD submodel for population-level elements of the screening process is the focus of the next sections.

4.2 SD Submodel

The purpose of the SD submodel is to represent population-level elements of the screening process, specifically those factors related to adherence to a prescribed screening policy. It may have been possible to capture these effects using discrete-event type logic, but this would likely have caused excessively long run times, and would have prevented us from

using optimization. Adherence to a screening policy is based on a number of factors, some at the population level, such as the level of congestion at screening facilities, and others at the individual level, such as the presence of other comorbid diseases for each woman. After developing the majority of the discrete-event logic, we formulated a combined DES/SD causal loop diagram, presented in Figure 4-6. The top half of the diagram contains characteristics of individual women, and these characteristics are modeled using DES logic; and the bottom half of the diagram contains characteristics of the population, and these population-level variables are modeled using SD logic. The SD logic and DES logic are related through a logistic regression model for adherence to breast cancer screening, and through hybrid SD levels that have a direct effect on DES logic. In the causal loop diagram, if factor A effects factor B, then there is an arrow originating at factor A and terminating at factor B. There is also a direction of influence, positive or negative, associated with each arrow. A positive influence means that as if the level of factor A increases, then the level of factor B increases as well. Negative influence means that if the level of factor A increases, then the level of factor B decreases.

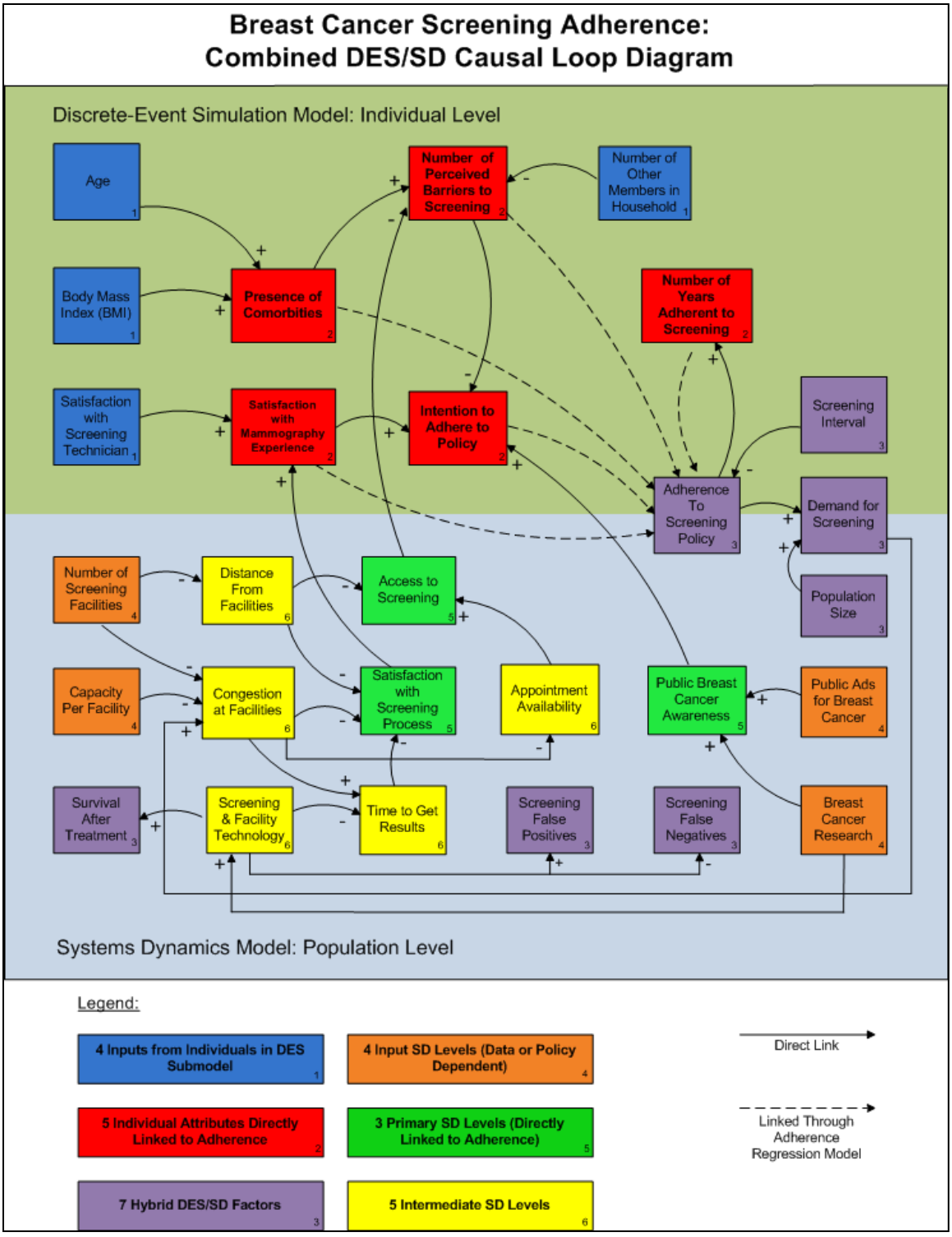


Figure 4-6. Combined DES/SD Causal Loop Diagram

In order to model adherence on an individual level, it was necessary locate a model that predicts adherence to a screening policy based on several individual factors. Gierisch [37] presents a logistic regression model that represents the probability of attending each annual screening appointment as a function of a number of individual factors. We added one factor to this regression model, screening interval; and based on discussions with our experts, we assumed that as the screening interval decreases (more frequent screening), the adherence to screening would increase. This was accomplished by simply adding a screening interval factor to the regression model and making some assumptions about the odds ratios for values greater than one (one is the default value for the model).

After finalizing the regression model, we formulated an SD submodel that represents elements of the screening process at the population level. First we determined the levels that effect the indivdual attributes directly linked to regression, and then we identified other intermediate levels which could potentially affect those attributes. Some levels are user inputs, and intermediate levels are computed based on the values of user input levels and other levels. "Hybrid levels" are defined at the population level but are directly linked to individuals. For example, the probability of a false positive is defined for the entire population, but it is also based on individual characteristics, and it can be affected by the level of screening technology in the SD submodel. These hybrid levels are described in detail in the next section focused on the linkage between the DES and SD submodels. While formulating this model, we realized that we would need to account for several individual attributes which were not included in the original DES submodel, such as the presence of comorbid diseases. These additional attributes and their definitions are given in Section 4.2.5.

4.2.1 State-Variable and Individual Attribute Declarations

Several SD state-variables and some new individual characteristics are incorporated into the combined DES/SD model. The following section lists these variables and individual characteristics, defines the units for each (if necessary), and defines the ranges of values for

each. It is important to understand these variables and their definitions so one can understand the mathematical development of the combined DES/SD model.

4.2.1.1 SD Input Levels (4)

$NumOfFacilities(t) \in [5000, 15000]$, expressed as the total number of screening facilities nationwide;

$CapacityOfFacilities(t) \in [12250, 20000]$, the service rate of all the facilities expressed as the total number of mammograms per year;

$PublicAds(t) \in \{1, 2, 3\}$, the overall level of public advertising; and

$BCResearch(t) \in \{1, 2, 3\}$, the overall level of research on breast cancer.

4.2.1.2 Other SD Levels (5)

$DistanceToFacilities(t) \in [5, 25]$, the patient's travel distance from home to a screening facility, in miles;

$CongestionAtFacilities(t) \in [0, 100]$, the patient's expected waiting time before being screened, in minutes;

$ScreeningTechnology(t) \in [0, 1]$, the current level of screening technology, where 0="completely ineffective" and 1="perfect";

$TimeToGetResults(t) \in [5, 35]$, the delay between completion of screening and notifying the patient of the results, expressed in days; and

$AppoinmentAvailability(t) \in [0, 1]$, the probability that the patient can make an appointment for screening.

4.2.1.3 SD Levels Linked to Individual Attributes Affecting Adherence (3)

$Access(t) \in [0,1]$, the degree to which patients get an appointment and their degree of difficulty in getting there, where 0="almost no access" and 1="excellent access";

$ScreenProcSat(t) \in [0,1]$, the patient's degree of satisfaction with the screening process, where 0="completely dissatisfied" and 1="completely satisfied"; and

$PublicAwareness(t) \in [0,1]$, level of public awareness of the importance of breast cancer screening.

4.2.1.4 Input Characteristics from Individuals from DES Submodel (4)

$Age(t) \in [65,110]$, expressed in years;

$InitialAge \in [65,95]$, expressed in years;

$BMI \in \{1,2,3,4\}$, expressed as follows: category=1 if patient's BMI $\leq$ 25; category=2 if patient's BMI $\in$ (25,30]; category=3 if patient's BMI $\in$ (30,35]; and category=4 otherwise;

$HouseSize \in \{0,1,2\}$, expressed as number of other members of the patient's household; and

$ScreenTechSat(t) \in \{0,1\}$, level of satisfaction with screening technician,where 0="completely unsatisfactory" and 1="completely satisfactory."

4.2.1.5 Individual Characteristics Directly Linked To Adherence (5)

$Comorbidity(t) \in \{0,1\}$, indicator of comorbid conditions where 0="no comorbid conditions," and 1="comorbid condition(s) present";

$MammSat(t) \in \{0,1\}$, indicator of satisfaction with previous mammograms, where 0="dissatisfied" and 1="satisfied";

$NumOfBarriers(t) \in \{0,1,2\}$, number of perceived barriers to screening;

$IntentToAdhere(t) \in \{0,1\}$, indicator of patient's intent to adhere to screening, where 0="do not intend to adhere" and 1="do intend to adhere"; and

$Last2Adherence(t) \in \{0,1,2\}$, number of times patient went to screening out of last two opportunities.

4.2.1.6 Personal Penalty Factors (5)

The following random variables represent variation among individual women in the weights they place on various factors in determining, for example, their overall satisfaction with previous mammographic experiences, etc, as detailed in Section 4.2.2 below.

$ScreenTechPen \sim UNIFORM\,(0.25,1)$
$ScreenProcPen \sim UNIFORM\,(0.25,1)$
$ComorbidPen \sim UNIFORM\,(0.25,1)$
$HouseSizePen \sim UNIFORM\,(0.25,1)$
$AccessPen \sim UNIFORM\,(0.25,1)$

4.2.1.7 Other Factors

The following auxiliary factors are explained in detail in Section 4.2.2 below.

$NumOfFacilitiesTrend(t) \in \{-2,-1,\ 0,\ 1,\ 2\}$;

$CapacityOfFacilitiesTrend(t) \in \{-2,-1,\ 0,\ 1,\ 2\}$;

NumberOfMammograms(t), the number of women who got a mammogram in year *t*;

DemandforScreening(t), the total demand for mammograms in year *t*;

DemandPerFacility(t), the average hourly demand per facility for mammograms in year *t*;

$ArrivalRatePerFacility(t) = \lambda(t)$, the average number of women needing a mammogram who arrive per hour at each facility during business hours in year *t*;

ServiceRatePerFacility(t), the average number of mammograms that can potentially be completed per hour at each facility during business hours in year *t*;

$MeanServiceTimePerFacility(t) = \mathrm{E}[S]$, the expected value of the duration (in hours) of each mammogram performed in year *t*;

$VarianceServiceTime(t) = \mathrm{Var}[S]$, the variance of the duration of each mammogram (expressed in hr^2) performed in year *t*;

$FalsePosPercent(t) \in [0,1]$, the fraction of mammograms given in year t that can be expected to yield a false positive;

$FalsePosPercentAdj(t) \in [0,1]$, *FalsePosPercent(t)* adjusted for changes in screening technology in year *t*;

$FalseNegPercent(t) \in [0,1]$, the fraction of mammograms performed in year t that can be expected to yield a false negative;

$FalseNegPercentAdj(t) \in [0,1]$, *FalseNegPercent(t)* adjusted for changes in screening technologies in year *t*;

$TreatmentSurvivalAdj(t) \in [0,1.8]$, the number of years until breast cancer death for women who are treated, adjusted for the level of technology in year *t*;

$StageFactor \in \{1,2,3\}$, represents the stage of breast cancer at diagnosis and is used to determine *TreatmentSurvivalAdj(t)* for a woman in a given stage;

4.2.2 SD Input Levels

Four of the SD levels are user inputs. They are levels that the users can control through the user interface, and they can be controlled to an even finer level via model logic. The number of screening facilities is set for 2001–2012, but the user can control the trend in the number of facilities (increasing, slightly increasing, constant, slightly decreasing, or decreasing) over the period 2012–2020. Similarly the user can control future trends in the capacity of these screening facilities. The level of advertising for breast cancer screening (low, medium, high) can be assigned by the user for both the past and the future. We select the default level (medium) and use it for both the past and the future in most experiments, but in further experimentation the user can explore how changes in these values impact the results. The level of breast cancer research is analogous to the level of advertising in that it has the same three levels, the same default values, and allows the user to input levels for both the past and future.

4.2.2.1 Number of Screening Facilities

The number of breast cancer screening facilities present in the US, *NumOfFacilities(t)*, directly impacts the average distance each woman must travel to get

screened as well as the congestion at each facility. As the number of facilities decreases, the distance to each facility will increase and the congestion will increase (unless the capacity of each facility increases). Breast cancer screening facilities must register with the FDA, and we were able to obtain data on the approximate number of facilities in 2001 from the BCSC and in 2012 by querying the FDA website. There were approximately 10,125 registered facilities in 2001 and there were approximately 8,444 registered at the start of 2012. However, the number of women in the US population is increasing, the average age is increasing, and public awareness regarding the importance of mammography has increased since 2001. All these facts lead to the conclusion that the demand has likely been increasing, while the number of facilities has been decreasing. Our breast cancer experts have suggested that this phenomenon has been caused by smaller facilities shutting down and joining other facilities, and the average capacity of registered facilities has increased, thus allowing the country's screening capacity to keep up with demand.

We assumed the number of facilities deceased linearly from 2001 to 2012, and the user does not have control over the number of facilities during this time period. However, beginning in 2013, the user is allowed choose one of the following five levels to govern how the number of facilities will change from 2013 to 2020: decreasing, slightly decreasing, constant (default), slightly increasing, and increasing. The state-variable $NumOfFacilitiesTrend(t)$ is defined according to Equation (4.5):

$$NumOfFacilitiesTrend(t) \equiv \begin{cases} -2, & \text{if decreasing,} \\ -1, & \text{if slightly decreasing,} \\ 0, & \text{if constant (default),} \\ 1, & \text{if slightly increasing,} \\ 2, & \text{if increasing.} \end{cases} \tag{4.5}$$

For the period 2013 to 2020, the number of facilities are time can be computed as a function of the number of facilities at time $t-1$ and the trend in the number of facilities at time $t-1$ according to Equation (4.6):

$$NumOfFacilities(t) = NumOfFacilities(t-1) + 50 * NumOfFacilitiesTrend(t-1). \quad (4.6)$$

Based on the data, it would be reasonable to believe either -1 or 0 will reflect the actual future trend, but it may be interesting to explore how other options affect the results. The default value of $NumOfFacilitiesTrend(t)$ is 0, that is, the number of facilities will remain constant during the period 2013–2020. However, the default value for the state-variable $CapacityOfFacilitiesTrend(t)$ is 1, that is, the capacity of screening facilities is slightly increasing during this period. It is known that demand for screening will increase, so we naturally assumed that the overall capacity for screening would at least slightly increase during the period 2013–2020.

4.2.2.2 Average Capacity of Screening Facilities

The average capacity of screening facilities, $CapacityOfFacilities(t)$, is defined in terms of the average number of mammograms per year that facilities can provide; and this can easily be translated into an hourly arrival rate with some basic assumptions. Demand has likely increased with the size and age of the US female population, and the number of facilities has decreased. Thus, the average capacity of screening facilities must be at least slightly increasing in order to maintain a reasonable level of service.

We assumed the average capacity of screening facilities increased from 2001–2012, and the user does not have control over the number of facilities during this time period. However beginning in 2013, the user is allowed choose one of the following five levels to govern how the average capacity of facilities will change from 2013–2020: decreasing, slightly decreasing, constant (default), slightly increasing, and increasing. The state-variable $CapacityOfFacilitiesTrend(t)$ is defined according to Equation (4.7):

$$CapacityOfFacilitiesTrend(t) \equiv \begin{cases} -2, & \text{if decreasing,} \\ -1, & \text{if slightly decreasing,} \\ 0, & \text{if constant,} \\ 1, & \text{if slightly increasing, (default)} \\ 2, & \text{if increasing.} \end{cases} \tag{4.7}$$

For the period 2013–2020, the capacity of facilities are time t can be computed as a function of the capacity of facilities at time $t-1$ and the trend in the capacity of facilities at time $t-1$ according to Equation (4.8),

$$\begin{aligned} CapacityOfFacilities(t) = {} & CapacityOfFacilities(t-1) \\ & + 0.1 * CapacityOfFacilitiesTrend(t-1). \end{aligned} \tag{4.8}$$

Based on the assumed past trend, it would be reasonable to believe that a value of either 1 or 2 in Equation (4.7) will reflect the actual future trend, but it may be interesting to explore how other options affect the results. The default value of $CapacityOfFacilitiesTrend(t)$ is 1, that is, the capacity of facilities will slightly increase during the period 2013–2020.

4.2.2.3 Level of Public Advertising for Breast Cancer Screening:

Public advertisements for breast cancer screening influences the level of breast cancer awareness in the population. The level of public advertising for breast cancer can be set by the user and can be constant or have some criterion for changing. We define the level of public advertising, $PublicAds(t)$, as either low, medium, or high according to Equation (4.9):

$$PublicAds(t) \equiv \begin{cases} 1, & \text{if low,} \\ 2, & \text{if medium (default),} \\ 3, & \text{if high.} \end{cases} \tag{4.9}$$

Currently, the level of public advertising is set to medium for both the past time periods and the future. However, the user does have control over the values for both of these periods. So the impact of increasing or decreasing the level of public advertising can be observed for the given time period by changing the input values for this level. It is difficult to quantify the

level of public advertising. One might think of using the number of local TV/radio ads or the total amount of money raised through charitable organizations, but it is difficult to find data on these quantities. Thus, we chose to go with a more subjective low-medium-high scale for this input.

4.2.2.4 Level of Breast Cancer Research

Breast cancer research also influences the level of breast cancer awareness in the population. Like the level of public advertising, the level of breast cancer research can be set by the user and can be constant or have some criterion for changing. We define the level of breast cancer research, *BCResearch(t)*, as either low, medium, or high according to Equation (4.10):

$$BCResearch(t) \equiv \begin{cases} 1, \text{ if low,} \\ 2, \text{ if medium (default),} \\ 3, \text{ if high.} \end{cases} \tag{4.10}$$

Currently, the level of breast cancer research is set to medium for both the past time periods and the future. However, the user does have control over the values for both of these periods. So the impact of increasing or decreasing the level of breast cancer research can be observed for the given time period by changing the input values for this level. It is difficult to quantify the level of breast cancer research. One might think of using the number of papers published or estimating the total research expenditures on breast cancer research, but it is difficult to find data on these quantities, so again we chose to go with a more subjective low-medium-high scale for this input.

4.2.3 Intermediate SD Levels

SD Levels which are not inputs and are not directly linked to adherence are considered intermediate SD levels. Intermediate levels are functions of input levels and in some cases, other intermediate levels. They serve as inputs to either hybrid levels that alter DES logic, or to primary SD levels directly affecting adherence. The average distance to a

screening facility is a function of the number of facilities present in the United States. The average level of congestion at screening facilities is a function of the demand for screening, the number of facilities, and the average capacity of those facilities. We use the somewhat crude measure of expected waiting time from queuing theory with several assumptions in place. While this may not reflect actual waiting times, it does provide a measure of congestion at screening facilities and thus serves our purpose well. The level of breast cancer screening technology is a simple zero-one level, which is a function of the level of breast cancer research; and it directly affects the hybrid levels for false positives and false negatives in the DES logic through adjustment factors. The average time to get results is a function of the level of technology and the level of congestion, and we use the number of days to get results as a unit of measure. Finally, appointment availability is a function of the congestion at screening facilities, and this is a simple level which varies between zero and one.

4.2.3.1 Average Distance to a Screening Facility

Dr. Michael Kay from the Industrial and Systems Engineering Department at NC State University specializes in logistics engineering; and using a facility location program [16] he developed, we determine the average distance to a screening facility as a function of the number of facilities present in the United States. He ran the code using between 5,000 and 15,000 facilities in increments of 1,000 to account for all extremes. We then fitted a function to these data points enabling us to estimate rapidly the average distance to a facility as a function of the total number of facilities. Figure 4-7 shows the data points, fitted function, and R^2 goodness-of-fit value. We use miles as the unit of measure for distance. Equation (4.11) shows how $DistanceToFacilities(t)$ is computed as a function of $NumOfFacilities(t)$:

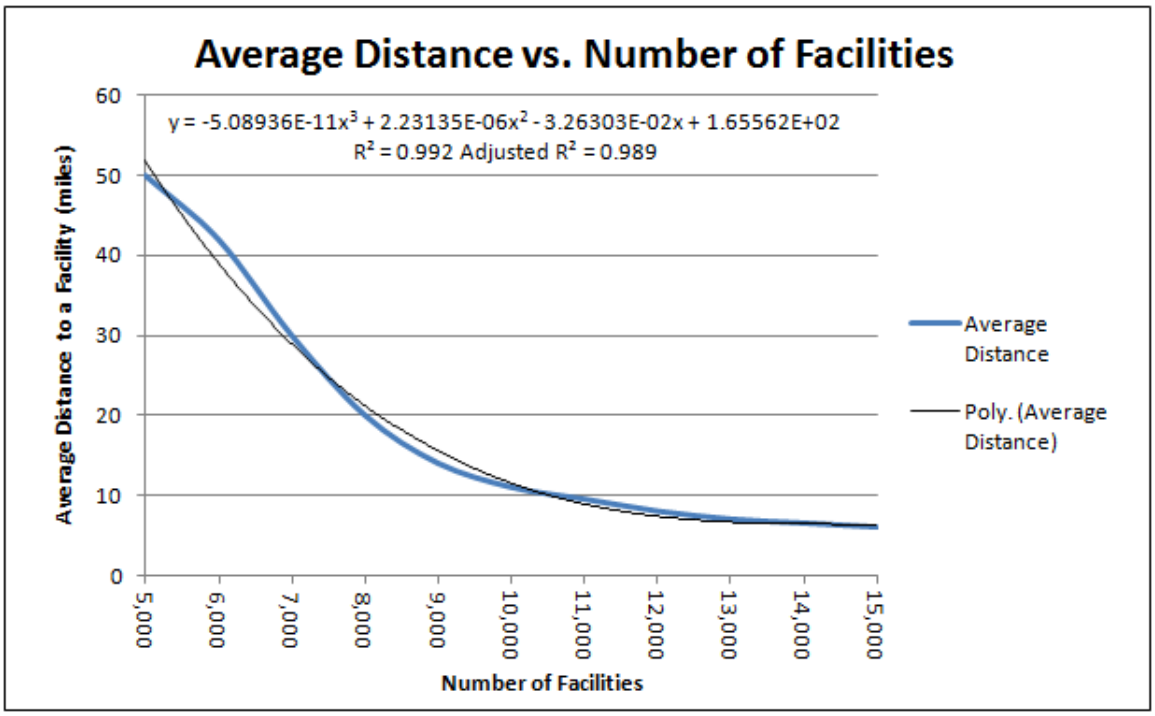


Figure 4-7. Distance to Facility as a Function of the Number of Facilities

$$DistanceToFacilities(t) = f\left(NumOfFacilities(t)^{-}\right),$$

$$\begin{aligned} DistanceToFacilities(t) = & -5.089*10^{-11}*\left[NumOfFacilities(t)\right]^{3} \\ & +2.231*10^{-6}*\left[NumOfFacilities(t)\right]^{2} \\ & -0.035*NumOfFacilities(t)+165.56. \end{aligned} \tag{4.11}$$

4.2.3.2 Level of Congestion at Screening Facilities

The level of congestion at screening facilities, *CongestionAtFacilities(t)*, is intended to reflect how crowded screening facilities may become with the aging female population. A measure of congestion often used in simulation models is waiting time. However, the logistics of having each individual woman queue up at a screening facility would cause the model to have an extremely long run time, making optimization routines difficult to run. Thus, we have chosen to model these on an aggregate population level. Each year, we know the number of screening facilities, *NumOfFacilities(t)*, the average service rate of these facilities, *ServiceRatePerFacility(t)*, and the number of women who got a mammogram in the model the previous year, $NumberOfMammograms(t-1)$. We can take the number of women who get a mammogram in the model each year, multiply by one thousand to scale up from 0.1% to the full US 65 and older female population size, and then scale up to all US women in the US population by multiplying by the inverse of the proportion of the female population that is 65 and older (derived from US population data). Thus, we have an estimate of the total demand for mammography in the US, *DemandforScreening(t)* (for the previous year as it would be impossible to know the demand for the current year as it occurs). Figure 4-8 shows a linear model fitted to the percentage of the US female population that is 65 and older, and Table 4-15 shows how we computed the scale factor, *ScaleFactor(t)*, for each year.

We make the assumption that the demand is equally distributed amongst the facilities and that each facility operates for 250 days per year and 8 hours per day, which allows us to compute $\lambda(t)$, the average hourly arrival rate for each facility, during business hours. We were able to run the model in a preliminary capacity to generate estimates of the demand for screening as a function of the screening interval; and using this data, we made some assumptions about the demand for screening in the year 2000, $NumberOfMammograms(0)$, so that congestion can be computed for the year 2001. Using these assumptions, a

mathematical formulation is given below. The units or measure for waiting time are minutes. The Pollaczek–Khinchine [48] formula is used to compute the long-run average waiting time, $\mathrm{E}[WQ_\infty]$, and we have assumed that each service time S~Uniform$(0.9\mu_s, 1.1\mu_s)$ so that the service time variance, $\mathrm{Var}[S]$, is given by $(1.1\mu_s - 0.9\mu_s)^2/12 = 0.00333\mu_s^2$. It is possible for the arrival rate to exceed the service rate, and that would cause the Pollaczek–Khinchine formula to produce incorrect negative values. To adjust for this possibility, we limit the maximum value of congestion to 100 minutes.

$$CongestionAtFacilities(t) = f\left(NumOfFacilities(t)^{-}, CapacityOfFacilities(t)^{-}, DemandforScreening(t)^{+}\right),$$

$$DemandforScreening(t) = NumberOfMammograms(t-1) * 1{,}000 * ScaleFactor(t), \qquad (4.12)$$

$$NumberOfMammograms(0) = 9600,$$

$$DemandPerFacility(t) = \frac{DemandforScreening(t)}{NumOfFacilities(t) * 2000}, \qquad (4.13)$$

$$\lambda(t) \equiv ArrivalRatePerFacility(t) = DemandPerFacility(t), \qquad (4.14)$$

$$ServiceRatePerFacility(t) = \frac{CapacityOfFacilities(t)}{2{,}000}, \qquad (4.15)$$

$$MeanServiceTimePerFacility(t) = \frac{1}{ServiceRatePerFacility(t)} = \mathrm{E}[S] = \mu_S \text{ hours}, \qquad (4.16)$$

$$\mathrm{Var}[S] = \sigma_S^2 = \frac{(0.2 * \mu_S)^2}{12} = 0.0033 * \mu_S{}^2, \qquad (4.17)$$

$$CongestionAtFacilities(t) \equiv \mathrm{E}[WQ_\infty] = \lambda(t) * \left\{\frac{\sigma_S^2 + \mu_S{}^2}{2 * [1 - \lambda(t) * \mu_S]}\right\} * 60 \text{ minutes}, \qquad (4.18)$$

Table 4-15. Demand Scale Factor Used to Compute Total Demand

Year	Total Female	Female 65+	Female 65+ % of Total Female	Multiplication Ratio*1000
2000	143,713,409	20,497,892	0.1426	7011.13
2001	145,083,005	20,590,658	0.1419	7046.06
2002	146,390,096	20,686,951	0.1413	7076.45
2003	147,649,491	20,838,602	0.1411	7085.38
2004	148,908,065	20,981,785	0.1409	7097.02
2005	150,192,384	21,180,464	0.1410	7091.08
2006	151,532,510	21,418,004	0.1413	7075.01
2007	152,967,793	21,736,924	0.1421	7037.23
2008	154,300,620	22,195,107	0.1438	6952.01
2009	155,557,060	22,569,111	0.1451	6892.48
2010	156,521,091	22,905,024	0.1463	6833.48
2011	157,509,714	23,409,500	0.1468*	6812.21
2012		23,913,500	0.1478*	6767.86
2013		24,456,200	0.1487*	6724.10
2014		25,070,900	0.1497*	6680.89
2015		25,744,300	0.1506*	6638.24
2016		26,446,100	0.1516*	6596.12
2017		27,237,700	0.1526*	6554.54
2018		28,102,300	0.1535*	6513.48
2019		29,033,600	0.1545*	6472.93
2020		30,090,700	0.1555*	6432.88

*Values were estimated from a the linear model in Figure 4-8

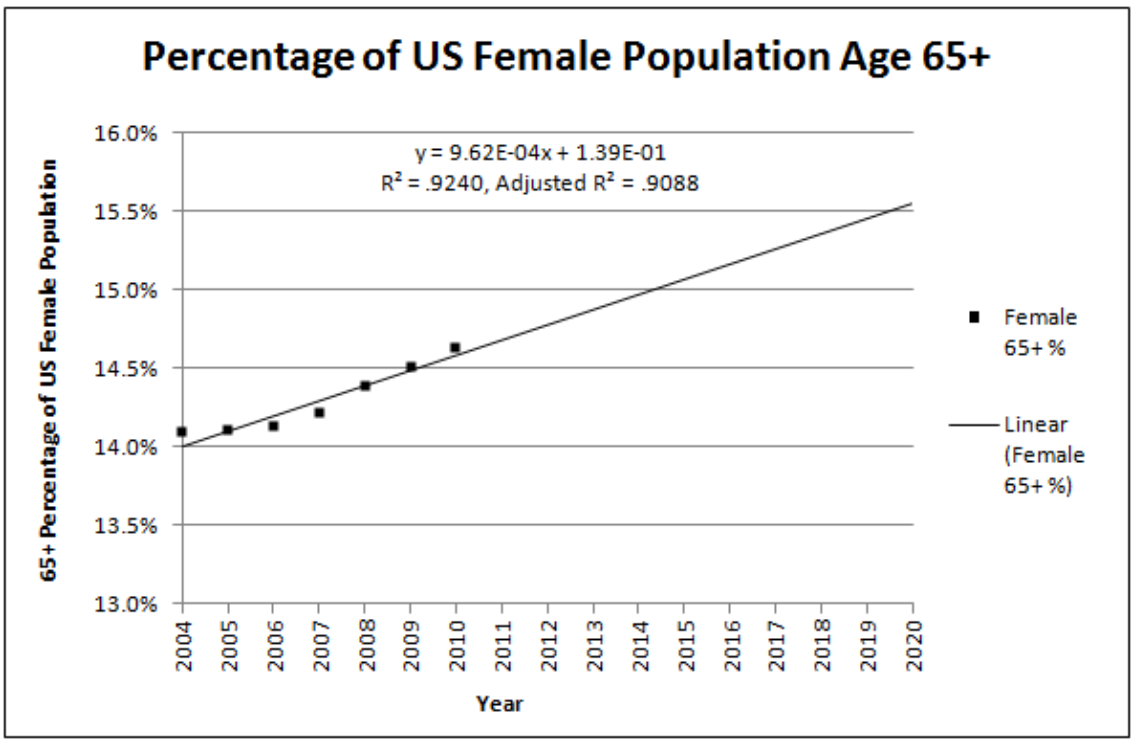


Figure 4-8. Linear Model Percentage of US Female Population of Age 65+

4.2.3.3 Level of Breast Cancer Screening Technology

The level of screening technology, *ScreeningTechnology(t)*, can take on values between 0 and 1, with 0 representing extremely poor technology, 0.4 representing technology in 2001, and 1 representing the highest possible level of technology that is achievable by the year 2020.

$$ScreeningTechnology(t) = f\left(BCResearch(t)^{+}\right),$$

$$ScreeningTechnology(0) = 0.40,$$

$$ScreeningTechnology(t) = ScreeningTechnology(t-1) + \frac{BCResearch(t-1)}{100}. \qquad (4.19)$$

The level of screening technology is assumed to have an impact on the percentage of screening and diagnostic exams with false positive results and on the percentage of screening exams with false negative results. We assumed that increases in technology would cause

false positives to increase (as seen over the last 10 years) but that it would cause false negatives to decrease. We also assumed a minimum adjustment of 0% and a maximum adjustment of 5% for both false positives and false negatives. Please refer to Table 4-5, Table 4-6, and Table 4-7 for the base values of *FalsePosPercent* and *FalseNegPercent* for both screening and diagnostic exams. We now introduce two adjustment factors, *FalsePosPercentAdj(t)* and *FalseNegPercentAdj(t)*, that will be used in the DES submodel and define them mathematically:

$$FalsePosPercentAdj(t) \equiv \begin{cases} 0, & \text{if } 0 \le ScreeningTechnology(t) \le 0.4, \\ 5/6, & \text{if } 0.4 < ScreeningTechnology(t) \le 0.5, \\ 10/6, & \text{if } 0.5 < ScreeningTechnology(t) \le 0.6, \\ 15/6, & \text{if } 0.6 < ScreeningTechnology(t) \le 0.7, \\ 20/6, & \text{if } 0.7 < ScreeningTechnology(t) \le 0.8, \\ 25/6, & \text{if } 0.8 < ScreeningTechnology(t) \le 0.9, \\ 30/6, & \text{if } 0.9 < ScreeningTechnology(t) \le 1.0, \end{cases} \tag{4.20}$$

$$FalseNegPercentAdj(t) \equiv \begin{cases} 0, & \text{if } 0 \le ScreeningTechnology(t) \le 0.4, \\ -5/6, & \text{if } 0.4 < ScreeningTechnology(t) \le 0.5, \\ -10/6, & \text{if } 0.5 < ScreeningTechnology(t) \le 0.6, \\ -15/6, & \text{if } 0.6 < ScreeningTechnology(t) \le 0.7, \\ -20/6, & \text{if } 0.7 < ScreeningTechnology(t) \le 0.8, \\ -25/6, & \text{if } 0.8 < ScreeningTechnology(t) \le 0.9, \\ -30/6, & \text{if } 0.9 < ScreeningTechnology(t) \le 1.0, \end{cases} \tag{4.21}$$

$$FalsePosPercent(t) = FalsePosPercent + FalsePosPercentAdj(t), \tag{4.22}$$

$$FalseNegPercent(t) = FalseNegPercent + FalseNegPercentAdj(t). \tag{4.23}$$

The level of technology is also assumed to affect survival after treatment. As technology improves, survival after treatment will improve. Thus, we introduce an

adjustment factor, *TreatmentSurvivalAdj(t)*, that describes the increase in survival (years) as function of the level of technology. We define another adjustment factor, *StageFactor*, which acts as a multiplier for *TreatmentSurvivalAdj(t)* based on the stage of cancer at diagnosis. There will be less improvement in survival for distant cancer than there would be for localized cancer, so *StageFactor* is 3 if cancer is in the local stage, 2 if in the regional stage, and 1 if in the distant stage, in order to make it dependent of the stage of cancer when treatment is received:

$$StageFactor \equiv \begin{cases} 1, & \text{if cancer stage at diagnosis is distant,} \\ 2, & \text{if cancer stage at diagnosis is regional,} \\ 3, & \text{if cancer stage at diagnosis is local,} \end{cases} \tag{4.24}$$

$$TreatmentSurvivalAdj(t) = \left(ScreeningTechnology(t) - 0.4\right) * StageFactor. \tag{4.25}$$

4.2.3.4 Average Time to Get Results

The time to get the results from a screening mammogram can affect each woman's satisfaction with the screening process. We assume the time to get results decreases as technology improves, and increases as facilities become more congested. The average time to get results is measured in days, and is expressed in the following way:

$$TimeToGetResults(t) = f\left(ScreeningTechnology(t)^{-}, CongestionAtFacilities(t)^{+}\right),$$

$$\begin{aligned} TimeToGetResults(t) = 5 + \left(\frac{CongestionAtFacilities(t)}{1440} \times 14.4\right) * 15 \\ + \left(1 - ScreeningTechnology(t)\right) * 15. \end{aligned} \tag{4.26}$$

4.2.3.5 Appointment Availability

Appointment availability can affect the level of access to screening for each woman in the population. We assume that appointment availability decreases as congestion at

screening facilities increases. Appointment availability is measured on a zero-one scale, with 1 being the best and zero being the worst, and is expressed mathematically as follows:

$$AppointmentAvailability(t) = f\left(CongestionAtFacilities(t)^{-}\right),$$

$$AppointmentAvailability(t) = 1 - \left(\frac{CongestionAtFacilities(t)}{100}\right). \tag{4.27}$$

4.2.4 Primary SD Levels (Directly Linked to Adherence)

Primary SD levels are those that are linked to attributes used in the adherence regression equation. Inputs to primary levels can be input levels or intermediate levels; and as the name indicates, primary levels are inputs to the attribute levels affecting adherence through the logistic regression equation. The level of access to screening, the level of satisfaction with the breast cancer screening process, and public breast cancer awareness are the three primary SD levels. Each of these levels is a zero-one level, with one being the best and zero being the worst. The variable *Access(t)* is a function of distance to facilities and appointment availability. Satisfaction with the screening process, *ScreenProcSat(t)*, is a function of the distance to facilities, congestion at the facilities, and the time it takes to get results. Public awareness, *PublicAwareness(t)*, is a function of breast cancer research and public advertising for breast cancer screening.

4.2.4.1 Level of Access to Screening

The level of access to screening in the population is assumed to increase as the average distance to facilities decreases, and is assumed to increase as appointment availability increases. The level of access to screening is measured on a zero-one scale, with 1 being the best and zero being the worst, and is expressed mathematically as follows:

$$Access(t) = f\left(DistanceToFacilities(t)^{-}, AppoinmentAvailability(t)^{+}\right),$$

$$Access(t) = \left(\frac{1}{2} - \frac{DistanceToFacilities(t) - 5}{40} \right) + \frac{AppoinmentAvailability(t)}{2}. \quad (4.28)$$

4.2.4.2 Level of Satisfaction with the Screening Process

The level of satisfaction with the screening process is assumed to decrease as the average distance to facilities, congestion at facilities, and average time to get results increase. The level of satisfaction with the screening process is measured on a zero-one scale, with 1 being the best and zero being the worst, and is expressed mathematically according to Equation (4.29). Note that $DistanceToFacilities(t)$ and $TimeToGetResults(t)$ have maximum values of 25 and 35 respectively, and that both have a minimum value of 5. $CongestionAtFacilities(t)$ has a minimum of zero and a maximum of one.

$$ScreenProcSat(t) = f(DistanceToFacilities(t)^{-}, CongestionAtFacilities(t)^{-}, TimeToGetResults(t)^{-}),$$

$$ScreenProcSat(t) = 1 - \frac{(CongestionAtFacilities(t)/100)}{3} - \frac{DistanceToFacilities(t) - 5}{60} - \frac{TimeToGetResults(t) - 5}{90}. \quad (4.29)$$

4.2.4.3 Level of Public Breast Cancer Awareness

The level of breast cancer public awareness in the population is assumed to increase with both the level of public advertising for breast cancer and the level of breast cancer research. The level of breast cancer public awareness is measured on a zero-one scale, with 1 being the best and zero being the worst, and is expressed mathematically as follows:

$$PublicAwareness(t) = f\left(PublicAds(t)^{+}, BCResearch(t)^{+}\right),$$

$$PublicAwareness(t) = \frac{\left[PublicAds(t-1) + BCResearch(t-1)\right] - 2}{4}. \quad (4.30)$$

4.2.5 Input Individual Characteristics from DES Submodel

Some individual characteristics are that are inputs to the factors in the adherence regression model are from the DES submodel. For example, the presence of comorbid diseases is a attribute used in the regression equation, and the probability of having comorbid diseases is dependent of both age and body mass index. Age, body mass index, the number of other people in the household, and the satisfaction with the individual screening technician are computed using DES logic.

4.2.5.1 Age

The minimum age in the model is 65, and the maximum possible age is 110, a function of the breast cancer–adjusted life tables. The maximum initial age was limited to 95, a function of the BCSC risk-estimation data set.

4.2.5.2 BMI

BMI is assigned at the beginning of the Phase I model and used in the Barlow risk model. However, the risk model allows for the possibility of BMI being unknown, and we need BMI to determine comorbidity (see the next section). For women who have unknown BMI, we use BCSC data from 2009 (the latest available year) to assign BMI, which is defined in the following way:

$$BMI \equiv \begin{cases} 1, & \text{if the individual's body mass index} \le 25, \\ 2, & \text{if } 25 < \text{the individual's body mass index} \le 30, \\ 3, & \text{if } 30 < \text{the individual's body mass index} \le 35, \\ 4, & \text{if the individual's body mass index} > 35. \end{cases} \tag{4.31}$$

The BCSC data yield the following probability distribution of BMI for a randomly sampled individual in the population of women of ages 65+:

$$\Pr\{\text{Value of } BMI \text{ assigned} = j\} = \begin{cases} 0.480, \text{ if } j = 1, \\ 0.281, \text{ if } j = 2, \\ 0.140, \text{ if } j = 3, \\ 0.099, \text{ if } j = 4. \end{cases} \tag{4.32}$$

Based on the BCSC data for the distribution of body mass index in the population of older US women, we see that the assigned values of 1, 2, 3, and 4 for the attribute BMI for an individual woman correspond respectively to the following ranges for the individual's body mass index: [0, 25]; (25, 30]; (30, 35]; and (35, ∞).

4.2.5.3 Number of Other Members in Household

The number of other members in each woman's household, *HouseSize*, is assigned upon entering the Phase II model. It is defined in the following way:

$$HouseSize \equiv \begin{cases} 0, & \text{if no other members in household,} \\ 1, & \text{if one other member in household,} \\ 2, & \text{if two or more other members in household.} \end{cases} \tag{4.33}$$

For a randomly sampled individual in the population of older US women, *HouseSize* is assigned according to the following distribution using data from Gierisch [37]:

$$\Pr\{HouseSize \text{ for an individual} = k\} = \begin{cases} 0.109, \text{ if } k = 0, \\ 0.527, \text{ if } k = 1, \\ 0.364, \text{ if } k = 2. \end{cases} \tag{4.34}$$

4.2.5.4 Satisfaction with Screening Technician

This individual characteristic is assigned each time a woman goes for screening. It is meant to capture each woman's level of satisfaction with her screening technician for that specific appointment. The highest level of satisfaction is one and the lowest level of

satisfaction is zero, and after each screening appointment this characteristic is assigned using a uniform distribution between zero and one:

$$ScreenTechSat(t) \sim UNIFORM(0,1).$$

We use the uniform distribution because the screening technician is typically not the same for a given woman each time she is screened: and according discussions with several women and our experts, there is significant variation in the level of satisfaction with a technician for a given appointment. The screening technician handles women's breasts, and this can be a uncomfortable experience for some women. Some technicians are better about being careful than others. In addition, some women do not like to engage in conversation and others do, and the screening technician may have a hard time judging this leading to a somewhat uncomfortable experience. This factor is an attempt to take into account these types of human interactions that take place during screening exams.

4.2.6 Individual Characteristics Linked Directly to Adherence

The individual characteristics that directly determine the probability of adherence to any screening appointment are a function of the Gierisch breast cancer aherence model [37]. These characteristics are functions of both individual characteristics from the DES model and population levels from the SD model. An attribute, "Comorbidity", representing each woman's health status was added to the model, and is a function of age and body mass index. Satisfaction with the screening process is a function of each woman's satisfaction with her individual screening techniican and her staisfaction with the overall screening process. The number of percieved barriers to screening is a function of each woman's house size, health status, and her access to screening. Each woman's intention to adhere to the screening policy is a function of her satisfaction with the screening process, her number of perceived barriers, and the level of public awareness amongst the population. Lastly, the number of times a woman was adherent to the screening policy out of the last two screening opportunities contributes to the regression equation.

4.2.6.1 Comorbidity

The comorbidity attribute is defined in Equation (4.35) below. When each woman enters the model, there is a probability she already has comorbidities present. If comorbidities are present, then it is assumed that they are not cured and persist for the duration of the woman's life. For women without comorbidities, each year there is some probability that comorbidities will become present. To describe this process, we use a two-state Markov chain as defined by Equation (4.36), with an initial distribution defined by Equation (4.37):

$$Comorbidity(t) = f\left(Age(t)^{+}, BMI^{+}\right),$$

$$Comorbidity(t) \equiv \begin{cases} 0, & \text{if Comorbid Conditions Not Present,} \\ 1, & \text{if Comorbid Condition(s) Present,} \end{cases} \tag{4.35}$$

where the 2x2 matrix $\mathbf{P_{co}}$ of one-step transition probabilities with (i, j) element $\Pr\{Comorbidity(t+1) = j \mid Comorbidity(t) = i\}$ has the form

$$\mathbf{P_{co}} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 1 - P_{CO}(t) & P_{CO}(t) \\ 0 & 1 \end{bmatrix} \end{array}, \tag{4.36}$$

where the variable $P_{CO}(t)$ is given in Table 4-16 as a function of age and body-mass index; and the 2x1 vector $\boldsymbol{\pi}_{\mathbf{co}}$ of initial probabilities with the j^{th} element $\Pr\{Comorbidity(0) = j\}$ has the form

$$\boldsymbol{\pi}_{\mathbf{co}} = \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 1 - \pi_{CO} & \pi_{CO} \end{bmatrix} \end{array}, \tag{4.37}$$

where the variable π_{CO} is also given in Table 4-16 as a function of age and body-mass index.

The relevant state transition probabilities and initial state probability distribution for each woman are a function age and body mass index (BMI).

$$\pi_{CO} = f\left(InitialAge^{+}, BMI^{+}\right),$$

$$P_{CO}(t) = f\left(Age(t)^{+}, BMI^{+}\right).$$

Table 4-16 describes the relationship between age, BMI, and the two relevant probabilities. Piecewise linear functions are used to compute the probabilities not explicitly stated in Table 4-16.

Table 4-16. Relationship Between Comorbidity, Age, and BMI

Age/BMI	$BMI = 1$		$BMI = 2$		$BMI = 3$		$BMI = 4$	
Age	π_{CO}	$P_{CO}(t)$	π_{CO}	$P_{CO}(t)$	π_{CO}	$P_{CO}(t)$	π_{CO}	$P_{CO}(t)$
65	.15	0.0100	.20	0.0133	.25	0.0167	.30	0.0200
85	.35	0.0200	.40	0.0267	.45	0.0333	.50	0.0400
95	.45	0.0300	.45	0.0400	.55	0.0500	.60	0.0600
100	-	0.0500	-	0.0667	-	0.0833	-	0.1000
110	-	0.1000	-	0.1333	-	0.1666	-	0.2000

For each woman in the sample, her associated Markov chain governs changes in her attribute *Comorbidity*(t). Her age will increase over time, but her *BMI* category will stay the same. The probability $P_{co}(t)$ of initially acquiring comorbid conditions at time *t* increases with age and *BMI*.

4.2.6.2 Satisfaction with Mammography

When a woman enters the model, an initial distribution, $\boldsymbol{\pi}_{\mathbf{SAT}}$, will be used to determine the each woman's satisfaction with her previous mammographic experiences. This distribution is given in Equation (4.39). The probabilities were derived from data in Gierisch(2010).

$$MammSat(t) = f\left(ScreenTechSat(t)^{+}, ScreenProcSat(t)^{+}\right),$$

$$MammSat(t) \equiv \begin{cases} 0, & \text{if somewhat satisfied, dissatisfied, or very dissatisfied,} \\ 1, & \text{if satisfied or very satisfied,} \end{cases} \tag{4.38}$$

$$\boldsymbol{\pi}_{\mathbf{SAT}} = \begin{bmatrix} 0 & 1 \\ 1-\pi_{SAT} & \pi_{SAT} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.111 & 0.889 \end{bmatrix}. \tag{4.39}$$

After each woman attends her first screening, her probability of being satisfied with the overall screening experience, $P_{SAT}(t)$, is a function of her experience with her technician on that visit and her satisfaction with the screening process,

$$P_{SAT}(t) = f\left(ScreenTechSat(t)^{+}, ScreenProcSat(t)^{+}\right).$$

Each time a woman goes for screening, we sample a value of *ScreenTechSat(t)* from the uniform distribution on the unit interval [0,1],

$$ScreenTechSat(t) \sim UNIFORM(0,1).$$

It is important to account for the fact that individual women may be heavily influenced by some factors and slightly influenced by others, and this may vary greatly from woman to woman. To account for this, two personal penalty factors, $ScreenTechPen$ and $ScreenProcPen$, are used as weights that describe how much emphasis each individual woman puts on that factor.

$$ScreenTechPen \sim UNIFORM\left(0.25,1\right) \text{ and } ScreenProcPen \sim UNIFORM\left(0.25,1\right).$$

Finally, the probability of being satisfied with the overall screening experience for each woman can be computed using Equation (4.40):

$$\begin{aligned} P_{SAT}(t) = 1 &- \left(ScreenTechPen\right) * \left(1 - ScreenTechSat(t)\right) * (0.5) \\ &- \left(ScreenProcPen\right) * \left(1 - ScreenProcSat(t)\right) * (0.5), \end{aligned} \tag{4.40}$$

$$MammSat(t) = \begin{cases} 0, & \text{with probability } 1 - P_{SAT}(t), \\ 1, & \text{with probability } P_{SAT}(t). \end{cases} \tag{4.41}$$

4.2.6.3 Perceived Number of Barriers to Screening

The definition of the number of perceived barriers to screening is given in Equation (4.42). We consider three possible barriers: health status, the number of other members in the household, and the level of access to screening. Each of these is weighted according to a penalty factor to account for the fact that some women may be influenced by some of these factors more than others.

$$NumofBarriers(t) = f\left(Comorbidity(t)^{+}, HouseSize^{+}, Access(t)^{-}\right),$$

$$NumofBarriers(t) \equiv \begin{cases} 0, & \text{if the woman has 0 perceived barriers,} \\ 1, & \text{if the woman has 1 perceived barrier,} \\ 2, & \text{if the woman has 2+ perceived barriers.} \end{cases} \tag{4.42}$$

Again, in order to account for individual emphasis on certain factors, we introduce three more penalty factors, *ComorbidPen*, *HouseSizePen*, and *AcessPen*. The perceived number of barriers to screening for each woman can be computed according to the Equation (4.43). This is translated into the three states {0, 1, 2} using the established thresholds in Equation (4.44):

$$\begin{aligned} &ComorbidPen \sim UNIFORM\left(0.25,1\right), \\ &HouseSizePen \sim UNIFORM\left(0.25,1\right), \\ &AccessPen \sim UNIFORM\left(0.25,1\right), \end{aligned}$$

$$\begin{aligned} NumofBarriers(t) = &\left(ComorbidPen\right)*\left(Comorbidity(t) == 1\right) \\ &+\left(HouseSizePen\right)*\left(HouseSize == 0\right)(1) \\ &+\left(HouseSizePen\right)*\left(HouseSize == 1\right)(0.5) \\ &+\left(AccessPen\right)*\left(1 - Access(t)\right), \end{aligned} \tag{4.43}$$

$$NumofBarriers(t) = \begin{cases} 0, & 0 \le NumofBarriers(t) \le 0.75, \\ 1, & 0.75 < NumofBarriers(t) \le 1.75, \\ 2, & NumofBarriers(t) > 1.75. \end{cases} \tag{4.44}$$

4.2.6.4 Intention to Adhere to Screening Policy:

Intention to adhere to the screening policy is defined in Equation (4.45). Upon entering the model, an initial distribution, $\boldsymbol{\pi}_{\mathbf{INT}}$, will be used to determine each woman's intention to adhere to the screening policy. This distribution is given below. The probabilities were derived from data in Gierisch [37].

$$IntentToAdhere(t) = f\left(MammSat(t)^{+}, NumOfBarriers(t)^{-}, PublicAwareness(t)^{+}\right),$$

$$IntentToAdhere(t) \equiv \begin{cases} 0, & \text{if woman does not intend to adhere,} \\ 1, & \text{if woman intends to adhere,} \end{cases} \tag{4.45}$$

$$\boldsymbol{\pi}_{\mathbf{INT}} = \begin{bmatrix} 0 & 1 \\ 1-\pi_{INT} & \pi_{INT} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.082 & 0.918 \end{bmatrix}. \tag{4.46}$$

After each woman attends her first screening, her probability of having the intent to adhere to the screening policy, $P_{INT}(t)$, is a function of her screening experience satisfaction, her perception of the number of barriers to screening, and public awareness of breast cancer. The intent for each woman is computed according to Equation (4.47):

$$P_{INT}(t) = f\left(MammSat(t)^{+}, NumOfBarriers(t)^{-}, PublicAwareness(t)^{+}\right),$$

$$P_{INT}(t) = \frac{MammSat(t) + \left(3 - NumOfBarriers(t)\right) + PublicAwareness(t)}{5}, \tag{4.47}$$

$$IntentToAdhere(t) = \begin{cases} 0, \text{ with probability } 1 - P_{INT}(t), \\ 1, \text{ with probability } P_{INT}(t). \end{cases} \tag{4.48}$$

4.2.6.5 Number of Times Adherent to Screening Policy

Each individual woman will have an attribute, $Last2Adherence(t)$, which designates the number of times that she actually went to screening when the screening policy suggested she should out of the last two opportunities. Possible values are zero, one, or two; and the definition is given in Equation (4.49):

$$Last2Adherence(t) = f\left(\text{last 2 times due for screening}\right),$$

$$Last2Adherence(t) \equiv \begin{cases} 0, & \text{if she adhered 0 times out of the last 2 opportunities,} \\ 1, & \text{if she adhered 1 times out of the last 2 opportunities,} \\ 2, & \text{if she adhered 2 times out of the last 2 opportunities.} \end{cases} \quad (4.49)$$

4.3 Linkage between DES and SD Submodels

There are two fundamental ways in which the DES submodel and the SD submodel interact, through hybrid DES/SD levels and through the Gierisch breast cancer aherence model [37]. The adherence regression model allows us to compute the probability of adhering to each screening appointment specified by the screening policy for each woman in the model. Individual characteristics assigned in the DES submodel are used in the regression model, and some are functions of SD levels; thus the DES and SD submodels are interacting through this adherence regression. In addition, adherence to the screening appointments directly determines demand in the DES submodel, and this value of demand is used to determine the level of congestion at faciltities in the SD submodel. Demand is an example of what we have called "hybrid levels." These hybrid levels can be functions of other SD levels or individual characteristics from the DES model, or both. They can have an effect on other SD levels, individual characteristics, or both. For example, the level of demand for screening is actually a function of numerous discrete-events, and it is also a function of adherence, which itself is a function of the SD levels. The same is true for the level of adherence.

4.3.1 The Gierisch Breast Cancer Adherence Model

The Gierisch breast cancer adherence model [37] was developed using the data from the study called Personally Relevant Information about Screening Mammography (PRISM), a health communication intervention study [25]. PRISM was a National Cancer Institute (NCI) funded intervention trial designed to enhance annual mammography adherence. It was conducted from October 2003 to September 2008 as part of the NIH Health Maintenance Consortium. The sampling frame for PRISM was female North Carolina residents, ages 40 to 75 years, who were enrolled in the North Carolina State Health Plan for Teachers and State Employees for two or more years before sampling. According to [37], 80% of the women in this study were 50 years or older. After reviewing relevant literature, we judged that this model presented the best possible match to our situation. We communicated with the author in an attempt to obtain the data and verify the model. Unfortunately the data were obtained as part of an agreement that has since been terminated, so the data are no longer available. However, we did verify that the model produced the given results, and our experts reviewed the results and deeemed them reasonable. Just as with the Barlow Risk Model, there is a dependent (response) variable, and for this model it is a binary variable for which the values 0 and 1 denote non adherence and adherence to annual screening, respectively; and the predictor variables or "covariates" are the categorical independent variables presented in Table 4-17. The level of each of these factors for each woman is dependent on SD levels and individual attributes.

Table 4-17. Characteristics of Individuals Directly Affecting Adherence form Gierisch [37]

Factor #	Factor	Factor Name	Factor Coding
1	x_1	Presence of Comorbidities	0 = No 1 = Yes
2	x_2	Satisfaction with Screening Process	0 = Not or Somewhat Satisfied 1 = Mostly or Totally Satisfied
3	x_3	Number of Perceived Barriers	0 = 0 Barriers 1 = 1 Barrier 2 = 2+ Barriers
4	x_4	Intention to Adhere to Screening Policy	0 = No 1 = Yes
5	x_5	Number of Times Adherent to Screening Appointment (of Last 2)	0 = 0 of Last 2 Appointments 1 = 1 of Last 2 Appointments 2 = 2 of Last 2 Appointments
6	x_6	Screening Interval	1 = Annual Screening 2 = Every 2 Years 3 = Every 3 Years 4 = Every 4 Years 5 = Every 5 Years

A *p*-value criterion of $p < 0.05$ was required for a factor to be statistically significant and included in the model. There were seven statistically significant risk factors: age, health status, satisfaction with the screening process, number of perceived barriers to screening, intention to adhere to screening, and the number times the woman has adhered to the screening policy (of the last two opportunities). The following risk factors were not statistically significant: race, education level, marital status, perceived financial situation, family history of breast cancer, doctor recommendation for mammogram within the last year, attitude about mammography, and perceived risk of getting breast cancer. The levels of age were under fifty and over fifty, so although this factor was significant, its level was constant for our designated population of women of age 65+ and thus was not relevant for our purposes.

Equation (4.50) is the regression equation, which was formed using the coefficients given in Table 4-17. Equation (4.51) is the explicit equation that can be used to calculate the probability of *nonadherence* to a screening appointment when given the values of the regression coefficients and the values of the categorical variables for a specific woman. The probability of adherence can be computed by taking one minus the probability of non-adherence. The logit can be calculated according to Equation (4.52), noting that the double equals sign is a true/false statement that evaluates to either zero or one:

$$L_{\text{AD}} = \text{Logit}(p_{\text{AD}}) = \ln\left(\frac{p_{\text{AD}}}{1-p_{\text{AD}}}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6, \quad (4.50)$$

$$p_{\text{AD}} = f(L_{\text{AD}}) = \frac{\exp(L_{\text{AD}})}{1+\exp(L_{\text{AD}})} = \frac{1}{1+\exp(-L_{\text{AD}})}, \quad (4.51)$$

$$\begin{aligned} L_{\text{AD}} = {} & -1.6852 + (\text{Comorbidity} == 1)*(0.5188) + (\text{Comorbidity} == 0)*(0) \\ & + (\text{MammSat} == 0)*(0.3577) + (\text{MammSat} == 1)*(0) + (\text{NumOfBarriers} == 2)*(0.392) \\ & + (\text{NumOfBarriers} == 1)*(0.3001) + (\text{NumOfBarriers} == 0)*(0) \\ & + (\text{IntentToAdhere} == 0)*(0.9783) + (\text{IntentToAdhere} == 1)*(0) \\ & + (\text{Last2Adherence})*(-0.1744) + (\text{ScreenFreq} == 5)*(0.47) \\ & + (\text{ScreenFreq} == 4)*(0.37156) + (\text{ScreenFreq} == 3)*(0.26236) \\ & + (\text{ScreenFreq} == 2)*(0.13976) \; + \; (\text{ScreenFreq} == 1)*(0). \end{aligned} \quad (4.52)$$

4.3.2 Hybrid Levels

Hybrid levels are those which really belong to both the DES submodel logic and the SD submodel logic, and these levels can be accessed by the respective submodels to perform computations whenever they are needed. Demand for screening is a hybrid level, because the DES submodel logic is actually counting the number of women attending screening appointments each year, but this value is shared with the SD submodel when computing the

congestion at facilities. In addition, demand is a function of adherence, which is a function of several SD levels. This is very similar to the level of adherence, where the adherence of individual women is controlled in the DES submodel logic, but the likelihood that individual women adhere to the screening policy is a function of several SD levels. The screening interval is assigned using a population-level screening policy; however this screening interval could be different for individual women based on their individual attributes. The population size was actually determined using DES logic in the Phase I natural history model, but it is a population-level characteristic, so that it is also considered to be a hybrid level. The levels of false positives, false negatives, and survival after treatment are in the form of adjustment factors that are determined by the level of screening technology in the SD submodel. These adjustment values are then used in real time to alter the probabilties pertaining to false positives, false negavites, and survival after treatment.

4.4 User Interface

The user interface was designed to make alternative screening policies very easy to implement in the integrated screening model. One of the goals of this research was to ensure that this model, or some form of it, could be understood by clinical experts and government policy makers. Because we own the rights to this model, we are able to share its abilities and results with whomever we please. With the appropriate Arena software (or access to it via a web applet of some sort), the user can simply click the run button, and a series of prompts are displayed that allow the user to choose several options related to the screening policies for both the past and future periods. Default values for the period 2001–2011 are suggested to users, and they are encouraged to choose those values unless they have a valid reason not to. In addition to a user interface for screening policies, one of the forms in the user interface allows the user to select values for the SD input variables, such as the trend in the number of screening facilities. Again default values are suggested, but users are allowed to change these values if they want to explore alternative scenarios. These prompts were designed in Microsoft Visual Basic, which has an editor embedded within Arena.

Upon clicking the run button the message box in Figure 4-9 appears with the following statement: "Unless you have communicated with the programmer, please select the default values for all prompts which have a default value. This will ensure the correct inputs are used for the past period 2001–2011, and accurate results are obtained for the future period 2012–2020."

Figure 4-9. User Message Box Appearing Before the User Interface Launches

After clicking the OK button, the user is given the form in Figure 4-10, which gives the user the opportunity to choose the screening interval for the period 2001–2011. The default value is one year, and this default value was chosen using our validation technique presented in the next section. The only reason for choosing a different value would be if the user wanted to determine what would have happened from 2001–2011 if a different screening policy had been chosen, but this has limited value since the population starts out cancer free, and it takes five to ten years (depending on the statistic examined) for most of the important performance measures to reach steady-state. If the user closes this box without making a selection, the value of the screening interval defaults to one year.

Figure 4-10. User Interface Prompt for Screening Interval during the Period 2001–2011

Following a selection by clicking the OK button (or by closing the prompt), the dialogue box in Figure 4-11 is shown regardless of the previous selection. The user is allowed to choose whether the stopping age is a deterministic value that stops screening for every woman, or is assigned according to a distribution so that each woman has her own individual stopping age. For the period 2001–2011, there was no definitive stopping age, so it makes sense to use a stochastic distribution to assign a stopping age to each woman individually. Thus, stochastic is the default for this prompt, and the user is encouraged to select this value.

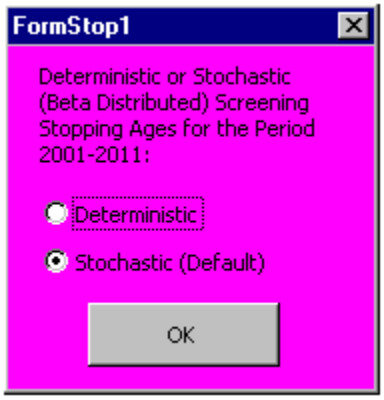


Figure 4-11. User Interface Prompt for Type of Screening Stopping Age during the Period 2001–2011

If “deterministic” is selected for the type of stopping age, then the prompt in Figure 4-12 will appear, which instructs the user to select a deterministic stopping age. If “stochastic” is chosen as the type of stopping age, then the prompt in Figure 4-13 will appear, which instructs the user to select the mode of a beta distribution for the stopping age with a minimum of 65 and a maximum of 100. The default is a stochastic stopping age with a mode of 75 years, because this was shown to best match the data for that time period (see the next section).

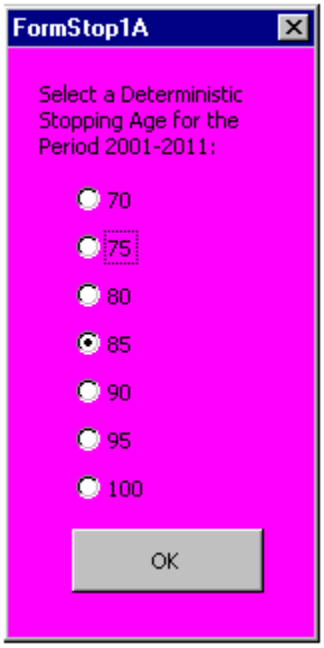


Figure 4-12. User Interface Prompt for Deterministic Screening Stopping Age during the Period 2001–2011

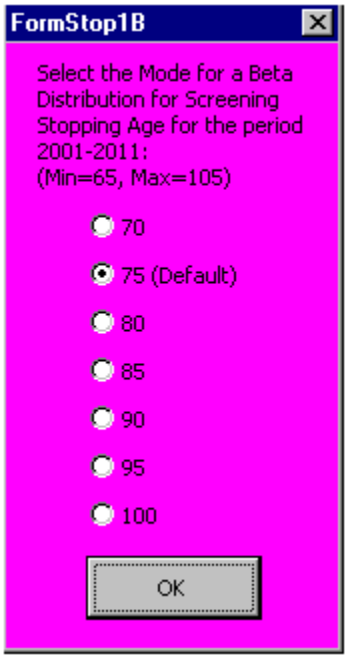


Figure 4-13. User Interface Prompt for Stochastic Screening Stopping Age during the Period 2001–2011

This concludes the prompts, which govern the screening policy for the period 2001–2011. The next series of prompts governs the screening policy for the period 2012–2020. First, the user is presented with the prompt in Figure 4-14, which allows the user to select one of three different types of screening policies: interval-based screening, risk-based screening, and factor based screening. The next prompt depends on the type of screening policy selected.

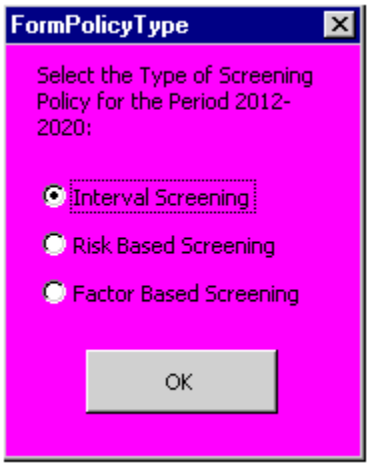


Figure 4-14. User Interface Prompt for the Type of Screening Policy during the Period 2001–2011

If the user selects “interval screening” as the policy type, then the same screening interval will be used for all women; and the prompt in Figure 4-15 will appear, allowing the user to choose the screening interval. After the user selects the screening interval, the stopping age options will appear.

If the user selects “risk-based screening” as the policy type, then the user is allowed to choose different screening intervals for high-risk and low-risk women. The user also chooses the definition of "high-risk." In the natural history model, we used the Barlow risk model to estimate the one-year risk of being diagnosed with cancer. We recorded individual observations, and then determined the 95^{th}, 90^{th}, 85^{th}, and 80^{th} percentiles of the data, which represent the top 5%, 10%, 15%, and 20% of women in terms of annual risk. Figure 4-16

appears first, and prompts the user to select a screening interval for high-risk women and whether high risk corresponds to the top 5%, 10%, 15%, and 20% of women in terms of annual risk. Figure 4-17 then appears and prompts the user to select the screening interval for low-risk women. After the user selects the screening interval for low-risk women, the stopping age options will appear.

If the user selects factor-based screening, then the prompt in Figure 4-18 will appear. The eleven risk factors are listed along the left side, and there is box next to each factor. Selecting the box next to a factor means the screening intervals that the user selects for different levels of that factor will be taken into account when determining the screening interval used for each woman. The user is allowed to select screening intervals of one, two, three, four, or five years for each level of every factor. If the user selects multiple factors, then each woman will be assigned the appropriate screening interval for each factor, and the smallest interval of that group is actually used. This is a rather conservative approach. Another approach could be to average the numbers and round up and/or round down, and this may be considered in future work. After the user clicks the OK button, the stopping age options will appear.

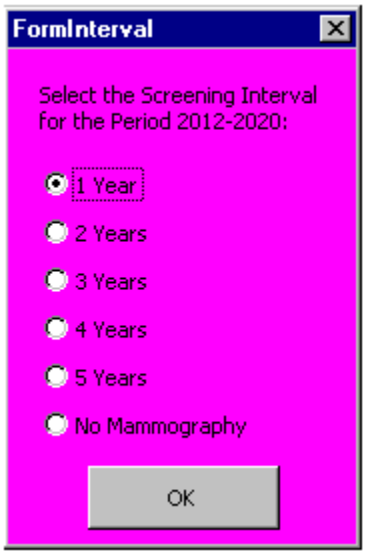


Figure 4-15. User Interface Prompt for Screening Interval for the Period 2012–2020

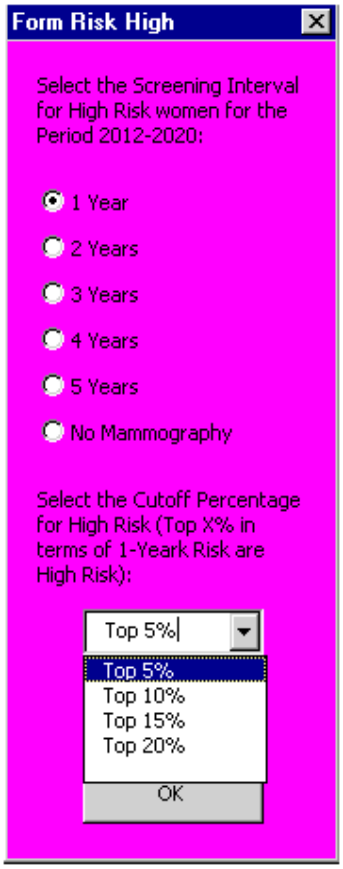


Figure 4-16. User Interface Prompt for Screening Interval for High Risk Women for the Period 2012–2020

Figure 4-17. User Interface Prompt for Screening Interval for Low Risk Women for the Period 2012–2020

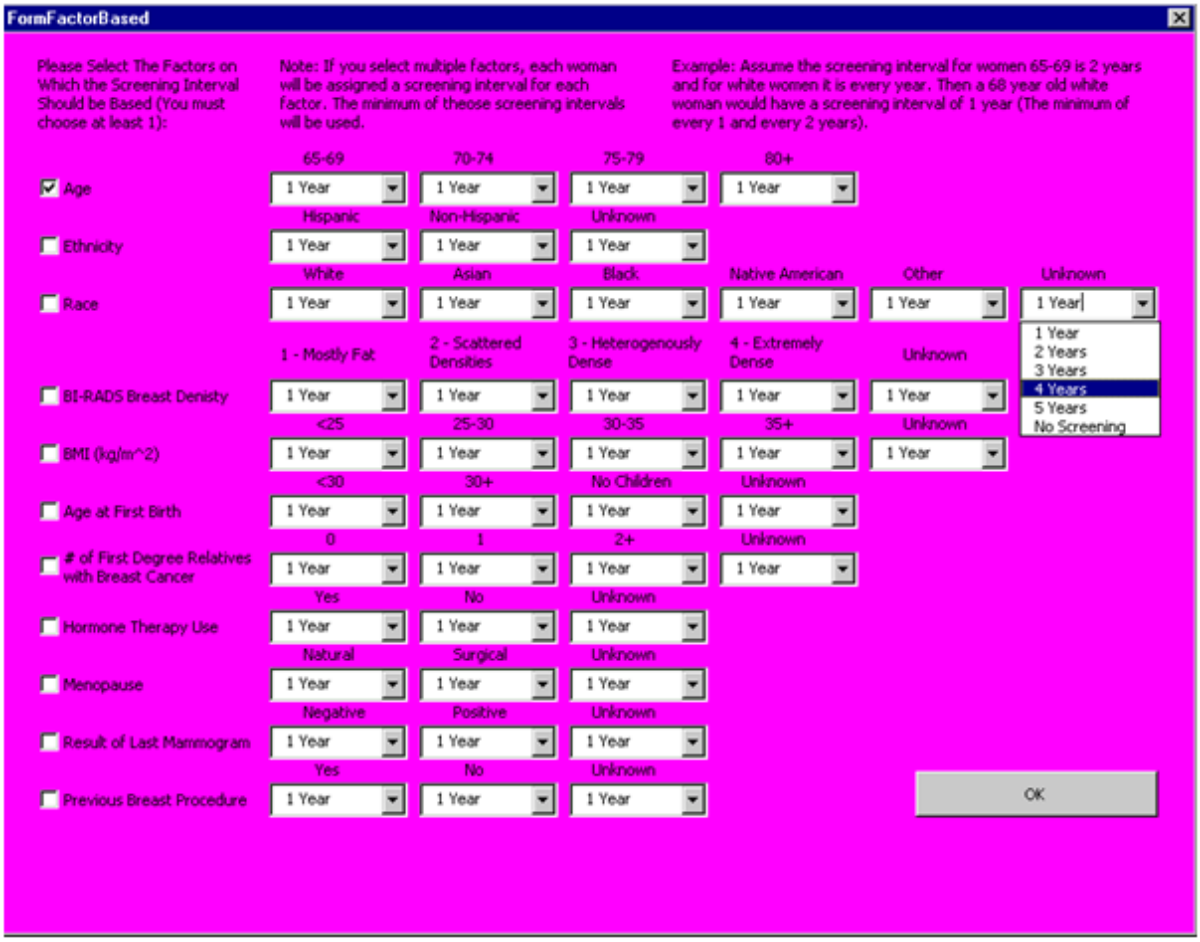


Figure 4-18. User Interface Prompt for Factor-Based Screening Intervals for the Period 2012–2020

The prompts for the type of stopping age (deterministic or stochastic) and then the appropriate prompts that follow are the same for the period 2012–2020 as they were for the period 2001–2011. The prompt in Figure 4-19 is shown first, prompting the user to select either a deterministic or stochastic stopping age. A stochastic stopping age was appropriate for the past when there was no actual policy; but for future periods, we would like to set a strict cutoff age at which people are no longer screened. Thus a deterministic stopping age is the default for the period 2012–2020. If the user selects a deterministic stopping age, then the prompt in Figure 4-20 appears and allows the user to select a deterministic stopping age; and if the user selects a stochastic stopping age, then the prompt in Figure 4-21 appears and allows the user to select the mode of a Beta distribution for the stopping age.

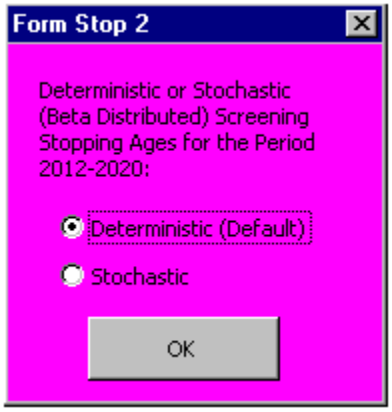


Figure 4-19. User Interface Prompt for Type of Screening Stopping Age for the Period 2011-2020

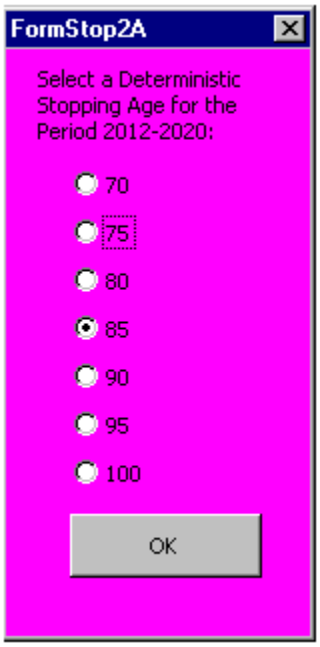


Figure 4-20. User Interface Prompt for Deterministic Screening Stopping Age for the Period 2012–2020

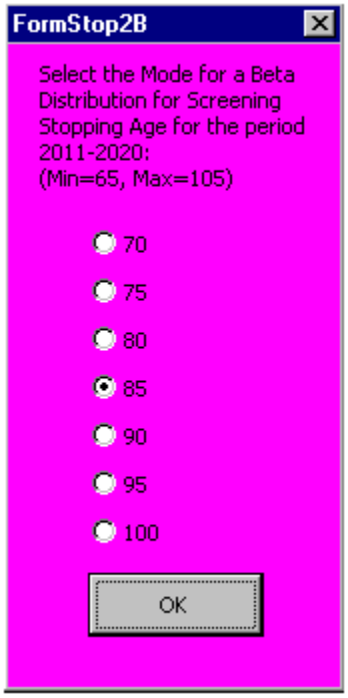


Figure 4-21. User Interface Prompt for Stochastic Screening Stopping Age for the Period 2012–2020

After the user selects a stopping age or mode for the stopping age and clicks the OK button, the message box in Figure 4-22 will appear with the following text: "Unless you have communicated with the programmer, please select default values for the Breast Cancer Research and Advertising for the period 2001–2011."

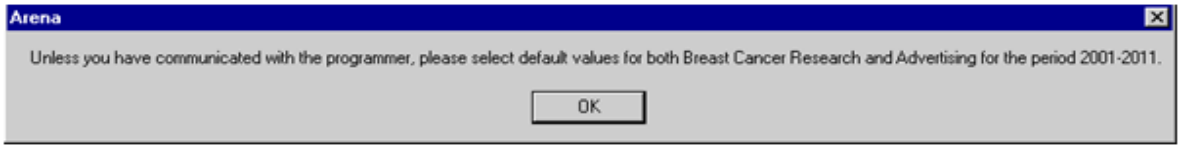


Figure 4-22. User Message Box Appearing Before SD Input Prompt

After the user clicks the OK button on the message box, the final prompt in Figure 4-23 appears, allowing the user to select the inputs for the SD submodel. The user is allowed to select the trend in the number of breast cancer screening facilities and the trend in the capacity of those facilities. The user is also allowed to define the level of public advertising and the level of breast cancer research for the periods 2001–2011 and 2012–2020 separately.

After the user clicks the OK button on this form, the model executes using the input values selected by the user. Extensive output reports are generated, with results for each of the 10 populations individually, and another report is generated that aggregates those results into 95% confidence intervals.

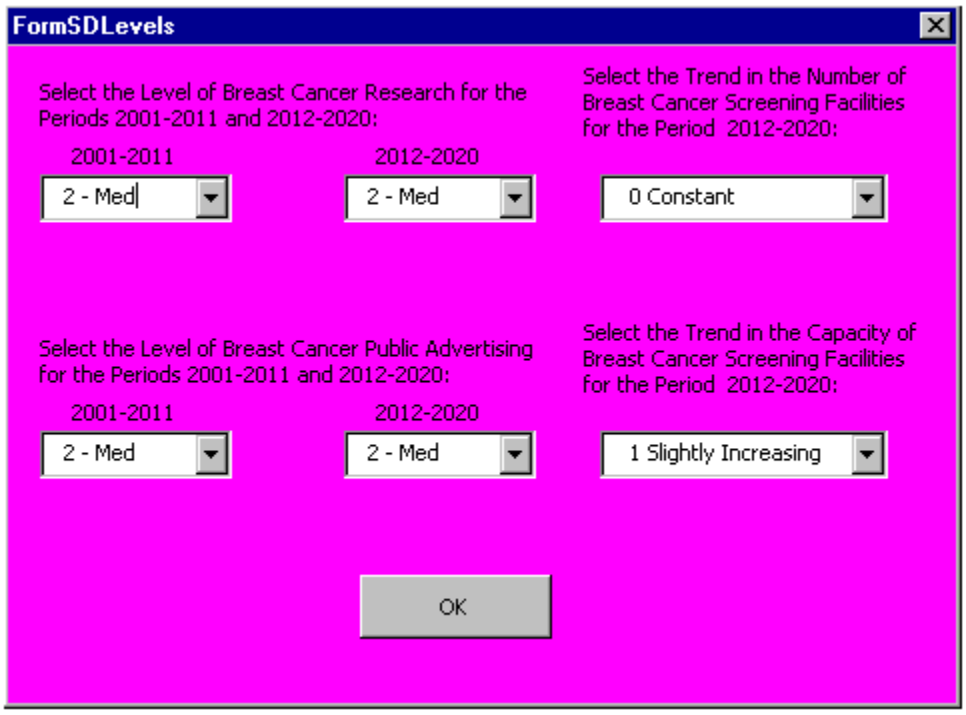


Figure 4-23. User Interface Prompt for SD Input Levels

4.5 Analysis of Results, Discussion, and Validation Techniques

This section presents the results of the validation process for the period 2001–2011, the results from simulation experiments with alternative screening policies for the period 2012–2020, and a discussion of all these results. Optimization was used in several contexts and the details of these optimization problems are also presented in this section. After using optimization to identify the 5 best screening policies for three separate optimizations, we use paired *t*-tests to identify the policy or group of policies that can be declared statistically better in terms of a given performance measure (possibly subject to some constraints). Although cost per QALY saved is a standard measure of cost-effectiveness, this may not be the best

measure of effectiveness for breast cancer screening policies. Thus, several optimization routines with different objectives and constraints were run in order to determine the best policies under different circumstances. A detailed discussion of these results and some insight into their meaning is given at the end of the section. In order to make a convincing case that the simulation-generated results for the future period 2012–2020 are a valid representation of what will actually happen, we validate the model output for the period 2001–2011 against SEER breast cancer data available for that time period. Our techniques for validating the model are explained in detail, and compelling evidence toward that the simulation model is a reasonably accurate representation of what actually happened from 2001–2011 is presented. In fact, we make the case several screening policies could have been considered valid for that past time period. Note that while it is not the main goal of this particular research, the model can easily be used to determine what would have happened if alternative screening policies had been used for the period 2001–2011.

4.5.1 Calibration and Validation

In order to provide a convincing case that this integrated simulation model is able to accurately predict the results of using alternative screening policies for the future period 2012–2020, we developed a formal method for validating certain performance measures for the period 2001–2009 for which data were available. Since there was no screening policy in place for this period, we cannot impose a policy and then match the results to data by tuning the adherence submodel. Instead, we must impose several polices and then determine which policy best matches the data, and this "historically calibrated" policy will be in effect from 2001 to 2011. Alternative screening policies are tested for the period 2011–2020. We define a screening policy in terms of a screening interval (which could be different for different women based on their individual attributes) and a stopping age. The screening intervals to be compared are screening every one, two, and three years. Since there was no set stopping age for screening during this time period, we do not want to force every woman to stop screening at some specific age. However, we need to have some control over when women stop screening, and we need to be able to vary that stopping age. There may be some women who

never go for screening after they turn 65, so the absolute minimum stopping age would be 65. The oldest possible age in the model is 110; however, it is extremely unlikely any woman reaches this age, and so we set the maximum stopping age at 105 years old. Using these values as the upper and lower limits, we fit seven Beta distributions with modes 70, 75, 80, 85, 90, 95, and 100. These Beta distributions are defined according to Equation (4.53).

$$\text{Stopping Age Distribution} = \begin{cases} 65 + 40 * BETA(1.480, 4.360), & \text{if mode} = 70, \\ 65 + 40 * BETA(2.200, 4.600), & \text{if mode} = 75, \\ 65 + 40 * BETA(3.117, 4.529), & \text{if mode} = 80, \\ 65 + 40 * BETA(4.000, 4.000), & \text{if mode} = 85, \\ 65 + 40 * BETA(4.529, 3.117), & \text{if mode} = 90, \\ 65 + 40 * BETA(4.600, 2.200), & \text{if mode} = 95, \\ 65 + 40 * BETA(4.360, 1.480), & \text{if mode} = 100. \end{cases} \tag{4.53}$$

This yields five screening intervals and seven potential distributions for stopping age, and there are a total of 21 combinations of these two inputs, leaving us with 35 potential screening policies that could be used for 2001–2011. We generated full sets of simulated results for all 21 policies, and then we performed our validation technique on selected performance measures for each of these 21 policies. The best policy was selected, and that policy was used for 2001–2011 regardless of the alternative policy in place from 2012–2020.

We have the following situation: the simulation model produces mean values and confidence interval half-lengths (a measure of the precision of the estimated mean value), but the available data only provide mean values, although we know there is some variation in these reported values. Thus, we develop a formal statistical method for comparing the simulation-generated observations to observed values, and we present a reasonable measure of overall closeness of the simulation results to an observed set of real-world observations. We use this overall measure to determine the screening policy that best matches the available

data. Ten randomly sampled populations were used to generate mean values and CI half-lengths. After experimenting with five simulated populations, we determined that in order to get CI half-lengths of appropriate size, ten simulated populations were necessary. However, as a result, the CI half-lengths for the percentages of women diagnosed in each stage of breast cancer and for the percentage of benign biopsies are very small. Thus, those measures were validated from a practical standpoint by simply declaring whether or not the mean simulation values were within ± 10% of the mean data values. SEER data [66] were used for measures 1–6 below, and literature [39] was used for measure 7. The following categories of performance measures could be used to compare simulation results with observed data:

1. US Female 65+ Population Size (2002–2010)

   Data were available for the period 2001–2010; but the initial population size in 2001 was set to its the actual value, so this comparison for the start of 2001 would have no meaning. The other nine data points are formally validated against simulation-generated observations using our revised *t*-statistic method.

2. US Female 65+ Invasive Cancer Incidence Rates (2005–2009)

   Data were available for the period 2001–2009; however it takes the simulation model until the end of 2004 to "warm up" and produce reasonable values. The five data points for the period 2005–2009 are formally validated against simulation-generated observations using our revised *t*-statistic method. The volatility in this data make fitting a statistical model to the data very difficult, and testing simulation-generated values against that model would not be a reliable validation technique.

3. US Female 65+ DCIS Incidence Rates (2005–2009)

   Data were available for the period 2001–2009; however it takes the simulation model until the end of 2004 to "warm up" and produce reasonable values. The five data points for the period 2005–2009 are formally validated against

simulation-generated observations using our revised *t*-statistic method. The volatility in this data make fitting a statistical model to the data very difficult, and testing simulation-generated values against that model would not be a reliable validation technique.

4. US Female 65+ Total Incidence Rates (2005-2009)

   Data were available for the period 2001–2009; however it takes the simulation model until the end of 2004 to "warm up" and produce reasonable values. The five data points for the period 2005–2009 are formally validated against simulation-generated observations using our revised *t*-statistic method. The volatility in this data make fitting a statistical model to the data very difficult, and testing simulation-generated values against that model would not be a reliable validation technique.

5. US Female 65+ Cancer Death Rates (2011-2020)

   Data were available for the period 2001–2009; however it takes the simulation model until the end of 2010 to "warm up" and produce reasonable values for cancer death rates. As expected, there is a longer warm up period needed for cancer death rates than for cancer incidence rates. This leaves us with no data points that can be compared with any meaning. However the observed data was relatively stable, which allowed us to fit a linear model to the data, predict values for the period 2011-2020, and test simulation-generated values against that model. The ten model-predicted points are formally validated against simulation-generated observations using our revised *t*-statistic method.

6. US Female 65+ Stage Distribution at Diagnosis

   Unlike the previous five performance criteria, the percentages of women diagnosed in each stage (local, regional, distant) of breast cancer are not available on a yearly basis; instead they are simply overall percentages based

on data from the period 2001–2009. Since these values are limited on the range [0,100] because they are percentages, the CI half-lengths generated by the simulation model are very small, on the order of 1%. Thus, this measure was validated from a practical standpoint by simply declaring whether or not the mean simulation values were within ± 10% of the mean data values. There are percentages for local, regional, and distant cancers, so there are three tests for practical significance performed using this data, one for each stage.

7. PPV3 (or Percentage of Benign Biopsies)

   The percentage of benign biopsies is not available in BCSC data. Thus, we used literature [39] to determine that about 80% of breast biopsies are negative. Since this value is also limited on the range [0,100], the CI half-lengths generated by the simulation model are very small, again on the order of 1%. Thus, this measure was validated from a practical standpoint by simply declaring whether or not the mean simulation value is within ± 10% of the value of 80% derived from literature.

Thus, there are 34 performance measures from categories 1–5 above for which statistical analysis is used to determine whether or not a significant difference exists between the data and the simulation-generated observations. The other four performance measures from categories 6 and 7 above were evaluated from a practical standpoint as previously stated. For several scenarios, all 34 statistical tests are passed, and thus our development of a measure of overall closeness. The following section presents the development of our validation techniques.

4.5.1.1 Development of Validation Techniques

Suppose we have k performance criteria $\{Y_i : i = 1,...,k\}$ that are observed on the real system, and for a given configuration of the simulation model we perform m independent replications so that for the i^{th} performance criteria $(i = 1,...,k)$, we have the independent and identically distributed (IID) simulation generated observations:

$$\{X_{ij} : j = 1,...,m\} \overset{IID}{\sim} \text{Normal}(\mu_i, \sigma_i^2) \tag{4.54}$$

with sample mean

$$\overline{X_i} = \frac{1}{m}\sum_{j=1}^{m} X_{ij} \tag{4.55}$$

and sample variance

$$S_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(X_{ij} - \overline{X_i}\right)^2. \tag{4.56}$$

Now the i^{th} performance measure Y_i observed on the real system is subject to random variation; however we have no data about the variance, only the mean values. Thus, we assume that

$$Y_i \sim \text{Normal}(\mu_i, \sigma_i^2), \tag{4.57}$$

the same distribution as for the i^{th} simulation-generated performance measure on each replication. This assumption may not provide a sufficiently accurate approximation to the variance of each system-performance estimator that is based on real-world data. In future research we will formulate more-accurate approximations to the variances of these point estimators so that we can perform a more-refined validation of the Tejada model versus the system-performance estimators that are based on real-world data. Based on our current variance assumption, we have

$$\mathrm{Var}\left[Y_i - \overline{X_i}\right] = \mathrm{Var}\left[Y_i\right] + \mathrm{Var}\left[\overline{X_i}\right] = \sigma_i^{\,2} + \frac{\sigma_i^{\,2}}{m} = \sigma_i^{\,2}\left(\frac{m+1}{m}\right) \tag{4.58}$$

and the sample variance estimate

$$\widehat{\mathrm{Var}}\left[Y_i - \overline{X_i}\right] = \left(m+1\right)S_i^{\,2}/m \tag{4.59}$$

has a scaled chi-squared distribution with $m-1$ degrees of freedom.

A natural measure of discrepancy between Y_i and $\overline{X_i}$ that is dimensionless (and thus can be combined with the same measure of discrepancy for other performance criteria) is

$$t_i = \frac{Y_i - \overline{X_i}}{\sqrt{\left(m+1\right)S_i^{\,2}/m}} = \frac{\left(Y_i - \overline{X_i}\right)\Big/\sqrt{\left(m+1\right)\sigma_i^{\,2}/m}}{\sqrt{\dfrac{\left(m+1\right)S_i^{\,2}/m}{\left(m+1\right)\sigma_i^{\,2}/m}}} = \frac{\left(Y_i - \overline{X_i}\right)\Big/\sqrt{\left(m+1\right)\sigma_i^{\,2}/m}}{\sqrt{S_i^{\,2}/\sigma_i^{\,2}}} = \frac{Z}{\sqrt{\chi^2_{m-1}}} \tag{4.60}$$

where $Z \sim Normal\left(0,1\right)$ is independent of $\chi^2_{m-1} \sim$ chi-squared random variable with $m-1$ degrees of freedom under the null hypothesis

$$\mathrm{H}_0 : E\left[X_{ij}\right] = \mu_i \equiv E\left[Y_i\right] \qquad \text{for } i = 1,\ldots,k. \tag{4.61}$$

It follows that under the null hypothesis, Equation (4.61), the i^{th} performance criterion,

$$t_i = \frac{Y_i - \overline{X_i}}{\sqrt{\left(m+1\right)S_i^{\,2}/m}} \tag{4.62}$$

has Students-*t* distribution with $m-1$ degrees of freedom.

As an overall measure of the discrepancy between the with $k \times 1$ vector

$$\mathbf{Y} \equiv \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} \tag{4.63}$$

of responses from the real system and the $k \times 1$ vector

$$\overline{\mathbf{X}} \equiv \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_k \end{bmatrix} \tag{4.64}$$

of average responses taken over m independent simulation runs, we use the squared length of the $k \times 1$ vector

$$\mathbf{t} \equiv \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_k \end{bmatrix} \tag{4.65}$$

of standardized discrepancies between the real and simulated responses

$$D^2 = \left\| \mathbf{t}^2 \right\| = \sum_{i=1}^{k} t_i^{\,2} = \sum_{i=1}^{k} \left[\frac{\left(Y_i - \overline{X}_i \right)}{\sqrt{(m+1) S_i^{\,2} / m}} \right]^2 = \left(\frac{m}{m+1} \right) \sum_{i=1}^{k} \left[\frac{Y_i - \overline{X}_i}{S_i} \right]^2 . \tag{4.66}$$

In vector notation we can write this compactly in a form the closely resembles Hotelling's-T^2 statistic. Let

$$\begin{aligned} \mathbf{V} &= \left[(m+1)/m \right] \operatorname{diag}\left(S_1^{\,2}, S_2^{\,2}, \ldots, S_k^{\,2} \right) \\ &= \left[(m+1)/m \right] \begin{bmatrix} S_1^{\,2} & 0 & \cdots & 0 \\ 0 & S_2^{\,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k^{\,2} \end{bmatrix}; \end{aligned} \tag{4.67}$$

then we see that

$$D^2 = \left(\mathbf{Y} - \overline{\mathbf{X}} \right)^T \mathbf{V}^{-1} \left(\mathbf{Y} - \overline{\mathbf{X}} \right). \tag{4.68}$$

We propose choosing the simulation configuration that minimizes D^2. Then, as an overall validation of the selected simulation configuration, we compute the *p*-values associated with the revised t_i statistics in Equation (4.62), feed the resulting *p*-values into $\Phi^{-1}(\bullet)$, the inverse of the standard normal CDF $\Phi(\bullet)$ to obtain the corresponding *Z*-values, combine the *Z*-values in an overall Shapiro-Wilk test [80, 81, 82], and create the corresponding Q-Q plot. Regardless of the final *p*-value of the overall Shapiro-Wilk test on the *Z*-values

$$\left\{Z_i = \Phi^{-1}(p_i) : 1, \ldots k\right\}, \tag{4.69}$$

visual inspection of the associated Q-Q plot is the main tool for validation of the selected simulation model. In terms of Q-Q plots, the more the plotted points resemble a straight line the better. Although this is somewhat subjective, this method of evaluating Q-Q plots to validate simulations has been used in many application contexts [44, 46, 56], and we are confident we can select the simulation model that best matches historical data for the period 2001–2011 using this method.

At first thought, it may seem appropriate to combine the *p*-values using the Fisher chi-squared test, which involves taking the natural log of each *p*-value, adding them up, and multiplying by –2. This statistic is chi-squared distributed with $n-1$ degrees of freedom where n is the number of *t*-tests. However, this method is only valid if the *p*-values are independent. Performing the von Neumann randomness test [4, 53, 54, 90, 104] on the Z_i's from Equation (4.69), we found that for each scenario, this test was failed with a very small level of significance. Thus, using the Fisher chi-squared test is inappropriate for this particular situation where the *p*-values are highly correlated.

It should be noted that this method can easily be extended to situations where there are a different number *t*-tests for each scenario simply dividing the overall effectiveness measure D^2 by the number of *t*-tests.

$$D^2_{\text{ADJ}} = \frac{D^2}{k} = \frac{1}{k}\sum_{i=1}^{k} t_i^{\,2} \tag{4.70}$$

4.5.1.2 Validation Results

Table 4-18 presents the results from implementing our validation technique. For each of the 21 scenarios, we give the number of *t*-tests passed (of 34), the D^2 value, the number of practical significance tests passed (of 4), and the von Neumann randomness test *p*-value. We also give the Shapiro-Wilk *p*-value for the top four policies, and present the Q-Q plots for each of those scenarios.

Table 4-18. Results from Validation Technique

Interval/ Measure	*t*-tests Passed	D^2		Practical Tests (±10%)	VN *p*-value	SW *p*-value
Interval = 1	**# Passed (34)**	**Value**	**Rank**	**# Passed (4)**	***p*-Value**	***p*-Value**
70	32	44.102	3	2	7.699E-09	0.2174
75	34	26.679	1	3	1.576E-08	0.3030
80	34	36.013	2	4	1.144E-08	0.5141
85	34	47.411	4	4	1.077E-08	0.6491
90	29	87.011	18	4	6.653E-09	-
95	29	93.408	19	4	5.690E-09	-
100	26	118.876	21	4	4.250E-09	-
Interval = 2	**# Passed (34)**	**Value**	**Rank**	**# Passed (4)**	***p*-Value**	***p*-Value**
70	31	64.607	13	0	7.605E-09	-
75	31	58.995	10	1	9.842E-09	-
80	30	57.736	9	2	7.291E-09	-
85	32	52.924	6	3	9.160E-09	-
90	33	57.182	8	3	1.725E-08	-
95	30	78.783	16	3	7.763E-09	-
100	31	78.714	15	3	8.854E-09	-
Interval = 3	**# Passed (34)**	**Value**	**Rank**	**# Passed (4)**	***p*-Value**	***p*-Value**
70	26	95.797	20	0	4.850E-09	-
75	30	77.150	14	0	5.641E-09	-
80	31	56.252	7	0	7.302E-09	-
85	30	64.325	12	0	7.575E-09	-
90	30	60.638	11	0	8.777E-09	-
95	33	49.447	5	1	4.453E-08	-
100	29	85.227	17	1	8.619E-09	-

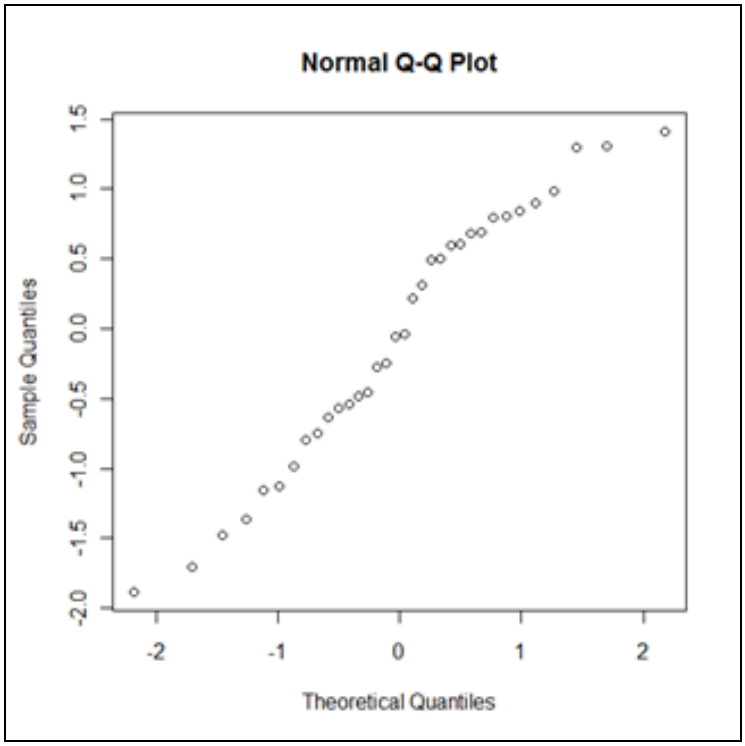


Figure 4-24. Q-Q Plot for Annual Screening with Stopping Age Mode of 70

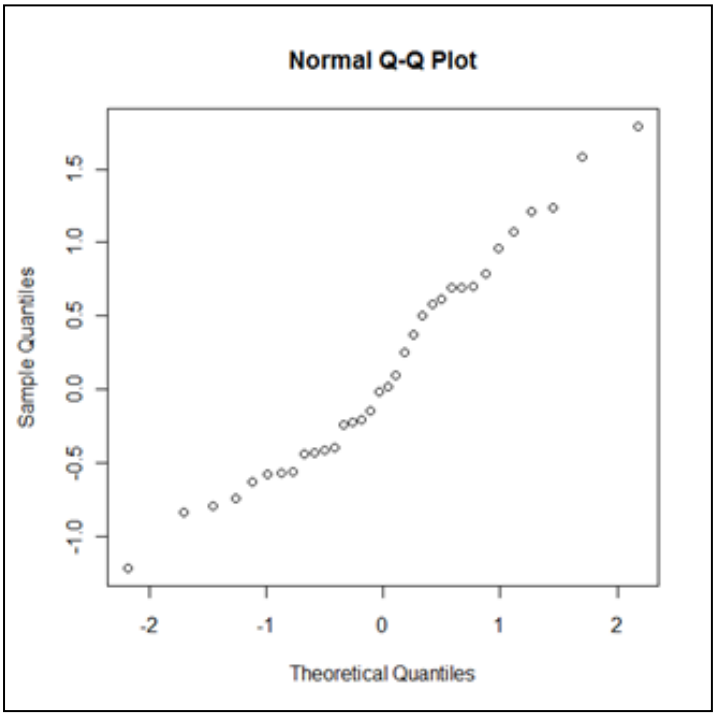


Figure 4-25. Q-Q Plot for Annual Screening with Stopping Age Mode of 75

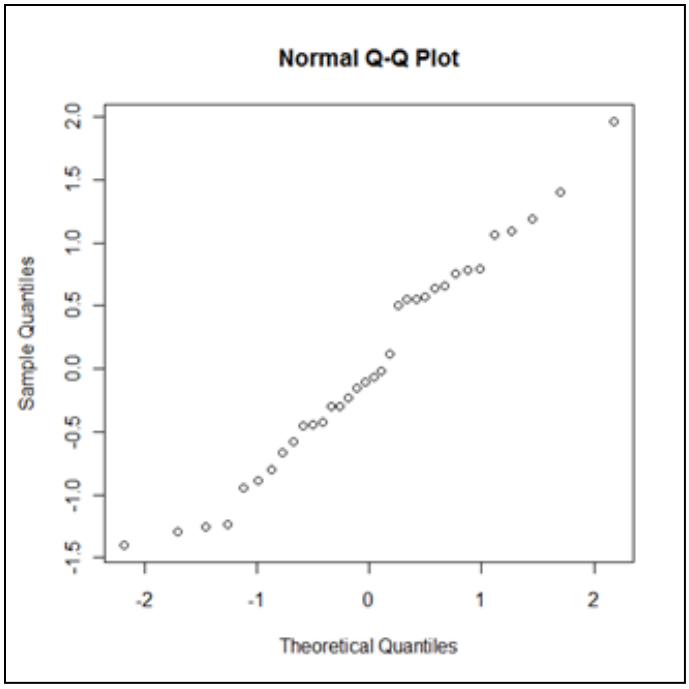


Figure 4-26. Q-Q Plot for Annual Screening with Stopping Age Mode of 80

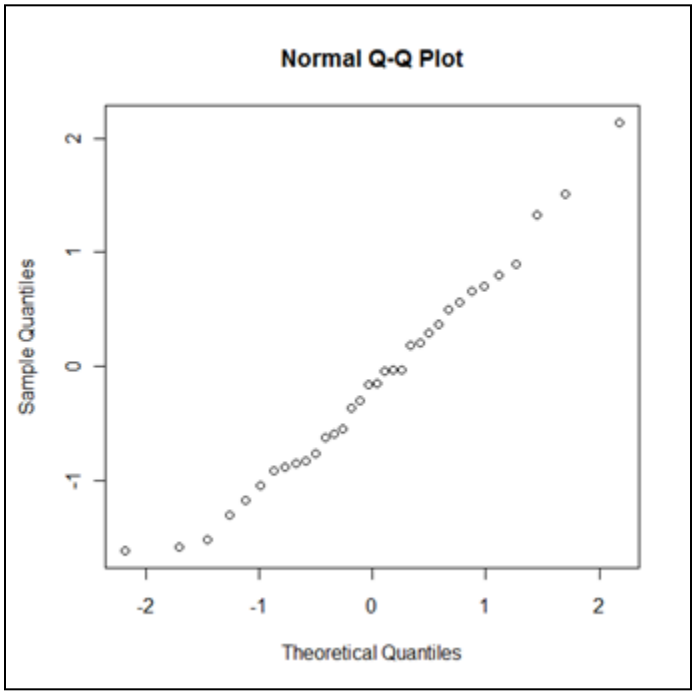


Figure 4-27. Q-Q Plot for Annual Screening with Stopping Age Mode of 85

Upon examination of the D^2 statistic and the associated Q-Q plots, we believe there is clear evidence that annual screening with a beta distributed stopping age with a mode of 75 matches the data the best. It ranks first in the D^2 statistic, and the associated Q-Q plot appears to be the closest to resembling a straight line. This policy also has a Shaprio-Wilk *p*-value that is not significant at the 0.05 level. Thus, we select that policy for use in the periods 2001–2011. Our breast cancer experts expected annual screening to match the data best, so this was expected to some degree. In fact, it seems that multiple screening policies could have been deemed "valid" representations of the past, and we feel this is further evidence of the validity of the model. Since none of these policies was actually in place, it makes sense that more than one policy may match the data reasonably well for a model which is accurately capturing the behavior of the actual system. With a valid screening policy in place for the past, we turn our attention to the results of experimenting with alternative screening policies for future periods.

4.5.2 Results and Discussion

Each time the model is executed, a plethora of results are generated. A full set of results includes several tables with yearly values for each SD level and SD performance measure, a table with yearly averages for individual characteristics of interest, the percentage of women treated, a table of yearly cancer incidence and death rates, graphs describing the stage distribution at diagnosis, graphs of cost and QALYs saved information by category, and graphs with information about methods of detection. Because the full set of results includes a large number of performance measures, we only present the full set of results for the screening policy which saves the most lives for the period 2012–2020. For other policies which are close (or not statistically different in terms of some selected performance measures), we provide the key performance measures only.

Three different optimization problems were considered, and to each of these problems we applied policies based on interval screening, risk-based screening, and factor-based

screening separately. The results across the different policy types were combined in order to determine the best policy or group of policies for each of the three optimization problems. By an optimization problem, we mean the specification of a performance measure to be optimized as well as the specification of constraints on input variables or on other performance measures (outputs). Before we formulate the three optimization problems, we provide a list of key performance measures.

4.5.2.1 Key Performance Measures

There are several performance measures of interest when considering the impact of alternative breast cancer screening policies on the designated population as a whole and on the individuals composing that population. The costs of the screening and treatment processes and the years of life saved are important performance measures for a screening policy, and these outputs can be broken down into categories for even more insight. Ultimately, the goal is to minimize breast cancer deaths, so another key performance measure is the number of deaths caused by breast cancer. The following list describes the key performance measures that are generated by the simulation and how each are categorized.

1. Total Cost
    a. By Past or Future Time Period (2001–2011, 2012–2020)
    b. By Detection Type (Clinical, Screening)
    c. By Procedure Type (Screening, Diagnostic, Biopsy, Treatment)
2. Cost/Life-Year Saved
    a. By Past or Future Time Period (2001–2011, 2012–2020)
    b. By Detection Type (Clinical, Screening)
3. Life-Years Saved
    a. By Past or Future Time Period (2001–2011, 2012–2020)
    b. By Detection Type (Clinical, Screening)
4. Cost/QALY Saved
    a. By Past or Future Time Period (2001–2011, 2012–2020)

    b. By Detection Type (Clinical, Screening)
5. QALYs Saved
    a. By Past or Future Time Period (2001–2011, 2012–2020)
    b. By Detection Type (Clinical, Screening)
6. Cost of False Positive Exams and Benign Biopsies
    a. By Past or Future Time Period (2001–2011, 2012–2020)
    b. By Procedure Type (Screening, Diagnostic, Benign Biopsy)
7. Number of False Negative Exams Where Detection was Possible
    a. By Past or Future Time Period (2001–2011, 2012–2020)
8. Number of Cancer Deaths
    a. By Past or Future Time Period (2001–2011, 2012–2020)
9. Stage of Cancer at Diagnosis (Local, Regional, Distant)
    a. By Past or Future Time Period (2001–2011, 2012–2020)
10. Type of Detection Percentages (Screening and Clinical)
    a. By Past or Future Time Period (2001–2011, 2012–2020)
11. Benign Biopsy Percentage
    a. By Past or Future Time Period (2001–2011, 2012–2020)

With the help of our breast cancer experts, we choose five of these performance measures as the most important subset, and we ranked each policy in terms of these five performance measures (with rank one denoting the best policy). The five most important performance measures are:

1. Number of breast cancer deaths during and after the year 2012;
2. Number of QALYs saved by screening during the period 2012–2020;
3. Percentage of cancers diagnosed in the distant stage during the period 2012–2020;
4. Cost/QALY saved by screening during the period 2012–2020;
5. The total cost of false positive exams and benign biopsies during the period 2012–2020.

Preventing deaths from breast cancer is the real objective of screening, and that could be considered the most important performance measure. Similarly, the number of QALYs saved is a measure of the amount of life saved and the quality of that life over the entire population, and this is another important performance measure. The cost-effectiveness of screening is a measure of how much we are paying to save one quality-adjusted year of a given woman's life. The percentage of cancers diagnosed in the distant stage is important because women diagnosed in the distant stage are not likely to live long and their quality of life will be poor. Cost per QALY saved by screening is something that should be considered, but we argue it is not as important as lives saved; and so long as the policy is reasonably cost-effective, the main objective should be to maximize the lives saved by the policy. The total cost of false positives is important, but not as important as lives saved or overall cost-effectiveness. The monetary cost of these exams to the population alone can be seen as a problem, but we must also consider the nonmonetary costs for individual women who endure substantial emotional stress because of a false diagnosis of breast cancer. In addition, the more incisions to a given woman's breasts, the more difficult is it is to identify breast cancers in the future (we did not account for this in our model); and a woman's future adherence to a screening protocol may be adversely affected because of a prior experience where the outcome was a false positive result. For these reasons, we included the total cost of false positive exams and benign biopsies in the top five performance measures; however we argue that this cost is the least important of the main performance measures.

4.5.2.2 Factorial Experiment Enumeration and Optimization Routines

The following three optimization problems were run for all three different types of screening:

1. Objective: Minimize cost/QALY saved for breast cancers detected by mammography during the period 2012–2020.
   Constraints:
   a. All screening intervals (including low and high risk specific intervals) are restricted to 1, 2, 3, 4, or 5 years.
   b. The stopping age is restricted to 70, 75, 80, 85, 90, 95, or 100.
2. Objective: Minimize breast cancer deaths during the period 2012–2020.
   Constraints:
   a. Total cost/QALY saved during the period 2012–2020 cannot exceed $50,000 (~GDP Per Capita).
   b. All screening intervals (including low-and high-risk specific intervals) are restricted to 1, 2, 3, 4, or 5 years.
   c. The stopping age is restricted to 70, 75, 80, 85, 90, 95, or 100.
3. Objective: Maximize QALYs saved during the period 2012–2020.
   Constraints:
   a. Total cost/QALY saved during the period 2012–2020 cannot exceed $50,000 (~GDP Per Capita).
   b. All screening intervals (including low-and high-risk specific intervals) are restricted to 1, 2, 3, 4, or 5 years.
   c. The stopping age is restricted to 70, 75, 80, 85, 90, 95, or 100.

It should be noted that for interval and risk-based screening, we were able to enumerate all possible screening policies and thus conduct a full designed experiment with all input variables at all values for all three optimization problems. However, the number of possible screening policies for factor based screening is over 16 million, and we could not

possibly test all possibilities. Thus, we allowed the optimization to run until 5,000 screening policies were tested for each of the three optimizations. While these optimizations are rather simple, extremely complex optimizations could be and may be performed as future work.

After examining the results, we concluded that factor-based screening was unable to compete with risk-based and interval screening in terms of the key performance measures (most likely because we did not find the best group of policies amongst the first 5,000); therefore we focus on the results of risk-based and interval screening. However, the ability to produce results for factor-based screening policies is a key feature of this model and should not be overlooked. The top five policies in terms of the objectives were ranked and placed into tables in order from best to worst. Upon examining the results, we found that optimizations 2 and 3 produced the same set of 5 best policies; and therefore we were left with two sets of 5 best policies—one set for optimization 1, and one set for optimizations 2 and 3.

After reviewing the full sets of results for the best policies from optimization 1 with our breast cancer experts, it quickly became clear that simply minimizing the cost per QALYs saved has detrimental effects on the other performance measures; and our experts stated that those effects could not be tolerated. For example, there was a 15% increase in women diagnosed in the distant stage of breast cancer. This type of shift in the stage distribution was deemed unacceptable. However for the best policy according to optimization 1, the cost-effectiveness of that policy was approximately $29,000 per QALY saved; and this is well below the threshold (GDP per capita) of $50,000 for a cost-effective screening policy. Thus, we decided that a more appropriate method for determining the best screening policy for older US women would be based on optimization problems 2 or 3, wherein an upper bound of $50,000 is imposed on the cost/QALY saved while respectively minimizing deaths due to breast cancer or maximizing QALYs saved.

There are trade-offs between these five main performance measures enumerated on p. 208, and no policy will be best in terms of all five measures. After reviewing the key performance measures for the best policies from optimization 3, we determined that only the top two policies merited comparison on a statistical basis. From a practical standpoint, the best policy was superior to the second best in terms of three of the five key performance measures; and, the other two performance measures were the less-important outputs of cost-effectiveness and total cost of false positives. It is important to remember that performance measures that are not rates or percentages, like the number of breast cancer deaths or the number of QALYs saved, are for 0.1% of the population; therefore these results should be multiplied by 1,000 in order to scale up to the entire population of women 65 and older. Paired Student *t*-tests are used to identify the policy that is statistically superior to the others in terms of the most important performance criterion, and we present the full results for our recommended screening policy in Section 4.5.4. We now present tables of key performance measure values for the top five policies from optimizations 1, 2, and 3, along with a discussion of the results and our recommended screening policy based on the results from optimizations 2 and 3. Note that in this dissertation, we only present the mean values of these performance measures over ten randomly sampled populations. However, the means and CI half-lengths for all performance measures can be obtained by contacting the author.

4.5.2.3 Optimization Routine 1 Results

The best five screening policies in terms of all five performance measures from optimization 1 are:

1. Annual screening for the top 5% in terms of risk, every four years for everyone else, and screening stops at age 70;
2. Annual screening for the top 10% in terms of risk, every two years for everyone else, and screening stops at age 70;
3. Annual screening stops at age 70;
4. Annual screening stops at age 75; and

5. Annual screening stops at age 80.

The following series of tables are used to present full sets of results (mean values) for all performance measures of interest.

Table 4-19. Optimization 1 Costs by Category for the Period 2012–2020

Policy	Total Cost 2012-2020	Screening Costs 2012-2020	Diagnostic Costs 2012-2020	Work-Up Costs 2012-2020	Treatment Costs 2012-2020	Screening Detection Costs 2012-2020	Clinical Detection Costs 2012-2020
1	$64,625,871	$11,093,776	$1,140,998	$7,097,373	$45,293,725	$40,655,903	$23,969,968
2	$73,456,854	$13,810,013	$1,380,993	$8,439,582	$49,826,266	$54,254,288	$19,202,567
3	$60,168,015	$7,937,157	$872,744	$4,785,568	$46,572,547	$36,021,188	$24,146,827
4	$60,604,416	$8,214,607	$899,738	$4,965,345	$46,524,725	$36,560,556	$24,043,859
5	$58,042,040	$7,212,167	$820,324	$4,342,519	$45,667,030	$32,828,957	$25,213,082

Table 4-20. Optimization 1 Life-Years Saved and Cost Per Life-Year Saved by Category for the Period 2012–2020

Policy	Life-Years Saved 2012-2020	Life-Years Saved Screening 2012-2020	Life-Years Saved Clinical 2012-2020	Cost Per Life-Year Saved 2012-2020	Cost Per Life-Year Saved Screening 2012-2020	Cost Per Life-Year Saved Clinical 2012-2020
1	3100.9	1984.0	1116.9	$20,875	$22,330.23	$21,623
2	3232.3	2262.3	970.0	$22,773	$24,752.90	$20,019
3	2314.1	479.9	1834.1	$18,112	$15,805.43	$18,794
4	2486.6	874.2	1612.4	$18,537	$16,224.00	$19,886
5	2801.8	1392.5	1409.3	$19,082	$17,604.59	$20,723

Table 4-21. Optimization 1 QALYs Saved and Cost Per QALY Saved by Category for the Period 2012–2020

Policy	QALYs Saved 2012-2020	QALYs Saved Screening 2012-2020	QALYs Saved Clinical 2012-2020	Cost Per QALY Saved 2012-2020	Cost Per QALY Saved Screening 2012-2020	Cost Per QALY Saved Clinical 2012-2020
1	1695.3	1088.8	606.6	$38,184	$37,410	$39,826
2	1765.9	1238.2	527.7	$41,687	$43,908	$36,795
3	1260.0	264.9	995.1	$33,263	$28,665	$34,642
4	1356.4	483.1	873.3	$33,983	$29,230	$36,715
5	1531.1	768.4	762.7	$34,922	$31,974	$ 38,303

Table 4-22. Optimization 1 False Positive Costs by Category for the Period 2012–2020

Policy	False Positive Costs 2012-2020	False Positive Screening Costs 2012-2020	False Positive Diagnostic Costs 2012-2020	Benign Work-Up Costs 2012-2020	False Negative Screening Exams 2012-2020
1	$8,236,844	$1,076,982	$1,052,020	$6,107,842	360
2	$9,947,794	$1,316,920	$1,280,573	$7,350,301	499
3	$1,301,444	$197,719	$193,360	$910,365	67
4	$2,691,079	$402,405	$391,788	$1,896,887	122
5	$5,457,013	$701,791	$687,035	$4,068,187	199

Table 4-23. Optimization 1 Cancer Deaths, Stage Distribution, and Method of Detection for the Period 2012–2020

Policy	Cancer Deaths 2012-2065	Stage At Diagnosis Local % 2012-2020	Stage At Diagnosis Regional % 2012-2020	Stage At Diagnosis Distant % 2012-2020	Screening Detection % 2012-2020	Clinical Detection % 2012-2020	Benign Biopsy % 2012-2020
1	522.9	34.7	39.3	26.0	44.1	55.9	86.3
2	513.6	43.2	35.6	21.2	58.2	41.8	87.3
3	538.4	18.0	46.4	35.6	10.0	90.0	51.8
4	537.4	21.6	43.8	34.6	17.1	82.9	68.7
5	532.1	25.5	43.5	31.0	26.3	73.7	82.1

As previously stated, we choose five of these measures as the most important and then ranked each of these five policies according to each of the five performance measures. Table 4-24 and Table 4-25 present these measures and their relative ranks. These tables allow one to easily see which policies are best for certain measures, and to make a decision about which policy should be recommended.

Upon reviewing these tables, we can immediately see that the risk-based policies and screening policies with a stopping age of 70 which are best in terms of cost-effectiveness and the cost of false positives are the worst policies in terms of saving life. The magnitude of the shifts in the performance measures that quantify the saving of life are much too large to consider using one of the most cost-effective policies. We now move to the results from the best group of policies identified by optimizations 2 and 3, and we note that two of the

policies that were the best in terms of all five measures for optimization 1 (policies 4 and 5) are also in the set of the best of policies for optimizations 2 and 3.

Table 4-24. Top 5 Policies Ranked in Terms of Top 5 Performance Measures for Optimization 1 Part I

Policy	Cancer Deaths 2012-2065 Rank	Cancer Deaths 2012-2065	QALY Saved Screening 2012-2020 Rank	QALY Saved Screening 2012-2020	% Distant Stage 2012-2020 Rank	% Distant Stage 2012-2020
1	5	538.4	5	264.9	5	35.6
2	4	537.4	4	483.1	4	34.6
3	3	532.1	3	768.4	3	31
4	2	522.9	2	1088.8	2	26
5	1	513.6	1	1238.2	1	21.2

Table 4-25. Top 5 Policies Ranked in Terms of Top 5 Performance Measures for Optimization 1 Part II

Policy	Cost/QALY Saved Screening 2012-2020 Rank	Cost/QALY Saved Screening 2012-2020	Cost of False Positives 2012-2020 Rank	Cost of False Positives 2012-2020	Sum of Ranks
1	1	$28,665	1	$1,301,443	17
2	2	$29,230	2	$2,691,079	16
3	3	$31,974	3	$5,457,012	15
4	4	$37,410	4	$8,236,843	14
5	5	$43,908	5	$9,947,793	13

4.5.2.4 Optimization Routine 2 and 3 Results

The best five screening policies in terms of all five performance measures from optimizations 2 and 3 are:

1. Annual screening stops at age 80;
2. Annual screening stops at age 75;
3. Annual screening for top 10% in terms of risk, every two years for everyone else, and screening stops at age 80;

4. Annual screening for top 5% in terms of risk, every two years for everyone else, and screening stops at age 80; and
5. Biennial screening stops at age 80.

The following series of tables are used to present full sets of results (mean values) for all performance measures of interest.

Table 4-26. Optimization 2 Costs by Category for the Period 2012–2020

Policy	Total Cost 2012-2020	Screening Costs 2012-2020	Diagnostic Costs 2012-2020	Work-Up Costs 2012-2020	Treatment Costs 2012-2020	Screening Detection Costs 2012-2020	Clinical Detection Costs 2012-2020
1	$64,625,871	$11,093,776	$1,140,998	$7,097,373	$45,293,725	$40,655,903	$23,969,968
2	$73,456,854	$13,810,013	$1,380,993	$8,439,582	$49,826,266	$54,254,288	$19,202,567
3	$41,719,686	$1,821,079	$261,396	$1,749,956	$37,887,255	$7,475,974	$34,243,712
4	$45,985,813	$3,911,390	$466,534	$2,763,279	$38,844,611	$14,099,359	$31,886,454
5	$53,339,837	$7,069,421	$765,449	$4,962,625	$40,542,342	$24,310,164	$29,029,673

Table 4-27. Optimization 2 Life-Years Saved and Cost Per Life-Year Saved by Category for the Period 2012–2020

Policy	Life-Years Saved 2012-2020	Life-Years Saved Screening 2012-2020	Life-Years Saved Clinical 2012-2020	Cost Per Life-Year Saved 2012-2020	Cost Per Life-Year Saved Screening 2012-2020	Cost Per Life-Year Saved Clinical 2012-2020
1	3100.9	1984.0	1116.9	$20,875	$ 24,029	$ 21,623
2	3232.3	2262.3	970.0	$22,773	$ 22,724	$ 20,019
3	2314.0	479.9	1834.1	$18,112	$ 17,643	$ 18,793
4	2486.6	874.2	1612.4	$18,536	$ 23,400	$ 19,885
5	2801.8	1392.5	1409.3	$19,081	$ 21,173	$ 20,723

Table 4-28. Optimization 2 QALYs Saved and Cost Per QALY Saved by Category for the Period 2012–2020

Policy	QALYs Saved 2012-2020	QALYs Saved Screening 2012-2020	QALYs Saved Clinical 2012-2020	Cost Per QALY Saved 2012-2020	Cost Per QALY Saved Screening 2012-2020	Cost Per QALY Saved Clinical 2012-2020
1	1765.9	1238.2	527.7	$41,687	$43,907	$36,795
2	1695.3	1088.8	606.6	$38,183	$37,409	$39,825
3	1517.1	814.9	702.2	$40,028	$45,086	$34,337
4	1535.1	803.5	731.7	$39,323	$45,053	$ 33,170
5	1538.8	773.4	765.3	$37,892	$42,843	$ 33,160

Table 4-29. Optimization 2 False Positive Costs by Category for the Period 2012–2020

Policy	False Positive Costs 2012-2020	False Positive Screening Costs 2012-2020	False Positive Diagnostic Costs 2012-2020	Benign Work-Up Costs 2012-2020	False Negative Screening Exams 2012-2020
1	$9,947,793	$1,316,919	$1,280,573	$7,350,301	499
2	$8,236,843	$1,076,982	$1,052,020	$6,107,841	360
3	$5,581,041	$827,949	$808,131	$3,944,960	340
4	$5,361,173	$804,058	$782,947	$3,774,166	320
5	$4,822,807	$747,013	$730,608	$3,345,185	283

Table 4-30. Cancer Deaths, Stage Distribution, and Method of Detection for the Period 2012–2020

Policy	Cancer Deaths 2012-2065	Stage At Diagnosis Local % 2012-2020	Stage At Diagnosis Regional % 2012-2020	Stage At Diagnosis Distant % 2012-2020	Screening Detection % 2012-2020	Clinical Detection % 2012-2020	Benign Biopsy % 2012-2020
1	513.6	43.2	35.6	21.2	58.2	41.8	87.3
2	522.9	34.7	39.3	26	44.1	55.9	86.3
3	525.7	36.7	38.2	25.1	46.2	53.8	79.6
4	527.5	36.1	38.5	25.4	45.7	54.3	79
5	531.8	33.5	40	26.4	42.5	57.5	77.2

Again we ranked each of these five policies according each of the five performance measures. Table 4-31 and Table 4-32 present these measures and their relative ranks. Upon reviewing these tables, we can see that in terms of cancer deaths, QALYs saved, and percentage of cancers diagnosed in the distant stage, the best screening policy is annual

screening with a stopping age of 80. Although this policy ranks third in cost-effectiveness and last in cost of false positives, these are the less important measures; and from a practical standpoint we do not feel the negative impact on life saved cause by using annual screening stopping at age 75 is justified by the increase in cost-effectiveness or by the decrease in cost of false positives. For example, on average there are 9.3 fewer cancer deaths for annual screening stopping at age 80 as compared with stopping age 75; however, we must remember that since we are only simulating 0.1% of the population, this is actually 9,300 fewer cancer deaths, which we consider practically significant.

Table 4-31. Top 5 Policies Ranked in Terms of Top 5 Performance Measures for Optimization 2 Part I

Policy	Cancer Deaths 2012-2065 Rank	Cancer Deaths 2012-2065	QALY Saved Screening 2012-2020 Rank	QALY Saved Screening 2012-2020	% Distant Stage 2012-2020 Rank	% Distant Stage 2012-2020
1	1	513.6	1	1238.1	1	21.2
2	2	522.9	2	1088.8	4	26
3	3	525.7	3	814.9	2	25.1
4	4	527.5	4	803.5	3	25.4
5	5	531.8	5	773.4	5	26.4

Table 4-32. Top 5 Policies Ranked in Terms of Top 5 Performance Measures for Optimization 2 Part II

Policy	Cost/QALY Saved Screening 2012-2020 Rank	Cost/QALY Saved Screening 2012-2020	Cost of False Positives 2012-2020 Rank	Cost of False Positives 2012-2020	Sum of Ranks
1	3	$43,908	5	$9,947,794	11
2	1	$37,410	4	$8,236,844	13
3	5	$45,086	3	$5,581,041	16
4	4	$45,053	2	$5,361,173	17
5	2	$42,843	1	$4,822,807	18

After selecting annual screening with a stopping age of 80 as the potential best policy, we performed paired-comparison Student *t*-tests using information from each of the ten randomly sampled populations, and across each of the five major performance measures. Table 4-33 shows the differences across populations for all performance measures, and presents the results of the paired *t*-tests.

Table 4-33. Results from Paired *t*-tests on Two Best Policies

Differences Between Annual Screening Stopping at 80 vs. Annual Screening Stopping at 75					
Population	**BC Deaths**	**QALY Saved**	**Cost/ QALY Saved**	**% Distant**	**False Positive Costs**
1	10.0	-36.1	$12,010	-6.7	$1,761,258
2	-5.0	81.3	$5,919	-3.7	$1,663,328
3	7.0	52.0	$6,902	-3.4	$1,869,071
4	-49.0	265.6	$3,517	-6.6	$1,785,903
5	-33.0	106.3	$3,219	-6.6	$1,789,784
6	8.0	98.8	$4,295	-4.3	$1,695,975
7	-8.0	-21.6	$7,574	-3.5	$1,752,920
8	-18.0	44.1	$9,377	-3.4	$1,376,040
9	-11.0	133.9	$5,846	-4.3	$1,728,794
10	6.0	-18.9	$6,322	-5.1	$ 1,686,427
Mean Difference	-9.3	70.5	$6,498	-4.8	$1,710,950
Standard Deviation	19.5	90.2	$2,693	1.4	$131,953
95% CI Half-Width	13.9	64.5	$1,927	1.0	$94,393
95% CI Lower Limit	-23.2	6.1	$4,571	-5.7	$1,616,557
95% CI Upper Limit	4.6	135.0	$8,425	-3.8	$1,805,343
Percentage Difference	1.8%	4.0%	14.8%	22.5%	17.2%
Statistical Difference?	**No**	**Yes**	**Yes**	**Yes**	**Yes**
***p*-value**	**0.165**	**0.035**	**3.225E-05**	**1.682E-06**	**1.521E-11**

There was not a statistically significant difference in the number of breast cancer deaths, but there was a statistically significant different for all other performance measures. For QALYs saved and the percentage diagnosed in the distant stage, annual screening stopping at age 80 is statistically better. However, for cost-effectiveness and the cost of false positives, annual screening stopping at age 75 is statistically better. As previously discussed,

we put more weight on saving lives than on cost-effectiveness and false positives; and in terms of saving lives, we can declare annual screening stopping at age 80 the statistically best policy, and the cost-effectiveness of this policy is under the cost-effectiveness threshold. The only characteristics that could be perceived as negative is the high cost and number of false positive exams and benign biopsies. However, there would not be a significant increase in these parameters over what they were from 2001–2011; and for that period we tolerated a similar number of false positives. Decreasing false positives comes at the cost of saving lives, which we argue is not worth the trade-off.

4.5.3 Full Set of Results for Statistically Best Screening Policy

We now present a full set of results for annual screening stopping at age 80, our recommended screening policy. For most measures we included the results for 2001–2011 under the calibrated policy, the results for 2012–2020 under the calibrated policy (if we continued as we did in the past), and the results for 2012–2020 under our recommended best policy. The results are divided up into five sections: population demographics, average values of individual characteristics, SD levels and adherence results, cancer incidence, death, diagnosis, cost, and life-years.

4.5.3.1 Population Demographics

Although we sampled from the BCSC database, which we claim is an accurate cross-section of the US population, information about the sampled populations is still of interest. The following four figures show the distribution of race, ethnicity, body-mass index, and breast density.

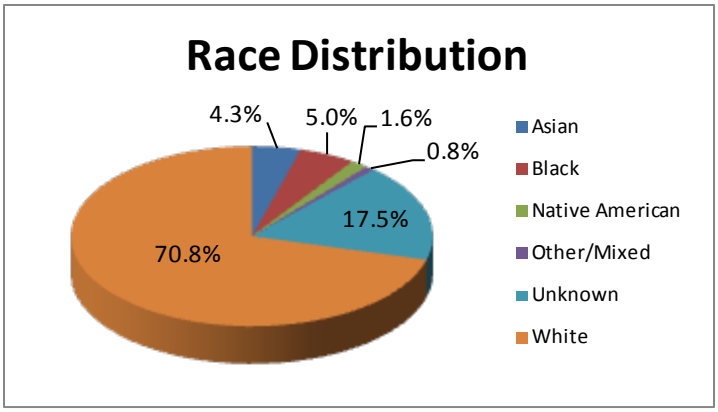


Figure 4-28. Distribution of Race for the 10 Sampled Populations

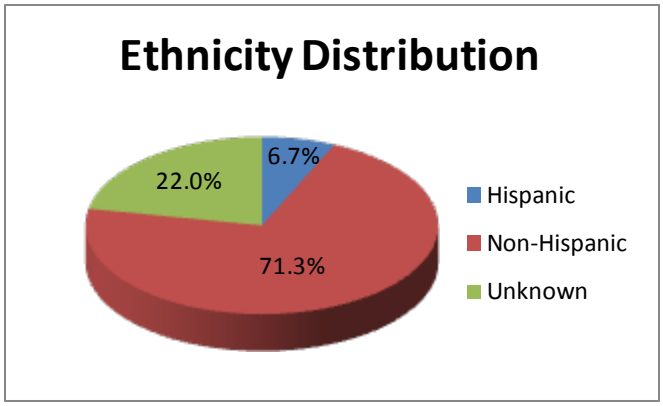


Figure 4-29. Distribution of Ethnicity for the 10 Sampled Populations

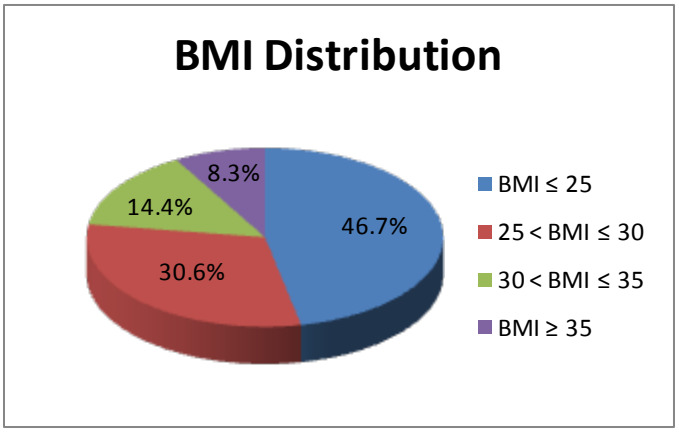


Figure 4-30. Distribution of Body-Mass Index for the 10 Sampled Populations

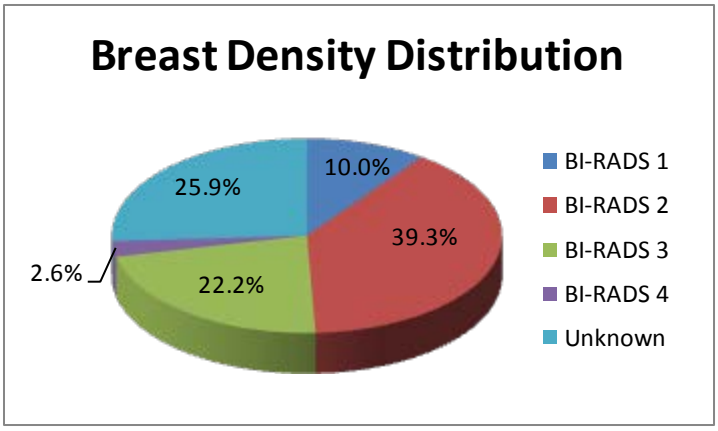


Figure 4-31. Distribution of Breast Density for the 10 Sampled Populations

4.5.3.2 Average Values of Individual Characteristics

There are several individual characteristics that change for each woman over the course of the time horizon and whose average values across the population may be of interest, including: average age, percentage of women with comorbid conditions, satisfaction with the mammography experience, the number of perceived barriers to screening, the intention to adhere to screening, and the number of times adherent out of the last two opportunities. Many of these characteristics affect adherence in the SD submodel, so their average values across the population may be useful to refer back to when examining SD and adherence results. Table 4-34 shows the average values for each year in the model time-horizon. Figure 4-32 and Figure 4-33 are plots of the average values in Table 4-34.

Table 4-34. Average Values of Individual Characteristics by Year for Statistically Best Policy

Year	Averages of Individual Characteristics Over All Individuals						
	Average Age	Comorbidity Percentage	Satisfaction with Experience	# of Perceived Barriers	Intention To Adhere	Last Two Adherence	Population Size
2001	75.87	29.91	0.889	0.680	0.850	1.500	20582.0
2002	75.98	30.41	0.828	0.669	0.838	1.532	20692.5
2003	76.04	30.87	0.810	0.644	0.826	1.560	20844.9
2004	76.06	31.16	0.801	0.617	0.822	1.558	21031.7
2005	76.05	31.45	0.795	0.592	0.819	1.557	21272.8
2006	76.02	31.66	0.790	0.569	0.817	1.558	21550.2
2007	75.98	31.83	0.785	0.545	0.814	1.560	21835.8
2008	75.89	31.89	0.780	0.525	0.813	1.561	22176.3
2009	75.78	31.91	0.775	0.506	0.811	1.562	22575.2
2010	75.66	32.02	0.772	0.489	0.809	1.563	23016.0
2011	75.55	31.90	0.766	0.475	0.807	1.563	23492.4
2012	75.43	31.88	0.761	0.464	0.804	1.563	24009.6
2013	75.30	31.86	0.753	0.464	0.797	1.560	24573.4
2014	75.16	31.74	0.746	0.460	0.794	1.559	25205.2
2015	75.03	31.68	0.739	0.462	0.788	1.557	25890.4
2016	74.92	31.61	0.734	0.454	0.787	1.555	26606.6
2017	74.79	31.45	0.733	0.451	0.786	1.555	27416.1
2018	74.66	31.36	0.732	0.446	0.784	1.555	28291.8
2019	74.55	31.28	0.729	0.444	0.784	1.555	29252.3
2020	74.44	31.16	0.727	0.445	0.781	1.552	30310.0

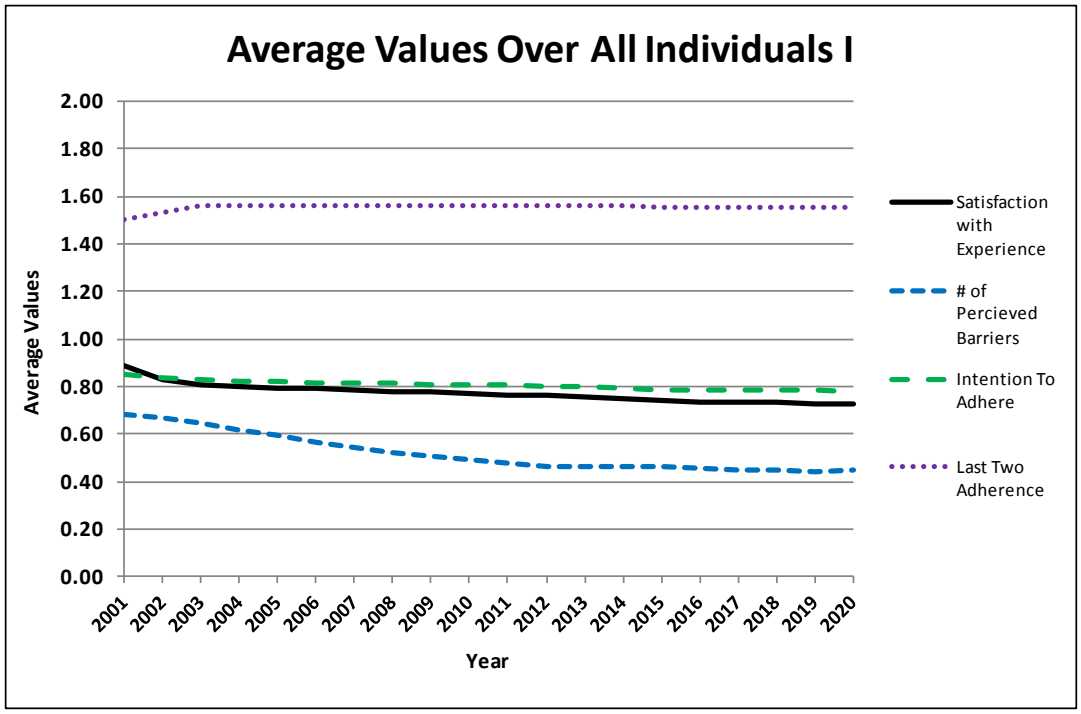


Figure 4-32. Averages of Individual Characteristics by Year I for Statistically Best Policy

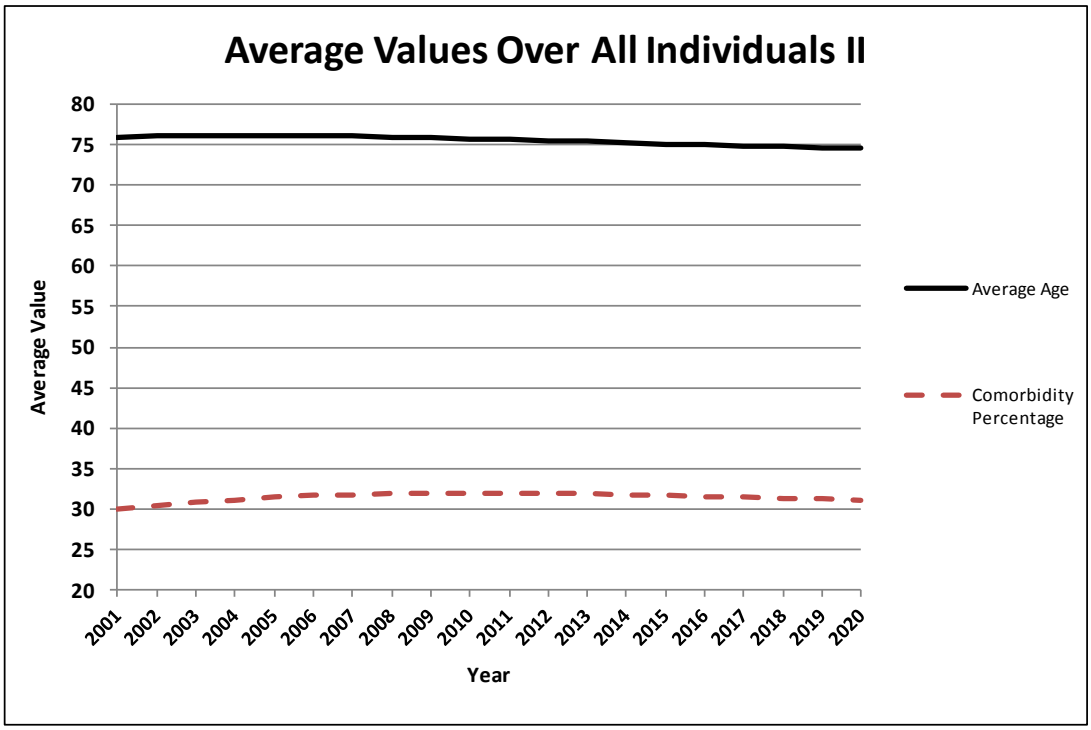


Figure 4-33. Averages of Individual Characteristics by Year II for Statistically Best Policy

The results from Figure 4-32 show that average values for the intention to adhere and the number of screening appoints attended of the last two opportunities across the population are relatively constant over time. One thing to note about Figure 4-33 is the decrease in average age over time. More and more women are turning 65 and coming into this population, and thus the number of younger women will tend to exceed the number of older women in the population as time goes on. This also means that a greater percentage of women will be eligible for screening as time goes on, and as a result the demand for screening will increase as more women will be younger than the stopping age for screening.

4.5.3.3 SD Levels and Adherence Results

The values of the levels from the SD submodel are important results from this research. These results are typical of SD type simulation models, plots of the levels of interest over the given time-horizon. We present tables and graphs for the SD input levels, intermediate SD levels, primary SD levels, and SD outputs related to adherence to the screening policy.

Table 4-35. SD Input Levels by Year for Statistically Best Policy

Year	SD Input Levels			
	# of Facilities	Capacity of Facilities	Breast Cancer Research	Breast Cancer Ads
2001	10125	6.125	2	2
2002	10000	6.250	2	2
2003	9850	6.375	2	2
2004	9700	6.500	2	2
2005	9550	6.625	2	2
2006	9400	6.750	2	2
2007	9250	6.875	2	2
2008	9100	7.000	2	2
2009	8950	7.125	2	2
2010	8800	7.250	2	2
2011	8650	7.375	2	2
2012	8500	7.500	2	2
2013	8500	7.600	2	2
2014	8500	7.700	2	2
2015	8500	7.800	2	2
2016	8500	7.900	2	2
2017	8500	8.000	2	2
2018	8500	8.100	2	2
2019	8500	8.200	2	2
2020	8500	8.300	2	2

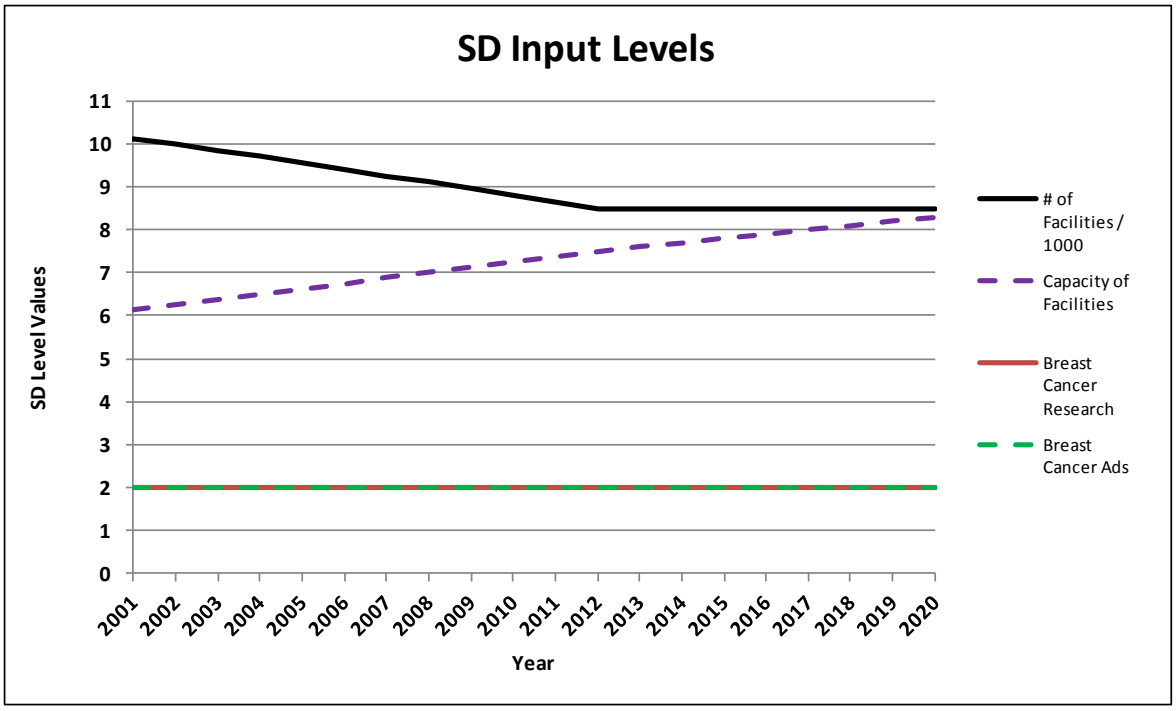


Figure 4-34. Plots of SD Input Levels for Statistically Best Policy

Table 4-36. Intermediate SD Levels by Year for Statistically Best Policy

Year	Intermediate SD Levels				
	Distance To Facilities	Congestion	Technology	Time to Get Results	Appointment Availability
2001	11.10	9.47	0.40	15.421	0.905
2002	11.50	9.82	0.42	15.174	0.902
2003	12.01	9.36	0.44	14.805	0.906
2004	12.55	9.08	0.46	14.462	0.909
2005	13.12	8.93	0.48	14.140	0.911
2006	13.73	8.88	0.50	13.832	0.911
2007	14.37	8.93	0.52	13.539	0.911
2008	15.05	8.80	0.54	13.220	0.912
2009	15.77	8.88	0.56	12.932	0.911
2010	16.53	9.07	0.58	12.660	0.909
2011	17.33	9.46	0.60	12.420	0.905
2012	18.16	9.83	0.62	12.174	0.902
2013	18.16	16.04	0.64	12.806	0.840
2014	18.16	16.38	0.66	12.557	0.836
2015	18.16	16.92	0.68	12.337	0.831
2016	18.16	17.74	0.70	12.161	0.823
2017	18.16	18.98	0.72	12.047	0.810
2018	18.16	20.59	0.74	11.989	0.794
2019	18.16	22.53	0.76	11.980	0.775
2020	18.16	24.73	0.78	12.009	0.753

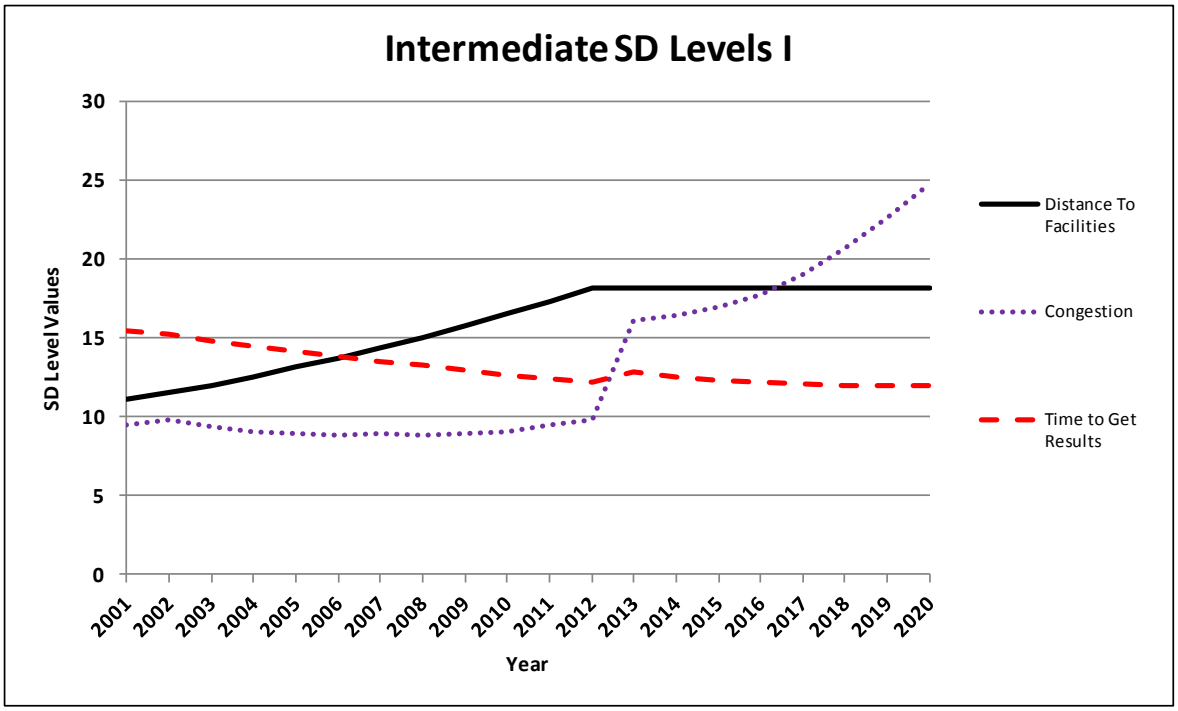


Figure 4-35. Plots of Intermediate SD Levels I for Statistically Best Policy

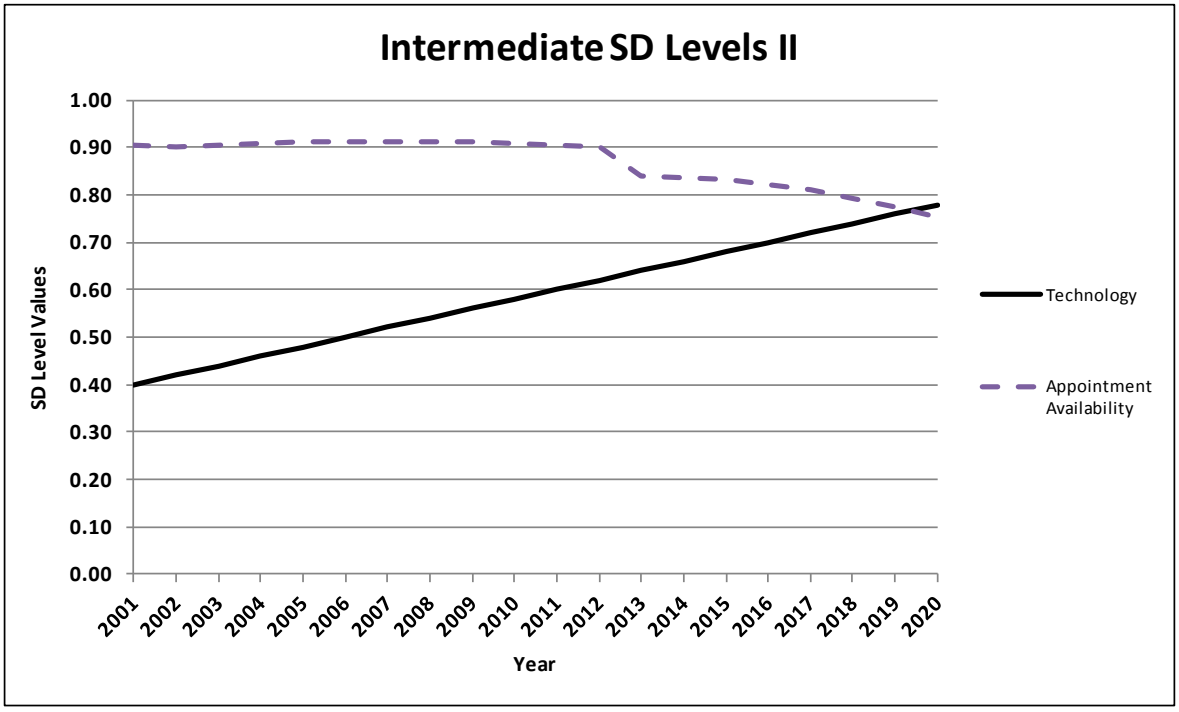


Figure 4-36. Plots of Intermediate SD Levels II for Statistically Best Policy

Table 4-37. Primary SD Levels by Year for Statistically Best Policy

Year	Primary SD Levels		
	Satisfaction with Screening Process	Public Awareness	Access
2001	0.751	0.500	0.800
2002	0.746	0.500	0.788
2003	0.743	0.500	0.778
2004	0.739	0.500	0.766
2005	0.733	0.500	0.752
2006	0.727	0.500	0.737
2007	0.719	0.500	0.721
2008	0.712	0.500	0.705
2009	0.703	0.500	0.686
2010	0.693	0.500	0.667
2011	0.681	0.500	0.645
2012	0.668	0.500	0.622
2013	0.640	0.500	0.591
2014	0.642	0.500	0.589
2015	0.643	0.500	0.586
2016	0.642	0.500	0.582
2017	0.639	0.500	0.576
2018	0.634	0.500	0.568
2019	0.628	0.500	0.558
2020	0.620	0.500	0.547

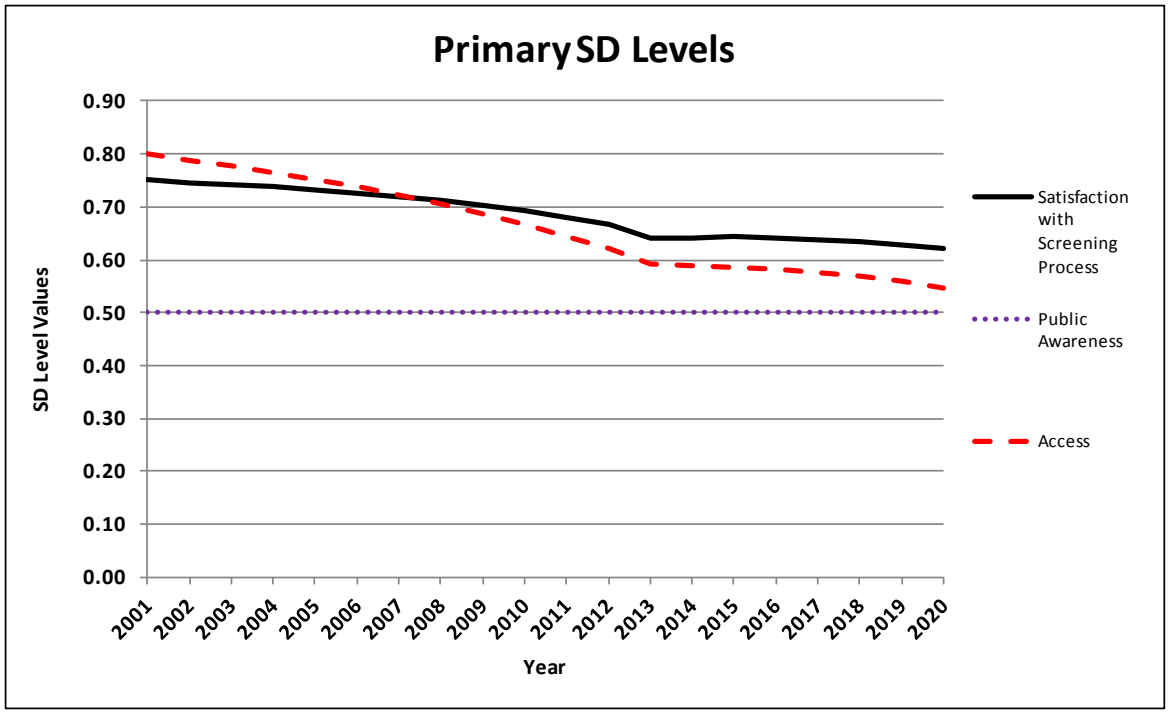


Figure 4-37. Plots of Primary SD Levels for Statistically Best Policy

Table 4-38. SD Population Results

Year	SD Population Results				
	# of Mammograms	Appointments Possible	Adherence Percent	Percentage of Population Screened	Percentage of Population Eligible
2001	9846.3	12196.6	80.73	47.84	59.26
2002	9722.3	12137.6	80.10	46.98	58.66
2003	9661.4	12121.9	79.70	46.35	58.15
2004	9681.3	12169.9	79.55	46.03	57.86
2005	9755.2	12279.8	79.44	45.86	57.73
2006	9901.4	12434.9	79.63	45.94	57.70
2007	10023.1	12612.9	79.47	45.90	57.76
2008	10211.6	12864.9	79.38	46.05	58.01
2009	10446.0	13188.3	79.21	46.27	58.42
2010	10720.0	13553.7	79.09	46.58	58.89
2011	11000.3	13920.8	79.02	46.82	59.26
2012	13406.9	16998.6	78.87	55.84	70.80
2013	13785.9	17492.6	78.81	56.10	71.19
2014	14218.5	18109.5	78.51	56.41	71.85
2015	14717.3	18768.5	78.42	56.84	72.49
2016	15305.0	19492.0	78.52	57.52	73.26
2017	15946.1	20332.8	78.43	58.16	74.16
2018	16619.4	21206.2	78.37	58.74	74.96
2019	17302.0	22118.3	78.22	59.15	75.61
2020	18072.9	23100.1	78.24	59.63	76.21

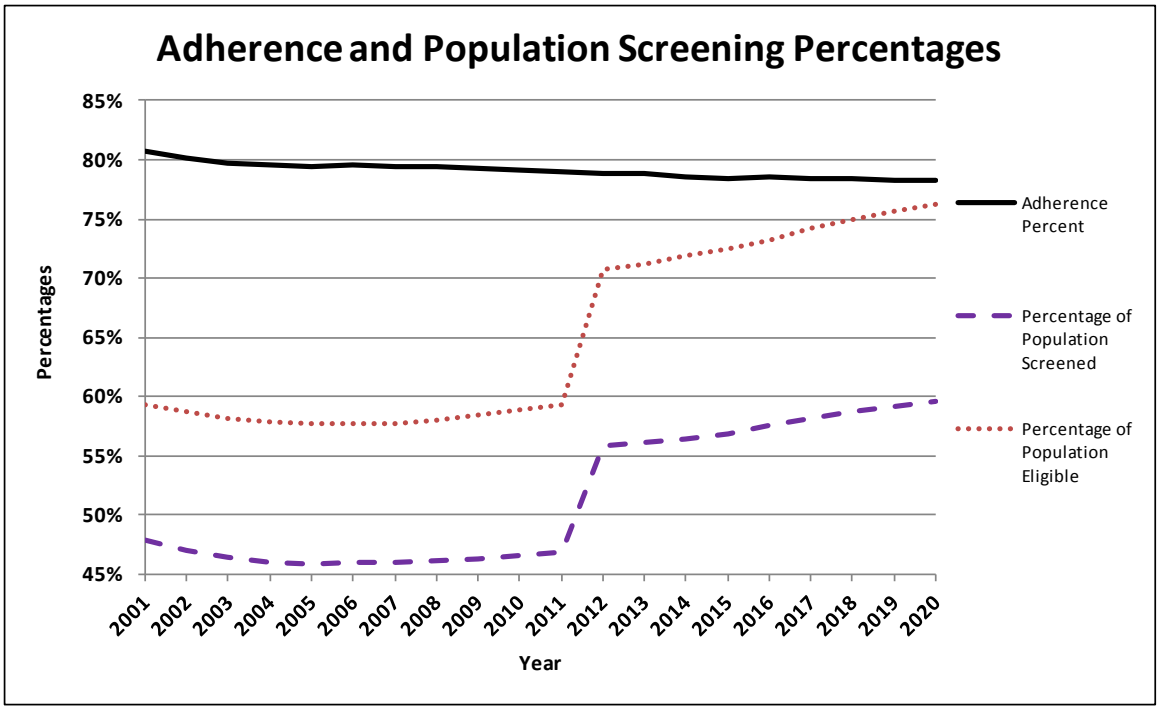


Figure 4-38. Plots of SD Outputs for Statistically Best Policy

Note that adherence trending very slightly downwards from 2001 to 2011, and then starts to trend downwards at a faster rate in 2012. This effect comes from the fact that the number of screening policies is decreasing from 2001 to 2011 and then remains constant for 2012 and beyond; however the demand for screening is increasing for the entire period 2001 to 2020, and especially during the period 2012 to 2020 when the population begins to grow at a faster rate. This increase in demand accompanied by a lack of increase in the number of screening facilities causes congestion to increase and satisfaction with the screening process to decrease.

4.5.3.4 Incidence, Deaths, and Stage Distribution, Method of Detection, and Treatment

In this section we present the results for cancer incidence, cancer deaths, stage distribution at diagnosis, method of detection, and the percentage of women treated. We present each of these results for both the historically calibrated screening policy and the statistically best screening policy for the years 2012–2020.

Table 4-39. Incidence and Breast Cancer Death Rates for Statistically Best Policy

Year	Rates are Per 100,000 Women			
	DCIS Incidence Rate	Invasive Incidence Rate	Breast Cancer Death Rate	Breast Cancer Deaths
2001	57.3	154.5	0.00	0.0
2002	86.0	286.6	5.32	1.1
2003	72.0	344.0	15.84	3.3
2004	74.2	392.3	20.45	4.3
2005	70.0	422.6	34.80	7.4
2006	71.9	443.1	52.90	11.4
2007	66.8	436.9	64.55	14.1
2008	74.4	455.9	65.39	14.5
2009	75.8	434.1	78.85	17.8
2010	68.7	455.2	82.54	19.0
2011	67.3	435.0	87.72	20.6
2012	136.6	578.9	93.30	22.4
2013	104.2	466.7	85.06	20.9
2014	91.3	461.8	105.10	26.5
2015	89.3	460.4	94.60	24.5
2016	104.1	439.8	105.50	28.1
2017	83.9	457.0	107.60	29.5
2018	100.1	429.6	83.06	23.5
2019	83.4	461.8	98.80	28.9
2020	95.0	455.4	101.40	30.7
2001-2011	-	-	-	113.5
2012-2020	-	-	-	235.0
Post 2020	-	-	-	278.6

Table 4-40. Incidence and Breast Cancer Death Rates for Historically Calibrated Policy

Year	Rates are Per 100,000 Women			
	DCIS Incidence Rate	Invasive Incidence Rate	Breast Cancer Death Rate	Breast Cancer Deaths
2001	54.4	139.0	0.97	0.2
2002	80.7	282.2	3.87	0.8
2003	73.9	364.6	11.04	2.3
2004	84.6	389.4	23.76	5.0
2005	60.6	436.7	39.99	8.5
2006	71.5	432.0	55.23	11.9
2007	72.8	423.7	60.93	13.3
2008	70.4	450.6	66.30	14.7
2009	68.2	451.0	78.02	17.6
2010	71.3	441.4	78.19	18.0
2011	60.9	462.2	77.51	18.2
2012	72.5	439.0	94.10	22.6
2013	72.1	452.5	101.30	24.9
2014	77.4	449.5	96.40	24.3
2015	83.8	441.3	97.80	25.3
2016	74.8	444.4	105.90	28.2
2017	68.6	442.1	109.40	30.0
2018	77.4	422.7	101.00	28.6
2019	76.2	448.5	90.30	26.4
2020	80.8	417.9	102.90	31.2
2001-2011	-	-	-	110.5
2012-2020	-	-	-	241.5
Post 2020	-	-	-	284.0

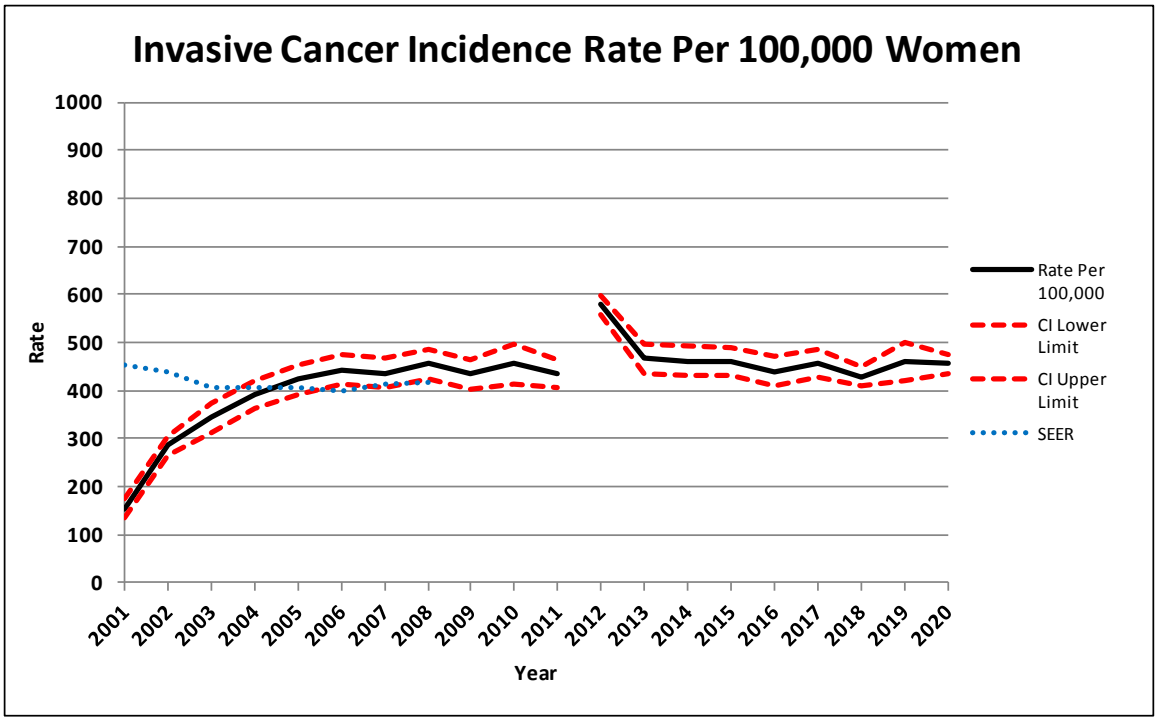


Figure 4-39. Invasive Cancer Incidence Rates for Statistically Best Policy

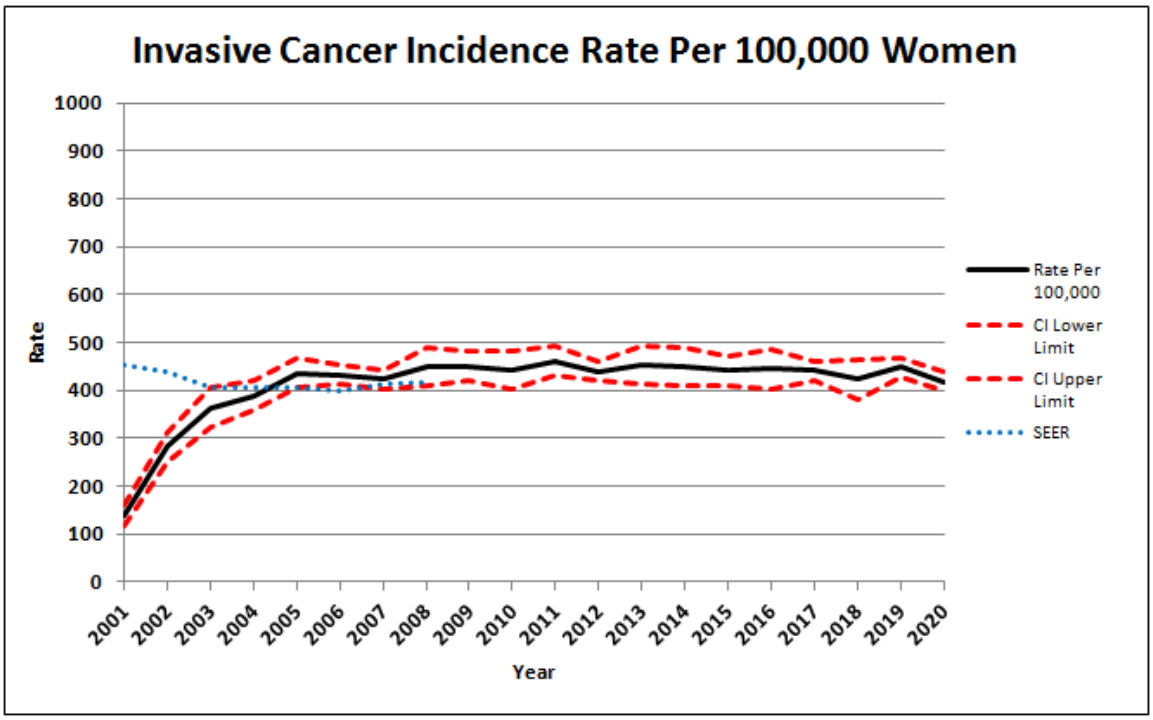


Figure 4-40. Invasive Cancer Incidence Rates for Historically Calibrated Policy

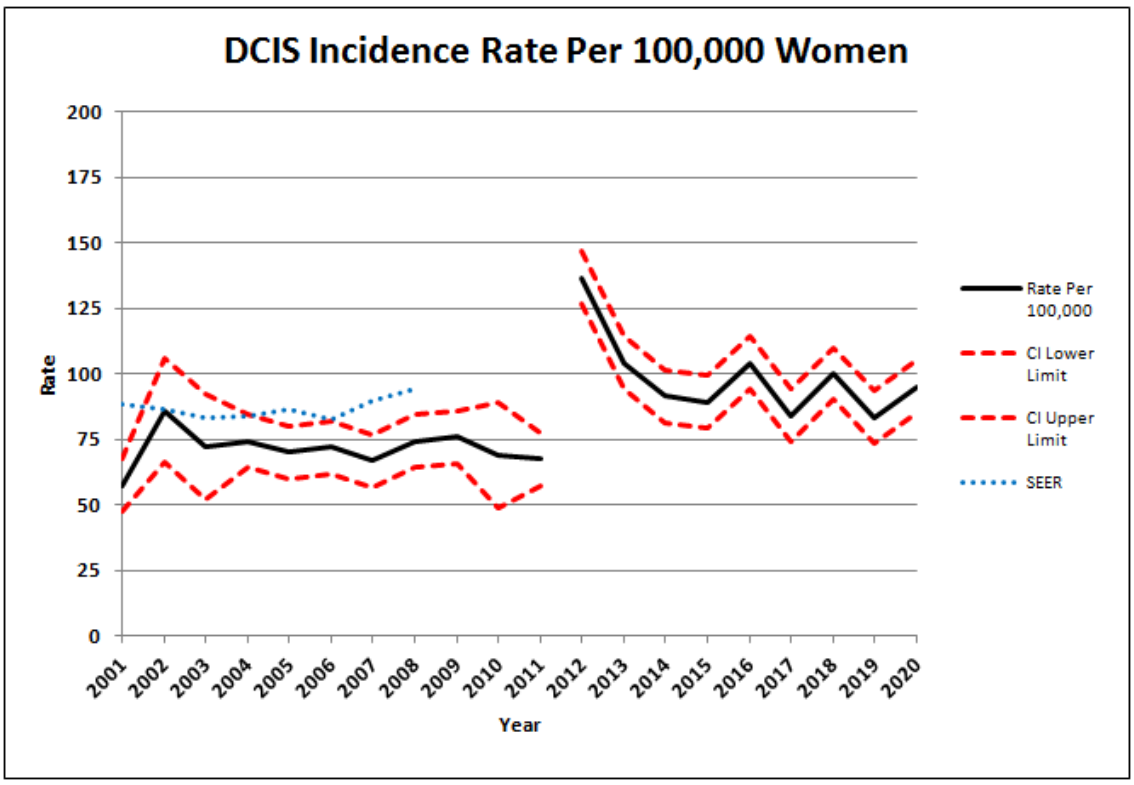


Figure 4-41. DCIS Incidence Rates for Statistically Best Policy

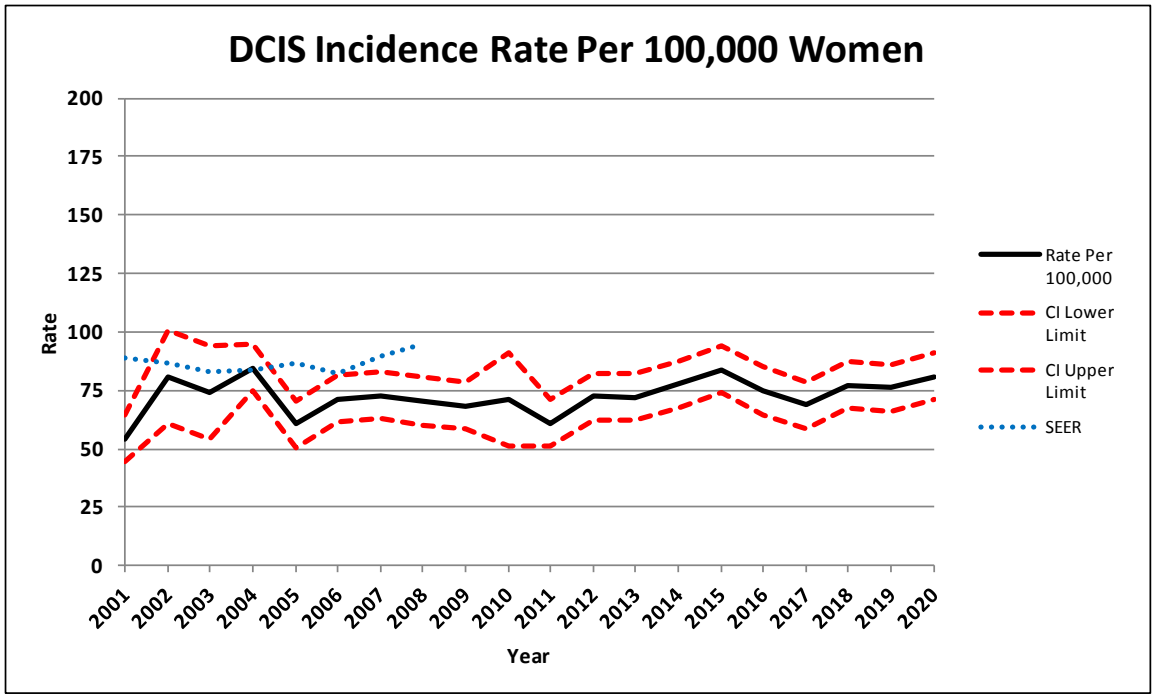


Figure 4-42. DCIS Incidence Rates for Historically Calibrated Policy

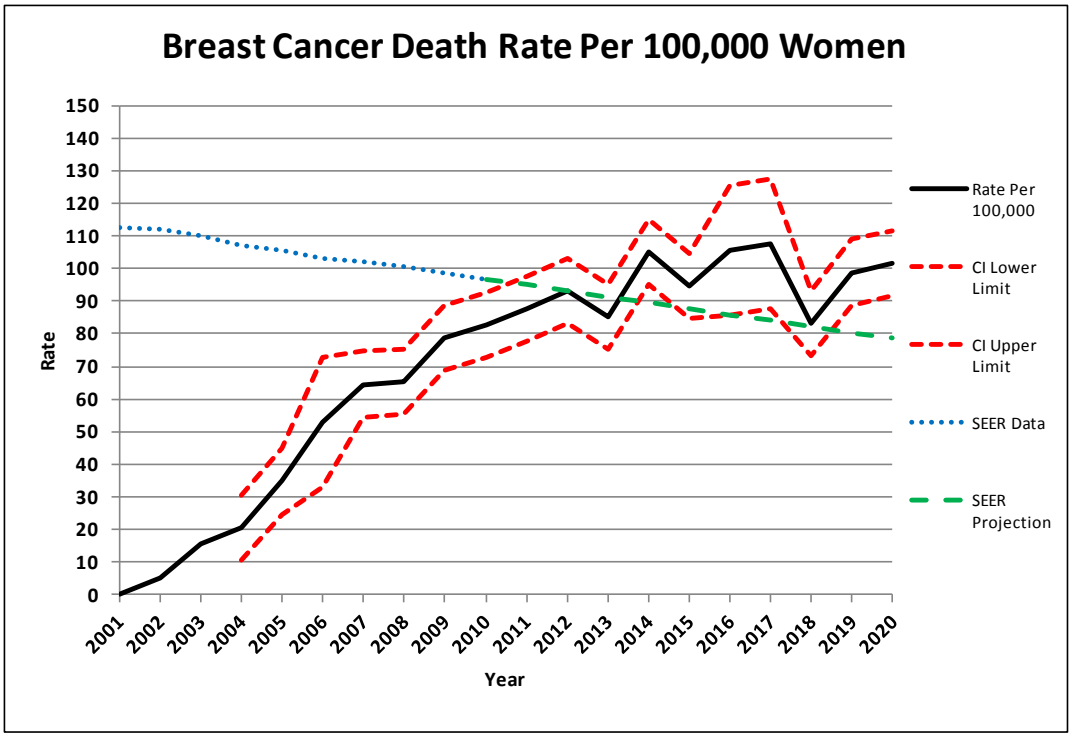


Figure 4-43. Breast Cancer Death Rates for Statistically Best Policy

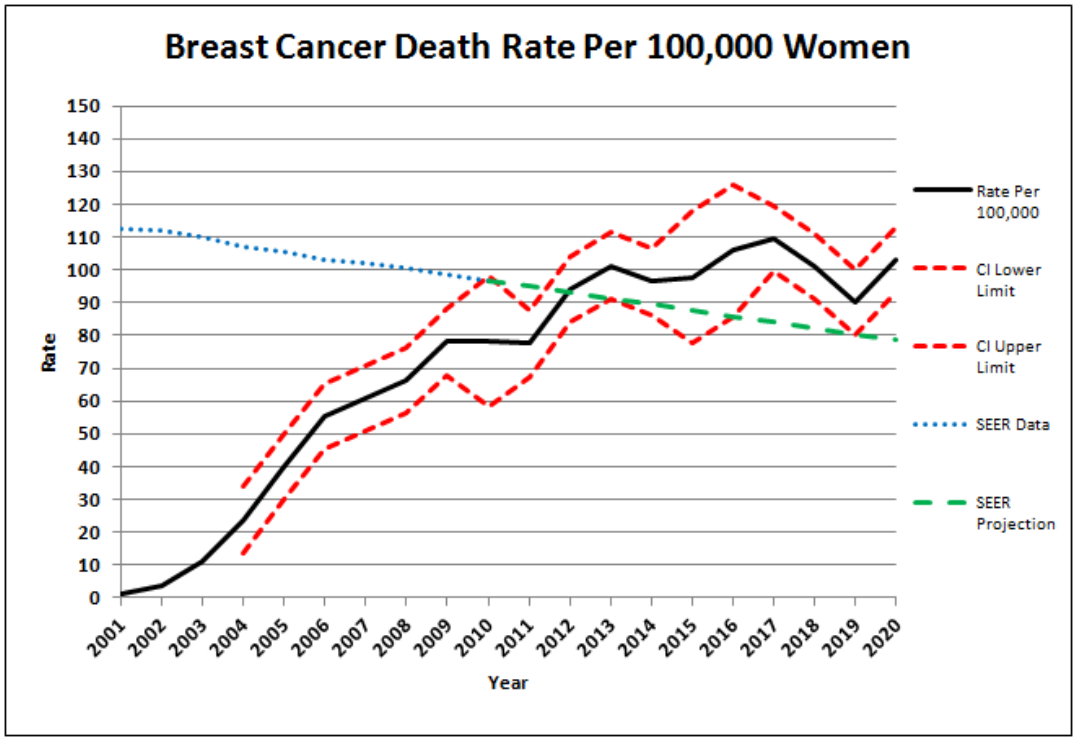


Figure 4-44. Breast Cancer Rates for Historically Calibrated Policy

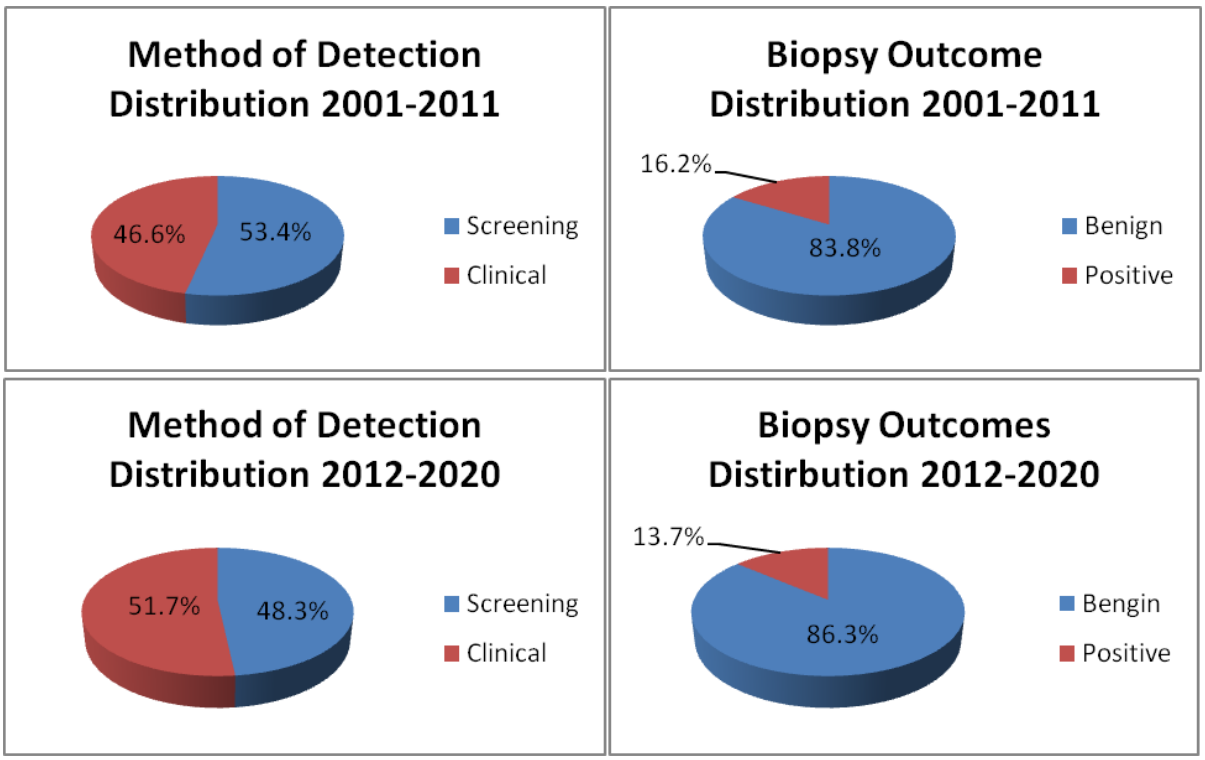


Figure 4-45. Method of Detection and Benign Biopsy Percentages for Historically Calibrated Policy

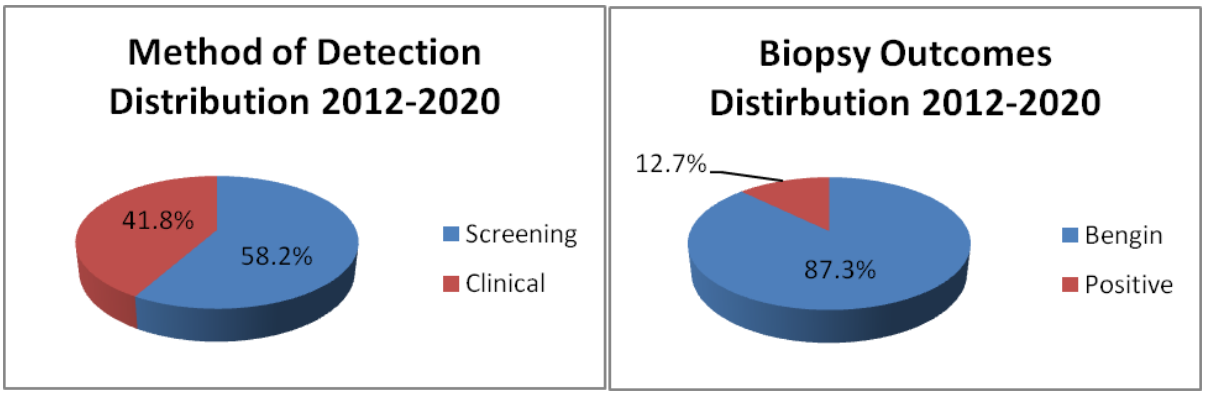


Figure 4-46. Method of Detection and Benign Biopsy Percentages for Statistically Best Policy

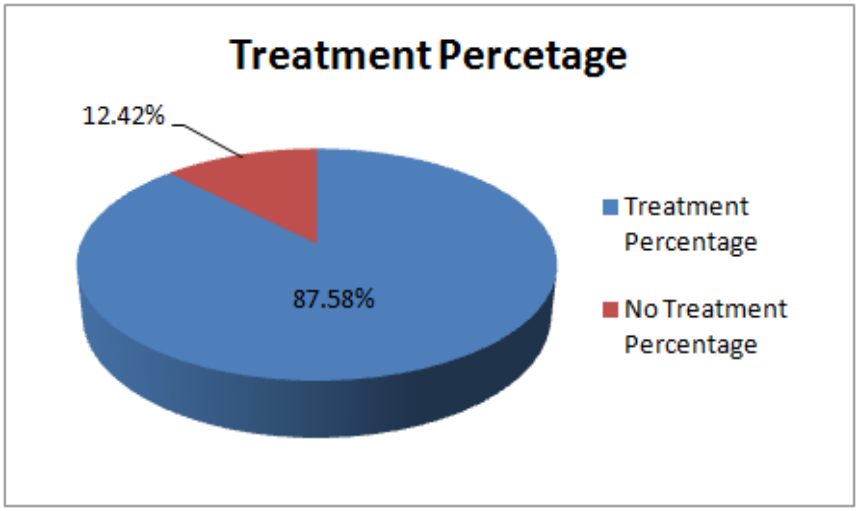


Figure 4-47. Percentage of Women Treated for Breast Cancer

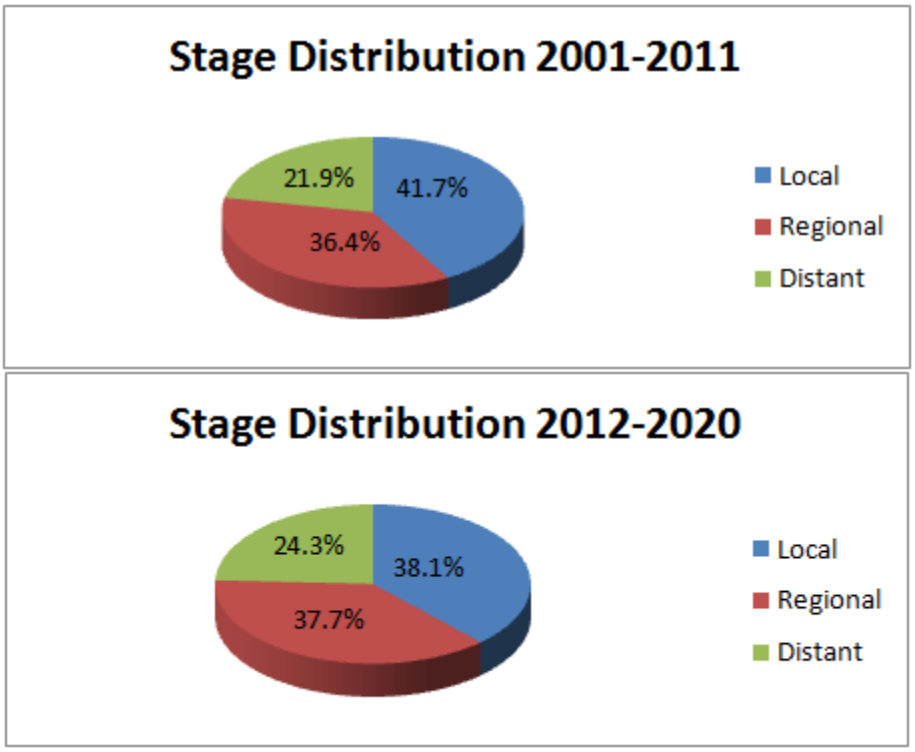


Figure 4-48. Stage Distribution for Historically Calibrated Policy

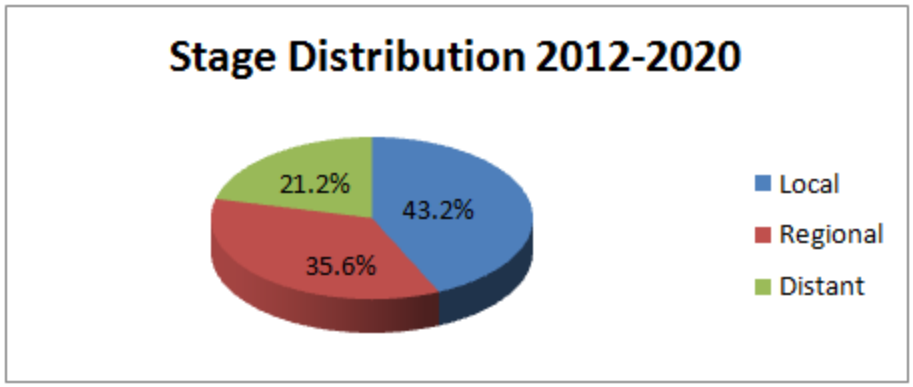


Figure 4-49. Stage Distribution for Statistically Best Policy

4.5.3.5 Cost and Life-Years

In this section we present all the results related to costs and life-years saved. We graph cost by procedure type, costs by method of detection, costs of false positive exams by procedure type, QALYs saved by method of detection, cost per QALY saved by method of detection, life-years saved by method of detection, and cost per life-year saved by method of detection. We present each of these results for both the historically calibrated screening policy and the statistically best screening policy for the years 2012–2020.

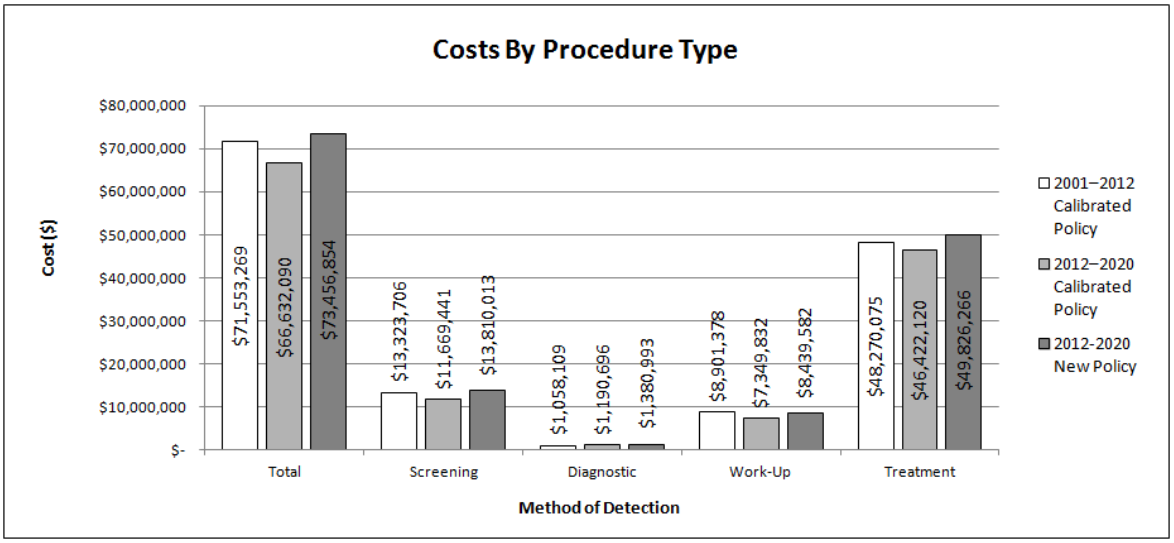


Figure 4-50. Costs by Procedure Type

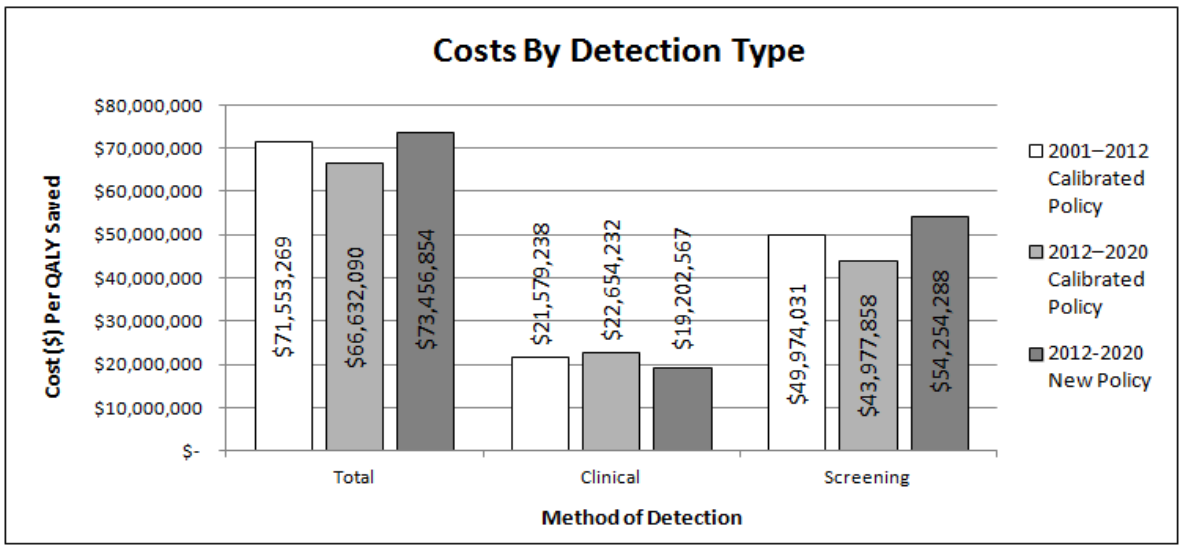


Figure 4-51. Costs by Method of Detection

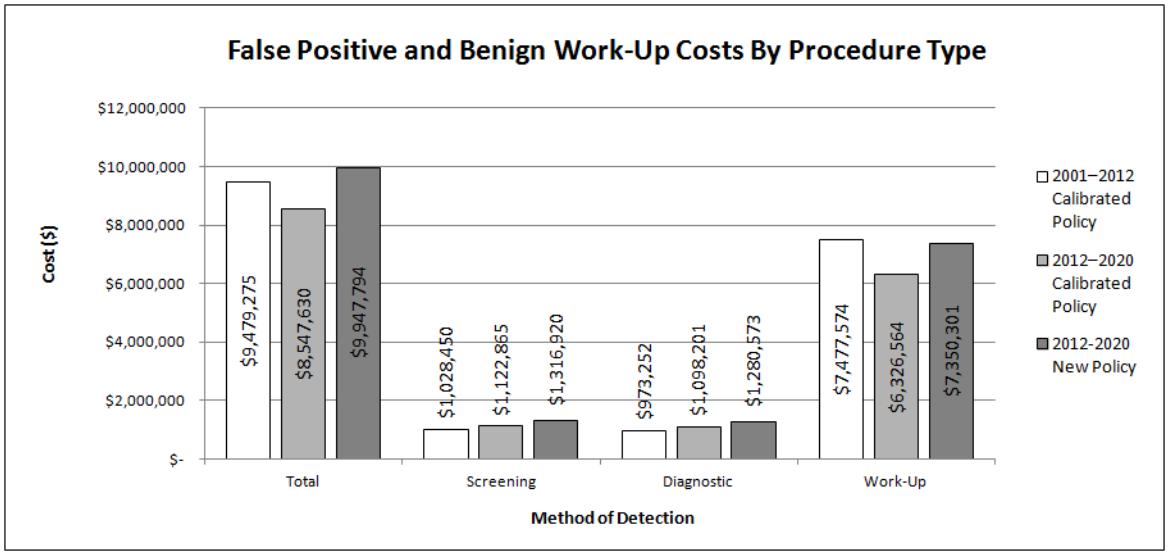


Figure 4-52. False Positive Costs by Procedure Type

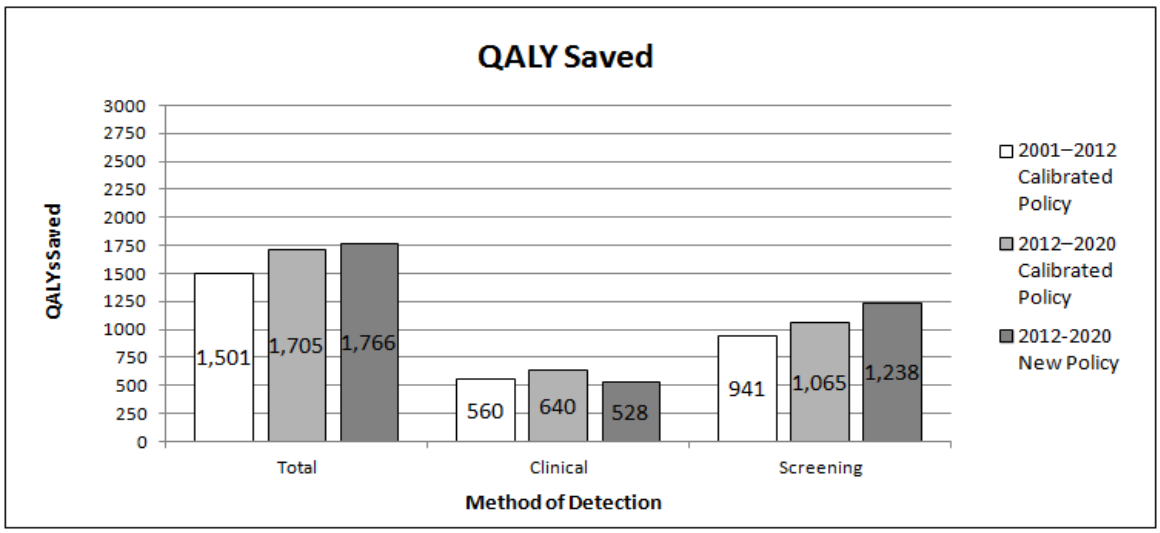


Figure 4-53. QALYs Saved

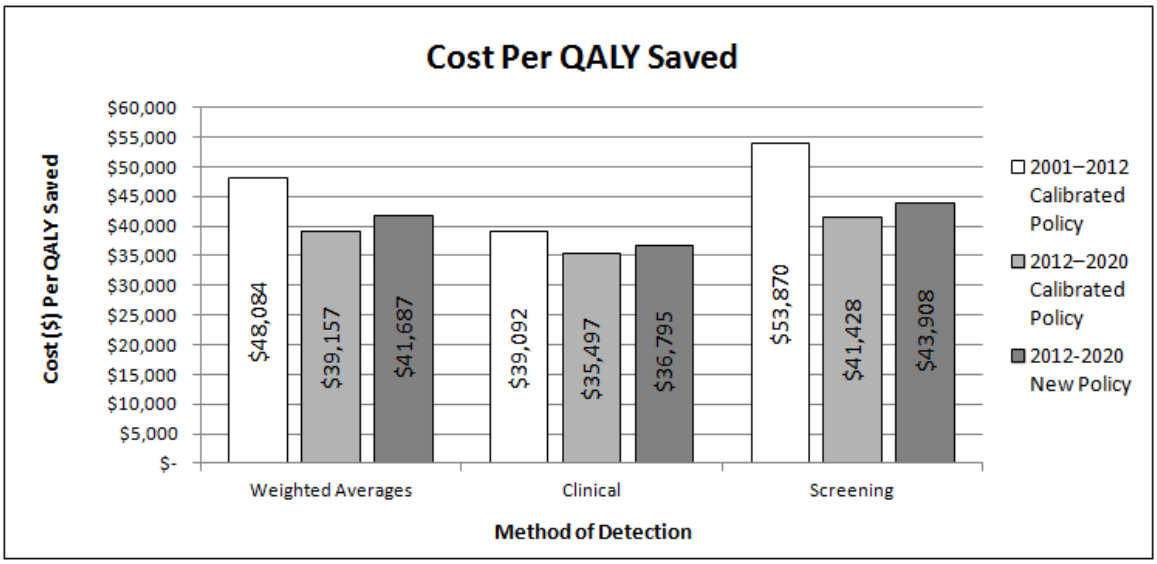


Figure 4-54. Cost Per QALY Saved

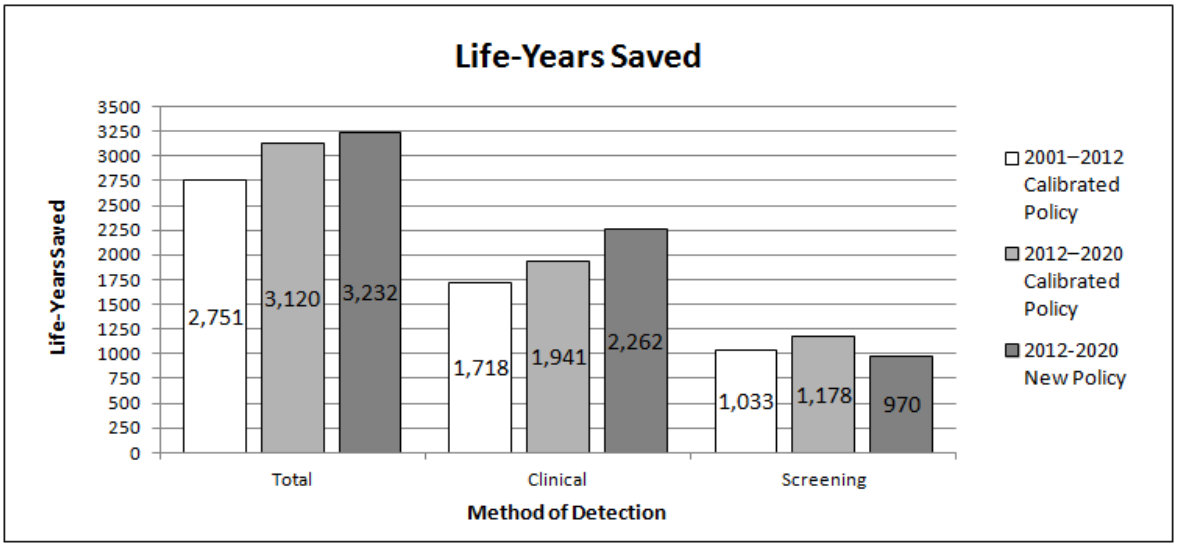


Figure 4-55. Life-Years Saved

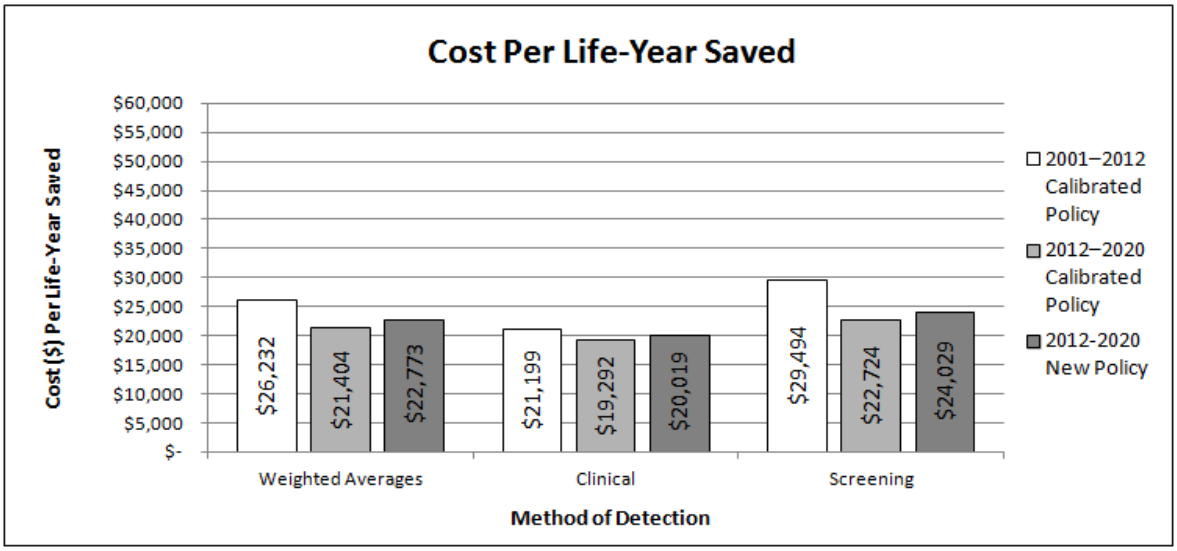


Figure 4-56. Cost Per Life-Year Saved

4.5.4 SD Input Experimentation

In addition to the experimentation with screening policies, we also conducted a small experiment with the SD input levels in order to demonstrate the effects of changes in the SD input levels on the other SD levels and outputs. There are two SD inputs with three levels, and two other SD inputs with five levels, yielding a total of 225 $(3^2 \times 5^2)$ combinations of SD inputs. Given that SD type results, which are typically yearly values of numerous levels, require a large number graphs in our particular case, we chose to simply demonstrate the ability of changes in the SD inputs to affect what is happening in the model. Adherence is a function of the both SD levels and individual DES attributes, so when changes in the SD input levels affect the adherence, this is evidence that the SD and DES submodels are interacting as they were designed to do. In order to demonstrate this, we choose to take our statistically best screening policy and run best-case and worst-case scenarios with the input levels with that same screening policy in effect. This strategy is a type of blocking which removes the effects of the screening policy, and simply shows the effects of changing the SD input levels on the other SD levels and model adherence related model outputs. What follows is a series of graphs for both best-case and worst-case scenarios for SD input levels, intermediate SD levels, primary SD levels, and adherence outputs. These could be compared with the graphs of these measures for the default levels of the SD inputs given in the previous section.

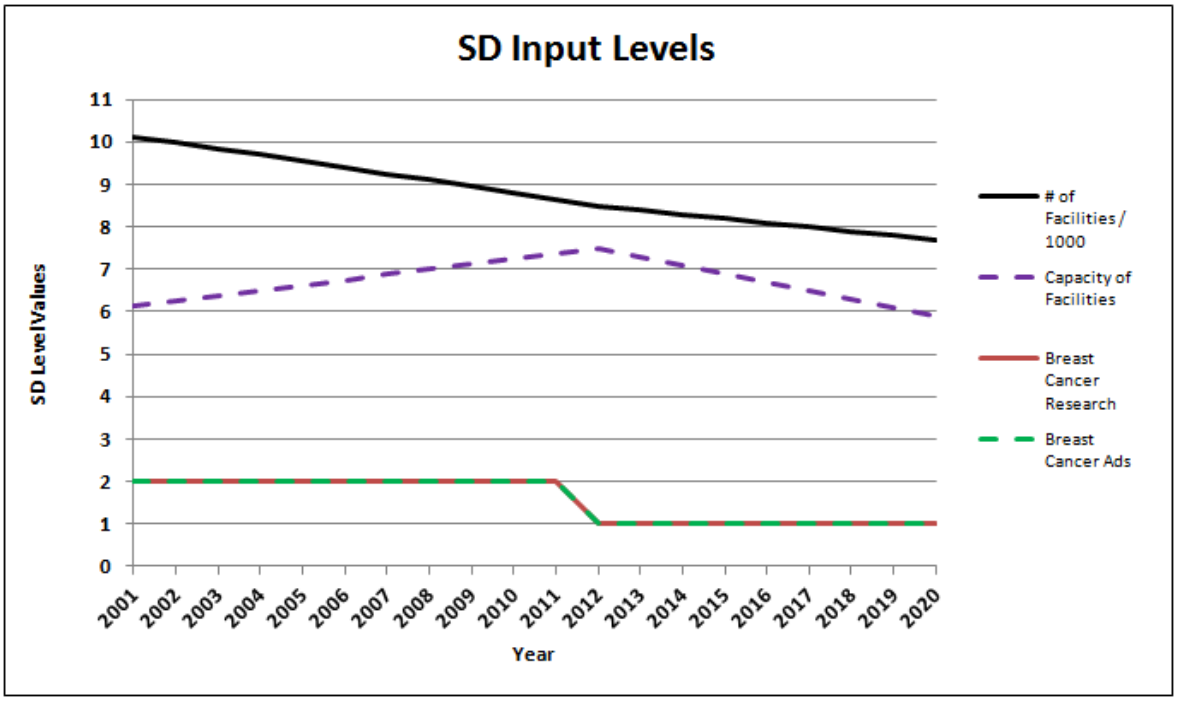


Figure 4-57. SD Input Levels for Worst-Case Scenario and Statistically Best Policy

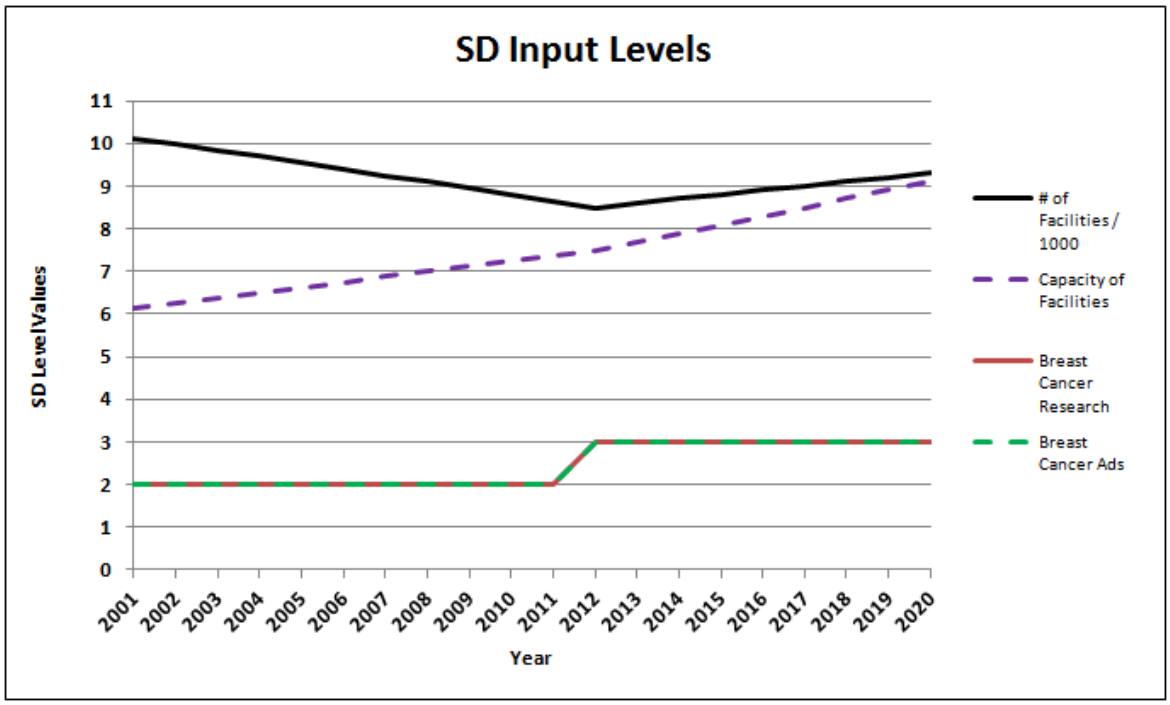


Figure 4-58. SD Input Levels for Best-Case Scenario and Statistically Best Policy

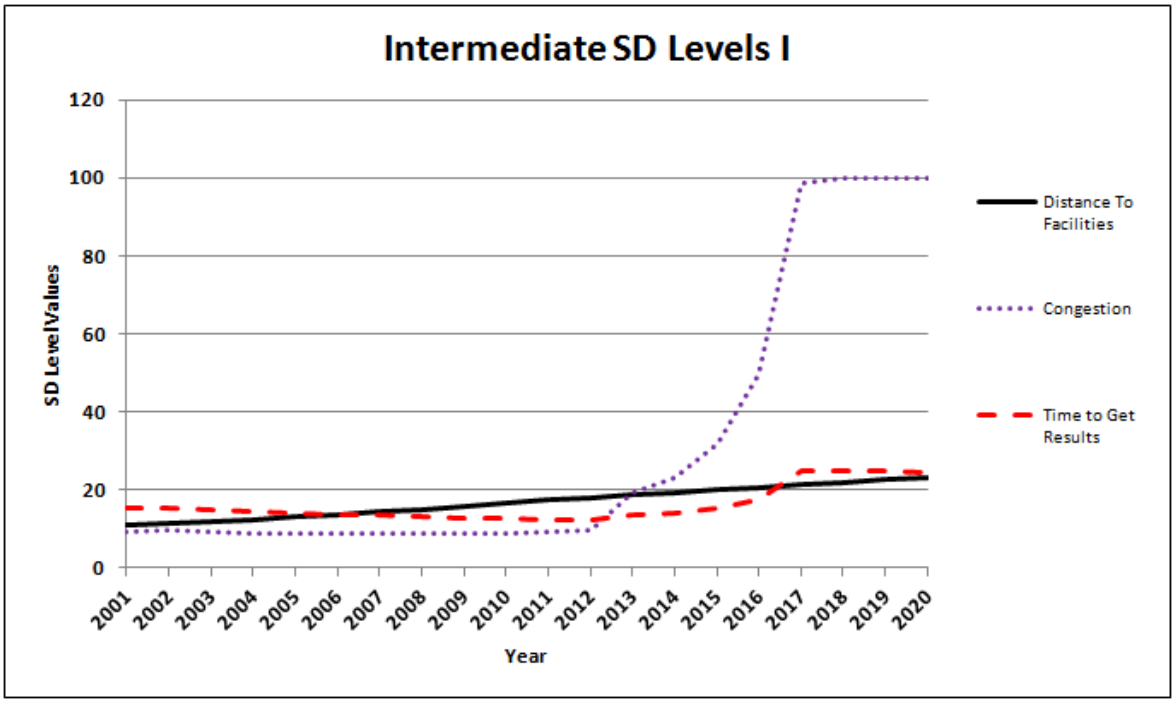


Figure 4-59. Intermediate SD Levels I for Worst-Case Scenario and Statistically Best Policy

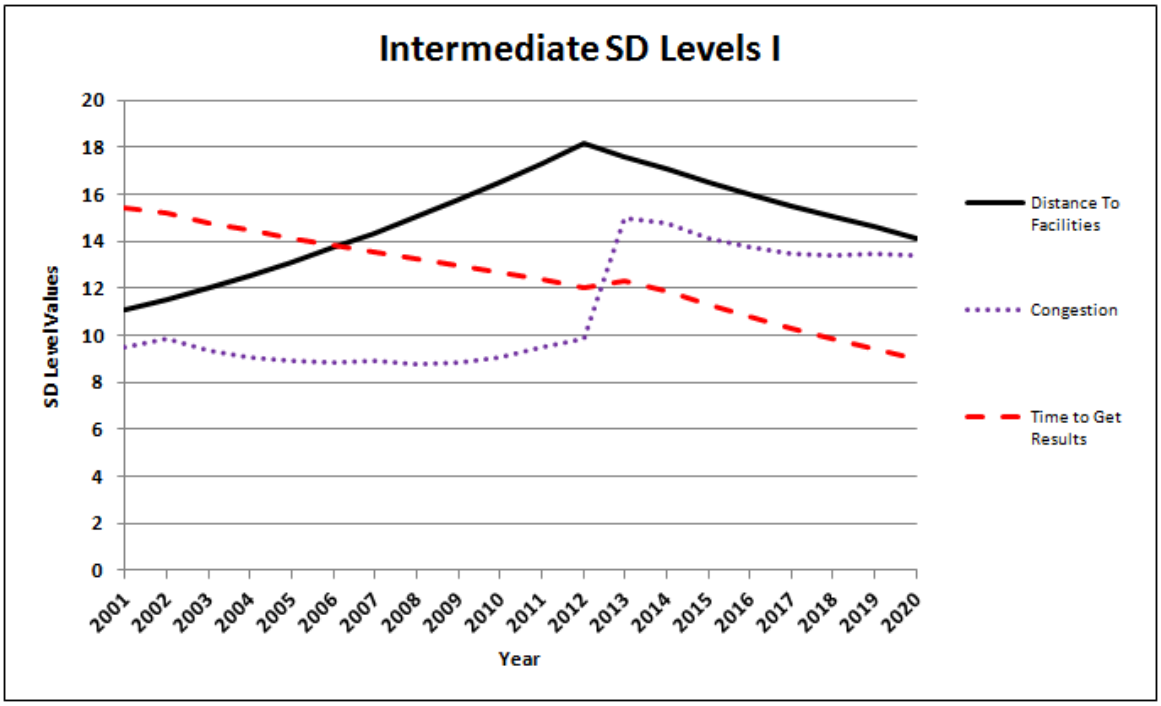


Figure 4-60. Intermediate SD Levels I for Best-Case Scenario and Statistically Best Policy

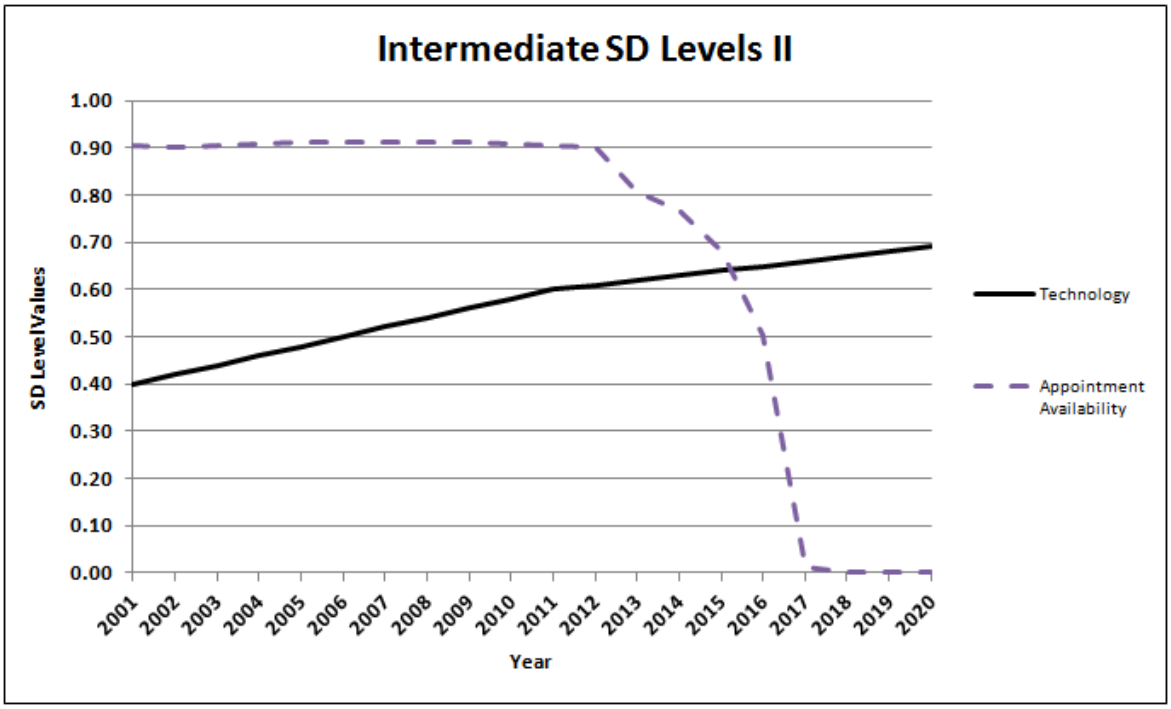


Figure 4-61. Intermediate SD Levels II for Worst-Case Scenario and Statistically Best Policy

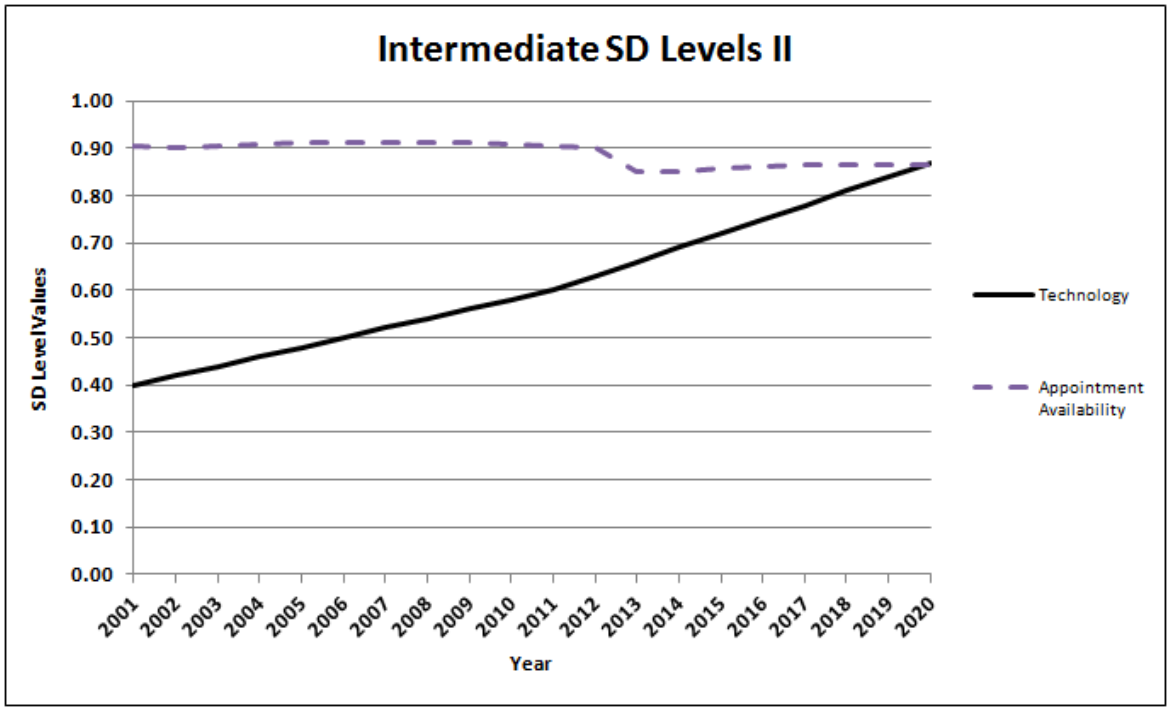


Figure 4-62. Intermediate SD Levels II for Best-Case Scenario and Statistically Best Policy

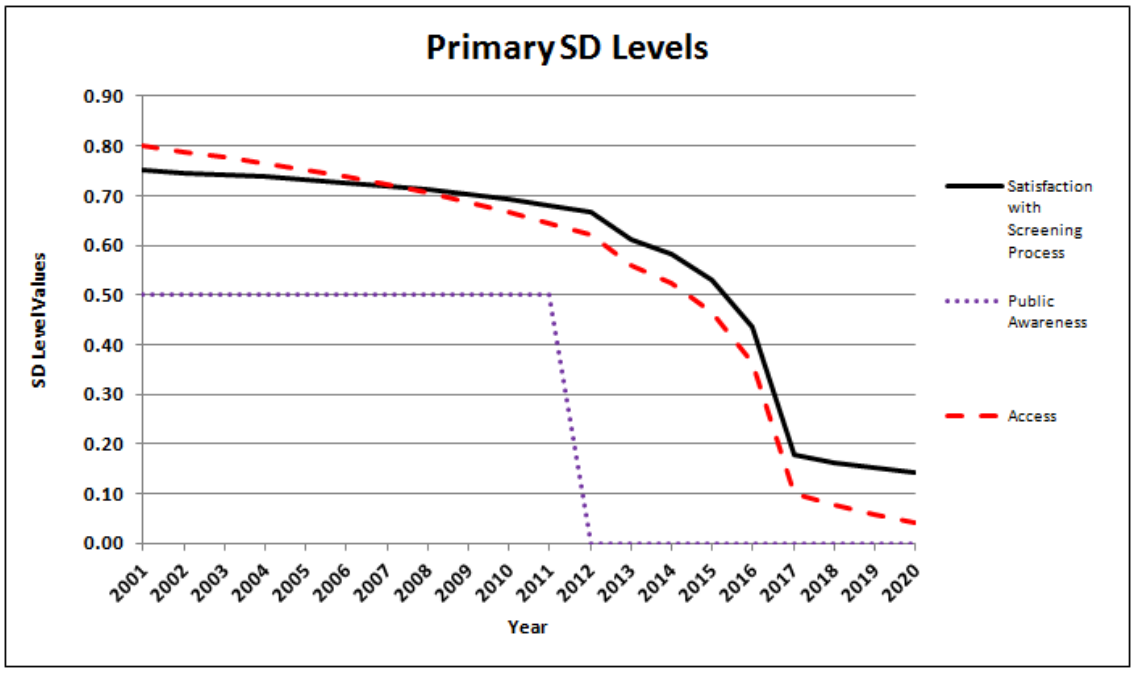


Figure 4-63. Primary SD Levels for Worst-Case Scenario and Statistically Best Policy

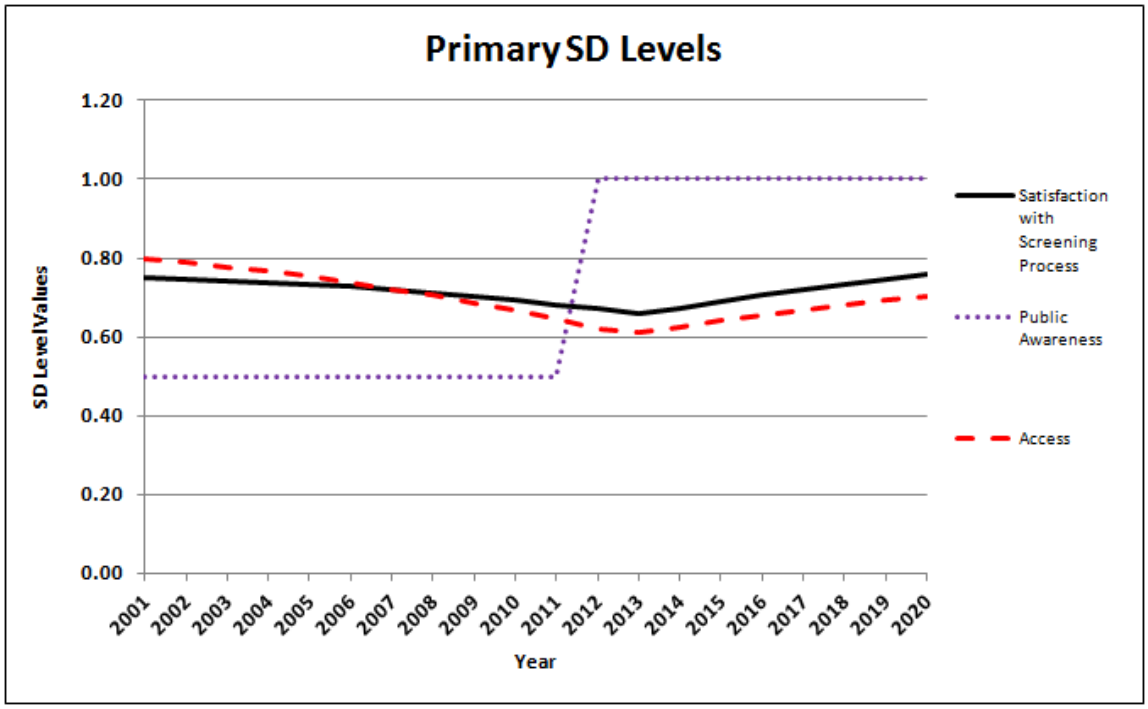


Figure 4-64. Primary SD Levels for Best-Case Scenario and Statistically Best Policy

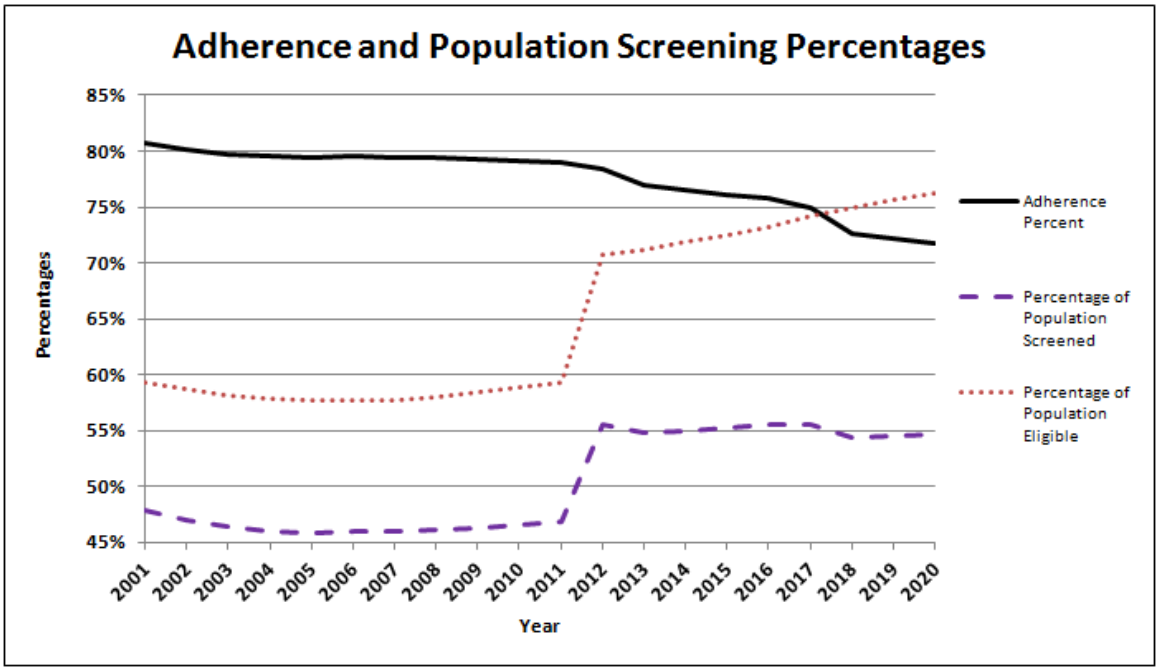


Figure 4-65. SD Outputs for Worst-Case Scenario and Statistically Best Policy

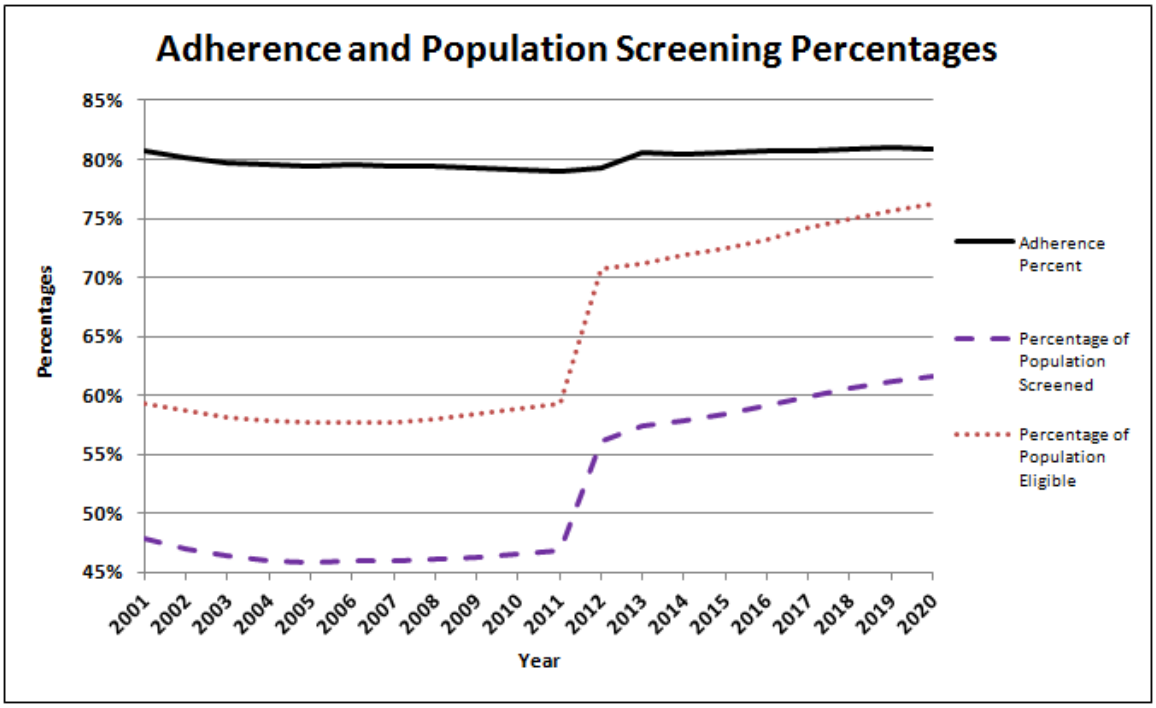


Figure 4-66. SD Outputs for Best-Case Scenario and Statistically Best Policy

4.5.5 Discussion

We have presented a great number of results from the integrated screening model in this section. The results we presented here were from a limited number of experiments performed with this two-phase simulation modeling methodology. In fact, numerous other experiments with factor-based screening are possible and may yield interesting results. In addition, experiments about how different screening policies for the past years could have led to different results could also be of interest. However, we were only able to present a limited number of results, but these were focused on the performance measures and types of screening policies we felt were most important. We feel these types of experiments demonstrate the power of this simulation methodology and its ability to combine population characteristics with the characteristics of individuals to produce a model that allows policy-makers to determine the effects of alternative screening policies and governmental policies towards breast cancer promotion on numerous key performance measures simultaneously.

One major objective of this research is to inform public breast cancer screening policy, and with the ability to evaluate the potential effects of nearly any screening policy or governmental policy on older women undergoing breast cancer screening in our population. For example, you cannot lower the death rate or detect more cancers in the local and regional states without increasing the percentage of benign biopsies and the total cost of false positives, unless technology dramatically improves. The ability to explore these types of trade-offs is a key feature of this model and was a goal of this research from the beginning.

At the outset of this research, we intended to make cost-effectiveness (measured by the cost per QALY saved) the primary measure of effectiveness for a screening policy, but this is a purely theoretical perspective that inadequately reflects several practical considerations. It is important to realize that we are not discussing the number of defective parts or customer satisfaction with a manufactured good, we are dealing with the impact of policy decisions on people's quality of life and ability to share it with their families and friends. The importance of considering this cannot be understated. Therefore, we ultimately

choose to make the number of QALYs saved the primary measure of importance, and simply managed cost-effectiveness by exploring policies which were at minimum still considered cost-effective by global disease screening standards.

We also point out that no model is a perfect representation of the real world, and this model is no exception. Should the screening policies suggested here be put into effect, it is likely the realizations of the performance measures would not be equal to the mean values produced by the simulation model. However, simulation is extremely effective at comparing the differences between two scenarios, and this model has the capability to quickly generate a full set of results for any almost any possible screening policy or government policy. At minimum, the effects of the different policies relative to one another provide information about the effects of those policies on individuals. Of course, in an ideal situation, we would screen every woman every year, detect almost every breast cancer early, and be able to treat most women and return them to a state of health and a full life. Unfortunately, that is just not practical, and thus it becomes important to explore alternatives that are effective at saving lives and keeping screening cost-effective for the entire population.

Combining DES and SD submodels into a single model that allows both to run simultaneously and interact with each other was another major goal of this research. As demonstrated in the previous chapter, the integrated DES/SD model provides insights into the effects of alternative screening and governmental policies concerning breast cancer for individual women, which is difficult to achieve if using one methodology alone. For example, if we had chosen to use only DES model logic, then we would have been forced to come up with a method for queuing up over 30,000 entities at different screening facilities; and this effect would have been extremely detrimental to overall runtime and would have rendered optimization of model parameters out of the question. In addition, using only SD model logic would have produced only population-level results, and would not have been able to create complex stochastic interacting submodels nor simulate the lives of individual women, which is what enables us to capture all of the important performance measures

simultaneously. Thus, using either one of these methodologies by itself would have limited the ability of the model to achieve its overall goal: to determine the effects of population-level policies and decisions and individuals. For this research, we took the approach of creating a complex DES submodel first, and then we incorporated the outside effects of population characteristics by adding and SD submodel which interacted with this DES submodel by altering its equations and logic. Adherence was the focus of the SD submodel, since it is a function of both individual attributes and population characteristics, providing one convenient way for linking DES and SD submodel logic through a logistic regression model for adherence. In addition, we formed several other hybrid SD levels whose values varied over time and altered the DES submodel logic, such as the adjustment for the percentage of false positives based on the increasing level of technology. These types of interactions characterize the exchange of information between the DES and SD submodels, and the results from a combined model are typical DES and SD results, only they are dependent on one another. The benefit of combining these models was well worth the cost of formulating a methodology for doing so. The effects of changing the level of advertising for breast cancer screening or the number of breast cancer screening facilities on the number of lives saved, the congestion at screening facilities, or the cost-effectiveness of screening are revealed quickly upon examining a full set of results from this integrated model. These types of results would not have been possible to obtain using only one of these methodology alone.

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

The final chapter of this dissertation presents conclusions, contributions to breast cancer screening policy, contributions to simulation modeling methodology, limitations of the model, and potential future work.

5.1 Conclusions

The most cost-effective screening policy for US women at least 65 is annual screening until age 70 for women who are in the top 5% in terms of risk, and a one-time screening at age 68 for all others. However, this optimal policy in terms of cost-effectiveness performs poorly in other areas of performance that are relevant to saving life. After reviewing with our experts the results from this policy and other policies with similar cost-effectiveness, we realized that simply finding the most cost-effective policy was not equivalent to finding the "best" policy to use for women in this age group. Instead, we examined the following alternate objectives while maintaining a reasonable level of cost effectiveness: (a) maximizing QALYs saved, and (b) minimizing the number of breast cancer deaths. We found that from the perspectives of both practical and statistical significance, annual screening for all women from age 65 until age 80 was a superior policy in terms of saving lives; and we presented a full set of results from using this policy. Nevertheless, some policy makers may not judge these performance measures to be the most important; and one of the key features of this model is its ability to evaluate alternate screening policies in terms of almost any relevant performance measure. Thus, others can review full sets of results generated by this model and make decisions about which policy they consider to be the best, and they will have numerical evidence from a statistically validated simulation model to support their perspective.

Combining the DES and SD simulation modeling methodologies provided a combination of population-level outputs and performance measures for individuals that would have been very difficult to achieve using only one of these methodologies in isolation.

The types of results produced give insight into impacts of policy on both individuals and the population as a whole, which was the initial motivation for combining these two types of methodologies. Many researchers and practitioners working at the interface between computer simulation and health care systems engineering have strongly supported the idea of combining these two types of simulation, and in this dissertation research we have taken a step towards creating simulation models that have a broad perspective and take into account effects beyond the boundaries of the system under immediate consideration. We feel that this integrated model demonstrates the values of using these two types of simulations models together, as opposed to only using one perspective when modeling a given system.

5.1.1 Contributions to Breast Cancer Screening Policy

The major contributions of this research to breast cancer screening are: (a) the development of the Tejada model, a combined DES/SD simulation tool for the analysis of screening policies and government policies for women at least 65; and (b) comprehensive experimentation with the Tejada model to estimate relevant performance measures for screening policies that have the potential to maximize the number of lives saved, to minimize the number of deaths caused by breast cancer, or to maximize system-wide cost effectiveness. We provided a full set of results for such screening policies by determining the proper screening interval and stopping age for individual patients. The Tejada model can be used by policymakers to inform their decisions about breast cancer screening policy, the number and capacity of breast cancer screening facilities, and the amount of advertising and research for breast cancer. The model has the ability to test individualized policies and determine their performance for the near future.

5.1.2 Methodological Contributions to Simulation Modeling

The development of the Tejada model has led to several methodological contributions to simulation modeling. The major contributions are listed below.

1. The Tejada model can be regarded as a template or a guide for how future combined DES/SD simulation models may be designed for other application domains. In addition, it provides an approach to modeling a complex disease and the screening and treatment of that disease in a population when several disparate performance measures are of key importance.
2. We have identified key performance characteristics to be estimated for breast cancer screening, and we have formulated appropriate point estimators and confidence interval estimators for the expected values of these performance measures.
3. We have formulated a method for sampling a Pearson IV random variable using a transformation of the standard Beta distribution. See Appendix F for a detailed explanation of this procedure and a proof its validity.

5.2 Limitations

As with all mathematical and computer-based models, effective use of the Tejada model requires a thorough understanding of the assumptions on which the model is based and of the resulting limitations on the model's applicability in practice. The key limitations of this model are the following.

1. We considered the population of US women at least 65; thus the model should only be applied to screening policies for women at least 65.
2. We assumed that all risk factors aside from age, including family history of breast cancer and body-mass index, remained constant over time. This may not be true of all risk factors for all women.

3. We assumed a Markov chain model for an individual woman's health status; and in doing so we assumed that her health status is a function of only her age and her health status in the previous year, when in fact there may be other key medical and socioeconomic variables on which her health depends. Moreover, the historical values of those variables out to lags greater than one year could have an effect on the probability of changing health states.
4. We assumed the tumor growth is a linear function of age, with no knowledge of the exact functional form for the dependence of tumor size on age.
5. We used expert opinion to determine the probability of being treated as a function of age and health status, and thus these estimates are subjective.
6. We simplified the SD model by using several unitless measures that had a maximum of zero and a minimum of one. We also used simple inputs for breast cancer advertising and research: low, medium, and high. What these levels actually mean may be subject to interpretation; however, the goal of the model is to determine the differences between scenarios, so it is valid to consider the relative difference between a low and high level of breast cancer research. The strategy for causing a shift in those input levels may prove to be complicated, and in the absence of detailed information about the mechanism governing changes in these state variables, we chose to use a minimum-information approach to modeling such changes.
7. The precise functional form of the interdependences by which some SD levels affect other levels was not known for all relationships, so assumptions about these mathematical relationships were made under the guidance of expert opinion when data were not available. While these assumptions are subjective, we are confident that the combination of our experts and available data provided enough information to construct a valid model, and we have mathematical evidence to support the validity our integrated model as a whole.
8. Logistic regression models from the literature were used to represent the following: (a) the probability of being diagnosed with breast cancer given that a screening exam occurred; and (b) the probability of adhering to a given screening appointment.

Probabilities (a) and (b) were each expressed as a function of their own set of variables and individual characteristics. It may not really be true that the logistic function is a linear model of the independent variables. In practice, such linear logistic models are analytically convenient, but they may not be adequate in some applications.

9. In the simulation, treatment is simply a treat or no-treat decision; and therefore the estimated impact on survival may not be as accurate as it would if we modeled the specific type of treatment being given and that treatment's impact on survival. However, such detailed survival data were not readily available from the literature or other sources.

5.3 Future Work

This modeling methodology as it stands leaves the opportunity for several areas of future work. The most obvious is a version of the model that supports the inclusion of women of all ages and provides the same measures of performance. There is considerable debate about how to screen middle-aged women, and a great deal of literature has focused on those women. Consequently, we choose to inform policy where there was a lack of consensus on the appropriate screening policy as well as a lack of strong evidence supporting what should be done. We could also include a more complicated model of treatment decisions, and a model of survival after treatment that accounts for the type of treatment given. DCIS incidence is accounted for, but our model does not have the ability to account for its progression to invasive cancer, which likely occurs over time with some probability. However, no data or reliable expert opinion was available to create a model for this progression of DCIS. Along similar lines, once a woman was afflicted with cancer, her survival was computed from survival curves and/or life-tables; the possibilities of curing the disease or recurrence of the disease were not explicitly considered. Future work could be to simulate whether or not the cancer was adequately treated using a treatment decision model; and then based on that information, we could simulate the probability of recurrence. A new

risk model for recurrent disease would need to be found in the literature or developed using existing data.

Further experimentation with SD model inputs is another possible area of future work. In fact, extending the SD model to include more levels supported by data and creating a less subjective SD model overall could be another potential area of future work.
With respect to types of screening policies, we can add as many different screening policies as we want by appropriately extending the model and the user interface. We can also set up a "Web App" to allow others to use this model or at least submit requests for results based on using different screening policies. Essentially anyone could go to a web site that accesses a server with Arena installed, so the user does not need Arena to run the model. The user would be able to interact with an interface that stores the user's inputs and submits them to the model; then after a few minutes, the user would receive an e-mail or hyperlink with the user-selected results from a displayed list. Another alternative to the web application is a regression metamodel that takes the values of inputs and predicts the values of the outputs without actually running the simulation model. However, classic polynomial metamodels are rarely adequate in practice [46], and more flexible parsimonious metamodeling techniques are required. Using neural networks is a popular approach to obtaining more flexibility, but this results in a massively overparameterized model (i.e., a model that is not parsimonious). Perhaps an approach based on wavelets and be made for flexible, parsimonious, and multiresponse simulation metamodeling [53].

BIBLIOGRAPHY

1. Badgwell, B.D., et al. "Mammography Before Diagnosis Among Women Age 80 Years and Older with Breast Cancer." *Journal of Clinical Oncology* 26, no. 15 (May 2008).

2. Baker, Rose D. "Use of a Mathematical Model to Evaluate Breast Cancer Screening Policy." *Health Care Management Science* 1 (1998): 103-113.

3. Barlow, William E, et al. "Prospective Breast Cancer Risk Prediction Model for Women Undergoing Screening Mammography." *Journal of the National Cancer Institute* 98, no. 17 (September 2006): 1204-1214.

4. Bartels, Robert. "The Rank Version of von Neumann's Ratio Test for Randomness." *Journal of the American Statistical Association* 77, no. 377 (1982): 40-46.

5. Berg, Wendie A. "Talored Supplemental Screening for Breast Cancer: What Now and What Next?" *American Journal of Radiology* 192 (February 2009): 390-399.

6. Bloom, H., M. Richardson, and B. Harries. "Natural History of Untreated Breast Cancer (1804-1933): Comparison of Treated and Untreated Cases According to Histological Grade og Malignancy." *British Medical Journal* 2 (1962): 213-221.

7. Boon, M. E., P. A. Trott, H. van Kaam, P. Kurver, A. Leach, and J. P. Baak. "Morphometry and Cytodiagnosis of Breast Legions." *Vrichows Archive* 396 (1982): 9-18.

8. Brailsford, Sally C. "System Dynamics: What's in it for Healthcare Simulation Modelers." *Proceedings of the 2008 Winter Simulation Conference*, 2008: 1478-1483.

9. Brailsford, Sally C. "Tutorial: Advances and Challenges in Healthcare Simulation Modeling." *Proceedings of the 2007 Winter Simulation Conference*, 2007: 1436-1448.

10. Brailsford, Sally C, and Andrew F Seila. "Opportunities and Challenges in Health Care Simulation." Chap. 1 in *Advancing the Frontiers of Simulation: A Festschrift in Honor of George Samuel Fishman*, by James R Wilson, Christos Alexopoulos and David Goldsman, 1-37. Springer, 2009.

11. Brailsford, Sally C, L Churilov, and S-K Liew. "Treating Ailing Emergency Departments with Simulation: an Integrated Perspective." *Proceedings of Western Multiconference on Health Sciences Simulation.* 2003.

12. Brailsford, Sally C, V A Lattimer, P Tarnarus, and J A Trunbull. "Emergency and On-Demand Health Care: Modelling a Large Complex System." *Journal of the Operational Research Society* 55 (2004): 34-42.

13. Brannick, Michael T. *Lositic Regression.* University of South Florida. June 29, 2007. http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html (accessed September 5, 2010).

14. Breast Cancer Surveillance Consortium (BCSC). "Performance Measures for 1,960,150 Screening Mammography Examinations from 2002 to 2006 by Age." *cancer.gov.* 2009. http://breastscreening.cancer.gov/data/performance/screening-/2009/perf_age.html (accessed December 7, 2011).

15. Breast Cancer Surveillance Consortium (BCSC). "Private Individualized Data Set with Risk Factor Attributes of U.S. Women Undergoing Mammography." Data Set, 2010.

16. Bucci, Michael J., Michael G. Kay, and Donald P. Warsing. "Comparison of Meta-Heuristics for Large Scale Facility Location Problems with Economies of Scale." *IIE Annual Conference and Expo 2007 - Industrial Engineering's Critical Role in a Flat World.* IIE, 2007. 1410-1415

17. Carter, Christine L., Carol Allen, and Donald E. Henson. "Relation of Tumor Size, Lymph Node Status, and Survival in 24,740 Breast Cancer Cases." *Cancer* 63 (January 1989): 181-187.

18. Center for Disease Control and Prevention. *Breast Cancer.* November 23, 2011. http://www.cdc.gov/cancer/breast/ (accessed March 27, 2010).

19. Chahal, Kirandeep, and Tillal Eldabi. "Applicability of Hybrid Simulation to Different Modes of Governance in UH Healthcare." *Proceedings of the 2008 Winter Simulation Conference*, 2008: 1469-1477.

20. Cramp, Derek G, and Ewart R Carson. "Assessing Health Policy Strategies: A Model Based Approach to Decision Support." *Unknown Journal*, Unknown Year: 1969-1983.

21. Crivellari, Dianna, et al. "Breast Cancer in the Elderly." *Journal of Clinical Oncology* 25, no. 14 (May 2007): 1882-1890.

22. Dangerfield, B C. "System Dynamics Applications to European Health Care Issues." *Journal of the Operational Research Society* 50, no. 4 (1999): 345-353.

23. Davies, R, and Sally C Brailsford. "Screening for Diabetic Retinopathy." In *Handbook of OR/MS Applications in Health Care*, 493-518. Norwell, Massachusetts : Kluwer, 2004.

24. Davies, R, P J Roderick, C R Canning, D N Crabbe, and Sally C Brailsford. "Using Simulation Modeling for Evaluating Screening Services for Diabetic Retinopathy." *Journal of the Operational Research Society* 51 (2000): 476-484.

25. DeFrank, JT, BK Rimer, JM Giersch, JM Bowling, D Farrell, and CS Skinner. "Impact of Mailed and Automated Telephone Reminders on Receipt of Repeat Mammograms; A Radomized Controlled Trial." *American Journal of Medicine* 36 (2009): 459-467.

26. Deo, N. *Graphy Theory with Application to Engineering and Computer Science.* Englewood Cliffs, New Jersey: Prentice Hall, 1974.

27. Diab, Sami G, Richard M Elledge, and Gary M Clark. "Tumor Characteristics and Clinical Outcome of Elderly Women with Breast Cancer." *Journal of the National Cancer Institute* 92, no. 7 (2000): 550-556.

28. Downey, Leona, Robert Livingston, and Allison Stopeck. "Diagnosing and Treating Breast Cancer in Elderly Women: A Call for Improved Understanding." *Journal of American Geriatrics Society* 55 (2007): 1636-1644.

29. Duffy, Stephen W., Laszlo Tabar, Bedrich Vitak, and Jane Warwick. "Tumor Size and Breast Cancer Detection: What Might Be the Effect of a Less Sensitive Screening Tool Than Mammography?" *The Breast Journal* 12, no. 1 (2006): S91–S95.

30. Eldabi, T, RJ Paul, and T Young. "Simulation Modelling in Healthcare: Reviewing Legacies and Investigating Futures." *Journal of the Operational Research Society* (Operational Research Society) 58 (2007): 262-270.

31. Evenden, D, P R Harper, Sally C Brailsford, and V Harindra. "System Dynamics Modelling of Chlamydia Infection for Screening Intervention Planning and Cost Benefit Estimation." *IMA Journal of Management Mathematics* 16 (2005): 265-279.

32. Fletcher, Suzanne W, and Joann G Elmore. "Mammographic Screening for Breast Cancer." *New England Journal of Medicine* 348, no. 17 (2003): 1672-1680.

33. Fone, David, et al. "Systematic Review of the Use and Value of Computer Simulation Modeling in Population Health and Health Care Delivery." *Journal of Public Health Medicine* 25, no. 4 (2003): 325-335.

34. Fryback, Dennis G, Natasha K Stout, Majorie A Rosenberg, Amy Trentham-Dietz, Vipat Kuruchittham, and Patrick L Remington. "The Wisconsin Brest Cancer Epidemiology Model." *Journal of the National Cancer Institute Monographs* 36 (2006): 37-47.

35. Gail, Michael H, et al. "Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually." *Journal of the National Cancer Institute* 81, no. 24 (1989): 1879-1886.

36. Geer Mountain Software Corporation. *Stat::Fit Distribution Fitting Software.* 2012. http://www.geerms.com/ (accessed January 15, 2011).

37. Gierisch, Jennifer M., Jo Anne Earp, Noel T. Brewer, and Barbara K. Rimer. "Longitudinal Predictors of Nonadherence to Maintenance of Mammography." *Cancer Epidemiology, Biomarkers & Prevention* 19 (March 2010): 1103-1111.

38. Griffiths, J D, M Price-Lloyd, M Smithies, and J E Williams. "Modelling the Requirement for Supplementary Nurses in an Intensive Care Unit." *Journal of the Operational Research Society* 56 (2005): 126-133.

39. Gutwein, Luke G., et al. "Utilization of Minimally Invasive Breast Biopsy for the Evaluation of Suspicious Breast Lesions." *The American Journal of Surgery*, 2011: 1-6.

40. Harper, P R, and A K Shahani. "Modelling for the Planning and Management of Bed Capacities in Hostpitals." *Journal of the Operational Research Society* 53 (2002): 11-18.

41. Hoel, P. G. *Introduction to Mathmatical Statistics.* 5th Edition. New York: Wiley, 1984.

42. Holmes, Chris E, and Muss B Hyman. "Diagnosis and Treatment of Breast Cancer in the Elderly." *CA Cancer Journal for Clinicians* 53, no. 4 (2003): 227-244.

43. Howden, Linday M., and Julie A. Meyer. "2010 Census Data." *Census.gov.* May 1, 2011. http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf (accessed September 2, 2011).

44. Humphrey, D. G., and James R. Wilson. "A Revised Simplex Search Procedure for Stochastic Simulation Response Surface Optimization." *INFORMS Journal on Computing* 12, no. 4 (2000): 272-283.

45. Humphrey, Linda L, Mark Helfand, Benjamin Chan, and Steven H Woolf. "Breast Cancer Screening: A Summary of the Evidence for the U.S. Preventative Services Task Force." *Annals of Internal Medicine* 137, no. 5-1 (2002): 347-360.

46. Irizarry, M., A. de los, James R. Wilson, and J. Trevino. "A Flexible Simulation Tool for Manufacturing Cell Design, II: Response Surface Analysis and Case Study." *IIE Transactions* 33, no. 10 (2001): 837-846.

47. Jun, J B, S H Jacobson, and J R Swisher. "Application of Discrete-Event Simulation in Health Care Clinics: A Survey." *Journal of the Operational Research Society* 50 (1999): 109-123.

48. Kelton, W David, Randall P Sadowski, and David T Sturrock. *Simulation with Arena.* New York, New Yotk: The McGraw-Hill Companies, 2007.

49. Kirch, R.L. A, and M Klein. "Surveillance Schedules for Medical Examinations." *Management Science* 20, no. 10 (June 1974): 1403-1409.

50. Kirch, Roberta L.A., and Morton Klein. "Examination Schedules for Breast Cancer." *Cancer* 33 (1974): 1444-1450.

51. Kirkwood, Craig W. "System Dynamics Resources." *System Dynamics Methods: A Quick Introduction.* Arizona State University. April 4, 1998. http://www.public.asu.edu/~kirkwood/sysdyn/SDIntro/SDIntro.htm (accessed July 5, 2010).

52. Kuhl, M. E., J. S. Ivy, E. K. Lada, N. M. Steiger, M. A. Wagner, and J. R. Wilson. "Univariate Input Models for Stochastic Simulation." *Journal of Simulation* 4 (2010): 81-97.

53. Lada, Emily K., and James R. Wilson. "A Wavelet-Based Spectral Procedure for Steady-State Simulation Analysis." *European Journal of Operational Research* 174 (2006): 1969-1801.

54. Lada, Emily K., Natalie M. Steiger, and James R. Wilson. "SBatch: A Spaced Batch Means Procedure for Steady-State Simulation Analysis." *Journal of Simulation* 2 (2008): 170-185.

55. Lane, DC, C Monefedlt, and JV Rosenhead. "Looking in the Wrong Place for Healthcare Improvements: A System Dynamics Study of an Accident and Emergency Department." *Journal of the Operational Research Society*, 2000: 518-531.

56. Lee, S., James R. Wilson, and M. M. Crawford. "Modeling and Simulation of a Nonhomogeneous Poisson Process Having Cyclical Behavior." *Communication in Statistics - Simulation and Computation* 20 (1991): 777-809.

57. Lee, Sandra, and Marvin Zelen. "Early Detection of Disease and Scheduling of Screening Examinations." *Statistical Methods in Medical Research* 13 (2004): 4433-456.

58. Louwman, W.J., J.C.M Vulto, R.H.A. Verhoeven, G.A.P. Nieuwenhuijzen, J.W.W. Coebergh, and A.C. Voogd. "Clinical Epidemiology of Breast Cancer in the Elderly." *European Journal of Cancer* 43 (2007): 2242-2252.

59. Maillart, Lisa M, Julie Simmons Ivy, Scott Ransom, and Kathleen Diehl. "Assessing Dynamic Breast Cancer Screening Policies." *Operations Research* 56, no. 6 (December 2008): 1411-1427.

60. Mandelblatt, Jeanne S, Mary E Wheat, Mark Monane, Rebecca Moshief, James P Hollenberg, and Jian Tang. "Breast Cancer Screening for Elderly Women with and without Comorbid Conditions." *Annals of Internal Medicine* 116 (1992): 722-730.

61. McManus, M J, and C Waelsch. "DNA Synthesis of Benign Human Breast Tumors in Untreated "Nude" Mouse: An In Vivo Model to Study Hormonal Influences on Growth of Human Breast Tissues." *Cancer* 25 (1980): 2160-2165.

62. Michaelson, James S, Elkan Halpern, and Daniel B Kopans. "Breast Cancer: Computer Simulation Method for Estimating Optimal Intervals for Screening." *Radiology*, 1999: 551-560.

63. Morecroft, John, and Stewart Robinson. "Explaining Puzzling Dynamics: Comparing the Use of System Dynamics and Discrete-Event Simulation." 2004.

64. Morecroft, John, and Stewart Robinson. "Explaining Puzzling Dynamics: Comparing the Use of System Dynamics and Discrete-Event Simulation." *Proceedings of the 23rd International Conference of the System Dynamics Society.* 2005. 1-32.

65. Murray, Talisa Marie. "Examining Intervention Schemes Through Simulation Modeling to Improve Access to Colorectal Cancer Care." Raleigh, 2004.

66. National Cancer Institute. *Cancer Query Systems.* 1970-2007. www.seer.cancer.gov (accessed March 21, 2011).

67. National Cancer Institute. "Costs of Cancer Care." *Cancer Trends Progress Report - 2009/2010 Update.* April 15, 2010.

68. National Institute of Health. "Menopause Age Page." *nih.gov.* April 20, 2010. http://www.nia.nih.gov/healthinformation/publications/menopause.htm (accessed December 12, 2009).

69. Norton, Larry. "A Gompertzian Model of Human Breast Cancer Growth." *Cancer Research* 48 (December 1988): 7067-7071.

70. Ozekici, Suleyman, and Stanley R Pliska. "Optimizing Scheduling of Inspections: A Delayed Markov Model with False Positives and Negatives." *Operations Research* 39, no. 2 (1991): 261-273.

71. Parmigiani, Giovanni, and Yu Shen. "Optimization of Breast Cancer Screening Modalities." *Cancer Epidemiology, Biomarkers, and Prevention*, 2003: Working Paper.

72. Pearlman, A W. "Breast Cancer: Influence of Growth Rate on Prognosis and Treatment Evaluation." *Cancer* 38 (1976): 1826-1833.

73. Peer, Petronela, Jos van Dijck, Jan Hendriks, Roland Holland, and Andre Verbeek. "Age-Dependent Growth Rate of Primary Breast Cancer." *Cancer* 71, no. 11 (1993): 3547-3551.

74. Pesce, C., and R. Colacino. "Morphometry of the Breast Fibroadenoma." *Pathology Research Practicum* 1986 (1986): 718-720.

75. Plevritis, Sylvia K, Peter Salzman, Bronislava M Sigal, and Peter W Glynn. "A Natural History Model of Stage Progression Applied to Breast Cancer." *Statistics in Medicine* 26 (2007): 581-595.

76. Rae-Venter, B, and L M Reid. "Growth of Breast Cancer Carcinomas in Nude Mice and Subsequent Establishmnet in Tissue Culture." *Cancer Research* 40 (1980): 95-100.

77. Resnick, Barbara, and Sandra W McLeskey. "Cancer Screening Across the Age Continuum." *The American Journal of Managed Care* 14, no. 5 (2008): 267-276.

78. Roberts, Stephen, Lijun Wang, Robert Klien, Reid Ness, and Robert Dittus. "Development of a Simulation Model of Colorectal Cancer." *ACM Transaction on Modeling and Computer Simulation* 18, no. 1 (December 2007).

79. Rosenberg, Marjorie A. "Competeting Risks to Breast Cancer Mortality." *Journal of the National Cancer Institute Monographs* 36 (2006): 15-25.

80. Royston, Patrick. "A Toolkit for Testing for Non-Normality in Complete and Censored Samples." *The Statistician* 42 (1993): 37-43.

81. Royston, Patrick. "Approximating the Shaprio-Wilk W-test for Non-Normality." *Journal of Statistics and Computing* 2 (1992): 117-119.

82. Royston, Patrick. "Corrigendum: A Toolkit for Testing for Non-Normality in Complete and Censored Samples." *The Statistician* 48, no. 1 (1999): 158.

83. Sanchez, SM, DM Ferrin, T Ogazon, and JA Sepulveda. "Emerging Issues in Healthcare Simulation." *Proccedings of the 2000 Winter Simulation Conference.* 2000.

84. Schwartz, Michael. "A Mathematical Model Used to Analyze Cancer Screening Strategies." *Operation Research*, 1978: 937-955.

85. Shumate, Seth D, and Magda El-Shenawee. "Computational Model of Breast Cancer Tumor Growth." *IEEE Region 5 Technical Conference.* Fayettville: IEEE, 2007. 13-15.

86. Smith, Robert A, Debbie Saslow, Kimberly Andrews Sawyer, Wylie Burke, and Mary E Costanza. "American Cancer Society Guidelines for Breast Cancer Screening: Update 2003." *CA A Cancer Journal for Clinicians* 53, no. 3 (2003): 141-169.

87. Social Security Administration. *Period Life Tables.* USA.gov. April 19, 2010. http://www.ssa.gov/OACT/STATS/table4c6.html (accessed November 2009).

88. Speer, John F, Victor E Petrosky, Michael W Retsky, and Robert H Wardwell. "A Stochastic Numerical Model of Breast Cancer Growth that Simulates Clinical Data." *Cancer Research* 44 (September 1984): 4124-4130.

89. Standridge, Charles R. "A Tutorial On Simulation in Health Care: Applications and Issues." *Proceedings of the 1999 Winter Simulation Conference*, 1999: 49-55.

90. Steiger, Natalie M., Emily K. Lada, James R. Wilson, Jeffrey A. Joines, Christos Alexopoulos, and David Goldsman. "ASAP3: A Batch Means Procedure for Steady-State Simulation Analysis." *ACM Transactions on Modeling and Computer Simulation* 15, no. 1 (January 2005): 39-73.

91. Sterman, John D. *Business Dynamics: Systems Thinking and Modeling for a Complex World.* Mcgraw-Hill, 200.

92. Surborne, A, and L Norton. "Kinetic Concepts in the Treatment of Breast Cancer." *Annals of NY Academic Science* 698 (1993): 48-62.

93. Sweester, A. "A Comparison of System Dynamics and Discrete Event Simulation." *Proceedings of the 17th International Confernce of the System Dynamics Society.* Wellington: System Dynamics Society, 1999.

94. Tafazzoli, Ali Yazdi. "A Comparison of Screening Methods for Colorectal Cancer." Masters Thesis, Industrial and Systems Engineering, North Carolina State University, Raleigh, 2004.

95. Tafazzoli, Ali, Stephen Roberts, Robert Klein, Reid Ness, and Robert Dittus. "Probabilistic Cost-Effectiveness Comparison of Screening Strategies for Colorectal Cancer." *ACM Transactions on Modeling and Computer Simulation* 19, no. 2 (2009).

96. Tako, A A, and S Robinson. "Comparing Discrete-Event Simulation and System Dynamics: Users Perceptions." *Journal of the Operational Research Society*, 2008: 1-17.

97. Tako, Antuela A, and Stewart Robinson. "Model building in System Dynamics and Discrete-Event Simulation: A Quantitative Comparison." *SystemsDynamics.org.* 2008.

98. Tosteson, Anna N.A., et al. "Cost-Effectiveness of Digital Mammography Breast Cancer Screening." *Annals of Internal Medicine* 148 (2008): 1-10.

99. U.S. Census Bureau. "Estimates Data." *Population Estimates.* 2009. http://www.census.gov/popest/national/asrh/NC-EST2009-asrh.html (accessed Novermber 11, 2010).

100. U.S. Department of Health and Human Services. *Physician Fee Schedule Search.* March 2012. http://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx (accessed May 10, 2011).

101. U.S. Department of Labor. *Consumer Price Indexes: CPI Databases.* March 2012. http://www.bls.gov/cpi/data.htm (accessed May 7, 2011).

102. Van der Linden, H. C., J. P. Baak, A. W. Smeulders, C. J. Meyer, and J. Lindeman. "Morphology of Breast Cancer: Comparison of the Primary Tumors and the Auxillary Lymph Node Metastases." *Pathology Research Practicum* 181 (1986): 236-242

103. Venkateswaran, J, and Y J Son. "Hybrid System Dynamic—Discrete Event Simulation-Based Architecture for Hierarchical Production Planning." *International Journal of Production Research* (Taylor & Francis Group) 43, no. 20 (October 2005): 4397-4429.

104. von Neumann, John. "Distribution of the Ratio of the Mean Square Succesive Difference to the Variance." *The Annals of Mathematical Statistics* 12, no. 4 (1941): 153-162.

105. Weiss, Marissa. *Understanding Breast Cancer.* November 26, 2008. http://www.breastcancer.org/symptoms/understand_bc/ (accessed March 3, 2009).

106. World Bank, World Development Indicators. *GDP Per Capita.* December 12, 2011. http://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&ctype=l&strail=false&bcs=d&nselm=h&met_y=ny_gdp_pcap_cd&scale_y=lin&ind_y=false&rdim=country&idim=country:USA&ifdim=country&tstart=284756400000&tend=1293080400000&hl=en&dl=en&q=us+gdp+per+capital (accessed December 23, 2011).

107. World Health Organization (WHO). "The World Health Report 2002 – Reducing Risks, Promoting Healthy Life." Geneva, 2002, 108-109.

"Data collection for this work was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium co-operative agreement (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040). The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: http://www.breastscreening.cancer.gov/work/acknowledgement.html. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. We thank the BCSC investigators, participating women, mammography facilities, and radiologists for

the data they have provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/."

APPENDICIES

Appendix A: SAS Code for Barlow Risk Model

```
data risk;
infile 'Risk Model Data set.txt' ;
input menopaus agegrp density race Hispanic bmi agefirst nrelbc brstproc
lastmamm surgmeno hrt invasive cancer training count ;
run;
/*
The logistic regression models below reproduce the odds ratios and
confidence intervals
shown in the article.
*/;
title 'Premenopausal Model - Table 3';
proc logistic data=risk descending simple ;
class agegrp density race Hispanic bmi agefirst nrelbc brstproc lastmamm
/ ref=first param=ref;
model cancer = agegrp density race Hispanic bmi agefirst nrelbc brstproc
lastmamm / nodesignprint NCONCORDBIN=20000;
freq count;
where (menopaus=0);
run;
title 'Postmenopausal Model - Table 5';
proc logistic data=risk descending simple ;
class agegrp density race Hispanic bmi agefirst nrelbc brstproc lastmamm
surgmeno hrt / ref=first param=ref;
model cancer = agegrp density race Hispanic bmi agefirst nrelbc brstproc
lastmamm surgmeno hrt / nodesignprint NCONCORDBIN=20000;
freq count;
where (menopaus=1);
run;
```

Appendix B: SAS Results from Barlow Risk Model

Premenopausal Model - Table 3
18:05 Thursday, September 2, 2010

The LOGISTIC Procedure

Model Information

Data Set	WORK.RISK
Response Variable	cancer
Number of Response Levels	2
Frequency Variable	count
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	52869
Number of Observations Used	52869
Sum of Frequencies Read	568215
Sum of Frequencies Used	568215

Response Profile

Ordered Value	cancer	Total Frequency
1	1	1726
2	0	566489

Probability modeled is cancer=1.

Frequency Distribution of Class Variables

		cancer		
Class	Value	1	0	Total
agegrp	1	75	36968	37043
	2	539	238734	239273
	3	693	193891	194584
	4	419	96896	97315
density	1	19	18164	18183
	2	306	146415	146721
	3	704	200192	200896
	4	251	61162	61413
	9	446	140556	141002
race	1	1239	410495	411734
	2	98	29539	29637
	3	79	28389	28468
	4	14	4465	4479
	5	28	6669	6697
	9	268	86932	87200
Hispanic	0	1295	426303	427598
	1	85	35205	35290
	9	346	104981	105327
bmi	1	460	141956	142416
	2	212	69264	69476
	3	92	30911	31003
	4	55	19455	19510
	9	907	304903	305810
agefirst	0	470	156454	156924
	1	226	65288	65514
	2	253	69796	70049
	9	777	274951	275728
nrelbc	0	1234	426422	427656
	1	332	74774	75106
	2	18	2805	2823
	9	142	62488	62630
brstproc	0	1269	448241	449510
	1	378	81006	81384
	9	79	37242	37321

lastmamm	0	1176	395507	396683
	1	47	9507	9554
	9	503	161475	161978

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	23458.937	23118.356
SC	23470.187	23444.614
-2 Log L	23456.937	23060.356

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	396.5809	28	<.0001
Score	404.6273	28	<.0001
Wald	391.6477	28	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
***agegrp**	3	129.5800	**<.0001**
***density**	4	98.5752	**<.0001**
race	5	4.0314	0.5449
Hispanic	2	7.0010	0.0302
bmi	4	1.2663	0.8671
agefirst	3	6.0509	0.1092
***nrelbc**	3	59.3483	**<.0001**
***brstproc**	2	43.1415	**<.0001**
lastmamm	2	10.2609	0.0059

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.6863	0.2686	818.5758	<.0001
agegrp	2	1	0.2085	0.1241	2.8233	0.0929
agegrp	3	1	0.6474	0.1229	27.7278	<.0001
agegrp	4	1	0.8808	0.1270	48.0840	<.0001
density	2	1	0.7078	0.2372	8.9077	0.0028
density	3	1	1.2307	0.2342	27.6214	<.0001
density	4	1	1.3899	0.2405	33.3894	<.0001
density	9	1	1.2160	0.2371	26.3058	<.0001
race	2	1	0.0379	0.1095	0.1195	0.7296
race	3	1	0.0113	0.1187	0.0091	0.9239
race	4	1	0.1169	0.2711	0.1859	0.6663
race	5	1	0.3770	0.1940	3.7769	0.0520
race	9	1	0.000701	0.0913	0.0001	0.9939
Hispanic	1	1	-0.1339	0.1166	1.3193	0.2507
Hispanic	9	1	0.1709	0.0813	4.4166	0.0356
bmi	2	1	0.0334	0.0842	0.1569	0.6920
bmi	3	1	0.0864	0.1163	0.5520	0.4575
bmi	4	1	0.1119	0.1457	0.5898	0.4425
bmi	9	1	-0.00884	0.0714	0.0154	0.9014
agefirst	1	1	0.1502	0.0821	3.3502	0.0672
agefirst	2	1	0.1738	0.0799	4.7307	0.0296
agefirst	9	1	0.0802	0.0735	1.1913	0.2751
nrelbc	1	1	0.4297	0.0625	47.1907	<.0001
nrelbc	2	1	0.7454	0.2386	9.7552	0.0018
nrelbc	9	1	-0.1111	0.0940	1.3957	0.2375
brstproc	1	1	0.3721	0.0597	38.8367	<.0001
brstproc	9	1	-0.1712	0.1213	1.9934	0.1580
lastmamm	1	1	0.3970	0.1504	6.9708	0.0083
lastmamm	9	1	0.1137	0.0546	4.3401	0.0372

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
agegrp	2 vs 1	1.232	0.966	1.571
agegrp	3 vs 1	1.911	1.501	2.431
agegrp	4 vs 1	2.413	1.881	3.095
density	2 vs 1	2.030	1.275	3.230
density	3 vs 1	3.424	2.164	5.418
density	4 vs 1	4.015	2.505	6.433
density	9 vs 1	3.374	2.120	5.369
race	2 vs 1	1.039	0.838	1.287
race	3 vs 1	1.011	0.801	1.276
race	4 vs 1	1.124	0.661	1.912
race	5 vs 1	1.458	0.997	2.132
race	9 vs 1	1.001	0.837	1.197
Hispanic	1 vs 0	0.875	0.696	1.099
Hispanic	9 vs 0	1.186	1.012	1.391
bmi	2 vs 1	1.034	0.877	1.219
bmi	3 vs 1	1.090	0.868	1.369
bmi	4 vs 1	1.118	0.841	1.488
bmi	9 vs 1	0.991	0.862	1.140
agefirst	1 vs 0	1.162	0.989	1.365
agefirst	2 vs 0	1.190	1.017	1.391
agefirst	9 vs 0	1.084	0.938	1.251
nrelbc	1 vs 0	1.537	1.359	1.737
nrelbc	2 vs 0	2.107	1.320	3.364
nrelbc	9 vs 0	0.895	0.744	1.076
brstproc	1 vs 0	1.451	1.291	1.631
brstproc	9 vs 0	0.843	0.664	1.069
lastmamm	1 vs 0	1.487	1.108	1.997
lastmamm	9 vs 0	1.120	1.007	1.247

Association of Predicted Probabilities and Observed Responses

Percent Concordant	62.9	Somers' D	0.271
Percent Discordant	35.9	Gamma	0.274
Percent Tied	1.2	Tau-a	0.002
Pairs	977760014	c	0.635

The LOGISTIC Procedure

Model Information

Data Set	WORK.RISK
Response Variable	cancer
Number of Response Levels	2
Frequency Variable	count
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	226021
Number of Observations Used	226021
Sum of Frequencies Read	1642824
Sum of Frequencies Used	1642824

Response Profile

Ordered Value	cancer	Total Frequency
1	1	9300
2	0	1633524

Probability modeled is cancer=1.

Frequency Distribution of Class Variables

		cancer		
Class	Value	1	0	Total
agegrp	3	361	126407	126768
	4	1022	267633	268655
	5	1795	332337	334132
	6	1576	261945	263521
	7	1467	230437	231904
	8	1420	201686	203106
	9	1089	144013	145102
	10	570	69066	69636
density	1	306	124171	124477
	2	2957	594402	597359
	3	2884	430174	433058
	4	434	64833	65267
	9	2719	419944	422663
race	1	7049	1202881	1209930
	2	315	65969	66284
	3	457	81491	81948
	4	61	18857	18918
	5	70	14208	14278
	9	1348	250118	251466
Hispanic	0	7091	1202244	1209335
	1	439	103243	103682
	9	1770	328037	329807
bmi	1	1864	330391	332255
	2	1408	235870	237278
	3	643	104955	105598
	4	334	53664	53998
	9	5051	908644	913695
agefirst	0	3093	543202	546295
	1	485	71862	72347
	2	798	121014	121812
	9	4924	897446	902370
nrelbc	0	6421	1192870	1199291
	1	1502	202067	203569
	2	121	12209	12330
	9	1256	226378	227634

brstproc	0	6023	1157091	1163114
	1	2345	318635	320980
	9	932	157798	158730
lastmamm	0	6914	1259229	1266143
	1	204	21006	21210
	9	2182	353289	355471
surgmeno	0	4316	713650	717966
	1	2180	425152	427332
	9	2804	494722	497526
hrt	0	3985	725211	729196
	1	3950	679400	683350
	9	1365	228913	230278

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	114788.58	113018.20
SC	114800.90	113473.74
-2 Log L	114786.58	112944.20

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1842.3888	36	<.0001
Score	1779.7483	36	<.0001
Wald	1728.1325	36	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
***agegrp**	7	745.8279	**<.0001**
***density**	4	464.3562	**<.0001**
***race**	5	40.8176	**<.0001**
***Hispanic**	2	36.0040	**<.0001**
***bmi**	4	65.0672	**<.0001**
***agefirst**	3	27.7638	**<.0001**
***nrelbc**	3	122.0142	**<.0001**
***brstproc**	2	111.6793	**<.0001**
***lastmamm**	2	99.3076	**<.0001**
***surgmeno**	2	38.4138	**<.0001**
***hrt**	2	52.5079	**<.0001**

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.0713	0.0852	6894.3004	<.0001
agegrp	4	1	0.2830	0.0614	21.2180	<.0001
agegrp	5	1	0.6743	0.0581	134.6825	<.0001
agegrp	6	1	0.8196	0.0589	193.9029	<.0001
agegrp	7	1	0.9034	0.0594	231.3044	<.0001
agegrp	8	1	1.0260	0.0598	294.7886	<.0001
agegrp	9	1	1.1086	0.0617	323.1100	<.0001
agegrp	10	1	1.2023	0.0683	309.6190	<.0001
density	2	1	0.7377	0.0603	149.5244	<.0001
density	3	1	1.0799	0.0609	314.4024	<.0001
density	4	1	1.1482	0.0759	228.8218	<.0001
density	9	1	1.0449	0.0614	289.5403	<.0001
race	2	1	-0.2251	0.0594	14.3654	0.0002
race	3	1	0.0935	0.0497	3.5432	0.0598
race	4	1	-0.6108	0.1302	22.0054	<.0001
race	5	1	-0.0269	0.1214	0.0493	0.8243

race	9	1	-0.00037	0.0421	0.0001	0.9929
Hispanic	1	1	-0.3072	0.0516	35.4777	<.0001
Hispanic	9	1	-0.0641	0.0371	2.9780	0.0844
bmi	2	1	0.1337	0.0357	14.0226	0.0002
bmi	3	1	0.2463	0.0465	28.0848	<.0001
bmi	4	1	0.3827	0.0606	39.8984	<.0001
bmi	9	1	0.0295	0.0319	0.8565	0.3547
agefirst	1	1	0.1871	0.0494	14.3321	0.0002
agefirst	2	1	0.1637	0.0406	16.2456	<.0001
agefirst	9	1	0.0203	0.0293	0.4826	0.4872
nrelbc	1	1	0.2722	0.0289	88.6320	<.0001
nrelbc	2	1	0.5080	0.0927	30.0578	<.0001
nrelbc	9	1	-0.0488	0.0351	1.9396	0.1637
brstproc	1	1	0.2620	0.0248	111.6785	<.0001
brstproc	9	1	0.0585	0.0411	2.0243	0.1548
lastmamm	1	1	0.5241	0.0717	53.4358	<.0001
lastmamm	9	1	0.1840	0.0251	53.9037	<.0001
surgmeno	1	1	-0.1701	0.0274	38.4115	<.0001
surgmeno	9	1	-0.0603	0.0297	4.1132	0.0425
hrt	1	1	0.1714	0.0240	51.1867	<.0001
hrt	9	1	0.1244	0.0376	10.9739	0.0009

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
agegrp	4 vs 3	1.327	1.177	1.497
agegrp	5 vs 3	1.963	1.751	2.199
agegrp	6 vs 3	2.270	2.022	2.547
agegrp	7 vs 3	2.468	2.197	2.773
agegrp	8 vs 3	2.790	2.482	3.137
agegrp	9 vs 3	3.030	2.685	3.420
agegrp	10 vs 3	3.328	2.911	3.805
density	2 vs 1	2.091	1.858	2.354
density	3 vs 1	2.944	2.613	3.318
density	4 vs 1	3.152	2.717	3.658
density	9 vs 1	2.843	2.521	3.207
race	2 vs 1	0.798	0.711	0.897
race	3 vs 1	1.098	0.996	1.210
race	4 vs 1	0.543	0.421	0.701
race	5 vs 1	0.973	0.767	1.235
race	9 vs 1	1.000	0.920	1.086
Hispanic	1 vs 0	0.736	0.665	0.814
Hispanic	9 vs 0	0.938	0.872	1.009
bmi	2 vs 1	1.143	1.066	1.226
bmi	3 vs 1	1.279	1.168	1.401
bmi	4 vs 1	1.466	1.302	1.651
bmi	9 vs 1	1.030	0.968	1.096
agefirst	1 vs 0	1.206	1.094	1.328
agefirst	2 vs 0	1.178	1.088	1.275
agefirst	9 vs 0	1.021	0.964	1.081
nrelbc	1 vs 0	1.313	1.240	1.389
nrelbc	2 vs 0	1.662	1.386	1.993
nrelbc	9 vs 0	0.952	0.889	1.020
brstproc	1 vs 0	1.300	1.238	1.364
brstproc	9 vs 0	1.060	0.978	1.149
lastmamm	1 vs 0	1.689	1.468	1.944
lastmamm	9 vs 0	1.202	1.144	1.263
surgmeno	1 vs 0	0.844	0.799	0.890
surgmeno	9 vs 0	0.941	0.888	0.998
hrt	1 vs 0	1.187	1.133	1.244
hrt	9 vs 0	1.132	1.052	1.219

Association of Predicted Probabilities and Observed Responses

Percent Concordant	62.1	Somers' D	0.248
Percent Discordant	37.3	Gamma	0.250
Percent Tied	0.6	Tau-a	0.003
Pairs	15191773200	c	0.624

Appendix C: Raw Data for Graphical Results from the Natural History Model

C.1 Breast Cancer Incidence

Table 6-1. Incidence Counts CI Data

Invasive Caner Incidence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	121.5	8.07	113.4	129.6
2002	122.2	11.11	111.1	133.3
2003	118.2	6.55	111.7	124.8
2004	121.8	9.24	112.6	131.0
2005	116.0	5.23	110.8	121.2
2006	124.4	6.36	118.0	130.8
2007	121.6	5.68	115.9	127.3
2008	117.1	3.43	113.7	120.5
2009	123.8	7.44	116.4	131.2
2010	128.0	6.89	121.1	134.9
2011	130.0	7.50	122.5	137.5
2012	134.6	7.47	127.1	142.1
2013	133.0	9.21	123.8	142.2
2014	133.4	7.92	125.5	141.3
2015	137.7	9.71	128.0	147.4
2016	138.1	6.50	131.6	144.6
2017	143.1	4.73	138.4	147.8
2018	154.0	11.02	143.0	165.0
2019	157.0	11.64	145.4	168.6
2020	158.4	9.56	148.8	168.0

DCIS Cancer Incidence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	26.3	2.52	23.8	28.8
2002	31.5	5.74	25.8	37.2
2003	29.5	6.58	22.9	36.1
2004	33.7	2.84	30.9	36.5
2005	25.0	2.90	22.1	27.9
2006	27.1	4.57	22.5	31.7
2007	29.5	3.56	25.9	33.1
2008	33.5	2.87	30.6	36.4
2009	30.1	4.86	25.2	35.0
2010	27.0	4.83	22.2	31.8
2011	28.1	4.19	23.9	32.3
2012	32.8	4.78	28.0	37.6
2013	33.1	3.92	29.2	37.0
2014	32.4	3.74	28.7	36.1
2015	35.6	2.88	32.7	38.5
2016	39.9	4.76	35.1	44.7
2017	33.6	3.94	29.7	37.5
2018	36.5	5.96	30.5	42.5
2019	36.9	4.16	32.7	41.1
2020	39.6	3.18	36.4	42.8

Table 6-2. Incidence Percentages CI Data

Invasive Cancer Incidence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.5903	0.04	0.5503	0.6303
2002	0.5906	0.05	0.5406	0.6406
2003	0.5671	0.03	0.5371	0.5971
2004	0.5792	0.04	0.5392	0.6192
2005	0.5455	0.02	0.5255	0.5655
2006	0.5777	0.03	0.5477	0.6077
2007	0.5577	0.03	0.5277	0.5877
2008	0.5289	0.01	0.5189	0.5389
2009	0.5496	0.03	0.5196	0.5796
2010	0.5578	0.03	0.5278	0.5878
2011	0.5554	0.03	0.5254	0.5854
2012	0.5628	0.03	0.5328	0.5928
2013	0.5438	0.04	0.5038	0.5838
2014	0.5320	0.03	0.5020	0.5620
2015	0.5349	0.04	0.4949	0.5749
2016	0.5222	0.03	0.4922	0.5522
2017	0.5254	0.02	0.5054	0.5454
2018	0.5479	0.04	0.5079	0.5879
2019	0.5410	0.04	0.5010	0.5810
2020	0.5265	0.03	0.4965	0.5565

DCIS Cancer Incidence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.1278	0.01	0.1178	0.1378
2002	0.1523	0.03	0.1223	0.1823
2003	0.1415	0.03	0.1115	0.1715
2004	0.1602	0.01	0.1502	0.1702
2005	0.1176	0.01	0.1076	0.1276
2006	0.1259	0.02	0.1059	0.1459
2007	0.1352	0.02	0.1152	0.1552
2008	0.1513	0.01	0.1413	0.1613
2009	0.1337	0.02	0.1137	0.1537
2010	0.1176	0.02	0.0976	0.1376
2011	0.1201	0.02	0.1001	0.1401
2012	0.1371	0.02	0.1171	0.1571
2013	0.1354	0.02	0.1154	0.1554
2014	0.1293	0.02	0.1093	0.1493
2015	0.1383	0.01	0.1283	0.1483
2016	0.1509	0.02	0.1309	0.1709
2017	0.1234	0.01	0.1134	0.1334
2018	0.1298	0.02	0.1098	0.1498
2019	0.1271	0.01	0.1171	0.1371
2020	0.1316	0.01	0.1216	0.1416

Table 6-3. Incidence Rates Per 100,000 CI Data

Invasive Cancer Incidence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	590.3	40.0	550.3	630.3
2002	590.6	50.0	540.6	640.6
2003	567.1	30.0	537.1	597.1
2004	579.2	40.0	539.2	619.2
2005	545.5	20.0	525.5	565.5
2006	577.7	30.0	547.7	607.7
2007	557.7	30.0	527.7	587.7
2008	528.9	10.0	518.9	538.9
2009	549.6	30.0	519.6	579.6
2010	557.8	30.0	527.8	587.8
2011	555.4	30.0	525.4	585.4
2012	562.8	30.0	532.8	592.8
2013	543.8	40.0	503.8	583.8
2014	532.0	30.0	502.0	562.0
2015	534.9	40.0	494.9	574.9
2016	522.2	30.0	492.2	552.2
2017	525.4	20.0	505.4	545.4
2018	547.9	40.0	507.9	587.9
2019	541.0	40.0	501.0	581.0
2020	526.5	30.0	496.5	556.5

DCIS Cancer Incidence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	127.8	10.0	117.8	137.8
2002	152.3	30.0	122.3	182.3
2003	141.5	30.0	111.5	171.5
2004	160.2	10.0	150.2	170.2
2005	117.6	10.0	107.6	127.6
2006	125.9	20.0	105.9	145.9
2007	135.2	20.0	115.2	155.2
2008	151.3	10.0	141.3	161.3
2009	133.7	20.0	113.7	153.7
2010	117.6	20.0	97.6	137.6
2011	120.1	20.0	100.1	140.1
2012	137.1	20.0	117.1	157.1
2013	135.4	20.0	115.4	155.4
2014	129.3	20.0	109.3	149.3
2015	138.3	10.0	128.3	148.3
2016	150.9	20.0	130.9	170.9
2017	123.4	10.0	113.4	133.4
2018	129.8	20.0	109.8	149.8
2019	127.1	10.0	117.1	137.1
2020	131.6	10.0	121.6	141.6

Table 6-4. SEER Breast Cancer Incidence Rates Per 100,000

SEER Invasive Cancer Incidence	
Year	**Mean**
2001	454.7
2002	439.6
2003	405.3
2004	405.0
2005	407.2
2006	399.2
2007	413.7
2008	418.1

SEER DCIS Incidence	
Year	**Mean**
2001	88.5
2002	86.4
2003	83.2
2004	83.8
2005	86.3
2006	82.3
2007	89.9
2008	94.5

C.2 Breast Cancer Prevalence

Table 6-5. Prevalence Counts CI Data

Invasive Caner Prevalence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	115.0	7.05	108.0	122.1
2002	223.2	15.02	208.2	238.2
2003	316.2	14.76	301.4	331.0
2004	404.1	20.03	384.1	424.1
2005	474.6	17.69	456.9	492.3
2006	539.0	14.07	524.9	553.1
2007	593.6	12.50	581.1	606.1
2008	638.8	11.98	626.8	650.8
2009	679.7	12.30	667.4	692.0
2010	718.3	10.63	707.7	728.9
2011	752.7	9.04	743.7	761.7
2012	774.8	12.00	762.8	786.8
2013	798.1	20.22	777.9	818.3
2014	820.5	23.21	797.3	843.7
2015	843.4	23.45	820.0	866.9
2016	853.5	25.95	827.6	879.5
2017	873.1	23.31	849.8	896.4
2018	896.7	19.44	877.3	916.1
2019	925.4	23.19	902.2	948.6
2020	953.2	23.08	930.1	976.3

DCIS Cancer Prevalence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	25.1	2.51	22.6	27.6
2002	53.7	6.67	47.0	60.4
2003	79.0	7.70	71.3	86.7
2004	108.3	6.11	102.2	114.4
2005	126.8	5.90	120.9	132.7
2006	145.0	8.53	136.5	153.5
2007	165.2	7.99	157.2	173.2
2008	186.9	8.88	178.0	195.8
2009	203.9	11.39	192.5	215.3
2010	217.5	15.86	201.6	233.4
2011	230.4	13.34	217.1	243.7
2012	246.8	12.67	234.1	259.5
2013	263.4	13.89	249.5	277.3
2014	275.8	14.52	261.3	290.3
2015	289.9	15.07	274.8	305.0
2016	307.8	15.18	292.6	323.0
2017	319.9	14.99	304.9	334.9
2018	332.9	14.83	318.1	347.7
2019	342.4	13.34	329.1	355.7
2020	358.3	14.27	344.0	372.6

Table 6-6. Prevalence Percentages CI Data

Invasive Caner Prevalence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.5587	0.03	0.5287	0.5887
2002	1.0787	0.07	1.0087	1.1487
2003	1.5170	0.07	1.4470	1.5870
2004	1.9218	0.10	1.8218	2.0218
2005	2.2320	0.08	2.1520	2.3120
2006	2.5031	0.07	2.4331	2.5731
2007	2.7222	0.06	2.6622	2.7822
2008	2.8858	0.06	2.8258	2.9458
2009	3.0179	0.06	2.9579	3.0779
2010	3.1302	0.05	3.0802	3.1802
2011	3.2154	0.04	3.1754	3.2554
2012	3.2399	0.05	3.1899	3.2899
2013	3.2633	0.08	3.1833	3.3433
2014	3.2726	0.09	3.1826	3.3626
2015	3.2759	0.09	3.1859	3.3659
2016	3.2272	0.09	3.1372	3.3172
2017	3.2054	0.08	3.1254	3.2854
2018	3.1907	0.06	3.1307	3.2507
2019	3.1874	0.08	3.1074	3.2674
2020	3.1677	0.08	3.0877	3.2477

DCIS Cancer Prevalence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.1220	0.01	0.112	0.132
2002	0.2595	0.03	0.2295	0.2895
2003	0.3790	0.04	0.3390	0.4190
2004	0.5150	0.03	0.485	0.545
2005	0.5963	0.03	0.5663	0.6263
2006	0.6734	0.04	0.6334	0.7134
2007	0.7575	0.04	0.7175	0.7975
2008	0.8442	0.04	0.8042	0.8842
2009	0.9053	0.05	0.8553	0.9553
2010	0.9478	0.07	0.8778	1.0178
2011	0.9842	0.06	0.9242	1.0442
2012	1.0319	0.05	0.9819	1.0819
2013	1.0769	0.06	1.0169	1.1369
2014	1.1001	0.06	1.0401	1.1601
2015	1.1260	0.06	1.066	1.186
2016	1.1639	0.06	1.1039	1.2239
2017	1.1745	0.05	1.1245	1.2245
2018	1.1846	0.05	1.1346	1.2346
2019	1.1793	0.05	1.1293	1.2293
2020	1.1907	0.05	1.1407	1.2407

Table 6-7. Prevalence Rates Per 100,000 CI Data

Invasive Caner Prevalence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	558.7	30.0	528.7	588.7
2002	1078.7	70.0	1008.7	1148.7
2003	1517.0	70.0	1447.0	1587.0
2004	1921.8	100.0	1821.8	2021.8
2005	2232.0	80.0	2152.0	2312.0
2006	2503.1	70.0	2433.1	2573.1
2007	2722.2	60.0	2662.2	2782.2
2008	2885.8	60.0	2825.8	2945.8
2009	3017.9	60.0	2957.9	3077.9
2010	3130.2	50.0	3080.2	3180.2
2011	3215.4	40.0	3175.4	3255.4
2012	3239.9	50.0	3189.9	3289.9
2013	3263.3	80.0	3183.3	3343.3
2014	3272.6	90.0	3182.6	3362.6
2015	3275.9	90.0	3185.9	3365.9
2016	3227.2	90.0	3137.2	3317.2
2017	3205.4	80.0	3125.4	3285.4
2018	3190.7	60.0	3130.7	3250.7
2019	3187.4	80.0	3107.4	3267.4
2020	3167.7	80.0	3087.7	3247.7

DCIS Cancer Prevalence				
Year	**Mean**	**HW**	**LL**	**UL**
2001	122.0	10.0	112.0	132.0
2002	259.5	30.0	229.5	289.5
2003	379.0	40.0	339.0	419.0
2004	515.0	30.0	485.0	545.0
2005	596.3	30.0	566.3	626.3
2006	673.4	40.0	633.4	713.4
2007	757.5	40.0	717.5	797.5
2008	844.2	40.0	804.2	884.2
2009	905.3	50.0	855.3	955.3
2010	947.8	70.0	877.8	1017.8
2011	984.2	60.0	924.2	1044.2
2012	1031.9	50.0	981.9	1081.9
2013	1076.9	60.0	1016.9	1136.9
2014	1100.1	60.0	1040.1	1160.1
2015	1126.0	60.0	1066.0	1186.0
2016	1163.9	60.0	1103.9	1223.9
2017	1174.5	50.0	1124.5	1224.5
2018	1184.6	50.0	1134.6	1234.6
2019	1179.3	50.0	1129.3	1229.3
2020	1190.7	50.0	1140.7	1240.7

Table 6-8. SEER Prevalence Percentages by Age in U.S. Population

Age-Adjusted Prevalence Percent (US 2000)	
Age	**Prevalence Rate**
65-69	4.62%
70-74	4.83%
75-79	5.08%
80-84	5.34%
85+	4.97%

C.3 Breast Cancer and Non Breast Cancer Deaths

Table 6-9. Death Counts CI Data

Cancer Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.0	0.00	0.0	0.0
2002	1.2	0.74	0.5	1.9
2003	6.0	1.43	4.6	7.4
2004	11.6	2.39	9.2	14.0
2005	17.3	3.46	13.8	20.8
2006	27.4	4.41	23.0	31.8
2007	30.2	4.03	26.2	34.2
2008	34.4	4.49	29.9	38.9
2009	42.8	3.86	38.9	46.7
2010	45.7	6.09	39.6	51.8
2011	47.4	4.24	43.2	51.6
2012	57.1	3.51	53.6	60.6
2013	56.2	5.31	50.9	61.5
2014	56.7	5.92	50.8	62.6
2015	58.6	6.08	52.5	64.7
2016	67.5	6.13	61.4	73.6
2017	64.9	4.44	60.5	69.3
2018	70.1	7.59	62.5	77.7
2019	67.2	4.26	62.9	71.5
2020	72.5	6.63	65.9	79.1

Non Breast Cancer Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	945.9	25.99	919.9	971.9
2002	965.5	26.77	938.7	992.3
2003	972.7	28.29	944.4	1001.0
2004	972.3	29.25	943.1	1001.6
2005	986.0	18.91	967.1	1004.9
2006	999.2	20.70	978.5	1019.9
2007	1026.1	24.53	1001.6	1050.6
2008	1024.8	19.23	1005.6	1044.0
2009	1048.0	31.34	1016.7	1079.3
2010	1035.6	21.82	1013.8	1057.4
2011	1046.6	17.30	1029.3	1063.9
2012	1080.5	28.05	1052.5	1108.6
2013	1077.9	23.76	1054.1	1101.7
2014	1087.2	26.18	1061.0	1113.4
2015	1106.4	19.80	1086.6	1126.2
2016	1123.4	16.98	1106.4	1140.4
2017	1134.7	23.78	1110.9	1158.5
2018	1144.5	22.17	1122.3	1166.7
2019	1169.0	13.79	1155.2	1182.8
2020	1190.7	26.39	1164.3	1217.1

Table 6-10. Death Percentages CI Data

Cancer Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.0000	0.00	0.0000	0.0000
2002	0.0058	0.00	0.0058	0.0058
2003	0.0288	0.01	0.0188	0.0388
2004	0.0552	0.01	0.0452	0.0652
2005	0.0814	0.02	0.0614	0.1014
2006	0.1273	0.02	0.1073	0.1473
2007	0.1384	0.02	0.1184	0.1584
2008	0.1554	0.02	0.1354	0.1754
2009	0.1900	0.02	0.1700	0.2100
2010	0.1992	0.03	0.1692	0.2292
2011	0.2025	0.02	0.1825	0.2225
2012	0.2388	0.01	0.2288	0.2488
2013	0.2298	0.02	0.2098	0.2498
2014	0.2261	0.02	0.2061	0.2461
2015	0.2277	0.02	0.2077	0.2477
2016	0.2552	0.02	0.2352	0.2752
2017	0.2382	0.02	0.2182	0.2582
2018	0.2495	0.03	0.2195	0.2795
2019	0.2315	0.01	0.2215	0.2415
2020	0.2410	0.02	0.2210	0.2610

Natural Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	4.5958	0.13	4.4658	4.7258
2002	4.6658	0.13	4.5358	4.7958
2003	4.6663	0.13	4.5363	4.7963
2004	4.6238	0.14	4.4838	4.7638
2005	4.6367	0.08	4.5567	4.7167
2006	4.6402	0.10	4.5402	4.7402
2007	4.7054	0.12	4.5854	4.8254
2008	4.6294	0.09	4.5394	4.7194
2009	4.6530	0.14	4.5130	4.7930
2010	4.5132	0.10	4.4132	4.6132
2011	4.4710	0.08	4.391	4.551
2012	4.5184	0.12	4.3984	4.6384
2013	4.4076	0.10	4.3076	4.5076
2014	4.3363	0.10	4.2363	4.4363
2015	4.2980	0.09	4.208	4.388
2016	4.2477	0.05	4.1977	4.2977
2017	4.1657	0.08	4.0857	4.2457
2018	4.0725	0.07	4.0025	4.1425
2019	4.0264	0.05	3.9764	4.0764
2020	3.9570	0.08	3.877	4.037

Table 6-11. Death Rates Per 100,000 CI Data

Breast Cancer Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	0.0	0.0	0.0	0.0
2002	5.8	0.0	5.8	5.8
2003	28.8	10.0	18.8	38.8
2004	55.2	10.0	45.2	65.2
2005	81.4	20.0	61.4	101.4
2006	127.3	20.0	107.3	147.3
2007	138.4	20.0	118.4	158.4
2008	155.4	20.0	135.4	175.4
2009	190.0	20.0	170.0	210.0
2010	199.2	30.0	169.2	229.2
2011	202.5	20.0	182.5	222.5
2012	238.8	10.0	228.8	248.8
2013	229.8	20.0	209.8	249.8
2014	226.1	20.0	206.1	246.1
2015	227.7	20.0	207.7	247.7
2016	255.2	20.0	235.2	275.2
2017	238.2	20.0	218.2	258.2
2018	249.5	30.0	219.5	279.5
2019	231.5	10.0	221.5	241.5
2020	241.0	20.0	221.0	261.0

Natural Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	4595.8	130.0	4465.8	4725.8
2002	4665.8	130.0	4535.8	4795.8
2003	4666.3	130.0	4536.3	4796.3
2004	4623.8	140.0	4483.8	4763.8
2005	4636.7	80.0	4556.7	4716.7
2006	4640.2	100.0	4540.2	4740.2
2007	4705.4	120.0	4585.4	4825.4
2008	4629.4	90.0	4539.4	4719.4
2009	4653.0	140.0	4513.0	4793.0
2010	4513.2	100.0	4413.2	4613.2
2011	4471.0	80.0	4391.0	4551.0
2012	4518.4	120.0	4398.4	4638.4
2013	4407.6	100.0	4307.6	4507.6
2014	4336.3	100.0	4236.3	4436.3
2015	4298.0	90.0	4208.0	4388.0
2016	4247.7	50.0	4197.7	4297.7
2017	4165.7	80.0	4085.7	4245.7
2018	4072.5	70.0	4002.5	4142.5
2019	4026.4	50.0	3976.4	4076.4
2020	3957.0	80.0	3877.0	4037.0

Table 6-12. SEER Death Rates Per 100,000 CI Data

SEER Breast Cancer Deaths				
Year	**Mean**	**HW**	**LL**	**UL**
2001	112.4	2.8	110.9	113.8
2002	112.1	2.8	110.7	113.5
2003	110.0	2.8	108.6	111.4
2004	106.9	2.8	105.5	108.3
2005	105.6	2.8	104.2	107.0
2006	103.2	2.8	101.9	104.6
2007	102.3	2.6	101.0	103.6
2008	100.7	2.6	99.4	102.0

SEER Natural Deaths	
Year	**Mean**
2001	4316.0
2002	4290.1
2003	4214.9
2004	4040.9
2005	4042.6
2006	3894.3
2007	3800.9
2008	3818.0

C.4 Population Size

Table 6-13. Population Size CI Data with Census Data for Comparison

Population Size (65+)				
Year	**Mean**	**HW**	**LL**	**UL**
2001	20582.0	0.00	20582.0	20582.0
2002	20692.5	29.27	20663.2	20721.8
2003	20844.9	38.97	20805.9	20883.9
2004	21029.5	54.90	20974.6	21084.4
2005	21264.7	52.22	21212.5	21316.9
2006	21534.7	66.45	21468.3	21601.2
2007	21807.8	73.58	21734.2	21881.4
2008	22137.2	71.48	22065.7	22208.7
2009	22523.3	82.33	22441.0	22605.6
2010	22947.7	87.05	22860.7	23034.8
2011	23409.5	78.78	23330.7	23488.3
2012	23913.5	72.26	23841.2	23985.8
2013	24456.2	87.26	24368.9	24543.5
2014	25070.9	89.67	24981.2	25160.6
2015	25744.3	104.28	25640.0	25848.6
2016	26446.1	108.54	26337.6	26554.6
2017	27237.7	126.90	27110.8	27364.6
2018	28102.3	137.59	27964.7	28239.9
2019	29033.6	140.99	28892.6	29174.6
2020	30090.7	149.53	29941.2	30240.2

Population Size (65+)					
Year	**Value**	**Value/1000**	**Simulation Mean**	**Diff**	**% Diff**
2001	20590658.0	20590.66	20582.0	8.7	0.042%
2002	20686951.0	20686.95	20692.5	-5.5	0.027%
2003	20838602.0	20838.60	20844.9	-6.3	0.030%
2004	20981785.0	20981.79	21029.5	-47.7	0.227%
2005	21180464.0	21180.46	21264.7	-84.2	0.398%
2006	21418004.0	21418.00	21534.7	-116.7	0.545%
2007	21736924.0	21736.92	21807.8	-70.9	0.326%
2008	22195107.0	22195.11	22137.2	57.9	0.261%
2009	22569111.0	22569.11	22523.3	45.8	0.203%
2010	22905024.0	22905.02	22947.7	-42.7	0.186%

Appendix D: Inferred Fitted Distributions from Natural History Model

D.1 Distribution of the Number of Months it Takes Tumors to Grow to Minimum Size Detectable by Mammography (All Invasive Cancers)

- Best Distribution: Lognormal
- Distribution Parameters: μ = 38.6, σ = 29.4, *Min (Offset)* = 4.0
- Moments of the Fitted Lognormal Distribution
  - μ = 42.5, σ^2 = 864.4, γ_1 = 2.74, γ_2 = 15.8
- Moments of the Data
  - μ = 43.1, σ^2 = 1137.2, γ_1 = 0.53, γ_2 = 0.045
- Chi-Squared p-value < 0.005
- Kolmogrov Smirinov p-value < 0.005

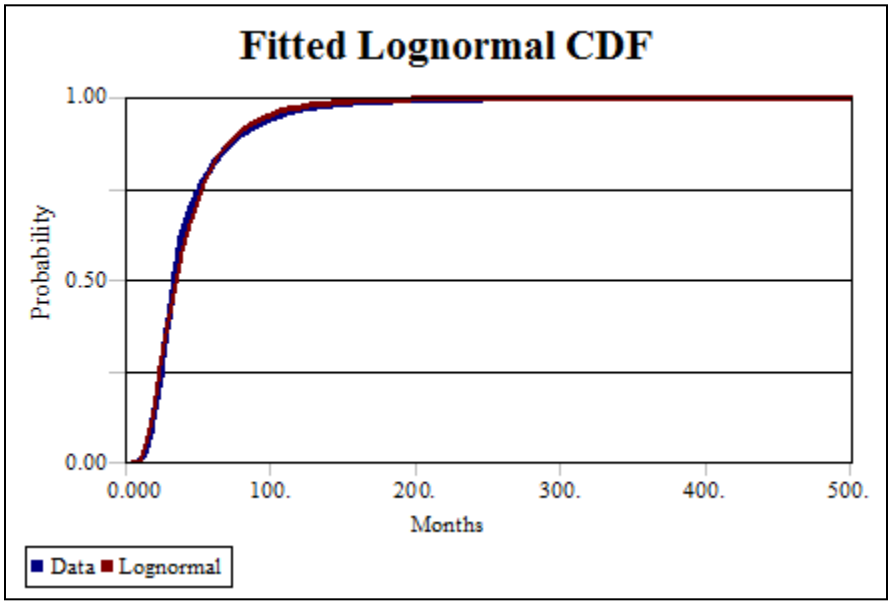


Figure 6-1. Months to Minimum Size Fitted Lognormal CDF vs. Empirical CDF

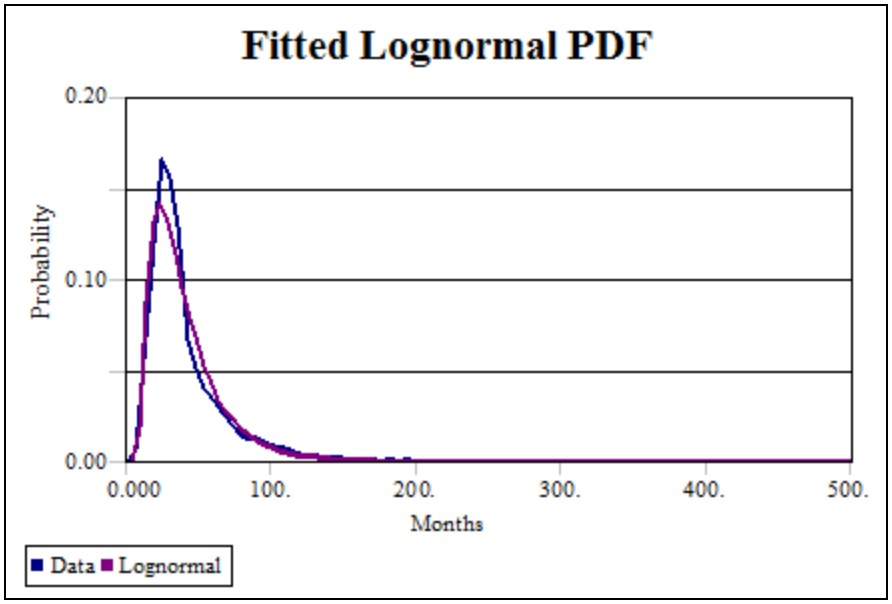


Figure 6-2. Months to Minimum Size Fitted Lognormal PDF vs. Empirical PDF

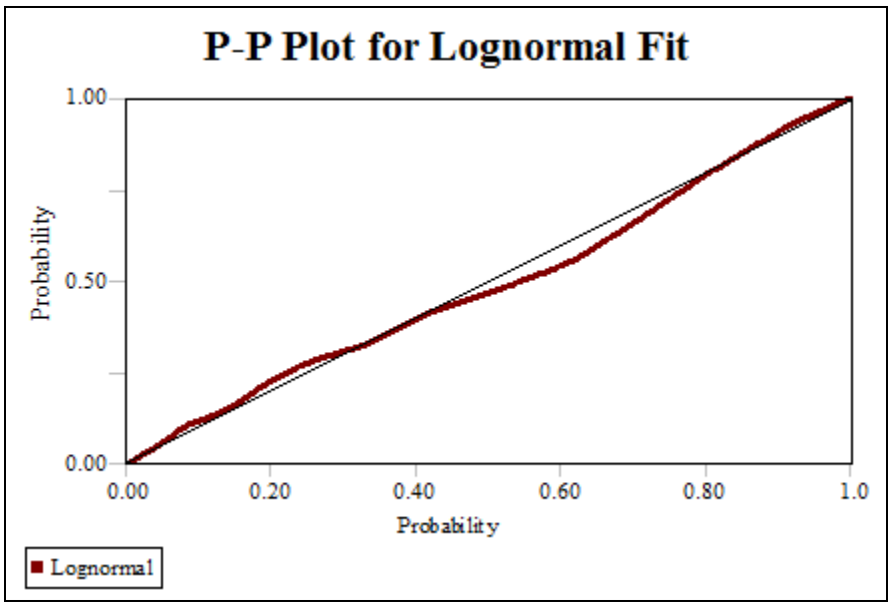


Figure 6-3. Months to Minimum Size P-P Plot of Lognormal Fit

D.2 Distribution of the Number of Months it Takes Tumors to Become Clinically Detectable or Symptomatic (All Invasive Cancers)

- Best Distribution: Pearson 6
- Pearson 6 Distribution Parameters: β = 106.4, p = 2.8, q = 4.7, *Min (Offset)* = 3.0
- Moments of the Fitted Pearson 6 Distribution
  - μ = 83.5, , σ^2 = 5580.1, γ_1 = 4.21, γ_2 = 73.3
- Moments of the Data
  - μ = 83.2, σ^2 = 5165.0, γ_1 = 0.53, γ_2 = 0.045
- Chi-Squared p-value = 0.904
- Kolmogrov Smirinov p-value = 0.885

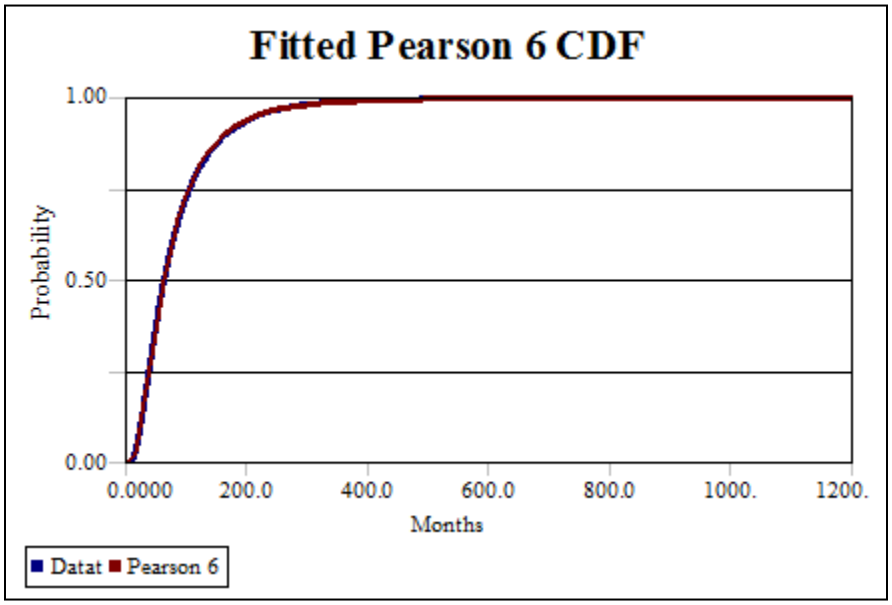


Figure 6-4. Months to Clinical Size Fitted Pearson 6 CDF vs. Empirical CDF

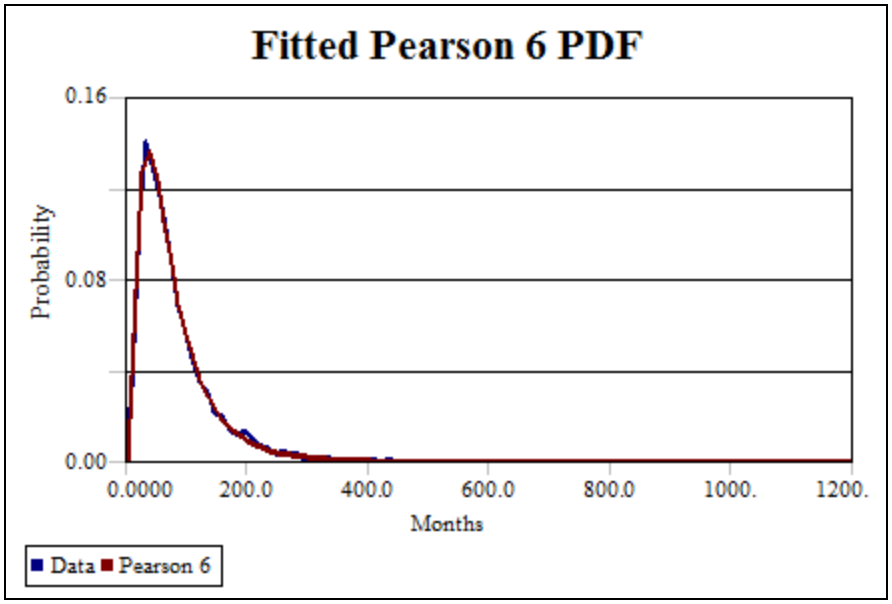


Figure 6-5. Months to Clinical Size Fitted Pearson 6 PDF vs. Empirical PDF

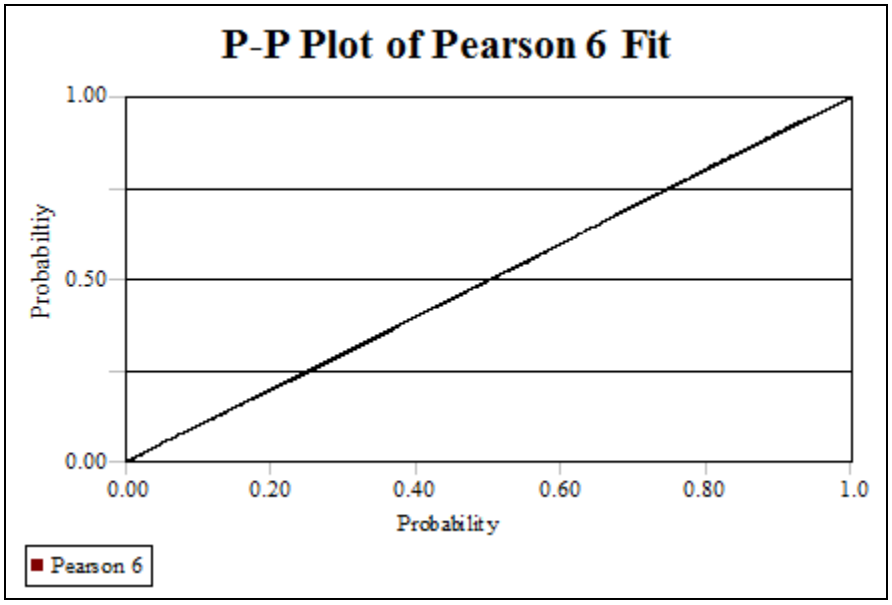


Figure 6-6. Months to Clinical Size P-P Plot of Pearson 6 Fit

D.3 Distribution of the Number of Months Until Tumors to Become Lethal (All Invasive Cancers)

- Best Distribution: Pearson 6
- Pearson 6 Distribution Parameters: β = 235.4, p = 2.82, q = 4.81, *Min (Offset)* = 7.0
- Moments of the Fitted Pearson 6 Distribution
  - μ = 181.0, , σ^2 = 25,281.0, γ_1 = 4.05, γ_2 = 62.3
- Moments of the Data
  - μ = 181.0, σ^2 = 24,116.5, γ_1 = 3.64, γ_2 = 23.64
- Chi-Squared p-value = 0.908
- Kolmogrov Smirinov p-value = 0.459

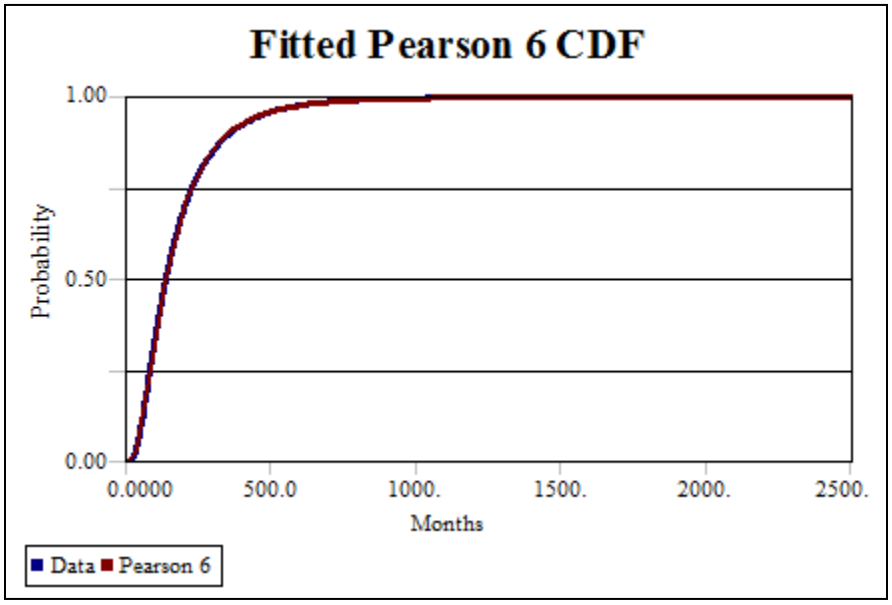


Figure 6-7. Months to Lethal Size Fitted Pearson 6 CDF vs. Empirical CDF

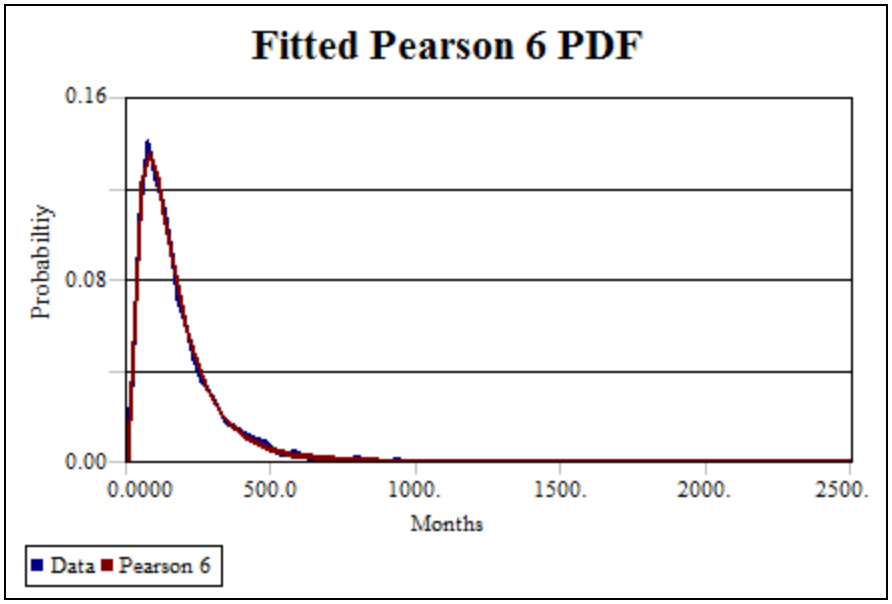


Figure 6-8. Months to Lethal Size Pearson 6 PDF vs. Empirical PDF

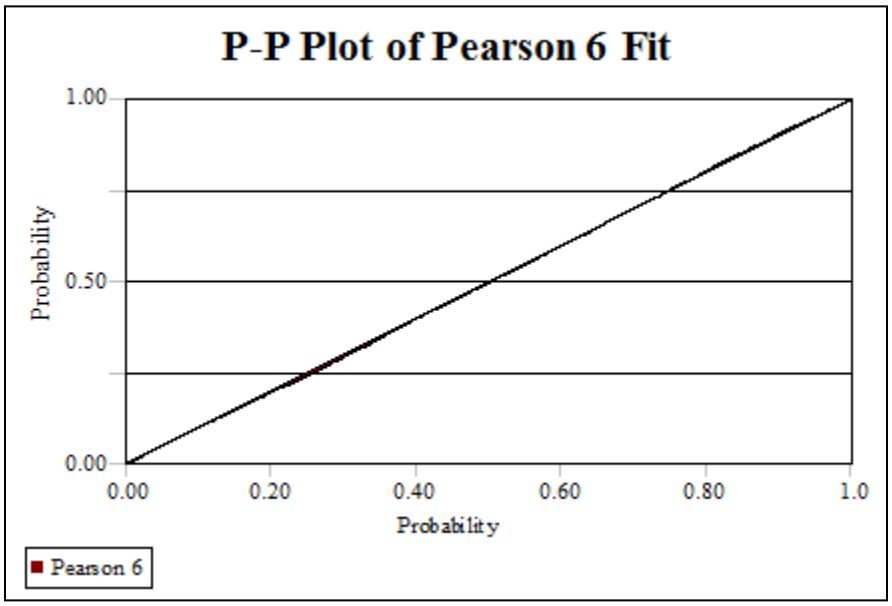


Figure 6-9. Months to Lethal Size P-P Plot of Pearson 6 Fit

D.4 Size When Tumor Becomes Detectable by Mammography (mm)

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = 2.27, σ = 0.639, *Min (Offset)* = 1.0
- Moments of the Fitted Lognormal Distribution
  - μ = 12.9, σ^2 = 71.2, γ_1 = 2.5, γ_2 = 12.7
- Moments of the Data
  - μ = 12.8, σ^2 = 73.1, γ_1 = 2.8, γ_2 = 11.8
- Chi-Squared p-value < 0.05
- Kolmogrov Smirinov p-value < 0.05

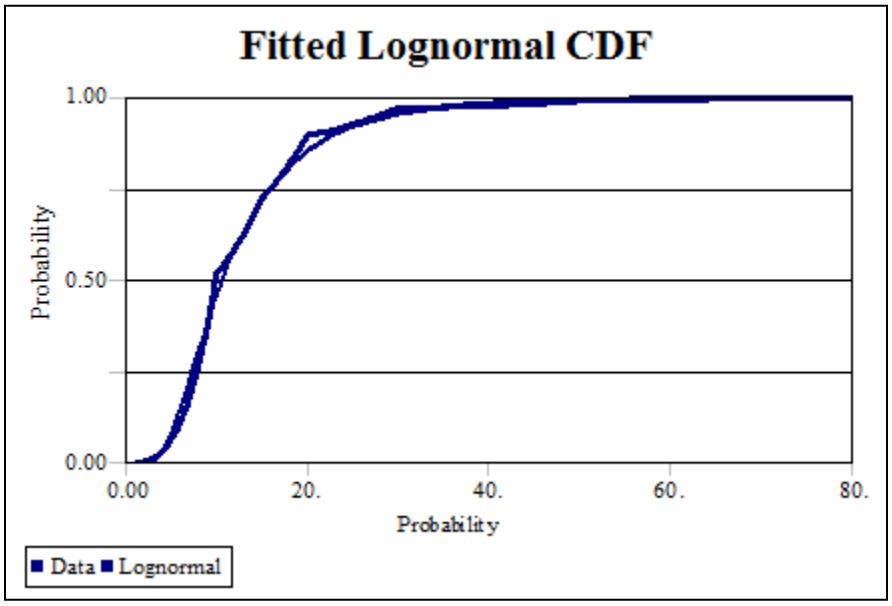


Figure 6-10. Minimum Mammography Detectable Size Fitted Lognormal CDF vs. Empirical CDF

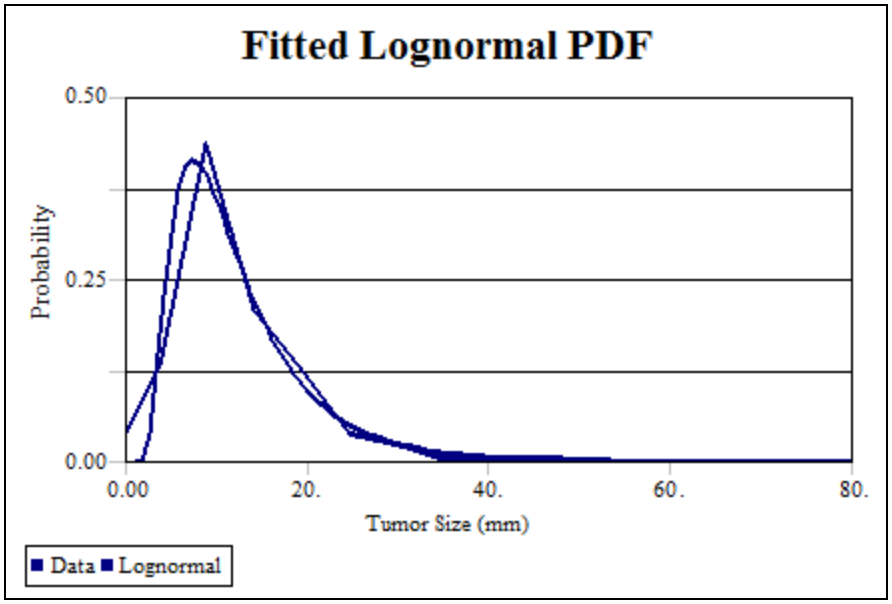


Figure 6-11. Minimum Mammography Detectable Size Fitted Lognormal PDF vs. Empirical PDF

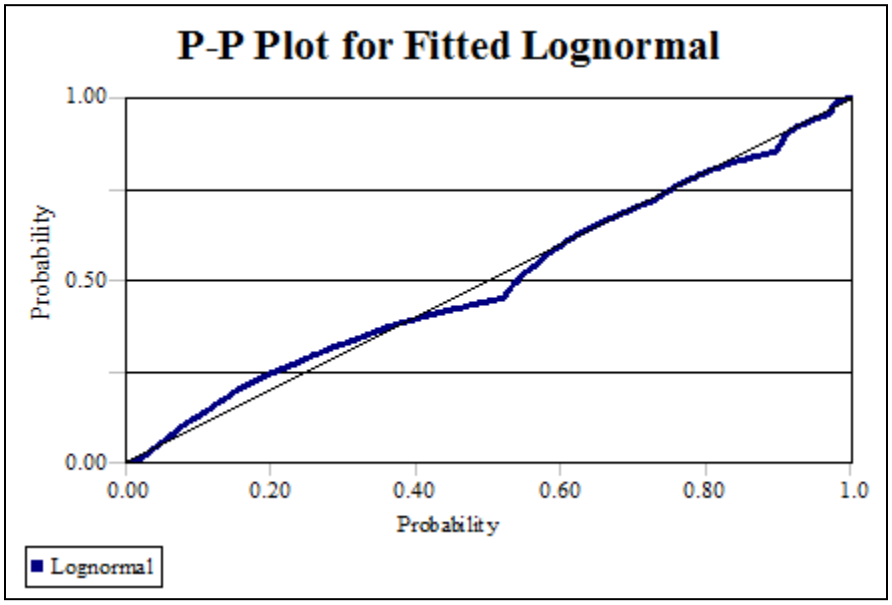


Figure 6-12. Minimum Mammography Detectable Size P-P Plot of Lognormal Fit

D.5 Size When Tumor Becomes Clinically Detectable (mm)

- Best Distribution: BETA
- Lognormal BETA Parameters: α_1 = 2.27, α_2 = 0.639, *Min (Offset)* = 1.0
- Moments of the Fitted BETA Distribution
  - μ = 47.1, σ^2 = 96.8, γ_1 = -0.04, γ_2 = -0.73
- Moments of the Data
  - μ = 47.1, σ^2 = 97.6, γ_1 = -0.09, γ_2 = -0.70
- Chi-Squared p-value = 0.129
- Kolmogrov Smirinov p-value = 0.130

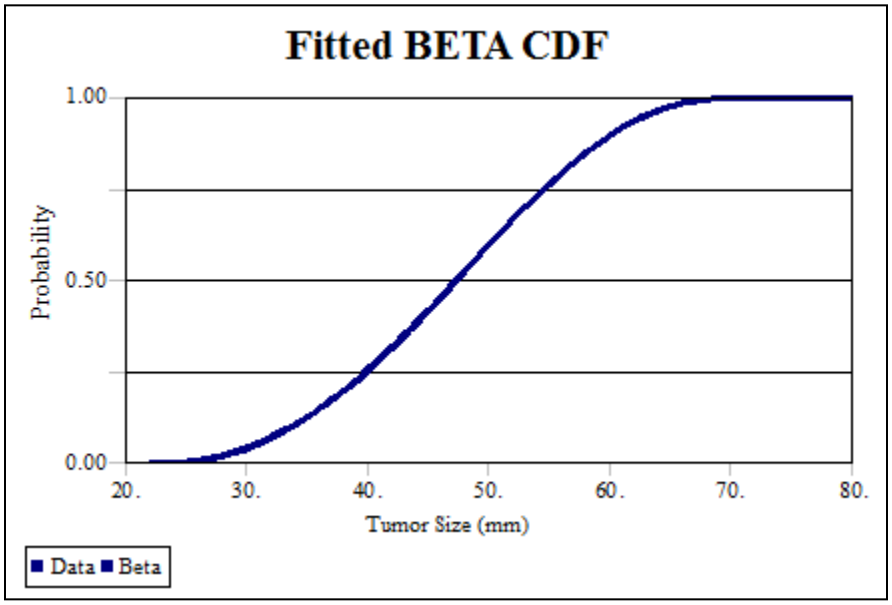


Figure 6-13. Minimum Clinical Detectable Size Fitted BETA CDF vs. Empirical CDF

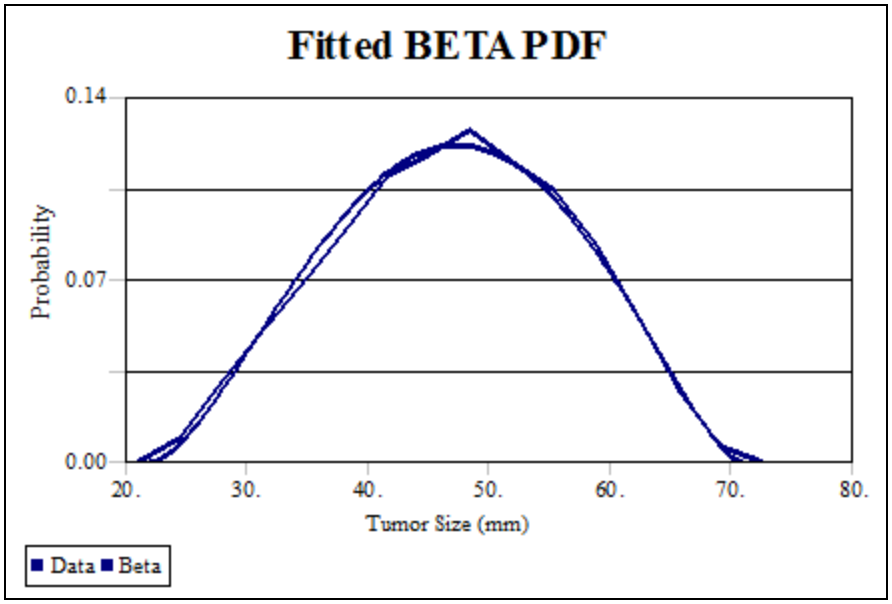


Figure 6-14. Minimum Clinical Detectable Size Fitted BETA PDF vs. Empirical PDF

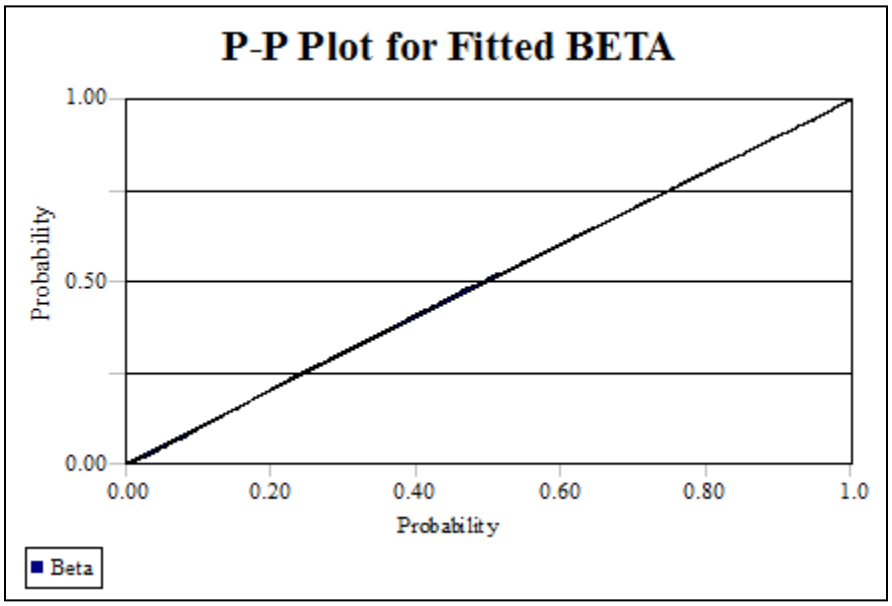


Figure 6-15. Minimum Clinical Detectable Size P-P Plot of BETA Fit

D.6 Cancer Onset Age for Women 65 and Older (All Invasive Cancers)

- Best Distribution: Pearson 6
- Pearson 6 Distribution Parameters: β = 383.7, p = 23.8, q = 248.6, *Min (Offset)* = 36.0
- Moments of the Fitted Pearson 6 Distribution
  - μ = 72.9, , σ^2 = 62.9, γ_1 = 0.47, γ_2 = 0.357
- Moments of the Data
  - μ = 72.9, σ^2 = 61.9, γ_1 = 0.53, γ_2 = 0.045
- Chi-Squared p-value < 0.05
- Kolmogrov Smirinov p-value < 0.05

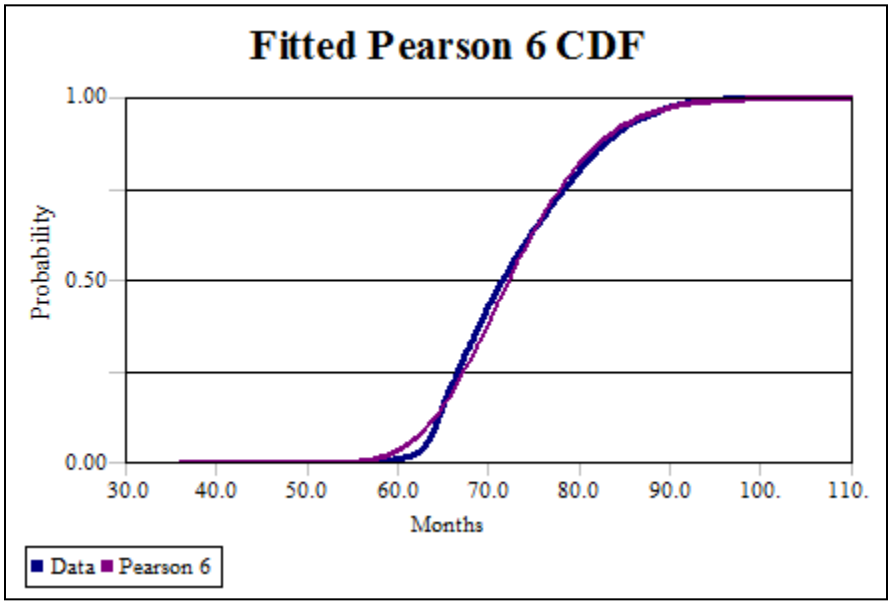


Figure 6-16. Onset Age Fitted Pearson 6 CDF vs. Empirical CDF

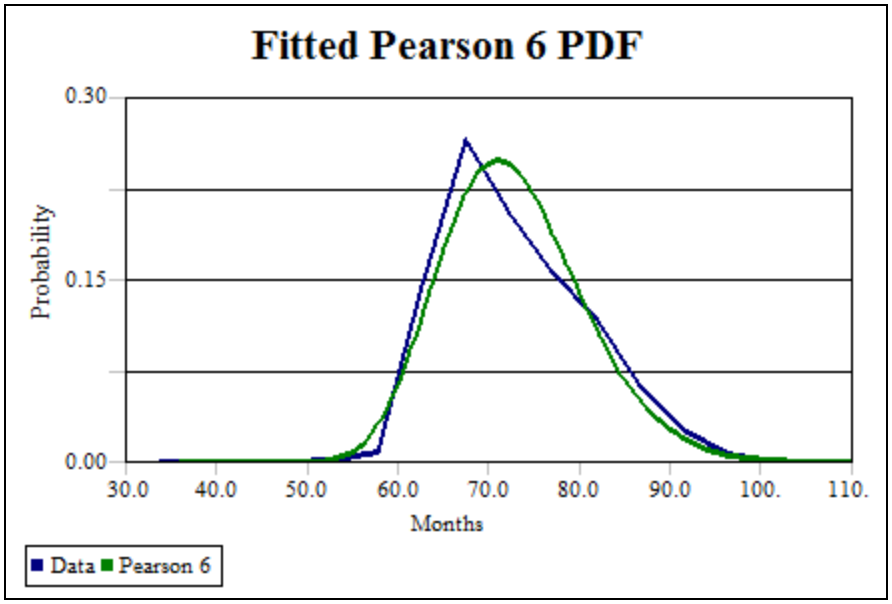


Figure 6-17. Onset Age Fitted Pearson 6 PDF vs. Empirical PDF

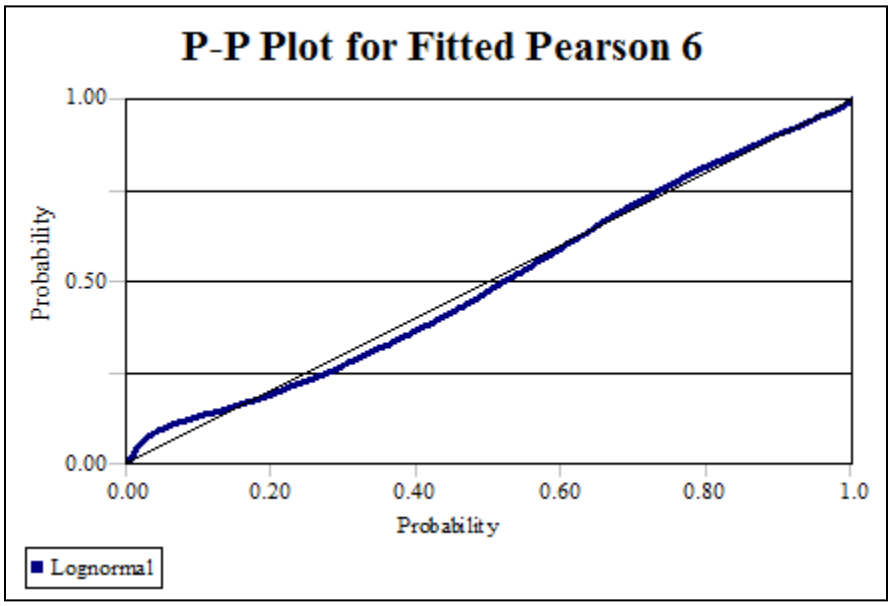


Figure 6-18. Onset Age P-P Plot of Pearson 6 Fit

D.7 Tumor Growth Rate for Women 65 and Older

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = -3.84, σ = 0.75, *Min (Offset)* = 0.0
- Moments of the Fitted Lognormal Distribution
  - μ = 0.0285, σ^2 = 0.0247, γ_1 = 3.26, γ_2 = 23.5
- Moments of the Data
  - μ = 0.0284, σ^2 = 0.00622, γ_1 = 3.2, γ_2 = 19.3
- Chi-Squared p-value = 0.894
- Kolmogrov Smirinov p-value = 0.930

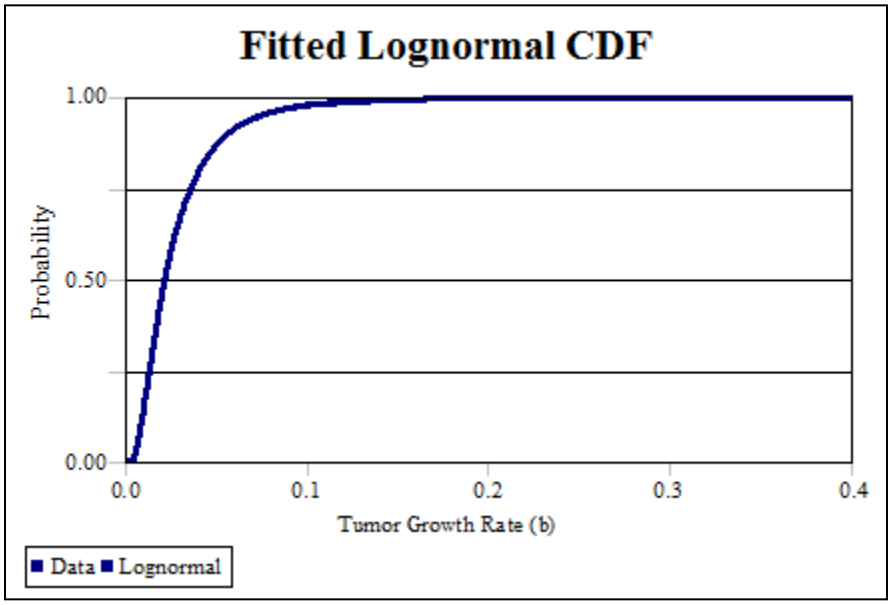


Figure 6-19. Tumor Growth Rate Fitted Lognormal CDF vs. Empirical CDF

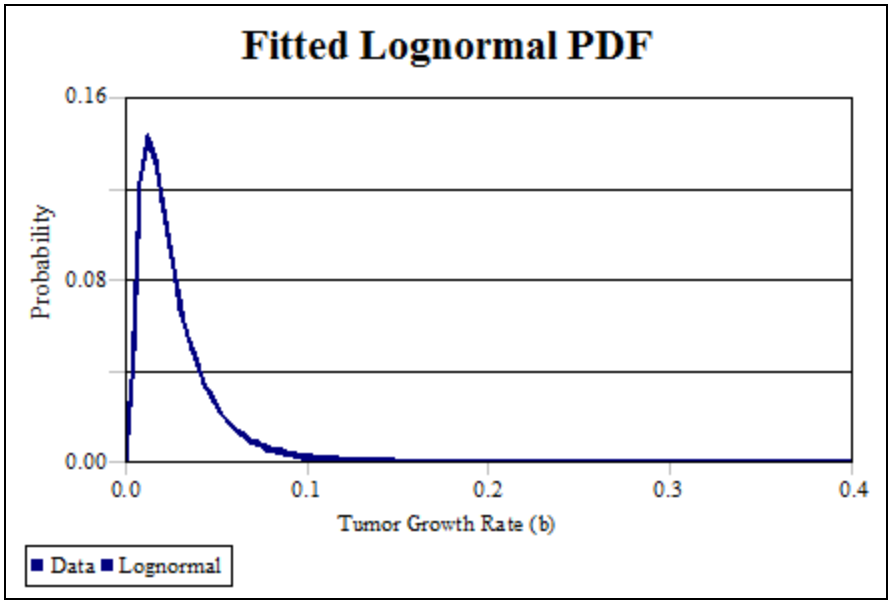


Figure 6-20. Tumor Growth Rate Fitted Lognormal PDF vs. Empirical PDF

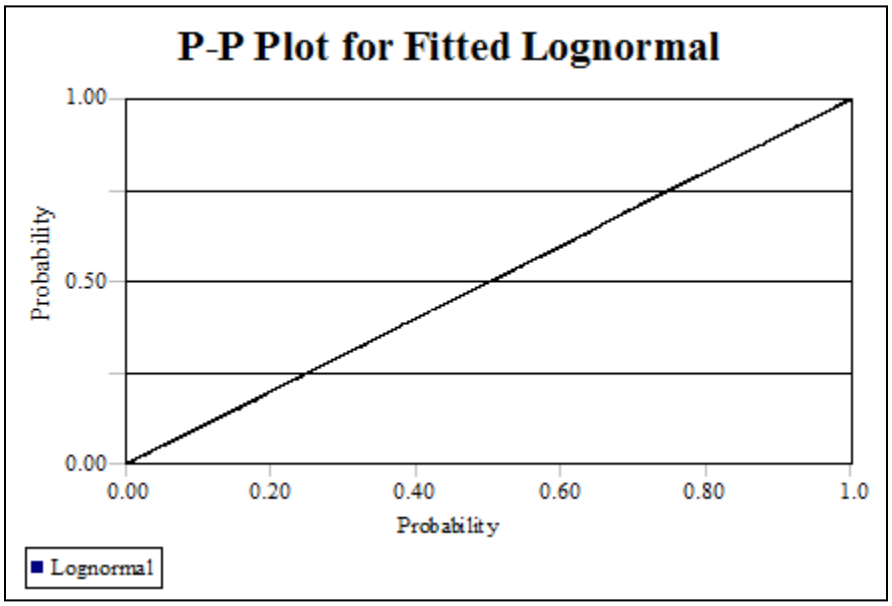


Figure 6-21. Tumor Growth Rate P-P Plot of Lognormal Fit

Appendix E: Distributions Fitted to Costing Data

E.1 Distribution for Cost of Film Screening Mammogram

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = 2.34, σ = 0.618, *Min (Offset)* = 71.0
- Moments of the Fitted Lognormal Distribution
  - μ = 83.6, σ^2 = 73.4, γ_1 = 2.36, γ_2 = 11.3
- Moments of the Data
  - μ = 83.3, σ^2 = 51.7, γ_1 = 1.29, γ_2 = 0.81
- Anderson-Darling p-value = 0.0855
- Kolmogrov Smirinov p-value = 0.126

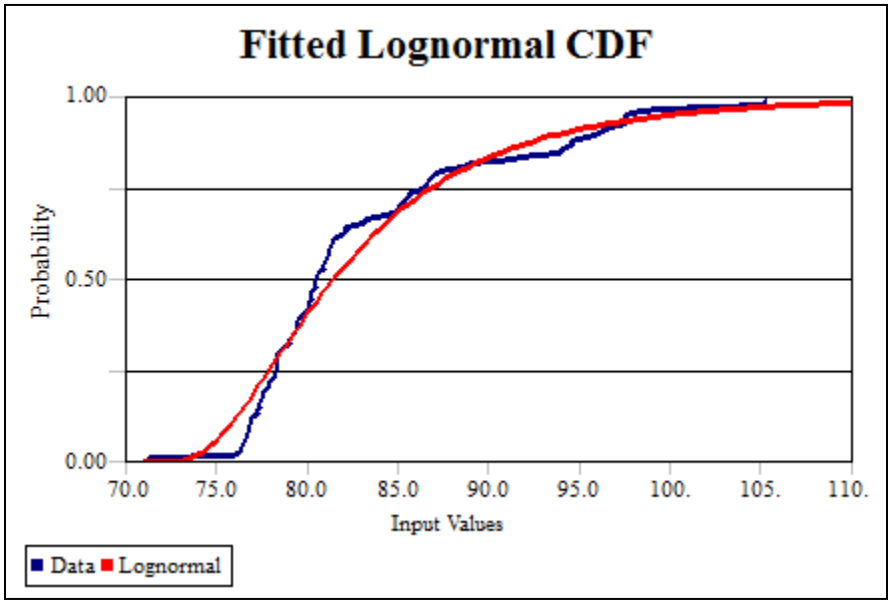


Figure 6-22. Cost of Film Screening Mammogram Fitted Lognormal CDF vs. Empirical CDF

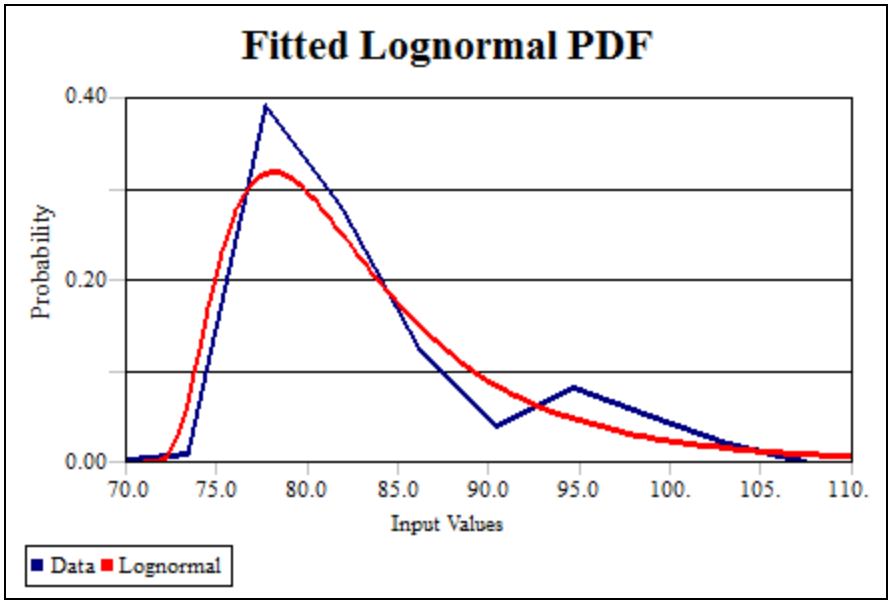


Figure 6-23. Cost of Film Screening Mammogram Fitted Lognormal PDF vs. Empirical PDF

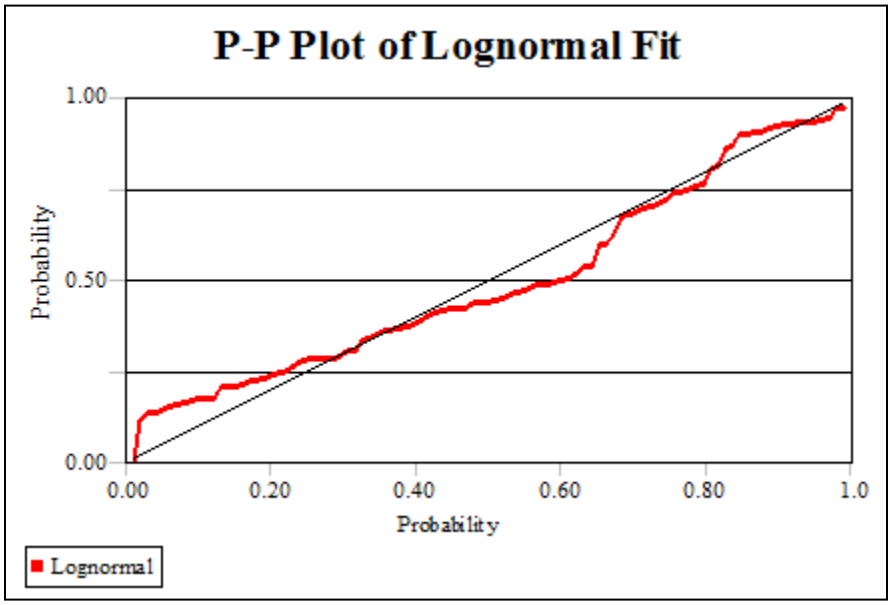


Figure 6-24. Cost of Film Screening Mammogram P-P Plot of Lognormal Fit

E.2 Distribution for Cost of Digital Screening Mammogram

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = 2.97, σ = 0.678, *Min (Offset)* = 121.0
- Moments of the Fitted Lognormal Distribution
  - μ = 146.0, σ^2 = 349.7, γ_1 = 2.74, γ_2 = 15.8
- Moments of the Data
  - μ = 144.37, σ^2 = 199.5, γ_1 = 1.35, γ_2 = 1.09
- Anderson-Darling p-value = 0.0529
- Kolmogrov Smirinov p-value = 0.077

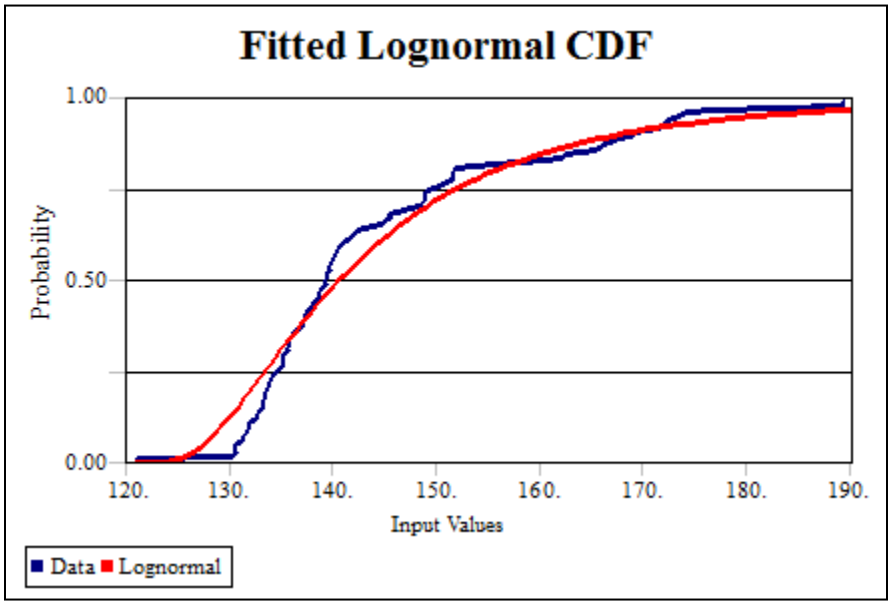


Figure 6-25. Cost of Digital Screening Mammogram Fitted Lognormal CDF vs. Empirical CDF

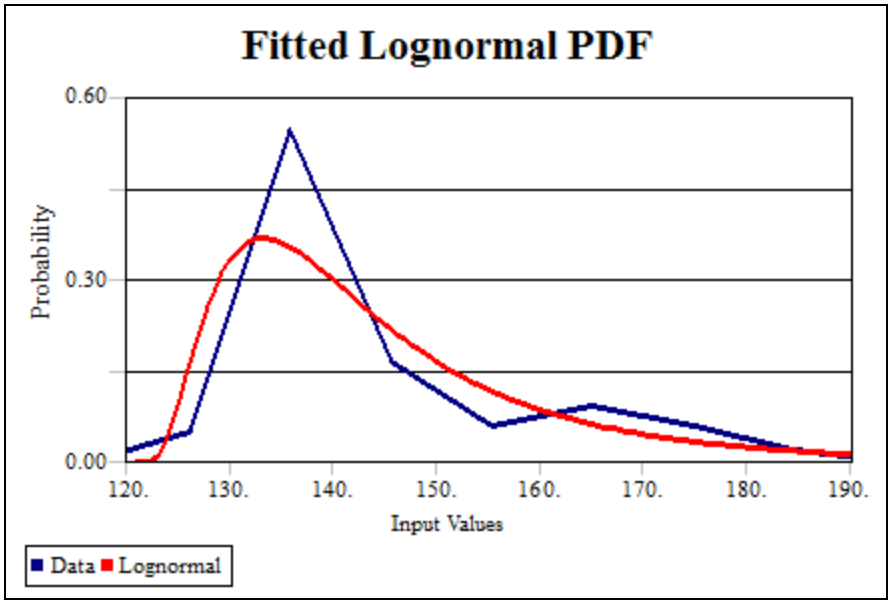


Figure 6-26. Cost of Digital Screening Mammogram Fitted Lognormal PDF vs. Empirical PDF

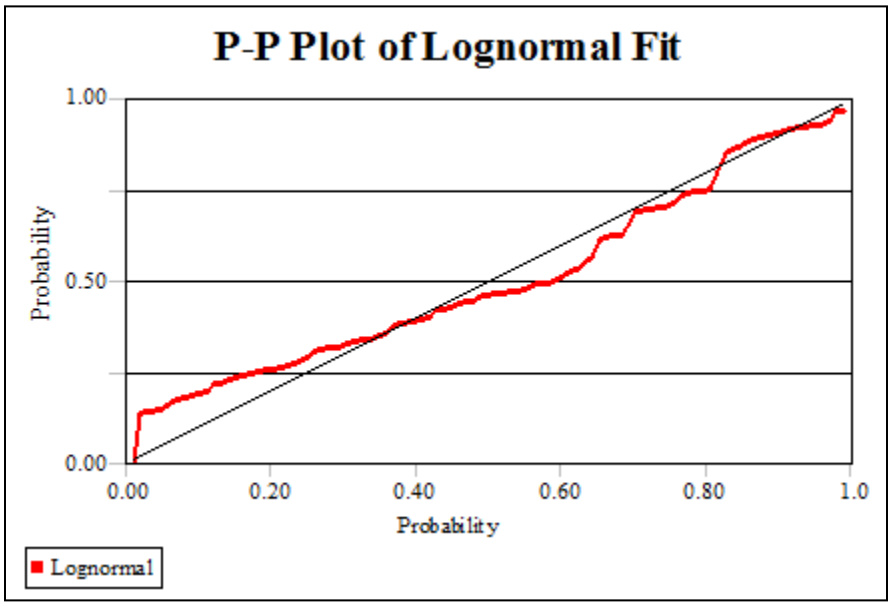


Figure 6-27. Cost of Digital Screening Mammogram P-P Plot of Lognormal Fit

E.3 Distribution for Cost of Film Diagnostic Mammogram

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = 2.49, σ = 0.548, *Min (Offset)* = 75.0
- Moments of the Fitted Lognormal Distribution
  - μ = 89.0, σ^2 = 68.72, γ_1 = 1.98, γ_2 = 7.72
- Moments of the Data
  - μ = 88.9, σ^2 = 61.16, γ_1 = 1.29, γ_2 = 0.85
- Anderson-Darling p-value = 0.115
- Kolmogrov Smirinov p-value = 0.0945

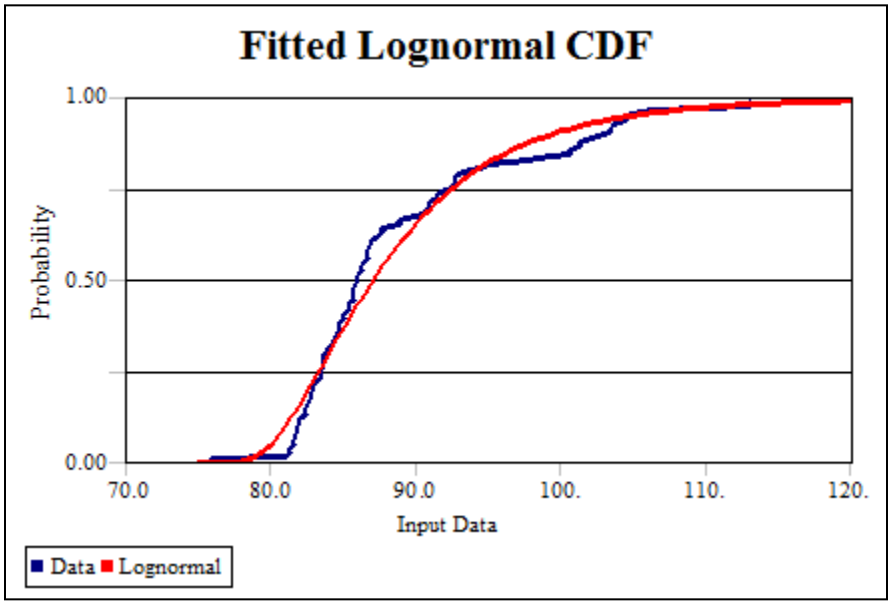


Figure 6-28. Cost of Film Diagnostic Mammogram Fitted Lognormal CDF vs. Empirical CDF

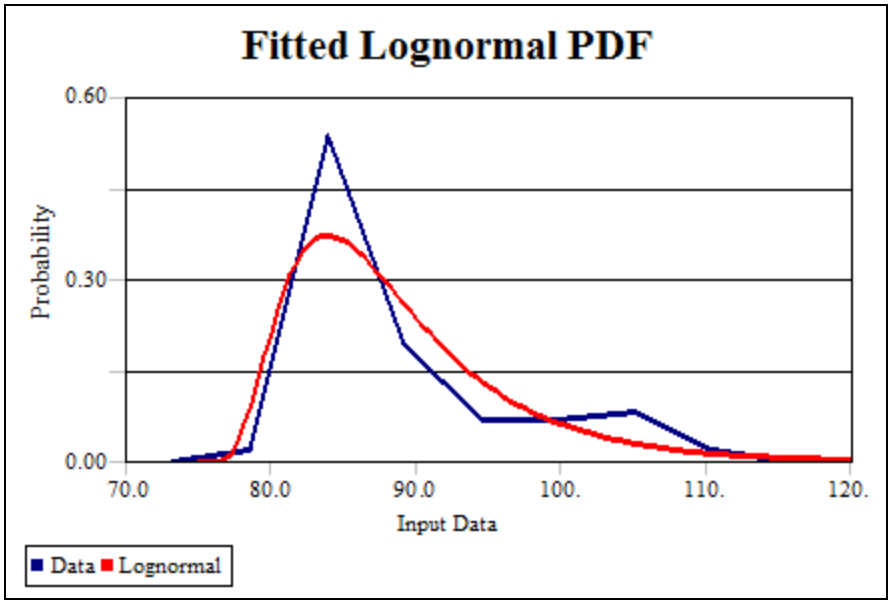


Figure 6-29. Cost of Film Diagnostic Mammogram Fitted Lognormal PDF vs. Empirical PDF

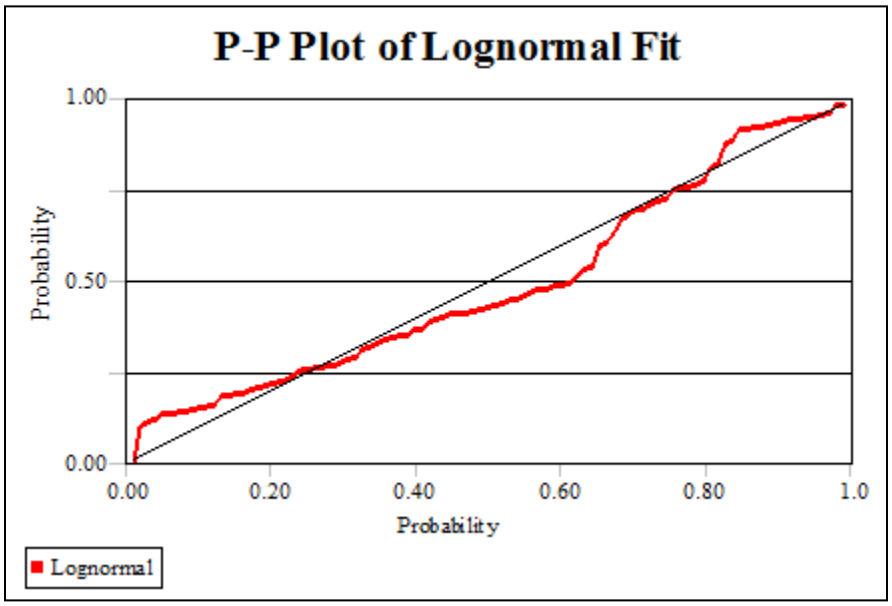


Figure 6-30. Cost of Film Diagnostic Mammogram P-P Plot of Lognormal Fit

E.4 Distribution for Cost of Digital Diagnostic Mammogram

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = 2.96, σ = 0.584, *Min (Offset)* = 114.0
- Moments of the Fitted Lognormal Distribution
  - μ = 137.0, σ^2 = 213.2, γ_1 = 2.17, γ_2 = 9.40
- Moments of the Data
  - μ = 136.6, σ^2 = 175.4, γ_1 = 1.35, γ_2 = 1.07
- Anderson-Darling p-value = 0.124
- Kolmogrov Smirinov p-value = 0.201

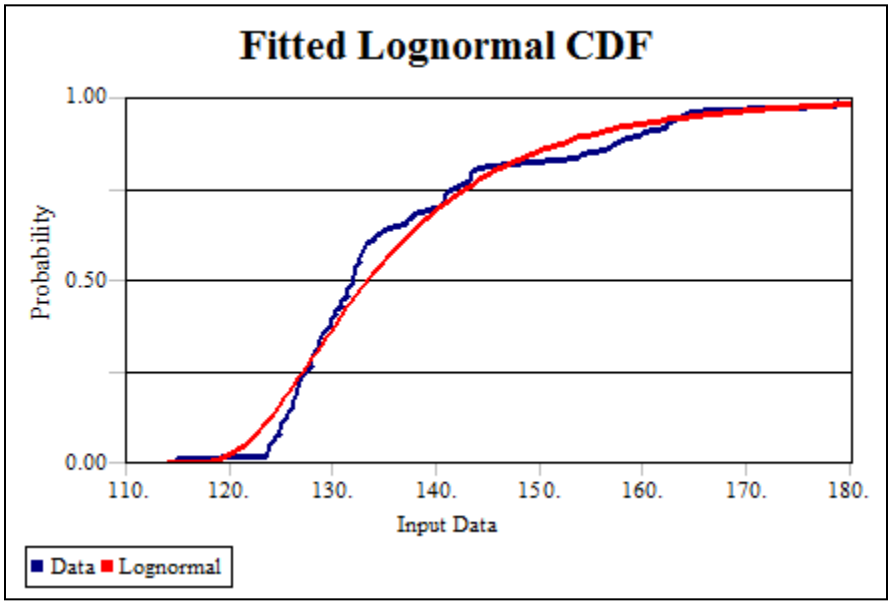


Figure 6-31. Cost of Digital Diagnostic Mammogram Fitted Lognormal CDF vs. Empirical CDF

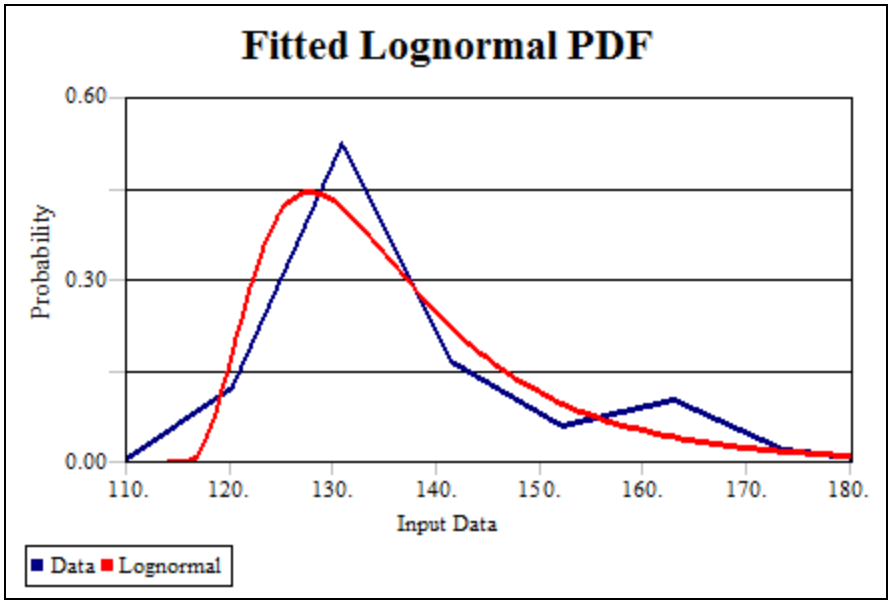


Figure 6-32. Cost of Digital Diagnostic Mammogram Fitted Lognormal PDF vs. Empirical PDF

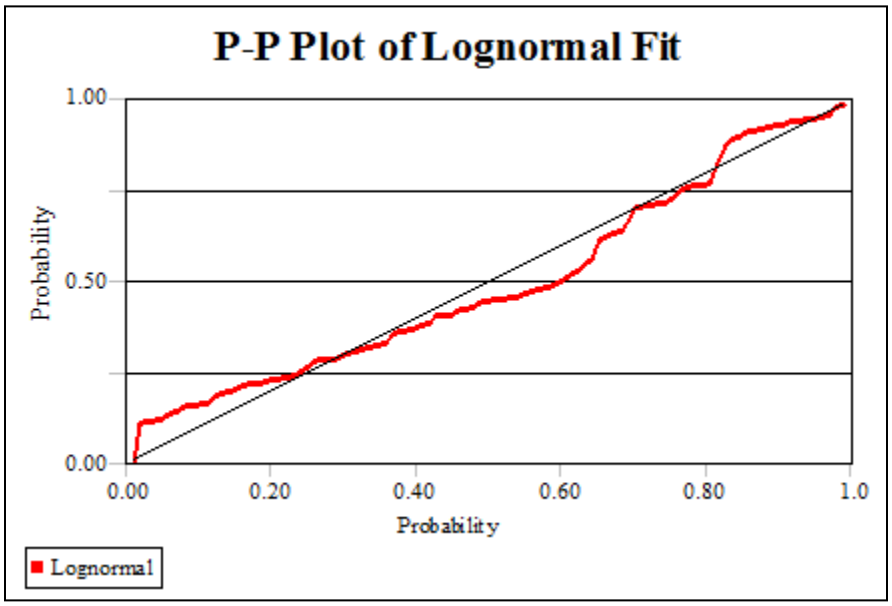


Figure 6-33. Cost of Digital Diagnostic Mammogram P-P Plot of Lognormal Fit

E.5 Distribution for Cost of Diagnostic Ultrasound

- Best Distribution: Lognormal
- Lognormal Distribution Parameters: μ = 2.62, σ = 0.62, *Min (Offset)* = 83.0
- Moments of the Fitted Lognormal Distribution
  - μ = 99.6, σ^2 = 130.0, γ_1 = 2.37, γ_2 = 11.5
- Moments of the Data
  - μ = 99.2, σ^2 = 89.8, γ_1 = 1.31, γ_2 = 0.97
- Anderson-Darling p-value = 0.0922
- Kolmogrov Smirinov p-value = 0.155

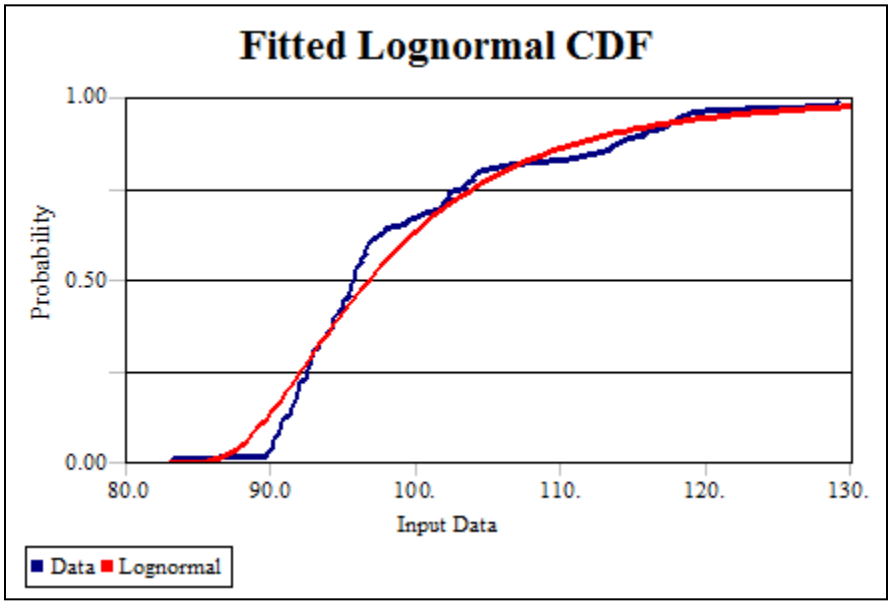


Figure 6-34. Cost of Diagnostic Ultrasound Fitted Lognormal CDF vs. Empirical CDF

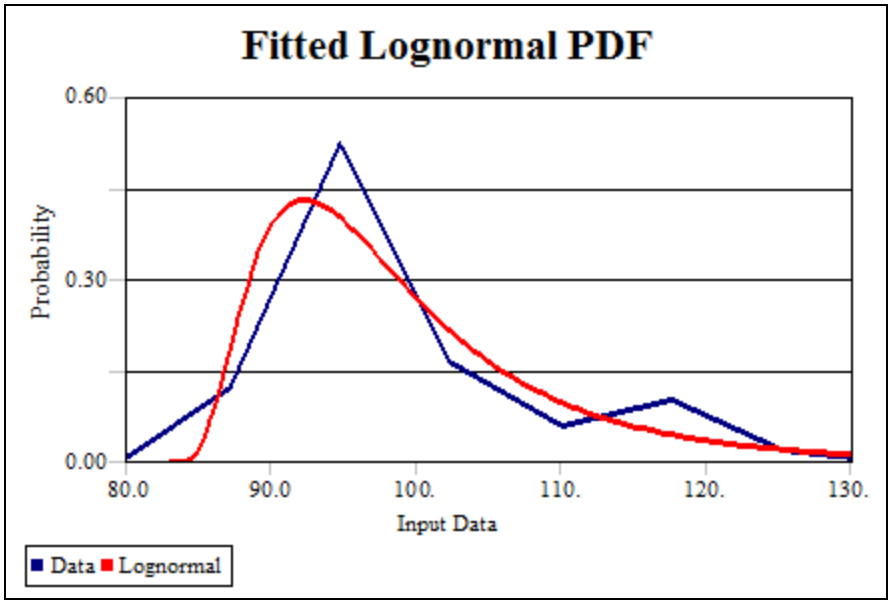


Figure 6-35. Cost of Diagnostic Ultrasound Fitted Lognormal PDF vs. Empirical PDF

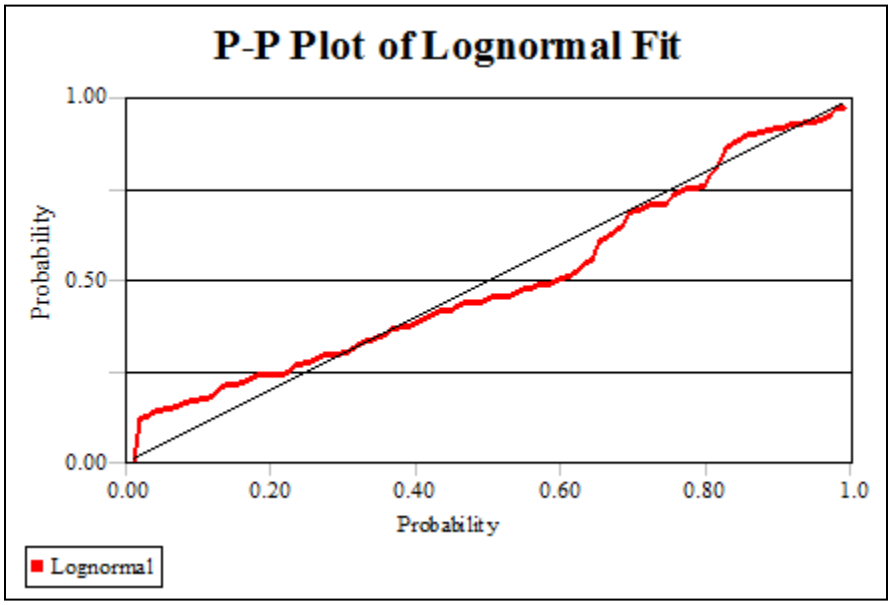


Figure 6-36. Cost of Diagnostic Ultrasound P-P Plot of Lognormal Fit

E.6 Distribution for Cost of Core Needle Biopsy (Ultrasound Guided)

- Best Distribution: Pearson 6
- Pearson 6 Distribution Parameters: β = 4872.7, p = 3.04, q = 115.4, *Min (Offset)* = 706.0
- Moments of the Fitted Pearson 6 Distribution
  - μ = 836.0, σ^2 = 5715.4, γ_1 = 1.21, γ_2 = 2.26
- Moments of the Data
  - μ = 839.4, σ^2 = 6212.6, γ_1 = 1.29, γ_2 = 0.877
- Anderson-Darling p-value = 0.0510
- Kolmogrov Smirinov p-value = 0.0875

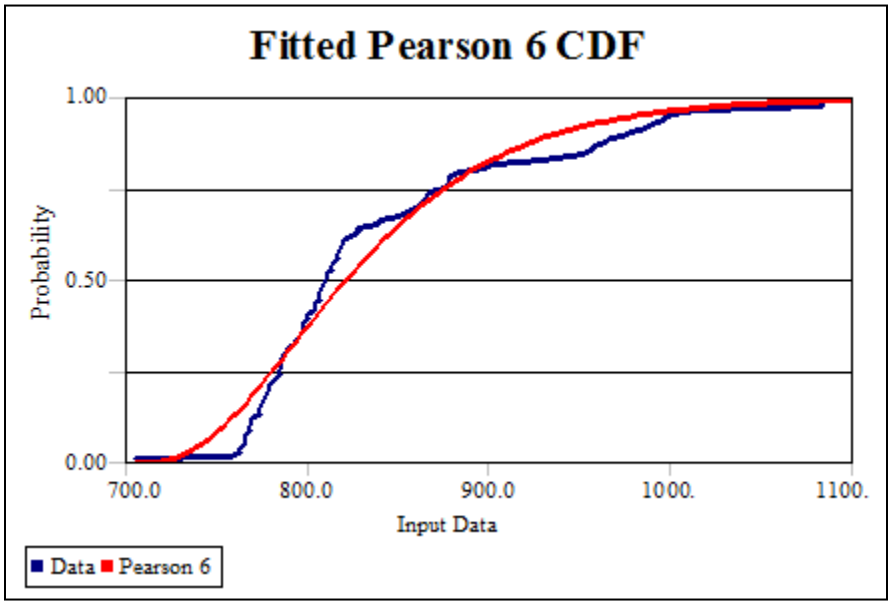


Figure 6-37. Cost of CNB (Ultrasound Guided) Fitted Pearson 6 CDF vs. Empirical CDF

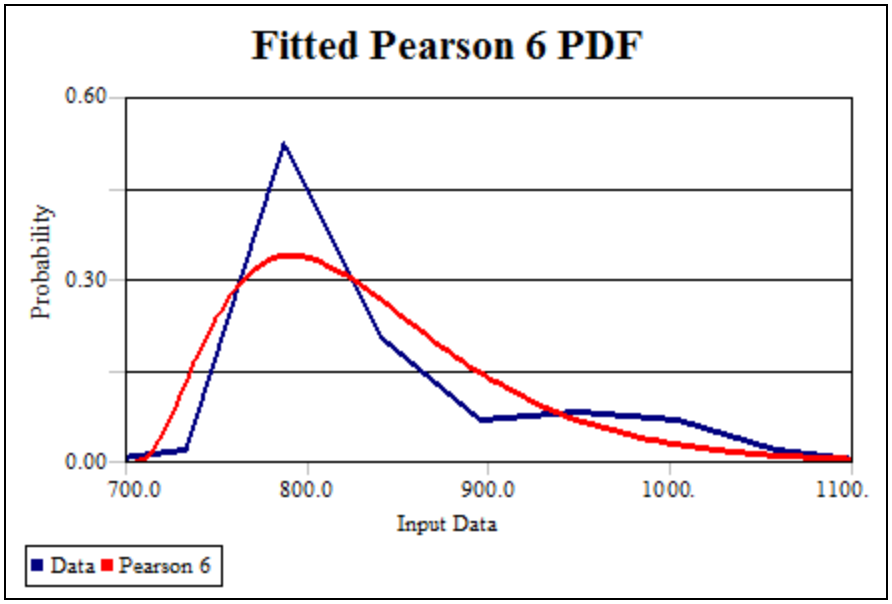


Figure 6-38. Cost of CNB (Ultrasound Guided) Fitted Pearson 6 PDF vs. Empirical PDF

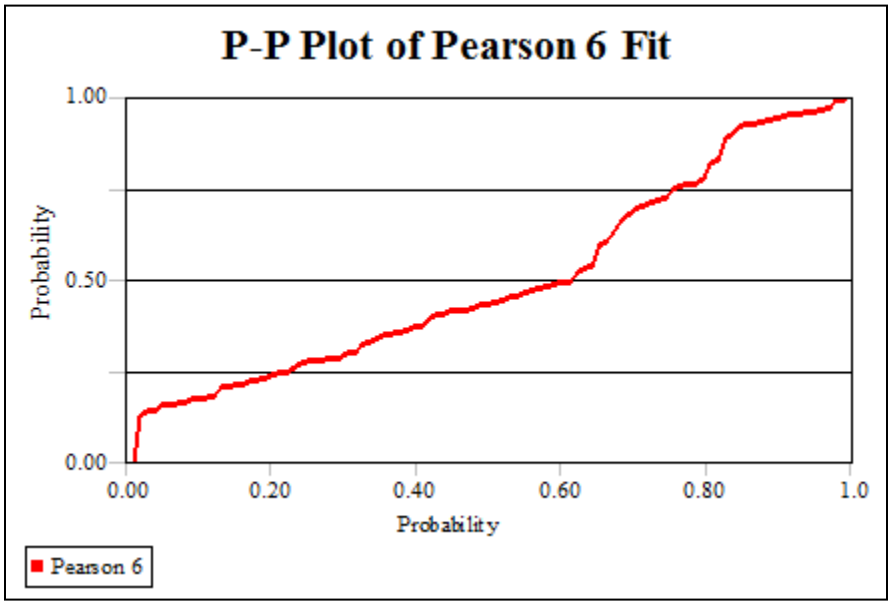


Figure 6-39. Cost of CNB (Ultrasound Guided) P-P Plot of Pearson 6 Fit

E.7 Distribution for Cost of Fine Needle Aspiration (FNA)

- Best Distribution: Pearson 6
- Pearson 6 Distribution Parameters: β = 739.5, p = 3.03, q = 29.2, *Min (Offset)* = 457.0
- Moments of the Fitted Pearson 6 Distribution
  - μ = 536.0, , σ^2 = 2391.2, γ_1 = 1.40, γ_2 = 3.31
- Moments of the Data
  - μ = 536.9, σ^2 = 2325.8, γ_1 = 1.31, γ_2 = 0.916
- Anderson-Darling p-value = 0.0523
- Kolmogrov Smirinov p-value = 0.0827

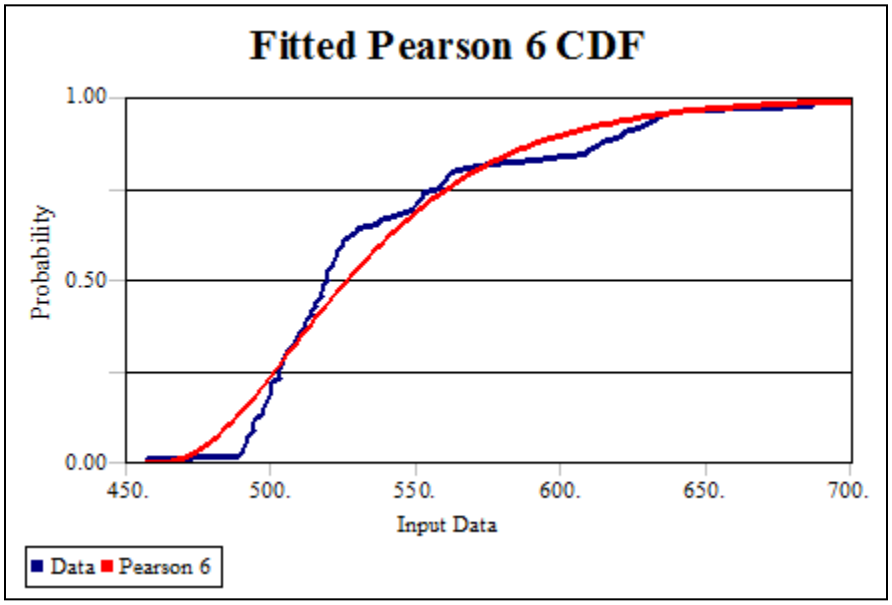


Figure 6-40. Cost of FNA Fitted Pearson 6 CDF vs. Empirical CDF

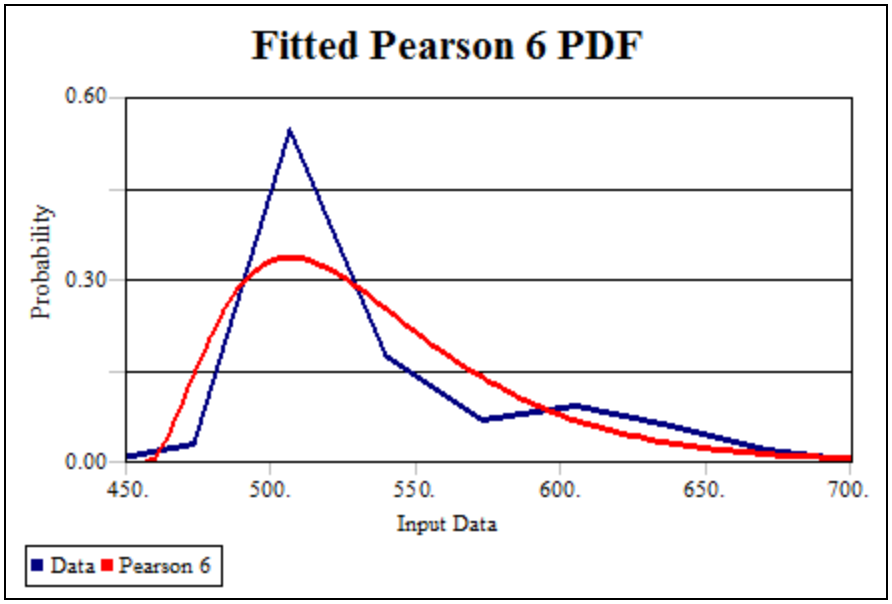


Figure 6-41. Cost of FNA Fitted Pearson 6 PDF vs. Empirical PDF

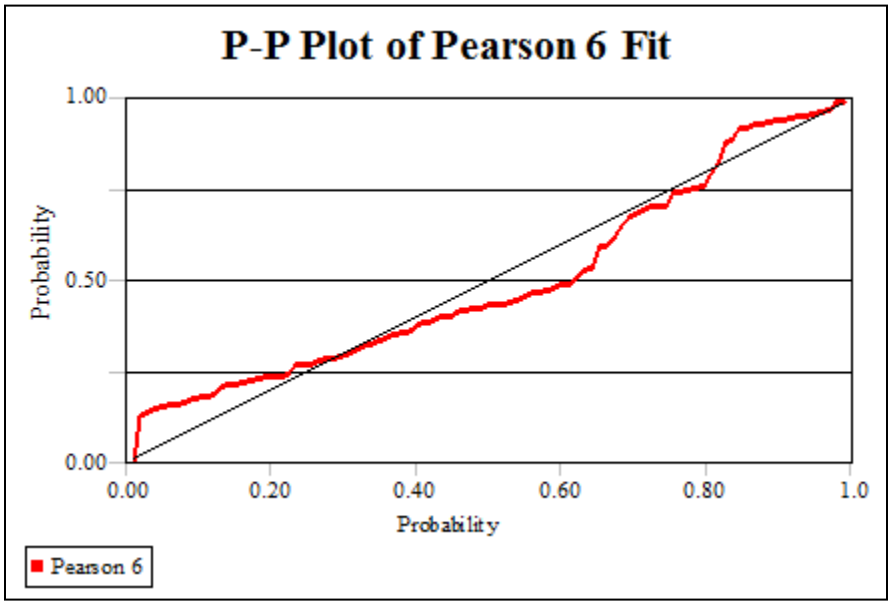


Figure 6-42. Cost of FNA P-P Plot of Pearson 6 Fit

Appendix F: Method for Sampling a Pearson Type VI Distribution

A Pearson Type IV random variable X with lower limit a, scale parameter β, and shape parameters α_1 and α_2 has the form

$$f_X(x) = \frac{\left[(x-a)/\beta\right]^{\alpha_1 - 1}}{\beta\left[1+(x-a)/\beta\right]^{\alpha_1+\alpha_2}\,\mathrm{B}(\alpha_1,\alpha_2)} \quad \text{for } x \geq a \tag{6.1}$$

where

$$\mathrm{B}(\alpha_1,\alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \tag{6.2}$$

is the beta function. The appearance of $\mathrm{B}(\alpha_1,\alpha_2)$ in Equation (6.1) strongly suggests that a simple transformation of $(x-a)/\beta$ must have a standard beta distribution. The simplest such transformation from $[a,\infty]$ to $[0,1]$ is

$$Y = \theta(X) = \frac{(x-a)/\beta}{1+(x-a)/\beta} \tag{6.3}$$

with inverse transformation

$$X = \theta^{-1}(Y) = a + \beta\left[y/(1-y)\right]. \tag{6.4}$$

Thus we see by the change of variables formula that if Y is a standard beta random variable with shape parameters α_1 and α_2 so that

$$f_Y(y) = \frac{y^{\alpha_1-1}(1-y)^{\alpha_2-1}}{\mathrm{B}(\alpha_1,\alpha_2)} \quad \text{for } 0 \leq y \leq 1, \tag{6.5}$$

then $X = \theta^{-1}(Y)$ has p.d.f

$$f_X(x) = f_Y\left[\theta(x)\right]\left|\frac{d}{dx}\theta(x)\right|$$

$$= \frac{1}{\mathrm{B}(\alpha_1,\alpha_2)} \cdot \left\{\frac{(x-a)/\beta}{\left[1+(x-a)/\beta\right]}\right\}^{\alpha_1-1} \left\{1-\frac{(x-a)/\beta}{\left[1+(x-a)/\beta\right]}\right\}^{\alpha_2-1} \times$$
$$\left|\frac{1}{\beta\left[1+(x-a)/\beta\right]} - \frac{(x-a)/\beta}{\left[1+(x-a)/\beta\right]^2}\cdot\frac{1}{\beta}\right|$$

$$= \frac{1}{\mathrm{B}(\alpha_1,\alpha_2)} \cdot \frac{\left[(x-a)/\beta\right]^{\alpha_1-1}}{\left[1+(x-a)/\beta\right]^{\alpha_1+\alpha_2-2}} \times \left|\frac{1}{\beta}\cdot\frac{1}{\left[1+(x-a)/\beta\right]^2}\right|$$

$$= \frac{1}{\mathrm{B}(\alpha_1,\alpha_2)} \cdot \frac{\left[(x-a)/\beta\right]^{\alpha_1-1}}{\beta\left[1+(x-a)/\beta\right]^{\alpha_1+\alpha_2}}, \tag{6.6}$$

which is the Pearson Type VI density in Equation (6.1) as required.

Therefore, to generate a Pearson Type VI random variable with lower limit a, scale parameter β, and shape parameters α_1 and α_2 (p and q, respectively in **Stat::Fit** notation), generate

$$Y \leftarrow \mathrm{BETA}(\alpha_1,\alpha_2) \tag{6.7}$$

in Arena notation, and then deliver

$$X \leftarrow a + \beta\left[Y/(1-Y)\right]. \tag{6.8}$$