

The Metastatic Receptor Status Impact on First-Line Treatment Plans and Outcomes for Recurrent Metastatic Breast Cancer

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

T. Allen Pannell, Jr.
December 2015

Copyright © 2015 by T. Allen Pannell, Jr.
All rights reserved.

DEDICATION

To Amy

I love you, I miss you, I adore you.

ACKNOWLEDGEMENTS

There is an old saying that “If you ever see a turtle sitting on a fence post, you know it had help getting there.” I am that turtle. I can never fully thank everyone for the support I received during the research process and writing this dissertation. I am sure by the time this paper is accepted, I will realize people I have omitted. In no particular order, here goes; Jessica Pannell, Nicholas Pannell, Melissa Pannell, Amy Adele Foster, Russell Zaretski, Kitty Cornett, Jane Moser, Bobby Mee, Ken Gilbert, Alex Miller, Carolyn Cuddy, Steve Mangum, Tim Panella, Ellen Perkins, Ed Miles, Deanne Michaelson, H. Alan Lasater, Richard Sanders, Tom Pannell, Jean Pannell, Terri Beaudrot, Tom Beaudrot, Srini, Randy Bradley, Bogdan Bichescu, Bruce Behn, Amy Cathey, Jake Fait, James Hurley, Tyler Massaro and Anna Cormier.

ABSTRACT

Background: For more than two decades, breast cancer researchers have studied the benefits, risks and clinical importance of testing the receptor status of metastatic tumors. While there is a growing consensus that the status should be re-tested and under what circumstances that re-testing should occur, there is little to no evidence that utilizing test results for metastatic tumor receptor status improves the clinical outcomes of patients. In fact, there is evidence that changes to treatment plans based on this re-testing can be harmful to patient outcomes.

Objective: This dissertation evaluates the current state of evidence related to altering patient treatment plans based on the re-test results of metastatic tumors, offers an update to existing national and international standards and executes a retrospective observational study to provide data that supports better informed decisions on first-line metastatic treatment plans where retesting of tumors is an option.

Methods: A thorough literature review was performed on the topic. Afterwards, a retrospective observational study was performed at the University of Tennessee Cancer Institute.

Conclusions: The research outcomes documented in this dissertation demonstrate that basing first-line treatment plans for metastatic breast cancer patients on the receptor status of the primary tumor instead of the metastatic tumor receptor status extends the life expectancy of patients. A standard of care is proposed that impacts national and international guidelines and reflects the risks associated with changing the first-line treatment plans of metastatic breast cancer patients based on the receptor status of metastatic tumors.

PREFACE

The Scientific Method Begins With an Observation

Amy

On January 8th, 2014 I observed Amy Adele Foster die. Three months earlier, on our first wedding anniversary, I observed an x-ray of her lungs taken in the Emergency Room of the LeConte Medical Center in Sevierville, TN. That x-ray showed two liters of fluid in the pleural space of her lung. Four weeks earlier I observed a doctor at a world renowned medical center change Amy's medication from one that was working to a different medication based on the fact that "her receptor status has changed from PR- to PR+." One week prior to the change in medicine, I observed her primary oncologist tell her, "Good news, your tumors are shrinking. I would estimate your life expectancy to now be four years, instead of the one year we originally estimated."

Allen

If the above paragraph were read backwards, one would notice that four months after being given four years to live, my wife was dead. In the world of metastatic breast cancer, such outcomes occasionally occur without cause; but my instincts led me to a different conclusion. I saw correlation and wondered if it was cause and effect. I contacted the patient advocacy group at the medical center where her medicine was changed and asked for a scientific explanation of that decision. It took six months to get a phone call to discuss my concerns. In those six months, I prepared for the phone call by doing research on the topic. That research became this dissertation.

TABLE OF CONTENTS

CHAPTER 1 Introduction and General Information	1
Critical Medical Decision	1
Breast Cancer	1
Breast Cancer Staging	4
Receptor Status.....	4
Breast Cancer Treatments.....	5
Metastatic Breast Cancer	7
Impact.....	8
Problem Statement.....	8
Hypothesis	10
Research Purpose.....	11
CHAPTER 2 Breast Cancer Literature Review	12
Review Criteria and Process	12
Permissions.....	12
Standards	12
Advanced Breast Cancer (ABC) Standards.....	13
Levels of Evidence.....	14
NCCN Standards.....	14

Discussion of Standards.....	16
Retesting of Metastatic Tumor Receptors	16
Discordance	18
How is Discordance Measured?	18
What Causes Discordance?	19
Impact on Treatment Decisions	22
Clinical outcomes.....	24
Potential Contradictory Data	28
Summary of Clinical Implications	31
Challenges Interpreting the Literature.....	31
Progress in the Literature.....	35
Clinical Standards Recommendation.....	37
Statistical Literature Review	37
CHAPTER 3 Materials and Methods.....	39
Research Purpose	39
Key Research Questions.....	40
Hypothesis	41
Study Design.....	41
Discordance	42
Phenotypes and Discordance.....	42

Study Decision Flow	44
Data Parameters.....	46
Variables in the Tumor Registry.....	46
Variables in Electronic and Paper Charts	47
Quality Control	47
Record Keeping	47
Confidentiality and Compliance.....	48
Statistical Considerations.....	48
Sample Size and Power.....	48
Sample Size Discussion.....	53
A Priori Analysis Plans.....	54
Statistical Significance	55
Statistical Methods Review	56
Simulation.....	57
Simulation Results	58
CHAPTER 4 Results and Discussion	67
Data Acquisition	67
Data Cleaning	67
Limited Data Set	69
Finalizing the Data.....	70

Data Collection Summary	72
List of Variables in the Final Dataset	73
Preparing the Data for Analysis	73
Study Results	78
Kaplan-Meier Survival Analysis	78
Chi-Square Analysis (RECIST Standard)	83
Multivariate Analysis	84
Complete Dataset Results	85
Multivariate Results for Re-Tested Patients	87
Univariate Statistical Analysis.....	88
Sensitivity Analysis.....	93
CHAPTER 5 Conclusions and Recommendations.....	95
Conclusion	95
Restatement of Hypothesis	95
Recommendation.....	96
Proposed Standard of Care	96
Level of Evidence	97
Research Limitations.....	97
Sample Size	97
Data Quality.....	98

Inference Space	99
Limited Ability to Examine Covariates	99
Researcher Bias	100
Contribution	101
Future Research	102
Unanswered Question.....	103
Definition of Discordance	104
Impact.....	104
LIST OF REFERENCES	106
APPENDIX	112
Appendix 1: List of Abbreviations.....	113
Appendix 2: Incidence Rate.....	114
Appendix 3: Permissions	115
Periodical Permissions.....	115
NCCN Permission.....	115
Appendix 4: Operational Definitions	115
Appendix 5: Data Collection Sheet	118
Appendix 6: Data Breakdown Flowchart	119
Appendix 7: Variables Utilized.....	120
Appendix 8: Information Only Variables	124

Appendix 9: Discordant Patient Variables	126
Appendix 10: Initial Cox Regression Results	127
Appendix 11: Power Tables	128
Appendix 12: Simulation Definitions.....	141
Appendix 13: Table of MATLAB Programs.....	143
Appendix 14: MATLAB Program 1	144
Appendix 15: MATLAB Program 2	146
Appendix 16: MATLAB Program 3	148
Appendix 17: MATLAB Program 4	150
Appendix 18: MATLAB Program 5	151
Appendix 19: MATLAB Program 6	151
Appendix 20: MATLAB Program 7	153
Appendix 21: MATLAB Program 8	159
VITA	165

LIST OF TABLES

Table 1. Breast Cancer Staging	4
Table 2. ABC2 Relevant Standards	13
Table 3. ABC2 Levels of Evidence.....	15
Table 4. Receptor Discordance Rates	18
Table 5. Receptor Combinations.....	19
Table 6. Change in Treatment Plans.....	22
Table 7. Receptor Phenotypes	43
Table 8. Phenotype Treatments	43
Table 9. Literature Survival Data	50
Table 10. Power Table for L3.5 C0 C .01	60
Table 11. Power Table for L3.5 C0 C .05	60
Table 12. Power Table for L3.5 C0 B10 .05.....	61
Table 13. Power Table for L3.5 C0 B2 .05.....	62
Table 14. Power Table for L3.5 C50 C .05.....	64
Table 15. Power Table for L3.5 C0 C N85/15 .05.....	65
Table 16. Key Data File Names	67
Table 17. Data Collection Files	70
Table 18. Sample Size Completeness.....	74

Table 19. Sensitivity Analysis Records	75
Table 20. Sample Size Examples	77
Table 21. Univariate Analysis	89
Table 22. Appendix 1- List of Abbreviations	113
Table 23. Appendix - Seer Breast Cancer Incidence Rates	114
Table 24. List of Variables Utilized	120
Table 25. Information Variables	124
Table 26. Discordant Patients Variables	126
Table 27. L3.5 C0 C .01	129
Table 28. L3.5 C0 C .05	129
Table 29. L3.5 C0 B10 .01	130
Table 30. L3.5 C0 B10 .05	130
Table 31. L3.5 C0 B5 .01	131
Table 32. L3.5 C0 B5 .05	131
Table 33. L3.5 C0 B2 .01	132
Table 34. L3.5 C0 B2 .05	132
Table 35. L2.0 C0 C .01	133
Table 36. L2.0 C0 C .05	133
Table 37. L5.0 C0 C .01	134
Table 38. L5.0 C0 C .05	134

Table 39. L3.5 C25 C .01	135
Table 40. L3.5 C25 C .05	135
Table 41. L3.5 C25 B10 .01	136
Table 42. L3.5 C25 B10 .05	136
Table 43. L3.5 C50 C .01	137
Table 44. L3.5 C50 C .05	137
Table 45. L3.5 C50 B10 .01	138
Table 46. L3.5 C50 B10 .05	138
Table 47. L3.5 C0 C N70/30 .01	139
Table 48. L3.5 C0 C N70/30 .05	139
Table 49. L3.5 C0 C N85/15 .01	140
Table 50. L3.5 C0 C N85/15 .05	140
Table 51. MATLAB Programs Descriptions.....	143

LIST OF FIGURES

Figure 1. Breast Cancer Incidence Rates.....	3
Figure 2. Breast Cancer Mortality Rates	3
Figure 3. Liedtke Discordance Results (p. 1956).....	25
Figure 4. Niikura Discordance Results (p. 597).....	26
Figure 5. Karlsson (p. 6).....	27
Figure 6. Chang (p. 597)	28
Figure 7. Amir, Miller (p. 591).....	29
Figure 8. Wilking Survival Curves (p. 558)	32
Figure 9. Wilking Data Flowchart (p. 556).....	34
Figure 10. Study Decision Flow	44
Figure 11. Null Survival Curve.....	50
Figure 12. Null and Alternative Curves.....	51
Figure 13. SAS Power Curves.....	52
Figure 14. JMP Power Calculation.....	53
Figure 15. Simulation Flow	59
Figure 16. Actual Sample Sizes	72
Figure 17. Overall Survival Curve	79
Figure 18. Re-test versus Not Re-tested Survival Curve	80

Figure 19. Concordant versus Discordant Survival Curves	81
Figure 20. Metastatic Versus Primary.....	82
Figure 21. RECIST Chi Square Analysis.....	83
Figure 22. Whole Dataset Final Results.....	85
Figure 23. pPheno KM Analysis	86
Figure 24. Multivariate Results for Re-Tested Patients	87
Figure 25. Visceral versus Non-Visceral	90
Figure 26. Cox Regression with Visceral.....	91
Figure 27. Visceral Patients Only	92
Figure 28. Best Case Curves	93
Figure 29. Worst Case Curves	94
Figure 30. Periodical Permissions	115
Figure 31. Data Collection Sheet	118
Figure 32. Data Breakdown Flowchart	119
Figure 33. Initial Complete Cox Regression	127

CHAPTER 1

INTRODUCTION AND GENERAL INFORMATION

Critical Medical Decision

This research endeavor focuses on a medical decision. This decision is one that oncologists and breast cancer patients must make together regarding appropriate treatment plans upon the recurrence of metastatic breast cancer. Currently, there is no clear evidence based protocol that indicates how to determine the best first-line treatment plan after recurrence when a certain set of circumstances exist. These circumstances involve the discordance between protein receptors of the primary and metastatic tumors. The purpose of this research was to collect and analyze the data that would produce a definitive treatment protocol. To understand the complex nature of determining the appropriate treatment plan, it is important to understand some basic facts about breast cancer. A list of abbreviations is found in Appendix 1.

Breast Cancer

The American Cancer Society defines breast cancer as:

A malignant tumor that starts in the cells of the breast. A malignant tumor is a group of cancer cells that can grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body. The disease occurs

almost entirely in women, but men can get it, too ("Breast Cancer," 2015, p. 2).

This dissertation's focus was exclusively on breast cancer in women. Breast cancer generally develops in the milk ducts (ductal carcinoma) or the milk producing lobules (lobular carcinoma). Other, less frequent types of breast cancer also exist. Breast cancer is detected in several ways, ranging from mammography or other imaging technology, self-breast exams, or doctor's physical exams. Any of these methods may detect a potential tumor. Once a tumor is suspected, a biopsy is generally performed to determine whether the tumor actually exists and whether it is cancerous. In some cases, no tumor is found. In other cases, the tumor is benign (non-cancerous). In the remaining cases, breast cancer is diagnosed ("Breast Cancer," 2015).

Breast cancer is the most common form of cancer for women in the United States. A woman born today in the U.S. has a 1 in 8 lifetime risk of being diagnosed with breast cancer. In 2015, an estimated 232,000 case of breast cancer will be diagnosed in the U.S., affecting approximately 1,000 women, and a few men, per working day. It is likewise estimated that 40,000 women will die from breast cancer, actually metastatic breast cancer, this year ("Breast Cancer," 2015, pp. 9-10). Approximately 60,000 cases of non-invasive Stage 0 breast cancer, sometimes called pre-cancer, will also be diagnosed each year as well. These particular cases are not included in most incidence statistics because no medical consensus exists that Stage 0 is cancer or whether it will eventually transform into invasive cancer ("Breast Cancer," 2015).

Breast cancer trends can be somewhat difficult to estimate. Arguments among experts on measurement techniques and the impact of increased mammography testing can cloud the discussion. Figures 1 and 2 illustrate breast cancer trends since 1975. The data were obtained from the SEER database and are listed in

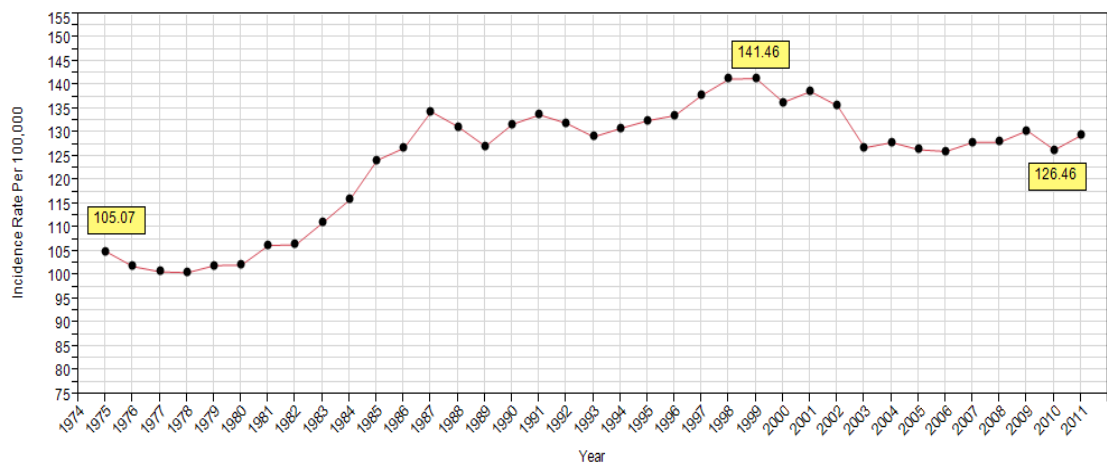


Figure 1. Breast Cancer Incidence Rates

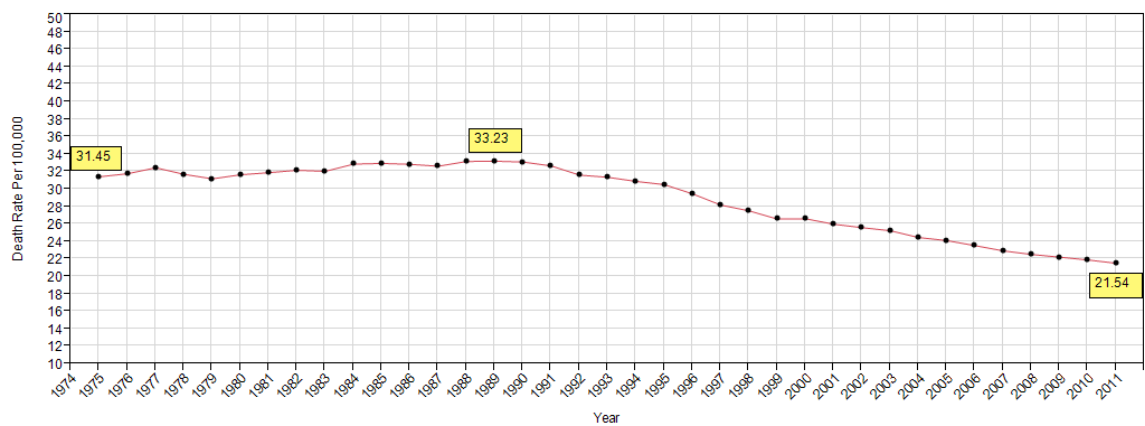


Figure 2. Breast Cancer Mortality Rates

Appendix 2 (Howlader N, 2015). Overall, the incidence rate of breast cancer in the United States has been increasing since 1975, with a peak in 1999 and a slight decline over the last ten years. The death rate peaked in 1989 at 33.23 deaths per 100,000 women. This rate has been decreasing for the last twenty five years, primarily due to early detection and chemotherapy (Berry et al., 2005; Howlader N, 2015). In 2011, the death rate reached a recent low of 21.5 deaths per 100,000 women in.

Breast Cancer Staging

Breast cancer is staged using the common Stage 0 – 4 scale which is summarized in Table 1 ("Breast Cancer," 2015, p. 42).

Table 1. Breast Cancer Staging

Stage	Description
0	Carcinoma In Situ. The cancer is in the original cell structure and has not invaded into nearby tissue. Considered "pre-cancer".
1	These cancer is still relatively small and either have not spread to the lymph nodes or have a tiny area of cancer spread in the sentinel lymph node.
2	These cancer is larger and/or have spread to a few nearby lymph nodes.
3	The tumor must be large (greater than 5 cm or about 2 inches across) or growing into nearby tissues (the skin over the breast or the muscle underneath), or the cancer has spread to nearby lymph nodes.
4	These cancer has spread beyond the breast and lymph nodes (at original diagnosis) to other parts of the body. Breast cancer most commonly spreads to the bones, liver, brain and lung.

Receptor Status.

A key diagnostic characteristic of breast cancer is determining the *receptor status* of the tumor ("Breast Cancer," 2015, p. 35). Receptors are proteins in or on human cells that promote the attachment of other proteins to the cell that fuel

growth. Two types of receptors that are of interest in breast cancer are progesterone (PR) and estrogen (ER). When present, these receptors, promote cancer cell growth by utilizing the naturally occurring hormones in a woman's body. If these receptors are present, the tumor is termed ER+ (ER positive) and/or PR+ (PR positive). The absence of these receptors are termed ER- and PR-, respectively. Another protein of interest is HER2 (human epidermal growth factor receptor 2), also referred to as HER2/neu. If this protein is overexpressed in a tumor, that tumor is termed HER2+, otherwise it is termed HER2- ("Breast Cancer," 2015, pp. 35-36).

When a women is first diagnosed with breast cancer, the tumor, referred to as the *primary tumor*, is tested to determine its receptor status. The receptor status is one aspect of the patient's diagnosis utilized to determine an appropriate treatment plan. If a patient's breast cancer recurs as a *metastatic tumor(s)*, it is now generally accepted practice to test the metastatic tumor's receptor status. When the metastatic tumor is tested (often referred to as *re-tested* in the literature), the metastatic tumor receptors are compared with the primary tumor receptor status. When the assessment of the primary tumor receptor status and metastatic tumor receptor status yield the same result, the tumor statuses are deemed concordant; when the results differ, the tumor statuses are deemed discordant.

Breast Cancer Treatments

Breast cancer is treated in a variety of ways. The determination of a treatment plan for a particular patient is based on the stage of the cancer at diagnosis and other physical and histological characteristics of the tumor. A comprehensive list of breast cancer treatments are outlined by the American Cancer Society and include those listed below ("Breast Cancer," 2015, pp. 49-92).

Breast Surgery. A lumpectomy can remove smaller tumors and conserves a majority of the breast, leaving a scar and some change in breast shape. The complete removal of the breast by a mastectomy may be required based on the size of the tumor or other complications. Occasionally, a patient chooses a prophylactic mastectomy (single or double) in order to lessen the risk of a recurrence of cancer in the breast. These options will often lead to some sort of elective reconstructive surgery (pp. 51-55).

Lymph Node Surgery. Lymph node surgery is performed to determine if the breast cancer has spread beyond the breast to lymph nodes underneath the armpit. One or more of these lymph nodes may be removed for evaluation. The presence of cancer cells in the lymph nodes indicates a higher probability that the breast cancer has spread to the bloodstream and thus to other locations in the body, such as the liver, lungs or brain. The more prevalent the cancer is in the lymph nodes, the more likely the patient will be prescribed chemotherapy as an adjuvant therapy ("Breast Cancer," 2015, pp. 57-59).

Radiation. Radiation therapy is used to destroy cancer cells with high-energy rays. Radiation is often administered to the breast and clavicle area after a lumpectomy and sometimes after a mastectomy to the breast and clavicle area to increase the probability that any unknown, but remaining cancer cells are destroyed (pp. 64–67).

Chemotherapy. Chemotherapy (chemo) is referred to as a *systemic* treatment. Systemic is a type of therapy that does not attack a known, specific cancer cell or tumor; rather it is utilized to reach cells throughout the body when cancer is known or suspected to have spread. Chemo is used in an attempt to kill cancer cells that may have spread from the breast, beyond the lymph nodes, to other areas of the body. Adjuvant chemotherapy is the precautionary use of chemo where there is no evidence the cancer has spread; however there exists a concern that it has spread, a decision that is generally based on the analysis of a

patient's lymph nodes (pp. 68-70).

Hormonal Therapy. Hormone therapy is another form of systemic therapy. It is most often used as an adjuvant therapy to help reduce the risk of the cancer recurrence after surgery if the patient is hormone receptor-positive. It is also used to treat cancer that has returned after the initial treatment. Approximately 2 out of 3 breast cancers are hormone receptor-positive. These cancers contain receptors for the hormones estrogen (ER+ cancers) and/or progesterone (PR+ cancers). Hormone therapy for breast cancer either lowers hormone levels or prevents hormones from assisting breast cancer cell growth. A woman's ovaries are the main source of estrogen until menopause. After menopause, smaller amounts of estrogen are still made in the body's fat tissue. Estrogen promotes the growth of cancers that are hormone receptor-positive. ("Breast Cancer," 2015, pp. 73-74). Tamoxifen is a common drug prescribed for hormone positive patients. Femara is a common drug prescribed for hormone positive, post-menopausal women.

Targeted Therapy. This terminology is used to describe systemic therapy for HER2+ patients. Herceptin (trastuzumab) is a common drug prescribed to breast cancer patients with HER/neu positive tumors (pp. 77-79).

Metastatic Breast Cancer

Metastatic breast cancer (MBC) is defined as breast cancer that has spread beyond the breast. If this spreading is observed when the breast cancer is first discovered, the cancer will be termed Stage 4. Traditionally, if the cancer recurs after the initial discovery and treatments, then it is typically referred to as recurrent metastatic breast cancer. Metastatic breast cancer has no cure. A person diagnosed with MBC will most likely die of MBC. The median life

expectancy of people diagnosed with MBC is between 2-3 years (Dawood, Broglio, Buzdar, Hortobagyi, & Giordano, 2010). Metastatic breast cancer is further defined by its location. If the cancer is confined to lymph nodes in the area of the breast, it is called *regional*. Otherwise, if it has spread to other organs, it is called *distant*. The statistic cited earlier that approximately 40,000 woman (and a few men) die annually from breast cancer is somewhat misleading. It is more accurate to state that yearly 40,000 people will die of *metastatic* breast cancer. In general, people do not die from cancer in the breast; they die when it has spread to other organs, thus metastatic breast cancer.

Impact

Currently, national databases do not track the number of women who are diagnosed with metastatic breast cancer. Given the constant mortality rate of approximately 40,000 women per year dying of MBC ("Breast Cancer," 2015), by extension, for each woman who dies another woman will be told she has metastatic breast cancer. Based on this logic, I estimate 40,000 U.S. women are diagnosed with metastatic breast cancer each year.

Problem Statement

For more than two decades, breast cancer researchers have studied the benefits, risks and clinical importance of testing the receptor status of metastatic breast cancer tumors. While a growing consensus exists that encourages re-testing of the metastatic receptor status, and the circumstances under which re-testing should occur, there is little to no evidence that changing first-line treatment plans based on the metastatic tumor status improves clinical outcomes. In fact,

there is evidence that changes to first-line treatment plans based on the re-tested tumor receptor status can indeed harm patient outcomes. The first-line treatment plan is the first treatment after diagnosis of metastatic breast cancer. It is common for MBC patients to cycle through many treatment plans over time as each treatment will eventually become ineffective. While extensive research on this topic exists, the most important question has yet to be addressed:

Does basing a patient's first-line treatment plan on the receptor status of metastatic tumors, when discordant with the receptor status of the primary tumor, improve patient outcomes?

The above critical question has not been directly or adequately addressed in the literature. In Chapter 2, I review the research related to this question and summarize both the research results and generally accepted practices that have evolved in the medical literature. I also propose an answer to my key question based on the literature review, and subsequently substantiate that answer with the data analysis provided in this dissertation.

The essence of this dissertation centers on the key decision that doctors and patients must make when all of the following conditions exist:

1. Metastatic breast cancer has recurred (this excludes an initial stage 4 diagnosis).
2. The metastatic cancer has been biopsied.
3. The receptor status of the metastatic tumor has been determined.
4. The receptor status of the metastatic tumor is discordant from that of the primary tumor.
5. The first-line treatment plan is being determined.

When all the above conditions occur, the oncologist and patient must make the following decision:

Should the first-line treatment plan be based on the receptor status of the primary tumor or on the receptor status of the metastatic tumor?

Currently, there is insufficient research based evidence to inform clear guidelines for national and international standards where discordance exists. The vague standards that do exist are discussed in the literature review section. I estimate 6,000 U.S. women a year are diagnosed with discordant receptor results and, along with their physicians, must decide among conflicting treatment plans indicated by the discordance. This estimate is based on my prior estimate that U.S. 40,000 women are diagnosed yearly with metastatic breast cancer and the 15% of those cases that I found in my data to be discordant. Around 521,900 people die of breast cancer each year world-wide ("Global Cancer Facts & Figures," 2012, p. 37). Using a conservative ratio of 10:1, I also estimate 60,000 women a year world-wide will have discordant receptor statuses and be faced with the decision investigated by this dissertation.

Hypothesis

Based on a thorough review of the literature, I hypothesized that:

In discordant cases, if the first-line treatment plan is based on the receptor status of the primary tumor, the median life expectancy of MBC patients will be longer than MBC patients whose first-line treatment plan is based on the receptor status of the metastatic tumor.

Research Purpose

The purpose of this dissertation was to determine the impact on *patient outcomes* in *recurrent* metastatic breast cancer diagnoses with discordant receptors when the *first-line* treatment plan is based on the receptor status of metastatic tumor(s) *instead of* the receptor status of the primary tumor(s). My ultimate goal is that the results of this dissertation leads to changes in international and national standards regarding the determination of first-line treatment plans in discordant cases of metastatic breast cancer, and to provide a higher level of evidence for those standards.

CHAPTER 2

Breast Cancer Literature Review

Review Criteria and Process

I conducted a thorough literature review on research related to metastatic breast cancer tumor retesting published prior to November 2014. Key phrases such as “metastatic receptor status,” “metastatic receptor test,” “discordance” and “hormone receptor” were queried during the electronic literature search. Searches for more recent and most cited publications were performed. The purpose of this critical literature review was to a) summarize the current data related to the risks and benefits of changing treatment plans based on metastatic tumor receptor status, b) assess the efficacy of the current clinical standards for assessing and utilizing metastatic tumor receptor status, and c) provide the data necessary to form my hypothesis.

Permissions

Permissions were obtained for all cited figures and extended quotes. Permission information is contained in Appendix 3.

Standards

Two medical standards organizations informed my research, the Advance Breast Cancer International Conference (Cardoso et al., 2012; Cardoso et al., 2014), and the National Comprehensive Cancer Network® (NCCN.org, 2014).

These organizations provide physicians with care standards recommendations based on the best available evidence. These standards often inform the clinical practice guidelines for medical facilities. Each standards organization has a defined level of evidence (LoE) scale to support and add context to their recommendations. As stated in Chapter 1, the goal of this dissertation is to impact the standards relevant to my hypothesis provided by these organizations.

Advanced Breast Cancer (ABC) Standards

“The ABC Consensus Conference was created by the European School of Oncology (ESO) with the ambitious goal of improving outcomes for all patients with ABC” (Cardoso et al., 2014, p. 1). This organization first met in 2011, resulting in ABC1 standards (Cardoso et al., 2012). The conference met again in 2013 (ABC2) and updated the standards (Cardoso et al., 2014). These standards focus specifically on patients with metastatic breast cancer. The current ABC2 standards (Cardoso et al., 2014, pp. 10-11) that are relevant to my research are found in Tables 2.

Table 2. ABC2 Relevant Standards

Guideline	LoE
A biopsy (preferably providing histology) of a metastatic lesion should be performed, if easily accessible, to confirm diagnosis particularly when metastasis is diagnosed for the first time	1C
Biological markers (especially HR and HER-2) should be reassessed at least once in the metastatic setting, if clinically feasible.	2C
If the results of tumour (sic) biology in the metastatic lesion differ from the primary tumour (sic), it is currently unknown [emphasis mine] which result should be used for treatment-decision making. Since a clinical trial addressing this issue is difficult to undertake [emphasis mine], we recommend considering [emphasis mine] the use of targeted therapy (ET and/or anti HER-2 therapy) when receptors are positive in at least one biopsy, regardless of timing.	Expert Opinion

Levels of Evidence

The relevant levels of evidence definitions used by ABC2 are listed in Table 3 (Cardoso et al., 2014). The levels of evidence definitions utilized by the ABC conference were originally published by Guyatt et al. (2006). The meaning of *expert opinion* was not defined in the literature I reviewed. I therefore inferred that expert opinion equates to no evidence available as defined by the LoE criteria, and the recommended standard is based solely on the clinical opinions of those assembled to reach consensus. On most guidelines where a level of evidence was defined, cited references supported the guideline. No references were cited to support the guideline listed as expert opinion in Table 2.

NCCN Standards

The National Comprehensive Cancer Network® (NCCN) is a not-for-profit organization consisting of 26 of the world's leading cancer centers. NCCN creates clinical practice guidelines for clinicians and healthcare professionals (NCCN.org). All NCCN guidelines I reference are reported at a 2A level of evidence, unless otherwise noted. NCCN defines a 2A level of evidence as one with a lower-level evidence, with uniform NCCN consensus that the intervention is appropriate (NCCN.org, 2014, p. MS1).

The relevant NCCN guidelines state:

The NCCN Panel recommends that metastatic disease at presentation or first recurrence of disease should be biopsied as part of the workup for patients with recurrent or stage IV disease. This ensures accurate determination of metastatic/recurrent disease and tumor histology, and allows for biomarker determination and selection of appropriate treatment.

Table 3. ABC2 Levels of Evidence

LoE	Quality of Supporting Evidence	Implications
1A	Randomized Clinical Trials (RCTs) without important limitations or overwhelming evidence from observational studies	Strong recommendation, can apply to most patients in most circumstances without reservation
1B	RCTs with important limitations or exceptionally strong evidence from observational studies	Strong recommendation, can apply to most patients in most circumstances without reservation
1C	Observational studies or case studies	Strong recommendation, but may change when higher quality of evidence becomes available
2A	Randomized Clinical Trials (RCTs) without important limitations or overwhelming evidence from observational studies	Weak recommendation, best action may differ depending on circumstances or patients'
2B	RCTs with important limitations or exceptionally strong evidence from observational studies	Weak recommendation, best action may differ depending on circumstances or patients' or societal values
2C	Observational studies or case studies	Very weak recommendation, other alternatives may be equally reasonable
Expert Opinion	No Evidence, just the best guess of those on the committee <i>(This is my definition based on researching the designation)</i>	Data is needed

Determination of hormone receptor status (ER and PR) and HER2 status should be repeated in all cases when diagnostic tissue is obtained. ER and PR assays may be falsely negative or falsely positive, and there may be discordance between the primary and metastatic tumors...

The NCCN Panel recommends that re-testing the receptor status of recurrent disease be performed, *especially* in cases when it was previously unknown, originally negative, or not overexpressed. For patients with clinical courses consistent with hormone receptor-positive breast cancer, or with prior positive hormone receptor results, the panel has noted that a course of endocrine therapy is reasonable, regardless of whether the receptor assay is repeated or the result of the most recent hormone receptor assay (NCCN.org, 2014, p. MS45).

Discussion of Standards

The level of evidence relating to a first-line treatment plan choice for MBC patients with discordant receptor results is insufficient, and lacks citations to literature to support decision making. ABC2 states the level of evidence is expert opinion. The NCCN standard states the level of evidence is 2A; however there are no citations supporting that recommendation. These standards also contain weak language such as *reasonable*, *currently unknown*, and *recommend considering*. Further research is needed to better inform standards, and to improve patient outcomes related to treatment decisions.

Retesting of Metastatic Tumor Receptors

Numerous papers have been published addressing the question of whether metastatic tumors should be biopsied and tested for receptor status

determinations. The overwhelming consensus in the literature points to re-testing and is best summarized as follows:

While biopsy of the metastatic site may not be prospectively able to demonstrate improved outcome for each patient – at least for the time being – it is paramount to introduce a culture of systematically biopsying (sic) metastases in order to gain information on the biology of the disease and, accordingly, select the best treatment option for our patients (Criscitiello et al., 2014, p. 6).

One of the key reasons re-testing has become more prevalent is the use of fine needle aspiration (FNA) which makes biopsies of smaller tumors more accessible and less painful to patients (Foukakis, Astrom, Lindstrom, Hatschek, & Bergh, 2012). Routine FNA, coupled with researchers' desire for more robust clinical data, has recently led to protocol revisions by ABC2 and NCCN that include recommendations for performing biopsies and comparing primary and metastatic tumors' receptor statuses. Until such metastatic biopsies became common, MBC patients' first-line treatment plans were based on the receptor status of the primary tumor, since the metastatic tumor status was not able to be assessed without the biopsy.

Criscitiello et al. (2014), provides an instructive review of the benefits, risks, and the how and when to biopsy metastatic tumors. These topics are also well covered by Foukakis et al. (2012). Articles by Penault-Llorca et al. (2013) and Turner and Di Leo (2013) provide excellent decision flowcharts for making biopsy and treatment determinations. Based on the weight of the evidence, biopsies of the metastatic tumor(s) and re-testing the receptor status has become standard practice. The NCCN guideline on re-testing metastatic tumors recommends "that metastatic disease ... be biopsied ... This ... allows for biomarker determination and selection of appropriate treatment" (Carlson et al., 2012, p. 3). As a

consequence of following the NCCN or ABC2 standards on re-testing, the determination of the proper treatment plan to prescribe when the receptor status of the metastatic tumor and the primary tumor are discordant is a common conundrum in current clinical practice.

Discordance

Many researchers study the percentage of discordance between the receptor status of the primary tumor and the metastatic tumor (Amir, Clemons, et al., 2012; Aurilio et al., 2014; Dieci et al., 2012; Farolfi, Ibrahim, Scarpi, & Amadori, 2013; Niikura et al., 2012). A recent meta-analysis by Aurilio et al. (2014) summarizes 48 articles totaling 3,000 – 4,000 tumors. Table 4 shows the pooled discordance rate found in that article.

Table 4. Receptor Discordance Rates

Receptor	Discordance
ER	20%
PR	33%
HER2	8%

How is Discordance Measured?

A cancer tumor contains three receptors of interest, and each can be positive or negative, thus there are eight (2^3) total combinations. In the literature, most authors call a difference in any receptor status, primary versus metastatic, a discordance. Thus, a change from one of the possible eight combinations to one of the remaining seven combinations, is considered a discordance. There are 56 (8×7) such discordances. However, not all of those discordances would dictate a

treatment change. For example, if a patient's receptor status is ER+, PR+ and HER2- in their primary tumor and ER+, PR-, HER2- in their metastatic tumor, the treatment plan will be the same, since the ER+ status indicates the same treatment for either set of results. Thus, not all discordances are clinically important for determining treatment plans. However, very few authors delineate between the possible discordant patterns. The receptor combinations and their corresponding descriptions are listed in Table 5. Clearly and strictly defining discordance and its impact on treatment plans is an important issue in this research, and a weakness in the current literature.

Table 5. Receptor Combinations

Receptor Combination	Description
ER-, PR-, HER2-	Triple Negative (TNBC)
ER+, PR+, HER2-	Hormone Receptor Positive (HR+)
ER+, PR-, HER2-	ER positive, PR negative (HR+)
ER-, PR+, HER2-	ER negative, PR positive (HR+)
ER-, PR-, HER2+	HER2 Positive, HR- (HER2+)
ER+, PR+, HER2+	Positive Breast Cancer (PBC)
ER-, PR+, HER2+	HR positive, HER2 positive (PBC)
ER+, PR-, HER2+	HR positive, HER2 positive (PBC)

What Causes Discordance?

Ultimately, it may or may not be important to understand the causes of discordance. Some theories in the literature as to the causes of discordance are outlined in this section for completeness. However, for my particular research, the cause of discordance is not considered. My research interest lies in whether there is discordance, not why, and whether that discordance leads to altered

patient treatment plans, and whether those altered plans lead to better or worse clinical outcomes.

One of the most common reasons proposed for discordance is the possibility of measurement error. That is, the two tumor receptor status results differ not because the tumors are different, rather there exists variation (inaccuracy) in the measurement process itself. The primary tumor can usually be measured based on the extraction of the full tumor. The metastatic tumor is measured on the basis of a FNA biopsy, given that a metastatic tumor can rarely be removed. A tendency in the literature is to assume the metastatic receptor status measurement is subject to more error (Foukakis et al., 2012).

Sighoko, Liu, Hou, Gustafson, and Huo (2014), use Bayesian methods to *extract* measurement error in order to estimate the *true* discordance rates. Their study provides some insight into the question of what percent of discordance may be due to measurement error. Liedtke et al. (2009), is an excellent example of a study where measurement error is contemplated as a key cause of discordance. The authors state:

We hypothesize that the poor survival outcome of patients with discordant receptor results may be due to both false-negative results that could lead to withholding endocrine therapy (or trastuzumab) and false-positive receptor results may also contribute some by leading to an initial period of ineffective therapy with targeted agents in patients who do not benefit. ...mistakes in receptor determination can lead to suboptimal therapy. ...our results illustrate the need to increase standardization and implementation of guidelines for hormone and HER2 receptor determination" (p. 1957).

There is not a consensus in the literature on the percentage of discordance that is due to true biological difference and the percentage due to measurement error. Based on the data I evaluated, I estimate that measurement error accounts for 33% to 50% of discordances. This range is a qualitative estimate only; thus it is not possible to put statistical confidence on the estimate. Discussions about other causes of discordance are less conclusive. A good example of other hypothesized reasons for discordance is found in Dieci et al. (2012) where they indicate that in addition to measurement errors, the heterogeneity of cancer tumors and the impact of previous treatments could also be reasonable explanations for a change in receptor status (p. 4).

Foukakis et al. (2012) also outlines the potential causes of discordance:

However, the reason for this discrepancy is less clear and both technical and biological explanations have been discussed. The use of various methods for ER (biochemical versus IHC assays) and HER2 (IHC versus FISH), the variability caused by sampling methods (fine-needle aspiration (FNA) or core biopsy versus surgical excision in the primary tumour (sic)) and differences in analysis of samples from different tissues are all possible technical caveats that could cause a false discrepancy (pp. X351-X352).

Hoefnagel et al. (2013), demonstrate that tumors among different metastases within the same patient have 10% discordance with clinical consequences. This result supports the heterogeneity of breast cancer and casts some concern about relying on just one metastatic sample site for testing receptor status when multiple sites exist. The data is clear that discordance exists, yet it is unclear as to why. However, for this dissertation, the key question remains whether discordances affect treatment decisions for patients and what are the implications such decisions have on clinical outcomes.

Impact on Treatment Decisions

First-line treatment plans in the metastatic setting are based on the receptor status of the disease, along with other factors. Several authors, directly or indirectly, explore the impact that retesting tumor receptors has on the establishment of first-line treatment plans. These authors estimate the percent of treatment plan decisions that occur based upon the receptor status of the metastatic tumor, when that status is discordant with the primary tumor. Table 6 summarizes five articles that expressly address the impact of discordance on first-line treatment plans.

Table 6. Change in Treatment Plans

Author	Year	Percent of First-line Treatment Plans Changed for MBC Patients Because of Discordance
(Lindstrom et al., 2012)	2012	estimated > 50%
(Aurilio et al., 2013)	2013	17 of 27 discordances
(Dieci et al., 2012)	2012	13 of 27 patients
(Thompson et al., 2010)	2010	24 of 34 patients
(Amir, Clemons, et al., 2012)	2012	41 of 113 discordances

The Aurilio et al. (2013) article reports 17 of 27 discordances resulted in changes to treatment therapies for MBC patients. Based on the reported data, it is not possible to determine whether some changes are for the same patient, as the data was reported by discordance, not by patient. While many changes to treatment plans are also made for concordant receptors, the authors imply that those changes are based on "... disease progression ..." instead of the re-tested

receptor results (Aurilio et al., 2013, p. 1651). Amir, Clemons, et al. (2012), indicated 41 patients had their treatment plans changed after a metastatic biopsy resulted in discordant results. One hundred thirteen changes in receptor statuses were documented in the study. Because discordance is not clearly defined by the authors, the 113 changes in receptor status can be interpreted in several ways. That is, some of the changes may not indicate a change in standard therapy and some changes may occur in the same patient (i.e. both ER and PR changed from plus to minus). Thus, the 113 changes do not necessarily indicate 113 patients had changes. The authors indicate that 14% of those patients whose tumors were biopsied had their treatment changed based on the receptor status of the metastatic tumor (Amir, Clemons, et al., 2012). It is important to note that this 14% is of patients whose tumors were biopsied, *not the percent of those with tumor receptor status discordance*. It is unclear what percent of patients with discordant results had their treatment plan changed.

Based on the body of research reviewed, I estimate that currently 60% to 75% of MBC patients with discordance have their treatment plans changed based on the receptor status of the metastatic tumor. Again, no statistical confidence can be provided to this range due to the variation of, and uncertainty in the reporting methods utilized by the authors included in this literature review. In chapter 4, I report the percentage of discordant cases where the treatment plans are changed in my retrospective study. Based on the current literature and recommended clinical standards, doctors are performing more biopsies of metastatic tumors, re-testing the receptor status and evaluating the discordance with the primary tumor when determining first-line treatment plans. This increased testing warrants the examination of clinical outcomes for patients with first-line treatment plans based on metastatic tumor receptor statuses.

Clinical outcomes

In the Conclusion sections of several articles, the authors summarize their findings regarding the *clinical outcome* of discordant receptors (Chang et al., 2011; Duchnowska et al., 2012; Idirisinghe et al., 2010; E. E. Lower, Glass, Blau, & Harman, 2009; Montagna et al., 2012; Wilking et al., 2011). These particular authors focus on the impact of discordance on life expectancy, without directly testing the impact of changing treatment plans based on that discordance.

Additionally, some researchers note the clinical implications of changing treatment plans in the Discussion sections of their papers, yet offer no definitive statements in the Conclusion sections of their research. The lack of clearly stated conclusions related to the clinical benefits of changing treatment plans make it difficult to quantitatively summarize the literature. I utilized all the information in the literature search, not just the authors' stated conclusions, to ensure a comprehensive analysis. The inclusion of all information led to clear, emergent patterns that inform the clinical consequences of changed treatment plans.

The research derived from my literature review provides clear evidence that *changing treatment plans based on the metastatic tumor receptor status is more harmful than helpful*. A strong illustration of this finding is found in Liedtke et al. (2009, p. 1953) where the authors assert that "discordant cases have poor survival probably due to inappropriate use of targeted therapies." The implication is that patients with discordant receptor statuses are given treatments contrary to what their primary tumor statuses indicate, and thus this change is the likely cause of the poor results in life expectancy. The authors report that the median post recurrence survival (PRS) rate for triple negative breast cancer (TNBC) patients who have concordant receptors (thus no change made to treatment based on metastatic tumor status) is 43 months. TNBC

patients (primary tumor) whose metastatic tumor receptors were discordant, experienced a 15.6 month median PRS (Liedtke et al., 2009, p. 1956). PRS represents the survival rate of patients over time, calculated from the time the metastatic recurrence is documented. This is the best and most common measurement for *clinical outcomes* in the literature. Graph A in Figure 3 depicts the TNBC patients only. Graph B represents the entire research sample.

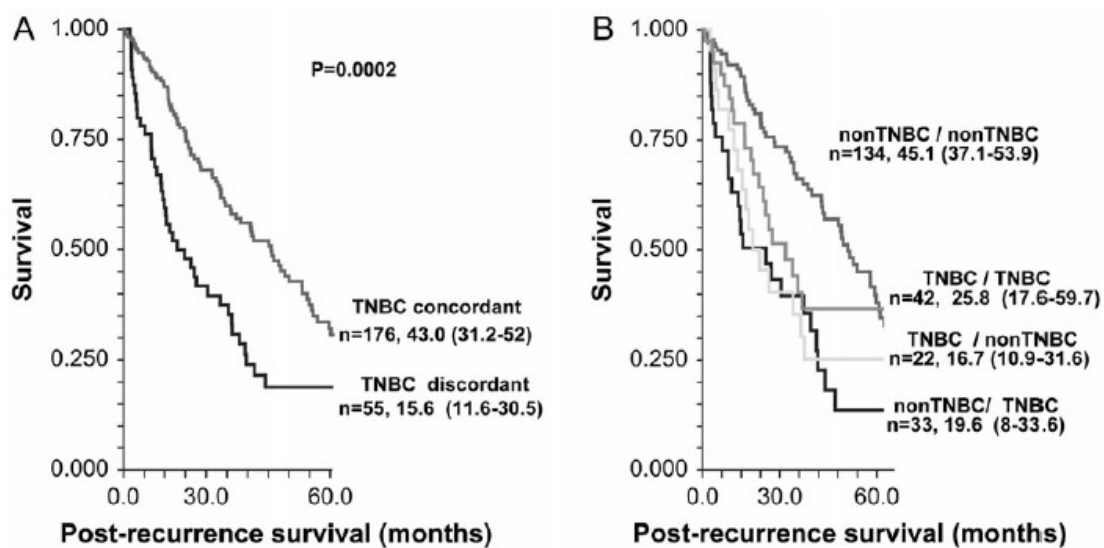


Figure 3. Liedtke Discordance Results (p. 1956)

The authors imply that altering treatment plans for discordant patients translates to poor clinical outcomes. Another example supporting the harmful outcomes that result from changing treatment plans based on discordance is discussed by Dieci et al. The authors indicate that “the impact of this approach on patient management and outcome is not yet clear, (and) induce(s) clinicians to modify the treatment choice in ~14% of the cases and that a change in the receptor status is associated with poorer survival” (Dieci et al., 2012, p. 1). As noted

earlier in this chapter, the 14% refers to the percentage of all biopsied patients, not the percentage of patients with discordant receptors.

Niikura et al. (2012, p. 598), finds that “patients with HER2 discordance between their primary and metastatic tumors have shorter OS. However, for our data, inaccurate testing may have caused discordance between primary and metastatic sites.” OS is the authors’ abbreviation for *Overall Survival*, which is equivalent to Post Recurrence Survival (PRS). The authors’ reference to *inaccurate testing* refers to measurement error. Figure 4 contains Niikura’s results.

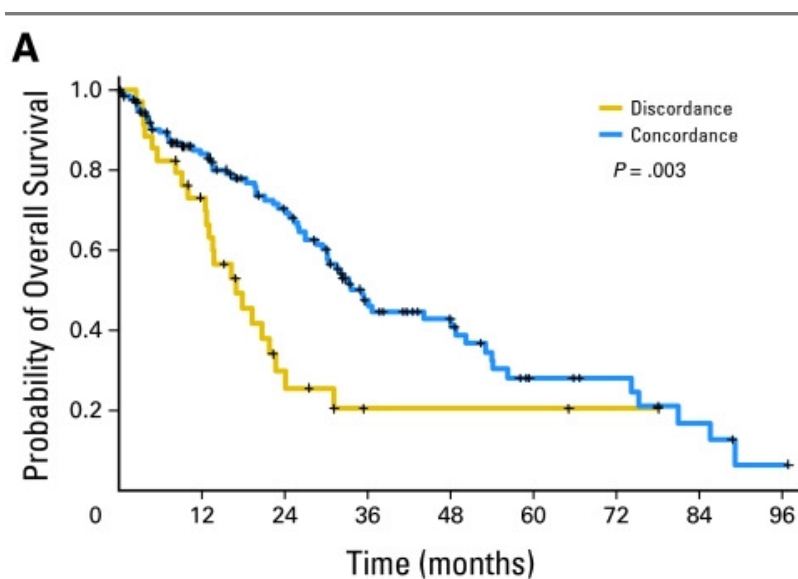


Figure 4. Niikura Discordance Results (p. 597)

Karlsson et al. (2014, p. 1), notes that changes from ER+ or PR+ in the primary tumor to ER- or PR- in the metastatic tumor result in an increased risk of death with hazard ratios of 3.62 and 2.34, respectively. These patients had their treatment plan changed and were *denied* endocrine therapy because of the receptor status of the metastatic tumor. The authors note that none of the

patients whose ER+ status changed to ER- received the endocrine therapy they would have received had their receptors been concordant. This important result supports my overall hypothesis and is shown in Figure 5. Note that in Figure 5 the worst performing patients are those whose ER status changed from positive to negative.

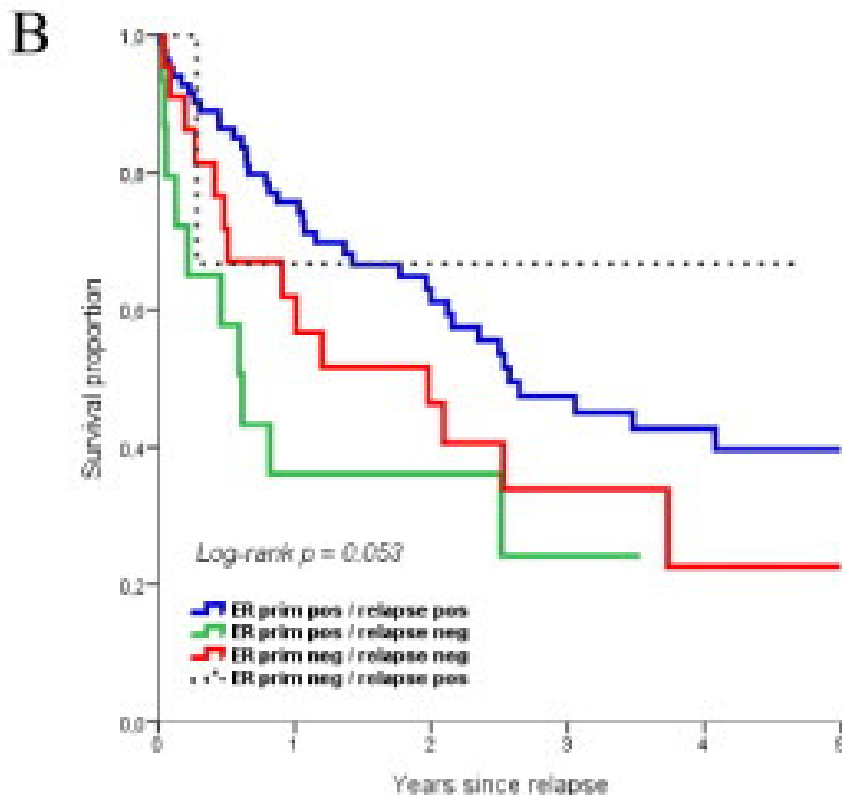


Figure 5. Karlsson (p. 6)

In a study on HER2 discordance, Chang et al. (2011, p. 598), states that changing treatment plans has "...an inferior outcome..." when compared to concordant cases. While their sample size is small ($n=18$), the data analysis supports my hypothesis. Those patients in Chang's sample whose treatment plan is based on the receptor status of the metastatic tumor in discordant cases had

negatively impacted survival rates. Chang's graphs are provided in Figure 6, where Group 2 has concordant results and Group 3 has discordant results (Chang et al., 2011, p. 597).

The weight of the literature review evidence indicates that changing treatment plans based on discordant receptor status results yields poorer clinical performance as measured by patient survival time post recurrence.

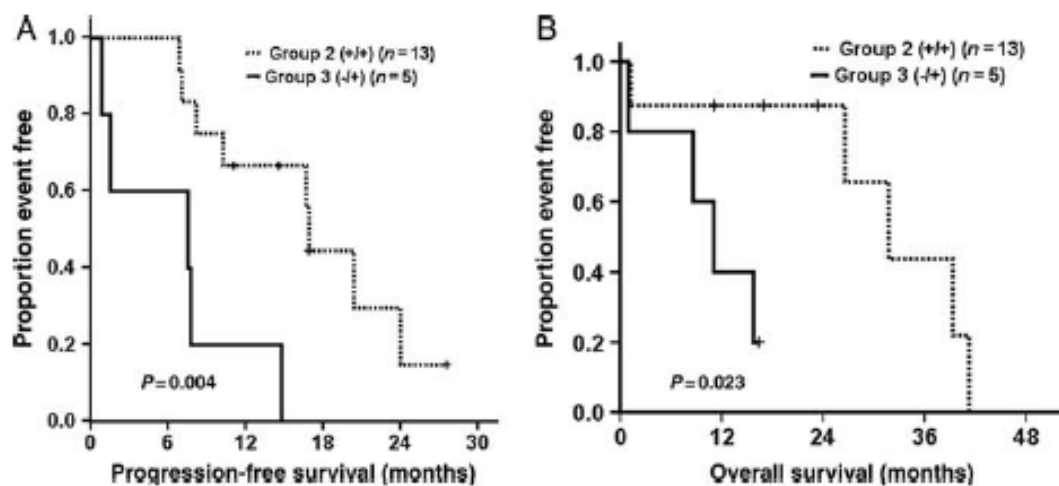


Figure 6. Chang (p. 597)

Potential Contradictory Data

Several articles provide some evidence that discordant results lead to equal or better outcomes for patients. Amir, Miller, et al. (2012), find no difference in survival rates for concordant and discordant results in MBC patients' tumors. They state "...after a median follow-up of 12 months, there were no trends for an association between receptor discordance and either time to treatment failure or overall survival" (p. 587). I have two concerns about their data. First, only

about 30% of the discordant cases had their treatment changed based on the receptor status of the metastatic tumor, as opposed to the 60% to 75% of cases found in most of the literature review. Therefore, discordant and change in treatment are not as confounded here as opposed to the typical studies in this review, thus supporting the notion that no change in treatment plans leads to better results. Additionally, the median patient follow-up is only 12 months, which is not sufficiently long enough for estimating 5 year survival curves. See Figure 7 for their results (p. 591).

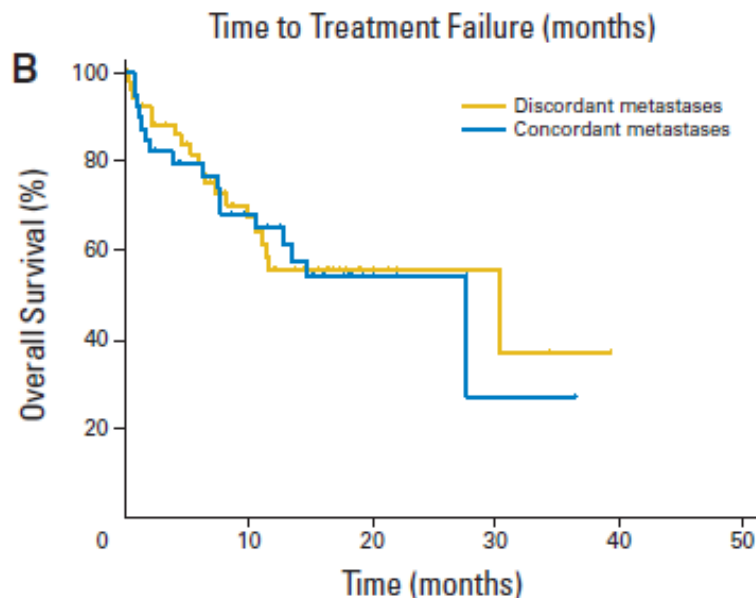


Figure 7. Amir, Miller (p. 591)

E. E. Lower et al. (2009), researching only HER2 status finds that "... fifteen percent of patients with primary tumors negative for HER-2/neu distant metastasis had HER-2/neu overexpression. These patients experienced improved survival compared to those patients with HER-2/neu negative primary and metastatic lesions" (p. 305). In other words, 15% of the patients changed from

HER2- in their primary tumor to HER2+ in their metastatic tumor. The authors do not describe how patients' metastatic treatment plans are determined. Therefore it is not possible to determine the effect of first-line treatment decisions in this study.

Duchnowska et al. (2012), find no difference in survival for discordant cases when measured from the time of the *initial diagnosis* with breast cancer. Evaluating survival from the time of initial diagnosis is not prudent in this research. There are too many variables that affect overall survival times before the recurrence, such as initial treatments, adjuvant therapies, age at diagnosis and BRCA status. Survival time from recurrence is the best metric and the one used in all other research I reviewed.

One article does conclude that *better* results are obtained by retesting receptors and utilizing the metastatic tumor results. Their conclusion is "we observed that 18 out of 100 biopsied patients (18.0%) had a conversion of predictive factors which allowed adjusting for therapy, ... Those 18 patients showed a better survival compared to the other 82 biopsied patients ..." (Botteri et al., 2012, p. 284). The authors imply that 18 patients have discordant results and that those patients' treatment plans are changed based on the metastatic tumor status, and those changes result in better clinical outcomes. My analysis of this article exposed several research flaws. First, this study is restricted to patients with liver metastases. Thus, it is not prudent to infer these results are applicable to other locations of metastases. Second, the sample size of patients with changed treatments is small, (n = 18). Third and most importantly, 10 of the 18 patients with adjusted treatment plans do not necessarily have a change in HER2 status. These 10 patients did not have their primary tumor tested for HER2 because HER2 testing was not yet developed at the time of their initial cancer diagnosis. These particular patients are HER2 positive in their metastatic tumor and they

received Herceptin for the first time. The authors believe it is reasonable to assume these patients were originally HER2- because about 80% of the current patient population is HER2- in their primary tumor (p. 287). It is unreasonable to make an assumption that impacts over half of the data (10 of 18). Additionally, one could argue that since these patients had not initially been treated with Herceptin and eventually had a recurrence, they are therefore more likely to have been HER2+ in their primary tumor.

Summary of Clinical Implications

Foukakis et al. (2012), notes that, "...carrying out a metastatic biopsy had an impact on treatment decisions ... no conclusions could be drawn from these analyses regarding the impact of this practice on survival" (p x351).

I disagree with Foukakis about drawing conclusions regarding the impact on survival. The weight of the evidence is clear and indicates that basing first-line treatment plans on the metastatic receptor status when discordant with the primary tumor receptor status has an *overall negative impact on survival*.

Challenges Interpreting the Literature

I encountered two primary difficulties interpreting the findings and conclusions of articles in the literature review. First, the definition of discordance often varies from author to author and is sometimes not clearly defined. Additionally, it is rare for an author to offer a definitive discussion about the impact of basing first-line treatment plans on metastatic receptors for patients with discordant receptors. Together, these two challenges make it impossible to conduct a precise quantitative meta-analysis of the literature review or easily make comparisons between studies. The impact of the variation and ambiguity in

discordance definitions has been discussed throughout this chapter. The difficulty measuring the impact of changes in treatment plans based on inconsistent and confusing reporting is best illustrated by Wilking et al. (2011). The key results from this study are reflected in their graphical analysis presented in Figure 8. The inserted note in Figure 8 is mine and not the authors.

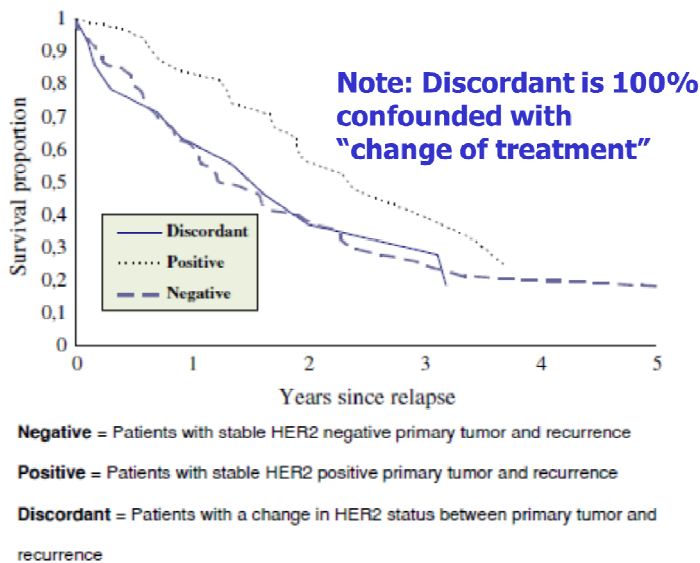


Figure 8. Wilking Survival Curves (p. 558)

The authors indicate that there is indeed discordance in their dataset with respect to HER2 and that changes in HER2 status resulted in a "significantly increased risk of dying" for those patients (Wilking et al., 2011, pp. 557-558). My interpretation of the authors' description of the data is that each patient with discordant metastatic receptors also had their first-line treatment plan based on the metastatic receptor status instead of the primary receptors. Thus, *change in treatment plan* is completely confounded statistically with discordance. A statistical finding about discordance is also a statistical finding about changing treatment plans based on the metastatic tumor receptors. This confounding is a

problem throughout the literature and highlighted by Turner and Di Leo's literature review when they state that the "interpretation of the prognostic impact of discordance is limited by the confounding influence of any treatment alteration based on the metastatic biopsy result" (Turner & Di Leo, 2013, p. 950). The curves in Figure 8 indicate worse survival results for discordant patients when compared to HER+ concordant patients. The key question is whether that difference in survival can be attributed to the discordance or to changes in patient treatment plans based on the discordance. That is, what would the survival curves demonstrate if they are plotted based on whether the patients' first-line treatment plan is based on the primary or metastatic tumor receptor statuses? In order to attempt to recast the analysis based on treatment plan determination, I carefully reread the article to determine the first-line treatment plans for the discordant patients; that is, are their treatments based on the metastatic tumor or the primary tumor receptor status? This analysis became a futile exercise. The authors' data flow diagram is reproduced in Figure 9 (Wilking et al., 2011, p. 556). I inserted the lettering in Figure 9 for discussion purposes. From the flowchart, it is clear that 43 patients (box E) have HER2+ primary tumors and 42 (boxes A and D) have HER2+ metastatic tumors. Fifteen patients (boxes B and D) have discordant tumors. Utilizing the authors' flowchart and their commentary about that flowchart, I first concluded that 27 to 29 of the 43 patients with HER2+ primary tumors received trastuzumab (prescribed for HER2+ cancer) after recurrence (p. 557). Based on this calculation, one would assume that a maximum of 29 of the 42 patients with HER2+ metastatic tumors received trastuzumab. The authors also indicate that *a//* patients (42) with HER2+ metastatic disease received trastuzumab (p. 557). The possible number of patients treated for HER2+ metastatic cancer is now 27, 29 or 42. In another section of this article, it is indicated that 23 to 26 patients are treated for HER2+ cancer (p. 557). After stating that all 15 discordant patients receive trastuzumab,

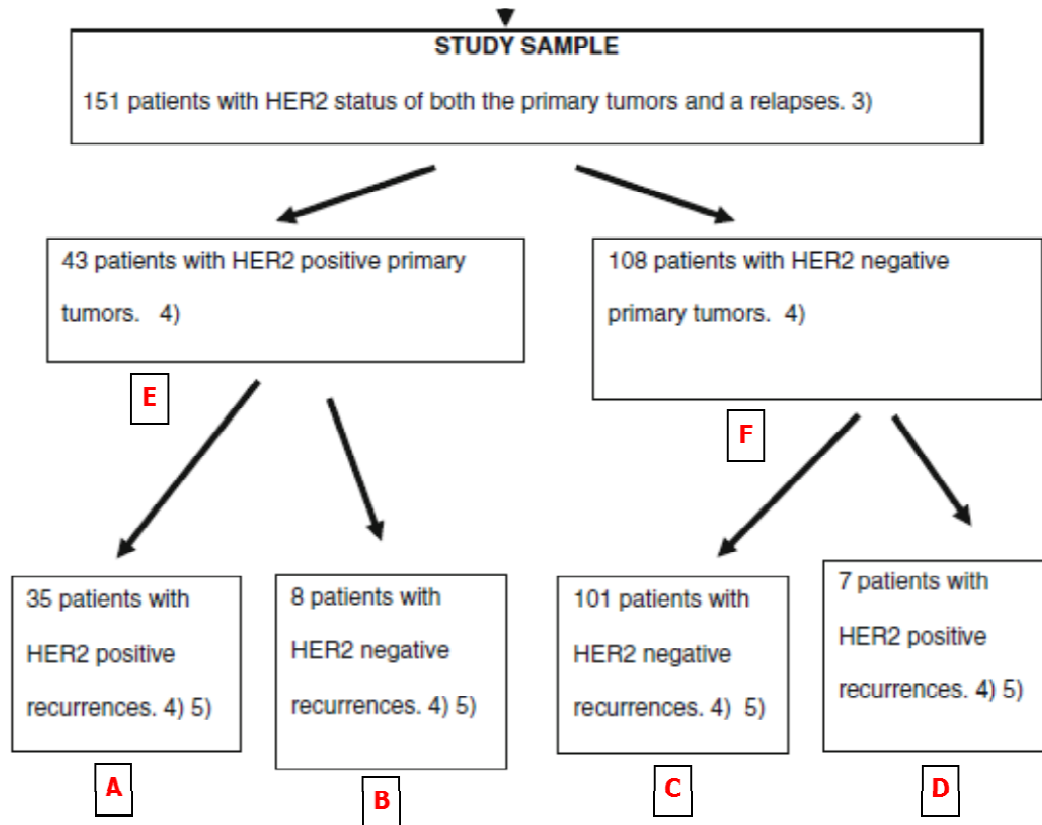


Figure 9. Wilking Data Flowchart (p. 556)

the paper indicates that "... of the 15 patients with change in HER2 status, 5 (33%) received trastuzumab" (Wilking et al., 2011, p. 557).

Based on the confusing information in the article, I cannot clearly determine the number of discordant patients who had their treatment plans changed. A reasonable interpretation is that most or all discordant patients have their treatments changed based on the metastatic tumor status instead of the primary tumor HER2 status. Figure 8 illustrates that discordant patients fared worse than concordant ones. Thus, the obvious conclusion is that retesting the HER2 receptor status and basing first-line metastatic treatments on that test is harmful to patients and should *not* be done. However, the authors state "the conclusion of our study is that a number of patients, who experience a recurrence, will not be managed correctly, if therapy is only based on characteristics of the primary tumor" (p. 559). This type of confusion and confounding found in the data of the relevant literature makes the study in this dissertation all the more necessary and important.

Progress in the Literature

When I began the literature review, I expected to find that someone had discovered the answer to my proposed research question: *Does basing a patient's first-line treatment plan on the receptor status of metastatic tumors, when discordant with the receptor status of the primary tumor, improve patient outcomes?* Most authors are researching the clinical impact of discordance, but are not investigating how clinical outcomes are affected when the first-line treatment plans of metastatic patients are based on the receptor status of the metastatic tumor instead of the primary tumor. The most current research is still

focused on the survival impact of discordance, rather than the survival impact of changing treatment plans based on the discordance.

One of the earliest articles I found on this subject was published in 1989 and the stated purpose of the study in the article is to determine the prognostic impact of ER discordance (Kamby, Rasmussen, & Kristensen, 1989, p. 1). Three years later, Spataro et al. (1992), seeks to "... determine the clinical relevance of the subsequent ER determination" (p. 733). Next, Kuukasjarvi, Kononen, Helin, Holli, and Isola (1996) attempt to determine if a failure to respond to endocrine therapy was due to discordance between primary and recurrent receptor status (p. 2584). In 2005, the impact of changes in metastatic receptors on "... change on survival" is investigated (Elyse E. Lower, Glass, Bradley, Blau, & Heffelfinger, 2005, p. 65).

Some 25 years after Kamby's research, Yang et al. state:

The aim of this study was to evaluate the frequency and prognostic impact of changes in the estrogen (ER), progesterone (PR) and human epidermal growth factor receptor 2 (HER2) status between primary and recurrent/metastatic lesions (RML) (Yang et al., 2014, p. 1).

After 25 years of research, Turner and Di Leo (2013) conclude that the data on the clinical impact of discordance and changing treatment plans is "...limited, inconclusive and conflicting" (p. 954). I disagree with this conclusion. Careful reading of the available literature, which is not limited, indicates a *conclusive* pattern that changing treatment plans based on the receptor status of the metastatic tumor results in more harm than good, shortening the life expectancy of metastatic breast cancer patients. The research in the literature that conflicts with this pattern is limited and weak.

Clinical Standards Recommendation

The current medical standards that relate to the topics of this dissertation are discussed earlier in this chapter. Based on my review of the literature, I offer the following over-arching standard of care for metastatic breast cancer patients with discordant receptor status results:

Where discordance between the primary and metastatic tumor receptor status would indicate different treatments, the status of the primary tumor should take precedent when developing the first-line treatment plan for a newly diagnosed recurrent metastatic breast cancer patient. Strong clinical evidence to the contrary must be present to warrant basing the treatment plan on the metastatic tumor receptor status.

This statement, combined with the non-contradictory standards of ABC2 and NCCN, coupled with the excellent decision flowcharts of Turner and Di Leo (2013) and Penault-Llorca et al. (2013) provide clear direction for the best course of action based upon current available literature and supporting data. My proposed standard is challenged for efficacy in the data analysis phase of my research.

Statistical Literature Review

In preparation for the analysis phase of this dissertation, I performed a brief literature review on the Kaplan-Meier (KM) method of survival analysis and the log rank method of testing KM survival curves (Lakatos, 1988). Most papers on this subject are either theoretical (the mathematical theory, or new theories

about calculating sample size) or theoretical with some practical examples. The methods referencing sample size estimation ultimately rely on some form of a normal approximation of the sum of hypergeometric probabilities based on the number expected deaths. The number of deaths is manipulated by the censor and death percentages to obtain an estimate of sample size and power calculations (Lakatos, 1988, pp. 223-224). Several practical points are found in the articles reviewed. With respect to the number of time intervals utilized in an analysis, "there are no big differences in power if the subjects are observed in 2, 4, or 8 time intervals" (Jozwiak & Moerbeek, 2011, p. 637). Additionally on time intervals, "...increasing the sample size generally has a greater effect on power than increasing the number of time points. ...it is better to recruit a large number of participants in the trial and observe them infrequently, for example, two or three times during the study" (Jozwiak & Moerbeek, 2011, p. 651). The impact that the hazard rate has on power depends on how close the rate is to zero or one, not just the difference between the two proportions (p. 650).

CHAPTER 3

Materials and Methods

Research Purpose

The goal of my research is to influence change in international and national medical standards that determine the first-line treatment plans in discordant cases of metastatic breast cancer, and provide a higher level of evidence for those standards. The purpose of this dissertation was to determine the impact on *patient outcomes* in recurrent metastatic breast cancer diagnoses with discordant receptors when the *first-line* treatment plan is based on the receptor status of metastatic tumor(s) *instead of* the receptor status of the primary tumor(s).

For my particular study, the *first-line treatment plan* was defined as the first treatment strategy administered after the diagnosis of a metastatic recurrence, when the receptor status of the metastatic tumor is discordant with the primary tumor, or the first treatment plan *after establishing discordance* if the testing of the metastatic receptors is performed routinely and not ordered because the first treatment plan is failing. (In the metastatic setting, treatment plans will inevitably change over time as the efficacy of one strategy fails and another is employed).

For the purpose of this study, *patient outcomes* was defined in two ways: 1) five year post recurrence survival time (PRS) curves, which will be calculated as the time from metastatic recurrence to the time of death or censoring, and 2) the first-line follow-up scan results using the RECIST (Eisenhauer et al., 2009) standard, defined in Appendix 4. The first follow-up scan typically occurs 90 days

after the first-line treatment is initiated. *Scan* refers to some form of imaging to assess the growth of the metastatic tumors.

It is important to note that in the absence of re-testing the metastatic tumor receptor status, first-line treatment plans are based on the receptor status of the primary tumor, previous treatments and the clinical observation of the patient. When the metastatic tumor is retested, *sometimes* the receptor status of the metastatic tumor is utilized *instead of* the status of the primary tumor when determining treatment plans. In this study, the first-line treatment plan was considered to be *based upon the test results of the metastatic tumor(s)* if: a) there was a discordance between the primary tumor receptor status and the metastatic tumor status, and b) that discordance indicated a different treatment plan, and c) the plan was based on that discordance. All three of these conditions must have existed to consider the first-line treatment plan to have been based on the metastatic tumor instead of the primary tumor.

Key Research Questions

Two research questions are postulated for my study:

1. Does basing a patient's first-line treatment plan on the receptor status of metastatic tumors, when discordant with the receptor status of the primary tumor, improve patient outcomes?
2. Should the first-line treatment plan in metastatic cases be based on the receptor status of the primary tumor or the metastatic tumor?

These two questions guided the research methodology of this study and are critical questions not addressed in the literature.

Hypothesis

In discordant case, if the first-line treatment plan is based on the receptor status of the primary tumor, the median life expectancy of MBC patients will be longer than MBC patients whose first-line treatment plan is based on the receptor status of the metastatic tumor.

Study Design

A retrospective observational study was designed utilizing the data of the patients diagnosed with recurrent metastatic breast cancer in The University of Tennessee Cancer Center's tumor registry. A retrospective study was the appropriate design since a prospective study could take 10-15 years to complete and ensuring sample randomness would be difficult. That is, the medical staff's awareness that a study was being conducted and exposure to the hypothesis could theoretically impact the decision making process and bias the data. A clinical trial would be possible, but again would likely take more than 10 years to complete. Additionally, ethical concerns have been proposed by other authors. Amir, Clemons et al. (2012) state "a more definitive assessment of biopsy would require a randomized trial... ...for such a trial would likely it (be) unfeasible" (p. 713). Aurilio et al. (2013) state "in our opinion, only high powered prospective and randomized trials could clarify the outcomes in this field and finally

demonstrate a survival benefit; however, they should involve large sample size and ethical controversies may be unsolvable” (p. 1655). I disagree with the statement that only prospective randomized trials could clarify the outcomes. I contend that my retrospective study, and other similarly executed retrospective studies, can provide the necessary data to address the hypothesis and research questions posed in this paper.

Discordance

Discordance was a critical concept introduced and discussed in Chapter 2. The most common definition of discordance in the literature is defined as a difference in the receptor status of the primary and metastatic tumor tests. Of the eight combinations of receptor status listed in Table 5, Chapter 2, there are only four combinations (called phenotypes) that traditionally dictate treatment decisions. Those four phenotypes are displayed in Table 7.

Phenotypes and Discordance

This study focused on patients whose metastatic tumors were reassessed and determined discordant. Discordance was strictly defined as a difference in the primary tumor receptor status and the metastatic tumor status that warrants a change in treatment plan based on current NCCN standards (NCCN.org, 2014, pp. BINV-4). The four phenotypes utilized for this study are TNBC (Triple Negative Breast Cancer), HR+ (Hormone Receptor positive and HER2 negative), HER2+ (HER2/neu positive and HR negative), and PBC (Positive Breast Cancer). Radiation treatment was not considered in my study because it is agnostic to receptor status. Table 8 delineates the basic treatment strategy descriptions by phenotype. Dieci et al. (2012), also uses these four phenotypes in their research.

Table 7. Receptor Phenotypes

Receptor Combination	Description	Phenotype
ER-, PR-, HER2-	Triple negative	TNBC
ER+, PR+, HER2-	Hormone Receptor (HR) positive	HR+
ER+, PR-, HER2-	ER positive, PR negative	HR+
ER-, PR+, HER2-	ER negative, PR positive	HR+
ER-, PR-, HER2+	HER2 positive, HR negative	HER2+
ER+, PR+, HER2+	Positive Breast Cancer	PBC
ER-, PR+, HER2+	HR positive, HER2 positive	PBC
ER+, PR-, HER2+	HR positive, HER2 positive	PBC

Table 8. Phenotype Treatments

	ER	PR	HER2	Phenotype	Primary Treatment
1	-	-	-	TNBC	Chemo
2	+	+	+	PBC	Chemo+Targeted, Endocrine
3	+	-	+	PBC	Chemo+Targeted, Endocrine
4	-	+	+	PBC	Chemo+Targeted, Endocrine
5	+	+	-	HR+	Chemo, Endocrine
6	+	-	-	HR+	Chemo, Endocrine
7	-	+	-	HR+	Chemo, Endocrine
8	-	-	+	HER2+	Targeted

Liedtke et al. (2009), report an analysis of TNBC and RPBC, which the authors define as any positive receptor and de Duenas et al. (2014), utilizes three phenotypes (HR+, HER2+, TNBC). In this dissertation, discordance was strictly defined as a change from one of the four phenotypes to one of the remaining three; thus there were 12 (4x3) possible discordances.

Study Decision Flow

Figure 10 is a graphic depiction of the study protocol flow and cohorts. This study began with isolating patients in the Tumor Registry who had a recurrence of metastatic breast cancer. The next step was to determine whether or not the MBC patient had their receptor status re-tested (cohorts A and B in Figure 10),

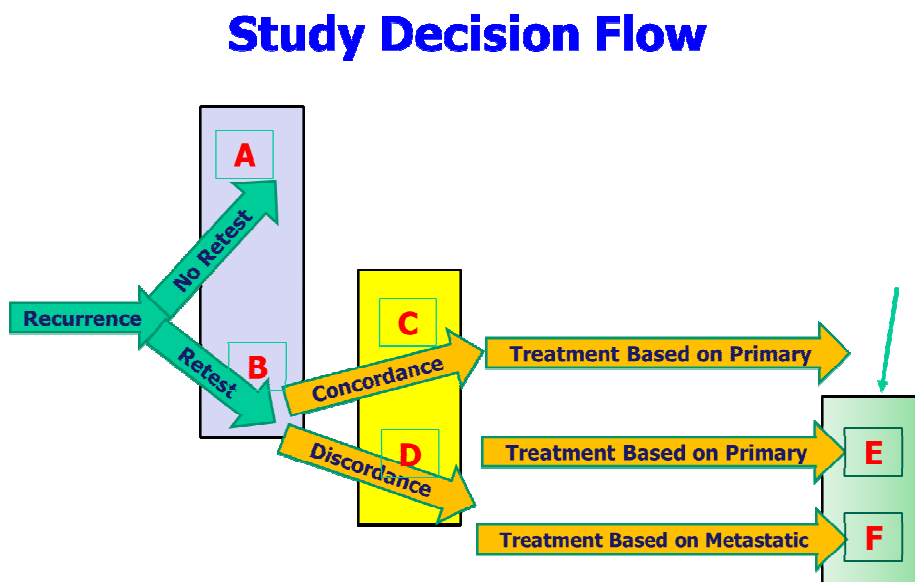


Figure 10. Study Decision Flow

and if available, the rationale for such re-testing. For retested patients, a determination was made to whether their primary and metastatic tumor receptor statuses were concordant or not (C and D). For patients who tested discordant, it was then determined whether their first-line medical treatment plan was based on the primary or the metastatic tumor receptor status. These two groups are represented as cohorts E and F in Figure 10.

The main objective of this study was to compare groups E and F (treatment plan based on primary versus metastatic tumor). My literature review revealed studies that compared groups A and B, and C and D as depicted in Figure 10, however no research explicitly carried out a protocol to compare survival rates of groups E and F. For example, one article on liver metastases compared the survival impact on patients with re-tested tumors to those patients whose tumors are not re-tested (A and B). The researchers found no difference in overall survival rate between these two groups (Botteri et al., 2012). My research project also evaluated groups A and B to determine any difference in survival rates, and to determine whether that difference indicated a potential bias in the data. My concern was that it might be possible that patients were re-tested because the doctor's clinical intuition indicated that a more aggressive treatment was needed. This type of bias was important to understand in a retrospective study because if patients in branch B were not representative of the population of all metastatic patients, then this bias could impact the interpretation of comparisons of groups C to D and E to F.

The articles in the Literature Review section that evaluate the clinical impact of discordance did so by comparing patients with concordant versus discordant tumors (groups C and D). In studies that evaluate the survival rates of patients in cohort C versus cohort D, cohort D patients are always, to some degree, confounded with cohort F. My analysis is unique since this confounding effect is

removed by specifically comparing the survival curves of groups E and F.

Data Parameters

Based on my protocol, I identified the list of variables necessary to conduct my research. The information needed for my variables was found in The University of Tennessee Cancer Center Tumor Registry, the electronic medical record and paper charts of the patients in the study.

Variables in the Tumor Registry

The data collection began with a request to the Tumor Registry manager to export variable information based upon my protocol criteria for patients with recurrent metastatic disease (local recurrence in the breast was not considered). I requested records on patients with recurrences after January 1st, 2000 and before November 1, 2014. This time period was chosen for my study because HER2 testing was not available until 2000. However, I suspected I would encounter missing data in the early records since HER2 testing was also not routinely performed until around 2005 (Harries & Smith, 2002). In order to ensure accuracy, I eventually asked the manager for access to *all* patient data in the tumor registry. I then personally managed the data to extract the variable information I sought. The manager informed me that some new information fields were added to the tumor registry database in 2010. Therefore, this additional information was missing in records dated prior to 2010, thus rendering the data problematic to utilize if it could not be found in other sources. Further discussion and definitions of the variables actually used in the analysis are provided in the Results section of this dissertation.

Variables in Electronic and Paper Charts

Some of the key variables necessary for my research were not routinely stored in the Tumor Registry. These particular variables were obtained from a review of patient charts (both electronic and paper). Additionally, I was able to verify and often complete missing variables in the Tumor Registry data through these chart records. I performed all the tumor registry, electronic records and paper chart reviews with assistance and quality control from the Tumor Registry manager. Further discussion and definitions of the variables actually used in the analysis are provided in the Results section of this dissertation.

Quality Control

Operational definitions of key variables were established and are found in Appendix 4. These definitions were vetted by the Tumor Registry manager, and aligned with industry standards. The registry manager and I reviewed the first ten patient records together to establish a common practice.

Record Keeping

Both paper and electronic records were created during the data collection process. I created a paper data collection form that listed the key variables sought from each patient record review. A copy of the form is found in Appendix 5. The form did not contain a direct patient identifier; only a code number, connecting the sheet to the patient, was employed. The code numbers and patient identifiers were maintained separately from the data collection sheets. Data collection forms will be destroyed once the electronic data set is complete, the quality control has been performed, and this dissertation is complete. Each electronic patient record's code number was assigned by random number to

ensure confidentiality. Separate Excel spreadsheets were created that contain the decoding information in one file and the raw data without identifiers in another file. The decoding file was password protected and kept separate from the raw data. JMP files for analysis will not contain patient identifiers.

Confidentiality and Compliance

Confidentiality is important in any medical research. I followed and completed the requisite procedures for IRB approval as required by both the University of Tennessee Medical Center and the University of Tennessee, Knoxville. Both IRB protocols required for my research were approved by expedited review.

Statistical Considerations

A priori statistical methods, hypotheses and sample size considerations are outlined in this section.

Sample Size and Power

It is important to determine the sample size in any research endeavor to ensure the objectives of the experiment are achieved and the hypothesis can be proven or not. In order to estimate sample size needs with respect to statistical power for survival analyses, several statistics needed to be known or estimated (power and other terms used in this section are defined in Appendix 12). Estimate of the survival curve characteristics under the null hypothesis, a hazard ratio (HR) for the alternative hypothesis, and the risk of a false positive were required to estimate the needed sample size. For my sample size considerations, I assumed

that a potential difference in median life expectancy greater than one year existed between those metastatic patients with discordant receptors whose first-line treatment plan was based on their primary tumor receptor status as opposed to the metastatic tumor receptor status. The hypothesized one year median difference was supported by the evidence in my Literature Review section (Chang et al., 2011; Liedtke et al., 2009; Niikura et al., 2012). This median difference translated to an approximate hazard ratio of 1.4. Therefore, at any point in time, a metastatic breast cancer patient was 40% more likely to die if their treatment plan was based on their metastatic tumor receptor status instead of their primary tumor results.

Major databases such as SEER do not currently collect data on recurrent metastatic breast cancer survival statistics. A few estimates of overall survival curves are found in the literature; however, I elected to combine and average data from five articles in the literature review to estimate the survival curve for recurrent metastatic breast cancer. The exact raw data required to calculate the survival curve was not always delineated in these papers. In some cases, I extrapolated data from graphical representations. It was important to recognize that this combined data produced a biased sample for the overall population because the samples in each study only contained cases where receptors were re-tested. However, this was a logical solution to establish survival curves for my research since I was only concerned with cases where tumors were re-tested. Table 9 summarizes the data extracted from the four articles and the (CTCA, 2014) website that was then utilized to establish an average survival curve. The CTCA data was the only available survival curve I located in my search that directly focused on metastatic breast cancer. The resultant average survival curve, depicted in Figure 11, had a median survival of 1.85 years and produced a five year survival rate (x-axis) of approximately 20% of the data population (y-axis).

Table 9. Literature Survival Data

CTCA	Tev	Daw 08	Chia	Daw 10	Year	Average
1	1	1	1	1	0	100.0%
0.78	0.8	0.78	0.79	0.81	0.5	79.2%
0.66	0.68	0.65	0.62	0.72	1	66.6%
0.57	0.57	0.54	0.5	0.62	1.5	56.0%
0.5	0.5	0.47	0.39	0.52	2	47.6%
0.42	0.39	0.42	0.32	0.47	2.5	40.4%
0.37	0.31	0.38	0.295	0.39	3	34.9%
0.32	0.28	0.32	0.22	0.32	3.5	29.2%
0.29	0.2	0.295	0.2	0.28	4	25.3%
0.25	0.19	0.25	0.18	0.26	4.5	22.6%
0.23	0.18	0.2	0.17	0.22	5	20.0%

Survival Curve for MBC

(Chia et al., 2007; CTCA, 2014; Dawood, Broglio, Ensor, Hortobagyi, & Giordano, 2010; Dawood et al., 2008; Tevaarwerk et al., 2013)

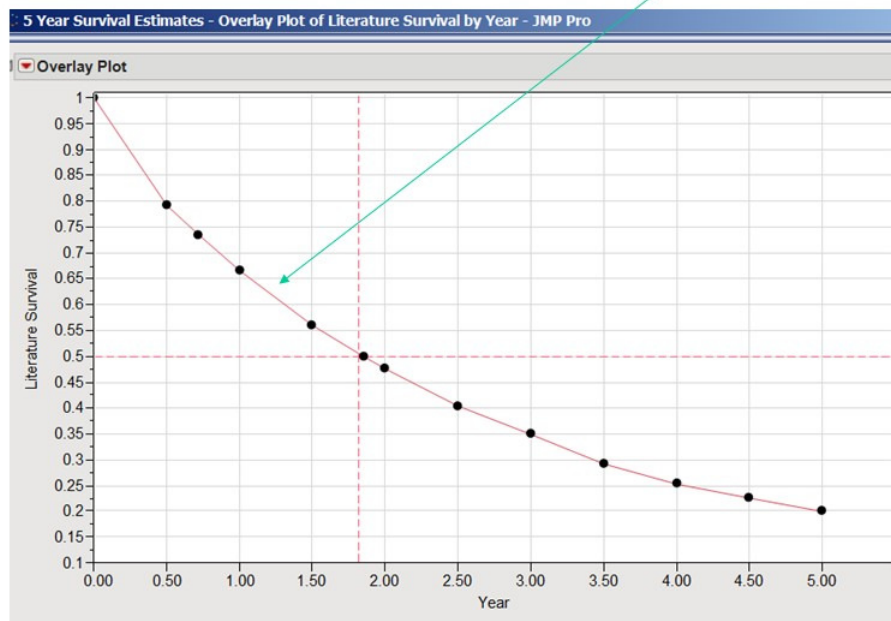


Figure 11. Null Survival Curve

The overall average survival curve postulated in Figure 11 was a combination of patients with concordant and discordant receptor statuses and different treatment plans based on either the primary or metastatic tumor receptor status. Because the overall average survival curve depicted patients with different receptor statuses and treatment plans, I then extrapolated two survival curves from the overall average survival curve. These curves represented the null and alternative hypotheses consistent with my estimate of one year median survival difference between patients whose treatment plans were based upon their primary or metastatic tumor receptor status, represented by cohorts E and F in Figure 10. Figure 12 illustrates the two hypothesized curves for groups E and F. I

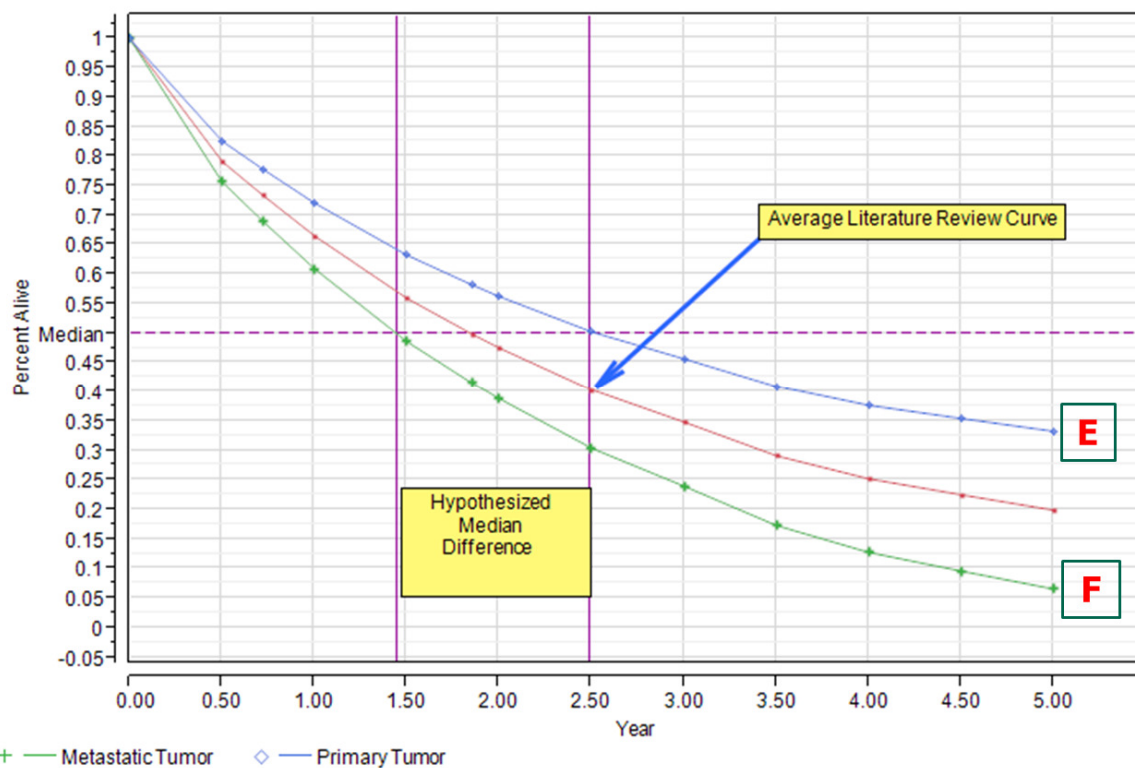


Figure 12. Null and Alternative Curves

chose .01 as the risk of a false positive in my analysis. With an estimated survival curve, a hypothesized HR and the risk of a false positive, the statistical power could be calculated for any given sample size; conversely, for a given power the required sample size could be calculated.

Both SAS and JMP software were utilized to determine sample size needs for my research. SAS software's two sample survival piecewise application allowed for calculating the optimal sample size estimate for my study. This methodology required the input of the average survival rates in half-year increments, utilizing data from Table 9. The resulting SAS power curve estimates yielded an optimal sample size of approximately 50 patients per cohort (E and F) of the study.

These curves are shown in Figure 13, Utilizing JMP, a sample size was estimated

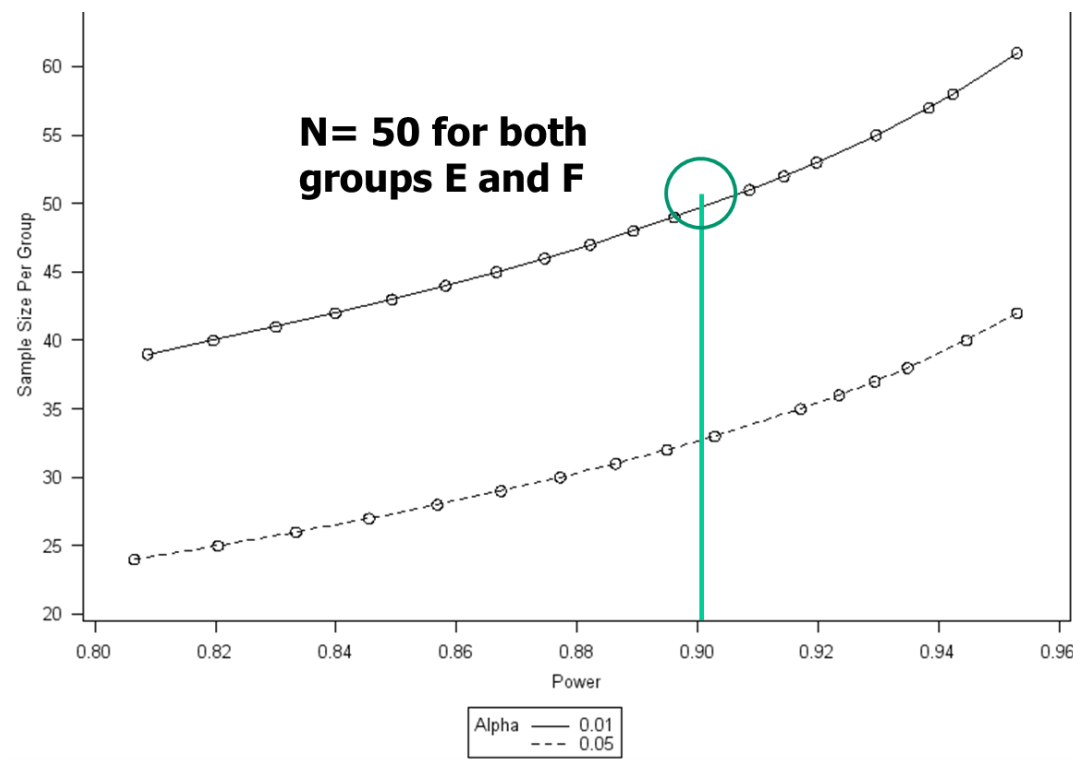


Figure 13. SAS Power Curves

using a two sample proportion test based on the percentage of patients surviving after five years under the null and alternative hypothesis and a chosen type I risk of .01 and power of .90. This calculation also yielded an optimal sample size of 50 for each cohort in the study (groups E and F), as shown in Figure 14.

The screenshot shows the 'Sample Size and Power' dialog box in JMP Pro. The 'Sample Size' tab is active, and the 'Two Proportions' section is selected. The dialog is configured for a two-sample proportion test where the null hypothesis is $H_0: P_1 - P_2 = \Delta_0$. The input values are: Alpha = 0.01, Proportion 1 = 0.325, Proportion 2 = 0.055, and the test is one-sided. The user has specified a Null Difference in Proportion of 0, Sample Size 1 of 50, Sample Size 2 of 50, and Power of 0.9. The calculated Actual Test Size is 0.0100217. The dialog includes 'Continue' and 'Back' buttons at the bottom.

Sample Size and Power - JMP Pro	
Sample Size	
Two Proportions	
Testing if two proportions are different from each other.	
Alpha	0.01
Proportion 1	0.325
Proportion 2	0.055
$H_0: P_1 - P_2 = \Delta_0$	
☐ Two-Sided ☒ One-Sided	
Supply two of (difference, sample sizes, power) to determine the third.	
When entering sample sizes, enter a value for both groups.	
Null Difference in Proportion	0
Sample Size 1	50
Sample Size 2	50
Power	0.9
Actual Test Size = 0.0100217	
Test size calculated holding P1 fixed and using $P_2 = P_1 - \Delta_0$	
Continue Back	

Figure 14. JMP Power Calculation

Sample Size Discussion

The above statistical exercise was important to determine the optimal sample size for this study. The fact that both calculations resulted in a consistent

answer of 50 patients per cohort was validating. At the time I wrote the research proposal to examine MBC patients, receptor statuses, and treatment plan decisions, the tumor registry manager estimated that my total sample size for the project would be between 100 and 200, yielding only 10 to 20 patients in each cohort of the study (E and F). These numbers were disappointing, but not a surprise. The available sample size for sub sample categories A-F in Figure 10 were certainly bound by the existing data in the Tumor Registry and the accessible paper and electronic charts at the time of data collection for this dissertation. The actual sample sizes are reported in Chapter 4.

A Priori Analysis Plans

During the proposal stage of this dissertation, my a priori analysis plans were as follows: *The first analysis will be a Chi-square test, analyzing the null hypothesis asking whether the two patient groups, E and F, have the same percentages of patients in five categories, 4 RECIST categories plus the deceased category I add to the study, (as defined in Chapter 1). Statistically significant results may lead to further refined tests. The five categories will then be combined into two categories: Complete Response (CR), Partial Response (PR) and Stable Disease (SD) will be combined into one category; Progressive Disease (PD) and Deceased (D) will be combined into a second category and analyzed with a Chi Square test. Next, utilizing the groups identified in the study decision flow, Kaplan-Meier survival curves will be calculated for all paired comparisons (A and B, C and D, E and F). The sample will also be examined using Cox Proportional regression in an effort to account for the effect of covariates, since this is not a randomized trial. All available covariates will be tested. The literature review indicates possible significant covariates will be age, time to recurrence, HR status, initial stage and initial lymph node involvement.*

Log Rank, Wald and other generally accepted tests for survival analysis will be executed. Univariate tests will be performed and reported on available covariates. Descriptive statistics will be reported on each group of patients (cohorts E and F). JMP Pro version 10.0 will be used for all analyses. Other software may be used for graphics and data management.

Statistical Significance

Surprisingly, a p-value (or critical value) of .05 was the most frequently utilized value of statistical significance in the literature. Given the randomization limitations of observational studies and the risk that non-significant results do not get published, a critical value of .01 would be a more prudent statistical measure in medical research. In other words, if 20 experiments are performed with a type 1 risk of .05, then one experiment is likely to be deemed statistically significant by chance. If the other 19 experiments are not published, a literature review would find one paper supporting a hypothesis and not reveal the 19 that do not support the hypothesis. One article recently reported a study showing that only 17% of clinical trials had been published one year after completion ("Spilling the Beans," 2015). The use of p-values has also come under scrutiny. Gigerenzer and Marewski (2014), offer several criticisms of p-values, including the cumulative risk of performing numerous tests and that spurious correlation can lead to statistically significant results, yet non-reproducible results. Based on the works just cited and holding myself to high standards, I elected a priori to pair evaluative statements with the varying degrees of possible statistical significance values my data might yield.

The following are my evaluative statements I developed based on possible p-values for key analyses of my research. Test results numerically consistent with my hypothesis that patients live longer when their first-line treatment plans are

based on the primary tumor receptor status instead of the metastatic tumor receptor status in discordant cases will be described as:

- *consistent with* the hypothesis for p-values greater than .05;
- *supporting* (providing corroborating evidence) the hypotheses for p-values between .01 and .05; and
- *demonstrating* (shown to be true by reasoning or adducing evidence) the hypotheses for p-values less than .01.

For sample data numerically inconsistent with my hypothesis showing that patients die sooner when their first-line treatment plans are based on the primary tumor receptor status instead of the metastatic tumor, my results will be described as:

- *not consistent* with the hypothesis for p-values greater than .05;
- *not supporting* the hypothesis for p-values between .01 and .05; and
- *refuting* (thus proving to be false or erroneous based on the data) the hypothesis for p-values less than .01.

Statistical Methods Review

The Kaplan-Meier (KM) method of survival analysis has been a standard in medical industry research literature (Richard Van Noorden & Brendan Maher, 2014). All articles that compare survival curves in my literature review utilize this method. In my review of the literature, statistical conclusions were being reached using small sample sizes; thus I decided to research the key features impacting the statistical power of the Kaplan-Meier method and perform Monte-

Carlo simulations to add to my practical knowledge of the KM technique. The research was performed and the simulations designed to answer several important questions:

- What is the power of KM tests with sample sizes smaller than 20?
- How is power affected by binning the data versus having continuous measurements?
- Does the number of bins (time intervals) affect power?
- How is power affected by censoring?
- How is power affected by unequal sample sizes when comparing two survival curves?

Simulation

While the statistical literature review provided some insight into the attributes of statistical power associated with the KM method of survival analysis, none of the articles reviewed provided a clear clarification of my questions about power. I therefore elected to utilize simulation to address my questions on the statistical power of the KM method. Key definitions for terminology used in the simulation are in Appendix 12. For simulation purposes, I translated the average survival curve previously discussed in this chapter to a statistical model utilizing an exponential distribution $[(1/\lambda) e^{-1/\lambda}]$ parameterized with $\lambda = 3.5$. This function provided a reasonable estimate of the hypothesized survival curve. In order to explore the power of the KM method using the log rank statistic, the following six attributes were manipulated: sample size ($n_1=n_2= 10, 20, 30, 50, 100, 200, 500, 1000$), sample size distribution ($p_1 = p_2 = 50\%$; $p_1 = 70\%, p_2 = 30\%$; $p_1=85\%, p_2=15\%$), hazard ratio ($\delta = 1.0, 1.25, 1.5, 1.75, 2.0, 3.0$), survival function (exponential, $\lambda = 2, 3.5, 5$), binning (continuous

recording, 10 bins, 5 bins, 2 bins) and censoring (none, 25%, 50%).

MATLAB R2015a was used to conduct the simulation. All programs were used with permission. Figure 15 provides a flowchart of the simulation process. For the power calculations, the simulation was executed 10,000 times, giving the maximum width 95% confidence interval for the estimates using the normal approximation to the binomial as $\pm .0049$ or .49%. Appendix 13 contains the MATLAB file names and descriptions used in the simulations. The code for each program is found in Appendices 14-21.

Simulation Results

The simulation results clearly indicated that sample sizes less than 20 per group yielded insufficient statistical power. The simulation results in Table 10 ($\alpha = .01$) and Table 11 ($\alpha = .05$) detail the power estimates with the following conditions: no censoring; $\lambda = 3.5$; equal sample sizes; and continuous recording of data. These tables show a hypothesized hazard ratio of two or more would be necessary to reach any reasonable level of power. From a practical standpoint in breast cancer research, an HR of 1.25 would be medically important. For instance, in the much cited Journal of American Medical Association (JAMA) article on hormone replacement therapy and breast cancer, the sample HR supporting hormone replacement therapy is a cause of breast cancer is 1.26 (Rossouw et al., 2002). In Table 11, for $n_1=n_2= 10$ and $HR = 1.25$, the assumed probability of a false positive ($\alpha = .05$) and the chance of a true positive (power) are both 5%. Thus $\frac{1}{2}$ of the positive results under these conditions would be false. To ensure the kind of power needed to produce a successful research outcome with $HR = 1.25$, sample sizes of greater than 500 sizes were needed. This demonstrated why important clinical trials involve large sample sizes, illustrated by the hormone replacement study that involved over 8,000 women in each cohort of the study (Rossouw et al., 2002). Observational

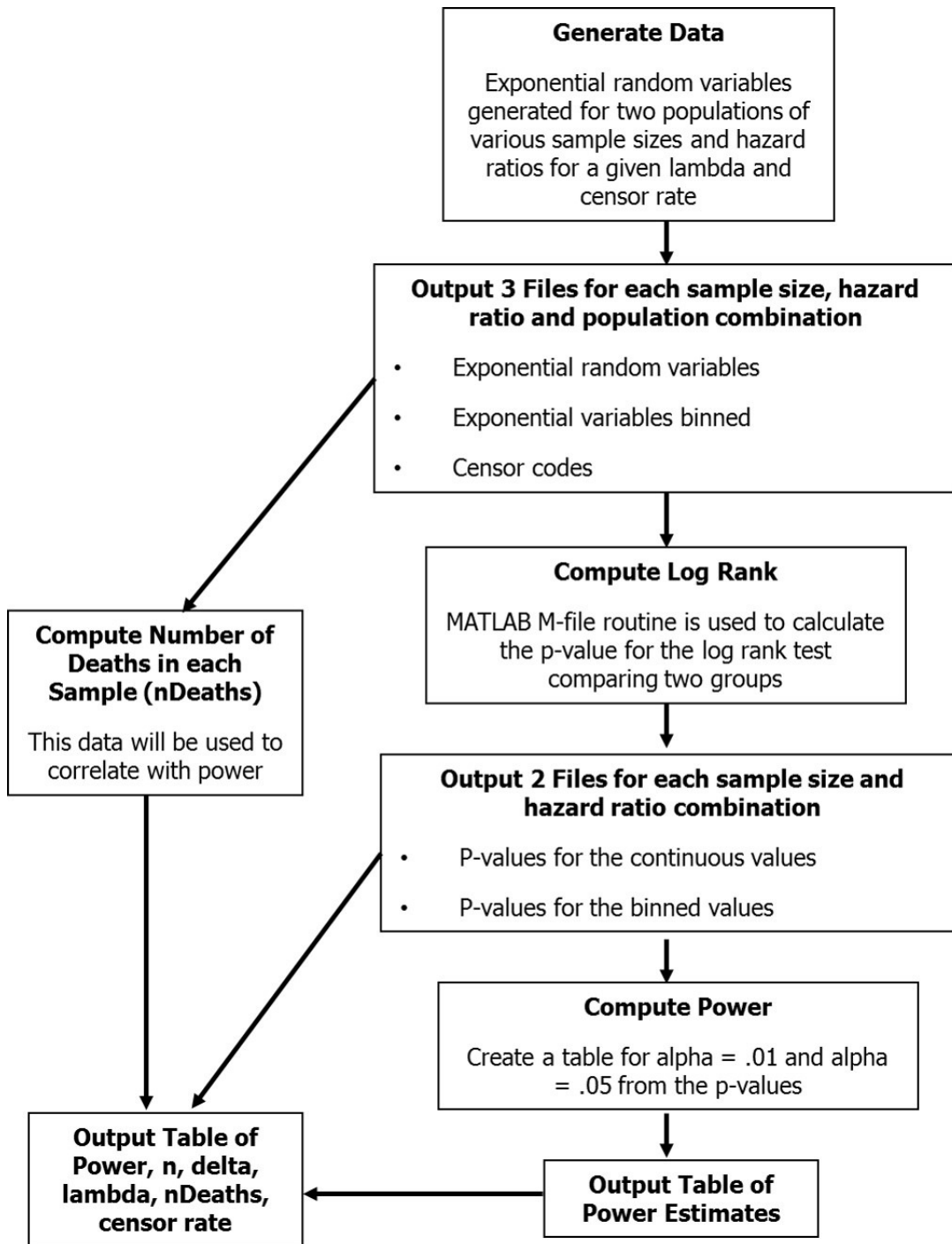


Figure 15. Simulation Flow

Table 10. Power Table for L3.5 C0 C .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0067	0.0116	0.0249	0.0467	0.0736	0.234
20	0.006	0.0179	0.0542	0.122	0.2256	0.6535
50	0.0088	0.0448	0.1994	0.4516	0.6989	0.9933
100	0.007	0.1021	0.4804	0.8282	0.9702	1
200	0.0086	0.2582	0.854	0.9929	1	1
500	0.0069	0.7015	0.9997	1	1	1
1000	0.0117	0.9693	1	1	1	1

Table 11. Power Table for L3.5 C0 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0342	0.047	0.0882	0.1408	0.2016	0.4819
20	0.0349	0.0744	0.1599	0.2992	0.4566	0.8585
50	0.0437	0.1439	0.4075	0.6893	0.8752	0.9994
100	0.0394	0.2584	0.7111	0.9447	0.9946	1
200	0.0434	0.4881	0.952	0.9989	1	1
500	0.0436	0.8761	0.9999	1	1	1
1000	0.0511	0.9931	1	1	1	1

studies tend to have smaller sample sizes and are generally based on the data available to the researcher, as Table 20 in Chapter 4 illustrates.

The simulation results indicated there were no practical differences in power when continuously recording each data point or binning the data. Tables 11 and 12 illustrate this result and show that any differences in the simulated power were either within the margin of error or of no real practical significance. As a scientist, I intuitively wanted to know the exact time of failure (death) for each data point, believing that information would have provided a more detailed analysis than just knowing the failure occurred within a certain time interval. However, since the ultimate measure was binary, it was important to realize that similar statistical results applied to both the percent of events that occurred over time versus at the end of a time period. Knowing exact time of death was

Table 12. Power Table for L3.5 C0 B10 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0315	0.0429	0.0808	0.131	0.1918	0.4611
20	0.0334	0.0697	0.1498	0.2899	0.4401	0.8472
50	0.0433	0.1415	0.4035	0.6843	0.8708	0.9994
100	0.0383	0.2559	0.7075	0.943	0.9938	1
200	0.0435	0.4829	0.9503	0.9988	1	1
500	0.0431	0.8735	0.9999	1	1	1
1000	0.0515	0.9931	1	1	1	1

therefore inconsequential to the power. A priori, I assumed that the loss of information resulting from utilizing less instead of more bins would have a tremendous impact on power. The literature and my simulation indicated that the number of bins was not that important to power.

The literature states larger samples observed less frequently (binned) are better than smaller samples observed more frequently. My simulation results supported the statistical literature review outlined in Chapter 2. While there were some differences in power based on the number of bins, those differences were negligible when compared to the impact sample size had on power. To illustrate this point, in Table 12 the power for $n_1 = n_2 = 20$ and a HR of 3 is 85% for ten bins and for the same set of variables with only two bins is 69% (Table 13). These differences warranted scrutiny, but were not as large as I expected. Once

Table 13. Power Table for L3.5 C0 B2 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.018	0.028	0.0489	0.083	0.1141	0.2697
20	0.0253	0.0513	0.1182	0.2218	0.3333	0.689
50	0.035	0.1145	0.3502	0.6027	0.8041	0.9947
100	0.0367	0.2317	0.6534	0.9081	0.9842	1
200	0.0401	0.4556	0.9289	0.9977	1	1
500	0.0436	0.8499	0.9999	1	1	1
1000	0.0469	0.9909	1	1	1	1

the sample size reached 50, the power differences between 10 and two bins was minimized. The primary concern with binning, (especially with less than 5 bins), is that the statistical significance between two populations can still be tested, but the visual evaluation of proportional hazards is hindered as the number of bins is reduced. The extreme case of utilizing only one bin in which results were observed once at the end of a time period provides excellent statistical knowledge when comparing two groups at that point in time, but would garner no information about the comparative shapes of the survival curves over time.

I had no experience with analyzing data that might be censored, so when I investigated survival techniques, censoring was a concern, especially its affect upon power. Descriptions of the KM method and the log rank test indicate that these methods were created for censored data, so I utilized the simulations to better understand the impact censoring had on data analysis results. Tables 11 and 14 best illustrated the effect of censoring. Table 14 was the result of 50% censoring; thus these numbers could be compared to Table 11 that had no censoring. Stated differently, sample sizes of 50 with no censoring could be compared to sample sizes of 100 with 50% censoring. These two results were nearly identical with respect to power. This outcome was consistent with the literature which indicates the real issue at stake is the number of events (deaths) not the pure sample size.

Common statistical calculations and concepts indicate that sample sizes of equal value are optimal when comparing two groups (Snedecor & Cochran, 1967, p. 104). In other words, if the total sample size is 20 for two groups, then it is statistically more effective to have an equal division between those two groups (i.e. 10 in each group), than any other combination. If the two groups' samples were split 16 and 4 respectively, the lack of data in the second group ($n_2=4$) would negatively impact the ability to compare that group with the other, even

Table 14. Power Table for L3.5 C50 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0277	0.0361	0.0512	0.0738	0.0966	0.2133
20	0.029	0.0482	0.0872	0.1497	0.2081	0.4711
50	0.0348	0.0856	0.209	0.3738	0.539	0.9059
100	0.0396	0.1429	0.4006	0.6915	0.8648	0.9981
200	0.0448	0.269	0.7135	0.9439	0.9926	1
500	0.0444	0.5867	0.9801	1	1	1
1000	0.0438	0.8722	0.9999	1	1	1

though the other group had more data ($n_1=16$). The simulation results strongly supported the concept of equally distributed sample sizes for comparative purposes. Tables 11 and 15 best illustrate this concept, where n_1 is the sample size of the null distribution. Comparing $n_1 + n_2 = 100$, where one sample is split evenly ($n_1+n_2=50$) and the other sample is split disproportionately ($n_1=85$ and $n_2= 15$), dramatic differences in power occurred. Power was reduced over 50% when employing unequal sample sizes in the simulation even though the total sample sizes were the same.

Table 15. Power Table for L3.5 C0 C N85/15 .05

	Hazard Ratio					
n1,n2	1	1.25	1.5	1.75	2	3
3, 17	0.035	0.031	0.0358	0.0486	0.0641	0.1487
6, 34	0.0377	0.0387	0.0628	0.1097	0.1666	0.4378
15, 85	0.0374	0.0622	0.1773	0.33	0.5009	0.9019
30, 170	0.0439	0.1194	0.3555	0.6332	0.825	0.9972
60, 340	0.0392	0.231	0.6571	0.9212	0.9858	1
150, 850	0.0423	0.5445	0.9714	0.9998	1	1
300, 1700	0.0479	0.843	0.9999	1	1	1

Type 1 Error

An interesting pattern appeared in all the data and is illustrated in Table 14. A hazard ratio of one meant that the two groups did not differ in hazard rate.

Thus, the type 1 error rate in the simulation was estimated where $HR = 1$. For all simulation runs, the actual type 1 error rate was lower than the declared rate for sample sizes under 1,000. The simulation's actual rate came close to the declared rate as $n_1 + n_2$ approaches 2,000. For example, in Table 14, with $n_1 = n_2 = 10$, the simulated alpha is .028 versus the declared rate of .05. That is, there were only 2.8% of the 10,000 p-values in the simulation exercise less than .05 where the $HR = 1$. For $n_1 = n_2 = 1,000$, the simulated rate is .044, approaching the declared rate of .05. These results, as simulated, would indicate the log rank test is conservative with respect to the type 1 error.

In summary, when comparing the data of two groups of subjects utilizing the KM and log rank methods, the key issue is the number of events (deaths) that will occur over the time of study. For estimating the sample size needed ahead of the research project's implementation, to calculate power or desired sample size, an estimate of the rate of events (deaths) occurring and the potential censoring rate are needed. Of less importance is the number of observation intervals, so that given the choice, it is better to have more subjects observed less often, than to have fewer subjects observed more often. However, for a clear picture of survival curve shapes, five to 10 bins are recommended. The most important factor to consider and attempt to control is obtaining even splits of the total sample size. This split is easier to control in a prospective study. In a retrospective study, the researcher may not be able to control the split of data at all, as the sample is based on the data available. All power tables for the various simulation combinations are shown in Appendix 11.

CHAPTER 4

RESULTS AND DISCUSSION

Data Acquisition

After receiving IRB approval, The University of Tennessee Cancer Institute provided access to the tumor registry data, electronic records and where possible, paper patient charts. The tumor registry was queried for recurrent metastatic breast cancer patients. Several initial queries were performed by the Tumor Registry manager that yielded varying sample size results each time. Based on these queries, I requested a wider query that allowed me to pare down the data manually where needed, yet was more likely contain all possible records for review. Appendix 6 contains a flowchart view of the data collection process.

Data Cleaning

All data analysis processes must begin with the cleaning of the data. Table 16 shows a listing of file names that represent the progression of the cleaning of the data. Each step in the data cleaning process is described below.

Table 16. Key Data File Names

Step	Excel File Name	Number of Rows
1	1 Copy of Breast Recurrs from 1990 As Received	3183
2	1.1 Copy of Breast Recurrs from 1990 SORTED	3182
3	1.3 Copy of Breast Recurrs from 1990 SORTED	605
4	1.4 Copy of Breast Recurrs from 1990 SORTED	587
5	1.521 Breast Recurs START	586

Step 1. A copy of the file received from the Cancer Center was made. This file had 79 columns (variables) of data.

Step 2. Two columns were added to the original file and one blank row deleted. A column to number (order) the rows as a safety precaution was added, in case the data was ever erroneously sorted in a way that could cause a loss of structure. An additional column entitled *Group* was inserted to combine records that appeared to cover several rows.

Step 3. Rows that were blank or duplicated based on the query output were removed. Records in which a diagnosis was initially Stage 4, thus not recurrent, were removed. A few records were deleted that had no recurrence data, incomplete data and contained phrases like “disease free” and “never” in the recurrence description. All records with the recurrence date prior to the year 2000 were deleted. Another column was created to store additional recurrence date(s) for patients with more than one recurrence. Patient records that appeared in more than one row were combined into one row.

Step 4. Additional duplicate patient records were discovered and removed. The 586 remaining records consisted of the following combination of records:

a. Recurrence dates

- i. 114 with 11/11/11 as the recurrence date (this was deemed a filler date for incomplete records)
- ii. 212 with missing recur date
- iii. 260 with a recur date

b. Recurrence location

- i. 151 distant

- ii. 41 distant with a long description
- iii. 115 unknown
- iv. 40 regional
- v. 30 local (not intended to be included in this study)
- vi. 210 never disease free (possibly originally stage 4)

Step 5. The medical record number for each patient was stripped from the original file and placed into a separate password protected file. New columns were added to match the variables that would be added based on chart reviews (see Appendices 4, 5 and 7). A column was added to indicate when the data collection was complete. Color coding was applied to indicate whether the data came from the registry, electronic records or paper records. With the data cleaning completed for the Tumor Registry data, the collection process from electronic records and paper patient charts began.

Limited Data Set

A small subset of electronic records were chosen to practice the data collection methods, learn the electronic medical record system and determine whether paper charts were necessary. This limited data set exercise revealed several challenges. First, around 30% of the records required paper charts to complete the record. Second, some data in the registry was found to be erroneous, such as a record showing that a recurrence had occurred, when in actuality there never was a recurrence at all. Last and importantly, I learned that some of my preconceived operational definitions would not handle all situations. In some cases, there was variation between lab reports and doctors' notes (to be discussed later) and a judgement had to be made on how to code the data and whether to use it in the analysis. Therefore, I decided to review those cases with

the Tumor Registry manager and adhered to her interpretation which she based on industry standards.

Finalizing the Data

The data collection process was slow and arduous. I double-checked records with the Tumor Registry manager to ensure quality and found that access to paper charts (when needed) was somewhat limited, since many reviews dated back to the year 2000 and most patients in this study were now deceased. The physicians' offices most closely associated with the Cancer Center and with the longest tenure were able to provide many of the charts I needed. Charts from private practice physicians or with patients pre-dating their association with UT were harder to obtain. While I did not get access to all the data in the patient population set, I was confident that the data I did obtain was a representative sample of that population, even though my sample was not a statistically designed random sample. The data collection progression is described in Table 17. The key activities for each step data collection are documented below.

Table 17. Data Collection Files

Step	Excel File Name	Number of Complete Records	Records to Exclude	Records with Data not Useable	Records that Remain to Review
1	1.521 Breast Recurs START	32	0	11	543
2	1.55 Breast Recurs START	66	98	74	305
3	1.60 Breast Recurs START	13	171	74	37
4	1.742 Breast Recurs START	No Change (NC)	NC	NC	NC
5	1.82 Breast Recurs FINAL	3	0	34	0
	Totals (586 records)	124	269	193	N/A

Step 1. The initial collection of data in the limited data set yielded 32 completed records and 11 unusable records. Unusable records fell into three categories: a) insufficient data, when the electronic record was woefully incomplete and the prospect of obtaining the paper chart was not deemed plausible; b) no evidence of recurrence, where a thorough review of the electronic records did not confirm a metastatic recurrence; and c) not applicable, in cases where cancers other than breast cancer presented, stage 4 was diagnosed initially, or men were included the data set.

Step 2. Several groupings of records were identified in the data that had similar characteristics. Records in these groups were eliminated, where as a group, there was a very low probability of obtaining complete and useful data. My decision to eliminate groups of records was based on a sample review of the group records. Seventy-four records were eliminated based on the analysis of these groups. Additionally, 98 total records were excluded where: a) 33 had a filler recurrence date of 11/11/11 and the last contact was before the year 2000; b) 42 had no recurrence data and last contact before the year 2000; and c) 23 had a filler recurrence date of 11/11/11, the patient was deceased, and the record did not fit the protocol of this study.

Step 3. One hundred and seventy-one records were excluded because a sampling of those records showed either no evidence of recurrence or insufficient data. The excluded records contained: a) 50 records with a recurrence date of 11/11/11 and the patient was listed as alive; b) 34 records that had a no recurrence date and the patient was listed as alive; and c) 87 records containing no recurrence date and no other data indicating a recurrence. This extensive effort produced only 13 useable records. The 37 remaining records required that I request paper copies of the patients' records.

Step 4. Columns with extensive missing values and un-needed columns were removed. New variables that were a combination of columns, like subtracting two dates, were created. Data in the spreadsheet was compared with the data collection sheets for an additional quality control check.

Step 5. Completion of the data record was possible for only three of the 37 remaining records. The other 34 were unusable because of insufficient data.

Data Collection Summary

The final dataset was named *1.9 Breast Recurs FINAL*. The breakdown of the 124 records in the completed dataset is illustrated in Figure 16 for each cohort of the study. As noted in Figure 16, 14 of 92 (15%) patients had discordant

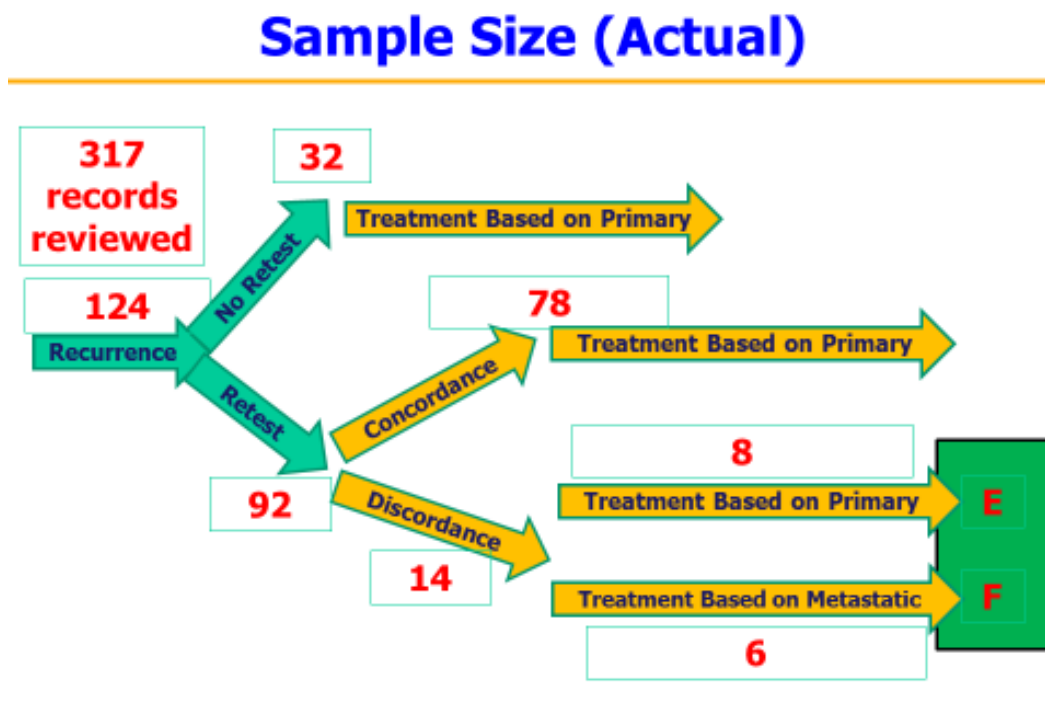


Figure 16. Actual Sample Sizes

receptor statuses. This sample result was slightly lower than the 20% I expected, but not out of line given that my definition of discordant was more restrictive than all but two of the relevant articles cited in the literature review. Of the 14 patients whose results were discordant, eight patients had their first-line treatment plan based on their primary tumor receptor status and six patients' treatment was based on the metastatic tumor receptor status. The sample result of 6 of 14 (43%) patients with a first-line treatment plan based on the metastatic tumor receptor status was lower than the 60% to 75% I estimated using data found in the current literature as discussed in Chapter 2. One reason for this discrepancy could have been the nature of variation in sample percentages estimated with small sample sizes. Additionally, my 60% to 75% estimate was based on current practices, while my data represented records evaluated over a 15 year period.

List of Variables in the Final Dataset

The list of variables utilized in the final dataset, the available source(s) of the data for each variable, and the purpose for each variable is shown in Appendix 7. Appendix 8 contains a list of variables maintained in the dataset for information purposes only, and not utilized in any analysis.

Preparing the Data for Analysis

Once the final Excel spreadsheet was complete, the data was moved to a JMP file for final processing and analysis. Histograms of all key variables were made as a visual quality check that ensured correct data entry and assessed the data for outliers. No outlier was revealed. Table 18 compares the number of valid variable responses against the total sample size available (n of N). Four records in the discordant cohort presented coding challenges when the records were investigated. That is, there were difficulties applying the a priori operational definitions to the actual data, such that alternative evaluations of the

Table 18. Sample Size Completeness

Variable Name	(n of N)
PRS	124 of 124
Status	124 of 124
FSD	124 of 124
Rdate	124 of 124
Retested	124 of 124
mPheno	14 of 14
pPheno	121 of 124
FLTP	14 of 14
FLD	13 of 14
LocoRegion	117 of 124
Visceral	116 of 124
rLocation	84 of 124
RECIST	14 of 14
Age	124 of 124
TTR	124 of 124
pER	119 of 124
pPR	119 of 124
pHER2	107 of 124
SSS	78 of 124
Hormone	114 of 123
Chemo	118 of 123
Rad	121 of 124
Herceptin	102 of 124
Grade	110 of 124
Size	120 of 124
nPos	121 of 124
AJCC	123 of 124
Type Chemo	113 of 124

values of key variables were possible. Those records and the challenges are outlined in Table 19. Those four records were first analyzed as they were coded with my best effort to apply the a priori definitions. In order to ensure an unbiased evaluation, a sensitivity analysis was performed based on these four data points. This sensitivity analysis was accomplished by changing the coding interpretation of those four records such that those records were *not supportive* of my hypothesis and then conducting the data analysis. Next, those four records were changed such that the coding interpretation was *supportive* of my hypothesis and reanalyzed. Those results are reported later in this section.

Table 19. Sensitivity Analysis Records

ID Code	Best Judgement Coding	<i>Not Support</i> Coding	<i>Support</i> Coding
103	Concordant	Discordant, Metastatic	Discordant, Primary
188	Discordant, Metastatic	As is	Concordant
312	Discordant, Primary	Concordant	As is
558	Discordant, Metastatic	Discordant, Primary	As Is

For the patient with ID code 103, the primary tumor was ER and PR negative. The metastatic ER and PR test results were 1% and 5% respectively. Historically, these values were interpreted as negative, but standards have changed over time such that these results would currently be deemed positive. The physician noted the results were *weak* in the patient chart. He prescribed chemotherapy as the first-line treatment plan and then switched to endocrine therapy, even though the chemo was working. In my judgement, this change in therapy was

planned ahead of time and not based on a patient assessment; however I was unable to verify my assumption. The record would have been coded concordant based on the lab report. The record would have been coded discordant based on the physician's interpretation. If I assume the physician's original plan was to switch to endocrine therapy, then the treatment plan was based on the primary tumor. If I assumed the switch occurred because of a clinical judgement after assessing the patient's progress, then the original plan would have been coded as based on the metastatic tumor receptor status.

The primary tumor in patient ID code 180 was reported equivocal at 2+ on HER2. General practice indicated an equivocal result as HER2-; however the patient received Herceptin per the doctor's orders, which indicated a HER+ judgement ("Breast Cancer," 2015, pp. 36-37). The metastatic tumor was TNBC, and Herceptin was not prescribed for treatment.

Her2 in the primary tumor for patient ID code 312 was measured as negative. Her2 in the metastatic tumor was measured as 2+, which is judged equivocal by current practices and requires an alternate test to confirm the results. The lab report called the metastatic result positive and noted that the tumor sample was not large enough to perform the alternate test.

Patient ID code 558 was ER/PR positive in her primary tumor. The recurrent tumor was TNBC. I was not able to review this chart, but the Tumor Registry manager did and indicated that endocrine therapy was not utilized upon recurrence. The lack of endocrine therapy indicated the treatment was based on the metastatic tumor. Given that I did not review the chart myself, and that the possibility existed that the patient was already on endocrine therapy at recurrence, I included this patient in the sensitivity analysis. Key variables for all the discordant patients are in Appendix 9, Table 26.

Comment on Sample Size

The final sample size of 14 discordant cases was disappointing, and on the lower end of the range I believed I would find a priori. Previous discussions in Chapter 3 on statistical power addressed the desired sample sizes for this study. I compare the sample size of 14 to sample sizes in the literature review in Table 20. These values are sample sizes for five key papers referenced in this dissertation and include two of the most cited articles in my review. In these papers, the primary statistical comparison the authors make is comparing concordant versus discordant patients (last column in Table 20). However, my hypothesis required a further breakdown of the discordant data into two parts (based on the first-line treatment plan), thus further reducing sample size of the two groups that were statistically compared. While my sample was small (n=14 discordant cases), it was consistent with sample sizes found in the literature.

Table 20. Sample Size Examples

Author	Number of Citations	Total Sample Size of Interest	Sample Sizes of Relevant Group Comparisons	Discordant Sample Size
Niikura(Niikura et al., 2012)	106 ^a	182	139,43	43
Liedtke(Liedtke et al., 2009)	127	231	134,42,22,33	22
Simmons(Simmons et al., 2009)	126	29	11,3	14
Wilking(Wilking et al., 2011)	49	151	136,15	15
Chang(Chang et al., 2011)	15	56	42,14	14

^anoted as a "Highly Cited Paper" on Web of Science.

Study Results

Fourteen records were found to be discordant, with eight patients treated in the metastatic setting based on their primary tumor status and six patients treated based on their metastatic tumor status. The survival data for these 14 patients was *consistent with* my hypothesis that patients with treatment plans based on the primary tumor receptor status have a longer median life expectancy.

Kaplan-Meier Survival Analysis

The Kaplan-Meier method was used as the primary tool to understand and compare survival curves. This method is a common technique in survival analysis and employed extensively in the articles reviewed for this dissertation (Richard Van Noorden & Brendan Maher, 2014). All p-values quoted for survival analyses were from the log rank test utilizing JMP 10.0 Pro.

Overall survival. Survival curves for metastatic breast cancer patients are difficult to find in the literature due to the fact that patients diagnosed with metastatic breast cancer are not tracked by the national databases. Although estimating a survival curve for patients with metastatic breast cancer was not one of the objectives of this analysis, a survival curve determination was important for context and for the cancer community as a whole, given the limited available data. The overall survival curve provided in Figure 17 represents all the complete records in the dataset (n=124). It shows a median survival time of 18 months and a five year survival rate of just over 20%. These numbers were consistent with the limited data available (CTCA, 2014). Additionally, I analyzed the difference between those patients who had their receptors re-tested and those who did not. My a priori assumption was that patients diagnosed to be in a “worse” condition might get their receptors re-tested in an effort to find a better treatment plan. Thus, those re-tested would have a shorter survival than those

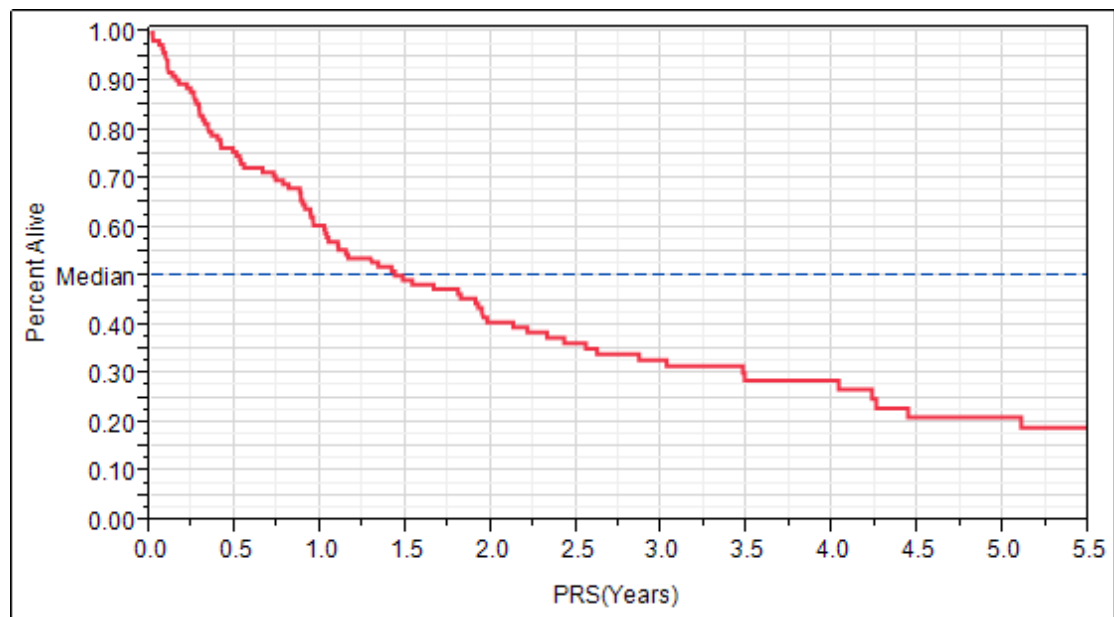


Figure 17. Overall Survival Curve

not re-tested. This assumption proved to be erroneous and the opposite result occurred as shown in Figure 18. While the literature review led me to believe that around 50% to 70% of patients may be re-tested based on current standards and less than 50% in the early 2000's, at the University of Tennessee Cancer Center it was a common practice to re-test throughout those years. In most cases where the patient was not re-tested, the patient presented in such a dire situation that there either was no time to re-test or it was deemed too late for any treatment to succeed. Thus, those not re-tested (median survival of about 5 months) showed a much worse outcome than those who were re-tested (median survival of nearly 24 months). The two survival curves re-tested (Yes) versus not re-tested (No) were evaluated for equality using the log rank method. The result of the test indicated the two curves did not have the same survival rate; therefore I rejected the null hypothesis of equality, with a p-value of .0002.

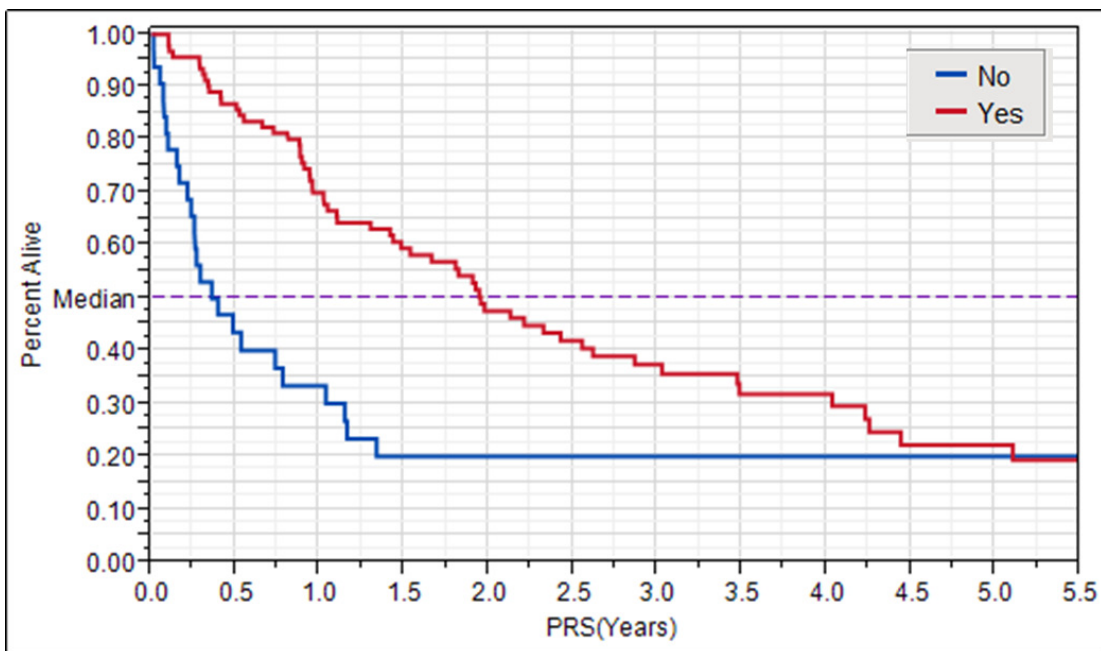


Figure 18. Re-test versus Not Re-tested Survival Curve

In order to judge discordance for a metastatic patient, the patient must be re-tested. Therefore, all remaining analyses were performed on patients who were re-tested, with the exception of one multivariate analysis performed on the complete data set.

Discordance survival. As discussed in the literature review, the survival outcome of patients with discordant receptor results tended to be less favorable than those with concordant results. However, in my data set, the results indicate the two groups had similar survival curves. That is, the null hypothesis that the two curves were the same was not rejected (p -value = .986). These results are shown in Figure 19. Sample size limitations (only 14 discordant cases) and the fact that only 6 of 14 (43%) patients had their treatment plan based on the metastatic tumor receptor status reasonably explained the results of similar survival curves. In Chapter 2, I estimated that 60% to 75% of patients had their treatment plans determined based on the metastatic tumor receptor status; my sample results proved to be lower than this, which yielded more patients in the

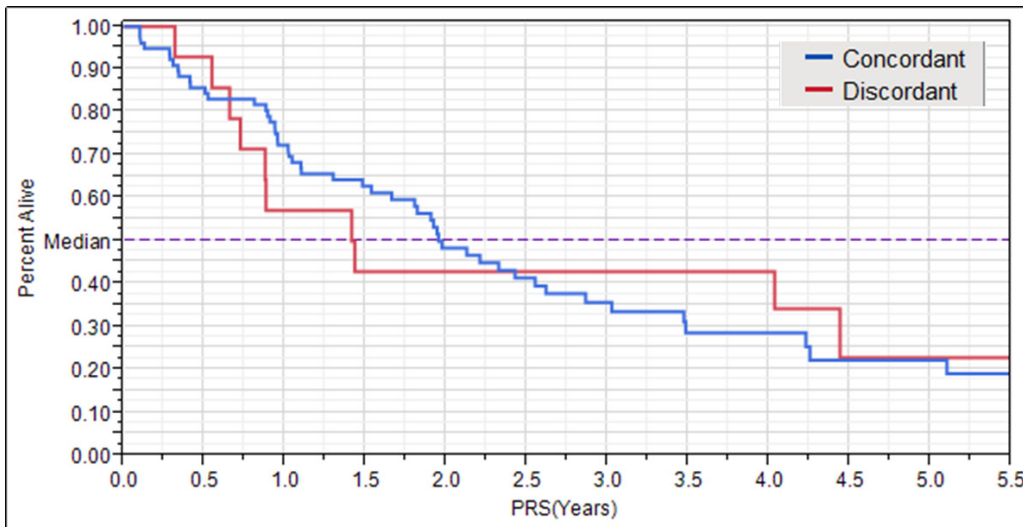


Figure 19. Concordant versus Discordant Survival Curves

discordant sample that performed well based on their treatment plans being determined by their primary tumor receptor status.

Treatment strategy survival. The primary purpose of this study was to compare the clinical outcomes of patients with discordant results on the basis of whether their primary or metastatic tumor receptor status was utilized to develop their first-line treatment plan. Figure 20 shows the results of this comparison. The graph clearly indicates that those patients whose first-line treatment plan was based on the receptor status of the primary tumor (n=8) had better survival outcomes than those who had their plans based on the metastatic tumor (n=6)

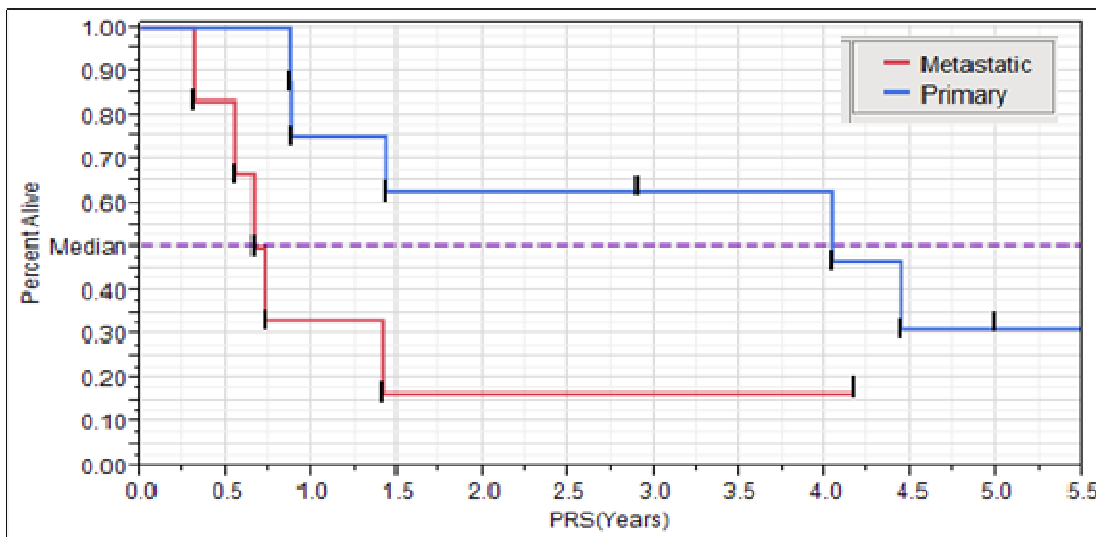


Figure 20. Metastatic Versus Primary

with a p-value of .049. The statistical results showed that metastatic breast cancer (MBC) patients who were treated based on the receptor status of their primary tumor had a median survival time of 48 months versus only eight months for those patients whose treatment plan was based on the metastatic

receptor status. This sample result of a 40 month difference in life expectancy was consistent with the important article by Liedtke et al. (2009) in the literature review and exceeded my qualitative a priori estimate of 12 months. Given the small sample size, other limitations discussed in Chapter 5 and the general poor prognosis of MBC patients, I would hypothesize the effect of utilizing the primary tumor receptor status to determine first-line treatment plans will likely be closer to 12 months than 40 months when larger samples sizes are studied. Regardless of the exact number, this finding provides a significant increase in life expectancy for metastatic breast cancer patients. Given the small sample size, and other limitations listed in Chapter 5, I described my results as *consistent* with my hypothesis rather than utilize the a priori decision that stated the results *support* my hypothesis when a p-value less than .05 was found.

Chi-Square Analysis (RECIST Standard)

Figure 21 shows the breakdown of discordant patients based on the RECIST standard and whether their treatment plan was based on their primary or

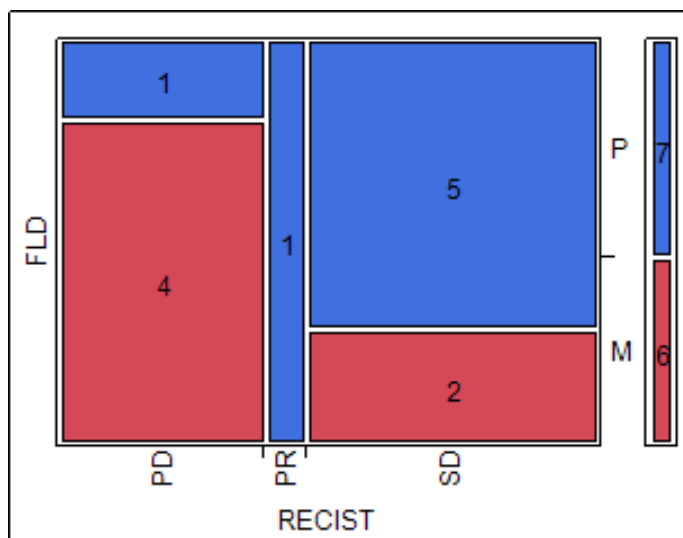


Figure 21. RECIST Chi Square Analysis

metastatic tumor receptor status. In chapter 3, I defined the RECIST standard and the intention to utilize that standard to code the results of the first follow-up examination performed by the patients' doctor, which occurs approximately 90 days after treatment begins for the metastatic disease. This metric was developed as a means to measure the immediate effect of the first-line treatment plan, since numerous plans might be implemented before the patient died, as all treatments lose efficacy over time. Based on the small sample size, I expected the Chi-Square analysis on the RECIST standard would not be statistically significant, and it was not with a p-value = .164 using Fisher's exact test for contingency tables. However, the results were consistent with my hypothesis, given that a) of the five patients who had progressive disease (PD) at the first re-scan, four were treated based on their metastatic tumor receptors; b) five of the seven patients with stable disease (SD) were treated based on their primary tumor; and c) the only patient with a partial response (PR) was treated based on their primary tumor receptor status.

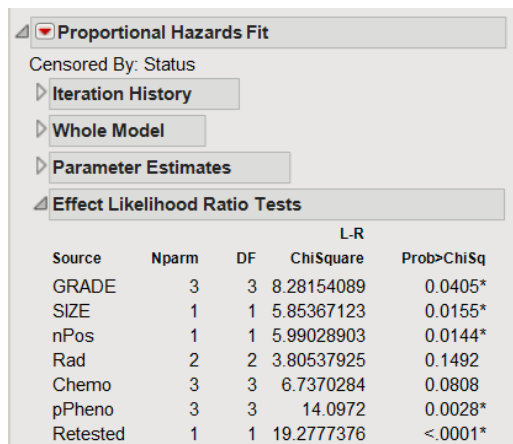
Multivariate Analysis

My statistical results were consistent with my hypothesis that metastatic patients with discordant receptor results whose first-line treatment plans utilize the primary tumor receptor status instead of the metastatic status had increased median life expectancy. To further substantiate these results, it was necessary to detect any covariates that might impact or influence the observed sample life expectancy. To examine the impact of covariates, the Cox Proportional Hazards method was used to analyze the data. Multivariate statistical analysis requires relatively larger sample sizes than univariate analyses. The number of variables that can be analyzed at once is strictly limited by the number of observations and

further impacted by nominal variables with multiple levels. Due to the small sample size of patients with discordant results in my study, effective multivariate analysis was not feasible on that subset of my data. Only a few variables could be tested at one time and the results were unlikely to be statistically significant. I employed various analyses with a small number of variables and verified the small sample size did not lead to statistically significant results. To determine whether potential covariates could impact the survival of patients and potentially bias the univariate results, I utilized the complete dataset and the re-tested patients to investigate the covariates. These findings are discussed below and in the Limitation section in Chapter 5.

Complete Dataset Results

The covariates available for analysis (referenced in Chapter 3) were utilized in a Cox regression analysis. Due to the relatively small sample size and the high number of nominal variables, I did not attempt to test any interactions between available covariates, only main effects were tested on the variables. The results of the initial analysis are shown in Appendix 10, Figure 33. The final model utilized is shown in Figure 22. This model was based on several iterations of the



Source	Nparm	DF	ChiSquare	Prob>ChiSq
GRADE	3	3	8.28154089	0.0405*
SIZE	1	1	5.85367123	0.0155*
nPos	1	1	5.99028903	0.0144*
Rad	2	2	3.80537925	0.1492
Chemo	3	3	6.7370284	0.0808
pPheno	3	3	14.0972	0.0028*
Retested	1	1	19.2777376	<.0001*

Figure 22. Whole Dataset Final Results

analysis where any variable with a p-value of .2 or less was initially kept in the model. Fourteen variables were found to be statistically significant with a p-value of .20 or less in the initial analysis. The 14 significant variables were iteratively re-analyzed, removing variables with a p-value greater than .10, and the final results indicated that six variables (Grade, Size, nPos, Chemo, pPheno, Retested) were statistically significant with a p-value of .10 or less. I kept variables with p-values less than .1 as a conservative measure to ensure I identified any covariates that could potentially impact the survival time of patients. For demonstration purposes, the graphical illustration of the effect of the variable pPheno, the receptor status of the primary tumor, on survival time is shown in Figure 23. The key point to note is that TNBC patients perform more poorly (shorter survival time) than do the other three phenotypes. It was important to review the breakdown of TNBC patients in my discordant sample, noting whether treatment were plans based on the primary tumor versus the metastatic tumor.

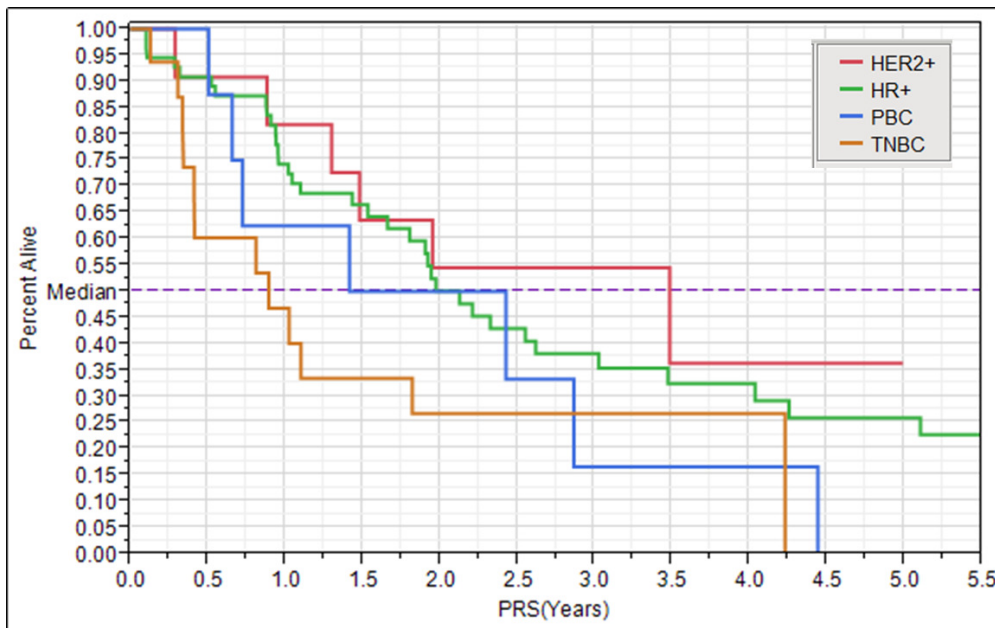


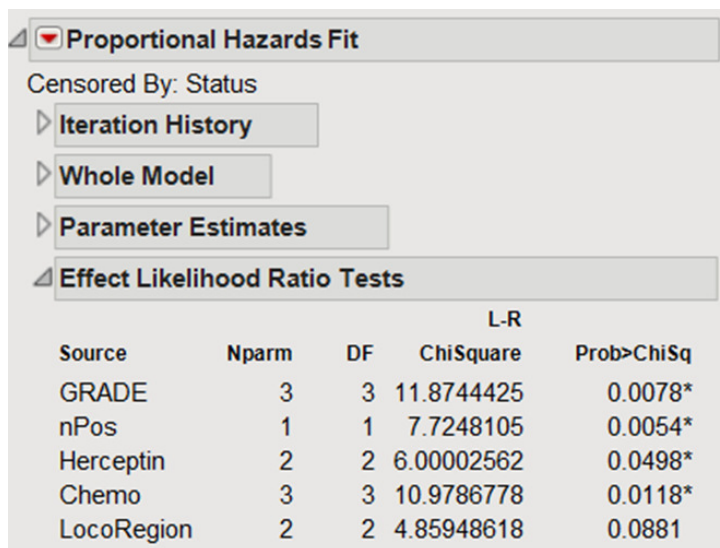
Figure 23. pPheno KM Analysis

If, for instance, there were significantly more patients in the metastatic tumor group who were TNBC than in the primary tumor group, a potential explanation for the differences between the groups would be the effect of the variable pPheno, rather than the impact the first-line treatment plan determination being based on the receptor status of the metastatic or primary tumor. This point is further explored in the Univariate Statistical Analysis section of this chapter.

Multivariate Results for Re-Tested Patients

A previous analysis indicated that those patients whose metastatic tumors were re-tested had better survival outcomes than those patients not re-tested.

Therefore, I conducted a multivariate analysis on the re-tested patients. Utilizing the Cox regression analysis on the re-tested patients, I employed the same procedure I used on the complete dataset. The final results of the analysis is shown in Figure 24. These results showed three of the six variables that were deemed statistically significant in the complete dataset, were also significant in the re-tested dataset. Additionally, two variables (Herceptin, LocoRegion) were



Source	Nparm	DF	ChiSquare	Prob>ChiSq
GRADE	3	3	11.8744425	0.0078*
nPos	1	1	7.7248105	0.0054*
Herceptin	2	2	6.00002562	0.0498*
Chemo	3	3	10.9786778	0.0118*
LocoRegion	2	2	4.85948618	0.0881

Figure 24. Multivariate Results for Re-Tested Patients

Statistically significant, yielding a total of eight variables identified as covariates that impacted patient survival and could potentially bias the univariate analysis performed on the discordant data. These eight variables warranted further univariate analysis.

Univariate Statistical Analysis

A two-sample univariate analysis was performed on each available covariate for the 14 discordant patients. Each analysis tested whether there were statistically significant differences between the subset of patients who had their treatment plan based on their primary tumor receptor status (n=8) and those who had treatment plans based on the metastatic tumor (n=6). For continuous variables, a two sample t-test was performed on the sample averages. For nominal variables, the exact Fisher test for contingency tables was used to test the equality of the percentage of patients in each nominal category. The analysis results are listed in Table 21. The eight covariates found in the multivariate analysis to be statistically significant are indicated with an asterisk. Given the small combined sample size of 14, the fact that no results were statistically significant was not unexpected.

To be thorough in the examination of the covariates, I reviewed the numerical differences between the two cohorts in spite of the lack of statistical significance for any variable. From a purely numerical standpoint, there was very little difference between the two samples for all but one covariate. For instance, the age at diagnosis for each sample differed numerically by only .2 years. Additionally, neither of the two cohorts had patients with TNBC primary tumors. Given that the multivariate analysis demonstrated that patients with TNBC

Table 21. Univariate Analysis

Variable	First-line Treatment Based on Metastatic Tumor Receptors (n=6)	First-line Treatment Based on Primary Tumor Receptors (n=8)	Statistical Significance (P-value)
Age	56.1	55.3	.890
TTR	1596.5	1390.9	.729
Grade*	2.6	2.6	.935
Size*	35.8	27.0	.463
nPos*	7.0	8.0	.832
Age DX	51.7	51.5	.975
pPheno*	Her2+(1), HR+(2), PBC(3), TNBC(0)	Her2+(1), HR+(6), PBC(1), TNBC(0)	.450
mPheno	Her2+(1), HR+(1), PBC(0), TNBC(4)	Her2+(0), HR+(1), PBC(2), TNBC(5)	.723
AJCC Simple	S1(0), S2(3), S3(3)	S1(2), S2(3), S3(3)	.600
LocoRegion*	Distant(6), Regional(0)	Distant(6), Regional(2)	.473
Visceral	Non(0), Visceral(6)	Non(4), Visceral(4)	.085
Hormone	Yes(3), No(3)	Yes(4), No(3)	.999
Chemo*	Yes(5), No(1)	Yes(5), No(3)	.580
Rad	Yes(5), No(1)	Yes(5), No(3)	.580
Herceptin*	Yes(4), No(1)	Yes(2), No(5)	.242

*Variables found statistically significant in one of the multivariate analyses.

primary tumors had shorter life expectancies, had the two cohorts contained a different number of TNBC patients there would have been cause for concern about a possible bias in the survival analysis. The one variable that did present some concern was Visceral. The Visceral variable is an indication of the location of the metastatic tumors. Visceral tumors are located in essential organs such as the lung, liver or brain. Non-visceral tumors are generally located in the bones and lymphatic system. All patients whose treatment plans were based on their metastatic tumor receptor status had metastatic tumors rated as visceral, while only 4 of 8 patients whose first-line treatment plan was based on their primary tumor were rated as visceral. Visceral and metastatic treatment plan were thus statistically confounded in this study. Figure 25 shows the KM survival analysis curve

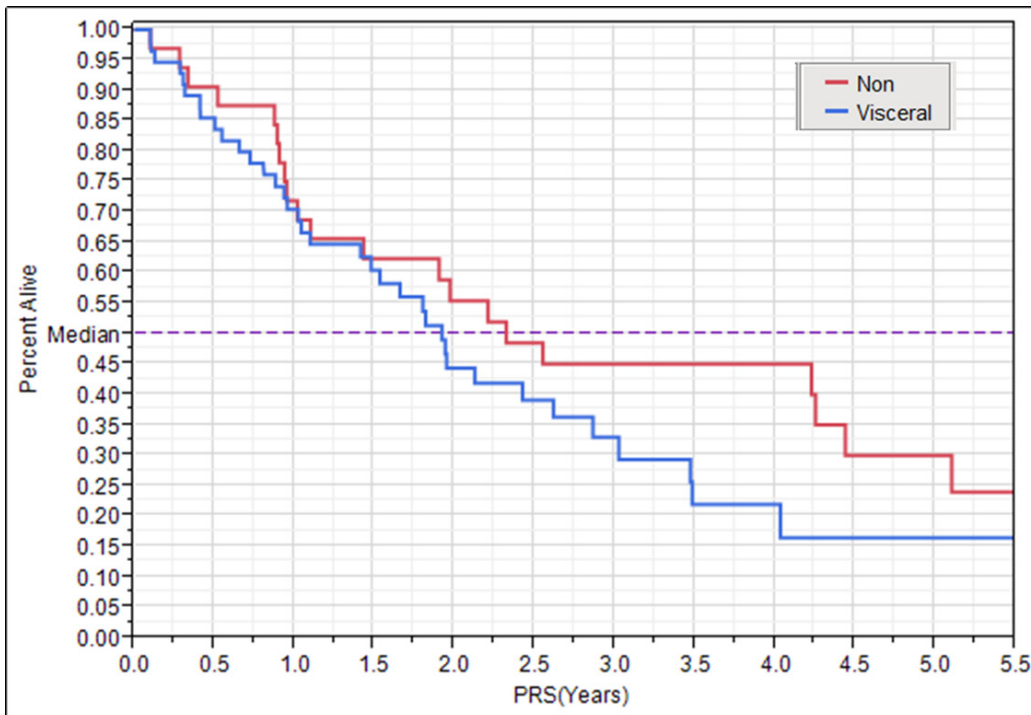


Figure 25. Visceral versus Non-Visceral

comparing visceral and non-visceral patients. The curves indicated that visceral patients performed worse (about 6 months shorter median PRS) than the non-visceral patients, although this result was not statistically significant with a p-value of .210. I made two evaluations of the impact the variable Visceral had on the primary versus metastatic tumor comparison results. First, I performed a Cox regression analysis on the discordant patients with the variable FLD (whether the primary or metastatic tumor receptor status was used to determine the first-line treatment plan) and Visceral in the model. The analysis results indicated that Visceral is not statistically significant and that even with Visceral in the model, FLD still had a hazard ratio of 4.2. These results are shown in Figure 26. This analysis indicated that patients whose first-line treatment plans were based on their metastatic tumor were 4.2 times more likely to die at a given point in time than those patients whose plans were based upon the primary tumor, even after accounting for the effect of the variable Visceral. Second, I performed a KM

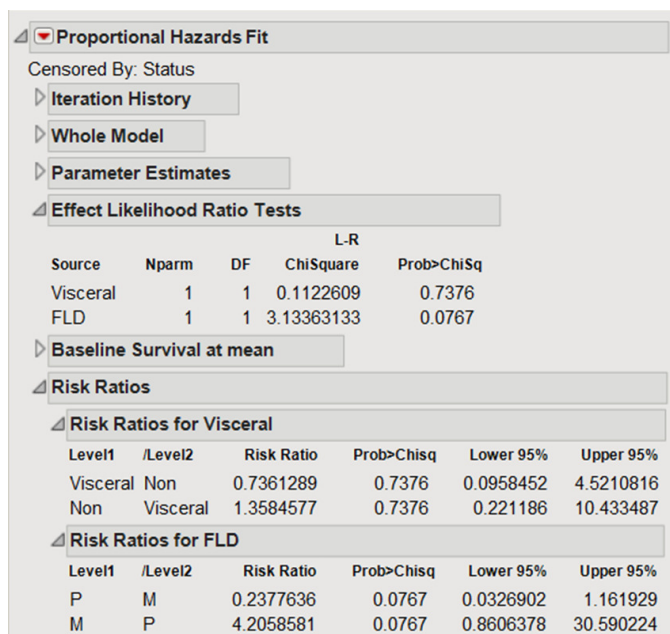


Figure 26. Cox Regression with Visceral

survival analysis on the discordant patients with only the visceral cases, which removed four of the patients whose first-line treatment plans were based on their primary tumor receptor status. These results are shown in Figure 27. The numerical and graphical results were nearly identical to the test results with the non-visceral patients included in the analysis. Of course, with the smaller sample size of 10 instead of 14, the p-value for the test of equal curves changed from .049 to .152. While there was certainly a difference in the two groups being compared based on the variable Visceral, I found no evidence that this fact was biasing or influencing the conclusions of this dissertation. Based on the entirety of the univariate analyses, I found no concern that a covariate biased the conclusions of this dissertation.

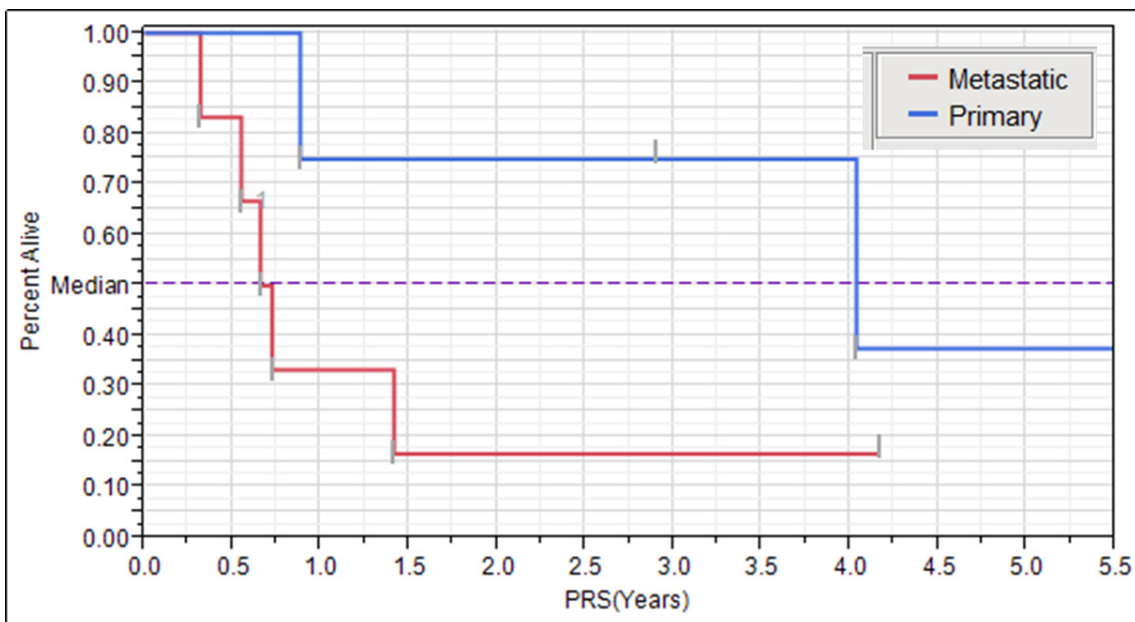


Figure 27. Visceral Patients Only

Sensitivity Analysis

As previously discussed, the four records in Table 19 represented data where the interpretation of the records could have yielded different variable values and thus impact the analysis of the discordant cases. I conducted two additional analyses with this data; one where the data was coded with a reasonable interpretation of the record such that the interpretation yielded data that was less supportive of my hypothesis, and a second analysis where each record was interpreted such that the coding resulted in data more supportive of my hypothesis. The analysis in Figure 28 (best case) represents the data where those four records were coded to support of my hypothesis. Obviously, the results showed a better numerical difference compared to the results with the originally coded data, with a best case median survival time of 42 months compared to the previously discussed estimate of 40 months shown in Figure 20. This result was statistically significant with a p-value of .0006.

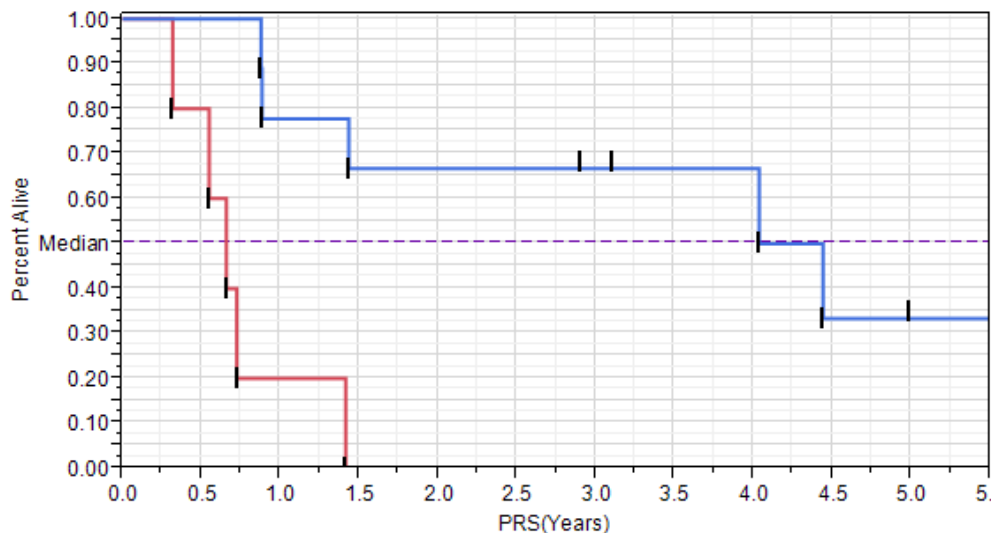


Figure 28. Best Case Curves

The analysis in Figure 29 (worst case) represents the data where the four records were coded to *not* be supportive of my hypothesis. Understandably, the results showed a worse numerical result compared to the originally coded data, with a worst case median survival time of 22 months compared to the previously discussed estimate of 40 months shown in Figure 20. These results were not statistically significant with a p-value of .41.

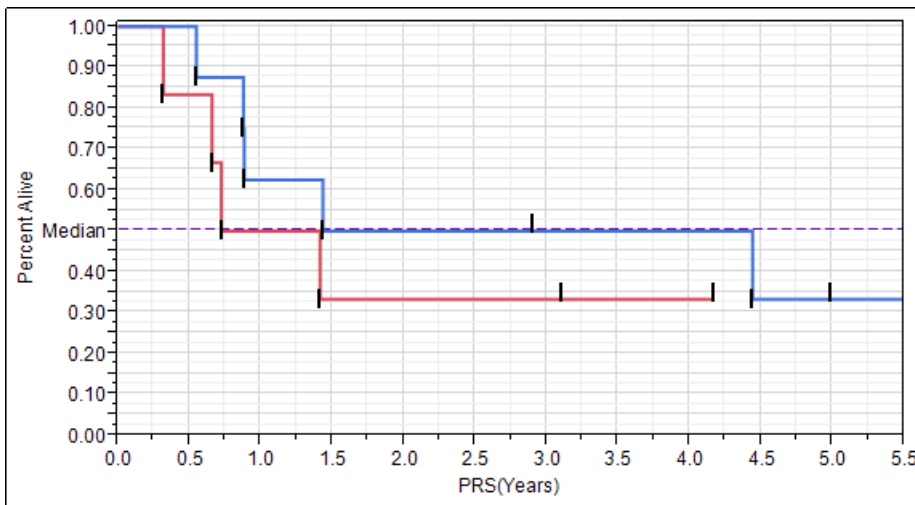


Figure 29. Worst Case Curves

CHAPTER 5

Conclusions and Recommendations

Conclusion

The purpose of this dissertation was to determine the impact on patient outcomes in recurrent metastatic breast cancer diagnoses when the first-line treatment plan was based on the receptor status of metastatic tumor(s) instead of the receptor status of the primary tumor(s).

Restatement of Hypothesis

For discordant cases, if the first-line treatment plan is based on the receptor status of the primary tumor, the median life expectancy of MBC patients will be longer than MBC patients whose first-line treatment plan is based on the receptor status of the metastatic tumor.

The weight of the evidence in this research is consistent with my hypothesis. The evidence is supported by: a literature review of historical data that built a strong case indicating that discordant MBC patients performed worse than concordant patients; the log rank test results with a p-value of .049; the covariate analysis; and the sensitivity analysis. A priori, based on the literature review, I expected a minimum of a one year median increase in life expectancy for patients with discordant receptor statuses whose first-line treatment plan was based on the primary tumor's receptor status instead of the metastatic receptor status. Ultimately, the sample data showed that discordant MBC patients, whose first-line treatment plan was based on their primary tumor rather than their metastatic tumor, had a longer median life expectancy of 40 months. The

sensitivity analysis resulted in a worst case scenario of a 22 month median difference, which still reinforces my hypothesis. Regardless of which life expectancy estimate is theorized, 12, 22, or 40 months, my research provides a distinct directive for determining first-line treatment plans for discordant MBC patients. I proposed that if the data analysis supported my hypothesis this would indicate a need for revised standards of care for MBC patients whose metastatic tumor receptor statuses differed from that of the primary tumor.

Recommendation

Since 1989, the clinical impact of testing metastatic tumors for receptor status to inform first-line treatment plans has been studied (Kamby et al., 1989). Prior to my research, researchers had drawn no definitive conclusions regarding the impact of re-testing receptor statuses on the clinical outcomes of metastatic breast cancer patients. In Chapter 1, I stated the ultimate goal of this dissertation was to influence changes in international and national standards regarding the determination of first-line treatment plans in discordant cases of MBC, and to provide a higher level of evidence for those standards.

Proposed Standard of Care

To assist physicians as they consult the relevant ABC2 and NCCN guidelines and levels of evidence concerning the care of metastatic breast cancer patients with discordant receptor statuses, I propose the implementation of the following overarching standard of care:

Where discordance between the primary and metastatic tumor receptor status would indicate different treatments, the status of the primary tumor should take precedent when developing the first-line treatment plan for a newly diagnosed

recurrent metastatic breast cancer patient. Strong clinical evidence to the contrary must be present to warrant basing the treatment plan on the metastatic tumor receptor status.

Level of Evidence

The NCCN and ABC2 standards discussed in the Literature Review section of my paper are not consistent with my research findings (Cardoso et al., 2014; NCCN.org, 2014). As reported in Chapter 2, neither entity supports the standards recommendations relevant to my hypothesis with citations from existing literature. The ABC2 committee states their recommendation is based on expert opinion only. Considering the ABC2 level of evidence definitions, I recommend the level of evidence for my proposed standard of care be rated at 2B. Level 2B is defined as “RCTs with important limitations or exceptionally strong evidence from observational studies” and clinical implications of a “weak recommendation, best action may differ depending on circumstances or patients’ or societal values” (Cardoso et al., 2014, p. 2).

Research Limitations

Below I discuss five limitations of this research that could potentially impede the application of my results to patient medical decisions. I then review how my research design minimized or compensated for these limitations. While other limitations may exist, I consider these five to be the most relevant.

Sample Size

My research sample size of discordant cases (n=14), included eight cases whose treatment plan was determined by the primary tumor receptor status, and six cases whose treatment plan was determined by the metastatic tumor receptor

status. This sample size was statistically small, and it provided little statistical power, as illustrated by the simulation results in Chapter 3. In addition to statistical matters, practical concerns arose about the impact that data quality could have on analyses, given that each record carried more weight in the statistical calculations since the sample size was small. While my total sample size ($n=124$) and the number of discordant cases ($n=14$) was small, such sample sizes are common in the published literature and several heavily cited articles on this topic (Table 20). For future studies, a larger sample size would be desirable and may be obtained by including more and larger cancer centers.

Data Quality

Working with three sources of data, the tumor registry, electronic records and paper patient charts, presented several data challenges. The most important data quality challenge was that of applying operational definitions when physicians utilized an interpretation of lab results that differed from the lab report itself. Additionally, incomplete electronic records and lack of access to paper charts diminished the available sample of patients. With a small sample size, data quality was even more important since one or two inaccurate observations could impact the analytical results and resultant conclusions. Data quality issues and challenges were discussed in the Data Acquisition and Sensitivity Analysis sections of Chapter 4. Understanding the importance of data quality when working with small sample sizes meant that I needed to employ safeguards as checks on data quality. I believe that the sensitivity analysis performed provided protection against any data quality issues and thus limited any impact on the accuracy of my results and conclusions. The best protection against data quality issues was clear operational definitions, which I had, and larger sample sizes, which I strongly recommend in future research.

Inference Space

The fact that this study was performed at only one cancer center was a limitation. Variation in cancer centers' patient populations, policies and procedures could potentially impact the ability to infer the results of my research to other centers. While the results of my retrospective study are statistically most valid for the population from which they were derived, I believe they can be applied more widely from a practical standpoint. The fact that there have been no clear standards of care regarding first-line treatment plans for MBC patients with discordant receptor results created the opportunity to observe varied strategies employed by the doctors at the UT Cancer Center. This variation, based on doctor preferences, provides logical support for applying my results beyond just one cancer center and one set of doctors. The best method for ensuring a wider inference space for future research is to involve several cancer centers in the protocol.

Limited Ability to Examine Covariates

In retrospective observational studies, it is important to evaluate the potential effect of covariates on the research results. Such concern arises from the researchers' lack of ability to control important variables when a study is retrospective rather than prospective. The importance of covariate analysis in the specific research performed in this dissertation is reviewed by several authors and discussed in the literature review (Dieci et al., 2012; Foukakis et al., 2012). In my research, the sample size hindered my ability to establish the impact of covariates on the results of this study. While I found no concerns for potential biasing of results based on the covariate analysis discussed in Chapter 4, it remains possible that confounding effects exist that bias the analysis results. Again, a larger sample size would help alleviate this limitation in future research.

Researcher Bias

Creswell (2009), suggests researchers clarify through self-reflection any potential for researcher bias in the design and implementation of their research projects (p. 192). As the researcher for this dissertation, I certainly had an emotional attachment to my hypothesis, given that it was motivated as a result of my wife's death. To avoid any bias or appearance of bias on my part, I applied various protections in the design and implementation of my research. The design of this research was reviewed by two IRB boards and my dissertation committee that had a practicing research oncologist as a member. Given the research design was fully vetted, the remaining potential for researcher bias was in the collection and analysis of the data. To eliminate bias in the analysis phase, the analysis plans and standards were established a priori. These plans and standards were ultimately implemented as the research design indicated.

My initial idea for removing potential researcher bias in the data collection phase was to have someone else collect the data, such as a medical resident. My committee chair and the oncologist member of my committee both recommended that I collect the data myself for two reasons: First, relying on someone else would mean the length of data collection time was out of my control. Second, the committee members believed collecting my own data provided an important learning experience, since I had never before collected medical data. Once the decision was made that I would collect the data, I developed three strategies to protect against potential data collection researcher bias. I first created detailed operational definitions of each data variable in an attempt to eliminate the need for judgement calls when evaluating patient records. Next, my quality review process involved an independent expert, the Tumor Registry manager, who reviewed all the discordant cases and other samplings of records. Lastly, I meticulously documented every step of the data acquisition, cleaning and collection process. My records allow an independent

observer to retrace any of my data records to the original patient records or to track any patient ID to the documented and analyzed data.

The advice of the committee members to collect my own data proved invaluable. Because of the complexity of the data collection process, I believe an independent person, not immersed in the topic of this research endeavor, could not have completed the collection process without being in constant consultation with me, thus clouding the intended independence. The discovery, outlined in Chapter 4, that those patients who did not have their metastatic tumor receptor status re-tested performed poorly, was a surprising and important finding. Because I personally conducted the record reviews, I was able to discern that in most cases where the patient was not re-tested, the patient had presented in such a dire situation that there either was no time to re-test or it was deemed too late for any treatment to succeed. An independent data collector would have likely missed this subtlety in the medical records since this information was not part of the pre-planned data collection process. Had I not collected the data myself, the interpretation of this one aspect of the data would have added an additional limitation to my results.

Contribution

At the time of completing this dissertation, I was unaware of any published scholarly paper or unpublished work that had expressly addressed the clinical impact of utilizing the metastatic receptor status of patients to determine their first-line treatment plan. For over 20 years, researchers have studied issues related to this topic; however, the precise question I examined in this dissertation was never evaluated. For the first time, the clinical impact of utilizing the receptor status of metastatic tumors, when determining MBC patients' first-

line treatment plan has been studied and reported. Additionally, my stated goal has always been to impact the national and international standards of care for discordant MBC patients. Proposing the over-arching standard recommending medical decisions for discordant MBC patients focus on the receptor status of the primary tumor and upgrading the level of evidence for that standard to 2B is a further contribution of my research. My intent is that this research leads to physicians consulting and implementing the over-arching standard of care I proposed, thus extending the lives of MBC patients.

Future Research

The primary need for future research is the engagement of additional cancer centers to utilize the protocols of this study to increase the sample size of data specifically targeting my hypothesis and expand the inference space for implementing the conclusions I have reached. Since the current medical standards on the appropriate treatment plan for discordant cases of recurrent metastatic cancer are weak, it would be beneficial to have data from various cancer centers that perhaps utilize different criteria about re-testing receptors and how that data is used to determine treatment plans. For example, in my personal experience with my wife and consulting the oncologist on my committee, I encountered three cancer centers (Piedmont in Atlanta, MD Anderson in Houston, and UT Knoxville) with at least three different approaches to determining first-line treatment plans. At one center, I witnessed a physician who did not make it a practice of re-testing the metastatic receptors to determine the first-line treatment plan. At a second center, the physician's standard practice was to re-test and utilize the results of that test to assist in treatment decisions. At the third center, there was no written standard; there

was a general practice of re-testing when possible, and each doctor then decided independently how to utilize the information on the metastatic tumor receptor status. I see two paths for expanding the number of cancer centers involved in my research. First, I will seek to engage other researchers, identified through my literature review, who have studied the impact of discordant receptor statuses. Based on my review, I believe some authors already have the data available from previous research to implement the analysis on cohorts E and F of my study. Other authors would need to add a new variable to their existing data identifying whether the first-line treatment plan was based on the primary or metastatic tumor receptor status. That additional variable would allow the researcher to determine the impact of the first-line treatment plan decisions. Second, I will seek to engage other cancer centers to grant me access to their data. With additional patients from different cancer centers, the amount of data would mount and contribute to an even higher level of evidence to inform the determination of first-line treatment plans for MBC patients with discordant tumor receptor statuses.

Unanswered Question

One unexplored question in my research and in the literature is: *Why does determining first-line treatment plans based on the primary tumor receptor status, as opposed to the metastatic tumor, result in better patient outcomes?*

The only answer proposed in the literature is measurement error. Liedtke et al. (2009), provides a good discussion of the potential impact of measurement error in the testing of MBC patients metastatic tumor receptor status. My belief is that measurement error is certainly part of the explanation, but to what extent is unknown. Presently I am content to know *what* the right decision is without having to understand *why* that decision is the correct one. If others wish to investigate this question, then that knowledge may lead to enhancements in the

first-line treatment plan decision making process and an increased understanding of the metastatic disease.

Definition of Discordance

One of the challenges during my literature review was that several definitions of discordance were utilized by researchers. I provided a clear and logical definition for discordance in Chapter 3. From a scientific point of view, having a universally applied definition of discordance is important. Such a definition would promote better research designs, produce clearer analyses, allow for more effective comparisons across studies and enable quantitative meta-analyses to be performed.

Impact

In September of 2013, Amy Foster was given four years to live. She died four months later after her treatment plan was modified based on the discordant receptor status of her metastatic tumor. As discussed in chapter 1, an estimated 60,000 women per year world-wide are faced with a decision because their test results show discordant receptor statuses. For the majority of those women, their first-line treatment plan will be based on their metastatic tumor and they will suffer the consequence of a shortened median life expectancy of one to three years. It is my hope that this dissertation will help affect a change in the standards of care for women with discordant receptor statuses. With the data I have, I will personally recommend to anyone faced with this particular medical decision to base their first-line treatment plan on the primary tumor receptor status. As new and better data is developed, then the level of evidence can be raised and the relevant standards updated appropriately. Cardoso (2009), calls metastatic breast cancer patients *the forgotten heroes* because today's breast

cancer research focuses on early detection and treatment of primary tumor breast cancer, not on challenges and issues facing metastatic patients. She laments, and I agree, "it is clearly a time for change!" (Cardoso, 2009, p. 271).

LIST OF REFERENCES

- Amir, E., Clemons, M., Purdie, C. A., Miller, N., Quinlan, P., Geddie, W., . . . Thompson, A. M. (2012). Tissue confirmation of disease recurrence in breast cancer patients: pooled analysis of multi-centre, multi-disciplinary prospective studies. *Cancer Treat Rev*, 38(6), 708-714. doi: 10.1016/j.ctrv.2011.11.006
- Amir, E., Miller, N., Geddie, W., Freedman, O., Kassam, F., Simmons, C., . . . Clemons, M. (2012). Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *J Clin Oncol*, 30(6), 587-592. doi: 10.1200/JCO.2010.33.5232
- Aurilio, G., Disalvatore, D., Prunerì, G., Bagnardi, V., Viale, G., Curigliano, G., . . . Nole, F. (2014). A meta-analysis of oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 discordance between primary breast cancer and metastases. *Eur J Cancer*, 50(2), 277-289. doi: 10.1016/j.ejca.2013.10.004
- Aurilio, G., Monfardini, L., Rizzo, S., Sciandivasci, A., Preda, L., Bagnardi, V., . . . Nole, F. (2013). Discordant hormone receptor and human epidermal growth factor receptor 2 status in bone metastases compared to primary breast cancer. *Acta Oncol*, 52(8), 1649-1656. doi: 10.3109/0284186X.2012.754990
- Berry, D. A., Cronin, K. A., Plevritis, S. K., Fryback, D. G., Clarke, L., Zelen, M., . . . Feuer, E. J. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*, 353(17), 1784-1792. doi: 10.1056/NEJMoa050518
- Botteri, E., Disalvatore, D., Curigliano, G., Brollo, J., Bagnardi, V., Viale, G., . . . Rotmensz, N. (2012). Biopsy of liver metastasis for women with breast cancer: impact on survival. *Breast*, 21(3), 284-288. doi: 10.1016/j.breast.2011.12.014
- Breast Cancer. (2015). *American Cancer Society*(Cancer Information Database).
- Cardoso, F. (2009). Metastatic breast cancer patients: the forgotten heroes! *Breast*, 18(5), 271-272. doi: 10.1016/j.breast.2009.09.001
- Cardoso, F., Costa, A., Norton, L., Cameron, D., Cufer, T., Fallowfield, L., . . . Winer, E. (2012). 1st International consensus guidelines for advanced breast cancer (ABC 1). *Breast*, 21(3), 242-252. doi: 10.1016/j.breast.2012.03.003
- Cardoso, F., Costa, A., Norton, L., Senkus, E., Aapro, M., André, F., . . . Winer, E. (2014). ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2). *Annals of Oncology*(00), 1-18. doi: 10.1093/annonc/mdu38510.1016/j.breast
- Carlson, R. W., Allred, D. C., Anderson, B. O., Burstein, H. J., Edge, S. B., Farrar, W. B., . . . Sheard, D. A. (2012). Metastatic Breast Cancer, Version 1.2012: Featured Updates to the NCCN Guidelines. *J Natl Compr Canc Netw*, 10(7), 821-829.

- Chang, H. J., Han, S. W., Oh, D. Y., Im, S. A., Jeon, Y. K., Park, I. A., . . . Kim, T. Y. (2011). Discordant human epidermal growth factor receptor 2 and hormone receptor status in primary and metastatic breast cancer and response to trastuzumab. *Jpn J Clin Oncol*, 41(5), 593-599. doi: 10.1093/jjco/hyr020
- Creswell, J. W. (2009). *Research Design: Qualitative, quantitative, and mixed methods approaches* (Third ed.): SAGE Publications, Inc.
- Criscitiello, C., André, F., Thompson, A. M., Laurentiis, M. D., Esposito, A., Gelao1, L., . . . Curigliano, G. (2014). Biopsy confirmation of metastatic sites in breast cancer patients: clinical impact and future perspectives. *Breast Cancer Research*, 16(205), 1-8.
- CTCA. (2014). Cancer Treatment Centers of America Website. Retrieved Comparison Graph to SEER, from <http://www.cancercenter.com/breast-cancer/statistics/>
- Dawood, S., Broglio, K., Buzdar, A. U., Hortobagyi, G. N., & Giordano, S. H. (2010). Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. *J Clin Oncol*, 28(1), 92-98. doi: 10.1200/JCO.2008.19.9844
- de Duenas, E. M., Hernandez, A. L., Zotano, A. G., Carrion, R. M., Lopez-Muniz, J. I., Novoa, S. A., . . . Gonzalez-Angulo, A. M. (2014). Prospective evaluation of the conversion rate in the receptor status between primary breast cancer and metastasis: results from the GEICAM 2009-03 ConvertHER study. *Breast Cancer Res Treat*, 143(3), 507-515. doi: 10.1007/s10549-013-2825-2
- Dieci, M. V., Barbieri, E., Piacentini, F., Ficarra, G., Bettelli, S., Dominici, M., . . . Guarnieri, V. (2012). Discordance in receptor status between primary and recurrent breast cancer has a prognostic impact: a single-institution analysis. *Ann Oncol*, 24(1), 101-108. doi: 10.1093/annonc/mds248
- Duchnowska, R., Dziadziuszko, R., Trojanowski, T., Mandat, T., Och, W., Czartoryska-Arlukowicz, B., . . . Polish Brain Metastasis, C. (2012). Conversion of epidermal growth factor receptor 2 and hormone receptor expression in breast cancer metastases to the brain. *Breast Cancer Res*, 14(4), R119. doi: 10.1186/bcr3244
- Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., . . . Verweij, J. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*, 45(2), 228-247. doi: 10.1016/j.ejca.2008.10.026
- Farolfi, A., Ibrahim, T., Scarpi, E., & Amadori, D. (2013). Biology matters: the clinical impact of single-receptor discordance on breast cancer. *Ann Oncol*, 24(3), 851. doi: 10.1093/annonc/mdt005

- Foukakis, T., Astrom, G., Lindstrom, L., Hatschek, T., & Bergh, J. (2012). When to order a biopsy to characterise a metastatic relapse in breast cancer. *Ann Oncol*, 23 Suppl 10, x349-353. doi: 10.1093/annonc/mds297
- Gigerenzer, G., & Marewski, J. N. (2014). Surrogate Science: The Idol of a Universal Method for Scientific Inference. *Journal of Management*, 41(2), 421-440. doi: 10.1177/0149206314547522
- Global Cancer Facts & Figures. (2012). *American Cancer Society*.
- Guyatt, G., Gutterman, D., Baumann, M. H., Addrizzo-Harris, D., Hylek, E. M., Phillips, B., . . . Schünemann, H. (2006). Grading strength of recommendations and quality of evidence in clinical guidelines*: Report from an american college of chest physicians task force. *Chest*, 129(1), 174-181. doi: 10.1378/chest.129.1.174
- Harries, M., & Smith, I. (2002). The development and clinical use of trastuzumab (Herceptin). *Endocrine-Related Cancer*, 9, 75-85.
- Hoefnagel, L. D., van der Groep, P., van de Vijver, M. J., Boers, J. E., Wesseling, P., Wesseling, J., . . . van Diest, P. J. (2013). Discordance in ERalpha, PR and HER2 receptor status across different distant breast cancer metastases within the same patient. *Ann Oncol*, 24(12), 3017-3023. doi: 10.1093/annonc/mdt390
- Howlader N, N. A., Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (2015). SEER Cancer Statistics Review, 1975-2012 (based on November 2014 SEER data submission, posted to the SEER web site, April 2015. ed.): National Cancer Institute. Bethesda, MD.
- Idirisinghe, P. K., Thike, A. A., Cheok, P. Y., Tse, G. M., Lui, P. C., Fook-Chong, S., . . . Tan, P. H. (2010). Hormone receptor and c-ERBB2 status in distant metastatic and locally recurrent breast cancer. Pathologic correlations and clinical significance. *Am J Clin Pathol*, 133(3), 416-429. doi: 10.1309/AJCPJ57FLLJRXKPV
- Jozwiak, K., & Moerbeek, M. (2011). Power Analysis for Trials With Discrete-Time Survival Endpoints. *Journal of Educational and Behavioral Statistics*, 37(5), 630-654. doi: 10.3102/1076998611424876
- Kamby, C., Rasmussen, B. B., & Kristensen, B. (1989). Oestrogen receptor status of primary breast carcinomas and their metastases. Relation to pattern of spread and survival after recurrence. *Br J Cancer*, 60(2), 252-257.
- Karlsson, E., Appelpgren, J., Solterbeck, A., Bergenheim, M., Alvariza, V., & Bergh, J. (2014). Breast cancer during follow-up and progression - A population based cohort on new cancers and changed biology. *Eur J Cancer*. doi: 10.1016/j.ejca.2014.08.014
- Kuukasjarvi, T., Kononen, J., Helin, H., Holli, K., & Isola, J. (1996). Loss of estrogen receptor in recurrent breast cancer is associated with poor response to endocrine therapy. *J Clin Oncol*, 14(9), 2584-2589.

- Lakatos, E. (1988). Sample Sizes Based on the Log-Rank Statistic in Complex Clinical Trials. *Biometrics*, 44(1), 229-241.
- Lee, E. T., & Wang, J. W. (2013). *Statistical Methods for Survival Data Analysis* (4th ed.): John Wiley & Sons, Inc.
- Liedtke, C., Broglio, K., Moulder, S., Hsu, L., Kau, S. W., Symmans, W. F., . . . Gonzalez-Angulo, A. M. (2009). Prognostic impact of discordance between triple-receptor measurements in primary and recurrent breast cancer. *Ann Oncol*, 20(12), 1953-1958. doi: 10.1093/annonc/mdp263
- Lindstrom, L. S., Karlsson, E., Wilking, U. M., Johansson, U., Hartman, J., Lidbrink, E. K., . . . Bergh, J. (2012). Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol*, 30(21), 2601-2608. doi: 10.1200/JCO.2011.37.2482
- Lower, E. E., Glass, E., Blau, R., & Harman, S. (2009). HER-2/neu expression in primary and metastatic breast cancer. *Breast Cancer Res Treat*, 113(2), 301-306. doi: 10.1007/s10549-008-9931-6
- Lower, E. E., Glass, E. L., Bradley, D. A., Blau, R., & Heffelfinger, S. (2005). Impact of metastatic estrogen receptor and progesterone receptor status on survival. *Breast Cancer Res Treat*, 90, 65 -70.
- Montagna, E., Bagnardi, V., Rotmensz, N., Viale, G., Renne, G., Canello, G., . . . Colleoni, M. (2012). Breast cancer subtypes and outcome after local and regional relapse. *Ann Oncol*, 23(2), 324-331. doi: 10.1093/annonc/mdr129
- NCCN.org. from <http://www.nccn.org/about/default.aspx>
- NCCN.org. (2014). NCCN Clinical Practice Guidelines in Oncology - Breast Cancer. *Version 3.2014*.
- Niikura, N., Liu, J., Hayashi, N., Mittendorf, E. A., Gong, Y., Palla, S. L., . . . Ueno, N. T. (2012). Loss of human epidermal growth factor receptor 2 (HER2) expression in metastatic sites of HER2-overexpressing primary breast tumors. *J Clin Oncol*, 30(6), 593-599. doi: 10.1200/JCO.2010.33.8889
- Penault-Llorca, F., Coudry, R. A., Hanna, W. M., Osamura, R. Y., Ruschoff, J., & Viale, G. (2013). Experts' opinion: Recommendations for retesting breast cancer metastases for HER2 and hormone receptor status. *Breast*, 22(2), 200-202. doi: 10.1016/j.breast.2012.12.004
- Richard Van Noorden, & Brendan Maher, R. N. (2014). The top 100 papers. *Nature*, 514, 550-553. doi: 10.1038/514550a
- Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., . . . Ockene, J. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*, 288(3), 321-333.

- Sharpe, N. R., Veaux, R. D. D., & Vellman, P. F. (2010). *Business Statistics* (Instructor, 1st ed.): Addison Wesley.
- Sighoko, D., Liu, J., Hou, N., Gustafson, P., & Huo, D. (2014). Discordance in hormone receptor status among primary, metastatic, and second primary breast cancers: biological difference or misclassification? *Oncologist*, *19*(6), 592-601. doi: 10.1634/theoncologist.2013-0427
- Simmons, C., Miller, N., Geddie, W., Gianfelice, D., Oldfield, M., Dranitsaris, G., & Clemons, M. J. (2009). Does confirmatory tumor biopsy alter the management of breast cancer patients with distant metastases? *Ann Oncol*, *20*(9), 1499-1504. doi: 10.1093/annonc/mdp028
- Snedecor, G. W., & Cochran, W. G. (1967). *Statistical Methods* (6th ed.): The Iowa State University Press.
- Spataro, V., Price, K., Goldhirsch, A., Cavalli, F., Simoncini, E., Castiglione, M., . . . Gelber, R. D. (1992). Sequential estrogen receptor determinations from primary breast cancer and at relapse: prognostic and therapeutic relevance. The International Breast Cancer Study Group (formerly Ludwig Group). *Ann Oncol*, *3*(9), 733-740.
- Spilling the Beans. (2015). *The Economist*, *416*, 62.
- Thompson, A. M., Jordan, L. B., Quinlan, P., Anderson, E., Skene, A., Dewar, J. A., . . . Breast Recurrence in Tissues Study, G. (2010). Prospective comparison of switches in biomarker status between primary and recurrent breast cancer: the Breast Recurrence In Tissues Study (BRITS). *Breast Cancer Res*, *12*(6), R92. doi: 10.1186/bcr2771
- Turner, N. H., & Di Leo, A. (2013). HER2 discordance between primary and metastatic breast cancer: assessing the clinical impact. *Cancer Treat Rev*, *39*(8), 947-957. doi: 10.1016/j.ctrv.2013.05.003
- Wilkings, U., Karlsson, E., Skoog, L., Hatschek, T., Lidbrink, E., Elmberger, G., . . . Bergh, J. (2011). HER2 status in a population-derived breast cancer cohort: discordances during tumor progression. *Breast Cancer Res Treat*, *125*(2), 553-561. doi: 10.1007/s10549-010-1029-2
- Yang, Y. F., Liao, Y. Y., Yang, M., Peng, N. F., Xie, S. R., & Xie, Y. F. (2014). Discordances in ER, PR and HER2 receptors between primary and recurrent/metastatic lesions and their impact on survival in breast cancer patients. *Med Oncol*, *31*(10), 214. doi: 10.1007/s12032-014-0214-2

APPENDIX

Appendix 1: List of Abbreviations

Table 22. Appendix 1- List of Abbreviations

Abbreviation	Definition
ABC1	First International Advanced Breast Cancer Conference
ABC2	Second International Advanced Breast Cancer Conference
ER	Estrogen Receptor status
HER2	Human Epidermal growth factor Receptor 2
HER2+	HER2 positive. HR Negative
HR+	Hormone receptor positive (ER and/or PR positive), and
LDS	Limited Data Set
LoE	Level of Evidence
MBC	Metastatic Breast Cancer
MDA	M.D. Anderson Medical Center
NCCN	National Comprehensive Cancer Network®
OS	Overall Survival
PBC	Positive Breast Cancer (HR+ and HER2+)
PR	Progesterone Receptor status
PRS	Post Recurrence Survival
SAS	Statistical Analysis System (statistical software)
TNBC	Triple Negative Breast Cancer
KM	Kaplan-Meier Survival analysis method
HR	Hazard Ratio
PHI	Protected Health Information
AJCC	American Joint Committee on Cancer

Appendix 2: Incidence Rate

Table 23. Appendix - Seer Breast Cancer Incidence Rates

Year of Diagnosis	All Races, Females, All Ages	Deaths All
1975	105.07	31.45
1976	101.92	31.8
1977	100.82	32.48
1978	100.64	31.73
1979	102.14	31.21
1980	102.26	31.68
1981	106.37	31.92
1982	106.5	32.19
1983	111.15	32.07
1984	116	32.9
1985	124.26	32.98
1986	126.86	32.87
1987	134.5	32.66
1988	131.33	33.2
1989	127.22	33.23
1990	131.86	33.14
1991	133.85	32.69
1992	132.05	31.64
1993	129.22	31.39
1994	131	30.92
1995	132.63	30.55
1996	133.7	29.49
1997	137.99	28.21
1998	141.39	27.54
1999	141.46	26.61
2000	136.48	26.64
2001	138.76	26.01
2002	135.73	25.62
2003	126.92	25.27
2004	128.03	24.49
2005	126.51	24.14
2006	126.14	23.56
2007	128.06	22.96
2008	128.11	22.55
2009	130.52	22.24
2010	126.46	21.92
2011	129.56	21.54

Appendix 3: Permissions

Periodical Permissions

Order Date	Article Title	Publication	Type Of Use	Order Status	Order Number
20-Oct-2015	Breast cancer during follow-up and progression – A population based cohort on new cancers and changed biology	European Journal of Cancer	reuse in a thesis/dissertation	Completed	3733270409478
20-Oct-2015	HER2 status in a population-derived breast cancer cohort: discordances during tumor progression	Breast Cancer Research and Treatment	Thesis/Dissertation	Completed	3733270102028
20-Oct-2015	Discordant Human Epidermal Growth Factor Receptor 2 and Hormone Receptor Status in Primary and Metastatic Breast Cancer and Response to Trastuzumab	Japanese Journal of Clinical Oncology	Thesis/Dissertation	Completed	3733261316261
20-Oct-2015	Prospective Study Evaluating the Impact of Tissue Confirmation of Metastatic Disease in Patients With Breast Cancer	Journal of Clinical Oncology	Dissertation/Thesis	Completed	3733261183978
20-Oct-2015	Loss of Human Epidermal Growth Factor Receptor 2 (HER2) Expression in Metastatic Sites of HER2-Overexpressing Primary Breast Tumors	Journal of Clinical Oncology	Dissertation/Thesis	Completed	3733260921377
20-Oct-2015	Prognostic impact of discordance between triple-receptor measurements in primary and recurrent breast cancer.	Annals of Oncology	Thesis/Dissertation	Completed	3733260686273
20-Oct-2015	ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2).	Annals of Oncology	Thesis/Dissertation	Completed	3733251502750

Figure 30. Periodical Permissions

NCCN Permission

NCCN allows students to utilize NCCN content without specific permission. The NCCN website states “Students using NCCN content in academic thesis or presentations for research purposes do not require written permission from NCCN. However, a credit line acknowledging NCCN, the name of the NCCN Guideline, and NCCN's website (www.nccn.org) should be included”(NCCN.org).

Appendix 4: Operational Definitions

1. Recurrence date

- a. Earliest known date of;
 - i. Symptom occurrence that led to doctor visit
 - ii. Date of doctor visit where indications of metastatic disease are identified
 - iii. Date of confirmation of metastatic disease

2. Description

- a. Loco-regional
- b. Distant
- c. Non-visceral
- d. Visceral
- e. Visceral Crisis

3. Receptors Retested? Reason Codes

- a. **A-** No, Not feasible/safe
- b. **B-** No, Standard practice
- c. **C-** No, No indication of need
- d. **D-** Yes, Primary results missing, old or suspect
- e. **E-** Yes, Standard practice
- f. **F-** Yes, Clinical indications

Result – as interpreted by the doctor

Location of biopsy? Tissue, location

4. Results Concordant? By 4 phenotype definition

- a. TNBC : Triple Negative Breast Cancer
- b. HR+ : Hormone receptor positive (ER and/or PR positive), **and** HER2 negative
- c. HER2+ : HER2/neu positive. HR Negative
- d. PBC: Positive Breast Cancer HR positive **and** HER2 positive

5. First-Line Treatment Plan based on? Primary Metastatic

- By doctor notes preferably. By doctor interview. By comparing receptor status with NCCN guidelines if necessary.

a. Why?

- i. **A-** Standard Practice
- ii. **B-** Doctor preference
- iii. **C-** Clinical indications
- iv. **D-** Primary results missing, old or suspect
- v. **E-** Other

6. First Line Treatment?

- a. **A-** Chemotherapy
- b. **B-** Endocrine Therapy
- c. **C-** Targeted Therapy

- d. **D**- Chemo + Targeted
- e. **E**- Other

7. First Reassessment Result (RECIST criteria)

- a. D – Deceased
- b. CR- Complete Response – Signifies that all target lesions have disappeared during the course of treatment.
- c. PR- Partial Response - Signifies that decreases of at least 30% have been noted in the lesion that has the largest diameter (LD)
- d. SD- Stable Disease – Signifies that there has been no significant decrease or increase in the size of target lesions
- e. PD- Progressive Disease – Signifies that there has been an increase of at least 20% in the sum of the LD of targeted lesions

8. Date Deceased or Last Contact

- a. Last contact verifying the whether the patient was alive or deceased. Could be a phone call, email or doctor's visit.
- b. Where doubt, obituary records may be searched
- c. Patients will not be contacted for this information for this study

Appendix 5: Data Collection Sheet

Code # _____	TNBC	HR+	HER2+	PBC
--------------	------	-----	-------	-----

Data Collection Sheet

Recurrent Metastatic Breast Cancer Retrospective Study

1. Recurrence Date: _____
2. Description (circle one in each row)

a. Loco-regional	Distant	Both
b. Non-visceral	Visceral	Visceral Crisis
3. Receptors Retested? Yes / No

a. Why? (circle reason code)	A	B	C	D	E	F	Other
i. If other, why?	_____						

(If No, proceed to item 6)

- b. Date _____
- c. Result (circle)

i. ER-	ER+
ii. PR-	PR+
iii. HER2-	HER2+
- d. Location of biopsy? _____

4. Results Concordant? Yes No (If Yes, proceed to item 6)
5. First-Line Treatment Plan based on? Primary Metastatic

a. Why?	A	B	C	D	E	Other
i. Other	_____					
6. First Line Treatment? A B C D E Other?

a. Other	_____
----------	-------
7. First Reassessment Result: D CR PR SD PD
8. Deceased Yes No
9. Date Deceased or Last Contact

a. Date:	_____
----------	-------

Figure 31. Data Collection Sheet

Appendix 6: Data Breakdown Flowchart

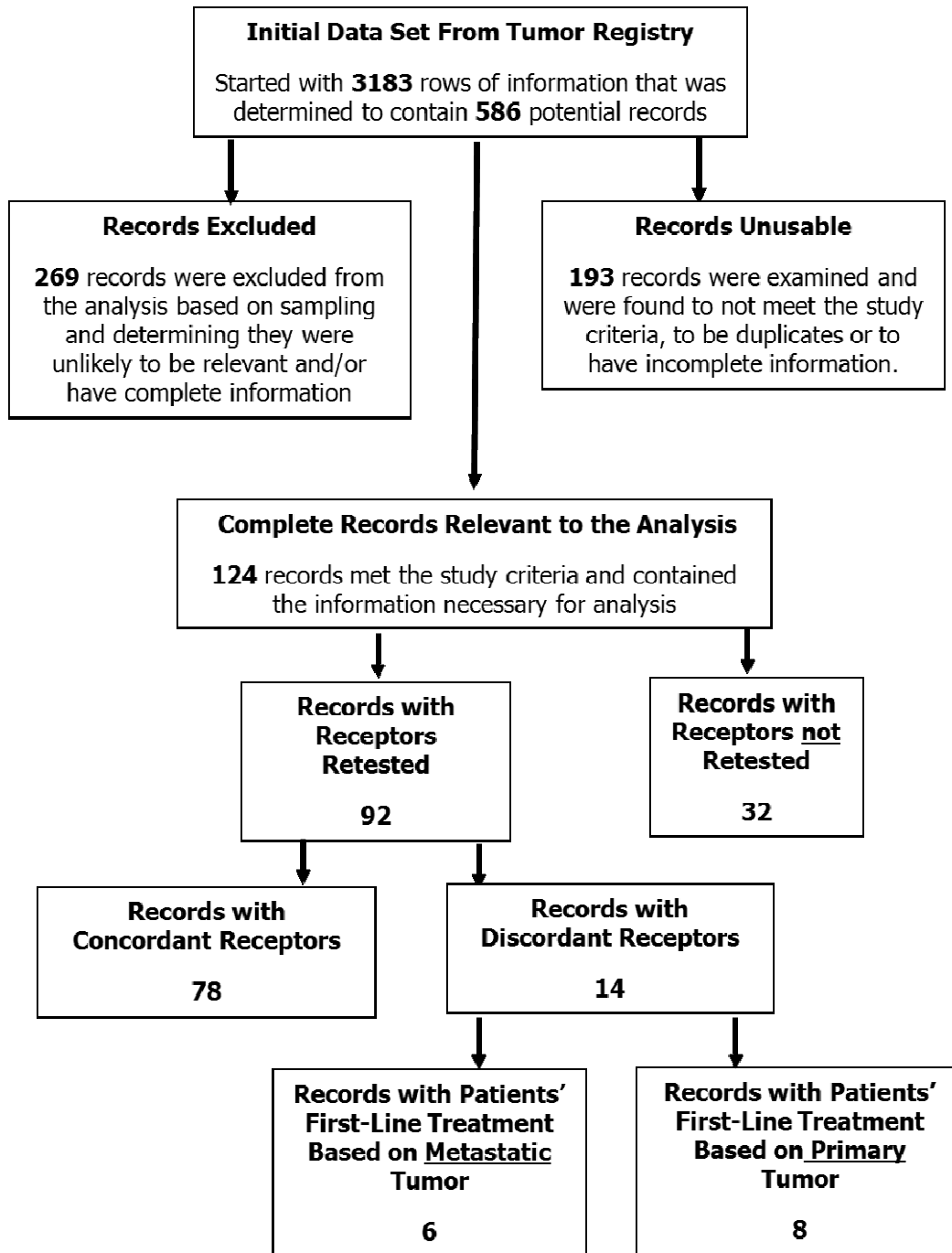


Figure 32. Data Breakdown Flowchart

Appendix 7: Variables Utilized

Table 24. List of Variables Utilized

Variable Name	Description	Source	Comment
PRS	Post Recurrence Survival – the length of time after the patient was diagnosed with a recurrence until death or last contact	Most of this data is from the Tumor Registry. The data was validated and updated if necessary during the data collection process	This is the main metric of the study. It is measured in days and will be graphically reported in years
Status	The final Status of the patient at last contact	Most of this data is from the Tumor Registry. The data was validated and updated if necessary during the data collection process	0= Deceased 1= Alive at last contact
FSD	Final Status Date – the date of last contact or death	Same as above	
Rdate	Recur Date – The date the metastatic recurrence was validated	Same as above	
Retested	Was the metastatic tumor(s) retested for the receptor status?	Found in the patient records. Assumed “No” if no evidence.	Yes/No
mPheno	Phenotype of the metastatic tumor when retested	Found in the patient records	Found in either a lab report or doctors’ notes or both.
pPheno	Phenotype of the primary tumor when retested	Found in the patient records and/or the tumor registry	Cross-checked if found in both sources

Table 24. List of Variables Utilized, Continued

Variable Name	Description	Source	Comment
Concordant	Where Retested = "Yes", do the primary and metastatic tumor receptor statuses match?	Found in the patient records.	Does mPheno match pPheno? Yes/No. No = Discordant
FLTP	First-line Treatment Plan. Only determined for discordant results.	Found in patient records	Coded per data collection sheet in Appendix 4
FLD	First-line Decision – Was the first-line treatment plan based on the metastatic or primary tumor. Only determined for discordant results.	Found in patient records.	Required comparing the plan to NCCN standard treatments by phenotype to the metastatic and primary phenotypes
LocoRegion	Is/are the metastatic tumor(s) local to the breast, distant or both?	Found in patient records	
Visceral	Based on the extent of the disease and the condition of the patient	Found in patient records	Judgement call by me
rLocation	Where in the body are metastatic tumors located?	Found in patient records	
RECIST	The patient status at their first post treatment evaluation, usually around 3 months after beginning treatment. Only determined for discordant results.	Found in doctors' notes in patient records	This required interpretation by me. The patient status using the RECIST terminology was not used in the doctors' notes.

Table 24. List of Variables Utilized, Continued

Variable Name	Description	Source	Comment
Age	Age at the time of recurrence. Calculated from age at primary diagnosis	Age at primary diagnosis was found in Tumor Registry in whole years.	
TTR	Time to recurrence. Calculated from primary diagnosis date and recurrence diagnosis date	Found in Tumor Registry, with recurrence date verified in patient records	Measured in days
pER	Estrogen receptor status of the primary tumor	Found in the Tumor Registry and patient records. Most data is from patient records. All data was verified by patient records	Used to determine phenotype and concordance. Not used in any analysis as an individual variable
pPR	Progesterone receptor status of the primary tumor	Found in the Tumor Registry and patient records. Most data was from patient records. All data was verified by patient records	Used to determine phenotype and concordance. Not used in any analysis as an individual variable
pHER2	HER2 Status of the primary tumor	Found in the Tumor Registry and patient records. Most data was from patient records. All data was verified by patient records	Used to determine phenotype and concordance. Not used in any analysis as an individual variable

Table 24. List of Variables Utilized, Continued

Variable Name	Description	Source	Comment
SSS	Documents the order of systematic therapy and surgery	Found in Tumor Registry. Not verified in patient records	Coded as Before, After, Both, None or Missing
Hormone	Did the patient receive hormone therapy for the <u>primary tumor</u>	Based on the review of several fields in the Tumor Registry	Yes, No or Missing
Chemo	Did the patient receive chemotherapy for the <u>primary tumor</u>	Based on the review of several fields in the Tumor Registry	Yes, No or Missing
Rad	Did the patient receive radiation therapy for the <u>primary tumor</u>	Based on the review of several fields in the Tumor Registry	Yes, No or Missing
Herceptin	Did the patient receive Herceptin therapy for the <u>primary tumor</u>	Based on the review of several fields in the Tumor Registry	Yes, No or Missing
Grade	Primary tumor grade	Tumor Registry coded field	Coded
Size	Primary tumor size	Tumor Registry field	
nPos	Number of positive lymph nodes associated with the primary tumor	Tumor Registry field	
AJCC Simple	AJCC cancer stage coding (pg. 47)("Breast Cancer," 2015)	Tumor Registry field	AJCC stage without the letter designation, utilizing just 0,1,2,3,4
Type Chemo	The type of chemotherapy provided for adjuvant therapy for the primary tumor	Created from a Tumor Registry text field	

Table 24. List of Variables Utilized, Continued

Variable Name	Description	Source	Comment
Age DX	Patient age at diagnosis of primary tumor	Tumor Registry field	Used to calculate other variables
dDate	Date of diagnosis of primary tumor	Tumor Registry field	Used to calculate other variables

Appendix 8: Information Only Variables

Table 25. Information Variables

Variable Name	Description	Source	Comment
Gen Stage	General staging of the primary disease	Tumor Registry field	Used to validate other variables
ID Code	Used to link patient ID to data records	Created by me	Patient ID kept in a separate file
Random	Used to link patient ID to data records	Created by me	Used to randomly assign codes so no pattern is present
SSF 15	HER2 status summary code	Tumor Registry field	Used to help validate administration of Herceptin for the primary diagnosis
SSF 16	Phenotype summary of the primary tumor	Tumor Registry field	Used to validate other variables
Rad Start	Radiation start date for primary tumor	Tumor Registry field	Used to validate other variables
Rad End	Radiation end date for primary tumor	Tumor Registry field	Used to validate other variables
Dose	Radiation dose for primary tumor	Tumor Registry field	Used to validate other variables
Chemo Text	Text field describing the chemotherapy treatment for the primary tumor	Tumor Registry field	Used to validate other variables

Table 25. Information Variables, Continued

Variable Name	Description	Source	Comment
Horm Start Date	Hormone treatment start date for primary tumor	Tumor Registry field	Used to validate other variables
RSN No Horm	Text field for reason hormone therapy was not performed	Tumor Registry field	Used to validate other variables
Recur Type	Type of recurrence by location	Tumor Registry field	Used to validate other variables
2nd Recur	Date of a 2 nd recurrence	Tumor Registry field	Used to validate other variables
2nd Type	Type of 2 nd recurrence	Tumor Registry field	Used to validate other variables
Order	Ties record to original dataset	Created by me	Used for quality control
Prim	Primary Surgeon of record	Tumor Registry field	Information purposed only
Mng	Managing doctor of record	Tumor Registry field	Information purposed only
Fol	Following doctor of record	Tumor Registry field	Information purposed only
Med	Oncologist of record	Tumor Registry field	Information purposed only
Complete	Used to indicate when the data collection on a given record is complete	Created by me	

Appendix 9: Discordant Patient Variables

ID CODE	AGE DX	pPheno	AGE at Recur	Rdate	mPheno	FLD	FLTP	Status	Visceral	PRS (Years)
6	53	PBC	55.5	9/9/2008	TNBC	M	Chemo	Deceased	Visceral	1.41
7	60	HR+	61.6	5/23/2003	TNBC	P	Endocrine	Deceased	Non	1.43
15	37	PBC	39.1	2/8/2005	HR+	P	Chemo + Targeted then Endocrine	Deceased	Non	4.44
99	61	HER2+	62.5	6/28/2010	TNBC	P	Therapy	Alive	Non	4.99
188	38	HER2+	41.4	2/21/2011	TNBC	M	Chemo	Alive	Visceral	4.17
233	46	HR+	49.2	7/19/2012	TNBC	P	Endocrine	Alive	Visceral	2.9
259	58	HR+	65.2	12/2/2005	PBC	P	Endocrine	Deceased	Visceral	0.88
312	35	HR+	44	1/12/2002	PBC	P	Chemo then Endocrine	Deceased	Visceral	4.03
355	56	HR+	56.2	10/23/2012	TNBC	P	Endocrine	Deceased	Non	0.88
399	59	HR+	64.7	1/4/2007	TNBC	P	Endocrine	Alive	Visceral	8.55
439	60	PBC	69.5	4/14/2011	HR+	M	Endocrine	Deceased	Visceral	0.66
535	59	PBC	61.2	10/4/2001	HER+	M	Chemo + Targeted	Deceased	Visceral	0.72
538	47	HR+	51.8	11/7/2014	TNBC	M	Chemo	Deceased	Visceral	0.32
558	53	HR+	56.9	1/5/2004	TNBC	M	No Endocrine	Deceased	Visceral	0.55

Table 26. Discordant Patients Variables

Appendix 10: Initial Cox Regression Results

Proportional Hazards Fit				
Whole Model				
Model	-LogLikelihood	ChiSquare	DF	Prob>ChiSq
Difference	65.9142	131.8285	52	<.0001*
Full	199.3863			
Reduced	265.3005			
Parameter Estimates				
Effect Likelihood Ratio Tests				
Source	Nparm	DF	ChiSquare	Prob>ChiSq
AGE DX	1	1	0.28838724	0.5913
GRADE	3	3	4.810042	0.1862
MORPHO	3	3	17.8530408	0.0005*
SIZE	1	1	11.5662712	0.0007*
nPos	1	1	2.01569236	0.1557
AJCC Simple	3	3	2.96606005	0.3969
Rad	2	2	8.93232289	0.0115*
Herceptin	2	2	8.72032953	0.0128*
Type Chemo	10	10	36.0345458	<.0001*
Chemo	3	1	4.60655692	0.0318*
Hormone	2	2	3.83801958	0.1468
SSS	4	4	8.72764314	0.0683
RECUR TYPE	12	12	20.5088434	0.0581
pPheno	3	3	20.8289353	0.0001*
Retested	1	1	13.6773311	0.0002*
Visceral	3	3	6.7372227	0.0808

Figure 33. Initial Complete Cox Regression

Appendix 11: Power Tables

Coding descriptions for the power tables:

- Lambda
 - L3.5 for Lambda = 3.5
 - L2.0 for Lambda = 2.0
 - L5.0 for Lambda = 5.0
- Binning
 - C = Continuous
 - B10 = 10 bins
 - B5 = 5 bins
 - B2 = 2 bins
- Alpha (Type 1 error)
 - .05
 - .01
- Censoring
 - C0 = no censoring
 - C25 = 25% censoring
 - C50 = 50% censoring
- Sample Size
 - $n_1 = n_2$ unless otherwise noted
 - N70/30 = 70%, 30% split of total N
 - N85/15 = 85%, 15% split of total N
- Hazard Ratios
 - 1 = both populations equal, thus type 1 error rate
 - 1.25, 1.5, 1.75, 2.0 and 3.0

Table 27. L3.5 C0 C .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0067	0.0116	0.0249	0.0467	0.0736	0.234
20	0.006	0.0179	0.0542	0.122	0.2256	0.6535
50	0.0088	0.0448	0.1994	0.4516	0.6989	0.9933
100	0.007	0.1021	0.4804	0.8282	0.9702	1
200	0.0086	0.2582	0.854	0.9929	1	1
500	0.0069	0.7015	0.9997	1	1	1
1000	0.0117	0.9693	1	1	1	1

Table 28. L3.5 C0 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0342	0.047	0.0882	0.1408	0.2016	0.4819
20	0.0349	0.0744	0.1599	0.2992	0.4566	0.8585
50	0.0437	0.1439	0.4075	0.6893	0.8752	0.9994
100	0.0394	0.2584	0.7111	0.9447	0.9946	1
200	0.0434	0.4881	0.952	0.9989	1	1
500	0.0436	0.8761	0.9999	1	1	1
1000	0.0511	0.9931	1	1	1	1

Table 29. L3.5 C0 B10 .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0055	0.0096	0.0213	0.0387	0.0615	0.207
20	0.0055	0.0173	0.0509	0.1148	0.2165	0.6346
50	0.0081	0.0435	0.1941	0.441	0.6876	0.9926
100	0.007	0.0993	0.4746	0.8231	0.9689	1
200	0.0083	0.2538	0.8521	0.9925	1	1
500	0.0068	0.6997	0.9997	1	1	1
1000	0.0118	0.9681	1	1	1	1

Table 30. L3.5 C0 B10 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0315	0.0429	0.0808	0.131	0.1918	0.4611
20	0.0334	0.0697	0.1498	0.2899	0.4401	0.8472
50	0.0433	0.1415	0.4035	0.6843	0.8708	0.9994
100	0.0383	0.2559	0.7075	0.943	0.9938	1
200	0.0435	0.4829	0.9503	0.9988	1	1
500	0.0431	0.8735	0.9999	1	1	1
1000	0.0515	0.9931	1	1	1	1

Table 31. L3.5 C0 B5 .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0064	0.008	0.0176	0.0362	0.0545	0.1788
20	0.0053	0.0137	0.0471	0.1102	0.2004	0.5981
50	0.0065	0.0414	0.1822	0.4166	0.6609	0.9911
100	0.0075	0.1044	0.4652	0.8211	0.9651	1
200	0.0095	0.2542	0.8431	0.9932	0.9997	1
500	0.0097	0.708	0.9996	1	1	1
1000	0.0085	0.9663	1	1	1	1

Table 32. L3.5 C0 B5 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.028	0.0409	0.0731	0.1255	0.182	0.4404
20	0.0331	0.0607	0.1526	0.284	0.4262	0.8272
50	0.036	0.1348	0.3917	0.6644	0.8533	0.9995
100	0.041	0.2623	0.702	0.9383	0.9926	1
200	0.0441	0.4854	0.9494	0.9989	0.9999	1
500	0.0462	0.8738	1	1	1	1
1000	0.0438	0.9928	1	1	1	1

Table 33. L3.5 C0 B2 .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.002	0.003	0.0061	0.0135	0.0183	0.0457
20	0.0042	0.0115	0.0342	0.0785	0.135	0.3964
50	0.0053	0.0346	0.1564	0.3537	0.5768	0.9663
100	0.0066	0.084	0.4038	0.7585	0.9398	0.9997
200	0.008	0.2243	0.8015	0.986	0.9994	1
500	0.0075	0.6644	0.9989	1	1	1
1000	0.0087	0.957	1	1	1	1

Table 34. L3.5 C0 B2 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.018	0.028	0.0489	0.083	0.1141	0.2697
20	0.0253	0.0513	0.1182	0.2218	0.3333	0.689
50	0.035	0.1145	0.3502	0.6027	0.8041	0.9947
100	0.0367	0.2317	0.6534	0.9081	0.9842	1
200	0.0401	0.4556	0.9289	0.9977	1	1
500	0.0436	0.8499	0.9999	1	1	1
1000	0.0469	0.9909	1	1	1	1

Table 35. L2.0 C0 C .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0072	0.013	0.0283	0.049	0.0779	0.2406
20	0.0077	0.021	0.0645	0.14	0.2358	0.6713
50	0.0087	0.0558	0.2339	0.5107	0.7465	0.9958
100	0.0067	0.1346	0.5486	0.8876	0.9814	1
200	0.0089	0.3163	0.9012	0.9981	1	1
500	0.01	0.7845	1	1	1	1
1000	0.0097	0.9851	1	1	1	1

Table 36. L2.0 C0 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0364	0.0549	0.0952	0.1556	0.224	0.5013
20	0.036	0.0804	0.1829	0.3243	0.4825	0.8694
50	0.0408	0.1655	0.4552	0.7447	0.9063	0.999
100	0.0415	0.3103	0.7702	0.9693	0.9965	1
200	0.0471	0.5577	0.971	0.9998	1	1
500	0.0504	0.9217	1	1	1	1
1000	0.0465	0.9972	1	1	1	1

Table 37. L5.0 C0 C .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0051	0.0107	0.023	0.0423	0.069	0.2288
20	0.0066	0.0159	0.0499	0.1063	0.1922	0.6151
50	0.0058	0.0399	0.1634	0.3818	0.6166	0.9869
100	0.0073	0.0867	0.4059	0.7686	0.9438	1
200	0.009	0.2175	0.7875	0.9865	0.9998	1
500	0.0096	0.6188	0.9982	1	1	1
1000	0.01	0.9354	1	1	1	1

Table 38. L5.0 C0 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0287	0.0452	0.0786	0.1301	0.1913	0.4714
20	0.0333	0.0634	0.1464	0.2635	0.3946	0.8225
50	0.0387	0.1239	0.3479	0.6206	0.8232	0.9979
100	0.0376	0.2221	0.644	0.9069	0.9874	1
200	0.0468	0.4232	0.9242	0.997	1	1
500	0.0456	0.8189	0.9999	1	1	1
1000	0.0475	0.981	1	1	1	1

Table 39. L3.5 C25 C .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0067	0.0078	0.0162	0.0338	0.0516	0.1437
20	0.0063	0.0151	0.0387	0.0871	0.1492	0.4595
50	0.0081	0.0329	0.1409	0.3163	0.514	0.9459
100	0.0065	0.0779	0.3364	0.6755	0.893	0.9998
200	0.0077	0.1781	0.7157	0.9625	0.998	1
500	0.0088	0.5412	0.9936	1	1	1
1000	0.0107	0.8896	1	1	1	1

Table 40. L3.5 C25 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0282	0.0388	0.0656	0.1069	0.1521	0.3431
20	0.0313	0.0624	0.1238	0.2261	0.3323	0.7033
50	0.0405	0.1115	0.3085	0.5509	0.7452	0.9897
100	0.0404	0.2048	0.5733	0.8605	0.9667	1
200	0.041	0.3816	0.8762	0.991	0.9999	1
500	0.0454	0.7639	0.999	1	1	1
1000	0.0469	0.9651	1	1	1	1

Table 41. L3.5 C25 B10 .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0053	0.0061	0.0125	0.0264	0.0415	0.1113
20	0.0056	0.0133	0.0349	0.0766	0.1274	0.402
50	0.0078	0.0297	0.1293	0.2869	0.4769	0.9226
100	0.0071	0.0729	0.3158	0.6415	0.8693	0.9996
200	0.0077	0.1676	0.6864	0.95	0.9967	1
500	0.0083	0.5156	0.9896	1	1	1
1000	0.0104	0.8722	1	1	1	1

Table 42. L3.5 C25 B10 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0262	0.0338	0.0584	0.0939	0.1356	0.2978
20	0.0287	0.0566	0.115	0.2096	0.3111	0.655
50	0.0391	0.1071	0.2884	0.5243	0.7163	0.9809
100	0.0411	0.1956	0.5506	0.8398	0.9578	0.9999
200	0.0396	0.3636	0.8603	0.9871	0.9996	1
500	0.0459	0.7451	0.9983	1	1	1
1000	0.0475	0.9579	1	1	1	1

Table 43. L3.5 C50 C .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0047	0.0084	0.0122	0.0193	0.0283	0.0737
20	0.006	0.0106	0.0258	0.0526	0.0823	0.2521
50	0.0075	0.0237	0.081	0.1744	0.3146	0.7703
100	0.0071	0.0473	0.1935	0.4579	0.6767	0.9879
200	0.0086	0.1112	0.4814	0.8258	0.9645	1
500	0.0091	0.3512	0.9328	0.9995	1	1
1000	0.0078	0.6991	0.9993	1	1	1

Table 44. L3.5 C50 C .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0277	0.0361	0.0512	0.0738	0.0966	0.2133
20	0.029	0.0482	0.0872	0.1497	0.2081	0.4711
50	0.0348	0.0856	0.209	0.3738	0.539	0.9059
100	0.0396	0.1429	0.4006	0.6915	0.8648	0.9981
200	0.0448	0.269	0.7135	0.9439	0.9926	1
500	0.0444	0.5867	0.9801	1	1	1
1000	0.0438	0.8722	0.9999	1	1	1

Table 45. L3.5 C50 B10 .01

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0024	0.0048	0.0078	0.0113	0.0191	0.0505
20	0.0044	0.0082	0.0192	0.0426	0.0654	0.1991
50	0.0061	0.0213	0.0697	0.1483	0.2687	0.6885
100	0.007	0.0412	0.1711	0.407	0.6147	0.9725
200	0.0083	0.0978	0.4331	0.7786	0.9443	1
500	0.0091	0.3176	0.9044	0.9974	1	1
1000	0.0075	0.6525	0.9985	1	1	1

Table 46. L3.5 C50 B10 .05

	Hazard Ratio					
n1=n2	1	1.25	1.5	1.75	2	3
10	0.0187	0.0272	0.0393	0.0575	0.0792	0.1637
20	0.0255	0.0411	0.0767	0.1304	0.1812	0.4133
50	0.0336	0.0786	0.1884	0.3398	0.4903	0.8686
100	0.0367	0.1318	0.3665	0.6438	0.8232	0.9936
200	0.0436	0.2477	0.67	0.9203	0.9856	1
500	0.0435	0.5521	0.9706	0.9998	1	1
1000	0.0435	0.8408	0.9999	1	1	1

Table 47. L3.5 C0 C N70/30 .01

	Hazard Ratio					
n1,n2	1	1.25	1.5	1.75	2	3
6, 14	0.0088	0.0093	0.0146	0.0267	0.0392	0.1273
12, 28	0.0062	0.012	0.0379	0.0762	0.1347	0.4558
30, 70	0.0085	0.0308	0.1257	0.2981	0.5105	0.9483
60, 140	0.0074	0.0649	0.3289	0.6795	0.8867	1
120, 280	0.0079	0.1777	0.7042	0.9632	0.9982	1
300, 700	0.0096	0.5342	0.991	1	1	1
600, 1400	0.0103	0.8946	1	1	1	1

Table 48. L3.5 C0 C N70/30 .05

	Hazard Ratio					
n1,n2	1	1.25	1.5	1.75	2	3
6, 14	0.0338	0.0399	0.0628	0.0979	0.1394	0.3377
12, 28	0.0366	0.0547	0.1227	0.2108	0.323	0.713
30, 70	0.0405	0.1098	0.3051	0.541	0.7449	0.9889
60, 140	0.0401	0.1983	0.5684	0.8608	0.9653	1
120, 280	0.0402	0.379	0.878	0.9923	0.9999	1
300, 700	0.0455	0.763	0.999	1	1	1
600, 1400	0.0467	0.969	1	1	1	1

Table 49. L3.5 C0 C N85/15 .01

	Hazard Ratio					
n1,n2	1	1.25	1.5	1.75	2	3
3, 17	0.0134	0.0031	0.0029	0.0009	0.001	0.0004
6, 34	0.0098	0.0074	0.0129	0.0255	0.0483	0.1861
15, 85	0.0078	0.0132	0.0544	0.1332	0.249	0.7401
30, 170	0.0088	0.0322	0.1535	0.3732	0.6184	0.9848
60, 340	0.0084	0.0829	0.4055	0.7743	0.9474	1
150, 850	0.0076	0.3005	0.9001	0.998	0.9999	1
300, 1700	0.0094	0.6544	0.9985	1	1	1

Table 50. L3.5 C0 C N85/15 .05

	Hazard Ratio					
n1,n2	1	1.25	1.5	1.75	2	3
3, 17	0.035	0.031	0.0358	0.0486	0.0641	0.1487
6, 34	0.0377	0.0387	0.0628	0.1097	0.1666	0.4378
15, 85	0.0374	0.0622	0.1773	0.33	0.5009	0.9019
30, 170	0.0439	0.1194	0.3555	0.6332	0.825	0.9972
60, 340	0.0392	0.231	0.6571	0.9212	0.9858	1
150, 850	0.0423	0.5445	0.9714	0.9998	1	1
300, 1700	0.0479	0.843	0.9999	1	1	1

Appendix 12: Simulation Definitions

Key words and concepts utilized in the simulation analysis are defined below:

- Censoring – A statistical term that describes that fact that not all subjects in a survival analysis can/are observed until the event (death) occurs. The subject may be alive at the end of the study or may be lost to observation for various reasons (Lee & Wang, 2013, pp. 2-5).
- Kaplan-Meier (KM) – A method of survival analysis created by Kaplan and Meier in 1958 that utilizes survival data that has censoring (Lee & Wang, 2013, pp. 68-92).
- Hazard Rate (Failure Rate) – The rate of failure (death) per a given unit of time (p. 10).
- Exponential Distribution – A common statistical distribution used in survival analysis with a constant failure rate (pp. 133–138).
- Lambda (λ) – The parameter of an exponential distribution where the mean time to failure (death) is λ .
- Survival Curve – A graphical curve with percent survival on the Y-axis and time on the x-axis showing the percent of subjects still alive at a given point in time (pp. 8-11).
- Log Rank – A semi-parametric statistical test used to evaluate the difference between two or more KM survival curves (pp. 112–113).
- Hazard Ratio (HR) – The ratio of two hazard functions. The ratio indicates the difference in the instantaneous risk of death between two groups. A HR of 2 would mean that one group has twice the risk of dying at any point in time than another group (pp. 282–284).

- Proportional Hazards – For the purposes of this dissertation, all analyses assume proportional hazards, that is, the HR is constant over time (pp. 282–284).
- Sample Proportions – The proportion of the total sample $N = n_1 + n_2$, that are in each group in an analysis (only two groups were considered in this analysis). For $n_1 = n_2$, the proportions are 50% and 50%.
- Delta – The hypothesized difference in the metric of interest between experimental groups that is of interest to the investigator. In the case of survival analysis the hazard ratio (HR) was used. For example, the experimenter may wish to statistically detect a difference between two groups of $HR = 1.4$ (Sharpe, Veaux, & Vellman, 2010, pp. 289-301).
- False Positive (Type 1 Error) – The probability that a statistical test will indicate a difference between groups (two in this case) on the metric of interest when in fact there is no difference. A priori, this risk is set by the experimenter and is commonly called alpha. A risk of .05 or 5% is a common risk taken by practitioners (Sharpe et al., 2010, pp. 289-301).
- P-value – After a statistical analysis is performed, this is the a posteriori probability of getting the observed result assuming the null hypothesis is true. Commonly reported to allow each person assessing a statistical analysis to apply their own type 1 risk (pp. 289-301).
- False Negative (Type 2 Error) – The probability that a statistical test will indicate there is not a difference between two groups when in fact there is. This risk is called beta. A priori, beta can be controlled by the experimenter and is impacted by the Delta chosen. A typical type 2 risk is 10% (pp. 289–301).

Power – Mathematically this number is $1 - \beta$. The interpretation is that power is the probability of discovering the desired Delta given the parameters of the experiment (pp. 289-301).

Appendix 13: Table of MATLAB Programs

Table 51. MATLAB Programs Descriptions

Program	Description
Generate_Data_v3_2p1p2	Exponential random variables generated for two populations of various sample sizes and hazard ratios for a given lambda, censor rate and sample size ratio. Lambda, censor rate and sample size ratio are manually changed for each run. Bin size can be changed as well. The code is found in Appendix 14.
Generate_Data_v3_2Bin	Bin size can be changed as well. Appendix 15.
Compute_log_rank_v2	This code calls a MATLAB M-file (Log Rank) routine to calculate the p-value for the log rank test comparing two groups. The p-values are stored in output files for each sample size/hazard ratio combination. Appendix 16.
Power_from_pval	An output table is create for $\alpha = .01$ and $\alpha = .05$ from the p-values for each combination of sample size and hazard ratio. See Appendix 17.
Compute_ndeaths	For each run of the simulation and each combination of variables, the number of actual "deaths" in the simulation is recorded. This data is used to correlate with various input variables. See Appendix 18.
ndeaths_vs_pval	This code combines the "deaths" with p-values and other input variables for analysis. See Appendix 19.
kmplot	This is a MATLAB provided file that plots KM curves. See Appendix 20 for the code.
logrank	This is a MATLAB provided file that calculates Log Rank statistics for comparing two survival curves. See Appendix 21 for the code description.

Appendix 14: MATLAB Program 1

```
% Copyright 2015, Tyler J. Massaro, all rights reserved.
% Date: 9 June 2015
%
% Generating exponential random survival data.
clc;
clear all;
close all;

% Allocate storage
N = [10 20 50 100 200 500 1000]; no_N = length(N);
p1=.30; p2=.70;% Create unequal sample size 70/30 or 85/15 split
p = 0.05;
lamda = 3.5; %similar to p = .05
censor_rate = 0.00;
delta = [1 1.25 1.5 1.75 2 3]; no_delta = length(delta);
ntrials = 10000;

% Create baseline survival data (exponential)
for i = 1:no_N

    NN = 2*p1*N(i);
    SURV = exprnd(lamda, NN, ntrials);%changed from 3 to lamda
    X = (12*SURV);
    binSURV = X;
    binSURV(X<=6) = 6;
    binSURV(X>6 & X<=12)= 12;
    binSURV(X>12 & X<=18)= 18;
    binSURV(X>18 & X<=24)= 24;
    binSURV(X>24 & X<=30)= 30;
    binSURV(X>30 & X<=36)= 36;
    binSURV(X>36 & X<=42)= 42;
    binSURV(X>42 & X<=48)= 48;
    binSURV(X>48 & X<=54)= 54;
    binSURV(X>54 )= 60;

    Y = rand(NN, ntrials);

    % Fixing the CENS data
    CENS = double(Y < censor_rate);
    CENS(SURV > 5) = 1;

    % Store information
```

```

NN = N(i);
csvwrite(['EXP_SURV_N_is_' num2str(NN) '.csv'], SURV);
csvwrite(['EXP_BIN_SURV_N_is_' num2str(NN) '.csv'], binSURV);
csvwrite(['EXP_CENS_N_is_' num2str(NN) '.csv'], CENS);

assignin('base', ['EXP_SURV_N_is_' num2str(NN)], SURV);
assignin('base', ['EXP_BIN_SURV_N_is_' num2str(NN)], binSURV);
assignin('base', ['EXP_CENS_N_is_' num2str(NN)], CENS);
end

% Create new data (exponential)
for i = 1:no_N

    NN = 2*p2*N(i);

    for d = 1:no_delta
        dd = delta(d);
        SURV = exprnd(lamda/dd, NN, ntrials);%changed from 3 to lamda
        X = (12*SURV);
        binSURV = X;
        binSURV(X<=6) = 6;
        binSURV(X>6 & X<=12)= 12;
        binSURV(X>12 & X<=18)= 18;
        binSURV(X>18 & X<=24)= 24;
        binSURV(X>24 & X<=30)= 30;
        binSURV(X>30 & X<=36)= 36;
        binSURV(X>36 & X<=42)= 42;
        binSURV(X>42 & X<=48)= 48;
        binSURV(X>48 & X<=54)= 54;
        binSURV(X>54 )= 60;
        Y = rand(NN, ntrials);

        % Fixing the CENS data
        CENS = double(Y < censor_rate);
        CENS(SURV > 5) = 1;

        % Store information
        NN = N(i);
        csvwrite(['EXP_SURV_N_is_' num2str(NN) ...
            '_d_is_' num2str(100*dd) '.csv'], SURV);
        csvwrite(['EXP_BIN_SURV_N_is_' num2str(NN) ...
            '_d_is_' num2str(100*dd) '.csv'], binSURV);
        csvwrite(['EXP_CENS_N_is_' num2str(NN) ...
            '_d_is_' num2str(100*dd) '.csv'], CENS);

        assignin('base', ['EXP_SURV_N_is_' num2str(NN) ...
            '_d_is_' num2str(100*dd)], SURV);
        assignin('base', ['EXP_BIN_SURV_N_is_' num2str(NN) ...

```

```

        '_d_is_' num2str(100*dd)], binSURV);
    assignin('base', ['EXP_CENS_N_is_' num2str(NN) ...
        '_d_is_' num2str(100*dd)], CENS);
end
end

```

Appendix 15: MATLAB Program 2

```

% Copyright 2015, Tyler J. Massaro, all rights reserved.
% Date: 9 June 2015
%
% Generating exponential random survival data.
clc;
clear all;
close all;

% Allocate storage
N = [10 20 50 100 200 500 1000]; no_N = length(N);
p1=.15; p2=.85;% Create unequal sample size 70/30 split
p = 0.05;
lamda = 3.5; %similar to p = .05
censor_rate = 0.00;
delta = [1 1.25 1.5 1.75 2 3]; no_delta = length(delta);
ntrials = 10000;

% Create baseline survival data (exponential)
for i = 1:no_N

    NN = 2*p1*N(i);
    SURV = exprnd(lamda, NN, ntrials);%changed from 3 to lamda
    X = (12*SURV);
    binSURV = X;
    % binSURV(X<=6) = 6;
    %binSURV(X>6 & X<=12)= 12;
    % binSURV(X<=12)= 12
    % binSURV(X>12 & X<=18)= 18;
    % binSURV(X>12 & X<=24)= 24;
    % binSURV(X>24 & X<=30)= 30;
    %binSURV(X>24 & X<=36)= 36;
    binSURV(X<=36)= 36;
    % binSURV(X>36 & X<=42)= 42;
    % binSURV(X>36 & X<=48)= 48;
    % binSURV(X>48 & X<=54)= 54;
    binSURV(X>36 )= 60;

```

```

Y = rand(NN, ntrials);

% Fixing the CENS data
CENS = double(Y < censor_rate);
CENS(SURV > 5) = 1;

% Store information
NN = N(i);
csvwrite(['EXP_SURV_N_is_' num2str(NN) '.csv'], SURV);
csvwrite(['EXP_BIN_SURV_N_is_' num2str(NN) '.csv'], binSURV);
csvwrite(['EXP_CENS_N_is_' num2str(NN) '.csv'], CENS);

assignin('base', ['EXP_SURV_N_is_' num2str(NN)], SURV);
assignin('base', ['EXP_BIN_SURV_N_is_' num2str(NN)], binSURV);
assignin('base', ['EXP_CENS_N_is_' num2str(NN)], CENS);
end

% Create new data (exponential)
for i = 1:no_N

    NN = 2*p2*N(i);

    for d = 1:no_delta
        dd = delta(d);
        SURV = exprnd(lamda/dd, NN, ntrials); %changed from 3 to lamda
        X = (12*SURV);
        binSURV = X;
        % binSURV(X<=6) = 6;
        % binSURV(X>6 & X<=12)= 12;
        % binSURV(X<=12)= 12
        % binSURV(X>12 & X<=18)= 18;
        % binSURV(X>12 & X<=24)= 24;
        % binSURV(X>24 & X<=30)= 30;
        % binSURV(X>24 & X<=36)= 36;
        binSURV(X<=36)= 36;
        % binSURV(X>36 & X<=42)= 42;
        % binSURV(X>36 & X<=48)= 48;
        % binSURV(X>48 & X<=54)= 54;
        binSURV(X>36)= 60;
        Y = rand(NN, ntrials);

        % Fixing the CENS data
        CENS = double(Y < censor_rate);
        CENS(SURV > 5) = 1;

        % Store information
        NN = N(i);

```

```

csvwrite(['EXP_SURV_N_is_' num2str(NN) ...
'_d_is_' num2str(100*dd) '.csv'], SURV);
csvwrite(['EXP_BIN_SURV_N_is_' num2str(NN) ...
'_d_is_' num2str(100*dd) '.csv'], binSURV);
csvwrite(['EXP_CENS_N_is_' num2str(NN) ...
'_d_is_' num2str(100*dd) '.csv'], CENS);

assignin('base', ['EXP_SURV_N_is_' num2str(NN) ...
'_d_is_' num2str(100*dd)], SURV);
assignin('base', ['EXP_BIN_SURV_N_is_' num2str(NN) ...
'_d_is_' num2str(100*dd)], binSURV);
assignin('base', ['EXP_CENS_N_is_' num2str(NN) ...
'_d_is_' num2str(100*dd)], CENS);
end
end

```

Appendix 16: MATLAB Program 3

```

% Copyright 2015, Tyler J. Massaro, all rights reserved.
% Date: 9 June 2015
% Compute_log_rank_v2

% Compute log-rank for binned and for continuous survival data. Also,
% storing p-values for later use.
clc;
clear all;
close all;

alpha = 0.05;
N = [ 10 20 50 100 200 500 1000]; no_N = length(N);
delta = [1 1.25 1.5 1.75 2 3]; no_delta = length(delta);
ntrials = 10000;
results_table = zeros(no_N, no_delta);
results_table_bin = zeros(no_N, no_delta);

for n = 1:no_N
    NN = N(n);

    Sbase = csvread(['EXP_SURV_N_is_' num2str(NN) '.csv']);
    Cbase = csvread(['EXP_CENS_N_is_' num2str(NN) '.csv']);

    Sbinbase = csvread(['EXP_BIN_SURV_N_is_' num2str(NN) '.csv']);

    for d = 1:no_delta
        dd = delta(d);

```

```

Sfname = ['EXP_SURV_N_is_' num2str(NN) '_d_is_' ...
          num2str(100*dd) '.csv'];
Cfname = ['EXP_CENS_N_is_' num2str(NN) '_d_is_' ...
          num2str(100*dd) '.csv'];

Sbinfname = ['EXP_BIN_SURV_N_is_' num2str(NN) '_d_is_' ...
             num2str(100*dd) '.csv'];

S = csvread(Sfname);
C = csvread(Cfname);
Sbin = csvread(Sbinfname);

nreject = 0;
nreject_bin = 0;

pvals = zeros(ntrials, 1);
pvals_bin = zeros(ntrials, 1);

for i = 1:ntrials
    X = [Sbase(:, i) Cbase(:, i)];
    Y = [S(:, i) C(:, i)];

    Xbin = [Sbinbase(:, i) Cbase(:, i)];
    Ybin = [Sbin(:, i) C(:, i)];

    pval = logrank(X, Y);
    pval_bin = logrank(Xbin, Ybin);

    pvals(i) = pval;
    pvals_bin(i) = pval_bin;

    if pval < alpha
        nreject = nreject + 1;
    end

    if pval_bin < alpha
        nreject_bin = nreject_bin + 1;
    end
end

pwr = nreject/ntrials;
results_table(n, d) = pwr;

pwr_bin = nreject_bin/ntrials;
results_table_bin(n, d) = pwr_bin;

csvwrite(['EXP_p_vals_N_is_' num2str(NN) '...

```

```

        '_d_is_' num2str(100*dd) '.csv'], pvals);
    csvwrite(['EXP_BIN_p_vals_N_is_' num2str(NN) ...
        '_d_is_' num2str(100*dd) '.csv'], pvals_bin);
end
end

```

Appendix 17: MATLAB Program 4

```

% Copyright 2015, Tyler J. Massaro, all rights reserved.
% Date: 12 June 2015
%
% Powers from p-values
clc;
clear all;
close all;

alpha = [0.01 0.05];
N = [10 20 50 100 200 500 1000]; no_N = length(N);
delta = [1 1.25 1.5 1.75 2 3]; no_delta = length(delta);
ntrials = 10000;

for i = 1:length(alpha)
    a = alpha(i);

    results_table = zeros(no_N, no_delta);
    results_table_bin = zeros(no_N, no_delta);

    for j = 1:no_N
        NN = N(j);

        for d = 1:no_delta
            dd = delta(d);

            Pbin = csvread(['EXP_BIN_p_vals_N_is_' num2str(NN) ...
                '_d_is_' num2str(100*dd) '.csv']);
            P = csvread(['EXP_p_vals_N_is_' num2str(NN) ...
                '_d_is_' num2str(100*dd) '.csv']);

            Xbin = Pbin < a;
            X = P < a;

            results_table(j, d) = sum(X)/ntrials;
            results_table_bin(j, d) = sum(Xbin)/ntrials;
        end
    end
end
end

```

```

    assignin('base', ['CONT_alpha_' num2str(100*a)], results_table);
    assignin('base', ['BIN_alpha_' num2str(100*a)], results_table_bin);
end

```

Appendix 18: MATLAB Program 5

```

% Copyright 2015, Tyler J. Massaro, all rights reserved
% Date: 27 July 2015
% Compute_ndeaths
% Compute the number of deaths based on censored data
clc;
clear all;
close all;

N = [10 20 50 100 200 500 1000]; no_N = length(N);
delta = [1 1.25 1.5 1.75 2 3]; no_delta = length(delta);

for i = 1:no_N
    n = N(i);

    for j = 1:no_delta
        d = delta(j);

        X = csvread(['EXP_CENS_N_is_' num2str(n) '_d_is_' ...
            num2str(100*d) '.csv']);

        ndeaths = n - sum(X);

        csvwrite(['ndeaths_N_is_' num2str(n) '_d_is_' ...
            num2str(100*d) '.csv'], ndeaths);
    end
end

```

Appendix 19: MATLAB Program 6

```

% Copyright 2015, Tyler J. Massaro, all rights reserved.
% Date: 27 July 2015
%
% Number of deaths vs p-values
clc;

```

```

clear all;
close all;

alpha = [0.01 0.05];
N = [10 20 50 100 200 500 1000]; no_N = length(N);
delta = [1 1.25 1.5 1.75 2 3]; no_delta = length(delta);

bin_corr_table = zeros(no_N, no_delta);
cont_corr_table = zeros(no_N, no_delta);

% Compute the correlation coefficient between p-val and ndeaths
for j = 1:no_N
    NN = N(j);

    for d = 1:no_delta
        dd = delta(d);

        Pbin = csvread(['EXP_BIN_p_vals_N_is_' num2str(NN) ...
            '_d_is_' num2str(100*dd) '.csv']);
        P = csvread(['EXP_p_vals_N_is_' num2str(NN) ...
            '_d_is_' num2str(100*dd) '.csv']);

        ndeaths = csvread(['ndeaths_N_is_' num2str(NN) '_d_is_' ...
            num2str(100*dd) '.csv']);

        rho_bin = corrcoef(Pbin, ndeaths);
        rho_cont = corrcoef(P, ndeaths);

        bin_corr_table(j, d) = rho_bin(1, 2);
        cont_corr_table(j, d) = rho_cont(1, 2);
    end
end

% Compute the variation
ndeaths_avg = zeros(no_N, no_delta);
ndeaths_var = zeros(no_N, no_delta);
for j = 1:no_N
    NN = N(j);

    for d = 1:no_delta
        dd = delta(d);

        ndeaths = csvread(['ndeaths_N_is_' num2str(NN) '_d_is_' ...
            num2str(100*dd) '.csv']);

        ndeaths_avg(j, d) = mean(ndeaths);
    end
end

```

```

        ndeaths_var(j, d) = var(ndeaths);
    end
end

% Compute the variation by normalizing wrt N
ndeaths_avg_norm = zeros(no_N, no_delta);
ndeaths_var_norm = zeros(no_N, no_delta);
for j = 1:no_N
    NN = N(j);

    for d = 1:no_delta
        dd = delta(d);

        ndeaths = csvread(['ndeaths_N_is_' num2str(NN) '_d_is_' ...
            num2str(100*dd) '.csv']);

        ndeaths_avg_norm(j, d) = mean(ndeaths/NN);
        ndeaths_var_norm(j, d) = var(ndeaths/NN);
    end
end

```

Appendix 20: MATLAB Program 7

```

function varargout=kmplot(varargin)
% KMPLOT Plot the Kaplan-Meier estimation of the survival function
% Survival times are data that measure follow-up time from a defined
% starting point to the occurrence of a given event, for example the %timefrom the beginning to
the end of a remission period or the time %from the diagnosis of a disease to death. Standard
statistical %techniques cannot usually be applied because the underlying %distribution is rarely
Normal
% and the data are often "censored". A survival time is described as
% censored when there is a follow-up time but the event has not yet
% occurred or is not known to have occurred. For example, if remission %time is being studied
and the patient is still in remission at the end %of the study, then that patient's remission
time would be censored. %If a patient for some reason drops out of a study before the end of
%the study period, then that patient's follow-up time would also be %considered to be
censored. The survival function S(t) is defined as %the probability of surviving at least to time t.
The graph of S(t) %against t is called the survival curve. The Kaplan's %Meier method can % be
used to estimate this curve from the observed survival times % without the assumption of an
underlying probability distribution.
%
% Syntax:  kmplot(x,alpha,censflag)
%
% Inputs:
%         X (mandatory)- Nx2 data matrix:

```

```

%           (X(:,1) = survival time of the i-th subject
%           (X(:,2) = censored flag
%           (0 if not censored; 1 if censored)
%       note that if X is a vector, all the flags of the second column
%       will be set to 0 (all data are not censored).
%       ALPHA (optional) - significance level (default 0.05)
%       CENSFLAG (optional) - Censored Plot flag (default 0). If 0
%       censored data will be plotted spreaded on the horizontal
%       segment; if 1 they will be plotted at the given time of censoring.
%   Outputs:
%       Kaplan-Meier plot
%
%   Example: (+ indicate that patient is censored)
%
%       -----
%       Patient   Survival
%              time
%       -----
%           1       7
%           2      12
%           3      7+
%           4     12+
%           5     11+
%           6       8
%           7       9
%           8       6
%           9      7+
%          10       2
%       -----
%   X=[7 0; 12 0; 7 1; 12 1; 11 1; 8 0; 9 0; 6 0; 7 1; 2 0];
%
%   Calling on Matlab the function: kmplot(X) the function will plot the
%   Kaplan-Meier estimation of the survival function
%
%       Created by Giuseppe Cardillo
%       giuseppe.cardillo-edta@poste.it
%
%   To cite this file, this would be an appropriate format:Curve
%   Cardillo G. (2008). KMPLOT: Kaplan-Meier estimation of the survival
%   function.
%   http://www.mathworks.com/matlabcentral/fileexchange/22293

%Input Error handling
args=cell(varargin);
nu=numel(args);
if isempty(nu)
    error('Warning: Data vectors are required')
elseif nu>3
    if nu>4
        error('Warning: Max two input data are required')
    end
end

```

```

end
default.values = {[7 0; 12 0; 7 1; 12 1; 11 1; 8 0; 9 0; 6 0; 7 1; 2 0],0.05,0,1};
default.values(1:nu) = args;
[x alpha cflag flag] = deal(default.values{:});
if ~all(isfinite(x(:))) || ~all(isnumeric(x(:)))
    error('Warning: all X values must be numeric and finite')
end
if isvector(x)
    x(:,2)=0;
else
    if ~isequal(size(x,2),2)
        error('KM PLOT requires Nx2 matrix data.');
```

end

```

    if ~all(x(:,2)==0 | x(:,2)==1)
        error('Warning: all X(:,2) values must be 0 or 1')
    end
end
if nu>1
    if isempty(alpha)
        alpha=0.05;
    else
        if ~isscalar(alpha) || ~isnumeric(alpha) || ~isfinite(alpha)
            error('Warning: it is required a numeric, finite and scalar ALPHA value.');
```

end

```

        if alpha <= 0 || alpha >= 1 %check if alpha is between 0 and 1
            error('Warning: ALPHA must be comprised between 0 and 1.')
        end
    end
end
if nu==3
    if isempty(cflag)
        cflag=0;
    else
        if ~isscalar(cflag) || ~isnumeric(cflag) || ~isfinite(cflag)
            error('Warning: it is required a numeric, finite and scalar CENSFLAG value.');
```

end

```

        if cflag~=0 && cflag~=1
            error('Warning: CENSFLAG value must be 0 or 1')
        end
    end
end
clear args default nu
%string for LEGEND function
str1=[num2str((1-alpha)*100) '% confidence interval'];

%sort data by survival time
x=sortrows(x,1);
%table of patients observed for each survival time
%the TABULATE function sets up this matrix:
%table1=[time count percent(on total)]
table1=[0 size(x,1) 1; tabulate(x(:,1))];

```

```

%if all observed time are integers remove not observed time added by
%TABULATE function
table1(table1(:,3)==0,:)=[];

%Table of censored data
table12=tabulate(x(x(:,2)==1));
if ~isempty(table12)
    % remove not observed time added by TABULATE function
    table12(table12(:,3)==0,:)=[];
    % setup the vector of the censored data
    [cens,loc]=ismember(table1(:,1),table12(:,1)); %find censored data
end

%the percents stored in the the third column are unuseful;
%so, place in the third column how many subjects are still alive at the
%beginning of the i-th interval.
a1=[table1(1,2); -1.*table1(2:end,2)];
table1(:,3)=cumsum(a1); table1(2:end,3)=table1(1:end-1,3);
%number of deaths in the intervals (don't take in account the censored
%data)
if ~isempty(table12)
    table1(cens,2)=table1(cens,2)-table12(loc(cens),2);
end
%finally, delete the first row that is now useless
table1(1,:)=[];

t1=[0;table1(:,1)]; %this is the x variable (time);
%this is the y variable (survival function)
T1=[1;cumprod(1-(table1(:,2)./table1(:,3)))];
if flag %if this function was not called by LOGRANK function
    %compute the standard error of the survival function
    SE=[0;T1(2:end).*sqrt(cumsum(table1(:,2)./(table1(:,3).* ...
        (table1(:,3)-table1(:,2))))));
end

%censored data plotting
if ~isempty(table12)
    %if there are censored data after max(t1), add a new cell into the t1,
    %T1 and SE arrays
    if table12(end,1)>=t1(end,1)
        t1(end+1,1)=table12(end,1)+1;
        T1(end+1,1)=T1(end,1);
        if flag %if this function was not called by LOGRANK function
            SE(end+1,1)=SE(end,1);
        end
    end
end
if ~cflag
    %vectors preallocation
    xcg=zeros(1,sum(table12(:,2))); ycg=xcg; J=1;
    %for each censored data into the i-th time interval...

```

```

for I=1:size(table12,1)
    %compute how many position into the array they must occupy
    JJ=J+table12(I,2)-1;
    %find the correct time interval in which censored data must be
    %placed
    A=find(t1<=table12(I,1),1,'last');
    B=find(t1>table12(I,1),1,'first');
    %equally divide this interval
    int=linspace(table12(I,1),t1(B,1),table12(I,2)+2);
    %put all in the vectors of the plotting variables
    xcg(J:JJ)=int(2:end-1);
    ycg(J:JJ)=T1(A);
    %update the counter
    J=JJ+1;
end
else
    xcg=table1(table1(:,2)==0,1);
    ycg=T1(table1(:,2)==0);
end
else
    if ~flag %if this function was called by LOGRANK function
        xcg=[]; ycg=[];
    end
end
%compute the hazard rate
c1=T1.*numel(x);
c2=-(diff(log(c1(1:end-1)))./diff(t1(1:end-1)));
lambda=mean(c2(c2~=0));

if flag %if this function was not called by LOGRANK function
    %compute the (1-alpha)*100% confidence interval curves
    cv=realsqrt(2)*erfcinv(alpha); %critical value
    %lower curve (remember that: the lower curve values can't be negative)
    lowc=max(0,T1-SE.*cv);
    %if the lower curve reaches the 0 earlier than survival %function, trim the data.
    if isequal(lowc(end-1:end),[0; 0])
        lowcend=find(lowc==0,1,'first');
    else
        lowcend=length(lowc);
    end
    %upper curve (remember that the upper curve values can't be >1)
    upc=min(1,T1+SE.*cv);
    %eventually, correct the data.
    if isequal(upc(end),1)
        cupend=find(upc<1,1,'last');
        upc(cupend:end)=upc(cupend);
    end

    %compute the median survival time (if exist...)
    if isempty(T1(T1==0.5)) %if there is not a point where T=0.5...

```

```

I=find(T1>0.5,1,'last'); %find the first point where T>0.5
J=find(T1<0.5,1,'first'); %find the first point where T<0.5
if isempty(J) %if all points are >0.5...
    mt=0; %...there is no median time
else
    %compute the median time by linear interpolation.
    p=polyfit([t1(I) t1(J)],[T1(I) T1(J)],1);
    mt=(0.5-p(2))/p(1);
    str2=['Median time ' num2str(mt)];
%string for LEGEND function
end
else
    mt=t1(T1==0.5);
    str2=['Median time ' num2str(mt)]; %string for LEGEND function
end

%plot all the data
clf
hold on
S2=stairs(t1(1:lowcend),lowc(1:lowcend),'g--');
%lower confidence interval curve
stairs(t1,upc,'g--'); %upper confidence interval curve
S1=stairs(t1,T1,'b'); %Kaplan-Meier survival function
if mt>0 %if exist a median time...
    S3=plot([0 mt mt],[0.5 0.5 0],'k:');
end
if ~isempty(table12) %if there are censored data...
    S4=plot(xcg,ycg,'r+');
else
    S4=[];
end
hold off

%set the axis properly
xmax=max(t1)+1;
axis([0 xmax 0 1.2]);
axis square
%add labels and legend
txt=sprintf('Kaplan-Meier estimate of survival function (hazard rate: %0.4f)\n',lambda);
title(txt,'FontName','Arial','FontSize',14,'FontWeight','Bold');
ylabel('Estimated survival function','FontName','Arial','FontSize',14,'FontWeight','Bold');
xlabel('Time','FontName','Arial','FontSize',14,'FontWeight','Bold');
if mt
    if isempty(S4)
        legend([S1 S2 S3],'Data',str1,str2)
    else
        legend([S1 S2 S3 S4],'Data',str1,str2,'Censored')
    end
else
    if isempty(S4)

```

```

        legend([S1 S2], 'Data', str1)
    else
        legend([S1 S2 S4], 'Data', str1, 'Censored')
    end
end
end
disp('HAZARD RATE IS AN EXPERIMENTAL FUNCTION!!!!')
end
if nargout
    varargout(1)={table1}
    varargout(2)={table12};
    varargout(3)={t1};
    varargout(4)={T1};
    varargout(5)={xcg};
    varargout(6)={ycg};
    varargout(7)={lambda};
end

```

Appendix 21: MATLAB Program 8

```

function out = logrank(varargin)
% LOGRANK Comparing survival curves of two groups using the log rank %test Comparison of
two survival curves can be done using a statistical
% hypothesis test called the log rank test. It is used to test the null
% hypothesis that there is no difference between the population %survival curves (i.e. the
%probability of an event occurring at any %time point is the same for each population). This
%function use the %Kaplan-Meier procedure to estimate the survival function, so it is
%mandatory to download
% KMPLOT (http://www.mathworks.com/matlabcentral/fileexchange/22293).
%
% Syntax: logrank(x1,x2,alpha,censflag)
%
% Inputs:
%     X1 and X2 (mandatory)- Nx2 data matrix:
%         (X(:,1) = survival time of the i-th subject
%         (X(:,2) = censored flag
%             (0 if not censored; 1 if censored)
%     note that if X is a vector, all the flags of the second column
%     will be set to 0 (all data are not censored).
%     ALPHA (optional) - significance level (default 0.05)
%     CENSFLAG (optional) - Censored Plot flag (default 0). If 0
%     censored data will be plotted spreaded on the horizontal
%     segment; if 1 they will be plotted at the given time of
%     censoring.
% Outputs:
%     Kaplan-Meier plot
%     Log-rank statistics
%

```

```

% Example:
%     load logrankdata x1 x2
%     logrank(x1,x2)
%
%LOG-RANK TEST FOR KAPLAN-MEIER SURVIVAL FUNCTIONS
%
%-----
%UL          S.E.          z          p-value (2-tailed test)  alpha
%-----
%6.57226      2.80788      2.16258      0.03057      0.050
%-----
% The survival functions are statistically different
%
% Created by Giuseppe Cardillo
% giuseppe.cardillo-edta@poste.it
%
% To cite this file, this would be an appropriate format:
% Cardillo G. (2008). LogRank: Comparing survival curves of two groups
% using the log rank test
% http://www.mathworks.com/matlabcentral/fileexchange/22317

%Input Error handling
args=cell(varargin);
nu=numel(args);
if isempty(nu) || nu<2
    error('Warning: Data vectors are required')
elseif nu>4
    error('Warning: Max four input data are required')
end
default.values = {[[],[],0.05,0];
default.values(1:nu) = args;
[x1 x2 alpha,cflag] = deal(default.values{:});
if ~all(isfinite(x1(:))) || ~all(isnumeric(x1(:))) ...
    || ~all(isfinite(x2(:))) || ~all(isnumeric(x2(:)))
    error('Warning: all X1 and X2 values must be numeric and finite')
end
if isvector(x1)
    x1(:,2)=0;
else
    if ~isequal(size(x1,2),2)
        error('LOGRANK requires Nx2 matrix data.');
```

```

end
if ~all(x2(:,2)==0 | x2(:,2)==1)
    error('Warning: all X2(:,2) values must be 0 or 1')
end
end
if nu>=3
    if isempty(alpha)
        alpha=0.05;
    else
        if ~isscalar(alpha) || ~isnumeric(alpha) || ~isfinite(alpha)
            error('Warning: it is required a numeric, finite and scalar ALPHA value.');
```

%check if alpha is between 0 and 1

```

        if alpha <= 0 || alpha >= 1
            error('Warning: ALPHA must be comprised between 0 and 1.')
        end
    end
end
if nu==4
    if isempty(cflag)
        cflag=0;
    else
        if ~isscalar(cflag) || ~isnumeric(cflag) || ~isfinite(cflag)
            error('Warning: it is required a numeric, finite and scalar CENSFLAG value.');
```

%Censored data for treatment 1

```

        if cflag~=0 && cflag~=1
            error('Warning: CENSFLAG value must be 0 or 1')
        end
    end
end
clear args default nu

%recall KM PLOT function to construct tables of data (table1 and table2),
%tables of censored data (table12 and table 22), Kaplan-Meier variables
%(t1, t2, T1 and T2) and Kaplan-Meier graphical data for censored data
%(xcg and ycg).
try
    [table1 table12 t1 T1 xcg1 ycg1 lambda1]=kmplot(x1,0.05,cflag,0);
    [table2 table22 t2 T2 xcg2 ycg2 lambda2]=kmplot(x2,0.05,cflag,0);
catch ME
    disp('Download KM PLOT: http://www.mathworks.com/matlabcentral/fileexchange/22293')
    rethrow(ME);
end

%plot both Kaplan-Meier curves
clf
hold on
S1=stairs(t1,T1,'b'); %Kaplan-Meier curve for treatment 1
if ~isempty(table12)
    S3=plot(xcg1,ycg1,'k+'); %Censored data for treatment 1
    % (if there are)
```

```

else
    S3=[];
end

S2=stairs(t2,T2,'r'); %Kaplan-Meier curve for treatment 2
if ~isempty(table22)
    S3=plot(xcg2,ycg2,'k+'); %Censored data for treatment 2
    %(if there are)
end
hold off
%set the axis properly
xmax=max([t1;t2])+1;
axis([0 xmax 0 1.2]);
axis square
%add labels and legend
title('Kaplan-Meier estimate of survival functions')
ylabel('Estimated survival functions')
xlabel('Time')
if isempty(S3)
    legend([S1 S2],'Treatment 1','Treatment 2')
else
    legend([S1 S2 S3],'Treatment 1','Treatment 2','Censored')
end

clear S1 S2 S3 xmax xcg1 ycg1 xcg2 ycg2 t1 t2 T1 T2

%Full-blown LOGRANK procedure
%Merge the first columns of Table1 and Table2 (time intervals)
%and pick-up unique values
A=unique([table1(:,1);table2(:,1)]);
table=zeros(length(A),9); %matrix preallocation
%Out in the first column the time intervals
table(:,1)=A;
%Put in the columns 2 and 3 and in the proper rows the deaths and alive
%taken from table1 columns 2 and 3
[~, ia ib]=intersect(table1(:,1),A);
table(ib,2:3)=table1(ia,2:3);
%Put in the columns 4 and 5 and in the proper rows the deaths and alive
%taken from table2 columns 2 and 3
[~, ia ib]=intersect(table2(:,1),A);
table(ib,4:5)=table2(ia,2:3);
%remove the rows where there aren't deaths in both treatments
table((table(:,2)==0 & table(:,4)==0),:)=[];
clear A c ia ib table1 table2
%fill the "pigeon-holes"
c=find(table(:,3)==0); %find the "pigeon-holes" of treatment 1
for I=1:length(c)
    if c(I)~=1
        %find the first interval time before the hole where there is almost 1
        %death
    end
end

```

```

J=find(table(1:c(I)-1,3)>0,1,'last');
table(c(I),3)=table(J,3)-table(J,2);
if ~isempty(table12)
    %find eventually censored data
    K=find((table12(:,1)<table(c(I),1) & table12(:,1)>=table(J,1)),1,'last');
    %Put in the hole how many subject were alive before the interval time
    %of the hole
    if ~isempty(K)
        table(c(I),3)=table(c(I),3)-sum(table12(K,2));
    end
end
else
    table(1,3)=length(x1);
end
end
%Do the same for tratment 2
c=find(table(:,5)==0);
for I=1:length(c)
    if c(I)~=1
        J=find(table(1:c(I)-1,5)>0,1,'last');
        table(c(I),5)=table(J,5)-table(J,4);
        if ~isempty(table22)
            K=find((table22(:,1)<table(c(I),1) & table22(:,1)>=table(J,1)),1,'last');
            if ~isempty(K)
                table(c(I),5)=table(c(I),5)-sum(table22(K,2));
            end
        end
    else
        table(1,5)=length(x2);
    end
end
clear c I J K table12 table22

%Fill the table and compute the statistic variable
%Compute the total deaths and alive before the i-th time interval
table(:,6:7)=[sum(table(:,[2 4]),2) sum(table(:,[3 5]),2)];
%Compute the difference between observed deaths for treatment 1 and
%expected deaths in the hyphthesis that the treatments are similar
table(:,8)=table(:,2)-table(:,3).*table(:,6)./table(:,7);
%Log-rank statistic is the sum of column 8 values
J=sum(table(:,8)); UL=abs(J);
%Compute the contribute to the standard error
table(:,9)=prod(table(:,[3 5 6]),2).*(table(:,7)-table(:,6)) ...
./((table(:,7).^2.*(table(:,7)-ones(size(table,1),1))));
%find if there is some NaN (i.e. 0/0)
loc=isnan(table(:,9));
if any(loc)
    table(loc,9)=0;
end
V=sum(table(:,9)); SUL=sqrt(V); %Compute the totale standard error
K=J/V; HR=exp(K); HRci=[exp(K-1.96/SUL) exp(K+1.96/SUL)];

```

[illegible]

VITA

Allen Pannell was born and raised in Knoxville Tennessee. He has three wonderful children, Melissa, Nicholas and Jessica. He received his B.S. and M.S in Statistics from the University of Tennessee in 1982 and 1984. His first job was as a staff statistician at Martin Marietta Energy Systems in Oak Ridge Tennessee. Allen was then hired by his former professors at UT as a consultant for Tennessee Associates International (TAI). Allen performed many roles at TAI, including account manager for major projects as President of the North American Operations. He led various projects saving clients millions of dollars. Allen transitioned to a faculty member in the Graduate and Executive Education (GEE) department at the University of Tennessee, where he was responsible for successfully leading consulting projects for clients, teaching and certifying Black Belts, instructing executive MBA candidates in statistical methods for business decision making and developing Analytics curriculum and programs. Today, Allen is the Director of Business Analytics at Lincoln Memorial University (LMU). His role as director and professor is to develop and deliver a Master of Science in Business Analytics program for the LMU Business School.