This is to certify that the
dissertation entitled

EVALUATION OF TIME TO RESPONSE IN SYMPTOM MANAGEMENT INTERVENTIONS FOR PAIN AND FATIGUE EXPERIENCED BY CANCER PATIENTS

presented by

Sangchoon Jeon

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Epidemiology

Matthew Reeves
Major Professor's Signature

April 30TH, 2009
Date

EVALUATION OF TIME TO RESPONSE IN SYMPTOM MANAGEMENT INTERVENTIONS FOR PAIN AND FATIGUE EXPERIENCED BY CANCER PATIENTS

By

Sangchoon Jeon

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Epidemiology

2009

ABSTRACT

EVALUATION OF TIME TO RESPONSE IN SYMPTOM MANAGEMENT INTERVENTIONS FOR PAIN AND FATIGUE EXPERIENCED BY CANCER PATIENTS

By

Sangchoon Jeon

This research was performed to understand how physical and psychological characteristics of cancer patients impact symptom response to management interventions for pain and fatigue. In addition, this research used an approach to define clinically meaningful reductions in pain and fatigue and identified important predictors of symptom response using several survival analysis approaches.

Six hundred and one cancer patients who were undergoing chemo-therapy were enrolled in one of two clinical trials (termed A and B) that tested interventions for managing cancer-related symptoms over an 8 week period. To define "symptom response" as a clinically meaningful change in symptoms, severity categories of mild score (≤1), moderate score (2-4), and severe score (5-10) were established based on the reported interference with the patient's daily life. Time to response was measured by counting the number of days from onset of symptoms to a symptom response, defined as a transition from severe to moderate, severe to mild, or moderate to mild severity.

Several survival analysis methods were implemented to identify important predictors of symptom response including age, comorbidity, and depression after adjusting for gender, site of cancer, physical function, and trial type (A or B). The survival analyses were performed under different assumptions regarding the proportional hazards assumption and the type of censoring. First, by assuming proportional hazards, the log-rank, Wilcoxon, and Cox proportional hazard model were used. Second, several alternative methods were implemented that do not require the proportional hazards assumption, including the Lin & Wang's test, Cox model with weighted estimations, and Rahbar's method. The impact of interval censoring was assessed by generating Accelerated Failure Time (AFT) models, and finally the effect of the correlation between pain and fatigue were explored using a marginal Cox regression model.

All final models found a significant comorbidity effect for pain and fatigue. Low comorbidity was significantly associated with shorter time-to-response for pain and fatigue in all applied models. There was no significant effect of age after adjusting for comorbidity in the Cox proportional hazard model or the marginal Cox regression model. The AFT model found that younger age and less depressed patients had shorter time-to-response for pain after adjusting for comorbidity as well as a priori confounders, including gender, site of cancer, physical function, and trial type (A or B). None of other survival models found significant association between depression and time-to-response for pain and fatigue.

ACKNOWLEDGMENTS

I wish to thank members of my guidance committee, Mathew Reeves, Ph.D., Charles Given, Ph.D., M. Hossein Rahbar, Ph.D., and Hwan Chung, Ph.D. who provided invaluable direction, intellectual support, and encouragement to complete this work. My thesis advisor, Dr. Mathew Reeves, has spent considerable time guiding me in the preparation of this thesis, improving my epidemiological concepts, and completing this process. I would like to express my appreciation to Dr. Charles and Barbara Given for sharing their valuable data and encouragement. I also appreciate to the valuable advices and intellectual supports from Dr. Alla Sikorskii. I deeply thank my wife, Hyesuk, and two daughters, Ashley and Claire, for their patience and support during this work.

This research was financially supported by the Family Home Care Study (principal investigator: Dr. Barbara Given) and the Automated Telephone Monitoring Symptom Management Study (principal investigator: Dr. Charles Given). The Family Home Care Study (RO1 CA79280) and the Automated Telephone Monitoring Symptom Management Study (RO1 CA30724) were funded by the National Cancer Institute. The Institutional Review Board (IRB) of Michigan State University approved the Family Home Care Study (IRB # 03-247) and the Automated Telephone Monitoring Symptom Management (IRB # 03-242 for). Yale University (IRB # 0810004338) has approved this secondary analysis using collected data from these two studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

OVERVIEW

Cancer patients may suffer from multiple symptoms that originate from the disease itself, the side effects of treatment, or comorbid conditions. Successful symptom management for cancer patients can help maintain therapeutically effective chemotherapy, physical and social functioning, and reduce emotional distress of patients (1-5). To maintain effective management of symptoms, it is important to understand how physical and emotional conditions of cancer patients are related to improvement of symptoms in response to interventions. Pain and fatigue are most prevalent and difficult to resolve among cancer patients. Severe pain and fatigue cause delay or premature termination of important therapies or treatments, impair physical function, and cause significant distress (6, 7). This delay or early termination of treatments may directly affect survival of cancer patients (8).

This study aims to make two contributions to the research program that evaluates symptom management for cancer patients. First, understanding the influence of patient socio-demographic, physical, and emotional conditions on resolution of symptoms will extend our knowledge of symptom management interventions by identifying which cancer patients will likely benefit more or less and sooner or later from symptom management. Second, by defining a clinical meaningful reduction in each symptom through analysis of time to symptom response, this study will help define a strategy for evaluating the effectiveness of interventions for symptom management in cancer patients experiencing pain and fatigue.

This study will address the following specific research questions:

1) Can clinically meaningful changes in pain and fatigue symptoms be measured using the four dimensions of interference (emotions, enjoyment of life, relations with others, and general daily activities) to define clinically meaningful cut-points that separate levels of symptom severity (mild, moderate, and severe)?

2) Using the clinically meaningful severity cut-points, which factors are predictors of time-to-response in pain and fatigue among cancer patients when using survival analysis techniques (the log-rank test, the Wilcoxon test, and the Cox proportional hazard model) that require the assumption of proportional hazards?

3) Do the findings based on survival analysis techniques appropriate for the proportional hazards assumption, hold when using alternative survival techniques (the Lin & Wang's test, the Rahbar's test, and the Cox model with weighted estimation) that do not require the proportional hazard assumption?

4) Do the findings based on survival analysis techniques that are appropriate for right censoring (the Cox proportional hazard model and the Cox model with weighted estimation) hold when using the Accelerated Failure Time model that accounts for interval censoring?

5) Do the findings from the separate models of pain and fatigue (the Cox proportional hazard model, the Cox model with weighted estimation, and the Accelerated Failure Time model) hold when using the marginal Cox model that accounts for the correlation between the two symptoms (pain and fatigue)?

Data used in this research are derived from 2 randomized clinical trials of symptom management. Cancer patients who were undergoing chemotherapy received symptom management interventions during 6 scheduled intervention contacts over an 8 week period. Chapter 1 is a comprehensive literature review that describes what is known about the burden of pain and fatigue among cancer patients and reviews factors related to severity of these symptoms and their management. Patient factors include; gender, age, stage of cancer, site of cancer, the number of comorbid conditions, and depressive symptoms. These patient factors are important because they may be associated with response of pain and/or fatigue when interventions are delivered. These factors could therefore influence the effectiveness of interventions and result in prolonging or shortening time to response of pain and fatigue.

Chapter 2 describes the methodological issues in assessing symptom change and addresses the difficulty of measuring a clinically meaningful symptom change with conventional methods. Supportive evidence for using interference based cut-points of symptom severity is described. Previous studies have suggested methods for developing and testing cut-points to establish symptom severity categories: such as “mild”, “moderate”, and “severe”. Symptom

response to interventions for managing pain and fatigue are defined using transitions among these categories, and used to measure time to symptom response.

In Chapter 3, several alternative survival analysis methods are introduced based on their underlying assumptions. The log-rank (Mantel and Haenszel 1959), Wilcoxon (Breslow 1974), and the Cox proportional hazard model (Cox 1972) are commonly used for survival analysis. When survival functions do not cross one another the hazard functions are proportional, and these methods are valid. To address how to evaluate time to response of pain and fatigue when the proportional hazard assumption is not satisfied in survival analysis, we applied 3 different approaches a nonparametric test for equality of survival function (Lin & Wang 2004), the Cox model with weighted estimation (Schemper 1992), and a nonparametric test for equality of survival mean (Rahbar 2007). To answer if the findings from these survival analysis methods that account for right censoring can be confirmed using a survival analysis method with interval censoring, the Accelerated Failure Time (AFT) model is tested. To answer if the identified factors still have impact on time to response of both pain and fatigue after accounting for the correlation between two symptoms within a patient, the marginal Cox model (Wei 1989) was used.

Chapter 4 describes the data and the methods used for the analyses in this study. Data from 601 cancer patients were collected from the two intervention trials for symptom management (the Family Home Care for Cancer project (Trial A) and the Automated Telephone Monitoring for Symptom

Management project (Trail B)). Patients rated their symptoms and multiple interference items on a 0 to 10 scale at 6 scheduled contacts. They received symptom management intervention when their symptom severity was rated at a 4 or higher. The optimal cut-points of severity were developed based on the sum of four interference items[1] at the first intervention contact. Severity categories[2] from the identified cut-points were examined to see if they consistently differentiated across the sum of the four interference items. Once longitudinal differentiation was confirmed, time to symptom response[3] was used as a measure of meaningful symptom change in response to interventions. The number of days from onset contact[4] to response contact was recorded as time to response, and time from onset to last contact without a response was considered the censoring time. To identify important covariates associated with time to response for pain and fatigue, several survival analysis methods introduced in Chapter 3 were employed and their underlying assumptions explicated.

In Chapter 5, results including the development of cut-points and assessing time to response are described. Based on the identified cut-points, three severity categories were defined: "mild" (sore of 1), "moderate" (score of 2 through 4), and "severe" (score of 5 or greater). These categories significantly differentiated the summed interference scores at each of the contacts. Comorbidity was consistently identified as an important independent covariate

---

[1] The 4 interference items include emotions, enjoyment of life, relations with others, and general daily activities.
[2] Severity categories include mild, moderate, and severe level of severity.
[3] Symptom response includes transition from severe to moderate, severe to mild, and moderate to mild categories of symptom severity.
[4] Onset contact is an intervention contact when patient initially reported a severity of 4 or higher for pain and fatigue.

associated with time to response for pain and fatigue by the different survival analysis methods used. Patients who had less than 3 comorbid conditions had shorter time to response for pain and fatigue compared to those who had less than 3 comorbid conditions. The effects of age and depressive symptom were not significant in three multivariable models (the Cox proportional hazard model, the accelerated failure time model, and the marginal Cox model).

In Chapter 6, the following points are summarized: 1) potential biases associated with combining data sets from two trials 2) evaluation of clinically meaningful changes in the severity of each symptom, 3) use of survival analysis to assess time-to-response among symptoms and conclude with an assessment of the conditions under which each survival technique would be appropriate for use.

This study demonstrates how to assess symptom changes in response to intervention in terms of measurement and analysis of symptom data leading to defining conditions under which time-to-response can be assessed and alternative approaches for evaluating time-to-response given different distributional properties of the responses. The developed measure of symptom response was reviewed as an evaluative measure based on a guideline for measure in clinical medicine (134). Using different survival analysis methods provided consistent statistical evidence for the effect of comorbidity on time-to-response for pain and fatigue. Comorbid conditions are considerable impediments to reducing pain and fatigue severity through interventions.

CHAPTER 1 BACKGROUND OF CANCER-RELATED PAIN AND FATIGUE

In this chapter, Section 1.1 will describe briefly the experience of pain and fatigue among cancer patients and its impact on the quality of patient's daily lives. The association between cancer-related pain and fatigue during treatments will also be discussed in this section. In Section 1.2, I will review potential factors associated with pain and fatigue among cancer patients, in order to evaluate if these factors are also related to the reduction of pain and fatigue in response to symptom management.

A literature review was conducted to investigate the burden of pain and fatigue among cancer patients and the characteristics associated with severity of pain and fatigue. Relevant literature was found in the MEDLINE database using the National of Library of Medicine PubMed with the following keywords; "cancer," "pain," "fatigue," and "chemotherapy." Combinations of each keyword were also used. The review articles were selected by reviewing abstracts; articles were restricted to the English language and published between 1960 and 2008.

1.1 Burden of Pain and Fatigue among Cancer Patients

This section is an overview of pain and fatigue in terms of the definitions, prevalence, pathologies, correlation with one another and associations with other cancer related symptoms. Prevalence describes how commonly cancer patients suffer from pain and fatigue at any given point in time; and is the preferred term to describe burden. Incidence measures the burden of symptoms over a specific

time window, however, only prevalence data are presented in the reviewed studies.

1.1.1 Cancer-Related Pain

The pain from treatment is greatly short term and its severity varies relative to the disease itself. According to a practical guideline from the National Comprehensive Cancer Network (NCCN) (9), pain can be classified into nociceptive and neuropathic pain according to the predominant mechanisms of pain pathophysiology. Nociceptive pain, which results from "injury to somatic and visceral structures and from activating nociceptors," is described as sharp, well localized, throbbing, and pressure-like. This type of pain usually occurs after surgical procedures or from bone metastasis. Neuropathic pain, which results from "injury to the peripheral or central nervous system," is described as burning, sharp, or shooting. It often occurs as an adverse effect of chemotherapy or radiation therapy (9).

Pain is one of most prevalent symptoms among cancer patients, being reported between 36% and 75% of patients. Patients with advanced cancer experience more severe pain (10). In a meta-analysis of fifty-two studies (11), Everdingen and colleagues estimated the pooled prevalence of pain among cancer patients for four different subgroups; patients after curative treatment (33%, 95% confidence interval (CI): 21% to 46%), patients under anticancer treatment (59%, 95% CI: 44% to 73%), patients with advanced/metastatic/ terminal disease (64%, 95%CI: 58% to 69%), and patients at all disease stages

(53%, 95% CI: 43% to 63%). In a 1991 study of a large population of cancer patients from seven hospices in Europe, the United States, and Australia, the prevalence of pain was 60% in breast cancer, 52% in lung cancer, 64% in colorectal cancer, and approximately 80% in gynecological cancers (12). According to one cancer pain study (13), 70% of cancer patients with advanced neoplasm reported pain, while 50% of patients at all stages of cancer reported pain. Pain is defined as "a sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage." (14) Cancer-related pain can occur due to the disease or the treatment. When tumor cells stimulate nerves or cause organ dysfunction patients feel severe pain that can be relieved by removing the tumor cells. Also, cancer patients often suffer from pain as a result of surgery, radiation, or chemotherapy.

The National Cancer Institute (NCI) recommends that the management of pain should be flexible and individualized according to the stage of the disease, personal preferences, and responses to pain interventions. Patient self-report is the standard assessment method for pain. Severity of pain is usually rated using a 0 to 10 point scale; but a categorical or pictorial scale, that uses pictures of faces for rating pain, is also available (9). Cancer-related pain is managed by providing psychological supports, specific educational materials, as well as pain-relieving drugs, such as nonsteroidal anti-inflammatory drugs (NSAIDs), opioids, and combination of analgesics. The NCCN practice guidelines suggest that management should be distinguished by three categories (mild, moderate, and severe) of pain intensity (9). It has been shown that cancer related pain has a

significant association with depression and anxiety. However, the causality among these symptoms has been debated. Cancer patients with pain reported severe depression, anxiety, and/or other psychosomatic symptoms in several studies (15-17). Chronic pain has been considered an important factor leading to severe depression in cancer patients (18, 19). Depression and/or anxiety from concern about the disease may worsen pain, and cancer patients with serious depression and/or anxiety tend to report more pain (15).

1.1.2 Cancer-Related Fatigue

The National Comprehensive Cancer Network (NCCN) describes cancer-related fatigue as "a persistent, subjective, sense of tiredness related to cancer or cancer treatment that interferes with usual functioning." Although cancer-related fatigue has been reported as the most important symptom that impairs a patient's quality of life and daily activities, it has received less attention in management compared with other symptoms such as pain, nausea, or vomiting (20). Fatigue is the most prevalent symptom among cancer patients - a recent report from the NCCN highlights that 70% to 100% of cancer patients experience fatigue (21). Most cancer patients suffer from fatigue while receiving cytotoxic chemotherapy, radiation therapy, bone marrow transplantation, or treatment with biological response modifiers[5] (22). Cancer patients experience fatigue resulting from the disease itself, cancer treatment, psychosocial burdens, and comorbid conditions: this fatigue worsens during the course of chemotherapy and persists

[5] Biological response modifier is a type of cancer treatments that enhances body's immune system.

for months after completing treatment (23). In a study conducted by Greene et al, 82% of breast cancer patients reported fatigue after the first course of chemotherapy (24). Among cancer patients receiving a course of chemotherapy and radiotherapy, 61% reported clinical fatigue (25). In other studies, 89% and 90% of cancer patients reported some degree of fatigue during their chemotherapy (26, 27).

Cancer-related fatigue is that most people generally suffer from in normal life, and it is not relieved by rest or sleep (28). Like pain, fatigue is also subjective and patient self-reports is the standard method for assessment. Additionally, the medical history, physical examination, laboratory data, and description of patient behavior by family members are all important sources of information to gauge the burden of fatigue, especially for children (22). The NCCN practice guideline recommends that fatigue be managed by an interdisciplinary institutional committee, comprised of representatives from medicine, nursing, social work, physical therapy, and nutrition (22).

Despite the high prevalence of fatigue among cancer patients, the biochemical, physiological, and behavioral mechanisms of this complex symptom are poorly understood, making it difficult to identify factors that are associated with fatigue. However, several risk factors associated with cancer-related fatigue have been proposed. Hwang and colleagues proposed a multidimensional conceptual model with situational, biological, physiological, and psychological dimensions that predict cancer-related fatigue (29). The situational dimension represents demographic information including age, gender, stage of cancer,

active cancer treatment, and caregiver status. The biological dimension can be described by serum chemistry profiles. There is evidence that anemia, which is a common side effect of chemotherapy or radiation therapy in cancer patients, is a major factor causing fatigue (30, 31). The impact of anemia on fatigue may be different depending on onset time, patient age, and comorbidity. Psychological factors, such as depression and anxiety, may contribute to the development of chronic fatigue before and after chemotherapy among patients with solid tumors (32). Distress after a diagnosis of cancer can be caused by the initial fatigue, and other side effects of upset, like insomnia which may also increase in patients undergoing chemotherapy. In a study of cancer patients with a history of chemotherapy, fatigue lasted the longest relative to other side effects and had the greatest impact on activities of daily living (5).

1.1.3 Correlations between Pain and Fatigue

Most cancer patients suffer from a number of symptoms which may impair function, therefore it is important to evaluate the impact of multiple symptoms on patient outcomes (33). Pain and fatigue are often observed along with other common symptoms, such as insomnia and depression in cancer patients (34-40). It has been observed that pain, fatigue, sleep disturbance, emotional distress, and poor appetite generally occur together (38). Significant negative effects of the symptom cluster of pain, fatigue and insomnia on physical function was found to be independent of the type of cancer, treatment, stage of cancer, or comorbid conditions (39, 40). Dodd et al. defined a "symptom cluster," as a group of

symptoms that are related to each other. They proposed four groups in cancer patients based on severity of pain and fatigue (i.e. group 1: high fatigue and low pain, group 2: low pain and low fatigue, group 3: low fatigue and high pain, group 4: high pain and high fatigue) (33).

Many studies have shown that pain and fatigue are significantly associated with other common symptoms, treatments, and other factors. For example, in a study with breast cancer patients, pain and fatigue are related to one another and their presence is associated with depression, insomnia, and menopausal symptoms such as hot flashes and night sweats (41). Kaasa et al. found, based on five studies, that pain and fatigue are more common in the more severity affected populations (i.e. palliative care and those with bone metastasis) (42). While chemotherapy alone or in combination with radiation has a significant impact on the level of fatigue, pain is more closely related to the timing of treatment or to the advanced nature of the disease (43). However, the assessment of cancer-related symptoms remains complex due to the multiple symptoms, the multiple etiologies, varying severity, duration, and, co-occurrence of symptoms.

1.1.4 Impact of Pain and Fatigue on Function and Quality of Life

Patients who are diagnosed with solid tumors and are undergoing chemotherapy treatment experience multiple symptoms which are a serious burden for patients as well as for their oncologists and primary care providers (1). These symptoms negatively impact dimensions of the quality of life, such as

physical functioning and depression, and are related to increased morbidity and health care costs (5, 44-46). It has been reported that the cancer-related symptoms have a positive association with negative emotions (47). Gift et al. (48) and Cooley (49) found that fatigue and pain are the most distressing symptoms for lung cancer patients. Pain, fatigue, and depression are recognized as prominent contributors in the suffering experienced by many cancer patients; and clinical studies have increasingly focused on obtaining a better understanding of these symptoms, as well as the development of new, more effective treatments.

1.2 Factors Related to Severity of Pain and Fatigue

Effective symptom management is defined as clinical interventions designed to reduce symptoms through a combination of drugs and other clinical treatments. Effective symptom management over time is a key to maintaining therapeutically effective dosing of chemotherapy agents, physical and social functioning, and to reducing the emotional distress in patients. To increase the effectiveness of symptom management, it may be important to identify the health conditions or patient characteristics which impact the management of symptoms. Identifying factors that predict change in these symptoms may also contribute to the design of effective symptom management studies. By evaluating the effects of interventions designed to manage pain and fatigue, we will be better able to identify those that actually help to resolve or to relieve these symptoms. Because pain and fatigue are key symptoms that may indicate the presence of other

symptoms (33), the factors associated with the reduction of pain and fatigue may be important in predicting change in the overall symptom burden during the period of active symptom management. In this section, potential factors that influence reduction of pain and fatigue in response to symptom management will be reviewed.

Gender and Age It is often recommended that the gender and age of patients should be considered in symptom management (50-53). Since socio-demographic factors are often associated with disease or other health outcomes, it may be difficult to interpret their role in symptom response.

It has been argued that female patients report pain and fatigue differently than male patients. In a recent study, younger patients and female patients had significantly higher fatigue levels (54). They suspected that higher severity in younger patients was due to underreporting by patients who were over 80 years of age. Sechzer and colleagues addressed inappropriate and questionable generalization of findings in cancer research due to sampling bias toward male patients. In response, Miaskowski suggested that this bias exists in symptom management research, and she reviewed articles related to difference in pain and fatigue according to gender among cancer patients (50). In her review of two published studies (44, 51) and her own unpublished studies (50), no gender difference was observed in the prevalence and severity of pain. Research performed by Cleeland and colleagues (44) found that female patients were more likely to be untreated for their cancer-related pain compared to male patients. In

reviews of gender differences in cancer-related fatigue, three studies that evaluated outpatients who were undergoing chemotherapy or radiation therapy (54-56) found that female patients reported higher severity and prevalence of fatigue. However, these gender differences in severity and prevalence could be explained by other factors associated with gender, such as site of cancer, and communication with caregiver. In previous research with 110 cancer patients receiving a 10 contact cognitive behavioral intervention (57), male patients were more likely to report necessary more time to resolve their fatigue compared with female patients in unadjusted analyses. However, resolution for gender was not significantly associated with time to resolve pain and fatigue after adjusting for site of cancer. More lung cancer in male patients may be responsible for poor results concerning fatigue.

Age is another factor potentially associated with pain in terms of its prevalence, severity, and duration. It has been observed that elderly patients are more likely to have high risk of comorbid conditions and late stage cancer (53, 58, 59). Elderly patients tend to attribute their pain as a normal part of aging and avoid reporting pain in order not to disrupt their cancer treatment (60). If elder patients rate the severity of their pain less than they actually feel, their pain might not be sufficiently managed by nurses or care providers. Because anxiety is recognized as a risk factor of fatigue (32), the association between age and anxiety is an important to account for fact, when evaluating the response of fatigue in different age groups. In a longitudinal study conducted with hospitalized cancer patients in Tokyo, younger patients were more distressed and reported

more anxiety than elderly patients (61). It is possible that relatively severe anxiety in younger patients decreases the effectiveness of fatigue management among younger patients. Also based on these findings, elderly patients were probably less likely to have intervention for pain and fatigue because they under-reported their symptoms and have higher comorbid conditions.

Cancer Site and Stage Since pain and fatigue can result from treatment of the disease, site and stage of cancer are important clinical factors likely to be associated with response to these symptoms. Site and stage of cancer may also influence the clinical strategy both for treating disease and managing symptoms. Paters and colleagues observed that patients with ovarian and lung cancer experience greater fatigue compared to those with breast and other cancers (54). In a longitudinal study of elderly cancer patients aged 65 or older (43), patients with lung cancer were significantly more likely to have pain and fatigue compared with those with breast cancer. Since the site of cancer is usually associated with several factors, including gender, age, and stage of cancer, it may be difficult to separate the effect of cancer site on the response time of fatigue from these other factors.

Significantly longer time in pain resolution was observed in patients with late stage of cancer compared with those with early stage of cancer (57). Advanced stages of cancers were more likely to be related to the occurrence of pain in two studies, but a significant association between the stage of cancer and the prevalence of fatigue was not observed (43, 62). In a survey of cancer-

related pain with the Brief Pain Inventory (BPI), 86% of patients with advanced cancer believed that their pain was caused by cancer itself (63). When patients believe their pain is caused by cancer, there is greater interference with their daily activities (10).

Burden of Comorbidity and Other Symptoms Comorbidity is usually measured by counting the number of different chronic conditions. Comorbid conditions are related to older age and chronic fatigue (64). It was found that both age and comorbidity strongly influence patient clinical decision-making (52, 65, 66); older patients with severe comorbid conditions are less likely to have intensive cancer treatments (53). According to the NCCN practice guideline; patient comorbidities are known to be associated with fatigue and they recommend more attention should be paid to chronic conditions in conjunction with the treatment of cancer-related fatigue (21). In a cohort study among cancer patients who were older than 64 years of age, high comorbidity, late stage of cancer, and lung cancer were associated with high risk of pain and fatigue (43).

It is observed consistently that high comorbidity is significantly associated with high prevalence and longer time to resolve fatigue among cancer patients undergoing chemotherapy (57). Based on these findings, more comorbid conditions result in more severe fatigue and may impede treatment of symptoms. That is, comorbidity is a risk factor for severe fatigue, as well as a modifier of the effect for symptom management.

Since cancer patients usually experience multiple symptoms after beginning chemotherapy, more symptoms could impede the effect of any single intervention. Multiple severe symptoms can also produce adverse psychological outcomes, including anxiety and depression which influence adherence to treatment (67-69). Co-occurrence of multiple severe symptoms may result in patients receiving a number of interventions not only for pain and fatigue, but also for other symptoms. For cancer patients who have poor health-related outcomes, a large number of interventions may reduce the effect on each of the symptoms.

Depressive Symptoms A strong interrelationship among pain, fatigue, and depression has been found (69-73). In a meta-analysis, DiMatteo and colleagues found that depressive patients are less likely to comply with medical treatment recommendations (67). They propose several explanations for the effect of depression on treatment. First, depression often causes a high level of hopelessness that treatment is not worthwhile. Second, depression may result in social isolation from individuals who could provide emotional support or assistance; and third, impairment of cognitive function could impair memory which can lead to less compliance among depressed patients.

In general, a measure of fatigue in cancer patients correlates positively with a measure of depression (74). However, Visser and colleagues suggest that there is no evidence of a causal relationship between fatigue and depression and that the same underlying pathology may be responsible for co-occurrence of

these two symptom (71). They observed the consistent correlation between fatigue measured by the Multidimensional Fatigue Inventory (MFI-20) and the mood component of the Center for Epidemiologic Studies of the Depression Scale (CES-D) at the start of treatment, 2 weeks after completion of radiotherapy, and 9 months later. Therefore, depressive symptoms represented by the CES-D may be a good predictor of change of fatigue during the intervention period.

CHAPTER 2 BACKGROUNDS FOR MEASURING SYMPTOM RESPONSE

To examine if the factors described in Chapter 1 have an impact on symptom response to treatment, it is important to define the change in pain and fatigue that represents "clinically meaningful change." The conventional conceptualization and operationalization of symptom responses have been conducted in an unsatisfactory manner. The question is, how can optimal cut-points for severity be determined based upon different magnitudes of interference with patient's daily life. Also, there is the question of whether the established cut-points can reliably and consistently differentiate interference scores over time.

Section 2.1 describes the important measurement issues in assessing pain and fatigue for cancer patients and addresses the difficulty of measuring clinically meaningful change in symptoms with conventional methods. Section 2.2 describes evidence to support the use of interference based cut-points linking differences in severity scores to levels of interference. The previously suggested methods for developing and testing cut-points to establish categories of symptom severity are also described in Section 2.2. A newly proposed method reduces the conventional ordinal symptom scale of 0 to 10 into a more clinically practical classification of three categories: "mild," moderate," and "severe." In Section 2.3 Response to interventions will be defined using transitions among these categories.

2.1 Conventional Measurement: Problems and Suggestions

By definition, symptoms are derived from the patient's perspectives (38). Patients are often asked to rate the severity of a symptom on a numerical scale. Since the American Pain Society used the 11-point scale (from 0=not present to 10=worst possible) for measuring pain (75), it has been extended to measure the severity of other cancer-related symptoms by the NCCN (21). The Brief Pain Inventory (BPI) (76) and Brief Fatigue Inventory (BFI) (4) are widely used instruments that assess sensory (severity) and reactive (interference) dimensions of pain and fatigue, respectively. These instruments define four aspects of severity (worst, least, average, and pain/fatigue right now) along with how greatly pain or fatigue interferes with general activities, mood, walking, normal work, relations with others, enjoyment of life, and sleep (Appendix A).

Once the numerical scales are established, the next step is taken to measure the changes in pain and fatigue. The ultimate purpose in measuring the severity of symptoms is to evaluate the relationship between the reduction of symptoms and the degree to which that has an impact on the patient's quality of life. More specifically, both practitioners and researchers are interested in the question of how the interventions prescribed by care providers ameliorate symptoms, and contribute to maintaining and improving patient's quality of life. For that purpose, it is necessary to have a valid and reliable method of measuring and interpreting changes in symptoms.

Clinically important changes in the severity of symptoms have been evaluated in several ways based on either relative difference or absolute differences (i.e. percent reduction in severity). For instance, Farrar et al. (77) proposed that an relative improvement of more than 30% on the pain intensity scale is a clinically important change among cancer patients. They observed pain intensity for patients who completed clinical trials of pregabalin[6]. The patients rated their pain on a 0 to 10 point scale at baseline and at the end of the clinical trial. After completing the clinical trial they also evaluated changes in pain intensity on a seven-point scale, includes "Very Much Improved", "Much Improved", "Minimally Improved", "No Change", "Minimally Worse", "Much Worse", and "Very Much Worse." Patients who had more than 30% reduction of pain intensity also evaluated their change in symptoms as "Very Much Improved" or "Much Improved." In other studies, a absolute reduction of approximately two points on a scale of 0 to 10 for pain intensity represented a clinically important difference ("Very Much Improved" and "Much Improved") (78). Cepeda et al. (79) performed a study with postsurgical patients to assess meaningful reduction of pain intensity. Patients rated pain intensity on a 0 to 10 point scale and 4-Likert scale (None, Mild, Moderate, and Severe) at baseline and every 10 minutes during administration of analgesic. They also reported the degree improvement in pain on a 5-point Likert scale. For patients with moderate pain, 35% and 45% reduction of pain intensity represented "Much Improved" and "Very Much Improved," respectively. These ordinal scales are relatively easy to be interpreted

[6] Pregabalin is an anticonvulsant drug for neuropathic pain

by clinicians compared with a 0 to 10 scale. Therefore, in this study, I assessed cut-points to categorize the severity of symptoms measured on a 0 to 10 scale.

However, the use of the 0 to 10 scale may be not sufficient to evaluate the impact of pain and fatigue on interferences. For example, Serlin (80) found that the absolute difference in the numerical rating on severity may not always bring about equal levels of differences in distress or functional impairments. Put differently, 20% reduction of severity of pain on a scale of 0 to 10 may not always represent the same percent of improvement in physical function or a reduction of interference scores. Although the validity and reliability of using a scale of 0 to 10 for measuring symptoms has been confirmed to some degree (81), one serious problem has yet to be addressed: patients tend to interpret differently the lower, middle, and higher sections of an 11 point scale (80, 82). For example, from a clinical perspective, the same thirty percent reduction, for instance, 9 to 6 and 3 to 2, may not be equivalent and thus should not be interpreted to have the same meaning.

In response, another approach has been developed to represent a clinically meaningful reduction by using a 3 level categorization scheme: mild, moderate, and severe levels of symptoms. Serlin et al. (80) proposed a threefold classification (mild, moderate, and severe) to describe pain severity, based on a set of interference items. They identified several advantages of the threefold categorization. First, patients are likely to use the lower, middle, and higher sections of the 0 to 10 point scale differently. Second, the threefold classification would be more useful to both clinicians and investigators than a finer

classification with more than three categories. That is, it could facilitate efficient communication between a patient and clinician. For instance, mild pain would not seriously distract patients. Moderate pain could be considered the level of pain that is hard to be ignored by patients. When patients feel their pain needs clinical attention, pain could be consider being severe. Third, it could reflect the fact that a non-linear relationship between severity and interference is better captured with a threefold classification. The non-linear association between severity of pain and fatigue and interference has been demonstrated (34, 76, 83-85). In some cancer pain and fatigue studies, the increased rates of interference were not equal across the 0 to 10 severity scale and relatively small increases were observed at two points of severity (4, 80). Therefore, the three levels of pain severity would be more informative than the finer quantitative gradation such as the 0 to 10 scale. Also the threefold classification has been often used for assessments of cancer-related pain by the National Comprehensive Cancer Network (NCCN), and the Agency for Health Care Policy and Research (AHCPR) (9, 21, 86). These findings are extended to assess the effect of intervention on fatigue in this study. Significant changes in relationship between interference and severity of fatigue were observed for the two cut-points that define the threefold classification (4).

2.2 Developing Cut-points and Testing Longitudinal Consistency of Cut-points

Patients can express the intensity of their pain. The difficulty is that intensity is an abstract concept that is likely to differ by patient. Cleeland et. al. found that the variance of patient self-report can be effectively captured by two dimensions ("sensory" and "reactive") in pain intensity (76). The sensory dimension refers to the severity of pain, while the reactive dimension refers to the interference with the patient's function and quality of life caused by pain. Cleeland defines (87) the measure of symptom burden as "a summative indicator of the severity of the symptoms that are most associated with a disease or treatment, and a summary of the patient's perception of the impact of these symptoms on daily living." He suggests that both symptom severity (sensory) and symptom interferences (reactive) constitute symptom burden. A variety of instruments are designed to assess both the severity of symptoms and their interference with daily life (4, 74, 88, 89).

Serlin suggests that severity is a primary factor in assessing a symptom, because it is more crucial to providing successful clinical management (80). Interference with daily life also is an important factor in understanding how much patients suffer from a symptom, because serious interference contributes to high distress in cancer patients (90). Although high correlations between severity and interference have been observed in several studies (4, 80, 85, 91), severity on 0 to 10 point scale does not always lead to greater interference. For example, patients who report a severity of 3 may not experience more interference compared with those who report a severity of 2 or 1. Therefore, categorization of

severity based on interference would be a more informative measure since it describes both symptom severity (sensory) and interference of daily life (reactive).

Anchoring individual patient reports of pain and fatigue severity to differences in interference using cut-points of mild, moderate, and severe categories has been shown to be appropriating for both research and clinical practice (4, 80, 92-94). It is suggested that these established categories, based on interference scores, provide more stable and more meaningful clinical interpretation in measuring the impact of behavioral interventions for management of pain and fatigue than simply calculating the percent of change using the conventional 0 to 10 scales (77, 78, 95).

There is no agreed-upon definition of the optimal cut-points that separates mild from moderate and moderate from severe pain or fatigue. Rather, different researchers define different cut-points. For instance, Serlin et al. (80) performed an analysis of variance (ANOVA) to find the optimal cut-points of symptom severity based on interferences in daily life. They defined the following range of each category using a 1 to 10 scale: 1–4 for mild pain, 5–6 for moderate pain, and 7–10 for severe pain. The other two studies found the cut-points of severity based on seven (general activity, mood, walking ability, sleep, enjoyment of life, normal work, and relations with others) and six interference items (general activity, mood, walking ability, normal work, relations with others, and enjoyment of life) using multivariate analysis of variance (MANOVA). In a study with oncology outpatients (Breast (52%), Prostate (11%), Lung (11%), and Other (26%)) who experienced pain from bone metastasis, Paul et al. (92) reported that

5 and 8 were optimal cut-points for moderate and severe pain, respectively. In the other study with oncology inpatients (Leukemia (33%), Lymphoma (43%), Breast (10%), Gastrointestinal (6%), Gynecologic (2%), Genitourinary (1%), and other (5%)) who experienced fatigue, Mendoza et al. (4) suggested that the cut-points for severity of fatigue, based on six interference items in the BFI, are 1–3 for mild fatigue, 4–6 for moderate fatigue, and 7–10 for severe fatigue in cancer patients. These studies had fairly consistent cut-points (4 or 5 for moderate, and 7 or 8 for severe) using the ANOVA or MANOVA. However, it is a question that the used interference scores had a normal distribution. If they did not have normal distribution, the result form ANOVA and MANOVA may not be valid.

Our research seeks to determine if, while receiving help with managing their pain and fatigue over time, patients continue to differentiate among levels of interference according to whether their symptom is categorized as mild, moderate, or severe. Patient can report their symptom differently by different measurements. It is known as shift in response which is a major threat to this argument. As patients implement strategies that lower the severity of pain or fatigue, they may "recalibrate" their definitions of severity, interference, or both. For example, patients may report declines in severity but continue to associate these new levels of severity with the same or increased interference. Retrospective assessments are likely to be adjusted by current patient perceptions. In this research patient responses to interventions are followed, compared with their reports of interference at each subsequent observation at which symptoms are rated as severe, moderate or mild. I, then, determine if

interference scores consistently differentiate between severe and moderate and moderate and mild levels of severity of pain and fatigue. If the integrity of the interference-based severity cut-points are preserved over time, then they can be used to measure patient's responses to these intervention strategies (96).

2.3 Defining Response of Pain and Fatigue

This study aims to identify factors, associated with reduction of pain and fatigue, in response to symptom-management interventions. In order to pursue this aim, it is important to define "response of symptom" as a meaningful reduction of symptom, using a reliable measurement of levels of intensity in symptoms. That is, a symptom response during the intervention would represent how successfully pain and fatigue were reduced by treatment. The symptom response can be effectively reflected by a dichotomous variable: "response" or "non-response". After the threefold categories of mild, moderate, and severe were developed to represent levels of severity in symptoms, their longitudinal consistency in differentiating interference scores were examined. If established severity categories differentiate the interference consistently over time, then the meaningful symptom reduction can be captured by transitioning from moderate or severe to a lower category level.

The underlying hypothesis therefore is that when pain and fatigue move from a higher category of severity to a lower category (e.g. Severe to moderate or mild, moderate to mild) they will exhibit substantial reductions in interference

scores. I believe that this would represent a clinically important improvement that oncologists would view as being meaningful. Reduction from moderate to mild may be less clinically important however. For instance, reduction in fatigue from 3 (moderate) to 1 (mild) may not be considered important by oncologists since moderate fatigue would not be considered serious enough to alter the treatment dosing or schedule. However, from a quality of life perspective, a reduction of 3 to 1 in the severity of fatigue corresponds to a decrease in the limitations of daily activities caused by fatigue and results in substantial improvement in physical function (21). Therefore, these three shifts ('severe to moderate', 'severe to mild', and 'moderate to mild') can be interpreted as clinically meaningful reductions in pain and fatigue, and will be called "symptom response" in this study. Non-response includes no shift ('moderate to moderate' and 'severe to severe') and transition from moderate to severe.

CHAPTER 3 SURVIVAL ANALYSIS FOR TIME TO SYMPTOM RESPONSE

Having defined and addressed the relevant measurement issues in Chapter 2, this chapter discusses statistical and technical issues regarding the use of survival analyses for analyzing time-to-response for pain and fatigue. To obtain valid results for testing time-to-response of pain and fatigue, it is important to understand the basic concepts of time-to-event (response) data and the underlying assumptions for survival analysis methods. In this chapter, I highlight the analytical problems associated with assessment of time-to-response of pain and fatigue. These problems include; 1) the proportional hazard assumption, 2) interval censoring, and 3) not accounting for the existence of correlations between two symptom outcomes.

Section 3.1 describes the basic concepts and terminology needed in time-to-event (response) data including censoring, survival function, and hazard rate. Commonly used survival analysis methods including the log-rank, Wilcoxon, and Cox proportional hazard methods are introduced and their underlying assumptions are discussed in section 3.2. When the underlying assumptions of these methods are not met, then, section 3.3 introduces three alternative methods including a nonparametric test for survival functions, a modified Cox model for a non-proportional hazard model, and a survival analysis method based on mean time-to-response. Section 3.4 introduces survival models for using interval censored data. Finally, a marginal Cox model accounting for correlation between pain and fatigue will be introduced in Section 3.5.

3.1 Survival Analysis for Assessing Time to Symptom Response

The concepts of censored data, survival functions, and hazard rates will be described in this section.

3.1.1 Censoring

The dichotomized outcome "symptom response" can be defined as the symptom reduction below a pre-specified level over a defined follow-up period, when exposed to an intervention. Cancer patients who participated in the trials analyzed in this thesis had their symptoms monitored and received intervention strategies at 6 contact periods (Case A in Figure 1). However, individual patients had different time periods over which their pain or fatigue was monitored. These different monitoring time intervals could have occurred due to the following two reasons. First, a patient did not complete the scheduled contacts due to lost to follow-up (for any number of reasons including death, becoming too sick, being busy, and not being interested) as shown by Case B in Figure 1. Second, the onset of symptoms was different for each patient. For example, patients could report their first pain 4 weeks after starting the 8 week study, as shown by Case C in Figure 1. This patient who reported pain at 4 week could be monitored only for 4 additional weeks. In contrast, a patient experiencing pain at the first intervention contact could be monitored for the full 8 weeks.

For these reasons, symptom response is not observed or the exact time-to-response is not known for some individuals. These types of data, including in-

completed observations, are called censored data and there are three possible censoring schemes; right-censoring, interval-censoring, and left-censoring.

Right-censoring is the most common censoring type. Due to loss to follow-up or death, monitoring of symptoms is often terminated before patients experience symptom response. The term of "right-censoring" implies that symptom response occurs some time after the termination of monitoring patients or the lost to follow-up. Additionally, even if patients complete the follow-up, there are some patients who do not experience a response to a symptom until the end of the study. Any termination of monitoring a symptom before a response occurs is called right-censoring. Time from onset to termination of monitoring a symptom or lost to follow-up is becomes the right-censored time for symptom response. Therefore, right-censored time is always shorter than actual time to response.

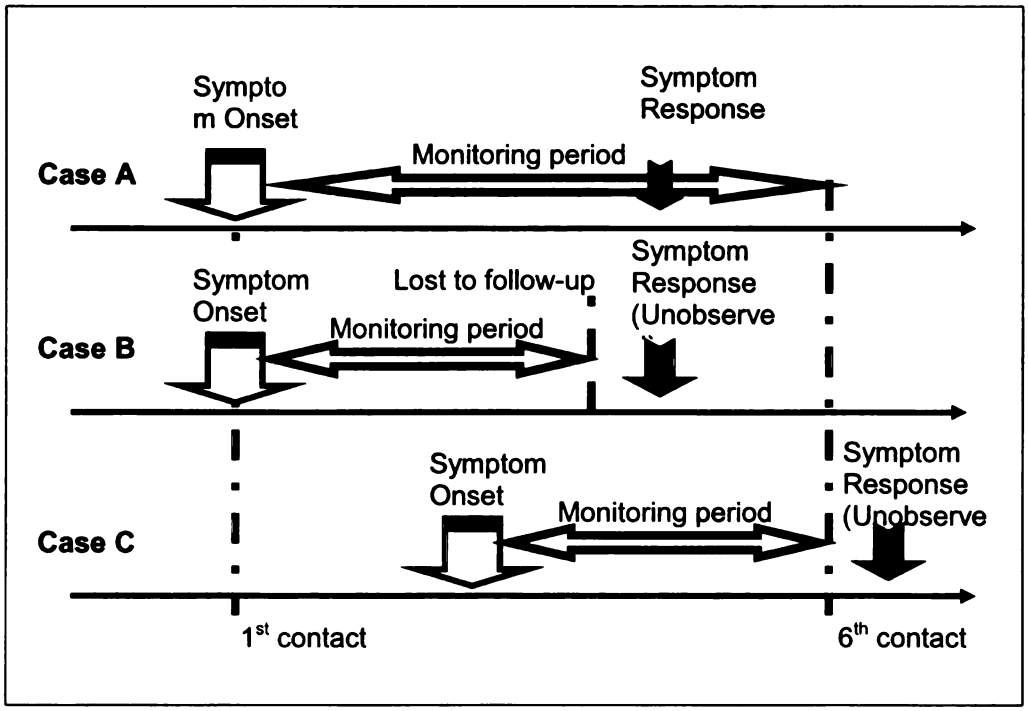
Interval-censoring is another form of censoring. This term reflects that symptom response occurs between two monitoring times but the exact time is unknown. If a symptom is monitored once every week, actual symptom response may occur within the interval between the two contacts. Since the exact time to symptom response is unobservable in this situation, it can be represented by an interval of time. For interval-censored time, actual time to response is between the two interval-censored times.

The third type of censoring is left-censoring, which is encountered when symptom response already occurred before monitoring symptom starts. If actual symptom response occurs before the first monitoring contact the time from onset

to the first monitoring symptom is left-censored time. The left-censoring was not observed in the data analyzed in this thesis.

With right-censored data, several survival analysis methods have been developed to estimate survival curves or assess the importance of covariates. However, few survival analysis methods and software packages are available for interval-censored and left-censored data.

Figure 1 Monitoring symptoms during intervention period



3.1.2 Survival Function and Hazard Rate

The following is a brief overview of the concepts of the survival function and hazard rate for time to pain and fatigue response. The time to pain and fatigue response is a primary outcome of this research. It can be interpreted as

the duration of pain and fatigue within the intervention period. This time to the clinical event (e.g. symptom response) can be summarized by the survival function ($S(t)$) and hazard rate ($h(t)$). The survival function ($S(t)$) can be defined as a probability that a patient's pain or fatigue does not respond during the length of time (in days), t , from onset, namely $S(t) = p[T > t]$. We assume that the distribution of survival time has a density function, f . Then the survival function can be written as

$$S(t) = 1 - \int_0^t f(u)du \qquad (3.1)$$

Greater values of survival function reflect longer time to pain and fatigue response during the follow-up period. In this study, the survival function, $S(t)$, of time to pain and fatigue response can be estimated by the empirical survival function given by

$$\text{S(t)} = \frac{\text{Number of patients with response time > t}}{\text{Number of patients with symptom at time t}} \qquad (3.2)$$

Kaplan and Meier (1958) introduced a nonparametric technique for estimating survival function. The product-limit estimator, which is called the Kaplan-Meier estimator, uses both censored and non-censored observations to estimate survival function. Let n_t be the number of patients with no response to pain or fatigue at contact t and d_j be the number of patients having symptom response at contact t, then we can obtain the Kaplan-Meier estimate of the survival function by

$$\hat{S}_{KM}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right), \tag{3.3}$$

for $t_{(k)} \le t \le t_{(k+1)}$ where $t_{(1)} < t_{(2)} < \cdots < t_{(k)} < t_{(k+1)} < \cdots < t_{(n)}$ are the numbers of days from onset.

The hazard rate ($h(t)$) represents the risk of event occurring during a specific time point. According to the Dictionary of Epidemiology published by the International Epidemiological Association (IEA), the hazard rate is defined as "a measure of the risk of occurrence of an event at a point in time t." (97) In this application, the hazard rate is interpreted as a probability that a symptom responds during time interval t, the conditional on the patient having symptoms at the start of time interval t. Therefore, it can be expressed by the conditional probability of response time T divided by time interval δt as follows;

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t \mid T > t)}{\delta t} \right\} = \frac{f(t)}{S(t)}, \tag{3.4}$$

where f and S are the probability distribution of symptom response time and survival function, respectively. Opposite to the survival function, the higher hazard rate represents a positive outcome for symptom response. That is, a greater hazard rate can be interpreted as a shorter time to symptom response. The cumulative hazard, $H(t) = \int_0^t h(u)du$, which integrates the hazard rate from onset time to a certain follow-up time t, is widely used for analyzing time to event data. It can be obtained from the survival function using the following equation; $H(t) = -\log(S(t))$.

The median and mean of time-to-event are also important statistics to evaluate time to symptom response. The median time to response is a time point when the survival function is 0.5. That is, the time point when half of the patients had responded to their symptoms. For an example, survival function shown in Figure 2, the median time to response is 14 days. The mean time to response is the expectation time that individuals experience response in their symptoms, and is obtained by calculating the area under the survival function, given by $\mu = \int_0^\infty S(u)du$. In Figure 2, the mean time to response is 24.6 days. To obtain the mean time to response, all censoring cases need to occur before the latest symptom response is observed. However, when there are individuals whose symptom did not respond throughout the whole duration of the follow-up, the mean of time-to-response is not available. In this situation, the restricted mean can be alternatively used, given by $\mu = \int_0^\tau S(u)du$ where $\tau > 0$ is an observed maximum time to response (98-101). Figure 3 describes the situation that about 20% of patients did not report symptom response until after 50 days, the reported maximum time to response. The restricted mean of time to response is an area under the survival function until 50 days. Therefore, the restricted mean (22.3 days) from Figure 3 is smaller than the mean of time to response (24.6 days) from Figure 3.

Figure 2 An example of median and mean time-to-symptom response

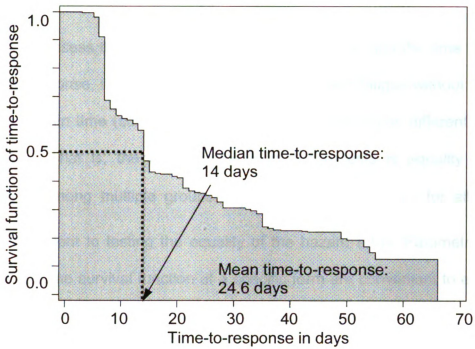
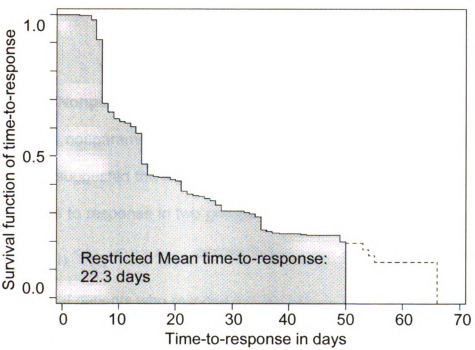


Figure 3 An example of restricted mean time-to-response



3.2 The log-rank, Wilcoxon Test, and the Cox Proportional Hazard Model

To assess the association between a covariate and the time to pain and fatigue response, the probabilities of having pain and fatigue without a response until a certain time (survival function) are compared across different levels of a covariate. That is, the null hypothesis of these tests is equality of survival functions among multiple groups, $H_0 : S_1(t) = S_2(t) = \cdots = S_k(t)$ for all t. This is also equivalent to testing the equality of the hazard rates. Parametric methods which assume survival function of a specific form are convenient to estimate the survival function or mean time to response. However, sometimes it is hard to know whether a survival time has a specific form such as exponential, lognormal, or Weibull. Frequently, nonparametric tests, which compare the empirical survival functions, are used to examine differences in time to event between covariate groups.

3.2.1 Nonparametric Tests for Equality of Survival Functions

One nonparametric method was developed by Mantel and Haenszel (102). They suggested the log-rank test to compare survival functions by deriving ranks of time to response in two groups. Suppose that there are ranked times to response, $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$, across two groups and that there are d_{1j} and d_{2j} individual patients who respond to their pain or fatigue in each of the groups for $j = 1, 2, \cdots, k$. Suppose further that n_{1j} and n_{2j} individuals with no responses

to pain or fatigue in each of the groups just before time $t_{(j)}$. Then there are $n_j = n_{1j} + n_{2j}$ patients who report their symptom just before time $t_{(j)}$ and $d_j = d_{1j} + d_{2j}$ patients who respond to the symptom at time $t_{(j)}$. If two groups have the same survival function, then the number of patients who are responders to their symptom in one group, d_{1j}, have a hypergeometric distribution. Therefore, the expectation and variance of d_{1j} is given by $e_{1j} = n_{1j} d_j / n_j$ and $v_{1j} = n_{1j} n_{2j} d_j (n_j - d_j) / \left\{ n_j^2 (n_j - 1) \right\}$ respectively. The log-rank test statistic can be obtained by summing the differences between the number of observed symptom responses and the expected number of symptom responses in one group, given by

$$\text{Log-rank Test Statistic} = \frac{\left[\sum_{j=1}^{k} (d_{1j} - e_{1j}) \right]^2}{\sum_{j=1}^{k} v_{1j}} \sim \chi^2(1), \qquad (3.5)$$

having a chi-square distribution with one degree of freedom.

Breslow developed the Wilcoxon test for comparing survival functions. In the Wilcoxon test, the difference between the number of observed responses and expected number of responses, $d_{1j} - e_{1j}$, is weighted by the number of patients with symptoms that do not respond just before time, $t_{(j)}$. Therefore, this test is more sensitive to the magnitude of differences in survival functions at earlier

times, compared with the log-rank test. That is, the difference in survival function between groups at one week from onset is more important than that at 4 weeks from onset. The Wilcoxon test statistic is given by

$$\text{Wilcoxon Test Statistic}=\frac{\left[\sum_{j=1}^{k} n_j(d_{1j}-e_{1j})\right]^2}{\sum_{j=1}^{k} n_j^2 v_{1j}} \sim \chi^2(1), \qquad (3.6)$$

where n_j is the total number of individuals with no responses to pain or fatigue just before time $t_{(j)}$. This also has a chi-square distribution with one degree of freedom.

Although the log-rank and Wilcoxon tests do not require assumptions regarding the distribution of the survival function, these tests are appropriate only when the estimated survival functions in the two groups do not cross. The log-rank test is more suitable when the proportional hazard assumption is satisfied. I will discuss in more detail the proportional hazard assumption in the next section.

3.2.2 The Cox Proportional Hazard Model

The nonparametric tests mentioned above are suitable only for univariate analysis. That is, they are available to compare time-to-response for two or more groups classified according to a single covariate. However, it is also important to know the simultaneous impact of patient's demographic variables or health related variables on time-to-symptom response. In order to assess associations between multiple variables and time-to-symptom response, a multivariable

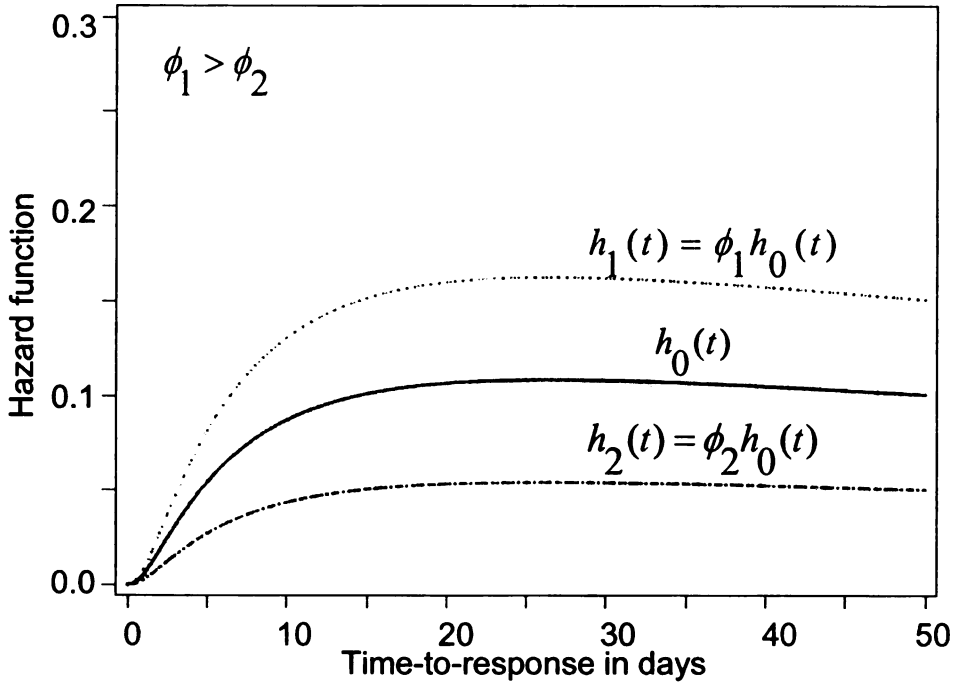
statistical model needs to be used. One widely used statistical model for survival data is the Cox proportional hazard model, developed by Cox (103). This model does not require any specific distribution of time to response, but it requires an assumption of proportional hazard between groups over time. Therefore, it is called a semi-parametric model.

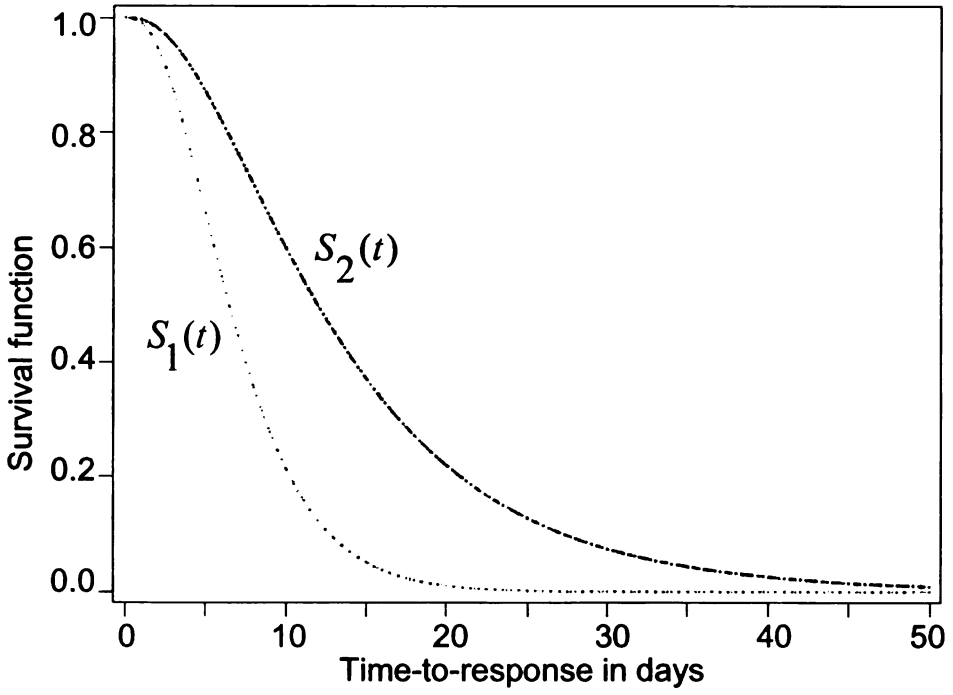
To employ the Cox proportional hazard model, it is important to understand the "Proportional hazard assumption". Let $h_i(t)$ be the hazard rate of symptom response at time t for patients in group i for $i = 1, 2, \ldots$. If the hazard rate of symptom response for patients in group i is proportional to the hazard for individuals in a reference group, then it can be expressed by $h_i(t) = \phi_i h_0(t)$ for $i = 1, 2, \ldots$, where ϕ_i is a constant for group i and $h_0(t)$ is a hazard rate of reference group (often called a baseline hazard). Using a lognormal distribution of survival time, two hazard functions, $h_1(t)$ and $h_2(t)$, are shown in Figure 4, where $\phi_1 > \phi_2$. In this figure, two hazard and survival functions proportionally increase over time and therefore their graphs do not cross.

Under the proportional hazard assumption, the ratio of the hazards for symptom response between group i and reference group, ϕ_i, can be set as a function of covariates $x_{i1}, x_{i2}, \cdots, x_{ip}$. The Cox proportional hazard model with multiple covariates can be written as

$$\frac{h_i(t)}{h_0(t)} = \phi_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_p x_{ip}). \qquad (3.7)$$

Figure 4 Hazard functions and survival functions under the proportional hazard assumption





Suppose $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$ are k individual time to response among $n(\geq k)$ cancer patients. Parameters are estimated from the partial likelihood function as follows (103);

$$L = \prod_{i=1}^{k} \left\{ \frac{\exp(x'_j \beta)}{\sum_{j \in R(t(i))} \exp(x'_j \beta)} \right\}^{\delta_i}, \tag{3.8}$$

where $R(t(i))$ is a set of individuals having no symptom response at time, t, and δ_i is an indicator of censoring (=1 if responded, 0 if lost to follow-up). To account for tied time to response, the approximate partial likelihood function is suggested by Breslow (104) as follows;

$$L = \prod_{i=1}^{k} \left\{ \frac{\exp(S'_i \beta)}{\sum_{j \in R(t(i))} \exp(x'_j \beta)} \right\}^{d_i}, \tag{3.9}$$

where d_i is the number of patients with time to response, $t(i)$ and S_i is the sum of covariates for d_i patients. The partial likelihood function is obtained by counting patient's symptoms with response and non-response at each ranked response time (103). Therefore, the partial likelihood function depends only on the ranked time to symptom response.

The proportional hazard assumption can be tested by adding an interaction term between a covariate and a function of time, $f(t) = \log(t / t^*)$, in the equation 3.7, where t^* is median time to response (105). For instance, if a

covariate, x_i, does not have a proportional hazard across time t, the ratio of hazard rates becomes $h(t)/h_0(t) = \exp(\beta x_i + \gamma x_i f(t))$, where $x_i f(t)$ is the interaction term between a covariate and a function of time which represents a change in the ratio of hazard rates over time. When the proportional hazard assumption is satisfied for covariate, x_i, the interaction parameter γ is zero. Therefore, testing the null hypothesis ($H_0 : \gamma = 0$ vs. $\gamma \neq 0$), determines whether the Cox proportional hazard model is appropriate for this covariate. A non-significant test ($p>0.05$) indicates that the proportional hazard assumption is satisfied.

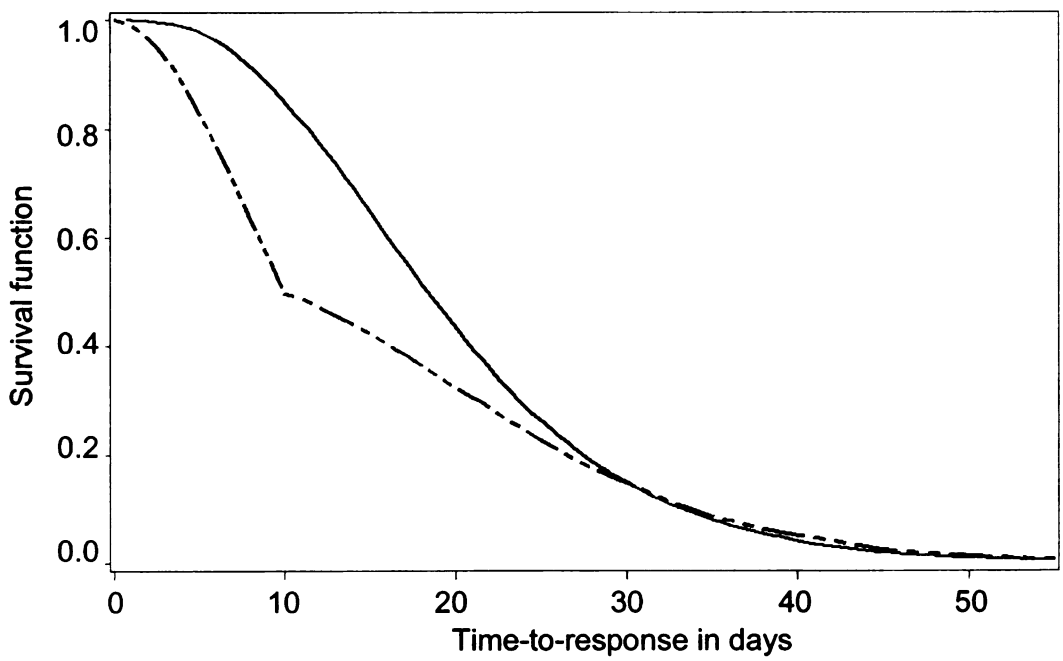
In clinical trials, the output of the Cox proportional model is the hazard ratio (HR), which indicates a treatment effect on survival or resolution of disease. For instance, if the outcome is survival time, then a hazard ratio of less than 1 for the treatment group indicates a longer survival time and therefore a positive outcome. However in symptom response studies, hazard ratios greater than 1 indicate a positive outcome, in which a patient in the intervention group resolves symptoms faster than the control group.

The Cox proportional hazard model has a close relationship with the log-rank test because the Cox proportional hazard model has the same result as that of the log-rank test when there are no tied observations (105). The log-rank test, Wilcoxon test, and the Cox proportional hazard model are non-parametric method and therefore do not require a specific distribution of symptom response times because they depend on the ranks of response or censored times rather

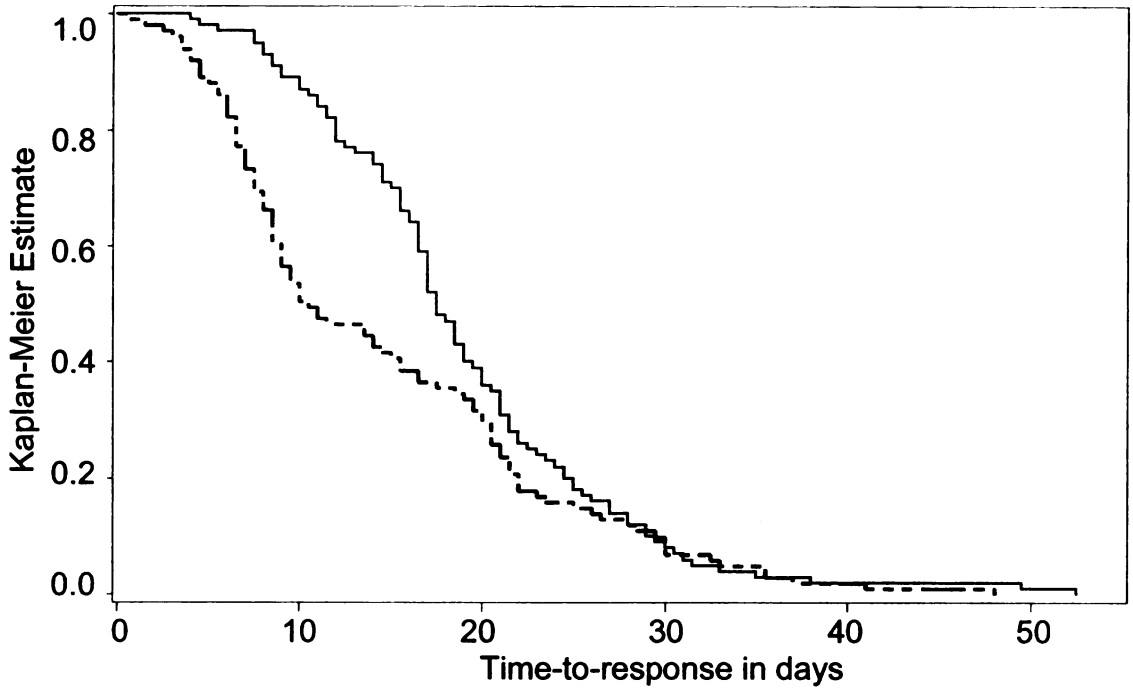
than an actual time interval. However, they still require the strong assumption of proportionality. Figure 5 part (a) illustrates on example where the actual survival curves cross one another. In part (b), the Kaplan-Meier estimate shows the same pattern of survival curves. However, the survival curves estimated from the Cox proportional hazard model (Part (c)) do not cross each other, because the survival curves are estimated under the proportional assumption. Therefore, even if the actual survival curves are different in part (a), the estimated survival curves from the Cox proportional hazard model are fairly similar in part (c).

These methods are not appropriate, when their underlying assumptions are not satisfied. Therefore, I will introduce recently proposed methods which are available for non-proportional survival functions in the next section.

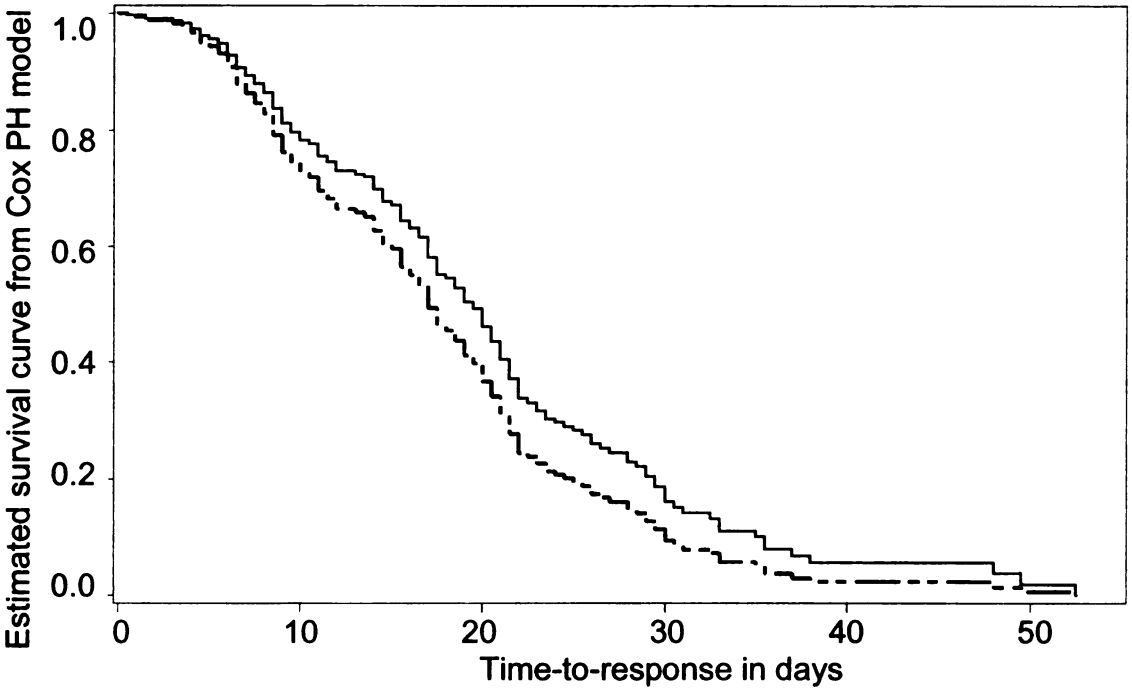
Figure 5 Example of survival functions when the proportional hazard assumption is not satisfied



(a) Actual Survival Curve



(b) Kaplan Meier Estimates



(c) Estimated Survival Curves under the proportional hazard assumption

3.3 Methods of Survival Analysis without the Proportional Hazard Assumption

The previously introduced methods are valid when survival functions have proportional hazards. While the Cox models are popular in health research, they are valid only under the proportional hazard assumption. Unfortunately, this assumption is often violated. Schemper (106) lists consequences of inappropriately using the Cox model when; 1) the power of testing effects of covariates with non-proportional hazards decreases, and 2) the relative risks for covariates with increasing or decreasing hazard ratios will be over- or under-estimated, respectively. The proportionality of hazards should therefore be checked before a survival analysis is carried out using a Cox model.

In addition, there is the question of how to deal with time-to-response data when the proportional assumption is not satisfied. In this section, three alternative methods are introduced.

The first method is a nonparametric test for equality of survival functions proposed by Lin and Wang (107). When the proportional hazard assumption is not satisfied, Lin and Wang (107) propose a new nonparametric test that has greater power to detect overall difference in time to events between groups when survival curves cross. Let d_{1j} and d_{2j} be the number of responses in Group I and II at each time t_j. The number of patients without any response to pain and fatigue before time t_j is denoted by n_{1j} and n_{2j} for Group I and II, respectively.

When there is no difference between the survival functions of the two groups, the number of responses, d_{1j}, has a hypergeometric given by;

$$f(d_{1j};d_{2j},n_{1j},n_{2j}) = \frac{\binom{n_{1j}}{d_{1j}}\binom{n_{2j}}{d_{2j}}}{\binom{n_{1j}+n_{2j}}{d_{1j}+d_{2j}}}. \tag{3.10}$$

The expected number of responses in Group I can be obtained from the hypergeometric distribution given by; $e_{1j} = n_{1j}d_j / n_j$ where $d_j = d_{1j} + d_{2j}$ and $n_j = n_{1j} + n_{2j}$. The new test statistic is obtained by the summation of the squared differences of the number of observed responses in symptom and the expected number of responses in symptom at each ranked time, $\Delta = \sum_{j=1}^{k}\left(d_{1j} - e_{1j}\right)^2$. The test statistic is obtained from the following equation;

$$T^* = \frac{\Delta - E(\Delta)}{\sqrt{Var(\Delta)}}, \tag{3.11}$$

where $E(\Delta)$ and $Var(\Delta)$ are calculated based on the hypergeometric distribution, and are given by

$$E(\Delta) = \sum_{j=1}^{k} \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)} \tag{3.12}$$

and

$$Var(\Delta) = \sum_{j=1}^{k} \{E(d_{1j}^4) - 4E(d_{1j}^3)E(d_{1j}) + 6E(d_{1j}^2)(E(d_{1j}))^2$$
$$-3(E(d_{1j}))^4 - (Var(d_{1j}))^2\}$$

(3.13)

According to their simulation study, this test has a higher power of detection of differences in survival curves between two groups than the log-rank and Wilcoxon test when the proportional hazard assumption does not hold. However, this test has less power than the log-rank and Wilcoxon test when survival curves do not cross each other. The Lin and Wang test is available for testing the survival function only between two groups.

The second method is a modified Cox model that weights a coefficient of a covariate which does not have a proportional hazard relative to the other covariates (106). This is a modified Cox model for non-proportionality of hazards. Several survival models have been developed for non-proportional hazards. Moreau proposed the piecewise proportional hazard model with separated parameters of a covariate for each interval of survival time (108). The Relative survival model was proposed by Esteve and colleagues, and to incorporate non-proportional hazards, it has different baseline hazards for each interval of survival time (109). Also, Cox suggested an alternative method of including a time by covariate interaction term in the Cox proportional hazard model (103). Schemper proposed a Cox Model with weighted estimation (WCM) for non-proportionality of hazards (106). For parameter estimation for WCM, the partial likelihood function of the Cox proportional hazard model (Equation 3.9) is modified. In general, the

maximum likelihood estimation (MLE) of parameter, β, for the Cox proportional hazard model is obtained by maximizing the partial likelihood function (Equation 3.9). The partial likelihood function for the Cox proportional hazard model has a maximum value, when the first derivative of log of the partial likelihood function is zero as follows;

$$\frac{\partial \log L}{\partial \beta_r} = \sum_{i=1}^{k} \left[S_{ir} - \frac{\delta_i \sum_{j \in R(t(i))} x_{lr} \exp(x'_j \beta)}{\sum_{j \in R(t(i))} \exp(x'_j \beta)} \right] = 0. \qquad (3.14)$$

For parameter estimation for WCM, the above equation is weighted by a function of time, $f_r(t_i)$, for each parameter as follows;

$$\frac{\partial \log L}{\partial \beta_r} = \sum_{i=1}^{k} f_r(t_i) \left[S_{ir} - \frac{\delta_i \sum_{j \in R(t(i))} x_{lr} \exp(x'_j \beta)}{\sum_{j \in R(t(i))} \exp(x'_j \beta)} \right] = 0 \qquad (3.15)$$

where $f_r(t_i)$ is a weighted function for parameter β_r, and can be replaced by the number of patients with a non-responded symptom, $R(t(i))$ (Gehan score) (110), or total sample size times Kaplan Meier Estimator, nF (Prentice) (111). Schemper demonstrated better power to detect the effect of a covariate with non-proportional hazard, compared to the Cox proportional hazard model. Heinze and his colleagues developed a SAS macro for the Cox model with weighted estimation (WCM) (112).

Rahbar et al. developed new methods to compare mean of survival time between groups (113-116). With respect to evaluating time-to-response, the previously discussed methods examine a null hypothesis that there is no

difference in survival functions among different levels of covariate across time. However, Rahbar and his colleagues suggested testing the equality of mean of time-to-response among different levels of covariate during the follow-up period. Regardless of difference between tests for survival functions and means of survival (response) time, the equality test for means in response time is a feasible way to compare times of response among groups. Also it does not require the assumption that survival functions do not have proportional hazards.

Rahbar et al. (113) proposed a new nonparametric test for mean survival time for multiple groups. Let θ_i be a mean time-to-response in group i. The null hypothesis of this test is $H_0 : \theta_1 = \theta_2 = \cdots = \theta_k$. Let $t_{i(1)} < t_{i(2)} < \cdots < t_{i(n_i)}$ be observed the time in response or censored time in group i, and $\delta_{i(j)}$ be an indicator variable of censoring (1 for response, 0 for censoring). The mean time to response for group i can be estimated by integrating the estimated survival function over time, given by;

$$\hat{\theta}_i = \sum_{k=1}^{n} [t_{i(k+1)} - t_{i(k)}] \cdot \hat{S}_{KM(i)}(t_{(k)}). \qquad (3.16)$$

The pooled or combined estimator of mean time-to-response among k groups is given by

$$\hat{\theta}_n^C = \sum_{i=1}^{k} \hat{L}_i \hat{\theta}_i, \qquad (3.17)$$

where $\hat{L}_i = \dfrac{\hat{\sigma}^2_{n_i} n_i}{\hat{\sigma}^2_{n_1} n_1 + \hat{\sigma}^2_{n_2} n_2 + \cdots + \hat{\sigma}^2_{n_k} n_k}$. For group i, Kaplan Meier mean, $\hat{\theta}_i$, can be obtained from the equation (3.16). The estimator of variance, $\hat{\sigma}^2_{n_i}$, is calculated by the following equation;

$$\hat{\sigma}^2_{n_i} = n_i \sum_{j=1}^{n_i} \delta_{i(j)} \left(\frac{\hat{A}(t_{i(j)})}{K(t_{i(j)})} \right)^2 , \tag{3.18}$$

where $K(t) = 1 + \sum_{j=1}^{n_i} I(t_{i(j)} > t)$, $\hat{A}(t_{i(s)}) = \sum_{j=s}^{n_i} \hat{F}(t_{i(j)})(t_{i(j+1)} - t_{i(j)})$.

With the estimated mean and variance, a new test statistic (113), Λ_n, is proposed to test the difference of survival means among different groups, given by:

$$\Lambda_n = n(\hat{\theta} - \hat{\theta}^C)' \hat{\Gamma}_n^{-1} (\hat{\theta} - \hat{\theta}^C), \tag{3.19}$$

where $n = n_1 + n_2 + \cdots + n_k$ is the total number of observations, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)'$ is a vector of Kaplan Meier means, and $\hat{\theta}^C = (\hat{\theta}^C_n, \hat{\theta}^C_n, \cdots, \hat{\theta}^C_n)'$ is a vector of the combined estimators of means of time-to-response among k groups, which is defined in equation 3.17, among k groups. The estimated variance-covariance matrix of a vector of the estimated mean of time-to-response, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)'$, is obtained by the following equation;

$$\hat{\Gamma}_n = \begin{pmatrix} \hat{\gamma}_{n_1} & \hat{\gamma}_{n_{21}} & \cdots & \hat{\gamma}_{n_{k1}} \\ \hat{\gamma}_{n_{12}} & \hat{\gamma}_{n_2} & \cdots & \hat{\gamma}_{n_{k2}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{n_{1k}} & \hat{\gamma}_{n_{2k}} & \cdots & \hat{\gamma}_{n_k} \end{pmatrix}, \tag{3.20}$$

where $\hat{\gamma}_{n_i} = \hat{\sigma}_i^2 \omega_{i,n}^{-1}(1-\hat{L}_i)^2 + \sum_{j\neq i}^{k} \hat{\sigma}_j^2 \omega_{j,n}^{-1} \hat{L}_j^{\,2}$,

$\hat{\gamma}_{n_{ij}} = -\hat{\sigma}_i^2 \omega_{i,n}^{-1} \hat{L}_i - \hat{\sigma}_j^2 \omega_{j,n}^{-1} \hat{L}_j + \sum_{m=1}^{k} \hat{\sigma}_m^2 \omega_{m,n}^{-1} \hat{L}_m^{\,2}$, and $\omega_{i,n} = n_i / n$, $i, j = 1, 2, \cdots, k$.

Rahbar et al. (113) showed that under the null hypothesis of equality of mean times to response, the sampling distribution of Λ_n converges to a chi-square distribution with k-1 degrees of freedom.

In the extension of this idea, Rahbar and colleagues have developed a nonparametric regression model of mean survival time (114-116). The regression model accounts for effects of two discrete covariates on time-to-response without the proportional hazard assumption. Rahbar and Gardiner (114, 115) proposed a linear regression model for the additive effect of discrete covariates on mean of time-to-event by estimating average increases of mean time-to-event among different levels of a covariate.

Suppose that there is a demographic variable such as patient age categorized into $m1$ levels with values, $c_1, c_2, \cdots, c_{m1}$ and a health related variable such as comorbidity categorized into $m2$ levels with values, $d_1, d_2, \cdots, d_{m2}$. Let

T_{ijk} be time-to-response of pain and fatigue for patient i with age of c_j and comorbidity of d_k . Then a linear regression model for time-to-response is expressed by

$$T_{ijk} = \alpha + \beta_1 c_j + \beta_2 d_k + \varepsilon_{ijk} \tag{3.21}$$

where α, β_1, β_2 are unknown parameters and ε_{ijk} is unobservable error. It is assumed that ε_{ijk} is independent and identically distributed with zero mean and common variance. It is different from a typical linear regression model, because the distribution of error, ε_{ijk}, is not assumed to be a normal distribution. Then the mean time-to-response, θ_{ijk}, is estimated by two covariates, c_j and d_k, given by

$$\theta_{jk} = \alpha + \beta_1 c_j + \beta_2 d_k \tag{3.22}$$

Rahbar and Gardiner (115) propose the estimators of coefficients, α, β_1, β_2 by a linear combination of Kaplan Meier means for each level of the covariates, $\hat{\theta}_{jk}$ (e.g. jth level of age and kth level of comorbidity), given by

$$\hat{\beta}_1 = (m2M1)^{-1} \sum_{k=1}^{m_2} \sum_{j \neq j'} \frac{\hat{\theta}_{jk} - \hat{\theta}_{j'k}}{d_j - d_{j'}}, \tag{3.23}$$

$$\hat{\beta}_2 = (m1M2)^{-1} \sum_{j=1}^{m_1} \sum_{k \neq k'} \frac{\hat{\theta}_{jk} - \hat{\theta}_{jk'}}{c_k - c_{k'}}, \text{ and} \tag{3.24}$$

$$\hat{\alpha} = n^{-1} \sum_{j=1}^{m1} \sum_{k=1}^{m2} n_{jk} \hat{\theta}_{jk} - \hat{\beta}_1 \left(n^{-1} \sum_{j=1}^{m1} n_{j.} d_j \right) - \hat{\beta}_2 \left(n^{-1} \sum_{k=1}^{m2} n_{.k} c_k \right), \quad (3.25)$$

where $M1 = \frac{1}{2} m1(m1-1)$, $M2 = \frac{1}{2} m2(m2-1)$, $n = \sum_{j=1}^{m1} \sum_{k=1}^{m2} n_{jk}$, $n_{j.} = \sum_{k=1}^{m2} n_{jk}$,

$n_{.k} = \sum_{j=1}^{m1} n_{jk}$, and n_{jk} is the number of patients at the j th level of age and the k th level of comorbidity.

An extension of this model, the linear model with an interaction effect between two discrete covariates, was suggested by Rahbar et al. (116). Using this method, it is possible to identify multiple factors associated with the time of response in pain and fatigue without a restriction of the proportional hazard assumption.

In a simulation study, it was shown that this regression model had better power to detect the interaction effect than the Cox proportional hazard model, since the rate of censoring before the end of follow-up increases. However, this method has some drawbacks. First, the largest censoring (lost to follow-up) time should not be larger than the largest time to response. If this condition is not satisfied, then the sample mean of time-to-response cannot be estimated. Second, this method may require a large sample size at each level. Third, this method does not easily extend to more than two covariates. Fourth, only discrete covariates are available in this method.

Three methods, Lin & Wang's test, the Cox model with weighted estimation (WCM), and Rahbar's method, will be employed in this study to

determine if they produce results different from the methods of proportional hazard based (i.e. the log-rank, Wilcoxon test, and the Cox proportional hazard model). Since there are no commercial software packages for these methods, a SAS macro was developed for Lin & Wang's test and Rahbar's methods for this dissertation (Appendix B).

3.4 Methods of Survival Analysis for Interval Censored Data

All of the previously discussed conventional and alternative survival methods are developed for "right-censored time data" when exact time-to-event (response) is known. However, it is often difficult to obtain exact time to response in clinical trials and observational studies, which are usually designed for monitoring patients at scheduled times. Therefore, it is important to understand survival analysis methods for interval censoring when all that is known is that a symptom response occurred between two scheduled intervals. This next section reviews the history of developing methods for interval censored data, and introduces the Accelerated Failure Time (AFT) model for interval censoring that will be used for this analysis.

In this study, patients were monitored at scheduled contacts for an observed pain and fatigue response. Because the question was that a patient has response in the last 7 days, time is observed at intervals so the exact time of onset or time of response is not known. For example if a patient reported a pain response at the 4^{th} intervention contact, the actual response time could have

occurred any where between the 3rd and 4th contact. This is a type of interval censoring. Since methods for interval censoring are more complicated and available software is limited, the typical approach in previous studies is to use either the lower or upper bound of the interval, or the middle value of the interval as the exact time of the event. However, these approaches have been criticized because it may produce biased estimations (117, 118). Therefore, methods for interval censoring will be examined and their results compared to determine if they are different from other methods with right censoring.

The Accelerated Failure Time (AFT) model is an available method incorporating interval censoring to evaluate time-to-response of pain and fatigue, and it is defined by the transformation:

$$T = T_0 \, \mathrm{e}^{x'\beta}, \tag{3.26}$$

where T is time to response for a patient with covariates $x' = (x_1, x_2, \cdots, x_p)$ and T_0 is time-to-response from baseline distribution corresponding to a covariate value of zero. The AFT model assumes that the parameters β's can accelerate or decelerate time-to-response for a patient with covariates, $x' = (x_1, x_2, \cdots, x_p)$. That is, the effect of covariates in this model is to change the scale but not the location of a baseline distribution of time to response. From this transformation, a survival function can be obtained by

$$S(t \mid x) = P[T > t \mid x] = P[T_0 > te^{-x'\beta}] = S_0(te^{-x'\beta}), \tag{3.27}$$

where S_0 is the survival function for patients with covariate value 0. If $\log T_0$ is $\sigma\varepsilon$, then the ATF model is expressed by logarithm form

$$\log T = x'\beta + \log T_0 = x'\beta + \sigma\varepsilon \tag{3.28}$$

The error term, $\varepsilon = (\log T - x'\beta)/\sigma$, is a random variable that can be assumed to have one of a variety of distributions including the standard normal, standard extreme value, and the logistic distribution. If $S(t) = P(\varepsilon_i > t)$, $F(t) = P(\varepsilon_i \le t)$, and $f(t) = dF(t)/dt$, parameters β's and σ can be estimated from the log-likelihood function as following (119);

$$\log l = \sum \log\left(\frac{f(u_i)}{\sigma}\right) + \sum \log\left(S(u_i)\right) + \sum \log\left(F(u_i)\right) + \sum \log\left(F(u_i) - F(v_i)\right), \tag{3.29}$$

Where $u_i = (\log t_i - x_i'\beta)/\sigma$. In this equation, the first sum of log-likelihood includes uncensored observations, the second sum includes right-censored observations, the third sum includes left-censored observations, and the last sum includes interval censored observations. Therefore, the AFT model can be applied to left-, right-, and interval censored data. If an appropriate distribution is used for time-to-response with interval censoring, it may provide a more accurate result than that of a Cox proportional hazard model. Advantages of the AFT model are to incorporate interval censored data and not to be restricted by the proportional hazard assumption.

Unlike the Cox proportional hazard model, the AFT model does not provide a hazard ratio. Therefore, there is an alternative interpretation of results

from the AFT model. From a AFT model given by $T_x = T_0^{\sigma} \exp(\beta_0 + \beta_1 X)$, the survival time ratio (or the response time ratio), $\exp(\beta) = T_{x=1} / T_{x=0}$, can be used to measure the effect of a covariate (x) on time-to-response (120). A positive value of coefficient β indicates that a patient who had a higher value of a covariate (x) took a longer time to response. For example, when a covariate indicates high comorbidity and β is positive, the response time ratio is interpreted as a patient with a high comorbidity took $a(=\exp(\beta))$-fold longer to response than one who had a low comorbidity.

3.5 Methods of Survival Analysis When Outcomes Are Not Independent

Unlike the survivorship of patients, in survival analysis of symptom response in each individual could have multiple events (e.g. response of pain and fatigue). Since cancer patients usually experience both pain and fatigue (34-40), time-to-response of pain and fatigue are unlikely to be independent in this analysis. Single symptom assessment ignores the associations between pain and fatigue and may underestimate the patient's total symptom severity burden. For example, when the effects of symptom management interventions are evaluated, pain may be more responsive than fatigue and, depending upon the association between the two symptoms within individuals, may, indirectly, decrease in severity more than fatigue. Since pain and fatigue may be interrelated, as a function of disease, treatment, or patient characteristics, the

possible correlation should be taken into account in evaluating times-to-response. Therefore, the marginal Cox model incorporating correlations between pain and fatigue will be discussed in this section.

Wei and Lin (121) developed the marginal Cox model for multivariable failure time observations that accounts for correlation among multiple event times within any subject. They applied this model to an example where multiple episodes of viral infections could occur in AIDS patients. To account for the correlations among multiple infections, a robust sandwich covariance matrix estimator was used. Suppose that n cancer patients can experience up to 2 potential events (pain and fatigue responses). Let $x_{ki1}, x_{ki2}, \ldots, x_{kip}$ be covariates for ith cancer patient with symptom k (k =1 for pain, k =2 for fatigue). The marginal Cox model is given by

$$h_k(t \mid x_{ki1}, x_{ki2}, \ldots, x_{kip}) = \exp(\beta_{k1} x_{ki1} + \beta_{k2} x_{ki2} + \ldots + \beta_{kp} x_{kip}) h_{k0}(t), \quad (3.30)$$

where h_{10} and h_{20} are baseline hazard functions for pain and fatigue response. Parameters, $\beta_{k1}, \beta_{k2}, \ldots, \beta_{kp}$, are estimated by a method developed by Wei, Lin, and Weissfeld (WLW method) (121). The robust sandwich covariance matrix estimator (the empirical covariate matrix estimator), which is widely used for the generalized estimating equations (GEE) (122), can be used to account for the correlation between pain and fatigue within an individual.

The marginal Cox proportional hazard model can be used to examine the different effects of a covariate on each outcome, e.g. pain and fatigue, by testing $\hat{\beta}_{11} = \hat{\beta}_{21}$. Suppose that $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$ are estimated effect of age on time-

to-response in pain and fatigue, respectively. Significant differences between the two estimates, $\hat{\beta}_{11} \neq \hat{\beta}_{21}$, indicate that patient age has a different impact on the response to pain and fatigue. Therefore, the marginal Cox model allows for either an overall effect of a covariate on both pain and fatigue (if $\hat{\beta}_{11} = \hat{\beta}_{21}$), or two separate effects ($\hat{\beta}_{11} \neq \hat{\beta}_{21}$). Most cancer patients undergoing chemotherapy experience both pain and fatigue, therefore the marginal Cox model is appropriate to account for the correlation between the two symptoms.

In summary, time-to-response of pain and fatigue can be evaluated by using appropriate survival analysis methods; and it is important to understand the underlying assumptions of each method. In this analysis, I will demonstrate how to detect meaningful reduction (response) of pain and fatigue, how to define time in their response, and how to implement survival analysis methods appropriately according to underlying assumptions.

Specific Aims of the study

The primary question of this study is what factors are associated with time-to-response of pain and fatigue among cancer patients who are receiving symptom management interventions in two clinical trials. The interventions are designed to reduce the severity of symptoms in patients undergoing chemotherapy by a nurse, coach, or the AVR system during an 8 week interval period.

This study will address the following specific research questions:

1) Can clinically meaningful changes in pain and fatigue symptoms be measured using the four dimensions of interference (emotions, enjoyment of life, relations with others, and general daily activities) to define clinically meaningful cut-points that separate levels of symptom severity (mild, moderate, and severe)?
2) Using the clinically meaningful severity cut-points, which factors are predictors of time-to-response in pain and fatigue among cancer patients when using survival analysis techniques (the log-rank test, the Wilcoxon test, and the Cox proportional hazard model) that require the assumption of proportional hazards?
3) Do the findings based on survival analysis techniques appropriate for the proportional hazards assumption, hold when using alternative survival techniques (the Lin & Wang's test, the Rahbar's test, and the Cox model with weighted estimation) that do not require the proportional hazard assumption?
4) Do the findings based on survival analysis techniques that are appropriate for right censoring (the Cox proportional hazard model and the Cox model with weighted estimation) hold when using the Accelerated Failure Time model that accounts for interval censoring?
5) Do the findings from the separate models of pain and fatigue (the Cox proportional hazard model, the Cox model with weighted estimation, and the Accelerated Failure Time model) hold when using the Cox

marginal model that accounts for the correlation between the two symptoms (pain and fatigue)?

CHAPTER 4 METHODS

Section 4.1 describes the two symptom intervention trials and illustrates the method used to measure the severity of pain and fatigue, and patient response to interventions designed address to these symptoms. Section 4.2 identifies the independent and dependent collected as part of the two trials. Finally, the analytical strategy for identifying important factors affecting time to response among patients with pain and/or fatigue in these trials is presented in section 4.3.

4.1 Clinical Intervention Trials for Symptom Management

4.1.1 Subjects and Settings

This study utilizes information from two large randomized clinical trials of symptom management interventions for cancer patients. These trials were funded by the National Cancer Institute (R01 CA79280, Trial A, Family Home Care for Cancer: A Community Based Model and R01 CA30724, Trial B, Automated Telephone Monitoring for Symptom Management) (123, 124). Each trial compared two symptom management intervention arms. Trial A compared nurse and non-nurse (Coach) intervention groups, while Trial B compared a nurse intervention to an Automated Voice Response (AVR) system intervention. Trial A required that patients should have a caregiver who agreed to participate in the study. Trial A and B were performed with the same timelines.

The same recruitment criteria were used for both trials. Cancer patients were recruited from one community cancer oncology program, two comprehensive cancer centers, and six hospital affiliated-community oncology centers in Michigan, Indiana, Connecticut, and Maryland. From October, 2003 until July, 2006, 1605 eligible patients were approached. The study inclusion criteria were; 1) 21 years of age or older, 2) have a diagnosis of a solid tumor cancer or non-Hodgkin's lymphoma, 3) be undergoing a course of chemotherapy, 4) be ability to speak and read English, 5) possess a touchtone telephone, and 6) have no cognitive deficits. If a patient had a caregiver, then his/her caregiver should be able to speak and read English and not have cognitive deficits. Among the eligible patients, 815 patients consented to participate in either Trial A or B. The participants were not assigned to any trial particular at the time of initial recruitment. The trials were approved by the institutional review board (IRB) of the Michigan State University, as well as IRB's of the participating sites. The entry criteria were different between the two trials. The only difference between the two studies at this stage was that Trial A requires a participation of caregiver while Trial B does not require.

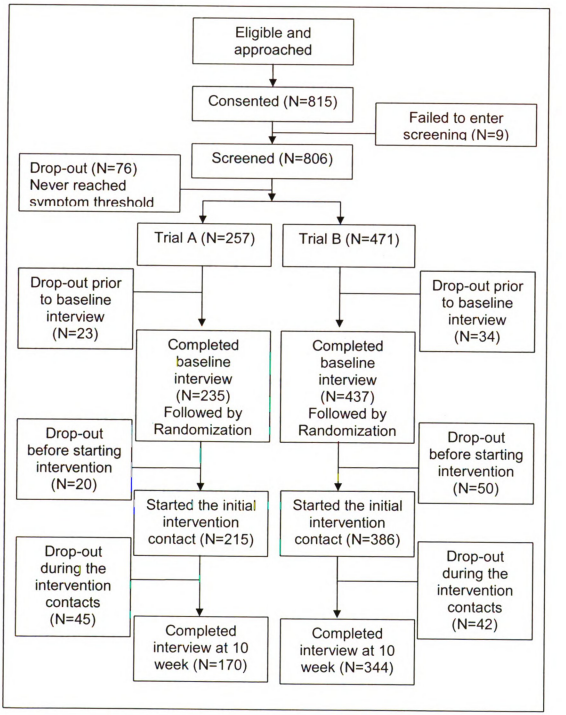
Second, symptom screening was performed to determine when a patient would enter into one of the two intervention trials. All enrolled patients underwent symptom screening, using the automated telephone response, during the same time period. Of the 815 recruited, 806 participants were screened to assure that they met minimum symptom severity criteria. The M.D Anderson symptom inventory (MDASI) was used for this process. This instrument is a widely

implemented instrument, designed to assess on a 0-10 scale the severity of multiple symptoms and the impact of symptoms on daily functioning in cancer patients (Appendix C). The reliability and validity of this instrument was demonstrated by Cleeland (34). Based on the MDASI, patients were asked to rate the severity of 13 cancer-related symptoms (i.e. pain, fatigue, nausea, disturbed sleep, distressed, shortness of breath, remembering things, lack of appetite, drowsy, dry mouth, sad, vomiting, and number or tingling), and to describe the impact on 6 items of symptom-related interference with daily functioning (i.e. general activity, mood, work around house, relations with other people, walking, and enjoyment of life) during the last 24 hours. In this screening, an automated voice response version of MDASI (4) called all patients at home twice weekly for up to six weeks. Patients with caregivers were entered into Trial A, when they first scored a 2 or higher on both pain and fatigue or a 3 or higher on pain or fatigue. Patients, who did not have caregiver or whose caregivers did not agree to participate, were entered into Trial B, when they first scored a 2 or higher on at least one symptom. Based on this screening, 257 and 417 patients were entered into Trial A and B, respectively. During the screening, 76 patients dropped out and were never entered into the intervention phase. Only 2 patients never reached the symptom threshold for either of the two trials.

Third, the trial participants completed a baseline interview a week after they were screened and entered. At the baseline interview, patients reported their symptoms, but were not given any intervention. Among them, 235 and 437 patients completed their baseline interviews in Trial A and B, respectively. Prior

to the baseline interview 23 and 34 patients dropped out from Trial A and B, respectively. After completing the baseline interview, participants in Trial A were randomized to a nurse or non-nurse (Coach), and in Trial B they were randomized to a nurse or automated-voice-response (AVR) system. Following randomization, intervention groups received 6 intervention contacts during an 8 week intervention period (See section 4.1.2 for further details). At each intervention contact patients reported symptoms and received interventions, if symptom severity reached a 4 or higher at the first intervention contact a week later. Figure 6 summarizes the flow of patients beginning with eligibility and proceeding through both trials including lost follow-up, as it occurred after the initial contact point.

Figure 6 Flowchart of patient accrual and retention in the Trial A and B



4.1.2 Intervention Trials for Symptom Management

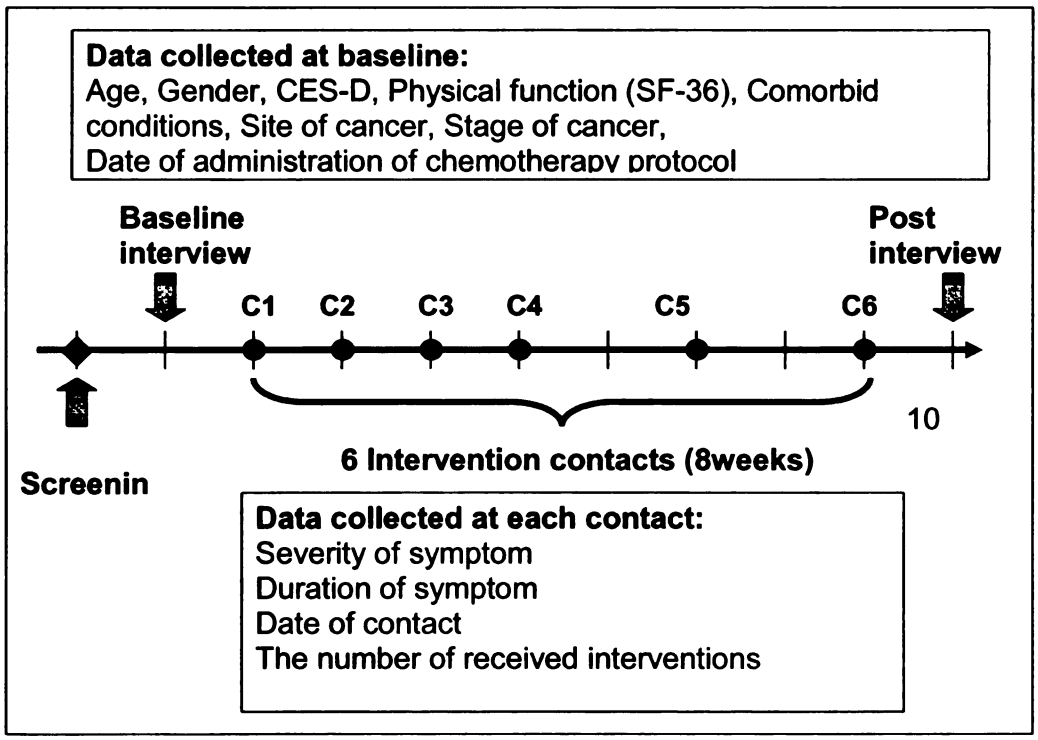
Trial A was designed to compare a nurse-based intervention with an intervention delivered by a non-nurse coach. In Trial B, the nurse-based arm was compared with an Automated Voice Response (AVR) system. Each trial was designed to compare the impact of intervention strategies in reducing the severity of symptoms for cancer patients undergoing chemotherapy. A nurse, coach, or the AVR system reached patients by telephone at pre-scheduled time points. The first four contacts were scheduled weekly, while the 5th and 6th contacts were every other week (Figure 7). At each intervention contact, the patients were asked about each of 16 symptoms[7], which included 12 symptoms (without drowsy) from the 13 symptoms at screening, as well as additional 5 symptoms (i.e. alopecia, constipation, cough, diarrhea, and weakness). In the intervention contact, nausea and vomiting were measured as one symptom.

In the nurse-based intervention of both Trial A and B, the trained nurse provided the patients with cognitive behavioral interventions that involved up to four strategies for dealing with each of their symptoms. The goal of this cognitive behavioral intervention was to assist patients in acquiring self-management knowledge, skills, and behaviors to address symptom problems. Cognitive behavioral strategies were improved to reduce the severity of symptoms, the impact on emotional distress, and the physical function. In this nurse-based arm, if a patient reported a symptom severity of 4 or higher, the nurse, with agreement from the patient, selected one of the 4 cognitive behavioral intervention strategies

[7] The 16 symptoms include alopecia, anxiety, constipation, cough, depression, diarrhea, difficult to breath, dry mouth, fatigue, insomnia, nausea/vomiting, pain, peripheral neuropathy, poor appetite, difficult to remember, and weakness.

classified as coping, reframing, education, or eliciting support to overcome the symptom. Each patient used a Symptom Management Guide (SMG) which described the cause of each symptom and strategies for managing each symptom. This manual, written at an 8th grade level, also included information on when to call the oncologist and how to access other sources of information. The patients in the non-nurse coach and the AVR arms were referred only to the SMG, when they reported symptoms at a severity of 4 or higher.

Figure 7 Time line of screening, baseline interview, intervention contacts, and post interview



4.1.3 Justifications for combining data from Trial A and B

In this study, I combined the data sets from Trial A and B into a single analytic file in order to achieve a larger sample size. While combining the two data sets can leads to a selection bias, I believe this possibility is minimal, and the differences between the two trials are as follows:

First, Trial A required consent from caregivers, whereas Trial B did not require caregiver participation. Since patients who had impairment in their physical functioning or who had adverse health conditions might be more likely to need assistance from caregivers, there is the potential bias that patients in Trial A might have poorer health compared with those patients in Trial B. Second, Trial A had somewhat different criteria for the symptom screening than Trial B. When eligible patients (i.e. those not having a caregiver) reported 2 or higher on severity of at least one of the 13 symptoms, they were entered into Trial B. In Trial A, however, patients were screened based on only pain and fatigue. For Trial A, patients needed to score 2 or higher on severity for both pain and fatigue, or 3 or higher on either pain or fatigue. Accordingly, patients in Trial A might have a greater opportunity to report pain and fatigue at early intervention contacts, compared with those in Trial B. However, despite differences in criteria for the presence of pain and fatigue, the prevalence of pain and fatigue was not significantly different between the two trials. Ninety-six and ninety-four percent of the patients reported a 4 or higher on severity of fatigue in Trial A and B, respectively. Prevalence of pain in Trial A (54.6%) was a little higher than that in Trial B (42.8%). This difference no longer existed among the analyzed samples,

because the survival analysis of time to response included only patients who had pain or fatigue during intervention period.

Regardless of the design difference between the two trials, there is evidence to support combining these data sets in this study.

First, Trial A and B were performed concurrently with the same study protocol. Patients were screened using the same system. The same nurse delivered the intervention protocols in nurse-assisted arms of the two trials. In the education-information arms, Trial A was delivered by a masters prepared social worker (non-nurse coach arm) and Trial B was delivered by an automated voice response system (AVR arm). While the mode was different, the protocol for these arms was identical in terms of the numbers of contacts, amount of time spent with each patient (approximately 18 minutes per contact), and the use of the same written materials to deliver symptom management strategies. The time frame from the recruitment to the end of intervention period was also identical between the two studies.

Second, the characteristics of the patients were not different between the two studies. Although more adverse health conditions were expected in patients with caregivers in Trial A, I did not find differences in comorbid conditions, depression, or age between the two trials (Table 1 in Chapter 5).

Third, the same instruments were used for assessing symptoms and other health conditions in both Trial A and B. Therefore, the two trials would be expected to have similar levels of measurement errors in symptom assessments.

Fourth, the different entry criteria for symptom screening did not affect the demographic difference between the two trials. Although Trial A had different criteria than Trial B, only two patients were excluded in screening because of the symptom severity requirements in both Trial A and B. (This in part was due to a fact that fatigue is the most prevalent symptom experienced by patients undergoing chemotherapy (43), affecting close to 80% of patients.)

As to the presence of caregivers in Trial A, there were remaining concerns about differences in gender and site of cancer between the two trials. More female and breast cancer patients were likely to be in Trial B compared with Trial A. It has been observed in a prior research that lung cancer patients experienced relatively severe pain and fatigue compared with patients with breast or other cancers (54). There is a question if as to whether differences in cancer type led to unbalanced patients health conditions at baseline between the two trials. Trial A has more Lung cancer patients (i.e. Breast (24.7%), Colon (7.4%), Lung (29.3%), Genitourinary (10.7%), Gastrointestinal (6.5%), Gynecological (6.5%), Pancreatic (2.3%), Non-Hodgkins lymphoma (7.0%), Myeloma (1.4%), and Other (4.6%)). Trial B has more breast cancer patients (Breast (41.5%), Colon (13.7%), Lung (15.0%), Genitourinary (6.9%) Gastrointestinal (3.6%), Gynecological (8.5%), Pancreatic (3.6%), Non-Hodgkins lymphoma (5.2%), Myeloma (1.0%), and Other (1.0%)).

At the baseline interview, there were no differences in comorbid conditions, total number of symptoms or depression. However, patients in Trial A had lower levels of physical functioning (SF-36) at baseline compared with those patients in

Trial B. To control for the baseline differences, the final models were adjusted for cancer site, sex, and physical functioning at baseline. If the effects of cancer site are different between the two trials, then the effect of cancer site could influence by the effects of intervention. For example, if the non-nurse coach intervention in Trial A is more effective for lung cancer patients, than the AVR intervention in Trial B, the cancer site effect can be overestimated, because there were more lung cancer patients in Trial A. To correct on this concern, the interaction between trial arm and selected covariates (cancer site, gender, and physical function) that differ by trial will be explored in multivariable analysis. Finally, by comparing final models for combined data with separate models for each trial, it will be shown that the identified factors associated with pain and fatigue response are not due to the differences between the two trials.

4.2 Measures

Symptoms were monitored at the baseline interview, at the each of the 6 intervention contacts (covering 8 weeks), and at the ten week (post-intervention) interview (Figure 7). At each contact, symptom management interventions were given to patients if they reported severity of 4 or higher among the 16 symptoms. The dates of the two interviews and 6 intervention contacts were recorded. Demographic data, depressive symptom, physical function, comorbidity were collected at the baseline interview. Site and stage of cancer were obtained from medical records.

Outcome Variable At each of 6 intervention contacts, severity of 16 symptoms [8] and 5 interference [9] items were recorded using a developed instrument based on Brief Pain Inventory (76) and Brief Fatigue Inventory (4) (Appendix A). Patients were asked to rate the average severity of pain and fatigue in the past 7 days on a scale of 0 (not present) to 10 (worst). When pain or fatigue was present (i.e. severity > 0), patients rated on a scale of 0 to 10 the extent to which that symptom interfered with their enjoyment of life, relations with others, general daily activities, emotions, and sleep.

The primary outcome variable for these trials is time-to-response in days for pain and fatigue. In order to evaluate how long it took to lower the severity of pain and fatigue by a delivered intervention, the date of contact on which a symptom was reported to have a severity of 4 or higher was considered the onset time of the symptom. When patients rated lower severity level (mild or moderate) compared to the initial severity level (moderate or severe) at onset that date of contact was defined as the response time.

Two different styles of records for time-to-response will be used in this study. First, most survival analysis methods[10] developed for right censored data require an actual time to response. Since exact time of onset and response could not be observed, the date of the scheduled contact was used as a proxy for the exact time of the events. That is, the number of dates from the contact where a

---

[8] The 16 symptoms includes pain, fatigue, nausea, insomnia, distress, difficult to breath, difficult to remember, poor appetite, dry mouth, vomiting, numbness or tingling, diarrhea, fever, cough, constipation, and weakness

[9] The 5 interference items include emotions, enjoyment of life, relations with others, general daily activities, and sleep.

[10] The implemented survival analysis methods for right censoring in this study includes the log-rank, Wilcoxon test, Cox proportional hazard model, Lin & Wang's test, Rahbar's methods, WCM, and marginal Cox model.

severity of pain or fatigue was 4 or higher to the contact where a response to the symptom was first identified was defined as the actual time to response. Second, a survival method developed for interval censored data (i.e. the Accelerated Failure Time model) requires an interval time for symptom response. Consider a hypothetical situation. Let us say a patient reported onset of pain at 1st contact and symptom response at 5th contact, and the actual response occurred between 4th and 5th contacts. Therefore, a lower bound of time to response should be the number of days from 1st to 4th contact, while the upper bound must be the number of days from 1st to 5th contact.

Independent Variables Patient depression was rated using the Center for Epidemiologic Studies – Depression (CES-D) scale (125). The CES-D instrument consists of 20-items rated on a 4 – point Likert-type scale (almost all of the time, most of the time, some of the time, and little/none of the time) (see Appendix D). The CES-D scale is a widely used reliable measure of depressive symptoms with an established cut-off of 16 or greater indicating the potential for clinical depression (126). Physical function was assessed using the 10-item physical function subscale from the Short Form 36 (SF-36), a widely used QOL instrument (see Appendix E). The subscale SF-36 has a standardized score of 0-100 scale, with higher scores representing better physical function. Patient comorbidity was derived from patient self-report using Katz instrument (127). To obtain comorbid conditions, patients were asked "Has a doctor ever told you that you have (a chronic disease)?" (See Appendix F). Comorbidity was measured by

the total number of 15 chronic conditions including high blood pressure, diabetes, other cancer, chronic bronchitis or emphysema, heart problem, angina or chest pains, stroke, emotional problems, arthritis or Rheumatism, fractured hip, surgical replacement of joint, incontinence, cataract surgery, hearing aid, and other major health problems at the intake interview. Patient comorbidity was dichotomized at the median into 0-2 versus 3 or higher categories.

Cancer site was categorized as breast, lung, colon, and others—where the pattern includes gastrointestinal, gynecological, genitourinary, pancreatic, non-Hodgkins lymphoma, myeloma, and others. Using information in the medical record, stage of cancer was categorized according to the tumor-node-metastasis staging criteria of the American Joint Committee on Cancer. Based on this scale, stage of cancer was collapsed into early stage (in situ zero, and stage I and II) and late stage (stage III and IV).

4.3 Data Analysis for Identifying Factors Associated Time to Pain and Fatigue Response

Data analysis was performed for developing optimal cut-points to define the level of pain and fatigue severity (mild, moderate, and severe) and testing consistency of the cut-points. The shifts ('severe to moderate', 'severe to mild', and 'moderate to mild') were defined as symptom response. Time-to-response was measured by the number of days from the onset contact to the response contact. To identify important factors impacting time to response of pain and

fatigue, several analysis methods were implemented. First, with the proportional hazard assumption, the log-rank, Wilcoxon test, and Cox proportional model were used. Second, the results from the above methods were confirmed by using the Lin & Wang's Test, WCM, and Rahbar's methods without the proportional hazard assumption. Third, using the lower and upper bounds of time to response, the Accelerated Failure Time (AFT) model was applied. Finally, I tested, using a marginal Cox model, to determine if the identified factors in the previous model were still significant, given the correlation between pain and fatigue symptoms within a same patient.

4.3.1 Developing Cut-points to Define the Severity Level of Pain and Fatigue

The first methodological aim of this study is to measure clinically meaningful changes of pain and fatigue in response to interventions. I developed optimal cut-points of pain and fatigue severity based on multiple interference items[11] for each symptom.

To use these four interference items (i.e. emotions, enjoyment of life, relations with others, and general daily activities) for finding cut-points of symptom severity, an internally consistent reliable interference scale for these multiple interference items was needed. Patients who reported pain and/or fatigue at a one or higher on the severity measure were then asked to report interference with daily activity, emotions, enjoyment of life, and relations with

[11] The multiple interference items include emotions, enjoyment of life, relations with others, and general daily activities.

others on the same 0-10 scale with 0 being not present and 10 worst possible. According to an explanatory factor analysis (123), the four interference items were highly and positively correlated with each other. The eigenvalues (λ) and the loadings of the four interference items on the first factor are listed in Table 1.The results support the fact that the four interference items are expected by one dimension and that a single summed score can be used to measure interference. Also the factor analysis with four items revealed that a single factor explained about 80% of the total variance for both pain-related interference (λ=3.27) and fatigue-related interference (λ=3.08). All four items in both cases had approximately equal (high) loadings on a single factor with the largest eigenvalue. Therefore, a single measure was used in this study by adding the four interference scores (0 to 40 scale). Using a single summed interference score to reflect the reactive dimension of symptom experience has been shown to be valid and reliable in other cancer studies (34).

Table 1 Factor loadings for the four interference items for each symptom at contact 1

	Pain	**Fatigue**
Interference with enjoyment of life	0.9138	0.9044
Interference with relations with others	0.8960	0.8558
Interference with general daily activities	0.9058	0.8885
Interference with emotions	0.9004	0.8619

To identify and differentiate mild, moderate, and severe categories of symptom severity, the optimal cut-points were selected by testing the difference in the summed interference score. In previous studies (4, 80, 92, 128),

Multivariate analysis of variance (MANOVA) was used to determine cut-points categorizing severity levels by maximizing symptom related outcomes (e.g. interference, emotional status, other functional status). In this study, since the distribution of interference sum skewed to the right side which is not a normal distribution, I used the generalized linear model, Let $1<c_1<c_2\leq 10$, where c_1 and c_2 are severity of pain or fatigue at the first contact. Since the distribution of interference sum has a gamma distribution, a generalized linear model with gamma distributed errors (129) was used as following;

$$\log(\mu)=\beta_0+\beta_1 S_1+\beta_2 S_2+\beta_3 T, \qquad (4.1)$$

where μ is the mean of summed interference scores; S_1 and S_2 are dummy variables for severe and moderate severity[12]; and T is total number of other symptoms at the first contact. This model was repeated 36 times with all possible combinations of c_1 and c_2 for each of pain and fatigue. If c_1 is 2, then there are 8 possible points for c_2 between 3 and 10. In the same way, 7, 6, 5, 4, 3, 2, and 1 points are available for c_2 when c_1 is 3, 4, 5, 6, 7, 8, and 9, respectively. This model was implemented via PROC GENMOD in SAS version 9.1(130). By seeking the largest size of the likelihood ratio (LR) of generalized linear model across all possible cut-points (c_1, c_2) for severity categories, optimal cut-points (c_1^*, c_2^*) separating moderate from mild and severe from moderate were selected.

---

[12] S_1 (severe) is 1 if $c_2 \leq$ severity, otherwise 0, and S_2 (moderate) is 1 if $c_1 \leq$ severity$< c_2$, otherwise 0

The distribution of the summed interference score could be influenced by other symptoms that were not included in the original four interference items. For instance, if the sample is divided into two groups according to the number of symptoms other than pain and fatigue, the distribution of the summed interference score in one group with a small number of symptoms is likely to be different from the distribution of the summed interference score in the other group with a large number of symptoms. Positive correlation coefficients between total number of other symptoms and the summed interference scores were observed for both pain ($r=0.3440$, p-value$<.0001$) and fatigue ($r=0.4427$, p-value$<.0001$). Therefore, the total number of other symptoms was adjusted for in the generalized linear models.

To provide additional evidence of validity of the cut-points, the summed interference scores, the duration of a symptom in the past 7 days, and the physical function (SF-36) were compared by mild, moderate, severe levels of pain and fatigue based on the selected cut-points.

4.3.2 Consistent Discrimination of Interference by the Cut-points

To use the selected cut-points defined above for measuring symptom change, the cut-points should reliably differentiate at every intervention contact. Therefore, I examined if the mild, moderate, and severe categories of symptom consistently differentiate the interference sums across 6 contacts for pain and fatigue. Since symptoms and interferences were repeatedly measured during multiple contacts, a longitudinal model should be used. By including the

interaction between time and severity categories (mild, moderate, severe), the means of the sums for the 4 interference score for three categories were compared at each contact.

For this analysis, the linear mixed effects (LME) models (131) with autoregressive covariance structure were employed. The LME model accommodates multiple missing data points in a longitudinal study. This consideration was important since several patients did not complete all 6 contacts due to missed contacts or because they dropped out. Further, the LME model incorporates correlations between contacts within the same patient. Let c_{1t}^{*} and c_{2t}^{*} be the selected cut-points of severity of pain or fatigue at the first contact. The model incorporating the interaction between the symptom severity categories (mild, moderate, severe) and time (contact number) was specified as follows;

$$\begin{aligned} y_t &= \beta_0 + \beta_1 S_{1t} + \beta_2 S_{2t} + \beta_3 T \\ &+ \gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_4 I_4 + \gamma_5 I_5 \\ &+ \delta_{11} S_{1t} I_1 + \delta_{12} S_{1t} I_2 + \delta_{13} S_{1t} I_3 + \delta_{14} S_{1t} I_4 + \delta_{15} S_{1t} I_5 \\ &+ \delta_{21} S_{2t} I_1 + \delta_{22} S_{2t} I_2 + \delta_{23} S_{2t} I_3 + \delta_{24} S_{2t} I_4 + \delta_{25} S_{2t} I_5 \end{aligned} \tag{4.2}$$

where y_t is the mean of summed interference scores at tth contact, S_{1t} and S_{2t} are dummy variables for severe and moderate severity at tth contact[13], T is total number of other symptoms at the first contact, $I_1,\ldots,I_5$ are dummy

[13] S_{1t} (severe) is 1 if c_{2t} ≤ severity, otherwise 0, and S_{2t} (moderate) is 1 if c_{1t} ≤ severity< c_{2t}, otherwise 0

variables[14] for 6 contacts. Parameters β_i s and γ_j s are coefficients for main effects of cut-points (S_{1t}, S_{2t}) and time ($I_1,\ldots,I_5$), respectively. Parameters δ_{ij} s are coefficients for interaction between cut-points and time. The LME model were fitted using PROC MIXED in SAS version 9.1 (130).

Least square (LS) means of the summed interference scores by the interaction terms were calculated from the model. The LS mean of the summed interference scores for mild, moderate, and severe symptom at tth contact can be estimated by $\hat{\beta}_0+\hat{\beta}_3\bar{T}+\hat{\gamma}_t$, $\hat{\beta}_0+\hat{\beta}_3\bar{T}+\hat{\gamma}_t+(\hat{\beta}_1+\hat{\delta}_{1t})$, and $\hat{\beta}_0+\hat{\beta}_3\bar{T}+\hat{\gamma}_t+(\hat{\beta}_2+\hat{\delta}_{2t})$, respectively, where $\bar{T}$ is the mean number of symptoms. To test if the mild, moderate, and severe symptoms have different summed interference score at t th contact, I tested a null hypothesis: $\hat{\beta}_1+\hat{\delta}_{1t}=\hat{\beta}_2+\hat{\delta}_{2t}=0$. If the null hypothesis is rejected for all 6 contacts, then the categories of mild, moderate, and severe may consistently differentiate the summed interference score over time.

Once consistent differentiation of interference by the cut-points is confirmed, symptom response can be defined by these cut-points as mentioned above. Finally time to response for pain and fatigue can be measured by counting the number of days from onset to response (See Outcome Variable section in Chapter 4.2 Measures).

---

[14] I_t is 1 if a patient was in t^{th} contact, otherwise 0, where $t=1,2,\ldots,5$

4.3.3 Evaluating Time to Response of Pain and Fatigue

To identify significant factors influencing time to response in the management of pain and fatigue, the combined data sets from Trial A and B were used. Since there were differences in gender, cancer site, and physical function at baseline between the two trials, the factors were tested after adjusting for these three covariates in multivariable analysis. In this analysis, all covariates were categorized because some tests are available only for categorized variables. To avoid multicolinearity between gender and cancer site, I created a combined variable, categorizing as female & Breast cancer, female & Lung cancer, female & other cancer, male & lung cancer, and male & other cancer. Age (60 or younger vs. older than 60 years old) and comorbidity (less than 3 vs. 3 or more comorbid conditions) were categorized using the median values. Since CES-D of 16 was considered as an important clinical threshold (126, 132). We used this value to create a binary variable.

The analyses were performed in the following steps: 1) Univariate analysis tested the effects of each of the covariates, including age, comorbidity, and depression without adjusting for any other covariate. 2) Multivariable models tested the effects of those covariates after adjusting for gender, cancer site, physical function, and trials (Trial A vs. B). The model selecting methods are described in next paragraph. 3) I tested the interaction between trials (Trial A and B) and the covariates.

4.3.4 Survival Analysis Incorporating Right-censoring

The first modeling strategy is to evaluate the time-to-response under the proportional hazard assumption by using the log-rank test, the Wilcoxon test, and the Cox proportional model. PROC LIFETEST (for the log-rank test and the Wilcoxon test: Univariate model) and PROC PHREG (for the Cox proportional hazard model: Multivariable model) procedures, SAS version 9.1, were implemented. The hazard ratio was presented to measure the effect of a covariate on time to symptom response.

Although these methods are most commonly used in survival analysis methods, they are valid only when the proportional hazard assumption is satisfied. Therefore, it is important to check the proportionality of hazards. To test the proportionality of hazards for each covariate, I performed the following steps (105); 1) created a non-decreasing log function of time ($g(t) = \log(t/t^*)$, where t is time to response and t^* is the median time to response, 2) added the interaction term between a covariate and the log function of time into the Cox proportional model given by $h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_1^* x_{i1}\log(t/t^*))$ for each covariate x_1, and 3) tested if the coefficient of interaction term, β_1^*, is statistically different from zero. If β_1^* is not zero (i.e. the term is $p<0.05$), then the covariate, x_1, has non-proportional hazard across time. This test was performed for both univariate and multivariable analysis.

The second modeling strategy is to evaluate time to response when non-proportional hazards exist by using the Lin & Wang's test, Rahbar's approach, and the WCM. The first modeling strategy that is based on the proportional hazard assumption might not appropriately detect the effect of a covariate on time-to-response, if the hazard functions are not proportional by the covariate. Therefore, non-proportional hazard based methods may have more valid results, since these methods do not require the proportionality of hazards. These methods are also available, when the proportional hazard assumption is satisfied. I investigated how the results of this modeling strategy are different from that of the first modeling strategy. Since the Lin & Wang's test is available only for two groups and Rahbar's approach is restricted to two covariates, these methods were performed for unadjusted analyses only. Only the WCM was used for a final multivariable model. For the Lin & Wang's test and Rahbar's methods, I developed SAS macros (See Appendix B) and a SAS macro developed by Heinze and his colleagues was used for the WCM.

4.3.5 Survival Analysis Incorporating Interval-censoring

The first and second modeling strategies can be used for right censored data, but these modeling strategies are not available when time to response is observed as an interval of time $[L, R]$, where $L \leq$ actual time to response $\leq R$. The third modeling strategy is to evaluate time to response when it was observed as an interval of time via the AFT model with a Weibull distribution. Since symptoms were monitored at the scheduled intervention contacts in Trial A and B,

actual symptom response occurred between the contact periods. Therefore, the ATF model with interval censoring would be more appropriate method for assessing this interval censored data. It would show the difference between methods with right censored[15] and interval censored[16] time data. Because the AFT model is a parametric model, an appropriate distribution of time-to-response needs to be selected. The Akaike Information Criterion (AIC) can be used for selecting a distribution in the AFT model (105). I compared four distributions: the Weibull, exponential, log-normal, and log-logistic distributions using the AIC. The largest AIC indicates that the best model for both pain and fatigue was based on the Weibull distributions. PROC LIFEREG, SAS version 9.1, will be used for the AFT model. Results from the AFT model will be compared that from the Cox proportional hazard model and the Cox model with weighted estimation.

4.3.6 Survival Analysis for Multiple Events Incorporating Right-censoring

For the above modeling strategies, the survival analysis models were performed separately for pain and fatigue. If time to response for pain were independent from that of fatigue, these separate models for pain and fatigue would be appropriate. The fourth modeling strategy is to identify the common effects of covariates on both pain and fatigue, while taking into account for a correlation between the two symptoms via the marginal Cox model. I produced a final multivariable model by using PROC PHREG procedure with the aggregate

[15] The survival analysis methods for right censoring includes the log-rank test, the Wilcoxon test, the Cox proportional hazard model, the Lin & Wang's test, Rahbar's methods, and the WCM.
[16] The survival analysis method for interval censoring includes the AFT model.

option in SAS version 9.1. The aggregate option was used to produce the empirical covariate matrix estimator which accounts for the correlation between pain and fatigue within an individual. Separate covariate effects for pain (β_{pain}) and fatigue ($\beta_{fatigue}$) were included this model. In the equality test ($H_0: \beta_{pain} = \beta_{fatigue}$) based on the empirical covariate matrix estimator, if the null hypothesis (H_0) is not rejected, then these two parameters can be replaced by a single parameter for the overall effect (β_{all}).

4.3.7 Diagnostic assessment of the final models

After final multivariable models were built, the adequacy of the model was assessed. In survival analysis, a standardized method of model diagnosis is to use residuals which are different from that in the linear regression model. A residual called the Cox-Snell residual (133) can be used to check the adequacy of the model. The basic idea of this diagnostic test is that a negative log in the survival function ($-\log(S(t))$) has an exponential distribution when survival time t is a continuous random variable. Because a survival function $S(t)$ has a uniform distribution between 0 to 1, $W = -\log(S(t))$ has an exponential distribution. If the correct model has been fitted, the Cox-Snell residual ($-\log(\hat{S}(t))$) has an exponential distribution. When a cumulative hazard of the Cox-Snell residuals are plotted against the Cox-Snell residuals, a straight line with a slop of 1 and zero intercept will indicate that the final model is adequate (105).

By seeking the significant covariates from the different modeling strategies, I identified covariates that all consistently recognized as being important factors for time to response of pain and fatigue. Different results using several other approaches will be discussed by considering their underlying assumptions

CHAPTER 5 RESULTS

This chapter presents the results from the evaluation of pain and fatigue specially the establishment of optimal cut-points for measuring the severity of pain and fatigue, and then the application of several different statistical methods for identifying factors associated with the responses to pain and fatigue. The analyses were performed to answer two major questions: first, how to define clinically meaningful changes in the response to symptom management interventions for pain and fatigue; and, second, how to analyze the time-to-response of pain and fatigue without violating the model assumptions regarding proportional hazards.

5.1 Demographic and clinical characteristics of the trials

Table 2 summarizes the characteristics of 601 patients in the two intervention trials. More women were assigned to Trial B (76.2%) than Trial A (59.1%) (p-value<0.0001). There was no difference in the age distribution between the two trials, however, differences in the distribution of cancer sites was observed between the two trials (p-value<0.0001). More breast-cancer cases were assigned to Trial B (41 vs. 24.7%) while more lung-cancer cases were assigned to Trial A (27.3 vs. 15.0%) (See Table 2). More patients had late stage cancer in Trial A (90.1%) than Trial B (82.3%), which is a reflection of the higher proportion of lung cancer in Trial A which is usually diagnosed later.

Based on the CES-D score, approximately 35% of patients qualified as being at risk for clinical depression (i.e. CES-D≥16) in either trial. There was no difference between Trial A and B in CES-D score (p-value=0.4381). There was also no difference in the number of comorbid conditions between Trial A and B (p-value=0.7907). Around 60% of patients reported two or less comorbid conditions.

Table 2 Characteristics of patients in two intervention trials

	Trial A (N=215) N (%)	Trial B (N=386) N (%)	P-values (Chi-square Test)
Patient gender			
Male	88 (40.9)	92 (23.8)	<.0001
Female	127 (59.1)	294 (76.2)	
Patient age			
25 ~ 44	24 (11.2)	57 (14.8)	
45 ~ 60	102 (47.4)	178 (46.2)	0.4520
61 ~ 74	70 (32.6)	113 (29.3)	
75 +	19 (8.8)	37 (9.6)	
Stage of disease			
Early	21 (9.9)	68 (17.7)	0.0112
Late	190 (90.1)	316 (82.3)	
Site of cancer			
Breast	53 (24.7)	160 (41.5)	
Lung	63 (29.3)	58 (15.0)	<.0001
Other	99 (46.0)	115 (43.5)	
CES-D			
Less than 16	137 (64.3)	257 (67.4)	0.4381
16 +	76 (35.7)	124 (32.6)	
Comorbidity			
Less than 3	133 (61.9)	243 (62.9)	0.7907
3 +	82 (38.1)	143 (37.1)	

5.2 Meaningful Reductions in Pain and Fatigue among Cancer Patients

To obtain a reliable definition of symptom response, the optimal cut-points of pain and fatigue will be identified in Section 5.2.1. We will determine if the set of cut-points consistently differentiate among patient interference over the 6 contacts in Section 5.2.2.

5.2.1 Optimal Cut-points of Pain and Fatigue Severity

To establish the optimal cut-points for differentiating patient interferences due to pain and fatigue, I explored how the severity of pain and fatigue are distributed across interference scores at the initial contact. Four of the interference items[17] were highly and positively correlated. We found that sleep was weakly correlated with the other four items and thus was removed from the final summed interference scale. The remaining four items were submitted to an exploratory factor analysis (See Table 1, Chapter 4). As severity of symptoms increased, the sum of interference score consistently increased. Optimal cut-points for severity need to be based on clinically important distinctions in summed interference scores.

Therefore, to establish possible cut-points a generalized linear model with a gamma distribution function (See Equation 4.1) was used to adjust for the total number of other symptoms at the first intervention contact. Since the models were repeated 36 times for all possible combinations of two cut-points (See

[17] The 4 interference items include emotions, enjoyment of life, relations with others, and general daily activities.

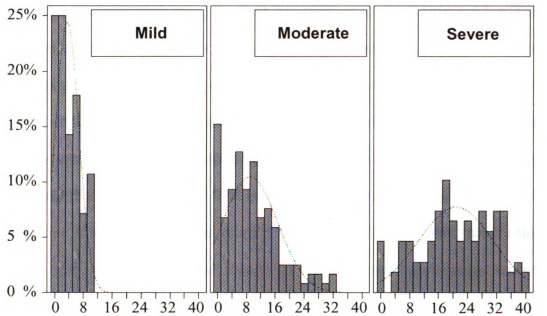
Chapter 4.3), 36 chi-square (Wald test) statistics were produced. The largest chi-square statistics for five pairs of two cut-points are listed in Table 3. Larger chi-square statistic represents better differentiation of interference, and the cut-points of 2 (for moderate) and 5 (for severe) produced the largest chi-square statistics in both pain and fatigue. Based on the identified cut-points, the severity levels, therefore, are defined as "mild" (score of 1), "moderate" (scores of 2 through 4), and "severe" (score of 5 and greater).

Table 3 Cut-points with five largest chi-square (Wald test) statistics

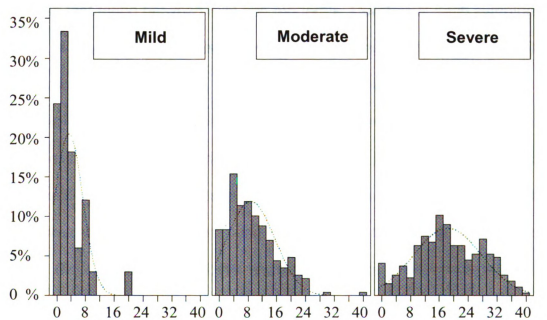
Rank	Pain		Fatigue	
	Cut-points	Chi-square	Cut-points	Chi-square
1	(2, 5)	92.35	(2, 5)	151.68
2	(2, 6)	82.13	(3, 5)	144.80
3	(2, 7)	81.57	(5, 9)	139.11
4	(5, 7)	75.57	(5, 7)	137.23
5	(3, 5)	75.20	(5, 8)	136.80

To ensure how appropriately the selected cut-points represent patient symptom burden, the distributions of interference sums for mild, moderate, and severe levels of pain and fatigue are compared in Figure 8. The histograms represent the distribution of interference scores among patients whose symptoms were classified as mild, moderate, and severe. Interference sums are distributed around 3 (median) at the mild level, while they are distributed around 8 and 20 (median) at the moderate and severe level, respectively.

Figure 8 Distribution of summed interference score (0~40) among mild, moderate, and severe level of pain and fatigue



(a) Distribution of interference sum by level of pain



(b) Distribution of interference sum by level of fatigue

In addition, Table 4 summarizes the cut-points for pain and fatigue. The means for the summed interference scores, duration of the symptom over the past seven days prior to delivering the first intervention, patient physical function (SF-36), and sum of severities of the 14 other symptoms assessed at the baseline interview are compared across the three severity levels for pain and fatigue, respectively. Among patients who reported pain or fatigue at the first contact, 42% and 50% were classified by severe pain and fatigue, respectively. Only 12% and 6% of patients had mild pain or fatigue, respectively. The means of the summed interference scores were 3.4, 9.1, and 20.8 for mild, moderate, and severe pain, respectively. The means of the summed interference scores were 3.4, 8.8, and 18.3 for mild, moderate, and severe fatigue, respectively. Greater differences in interference scores between moderate and severe are observed than that found between mild and moderate. Duration of each symptom (the number of days in the past week in which patients reported pain or fatigue), physical functioning score (SF-36), and the sum of symptom severities (i.e. sum of 14 possible other symptoms) exhibited greater differences between mild and moderate levels of severity for pain and fatigue than that between moderate and severe for each symptom. In the case of pain, there is no difference in physical function scores between moderate and severe severity. These differences indicate that the established cut-points represent not only interference with the patient's lives but also reflect the duration and the impact of pain and fatigue on their levels of physical function. The difference in physical function scores and the sum of symptom severities among patients classified as having mild,

moderate, and severe symptoms provide evidence that the established cut-points reflect real differences in health conditions.

Table 4 Cut-points for severity categories, number of cases reporting each symptom, unadjusted mean interference scores, duration before first contact, and physical function at baseline interview

Severity level of pain		**Mild (1)**	**Moderate (2–4)**	**Severe (5–10)**
No. of patients	N (%)	32 (12.4)	119 (45.9)	108 (41.7)
Interference sum	Mean (Std)	3.4 (3.3)	9.1 (7.7)	20.8 (10.3)
Duration[18]	Mean (Std)	2.9 (2.5)	4.7 (2.2)	5.1 (2.0)
Physical function[19]	Mean (Std)	63.0 (23.9)	49.5 (25.6)	48.9 (27.8)
Symptom severity sum[20]	Mean (Std)	29.5 (18.4)	40.7 (19.5)	50.6 (23.8)
Severity level of fatigue		**Mild (1)**	**Moderate (2–4)**	**Severe (5–10)**
No. of patients	N (%)	34 (6.4)	227 (42.9)	268 (50.7)
Interference sum	Mean (Std)	3.4 (3.9)	8.8 (6.7)	18.3 (9.3)
Duration[19]	Mean (Std)	2.8 (2.0)	4.7 (2.1)	5.5 (1.9)
Physical function[20]	Mean (Std)	77.0 (21.6)	61.9 (24.6)	48.4 (26.4)
Symptom severity sum[21]	Mean (Std)	25.1 (22.2)	32.2 (18.1)	43.8 (21.7)

Taking into consideration the magnitudes of interference for pain and fatigue categories, and how they might influence symptom response to management interventions, the stability of these cut-points over time will be

[18] Duration is the number of days when patient experience pain or fatigue in past 7 days prior to the first contact.
[19] Physical function is a standardized score on a 0 to 100 scale from (SF-36 subscale).
[20] Symptom severity sum is the sum of severities for the 16 symptoms assessed at baseline.

evaluated in the next section. If differences among these cut points for symptom severity are stable over successive intervention contacts then they may be used to determine if the management interventions successfully reduce patient burden of pain and fatigue.

5.2.2 Longitudinal Consistency of Cut-points for Pain and Fatigue

The question under examination is whether, while receiving interventions to manage pain and fatigue, the established cut-points continue to represent significantly different interference levels among patients over time. To use these cut-points for measuring meaningful change in pain and fatigue it is necessary to establish that these cut-points consistently differentiate among different levels of interference during the intervention contacts. This section will examine the magnitude of the difference in summed interference scores between moderate and mild and severe and moderate categories of pain and fatigue at each of the 6 contacts.

The least square (LS) means of the summed interference scores reflecting mild, moderate, and severe scores from the linear mixed model (See Equation 4.2) are presented across 6 intervention contacts in Table 5. The mean summed interference scores for mild pain and fatigue are less than 5 and they are 15 or greater for severe pain and fatigue across all contacts. In addition, the examination of the magnitude of the differences between moderate and mild, and severe and moderate categories consistently show large differences across all contacts. Table 5 contains the p-values for testing the equality of the means of

summed interference ($H_0: \hat{\mu}_{mi,t} = \hat{\mu}_{mo,t} = \hat{\mu}_{se,t} = 0$)[21] among the mild, moderate, and severe categories at each intervention contact *t*. If the null hypothesis (H_0) is rejected, then the three categories can discriminate the summed interference scores at contact *t*.

According to the estimated LS means for the summed interference scores in Table 5, there are consistent differences in scores by severity level across contacts. The estimated interference means in each severity level are similar over time. For example, the interference mean in mild, moderate, and severe fatigue are around 4, 7.5, 15, respectively, at most contact points. These results indicate that there is a consistent and logical relationship between symptom level and interference with milder symptom having lower interference scores. The decline in the mean of summed interference score for the same severity category is greatest between the first and second contact; and then no further significant declines in interference occurred between contact 2 through 6. This pattern may be due to a difference in the absence of intervention between the first contact and the later contacts. At each contact, patients reported their symptom severity and interferences first; and then they received interventions based on the reported symptom severity. Therefore, the symptom severity and interferences were measured at the first contact before receiving any intervention. Thus it is possible that the declines observed between contacts 1 and 2 were a function of the initial interventions. Because patients could receive any interventions after

[21] $\hat{\mu}_{mi,t}$, $\hat{\mu}_{mo,t}$, and $\hat{\mu}_{se,t}$ are the LS means of summed interference for mild, moderate, severe categories at *t*-th contact

reporting their symptoms at the first contact, the intervention could affect the severity and interference from the second contact. From contact 2 through 6, the difference in interference among the mild, moderate, and severe categories remained stable and distinct. However, the weight of evidence provided here indicates a strong and sustained relationship between interference scores and their corresponding symptom severity categories at all contacts.

Table 5 Estimated means and standards errors of summed interference score across 6 intervention contacts by severity level of pain and fatigue*

	Summed Interference Score Mean (Standard Error)					
Pain	Contact 1	Contact 2	Contact 3	Contact 4	Contact 5	Contact 6
Mild	5.5 (3.1)	1.4 (1.8)	5.3 (2.0)	5.7 (2.0)	6.5 (2.1)	5.2 (2.4)
Moderate	10.0 (1.0)	8.1 (0.9)	9.9 (0.9)	8.7 (0.9)	9.3 (0.9)	8.3 (1.0)
Severe	19.5 (0.8)	14.9 (1.0)	16.6 (1.1)	17.2 (1.1)	14.2 (1.1)	16.5 (1.2)
P-value	<.0001	<.0001	<.0001	<.0001	0.0003	<.0001
Fatigue	Contact 1	Contact 2	Contact 3	Contact 4	Contact 5	Contact 6
Mild	2.3 (2.3)	4.8 (1.2)	3.8 (1.3)	4.2 (1.4)	4.1 (1.5)	2.9 (1.4)
Moderate	10.5 (0.6)	8.0 (0.5)	7.8 (0.5)	7.8 (0.5)	7.5 (0.5)	6.2 (0.5)
Severe	17.3 (0.4)	14.9 (0.5)	14.9 (0.6)	15.0 (0.6)	15.0 (0.6)	15.4 (0.7)
P-value	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

* Severity level is determined at each contact point

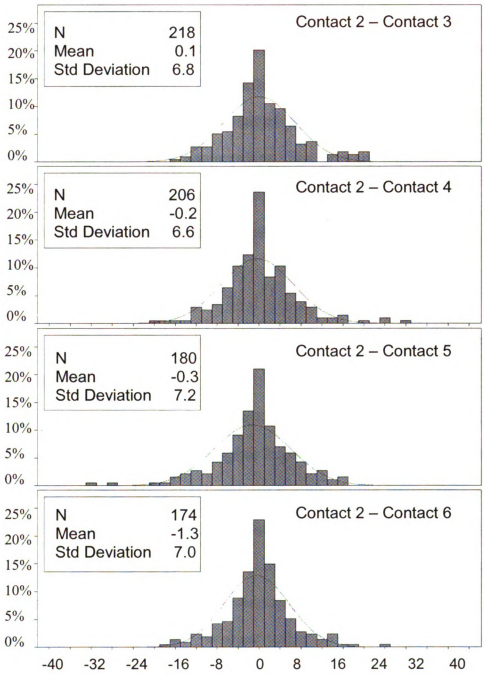
To use these categories (mild, moderate, severe) for evaluating change in pain and fatigue, it is important that individual variations in interference scores should be small, when patients remained in the same level of severity. I calculated the individual differences of summed interference scores between the

second contact and later contacts when a severity level remained the same. Because the association between severity and interference at contact 1 is different than that at later contacts, I compared the summed interference scores from contact 2 through 6. Figure 9 shows that distributions of differences between the summed interference score at the second contact and at later contacts within the same severity category (i.e. mild, moderate, or severe) over time. The graphs shown in Figure 9 were created from the data that is pooled across severity levels, however, there distributions were essentially the same when the data were stratified by severity. The individual differences of interference are symmetrically distributed around zero over time, when patients remain at the same severity level. These distributions show that a large number of patients report small differences in the summed interference scores at the same levels of severity. The paired t-test was performed to test the equality of summed interference scores between contacts at the same levels of severity for each symptom. The test did not reject the null hypothesis that the summed interference scores are equal at the same level of severity between contacts at 0.05 level of significant.

This consistent difference among mild, moderate, and severe categories supports the use of these interference based severity cut-points to define symptom response for three transitions: severe to moderate, severe to mild, and moderate to mild. The severity of zero (absence of symptoms) was not included in the mild category at each contact, because interference scores were not recorded when the symptom was not present. However, additional shifts

including severe (≥ 5) and moderate (2–4) to zero severity should be considered as meaningful change, and treated as a symptom response. Based on the defined symptom response categories from this section, I will identify important factors affecting time-to-response for pain and fatigue. This is presented in section 5.3. Several survival methods will be implemented to cope with the methodological problems discussed in Chapter 3.

Figure 9 Distributions of individual difference in summed interference scores (0-40) within the same level of severity for pain and fatigue[22]



[22] Data were generated only among patients who reported the mild, moderate, or severe categories at the second and each of later contacts

5.3 Evaluation of Time-to-Response for Pain and Fatigue Symptoms

Survival analysis is an appropriate technique to evaluate the time-to-response in pain and fatigue incorporating censored observations due to lost to follow-up. This section presents the results of unadjusted and adjusted analysis of the time-to-response for pain and fatigue symptoms using several methods each with different underlying assumptions. The primary exposure variables in this research are age, depression, and comorbidity. These variables were tested after adjusting for the following a priori confounders; gender, cancer site, physical function, and trial type (A or B).

5.3.1 Estimated Survival Functions for Time-to-Response for Pain and Fatigue

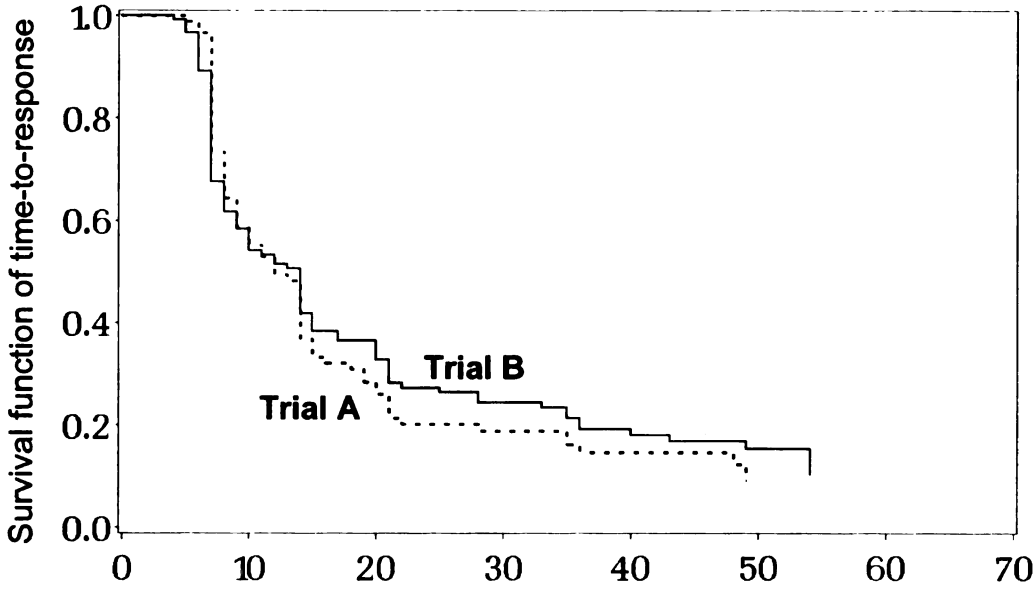
Among 601 patients who had the first contact, 212 (35%) and 451 (75%) patients reported a severity of 4 or higher in pain and fatigue sometime between the first and fifth contacts, respectively. By the end of the intervention contacts, 173 (81.6%) and 344 (76.3%) of these patients, achieved a response to pain and fatigue (i.e. severe to moderate, severe to mild, or moderate to mild), respectively. The number of days from the onset to response contact was recorded as time-to-response. The number of days from onset to last contact without a response was considered the censoring time.

Figure 10 shows the Kaplan-Meier survival function with 95% confidence intervals for time-to-response for pain and fatigue in Trial A (dash) and B (solid).

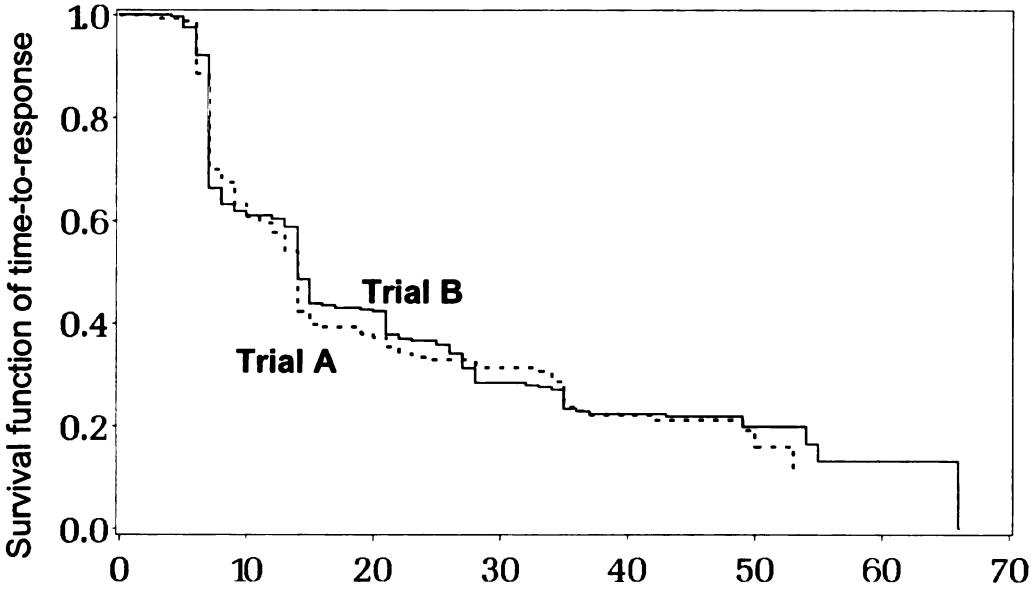
The survival functions for pain and fatigue are very similar between the two trials. Even though 6 intervention contacts were scheduled over 8 weeks, variations were observed. For example, some patients received intervention contacts for up to 66 days due to delays in the scheduled contacts. In Figure 10 the two survival functions for pain (panel a) and fatigue (panel b) have large declines around 1 week after onset. Based on the Kaplan Meier estimates, the median time-to-response for pain and fatigue are 13 and 14 days, respectively. That is, half of cases with pain and fatigue responded within 2 weeks. About 10% of pain cases were censored at the end of follow-up (i.e. these cases had not responded by the 6^{th} contact). Therefore, the overall mean for pain cannot be calculated, and therefore the restricted mean time[23] of 19.5 days was alternatively used for pain (See Figure 2 & 3 in Chapter 3.1.2). The estimated mean time-to-response in fatigue is 24.2 days.

[23] The restricted mean time is obtained by calculating an area under survival curve between zero time and the maximum time to response.

Figure 10 Estimated survival function of time to response of pain and fatigue[24]



(a) Time to response for pain in days



(b) Time to response for fatigue in days

[24] Response is defined as a reduction of severity levels (i.e. severe to moderate, severe to mild, or moderate to mild).

5.3.2 Testing for Time-to-Response by Assuming the Proportional Assumption Holds

This section identifies significant factors associated with time-to-response in pain and fatigue by assuming the proportional hazard assumption holds. The log-rank, Wilcoxon test, and Cox proportional hazard model were implemented to assess both unadjusted and adjusted effects. Patient age, depression, and comorbidity were first assessed without adjustment for other covariates in unadjusted analysis. In the multivariable model, each of the three covariates was tested after adjusting for gender, cancer site, physical function, and trial. Using the model selection strategy described in Chapter 4, a final model was developed.

Figures 11, 12, and 13 show the estimated survival functions for time-to-response for pain by age (60 or younger vs. older than 60), comorbidity (less than 3 comorbid conditions vs. 3 + comorbid conditions), and depression (CES-D < 16 vs. 16 +), respectively. In each of these figures, the estimated survival functions are less than 0.8 after about 7 days, indicating that more than 20% of subjects in pain responded within 7 days regardless of age, comorbidity, or depression. None of the survival functions cross one another in Figures 11, 12, and 13, supporting that the survival functions are likely to be proportional over time. To test the proportional hazard assumption, an interaction term between time and each covariate was tested in the Cox proportional hazard model. None of these tests were statistically significant and therefore the null hypothesis of proportional hazards for time-to-response in pain was not rejected (See Table 6). By holding the proportional hazard assumption, Table 6 summarizes tests for

equality in survival functions of time-to-response in pain and fatigue by patient characteristics using the log-rank test, Wilcoxon test, and Cox proportional hazard model. The median time-to-response and 95% confidence intervals (CI) calculated from the Kaplan-Meier survival functions are presented for each level of each covariate. All three methods identified age and comorbidity, as significant factors associated with time-to-response for pain. Patients who were older than 60 year old (Median of time-to-response=14 days, 95% CI=[14, 21]) had longer time-to-response for pain than younger patients (median of time-to-response =10 days, 95% CI=[9, 14]). Patients with 3 or more comorbid conditions (median time-to-response=18 days, 95% CI=[14, 21]) had a longer time-to-response for pain compared to those with fewer than three comorbid conditions (median time-to-response=9, days 95% CI=[8, 10]). Patients who reported CES-D of 16 or higher (median time-to-response=14 days, 95% CI=[10, 18]) had a little longer time-to-response for pain compared with those who reported a less than 16 on CES-D (median time-to-response=11 days, 95% CI=[8, 14]), but this depression effect was not significant in the log-rank test, Wilcoxon test, and the Cox proportional hazard model.

Figures 14, 15, and 16 show the estimated survival functions for time-to-response for fatigue by age, comorbidity, and depression, respectively. Fairly similar survival functions were observed by age in Figure 14. The survival functions by comorbidity categories (Figure 15) show that the median time to response is 14 days for both levels of comorbidity. After this time point (14 days), however, the difference between the two survival functions is greater than that

before the median time (14 days). The survival functions by depression almost overlap across time (Figure 16). Tests of the proportional hazard assumption using the interaction between time and a covariate conclude that all of covariates satisfied the proportional hazard assumption. Table 6 shows that comorbidity was the only factor associated with time-to-response in fatigue. Patients who were 60 years old or younger (median time-to-response=14 days, CI=[14, 15]) had same median time-to-response for fatigue compared with those who were older than 60 years old (median time-to-response=14 days, CI=[14, 21]). Patients with fewer numbers of comorbid conditions (median time-to-response=14, days CI=[13, 14]) had a shorter time-to-response compared with those with greater than three comorbid conditions (median time-to-response=15 days, CI=[14, 26]). The comorbidity effect was significant in all three statistical methods despite of the similarity of median time-to-response. Depression was not significantly associated with time-to-response for fatigue; median response times were 14 days for all categories of depression.

Figure 11 Estimated survival functions of time-to-response in pain by age

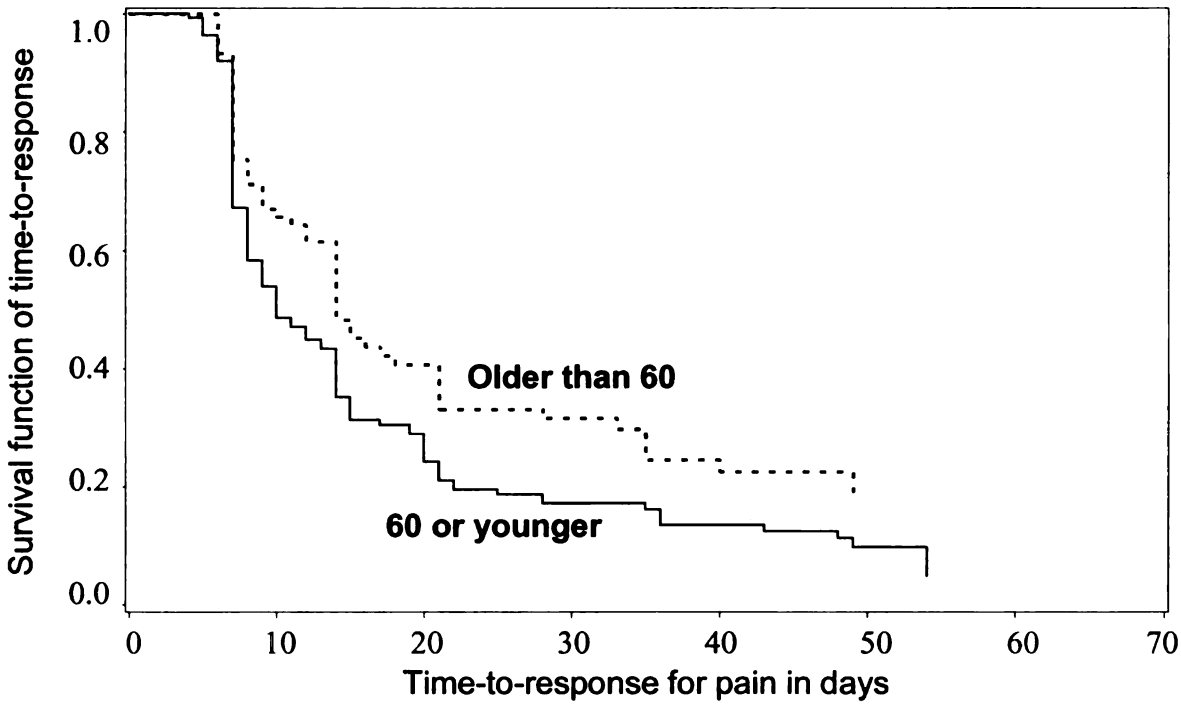


Figure 12 Estimated survival functions of time-to-response in pain by comorbidity

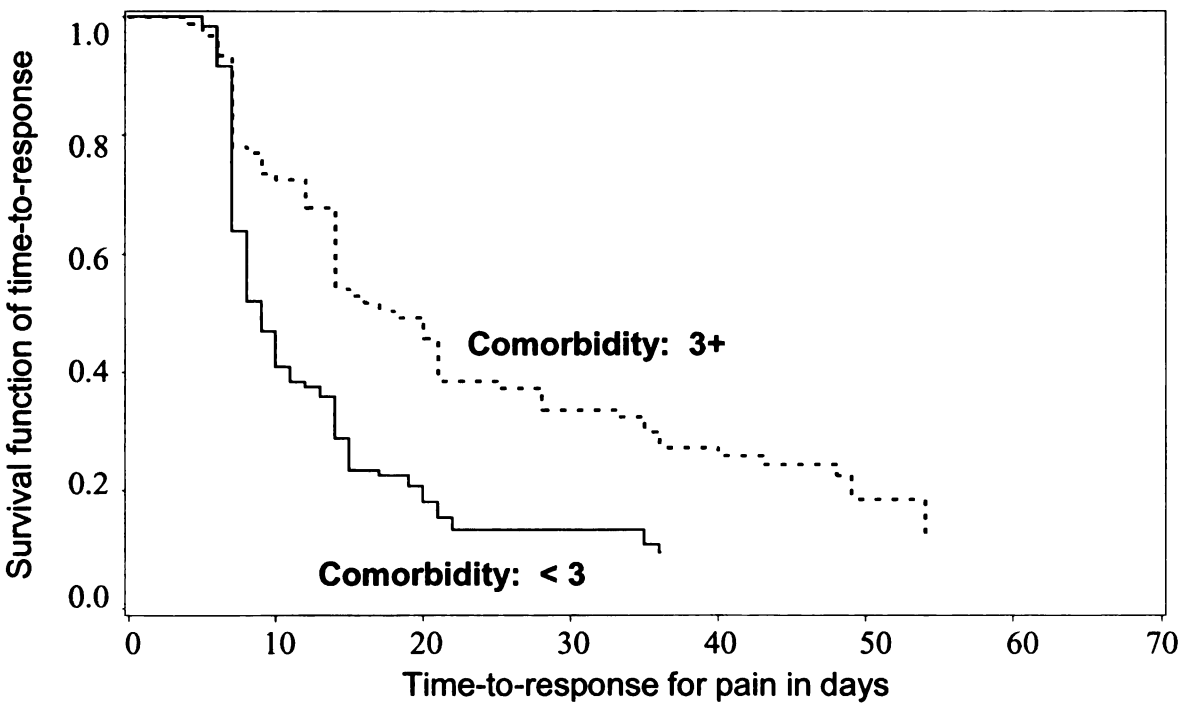


Figure 13 Estimated survival functions of time-to-response in pain by depression

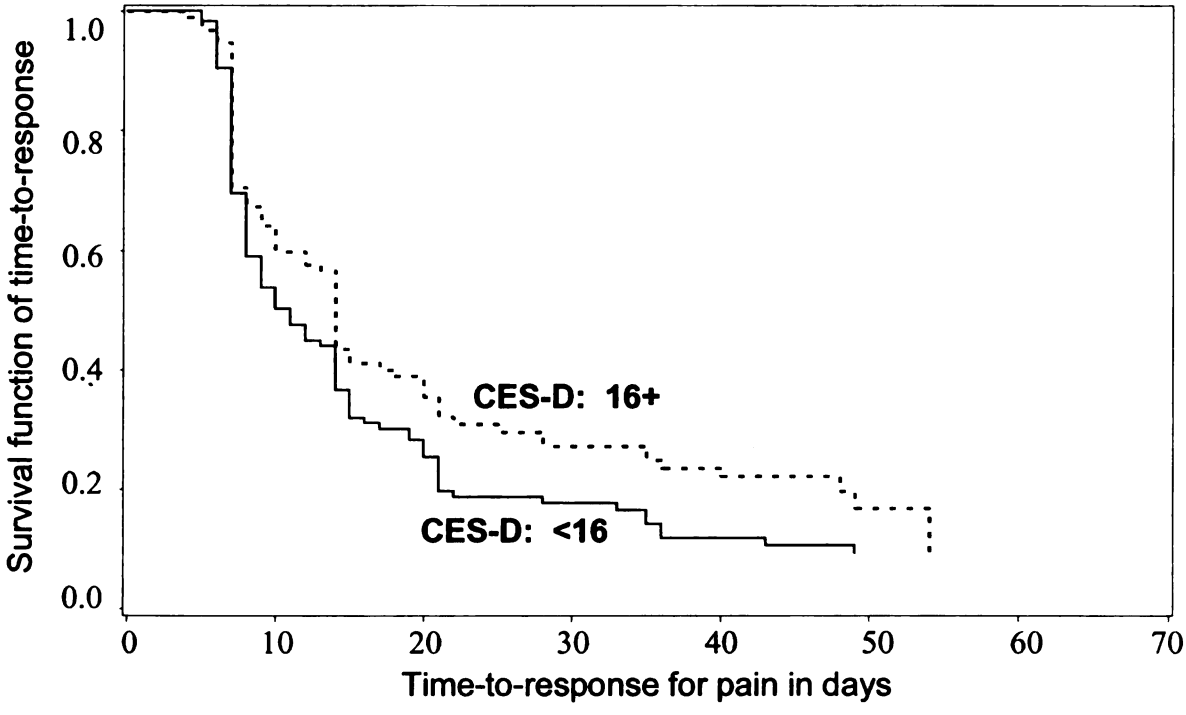


Figure 14 Estimated survival functions of time-to-response in fatigue by age

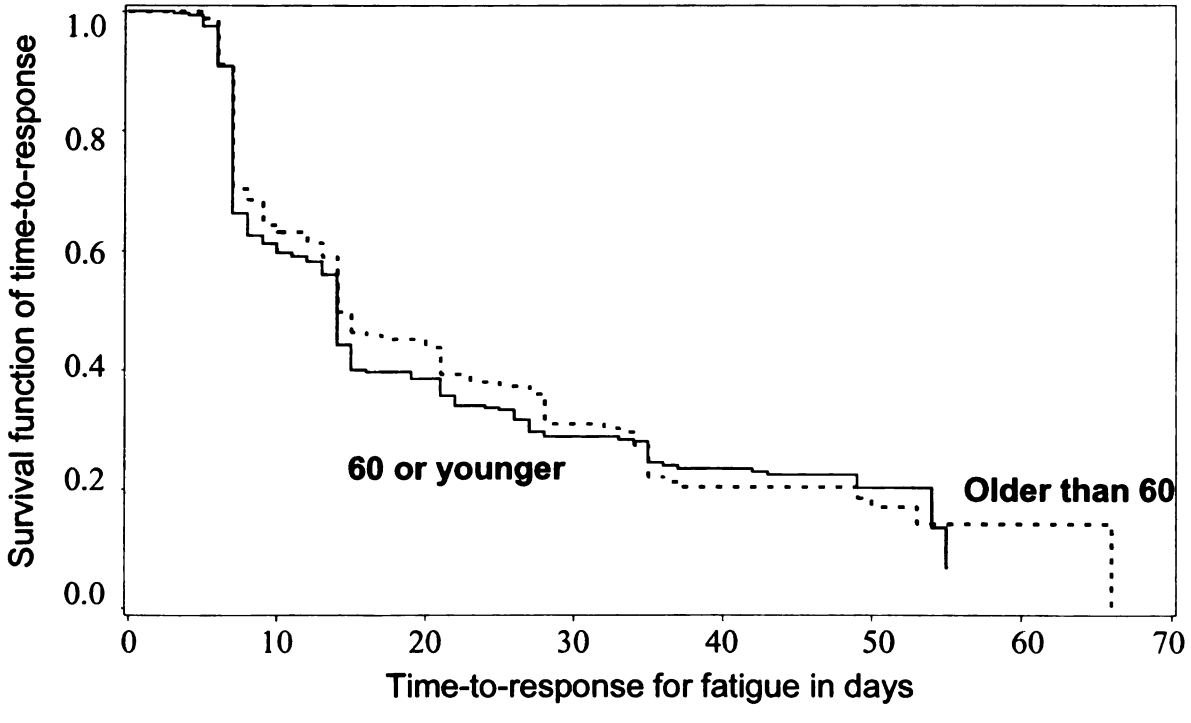


Figure 15 Estimated survival functions of time-to-response in fatigue by comorbidity

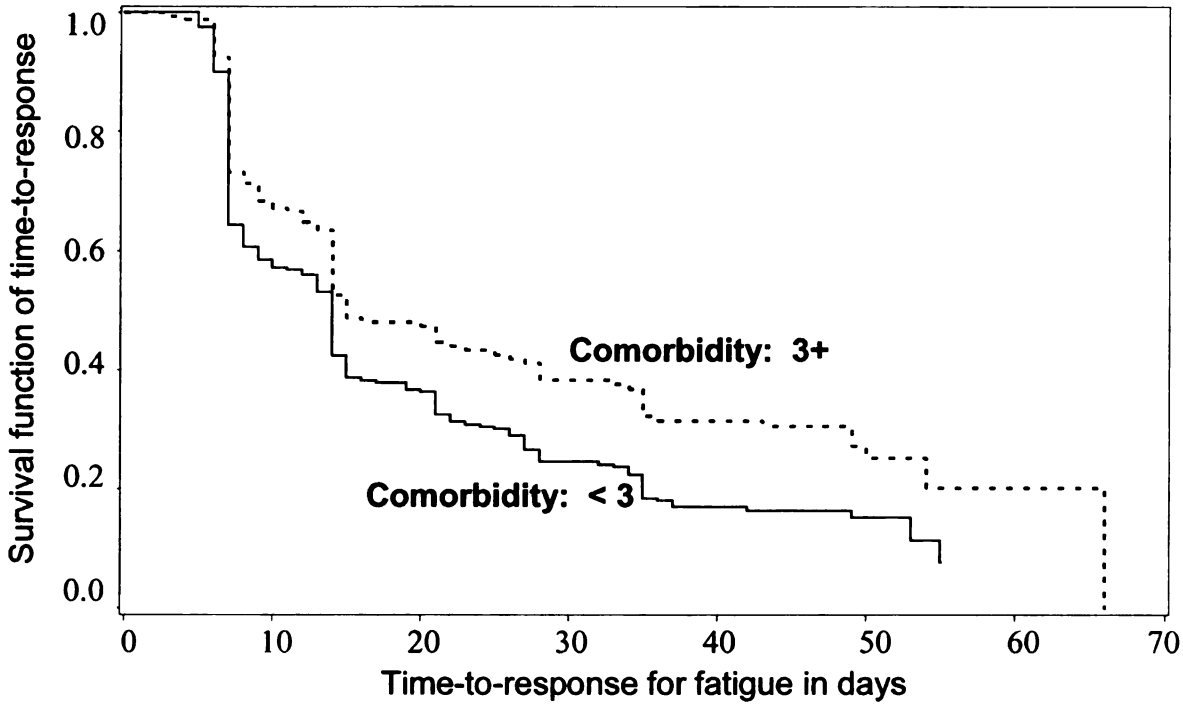


Figure 16 Estimated survival functions of time-to-response in fatigue by depression

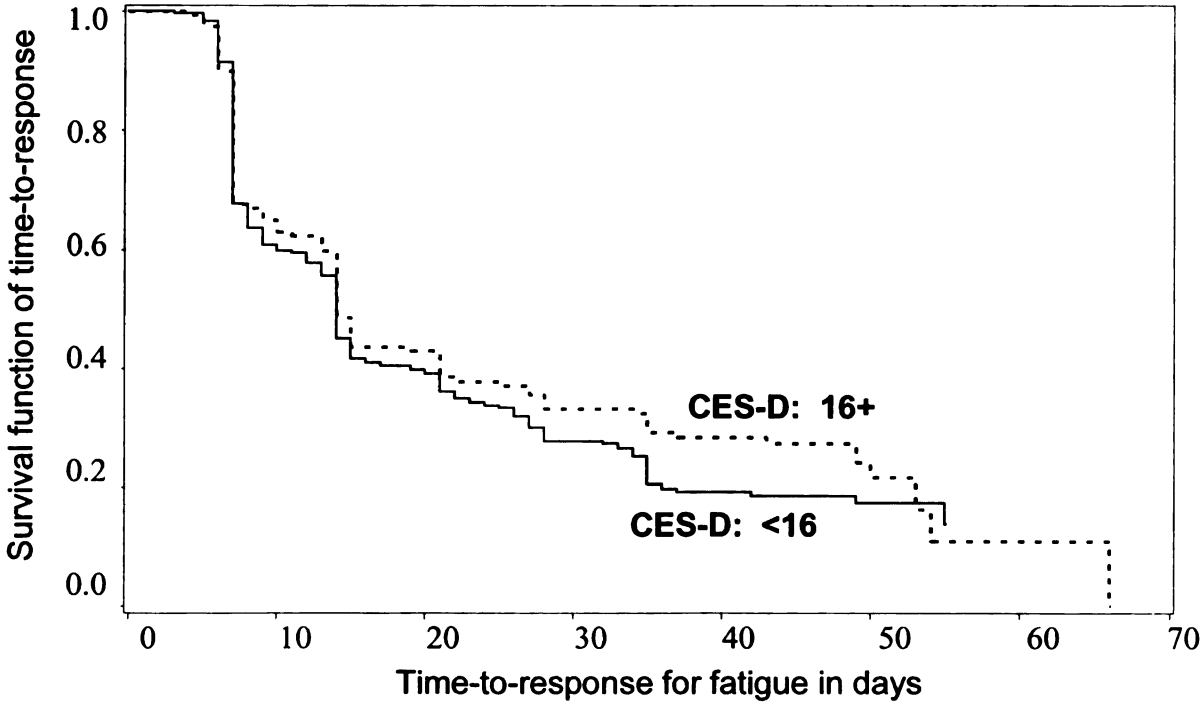


Table 6 Median time to response and unadjusted tests for time-to-response for pain and fatigue by the log-rank, Wilcoxon, and Cox proportional model

Time-to-Response in Pain		Median (days) (95% CI)	Log-rank P-value	Wilcoxon P-value	Cox PH Model P-value	Test for PH[25] P-value
Age	60 or younger	10 (9, 14)	.0190	.0293	.0293	.7476
	Older than 60	14 (14, 21)				
Comorbidity	Less than 3	9 (8, 10)	<.0001	<.0001	<.0001	.1321
	3 +	18 (14, 21)				
CES-D	Less than 16	11 (8, 14)	.0896	.1698	.1140	.9361
	16 +	14 (10, 18)				
Time-to-Response in Fatigue		Median (days) (95% CI)	Log-rank P-value	Wilcoxon P-value	Cox PH Model P-value	Test for PH[24] P-value
Age	60 or younger	14 (14, 15)	.6939	.4658	.7144	.3105
	Older than 60	14 (14, 21)				
Comorbidity	Less than 3	14 (13, 14)	.0031	.0140	.0060	.4584
	3 +	15 (14, 26)				
CES-D	Less than 16	14 (13, 15)	.3641	.6801	.3967	.4161
	16 +	14 (14, 21)				

[25] They are the results of testing whether the hazards are proportional for each covariate.

The multivariate models for time-to-response in pain and fatigue are shown in Table 7. The main effects including age, Comorbidity, and depression are tested after controlling for a priori confounders such as gender, cancer site, physical function, and trial in the final model. Only comorbidity remained as a significant covariate for both pain (HR=1.70, 95% CI=[1.18, 2.46]) and fatigue (HR=1.32, 95% CI=[1.02, 1.70]). Patients who had lower number of comorbid conditions (< 3) reported shorter time-to-response for both pain and fatigue.

To see if there exists any covariate that has a different effect between Trial A and B, I examined interaction effects between trial and each main effect in both final models. None of the interactions were significant. In addition, the final models were performed separately for Trail A and B. In the separate models, the hazard ratios for other covariates including gender, cancer site, and physical function were close to 1. The hazard ratios for comorbidity were greater than 1.3 in both models.

\

Table 7 Final models of time-to-response in pain and fatigue in the Cox proportional hazard model

		Cox Proportional Hazard Model		
Time-to-response in pain		Hazard Ratio	95% CI	p-value
Age	(60 yrs or less vs. > 60 yrs)	1.35	0.95, 1.90	.0909
Comorbidity	(<3 vs. 3+)	1.70	1.18, 2.46	.0045
CES-D	(<16 vs. ≥16)	1.18	0.83, 1.67	.3572
Gender	(Male vs. Female)	1.21	0.80, 1.83	.3650
Cancer	(Breast vs. Other)	0.94	0.63, 1.40	.9313
	(Lung vs. Other)	0.94	0.62, 1.41	
Physical function[26]		1.00	0.99, 1.01	.7651
Trial	(Trial B vs. Trial A)	1.03	0.73, 1.43	.8778
Time-to-response in fatigue		Hazard Ratio	95% CI	p-value
Age	(60 yrs or less vs. > 60 yrs)	0.92	0.73, 1.17	.4938
Comorbidity	(<3 vs. 3+)	1.32	1.02, 1.70	.0359
CES-D	(<16 vs. ≥16)	1.00	0.78, 1.27	.9831
Gender	(Male vs. Female)	0.99	0.75, 1.31	.9698
Cancer	(Breast vs. Other)	0.97	0.74, 1.28	.2738
	(Lung vs. Other)	0.78	0.57, 1.06	
Physical function [26]		1.00	0.99, 1.01	.1835
Trial	(Trial B vs. Trial A)	0.94	0.75, 1.17	.5637

5.3.3 Testing for Time-to-Response without Assuming the Proportional Hazards assumption holds

Although the proportional hazard assumption holds for three main effects of interest (age, comorbidity, and depression) in this data, this section will explore

[26] It is 1 unit change in a standardized physical function score on a 0 to 100 scale from (SF-36 subscale).

to see how the results are changed when using alternative survival analysis methods which do not require this assumption. Three recently proposed methods: including the Lin & Wang's test, a Cox model with weighted estimation (WCM), and the Rahbar's method were implemented to evaluate time-to-response. The first 2 methods are available only for univariate analysis, but WCM can be applied to multivariable analysis models. The three models, including the Lin & Wang's test, the WCM, and the Rahbar's method, are available for both proportional and non-proportional hazards. Since the proportional assumption was satisfied, the Cox proportional hazard model appears to be valid, therefore the WCM had the same result as the regular Cox proportional hazard model.

Table 8 summarizes the results of unadjusted tests for time-to-response for pain and fatigue by the Lin & Wang's test, the WCM, and the Rahbar's method. Interestingly, the Lin & Wang's test produced fairly large p-values, but failed to detect significant effects on time-to-response for pain and fatigue for most covariates except comorbidity. The WCM produces results similar to those produced by the Cox proportional hazard model in Table 6. The Rahbar's method produced different results than the other methods. This test identified age (p-value=0.0004), comorbidity (p-value<.0001), and depression (p-value =0.0001) as significant factors associated with time-to-response in pain. Rahbar's test also identified more significant covariates associated with time-to-response for fatigue including age (p-value=0.0304), comorbidity (p-value =<.0001), and depression (p-value=0.0014).

The smaller p-values generated by Rahbar's test may occur due to differences in the mean time-to-response. For example, in Figure 16 the estimated survival functions of time-to-response for fatigue by depression level are very close to each other until around 30 days. However, about 18% of patients, who had a CES-D of less than 16, are censored at 55 days, while more than 10% of patients, who had high depression, responded after 55 days. This difference results in a smaller mean time-to-response in fatigue in patients with low depression compared with those with high depression, and produces a significant p-value in the Rahbar's test.

The final multivariable WCM is not presented in this section, because the WCM is essentially identical to the Cox proportional hazard model when all covariates have proportional hazards.

Table 8 Unadjusted tests for time-to-response in pain and fatigue by the Lin & Wang's test, the WCM, and the Rahbar's method

Time-to-Response in Pain		Median (days) (95% CI)	Lin&Wang P-value	Rahbar P-value	WCM P-value
Age	60 or younger	10 (9, 14)	.6599	.0004	.0425
	Older than 60	14 (14, 21)			
Comorbidity	Less than 3	9 (8, 10)	<.0001	<.0001	<.0001
	3 +	18 (14, 21)			
CES-D	Less than 16	11 (8, 14)	.8106	.0001	.2023
	16 +	14 (10, 18)			
Time-to-Response in Fatigue		Median (days) (95% CI)	Lin&Wang P-value	Rahbar P-value	WCM P-value
Age	60 or younger	14 (14, 15)	.9343	.0304	.5211
	Older than 60	14 (14, 21)			
Comorbidity	Less than 3	14 (13, 14)	.4697	<.0001	.0206
	3 +	15 (14, 26)			
CES-D	Less than 16	14 (13, 15)	.2830	.0014	.6976
	16 +	14 (14, 21)			

5.3.4 Testing for Time-to-Response of Pain and Fatigue with Interval Censoring Type

In Section 5.3.2 and 5.3.3, all the methods account for right-censoring for time-to-response. However, since patients were monitored over two weeks at scheduled contact times, the date of contact during which patients reported response to the symptom may not correspond to the exact date of the response. Because the symptom response occurred between the prior contact and the current contact, time-to-response has an interval form. Thus, there is a question

as to how the findings from methods using right-censored data differ from those that account for interval censored time. To answer this question, the Accelerated Failure Time (AFT) model was implemented with interval time-to-response in pain and fatigue.

The results from the final multivariable models are summarized in Table 9. In contrast to the results of the Cox proportional model, age (TR=0.67, 95% CI=[0.46, 0.96]) and comorbidity (TR=0.56, 95% CI=[0.38, 0.82]) have significant effects in the final multivariable AFT model for pain. In this model, not only low comorbidity, but also younger age is associated with shorter time-to-response for pain. The AFT model of time-to-response for fatigue shows similar results to the Cox proportional model. Only comorbidity is significantly associated with time-to-response.

To test for differential effects between Trial A and B, I examined the interaction effect between trial and the three main effects on time-to-response for pain and fatigue in these AFT models. None of these interactions was significant which indicates that the identified effects did not differ between the two trials. Furthermore, when two separate models were generated for Trial A and B, the effects of the covariates are similar to the results shown in Table 9. Therefore, the difference between Trial A and B did not influence the results from the final multivariable AFT model.

Table 9 Final multivariable models of time-to-response in pain and fatigue in the AFT model with interval censoring

		Accelerated Failure Time Model		
Time-to-response in pain		Response Time Ratio	95% CI	p-value
Age	(60 yrs or less vs. > 60 yrs)	0.67	0.46, 0.96	.0289
Comorbidity	(<3 vs. 3+)	0.56	0.38, 0.82	.0031
CES-D	(<16 vs. ≥16)	0.68	0.47, 0.98	.0405
Gender	(Male vs. Female)	0.63	0.41, 0.97	.0343
Cancer	(Breast vs. Other)	0.87	0.55, 1.37	.7862
	(Lung vs. Other)	0.89	0.59, 1.34	
Physical function[27]		1.00	0.99, 1.01	.6537
Trial	(Trial B vs. Trial A)	1.05	0.74, 1.51	.7722
Time-to-response in fatigue		Response Time Ratio	95% CI	p-value
Age	(60 yrs or less vs. > 60 yrs)	1.18	0.91, 1.53	.2030
Comorbidity	(<3 vs. 3+)	0.75	0.56, 0.99	.0411
CES-D	(<16 vs. ≥16)	0.92	0.70 1.20	.5274
Gender	(Male vs. Female)	0.95	0.70, 1.20	.7597
Cancer	(Breast vs. Other)	0.90	0.67, 1.22	.4979
	(Lung vs. Other)	1.13	0.82, 1.57	
Physical function[28]		1.00	0.99, 1.00	.2303
Trial	(Trial B vs. Trial A)	1.11	0.87, 1.41	.4098

5.3.5 Testing for Time-to-Response in Pain and Fatigue while considering the correlation between the two symptoms

Effect of comorbidity was statistically significant in the three different modeling strategies in Sections 5.3.2, 5.3.3, and 5.3.4. The previous models were developed separately for pain and fatigue. Because of the correlations

[27] It is 1 unit change in a standardized physical function score on a 0 to 100 scale from (SF-36 subscale).

between pain and fatigue within an individual, the subjects included in the models for pain and fatigue are not independent. The validity of the statistical inference for the two Cox proportional hazard models for pain and fatigue could therefore be questioned because these time-to-response outcomes are correlated. Th marginal Cox model is able to account for the correlations between outcomes within an individual. The main advantage of the marginal Cox model is that it avoids the inflation of type I errors that occurs when running two separate models for pain and fatigue. Therefore, the marginal Cox model incorporating the correlation between pain and fatigue within an individual was used. The marginal Cox model can include multiple outcomes (i.e., response for pain and fatigue) for each patient and accounts for the correlation between pain and fatigue (See Chapter 3.5).

Because the marginal Cox model includes coefficients of each covariate for both pain and fatigue, the difference in the covariates between pain and fatigue can be tested. Based on the results of an equality test in the adjusted models, age (p-value=0.0848), comorbidity (p-value=0.1237), and depression (p-value=0.3541) were not significantly different and thus do not have significant different effects on the time-to-response for pain and fatigue. Therefore, using one coefficient for each covariate in the final model was appropriate strategy. The coefficients in Table 10 represent the common effect of each covariate on time-to-response on both pain and fatigue. This model also indicates that only comorbidity (HR=1.36, 95% CI=[1.11, 1.67]) is significantly associated with time-to-response.

Table 10 Final model of time-to-response in the marginal Cox model

	Hazard Ratio	95% Confidence Interval	p-value
Age (60 yrs or less vs. > 60 yrs)	0.96	0.79, 1.17	.7183
Comorbidity (<3 vs. 3+)	1.36	1.11, 1.67	.0030
CES-D (<16 vs. ≥16)	0.96	0.79, 1.17	.7115
Gender (Male vs. Female)	1.04	0.82, 1.33	.7391
Cancer (Breast vs. Other)	0.99	0.79, 1.24	.4475
(Lung vs. Other)	0.86	0.68, 1.09	
Physical function[28]	1.00	0.99, 1.01	.1440
Trial (Trial B vs. Trial A)	0.94	0.77, 1.13	.5045

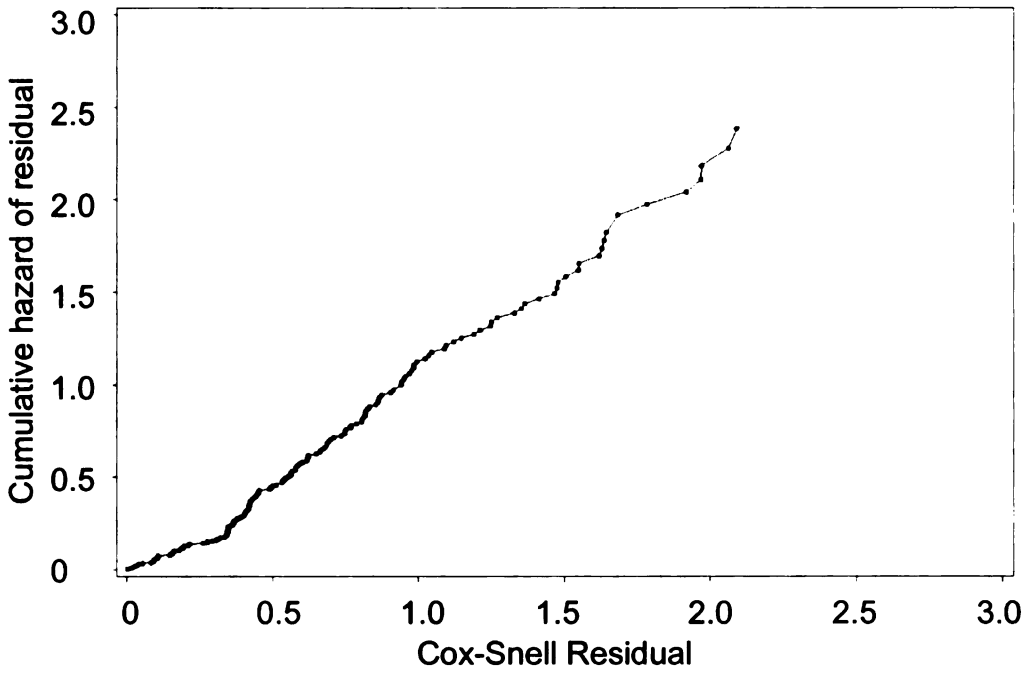
In the final mode, none of the other covariates (gender, cancer site, physical function, and trial) representing differences between Trial A and B are significant. Additionally, I also examined interactions between the indicator variable for trial (Trial A vs. B) and each covariate, and none was significant. In separate models for Trial A and B, the magnitude of the comorbidity effect (HR=1.42, 95% CI=[1.11, 1.82]) in Trial A is not seriously different from that (HR=1.34, 95% CI=[0.98, 1.83]) in Trial B.

[28] It is 1 unit change in a standardized physical function score on a 0 to 100 scale from (SF-36 subscale).

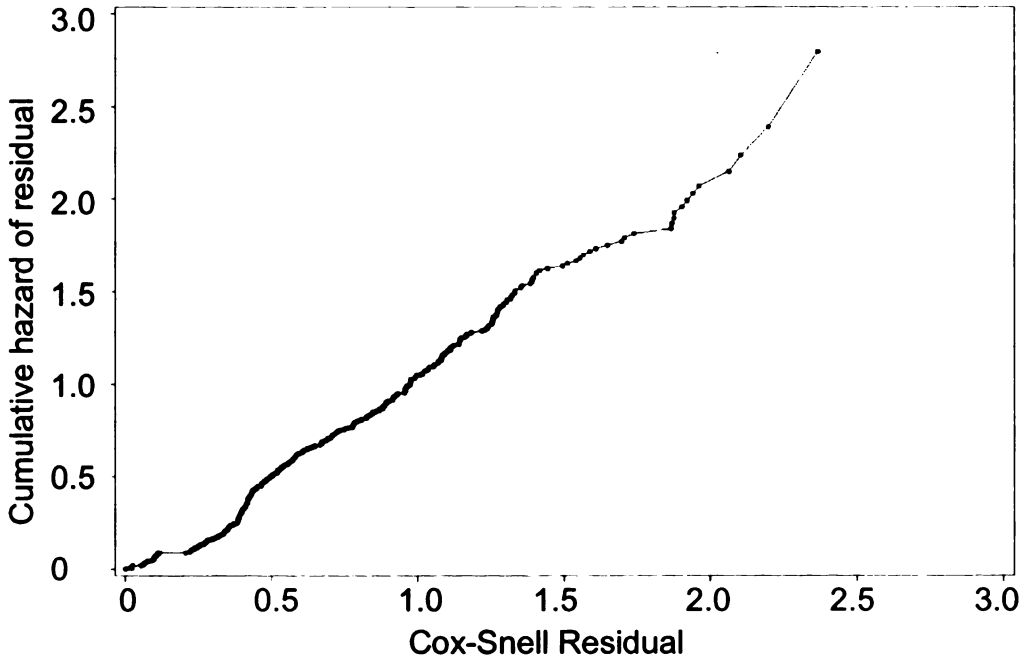
5.3.6 Diagnostic assessment of the final model

To undertake a diagnostic assessment of the final multivariable models including the Cox proportional hazard model, the AFT model, and the marginal Cox model, Cox-Snell residuals were estimated (See Chapter 4.3.7). The residual plots from the Cox proportional hazard model, the marginal Cox model, and the AFT model are presented in Figure 17 (panel a, b, and c), respectively. Because the plots are very similar for pain and fatigue for the same models, I am presenting only plots for pain. The Cox proportional model (panel a) and marginal Cox model (panel b) have a fairly straight line with a slope of one and zero intercept. These residual plots indicate that these two models fit the data fairly well. Although the plot of Cox-Snell residuals for the AFT model (panel c) is not straight as much as the Cox proportional hazard model and marginal Cox model (panel a and b), it is fairly linear line with a slope of one and zero intercept. Therefore, the results from the AFT model is also acceptable.

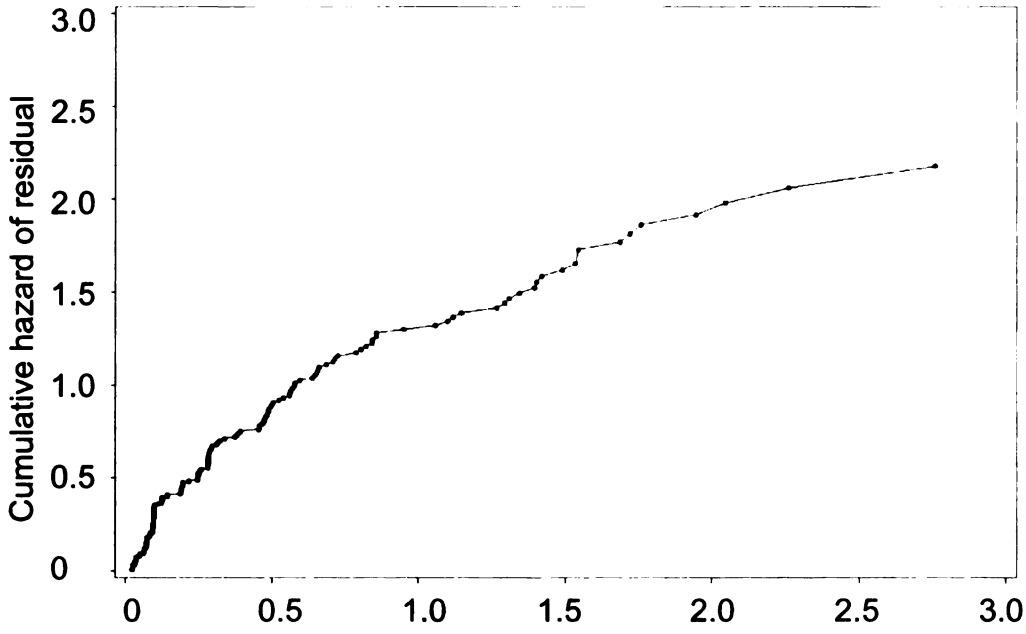
Figure 17 Cumulative hazard plot of the Cox-Snell residuals for three final models



(a) Cox proportional model



(b) Marginal Cox model



(c) AFT model with interval censoring

5.3.7 Recurrence of pain and fatigue

In this study, a symptom response is defined as the first episode of lower symptom severity level after onset (i.e. severe to moderate, severe to mild, or moderate to mild). That is, time-to-response was measured from time of onset to time of first response. As the results from several survival models show, we found that low comorbidity is significantly associated with shorter time of first response for pain and fatigue.

However, symptoms can obviously revert after an initial response (i.e. recurrence). For example, a patient who responded to pain at third contact could have severe pain again (i.e. recurrence of pain) at next contact. Thus this patient could have two episodes of response during the follow-up period. In this study, we ignored recurrent episodes thus the main effects were assessed only for the

first episode of response. The effects of multiple episodes on the results are unkown and would require a new analysis.

Table 11 is a summary of recurrence by covariates and the results of testing (Chi-square test), if recurrent cases are associated with any covariate. Thirty-four (16%) and eighty five (19%) of patients reported recurrence of pain and fatigue symptoms, resepectively, after initially responding. There is no significant difference in the recurrence of pain and fatigue between Trial A and B. Although a little more recurrences of pain occurred in younger and lower comorbidity groups, there is no significant difference in any of the covariates. Therefore, the recurrent symptoms would not seriously affect the covariate effects on time-to-response.

Table 11 Recurrence of pain and fatigue after responding to first episode of pain or fatigue

Pain	Recurrence (N=34) N (%)	No recurrence (N=178) N (%)	P-values (Chi-square Test)
Age			
< 60 yrs	26 (19.0)	111 (81.0)	0.1148
60 yrs +	8 (10.7)	67 (89.3)	
Comorbidity			
Less than 3	23 (19.3)	96 (80.7)	0.1398
3 +	11 (11.8)	82 (88.2)	
Depression			
Less than 16	16 (13.8)	100 (86.2)	0.3276
16 +	18 (18.7)	78 (81.3)	
Trial			
Trial A	21 (17.5)	99 (82.5)	0.5076
Trial B	13 (14.1)	79 (85.9)	
Fatigue	**Recurrence (N=85) N (%)**	**No recurrence (N=366) N (%)**	**P-values (Chi-square Test)**
Age			
< 60 yrs	50 (17.9)	230 (82.1)	0.4916
60 yrs +	35 (20.5)	136 (79.5)	
Comorbidity			
Less than 3	52 (18.8)	225 (81.2)	0.9593
3 +	33 (19.0)	141 (81.0)	
Depression			
Less than 16	54 (18.4)	239 (18.6)	0.7578
16 +	31 (19.6)	127 (80.4)	
Trial			
Trial A	56 (19.9)	226 (80.1)	0.4782
Trial B	29 (17.2)	140 (82.8)	

In summary, comorbidity has a strong independent effect on time-to-response in pain and fatigue among cancer patients undergoing chemotherapy. This result is consistent across different survival analysis methods. The proportional hazard models, including both the Cox proportional hazard model and the marginal Cox model, had a good model fit. The AFT model which handles interval censoring detected additional significant variable (i.e. age) on time-to-response for pain. The marginal Cox model provided an estimate of the overall effect of comorbidity (HR=1.36, 95% CI=[1.11, 1.67]) on time-to-response for both pain and fatigue when considering correlation between the two symptoms for each patient. In this data, low comorbidity was consistently associated with shorter time-to-response for both pain and fatigue with and without adjustment. Younger age was also statistically associated with shorter time-to-response for pain in both adjusted and unadjusted models. However, because younger age is associated with lower comorbid conditions, the effect of age was less significant when adjusting for comorbidity. Therefore, younger cancer patients who have a fewer comorbid conditions respond to their pain and fatigue in shorter time while they receive symptom managements.

CHAPTER 6 DISCUSSION

This research identified important factors associated with time-to-response for pain and fatigue among cancer patients undergoing chemotherapy. When assessing change in symptoms, it is important to know how to evaluate clinically meaningful changes in symptoms and what criteria to apply to assess which survival analysis methods are more appropriate in evaluating symptom response. This study demonstrates a process for evaluating clinically meaningful responses of pain and fatigue and compared alternative survival analysis models to evaluate time-to-response. In this chapter, I will discuss and summarize the following points: 1) potential biases associated with combining data sets from two trials 2) evaluation of clinically meaningful changes in the severity of each symptom, 3) use of survival analysis to asses time-to-response among symptoms and conclude with an assessment of the conditions under which each survival technique would be appropriate for use.

6.1 Considering Potential Bias from Combining Data Sets

Although pain and fatigue are relatively prevalent symptoms among cancer patients, the number of patients with symptoms reaching threshold (severity of 4 or greater) in pain and fatigue was not large in each of the trials. To gain larger sample size for increasing the statistical power, I used cases of pain and fatigue from two trials. The two trials were performed with identical timelines

and study designs. However, Trial A required the presence of a caregiver as one of entry criteria. More male patients were assigned to Trial A because mene were more likely to have a spouse who could act as a caretaker. More female patients were assigned to Trial B because they were less likely to have a spouse at home. Thus as a consequence of the different entry criteria, the higher proportion of male patients in Trial A increased the number of lung cancer patients in Trial A, and the higher proportion of women in Trial B resulted in more breast cancer patients being assigned in Trial B. The differences in cancer types between the trials resulted in more males and lower physical functioning in Trial A. These differences can lead to different survival functions for time-to-response for pain and fatigue between the two trials, which, also, may call into question the wisdom of combining data from the two trials.

Survival distributions for pain and fatigue were very similar between Trial A and B, and the coefficients for the trials (Trial A vs. B) were not significantly different in any of the survival models. This indicates that patterns of response in pain and fatigue are not different between Trial A and B. Only small changes in the hazard ratios and response time ratios were observed after controlling for site of cancer and other covariates. In addition, it was useful to compare the stratified hazard ratios by sites of cancer. The hazard ratios and their confidence intervals largely overlapped among patients with breast, lung, and other cancer sites (This is not presented in the result section.) For example, the hazard ratios for comorbidity for response in fatigue are 1.31 (95% CI=[0.87, 1.97]) for breast cancer, 1.45 (95% CI=[0.84, 2.49]) for lung cancer, and 1.20 (95% CI=[0.83,

1.74]) for other sites. Thus, despite differences in gender, site of cancer, and physical function between Trial A and B, these factors were not significantly associated with time-to-response for pain and fatigue. Further, there is no evidence that these factors modified or confounded the effects of age, comorbidity, and depression on time-to-response. Because the similar patterns of survival curves were observed between two trials in this study, it would be enough to examine the covariate effects after adjusting for the difference between trials.

6.2 Evaluating Clinically Meaningful Change on Symptom Severity

In this section, I discuss the first research question (Aim 1); can clinically meaningful changes in pain and fatigue symptoms be measured using the four dimensions of interference to define clinically meaningful cut-points that separate levels of symptom severity. To evaluate the effectiveness of symptom management it is important to use appropriate measures that are sensitive to detecting meaningful changes in symptoms. Kirshner and Guyatt (134) provided a guideline that create such measures in clinical medicine or the social science. They classified three types of measures by their purposes including discrimination, prediction, and evaluation. They define an "evaluative measure", as an instrument used to measure the individual or group change in a dimension of interest over time. Based on their guideline (134), I will review the categories of symptom severity (mild, moderate, severe) as an evaluative measure in terms

of item selection, item scaling, internal consistency, reliability, validity, and responsiveness.

First, an instrument should include items that represent clinically meaningful effects of interventions (item selection). The categories of mild, moderate, and severe are constructed using a combined interference score to separate severity measures into clinically meaningful categories. In the trials from which these data were drawn interventions were delivered to patients to reduce symptom severity. Successful reductions in symptom severity should lower interference with respect to patients' daily lives (135). That is, the changes in these interference items can represent the effect of interventions on reducing the severity of symptoms. Therefore, all items used satisfy the criterion for an evaluative measure.

Second, a scale should be sensitive to clinically important improvement or deterioration (item scaling). Usually a finer scale (0 to 10) than the threefold classification (mild, moderate, severe) is preferred as a evaluative measure (134). However, the absolute differences on a 0 to 10 scale of severity may not always represent an equal level in improvement or deterioration in interferences (80). Therefore, it is possible that there is no significant improvement in interferences while a patient lowers the severity of symptoms from 9 to 7. Although a 0 to 10 scale is more sensitive than the threefold classification, it does not have a clear threshold to identify clinically meaningful changes. According to Table 4 in Chapter 5, this threefold classification successfully discriminates the physical functioning, severities of other symptoms, as well as the summed score for

comparing the four items of the interference scale. The categories of mild, moderate, and severe may be sufficiently sensitive to clinically important changes and provide clear definitions and interpretations.

Third, Kirshner and Guyatt (134) suggest that all items in an evaluative instrument be internally consistent. The four interference items (emotion, relationship with others, daily activities, and enjoyment life) have consistently high correlations over time. The internal consistency among these items can be proved using the Cronbach's alpha[29]. Very high Cronbach's alpha (around 0.9) was observed at each of the 6 contacts. That is, it can be interpreted that about 90% of variance in the hypothetical measure (i.e., overall interference with patient's life) would be explained by these observed items. Therefore, the four interference items satisfy the criterion of internal consistency for evaluative measure.

Fourth, the replicated observations on each individual patient should remain stable over time (reliability). The evaluative measures need to remain small in magnitude on the within-patient variance (individual variance of repeated measures), while the discriminative measures focus on maximizing the between-patient variance. Figure 9 in Chapter 5 shows the distributions of differences in interference sums within patients whose severity category are very similar compared with previous contacts. These distributions which are symmetric around zero indicate small differences in interferences within a patient. That is, the expectation of differences in interference sum is zero when patients retain the

[29] Cronbach's alpha is a statistic on a 0 to 1 scale used to measure the consistency of multiple items in an instrument. When items are highly consistent, Cronbach's alpha is close to 1.

same level of severity. However, some large differences are observed in relatively small numbers of patients. These large differences occur, because interference may be associated with not only symptom severity but also with other psychological or environmental factors. In this study, these large differences are not associated with age, gender, cancer type, comorbidity, and depression at baseline. Further research needs to assess what other factors might explain these changes in interference that occur within patients over time.

Fifth, Kirshner and Guyatt (134) suggest that the measures need to be compared with some global standard methods for validity. The categories of mild, moderate, and severe are constructed by modifying the globally used instruments (Brief Pain Inventory (BPI) and Brief Fatigue Inventory (BFI)) whose validity has been proved (4, 88). This study proposed the severity of 2 and 5 as cut-points for categorizing mild, moderate, and severe in both pain and fatigue. In previous studies, Serlin et al. (80) proposed the cut-points of 5 and 7 for the categories in pain, and Mendoza et al. (4) proposed the cut-points of 4 and 7 for the categories in fatigue in cancer patients. These higher cut-points may be explained by the fact that these methods used the multivariate analysis of variance (MANOVA) to find the cut-points of severity in symptoms. This analytical technique assumes a normal distribution of interference items. In this study, however, interference sums were highly skewed to right and thus the generalized linear model with a gamma distribution was used to take into account the skewed distribution. This difference could lower the cut-points in this study below those of previous studies.

In this research the interference scale developed to establish interference based cut-points across the ten-point severity scale meet all five evaluation criteria set out by Kirshner and Guyatt (134). As a result, this scale represents a psychometrically sound indicator for separating meaningful interference based categories for defining severity, and its change (response) to interventions for managing pain and fatigue.

6.3 Applying Survival Analysis in the Assessments of Symptom

This study demonstrates application of survival analysis techniques in the assessment of pain and fatigue among cancer patients. Unlike survivorship or resolution of disease, more considerations and limitations exist, when using survival analysis models in symptom assessments. In this section, I discuss research questions (Aim 2 to 5) regarding the applications of survival analysis techniques.

Survival analysis under the proportional hazard assumption

There is the question as to which factors are predictors of time-to-response in pain and fatigue among cancer patients when using survival analysis techniques that require the assumption of proportional hazards (Aim 2). To obtain an answer for this question, we need to examine the proportional hazard assumption and to understand the distinctions among the three survival analysis methods which require the proportional hazard assumption.

In this study, the proportional hazard assumption was examined by assessing the interaction between time and each covariate in the Cox proportional hazard model. However, there are other methods for examining the proportional hazard assumption using graphical comparison and residuals rather than the interaction between time and each covariate in the Cox proportional hazard model (105, 136-139). The graphical method plots the cumulative hazard functions between groups. If the proportional hazard holds, then the log-cumulative hazard functions between groups with different levels of a covariate are parallel across log of time-to-response. Also, Grambsch and Therneau (140) proposed the use of Schoenfeld residuals for testing the proportionality. In recent research, Ng'andu (138) compared the power of detecting non-proportional hazard with these methods and found that the method using the interaction between time and covariate have equally reliable power compared to the other methods.

As long as the proportional hazard assumption holds, we need to understand the distinctions among the log-rank test, the Wilcoxon test, and the Cox proportional hazard model. In this study these three survival analysis methods had fairly consistent results that found significant effects of age and comorbidity on time-to-response for pain and comorbidity effects on the time-to-response for fatigue. In univariate analysis, the log-rank test is essentially the same as the partial likelihood function for the Cox proportional hazard model when there are no individuals having exactly the same time-to-event (105). In this study, fairly similar p-vales were obtained for the log-rank test and Cox

proportional hazard model. Ignorable differences in the p-values between the two methods may be due to the fact that several patients reported symptom responses at exactly the same time. It is known that the Wilcoxon test is more sensitive to differences in survival functions as time-to-response closes to zero (more number of individuals at risk) (105). However, the Wilcoxon test also produced similar p-values with the two methods. If investigators are more interested in the difference of time-to-response within relatively short periods then, the Wilcoxon test would be a better option. Regardless of these differences, the same results were obtained from these three methods using the univariate analysis.

In this study, the proportional hazard assumption holds for all covariates and only comorbidity has significant effect on time-to-response for both pain and fatigue in the final multivariable Cox proportional hazard model. It is possible that comorbidity is one of mediators between age and time-to-response for pain and fatigue. Based on the definition of a mediator that Baron and Kenny described (141), we can assess possibility of a mediation effect of comorbidity between age and time-to-response for pain. Baron and Kenny proposed that a mediator variable should satisfy the following conditions; 1) the predictor (i.e. age) causes the mediator variable (i.e. comorbidity), 2) the predictor is significantly associated with the outcome (i.e. time-to-response) without the mediator, 3) the mediator is significantly associated with the outcome without the predictor, 4) the predictor has smaller effect on the outcome in the model adjusting for the mediator than the model not adjusting for it. In several previous studies, it has been observed

that elderly patients are more likely to have high risk of comorbid conditions and late stage cancer (53, 58, 59). The significant effects of age and comorbidity on time-to-response for pain were observed in univariate analysis. The age effect was not significant after adjusting for comorbidity. Tein and Mackinnon (142) propose a method for estimating mediation effect with survival data. However, more research is needed to obtain a reliable confidence interval for this estimator.

Survival analysis ignoring the proportional hazard assumption

Second, the next question (Aim 3) was; whether findings based on survival analysis techniques that were appropriate for the proportional hazards assumption hold when using alternative survival techniques (the Lin & Wang's test, the Rahbar's test, and the Cox model with weighted estimation (WCM)) that do not require the proportional hazard assumption. Because the proportional hazard assumption holds for all covariates, it is not necessary to use these alternative methods in this study. The proposed methods (the Lin & Wang's test, the Rahbar's test, and the WCM) are available for both proportional and non-proportional hazards. It is worthy to discuss how differently these methods result compared to the log-rank, Wilcoxon, and Cox proportional hazard model.

Because the proportional hazard assumption held in this study, there was no difference between the Cox proportional hazard model and WCM. Because the WCM is a generalized form of the Cox proportional model to allow for the inconsistent hazard rates, the WCM produced similar results to the Cox proportional hazard model.

The Lin & Wang's test and Rahbar's method are available for both proportional and non-proportional hazards. However, there are some restrictions using these methods to assess our data. The Lin & Wang's test did not identify any significant effect of covariates (i.e. age, comorbidity, and depression) while the Rahbar's method and the WCM found significant effects of age and comorbidity. It is because large variances occur when large numbers of events (responses) occur at exactly same time. This negative aspect of Lin & Wang's test was not identified in their original study (107). The Rahbar's test had relatively very small p-values compared to other methods, because this method tests for mean survival time while other methods tests for survival functions. Test of mean survival time is relatively sensitive to small numbers of subjects with extremely long time-to-response (outliers) compared with that of survival function. The survival functions between age groups (Figure 14) or depression groups (Figure 16) are not significantly different until 55 days. Approximately 15% of the patients who were older or more depressed reported a response to their fatigue or were censored between 55 and 65 days. However, the patients who were younger or less depressed were censored at around 55 days. That is, the differences of survival functions were observed only between 55 and 65 days. The Rabhar's method is sensitive to this difference while other methods ignore that. Therefore, the Rabhar's method produced smaller p-values than others. Regardless of these differences, the significant comorbidity effect on time-to-response for pain and fatigue and the significant age effect on time-to-response for pain were observed using these alternative methods.

Survival analysis for interval censoring

The third, question was; whether the findings based on survival analysis techniques that were appropriate for right censoring hold when the Accelerated Failure Time (AFT) model is used that accounts for interval censoring. In this study, patients reported their experience with symptoms, such as severity, duration, and interference, during the past 7 days. Given that, the exact time of response cannot be observed, the time-to-response is observed at intervals. For example, if one patient reports severe pain over 3 weeks and reports mild pain at the 4^{th} week, then the time of response occurs between the 3^{rd} and 4^{th} week from the onset of the symptom. The AFT model incorporating interval censoring was used for this study and the finding was similar to the Cox proportional hazard model. One reason for the similar results between the AFT model and the Cox proportional hazard model is that the lengths of interval censoring are equal around 7 days in most cases due to prescheduled contacts. The other reason is the length of interval censoring is relatively small compared with the mean time-to-response. For example, the estimated restricted means of time-to-response are around 20 and 24 days for pain and fatigue, respectively. Therefore, difference within a length of interval censoring (one week) may not change the covariate effect compared to the Cox proportional hazard model with right-censored data. If the length of time between contacts is longer or is not scheduled on a consistent time interval, then in future research, it is worthy to compare the results from the Cox proportional hazard model and AFT model incorporating interval censoring.

Survival analysis for correlated events

Finally, the last question was; whether the findings from the separate models of pain and fatigue (the Cox proportional hazard model, the Cox model with weighted estimation, and the Accelerated Failure Time model) hold when using the Cox marginal model that accounts for the correlation between the two symptoms (pain and fatigue). Because pain and fatigue are highly correlated, the response for pain may be very close to that for fatigue within the same individual. For example, when a patient has a response to pain, he/she may, subsequently, experience a lower level of fatigue. Therefore, the time-to-response varies by individual characteristics rather than type of symptoms (pain or fatigue), and this trend resulted in fairly similar final models between pain and fatigue. The marginal Cox model determined if individual characteristics differentially affect time-to-response for pain and fatigue when considering a correlation between the two symptoms. This model consistently showed that comorbidity has a significant effect on time-to-response for both pain and fatigue with larger sample size.

In these data, I did not see considerable limitations from the non-proportional hazard or interval censoring models. Therefore, the marginal Cox model can provide the most reliable final model by incorporating the correlation between pain and fatigue. Although the limitations regarding survival analysis (Aims 2 to 5) did not seriously affect the final models in these data, this study shows what important assumptions need to be considered to use survival analysis techniques for assessing symptom response and what restrictions the used survival analysis methods have. The final models revealed a consistently

strong effect for comorbidity on time-to-response for pain and fatigue under the different conditions in terms of proportional hazard, interval censoring, and correlation between pain and fatigue. In these data shorter time-to-response for pain and fatigue among cancer patients who had fewer comorbid conditions could be due to less severe pain, fatigue, and other symptoms (21, 43) and more adherence to treatments (52, 65, 66) than those patients with more comorbid conditions. Because cancer patients who had fewer comorbid conditions may have fewer other symptoms that need to be managed, their responses to management strategies for pain and fatigue are less likely to be interfered with due to the lower overall symptom burden compared with patients who had several comorbid conditions. Patients who had fewer comorbid conditions could continue to follow strategies that the nurses or coach provided while the patients who had more comorbid conditions might have greater difficulty adhering to the larger numbers of interventions. Assuming the effectiveness of symptom management in achieving symptom responses, more comorbid conditions may interfere with adherence to symptom management thus reducing their effectiveness in producing symptom response. These differences could result in more effectively responding for pain and fatigue among patients with fewer comorbid conditions.

There is a potential that patients' age can be a confounding factor between comorbidity and time-to-response for pain and fatigue. However, the hazard ratios for comorbidity were not changed before and after adjusting for age. The hazard ratios for comorbidity without adjusting for age are 1.81 (95%

CI=[1.26, 2.59]) and 1.28 (95% CI=[2.01, 1.64]) in pain and fatigue, respectively. After adjusting for age, the hazard ratios are 1.70 (95% CI=[1.18, 2.46]) and 1.32 (95% CI=[1.02, 1.70]) which are quite similar to the hazard ratios prior to adjustment. Therefore, comorbidity appears to be a main effect and is not altered to any appreciable extent by age as a confounding effect. Symptom management for pain and fatigue in cancer patients needs to consider more intensive strategies for older patients who experience more comorbid conditions.

Appendix A The Symptom Assessment at Intervention Contact (Example for Pain)

PATIENT SYMPTOMS – PT VERSION

The next questions ask about the symptoms you have experienced in *the past 7 days* due to cancer or its treatment, and how they may have affected you.

1. During *the past 7 days*, on how many days did you experience pain? (If 0 skip to question 2)

   0 1 2 3 4 5 6 7
   Not present days

   A. On a scale of 0=not present to 10=the worst it could be, *how severe* is pain (for you)?

   0 1 2 3 4 5 6 7 8 9 10
   Not present worst possible

   B. Please rate your pain by telling me the one number that best describes your pain at its WORST in the past 7 days.

   0 1 2 3 4 5 6 7 8 9 10
   No pain as bad as you can imagine

   C. Please rate your pain by telling me the one number that best describes your pain at its LEAST in the past 7 days.

   0 1 2 3 4 5 6 7 8 9 10
   No pain as bad as you can imagine

   D. Please rate your pain by telling me the one number that best describes your pain at its AVERAGE in the past 7 days.

   0 1 2 3 4 5 6 7 8 9 10
   No pain as bad as you can imagine

   E. In the past seven days, how much reliefs have pain treatments or medications provided? Please tell me the one percentage that most represents how much reliefs you have received.

   0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
   No relief complete relief

   F. On a scale of 0=did not interfere to 10=completely interfered, overall, how much did pain interfere in your life?

   0 1 2 3 4 5 6 7 8 9 10
   Did not interfere completely interfered

G. Tell me the one number that describes how, during the past 7 days, pain has interfered with your GENERAL ACTIVITIY

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

H. Tell me the one number that describes how, during the past 7 days, pain has interfered with your MOOD

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

I. Tell me the one number that describes how, during the past 7 days, pain has interfered with your WORKING ABILITY

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

J. Tell me the one number that describes how, during the past 7 days, pain has interfered with your NORMAL WORK (includes both work outside the home and housework)

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

K. Tell me the one number that describes how, during the past 7 days, pain has interfered with your RELATIONS WITH OTHER PEOPLE

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

L. Tell me the one number that describes how, during the past 7 days, pain has interfered with your SLEEP

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

M. Tell me the one number that describes how, during the past 7 days, pain has interfered with your ENJOYMENT OF LIFE

0 1 2 3 4 5 6 7 8 9 10

Did not interfere completely interfered

Appendix B SAS macros

SAS Macro for the log-rank, Wilcoxon, and Ling & Wang's test

```
%Macro LinTest(Var=,Symptom=);

DATA Sample; Set Main(Keep= time1 &VAR CENSOR Symptom
where=(Symptom="&Symptom"));
 KEEP Z D G;
  Z=time1;
  D=Censor;
  G=&Var;

PROC RANK DATA=SAMPLE OUT=SAMPLES TIES=LOW;
VAR Z; RANKS KZ; RUN;

PROC FREQ DATA=SAMPLES NOPRINT;
 WHERE G=1; TABLE KZ / NOPERCENT OUT=T1 OUTCUM;

DATA T1;SET T1; KEEP KZ N1;N1=COUNT; RUN;
PROC FREQ DATA=SAMPLES NOPRINT;
 WHERE G=2; TABLE KZ / NOPERCENT OUT=T2 OUTCUM;
DATA T2;SET T2; KEEP KZ N2;N2=COUNT; RUN;


PROC FREQ DATA=SAMPLES NOPRINT;
 WHERE G=1 AND D=1; TABLE KZ / NOPERCENT OUT=D1 OUTCUM;
DATA D1;SET D1; KEEP KZ D1;D1=COUNT; RUN;
PROC FREQ DATA=SAMPLES NOPRINT;
 WHERE G=2 AND D=1; TABLE KZ / NOPERCENT OUT=D2 OUTCUM;
DATA D2;SET D2; KEEP KZ D2;D2=COUNT; RUN;


DATA COM;MERGE T1 T2 D1 D2; BY KZ;
 ARRAY O[4] N1 N2 D1 D2;
 DO I=1 TO 4; IF O[I]=. THEN O[I]=0;END;
 DROP I;
RUN;

PROC IML;
 USE COM;
```

```
 READ ALL VARc INTO N1;
 READ ALL VAR0 INTO N2;
 READ ALL VAR{'D1'} INTO D1;
 READ ALL VAR{'D2'} INTO D2;
 N=NROW(N1);
 M1=N1[+,];
 M2=N2[+,];
 S1=REPEAT(M1,N,1);
 S2=REPEAT(M2,N,1);
 K  =REPEAT(1,N,1);
 DO I=1 TO N-1;
    S1[I+1,1]=S1[I,1]-N1[I,1];
      S2[I+1,1]=S2[I,1]-N2[I,1];
  END;
 V=S1||D1||S2||D2;
 C={'N1','D1','N2','D2'};
 CREATE SAM FROM V [COLNAME=C];
 APPEND FROM V;
RUN;
QUIT;

DATA CAL;SET SAM;
 SumN=SUM(N1,N2);SumD=SUM(D1,D2);
  IF SumN=1 THEN DELETE;
RUN;

DATA CAL;SET CAL;
/*** Log Rank Test ***************************************/
 logE=D1-N1*SumD/SumN;
 logV=N1*N2*SumD*(SumN-SumD)/(SumN*SumN*(SumN-1));

/*** Wilcoxon Test ***************************************/
 WxE=SumN*(D1-N1*SumD/SumN);
 WxV=SumN*SumN*N1*N2*SumD*(SumN-SumD)/(SumN*SumN*(SumN-1));

/*** Lin_Test Statistic **********************************/

 E1=N1*SumD/SumN;
 V=N1*N2*SumD*(SumN-SumD)/(SumN*SumN*(SumN-1));
 E2=V+E1*E1;

IF SumD>=3 THEN C31=COMB(SumD,3);ELSE C31=0;
```

```
IF N1>=3 THEN C32=COMB(N1,3);ELSE C32=0;
IF SumN>=3 THEN C33=COMB(SumN,3);ELSE C33=0;

E3=3*E2-2*E1+GAMMA(4)*C31*C32/(C33+.000000001);

IF SumD>=4 THEN C41=COMB(SumD,4);ELSE C41=0;
IF N1>=4 THEN C42=COMB(N1,4);ELSE C42=0;
IF SumN>=4 THEN C43=COMB(SumN,4);ELSE C43=0;

E4=6*E3-11*E2+6*E1+GAMMA(5)*C41*C42/(C43+.000000001);

DELTA=(D1-E1)**2;
EDELTA=V;
VDELTA=E4-4*E3*E1+6*E2*(E1*E1)-3*(E1**4)-V*V;
RUN;

PROC MEANS DATA=CAL NOPRINT;
 VAR LogE LogV WxE WxV DELTA EDELTA VDELTA;
 OUTPUT OUT=LinTest SUM=;
RUN;
DATA LinTest ;SET LinTest;
  TestLogRank=LogE*LogE/LogV;
  TestWilcoxon=WxE*WxE/WxV;
  TestLin=(DELTA-EDELTA)/SQRT(VDELTA);
  LogRank=ROUND(1-CDF('CHISQUARE',TestLogRank,1),.0001) ;
  Wilcoxon=ROUND(1-CDF('CHISQUARE',TestWilcoxon,1) ,.0001) ;
  Lin=ROUND(2*(1-CDF('NORMAL',ABS(TestLin))),.0001) ;
RUN;


PROC PRINT DATA=LinTest;
 TITLE 'P-VALUES FOR HOMOGENIETY TEST FOR MEAN SURVIVAL TIME';
 VAR TestLogRank LogRank TestWilcoxon Wilcoxon TestLin Lin;
RUN;
%MEND;

%LinTest(Var=pSEX,Symptom=Pain);
%LinTest(Var=AGE,Symptom=Pain);
%LinTest(Var=StageCat,Symptom=Pain);
%LinTest(Var=Comorbidity,Symptom=Pain);
%LinTest(Var=CESD,Symptom=Pain);
%LinTest(Var=ChemoStart,Symptom=Pain);
```

```
%LinTest(Var=pSEX,Symptom=Fatigue);
%LinTest(Var=AGE,Symptom=Fatigue);
%LinTest(Var=StageCat,Symptom=Fatigue);
%LinTest(Var=Comorbidity,Symptom=Fatigue);
%LinTest(Var=CESD,Symptom=Fatigue);
%LinTest(Var=ChemoStart,Symptom=Fatigue);
```

SAS Macro for a nonparametric test for equality of mean survival time (Rahbar's Method)

```
%MACRO NEWTEST(G);
DATA SAM&G;SET SAMPLE;WHERE G=&G;J=1;
PROC RANK DATA=SAM&G OUT=SAM&G  DESCENDING TIES=LOW;
BY J;
VAR Z;
RANKS KZ;
RUN;
PROC SORT DATA=SAM&G ;BY J Z;RUN;
PROC LIFETEST DATA=SAM&G  NOPRINT OUTSURV=SURV&G;
 TIME Z*D(0);
 BY J;
DATA SURV&G ;SET SURV&G;
 LZ=LAG(Z);
 LSURVIV=LAG(SURVIVAL);
 BASE=Z-LZ;
 M=BASE*LSURVIV;
 IF SURVIVAL=. THEN M=.;
RUN;
PROC SORT DATA=SURV&G ;BY J Z;
DATA SURV&G ;MERGE SAM&G  SURV&G ;BY J Z;
 KZ2=LAG(KZ);
 IF  KZ2=KZ THEN DELETE;
RUN;
DATA UME&G ;SET SURV&G ; BY J;
 IF FIRST.J THEN TOTAL=0;
 TOTAL+M;

RUN;
```

```
PROC MEANS DATA=UME&G NOPRINT;
 VAR M;
 OUTPUT OUT=SUM&G  SUM=UE&G ;
 BY J;
 RUN;
DATA UME&G ;MERGE UME&G SUM&G;BY J ;
 AZ=UE&G -TOTAL;
 IF Z=0 THEN AZ=0;
   SIGMA&G =(1-_CENSOR_)*((AZ/KZ)**2);
 IF _CENSOR_=. THEN DELETE;
RUN;
PROC SORT DATA=UME&G ; BY J;
PROC MEANS DATA=UME&G  NOPRINT;
 VAR SIGMA&G;
 BY J;
 OUTPUT OUT=S&G SUM=;
RUN;
DATA UE&G ;MERGE SUM&G S&G;BY J;
 KEEP J  UE&G  SIGMA&G _FREQ_;
RUN;
PROC MEANS DATA=SAM&G NOPRINT;
 VAR KZ;
 OUTPUT OUT=NUM&G MAX=N&G;
 RUN;
DATA UE&G ;MERGE UE&G  NUM&G;
 SIGMA&G =N&G*SIGMA&G;
 STD&G =SQRT(SIGMA&G/N&G);
 G=&G;
 LABEL G='Negative';
 RUN;
PROC MEANS DATA=UE&G MEAN  MAXDEC=4;
 VAR G N&G UE&G SIGMA&G;
RUN;
%MEND;

/**** Test for Two Groups ****/

%Macro Rahbar(Var=,Symptom=);
DATA Sample;Set Main(Keep= time1 &VAR CENSOR Symptom
where=(Symptom="&Symptom"));
 KEEP Z D G;
  Z=time1;
```

```
  D=Censor;
  G=&Var;

%NEWTEST(1);
%NEWTEST(2);

DATA UES;MERGE UE1 UE2;RUN;
PROC DELETE DATA=UE1;PROC DELETE DATA=UE2;
PROC MEANS DATA=UEs N MEAN STD MAXDEC=4;
 VAR UE1 SIGMA1 UE2 SIGMA2;
RUN;

DATA ESTIMATE;SET UES;
 /***** COMBINED ESTIMATE *****/
 SUM1=SIGMA1*N1+SIGMA2*N2;
 L1=SIGMA1*N1/SUM1;
 L2=SIGMA2*N2/SUM1;

 CE1=L1*UE1+L2*UE2;
 CE2=L1*UE1+L2*UE2;

 N=N1+N2;

 OMI1=N1/N;OMI2=N2/N;

 RHO1=SIGMA1/OMI1;
 RHO2=SIGMA2/OMI2;

 COMBINE=RHO1*L1*L1+RHO2*L2*L2;
 STDCIMB=SQRT(COMBINE/N);

/*** LAMBDA                              ***/

 D1=(UE1-CE1);
 D2=(UE2-CE2);

 SUMD=SQRT(N)*D1+SQRT(N)*D2+SQRT(N);

 GAMMA1=((1-L1)**2)*RHO1+(L2**2)*RHO2;
 GAMMA2=(L1**2)*RHO1+((1-L2)**2)*RHO2;

 GAMMA12=-L1*RHO1-L2*RHO2+COMBINE;
```

```
RUN;

PROC IML;
 USE ESTIMATE;
 READ ALL VAR{'D1'} INTO D1;
 READ ALL VAR{'D2'} INTO D2;
 READ ALL VAR{'N'}  INTO N  ;
 READ ALL VAR{'GAMMA1'} INTO GAMMA1 ;
 READ ALL VAR{'GAMMA2'} INTO GAMMA2 ;
 READ ALL VAR{'GAMMA12'} INTO GAMMA12 ;
THETA=D1||D2;
C1=GAMMA1||GAMMA12;
C2=GAMMA12||GAMMA2;
V=C1`||C2`;
VO=GINV(V);
L=N*THETA*VO*THETA`;
C={'LAMBDA'};
LA=L;
CREATE LAMBDA FROM LA [COLNAME=C];
APPEND FROM LA;
RUN;
QUIT;

DATA PVALUE;SET LAMBDA;
  CHI=CINV(0.95,1);
  NEWTEST=ROUND(1-CDF('CHISQUARE',LAMBDA,1) ,.0001) ;
PROC MEANS DATA=ESTIMATE MEAN MAXDEC=3;
 VAR N1 UE1 STD1  N2 UE2 STD2  CE1 SIGMA1 SIGMA2 GAMMA1 GAMMA2
GAMMA12;
RUN;
PROC MEANS DATA=PVALUE  N MEAN MAXDEC=4;
 VAR LAMBDA NEWTEST;
 RUN;
%MEND;
%Rahbar(Var=pSEX,Symptom=Pain);
%Rahbar(Var=AGE,Symptom=Pain);
%Rahbar(Var=StageCat,Symptom=Pain);
%Rahbar(Var=Comorbidity,Symptom=Pain);
%Rahbar(Var=CESD,Symptom=Pain);
%Rahbar(Var=ChemoStart,Symptom=Pain);

%Rahbar(Var=pSEX,Symptom=Fatigue);
```

```
%Rahbar(Var=AGE,Symptom=Fatigue);
%Rahbar(Var=StageCat,Symptom=Fatigue);
%Rahbar(Var=Comorbidity,Symptom=Fatigue);
%Rahbar(Var=CESD,Symptom=Fatigue);
%Rahbar(Var=ChemoStart,Symptom=Fatigue);

/**** Tests for Three Groups ****/

%Macro Rahbar3(Var=,Symptom=);
DATA Sample;Set Main(Keep= time1 &VAR CENSOR Symptom
where=(Symptom="&Symptom"));
 KEEP Z D G;
  Z=time1;
  D=Censor;
  G=&Var;

%NEWTEST(1);
%NEWTEST(2);
%NEWTEST(3);

DATA UES;MERGE UE1 UE2 UE3;RUN;
PROC DELETE DATA=UE1;PROC DELETE DATA=UE2;PROC DELETE DATA=UE3;
PROC MEANS DATA=UES N MEAN STD MAXDEC=4;
 VAR UE1 SIGMA1 UE2 SIGMA2 UE3 SIGMA3;
RUN;

DATA ESTIMATE;SET UES;
 /***** COMBINED ESTIMATE *****/
 SUM1=SIGMA1*N1+SIGMA2*N2+SIGMA3*N3;
 L1=SIGMA1*N1/SUM1;
 L2=SIGMA2*N2/SUM1;
 L3=SIGMA3*N3/SUM1;

 CE1=L1*UE1+L2*UE2+L3*UE3;
 CE2=L1*UE1+L2*UE2+L3*UE3;
 CE3=L1*UE1+L2*UE2+L3*UE3;
 CE4=L1*UE1+L2*UE2+L3*UE3;

 N=N1+N2+N3;

 OMI1=N1/N;OMI2=N2/N;OMI3=N3/N;
```

```
 RHO1=SIGMA1/OMI1;
 RHO2=SIGMA2/OMI2;
 RHO3=SIGMA3/OMI3;

 COMBINE=RHO1*L1*L1+RHO2*L2*L2+RHO3*L3*L3;
 STDCIMB=SQRT(COMBINE/N);

  /*** LAMBDA                              ***/

 D1=(UE1-CE1);
 D2=(UE2-CE2);
 D3=(UE3-CE3);

 SUMD=SQRT(N)*D1+SQRT(N)*D2+SQRT(N)*D3;

 GAMMA1=((1-L1)**2)*RHO1+(L2**2)*RHO2+(L3**2)*RHO3;
 GAMMA2=(L1**2)*RHO1+((1-L2)**2)*RHO2+(L3**2)*RHO3;
 GAMMA3=(L1**2)*RHO1+(L2**2)*RHO2+((1-L3)**2)*RHO3;

 GAMMA12=-L1*RHO1-L2*RHO2+COMBINE;
 GAMMA13=-L1*RHO1-L3*RHO3+COMBINE;
 GAMMA23=-L2*RHO2-L3*RHO3+COMBINE;

 S1=SQRT(SIGMA1);
 S2=SQRT(SIGMA2);
 S3=SQRT(SIGMA3);
RUN;

PROC IML;
 USE ESTIMATE;
 READ ALL VAR{'D1'} INTO D1;
 READ ALL VAR{'D2'} INTO D2;
 READ ALL VAR{'D3'} INTO D3;
 READ ALL VAR{'N'}  INTO N  ;
 READ ALL VAR{'GAMMA1'} INTO GAMMA1 ;
 READ ALL VAR{'GAMMA2'} INTO GAMMA2 ;
 READ ALL VAR{'GAMMA3'} INTO GAMMA3 ;
 READ ALL VAR{'GAMMA12'} INTO GAMMA12 ;
 READ ALL VAR{'GAMMA13'} INTO GAMMA13 ;
 READ ALL VAR{'GAMMA23'} INTO GAMMA23 ;

THETA=D1||D2||D3;
```

```
C1=GAMMA1||GAMMA12||GAMMA13;
C2=GAMMA12||GAMMA2||GAMMA23;
C3=GAMMA13||GAMMA23||GAMMA3;

V=C1`||C2`||C3`;
V0=GINV(V);
L=N*THETA*V0*THETA`;
C={'LAMBDA'};
LA=L;
CREATE LAMBDA FROM LA [COLNAME=C];
APPEND FROM LA;
RUN;
QUIT;

DATA PVALUE;SET LAMBDA;
  CHI=CINV(0.95,2);
  NEWTEST=ROUND(1-CDF('CHISQUARE',LAMBDA,2) ,.0001) ;
PROC MEANS DATA=ESTIMATE MEAN MAXDEC=3;
 VAR N1 UE1 STD1  N2 UE2 STD2  N3 UE3 STD3   CE1
        SIGMA1 SIGMA2 SIGMA3
           S1 S2 S3
        GAMMA1 GAMMA2 GAMMA3
        GAMMA12 GAMMA13 GAMMA23 ;
RUN;
PROC MEANS DATA=PVALUE  N MEAN MAXDEC=4;
 VAR LAMBDA NEWTEST;
 RUN;
%mend;
%Rahbar3(Var=Cancer,Symptom=Pain);
%Rahbar3(Var=Cancer,Symptom=Fatigue);
```

Appendix C The M.D Anderson Symptom Inventory (MDASI) (143)

M.D. Anderson Symptom Inventory (MDASI) Core Items

Part I. How severe are your symptoms?

People with cancer frequently have symptoms that are caused by their disease or their treatment. We ask you to rate how severe the following symptoms have been ***in the last 24 hours***. Please fill in the circle below from 0 (symptom has not been present) to 10 (the symptom was as bad as you can imagine it could be) for each item.

1. Your pain at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

2. Your fatigue (tiredness) at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

3. Your nausea at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

4. Your disturbed sleep at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

5. Your feelings of being distressed (upset) at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

6. Your shortness of breath at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

7. Your problem with remembering things at its WORST?
   0 1 2 3 4 5 6 7 8 9 10
   Not present as bad as you can imagine

8. Your problem with lack of appetite at its WORST?

0 1 2 3 4 5 6 7 8 9 10

Not present as bad as you can imagine

9. Your feeling drowsy (sleepy) at its WORST?

0 1 2 3 4 5 6 7 8 9 10

Not present as bad as you can imagine

10. Your having a dry mouth at its WORST?

0 1 2 3 4 5 6 7 8 9 10

Not present as bad as you can imagine

11. Your feeling sad at its WORST?

0 1 2 3 4 5 6 7 8 9 10

Not present as bad as you can imagine

12. Your vomiting at its WORST?

0 1 2 3 4 5 6 7 8 9 10

Not present as bad as you can imagine

13. Your numbness of tingling at its WORST?

0 1 2 3 4 5 6 7 8 9 10

Not present as bad as you can imagine

Part II. How have your symptoms interfered with your life?

Symptoms frequently interfere with how we feel and function how much have your symptoms interfered with the following items in the last 24 hours:

14. General activity?

0 1 2 3 4 5 6 7 8 9 10

Do Not Interfere Interfered Completely

15. Mood?

0 1 2 3 4 5 6 7 8 9 10

Do Not Interfere Interfered Completely

16. Work (including work around the house)?

0 1 2 3 4 5 6 7 8 9 10

Do Not Interfere Interfered Completely

17. Relations with other people?

0	1	2	3	4	5	6	7	8	9	10
Do Not Interfere										Interfered Completely

18. Walking?

0	1	2	3	4	5	6	7	8	9	10
Do Not Interfere										Interfered Completely

19. Enjoyment of life?

0	1	2	3	4	5	6	7	8	9	10
Do Not Interfere										Interfered Completely

Appendix D The Center for Epidemiologic Studies – Depression (CES-D) scale (125)

Depression (CESD) Questions Your Feelings

These questions are about how you feel and how things have been with you within the past month. Please note the answer that comes closest to the way you have been feeling during the past month. The answer choices are "almost all of the time," "most of the time," "some of the time," "rarely or none of the time."

During the past month, how much of the time:

0) rarely/none of the time 1) some of the time 2) most of the time 3) all of the time

1. Were you bothered by things that usually don't bother you?
2. Have you not felt like eating; had a poor appetite?
3. Have you felt that you could not shake off the blues, even with the help of family or friends?
4. Have you felt that you were just as good as other people?
5. Have you had trouble keeping your mind on what you were doing?
6. Have you felt depressed?
7. Have you felt that everything you did was an effort?
8. Have you felt hopeful about the future?
9. Have you thought your life has been a future?
10. Have you felt fearful?
11. Have your sleep been restless?
12. Were you happy?
13. Have you talked less than usual?
14. Have you felt lonely?
15. Were people unfriendly?
16. Have you enjoyed life?
17. Have you had crying sells?
18. Have you felt sad?
19. Have you felt that people dislike you?
20. Could you not get going?

Appendix E The Short Form – 36 (SF-36) Physical Functioning Sub Scale

The following questions are about the activities you might do during a typical day.

1. Does your health limit your ability to do activities? If so, how much?
   a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports?
      Currently, does your health limit you in ----------?
      ------- Yes, limited a lot.
      ------- Yes, limited a little.
      ------- No, not limited at all.

   b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf?
      Currently, does your health limit you in ----------?
      ------- Yes, limited a lot.
      ------- Yes, limited a little.
      ------- No, not limited at all.

   c. Lifting or carrying groceries?
      Currently, does your health limit you in ----------?
      ------- Yes, limited a lot.
      ------- Yes, limited a little.
      ------- No, not limited at all.

   d. Climbing several flights of stairs?
      Currently, does your health limit you in ----------?
      ------- Yes, limited a lot.
      ------- Yes, limited a little.
      ------- No, not limited at all.

   e. Climbing one flight of stairs?
      Currently, does your health limit you in ----------?
      ------- Yes, limited a lot.
      ------- Yes, limited a little.
      ------- No, not limited at all.

f. Bending, kneeling, or stooping?
Currently, does your health limit you in ----------?
------- Yes, limited a lot.
------- Yes, limited a little.
------- No, not limited at all.

g. Walking more than a mile?
Currently, does your health limit you in ----------?
------- Yes, limited a lot.
------- Yes, limited a little.
------- No, not limited at all.

h. Walking several blocks?
Currently, does your health limit you in ----------?
------- Yes, limited a lot.
------- Yes, limited a little.
------- No, not limited at all.

i. Walking one block?
Currently, does your health limit you in ----------?
------- Yes, limited a lot.
------- Yes, limited a little.
------- No, not limited at all.

j. Bathing or dressing yourself?
Currently, does your health limit you in ----------?
------- Yes, limited a lot.
------- Yes, limited a little.
------- No, not limited at all.

Appendix F The Assessment for Comorbid Conditions (127)

HEALTH CONDITIONS

1. Has a doctor ever told you that you have high blood pressure of hypertension?
   a. Yes b. No

2. Do you have diabetes?
   a. Yes b. No

3. Has a doctor ever told you that you have cancer or a malignant tumor, other than the cancer for which you currently are being treated?
   a. Yes b. No

4. Not including asthma, has a doctor ever told you that have chronic lung disease such as chronic bronchitis or emphysema?
   a. Yes b. No

5. Have a doctor ever told you that you had a heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems?
   a. Yes b. No

6. Have you recently had any angina or chest pains due to your heart?
   a. Yes b. No

7. Has a doctor ever told you that you had a stroke?
   a. Yes b. No

8. Have you ever seen a doctor for emotional, nervous, or psychiatric problems?
   a. Yes b. No

9. During the last 12 months, have you seen a doctor specifically for arthritis or rheumatism?
   a. Yes b. No

10. Have you ever fractured your hip?
    a. Yes b. No

11. Have you ever had surgical replacement of a joint?
    a. Yes b. No

12. During the last 12 months, have you lost any amount of urine beyond your control?
    a. Yes b. No

13. Have you ever had cataract surgery?
    a. Yes b. No

14. Do you ever wear a hearing aid?
    a. Yes b. No

15. Do you have any other major health problems, which you haven't told me about?
    a. Yes b. No

    Specify ----------

BIBLOGRAPHY

1. Miaskowski C, Dodd M, Lee K. Symptom Clusters: The New Frontier in Symptom Management Research. JNCIMONO. 2004;2004(32):17-21.

2. Miaskowski C. Symptom clusters: establishing the link between clinical practice and symptom management research. Supportive Care in Cancer. 2006;14(8):792-4.

3. Chang VT, Thaler HT, Polyak TA, Kornblith AB, Lepore JM, Portenoy RK. Quality of life and survival. Cancer. 1998;83(1):173-9.

4. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. The rapid assessment of fatigue severity in cancer patients. Cancer. 1999;85(5):1186-96.

5. Curt GA, Breitbart W, Cella D, Groopman JE, Horning SJ, Itri LM, et al. Impact of Cancer-Related Fatigue on the Lives of Patients: New Findings from the Fatigue Coalition. Fatigue and Cancer. 2001.

6. Cleeland CS. Cancer-related symptoms. Seminars in Radiation Oncology. 2000;10(3):175-90.

7. Cleeland CS. Cancer-related fatigue: new directions for research. Introduction. Cancer. 2001;92(6 Suppl):1657-61.

8. Borden EC, Parkinson D. A perspective on the clinical effectiveness and tolerance of interferon-alpha. Semin Oncol. 1998;25(1 Suppl 1):3-8.

9. Adult Cancer Pain. National Comprehensive Cancer Network; 2008 [updated 2008; cited 2008 July 10]; Available from.

10. Weiss SC, Emanuel LL, Fairclough DL, Emanuel EJ. Understanding the experience of pain in terminally ill patients. The Lancet. 2001;357(9265):1311-5.

11. van den Beuken-van Everdingen MHJ, de Rijke JM, Kessels AG, Schouten HC, van Kleef M, Patijn J. Prevalence of pain in patients with cancer: a systematic review of the past 40 years. Annals of Oncology. 2007;18(9):1437.

12. Vainio A, Auvinen A. Prevalence of symptoms among patients with advanced cancer: An international collaborative study. Journal of Pain and Symptom Management. 1996;12(1):3-10.

13. Bonica JJ. Treatment of cancer pain: current status and future needs. Advances in pain research and therapy. 1985;9:589-616.

14. Merskey H. Classification of chronic pain: descriptions of chronic pain syndromes and definitions of pain terms. Pain Supplement(Amsterdam). 1986.

15. Spiegel D, Sands S, Koopman C. Pain and depression in patients with cancer. Cancer. 1994;74(9):2570-8.

16. Woodforde JM, Fielding JR. Pain and cancer. J Psychosom Res. 1970;14(4):365-70.

17. Ahles TA, Blanchard EB, Ruckdeschel JC. The multidimensional nature of cancer-related pain. Pain. 1983;17(3):277-88.

18. Massie MJ, Holland JC. The cancer patient with pain: psychiatric complications and their management. Med Clin North Am. 1987;71(2):243-58.

19. Hendler N. Depression caused by chronic pain. Journal of Clinical Psychiatry. 1984;45(3):30-6.

20. Ryan JL, Carroll JK, Ryan EP, Mustian KM, Fiscella K, Morrow GR. Mechanisms of Cancer-Related Fatigue. The Oncologist. 2007;12(Supplement 1):22.

21. Mock V, Atkinson A, Barsevick A, Cella D, Cimprich B, Cleeland C. NCCN Practice Guidelines for Cancer-Related Fatigue. Oncology (Williston Park). 2000;14(11A):151-61.

22. Cancer-Related Fatigue. National Comprehensive Cancer Network; 2008 [updated 2008 20/29/07; cited]; Available from: http://www.nccn.org/professionals/physician_gls/PDF/fatigue.pdf.

23. Mustian KM, Morrow GR, Carroll JK. Integrative Nonpharmacologic BehavioralInterventions fortheManagementofCancer-RelatedFatigue. The Oncologist. 2007;12:52-67.

24. Greene D, Nail LM, Fieler VK, Dudgeon D, Jones LS. A comparison of patient-reported side effects among three chemotherapy regimens for breast cancer. Cancer Pract. 1994;2(1):57-62.

25. Irvine D, Sc MH, Vincent L, Graydon JE, Bubela N, Thompson L. The prevalence and correlates of fatigue in patients receiving treatment with chemotherapy and radiotherapy; A comparison with the fatigue experienced by healthy individuals. Cancer Nursing. 1994;17(5):367.

26. Richardson A, Ream E. The experience of fatigue and other symptoms in patients receiving chemotherapy. European Journal of Cancer Care. 1996;5:24-30.

27. Richardson A, Ream E, Wilson-Barnett J. Fatigue in patients receiving chemotherapy: Patterns of change. Cancer Nursing. 1998;21(1):17.

28. Morrow GR, Shelke AR, Roscoe JA, Hickok JT, Mustian K. Management of Cancer-Related Fatigue. Cancer Investigation. 2005;23(3):229-39.

29. Hwang SS, Chang VT, Rue M, Kasimis B. Multidimensional independent predictors of cancer-related fatigue. Journal of Pain and Symptom Management. 2003;26(1):604-14.

30. Vogelzang NJ, Breitbart W, Cella D, Curt GA, Groopman JE, Horning SJ, et al. Patient, caregiver, and oncologist perceptions of cancer-related fatigue: results of a tripart assessment survey. The Fatigue Coalition. Semin Hematol. 1997;34(3 Suppl 2):4-12.

31. Glaspy J, Bukowski R, Steinberg D, Taylor C, Tchekmedyian S, Vadhan-Raj S. Impact of therapy with epoetin alfa on clinical outcomes in patients with nonmyeloid malignancies during cancer chemotherapy in community oncology practice. Procrit Study Group. Journal of Clinical Oncology. 1997;15(3):1218.

32. Reich SG. The tired patient: psychological versus organic causes. Hospital Medicine. 1986;22(7):142-54.

33. Miaskowski C, Cooper BA, Paul SM, Dodd M, Lee K, Aouizerat BE, et al. Subgroups of Patients With Cancer With Different Symptom Experiences and Quality-of-Life Outcomes: A Cluster Analysis. Oncology Nursing Forum. 2006;33(5):79-89.

34. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom distress in cancer patients. Cancer. 2000;89(7):1634-46.

35. Grond S, Zech D, Diefenbach C, Bischoff A. Prevalence and pattern of symptoms in patients with cancer pain: a prospective evaluation of 1635 cancer patients referred to a pain clinic. J Pain Symptom Manage. 1994;9(6):372-82.

36. Donnelly S, Walsh D, Rybicki L. The symptoms of advanced cancer: identification of clinical and research priorities by assessment of prevalence and severity. J Palliat Care. 1995;11(1):27-32.

37. Portenoy RK, Thaler HT, Kornblith AB, McCarthy Lepore J, Friedlander-Klar H, Coyle N, et al. Symptom prevalence, characteristics and distress in a cancer population. Quality of Life Research. 1994;3(3):183-9.

38. Cleeland CS. Symptom Burden: Multiple Symptoms and Their Impact as Patient-Reported Outcomes. JNCI Monographs. 2007;2007(37):16.

39. Dodd MJ, Miaskowski C, Paul SM. Symptom clusters and their effect on the functional status of patients with cancer. Oncol Nurs Forum. 2003;2001:465-70.

40. Given B, Given C, Azzouz F, Stommel M. Physical functioning of elderly cancer patients prior to diagnosis and following initial treatment. Nurs Res. 2001;50(4):222-32.

41. Bower JE, Ganz PA, Desmond KA, Rowland JH, Meyerowitz BE, Belin TR. Fatigue in Breast Cancer Survivors: Occurrence, Correlates, and Impact on Quality of Life. Journal of Clinical Oncology. 2000;18(4):743.

42. Kaasa S, Loge JH, Knobel H, Jordhoy MS, Brenne E. Fatigue. Measures and relation to pain. Acta Anaesthesiol Scand. 1999;43(9):939-47.

43. Given CW, Given B, Azzouz F, Kozachik S, Stommel M. Predictors of Pain and Fatigue in the Year Following Diagnosis Among Elderly Cancer Patients. Journal of Pain and Symptom Management. 2001;21(6):456-66.

44. Cleeland CS, Gonin R, Hatfield AK, Edmonson JH, Blum RH, Stewart JA, et al. Pain and Its Treatment in Outpatients with Metastatic Cancer. New England Journal of Medicine. 1994;330(9):592.

45. Gureje O, Von Korff M, Simon GE, Gater R. Persistent Pain and Well-being A World Health Organization Study in Primary Care. Am Med Assoc; 1998. p. 147-51.

46. Thase ME. Correlates and consequences of chronic insomnia. General Hospital Psychiatry. 2005;27(2):100-12.

47. Koller M, Heitmann K, Kussmann J, Lorenz W. Symptom reporting in cancer patients II. Cancer. 1999;86(8):1609-20.

48. Gift AG, Stommel M, Jablonski A, Given W. A cluster of symptoms over time in patients with lung cancer. Nurs Res. 2003;52(6):393-400.

49. Cooley ME. Patterns of symptom distress in adults receiving treatment for lung cancer. J Palliat Care. 2002;18(3):150?59.

50. Miaskowski C. Gender Differences in Pain, Fatigue, and Depression in Patients With Cancer. JNCIMONO. 2004;2004(32):139-43.

51. Turk DC, Okifuji A. Does sex make a difference in the prescription of treatments and the adaptation to chronic pain by cancer and non-cancer patients? Pain. 1999;82(2):139-48.

52. Newschaffer CJ, Penberthy L, Desch CE, Retchin SM, Whittemore M. The effect of age and comorbidity in the treatment of elderly women with nonmetastatic breast cancer. Archives of Internal Medicine. 1996;156(1):85-90.

53. Yancik R, Wesley MN, Ries LAG, Havlik RJ, Edwards BK, Yates JW. Effect of Age and Comorbidity in Postmenopausal Breast Cancer Patients Aged 55 Years and Older. Am Med Assoc; 2001. p. 885-92.

54. Pater JL, Zee B, Palmer M, Johnston D, Osoba D. Fatigue in patients with cancer: results with National Cancer Institute of Canada Clinical Trials Group studies employing the EORTC QLQ-C30. Supportive Care in Cancer. 1997;5(5):410-3.

55. Akechi T, Kugaya A, Okamura H, Yamawaki S, Uchitomi Y. Fatigue and Its Associated Factors in Ambulatory Cancer Patients A Preliminary Study. Journal of Pain and Symptom Management. 1999;17(1):42-8.

56. Redeker NS, Lev EL, Ruggiero J. Insomnia, fatigue, anxiety, depression, and quality of life of cancer patients undergoing chemotherapy. Sch Inq Nurs Pract. 2000;14(4):275-90.

57. Jeon S. Factors associated with time to resolution of pain and fatigue among cancer patients undergoing chemotherapy in a cognitive behavioral intervention group. East Lansing: Michigan State University; 2006.

58. Guralnik JM. Assessing the impact of comorbidity in the older population. Annals of Epidemiology. 1996;6(5):376-80.

59. Seeman TE, Guralnik JM, Kaplan GA, Knudsen L, Cohen R. The Health Consequences of Multiple Morbidity in the Elderly: The Alameda County Study. Journal of Aging and Health. 1989;1(1):50.

60. Ward SE, Goldberg N, Miller-McCauley V, Mueller C, Nolan A, Pawlik-Plank D, et al. Patient-related barriers to management of cancer pain. Pain. 1993;52(3):319-24.

61. Matsushita T, Matsushima E, Maruyama M. Anxiety and depression of patients with digestive cancer. Psychiatry and Clinical Neurosciences. 2005;59(5):576-83.

62. Daut RL, Cleeland CS. The prevalence and severity of pain in cancer. Cancer. 1982;50(9):1913-8.

63. Twycross R, Harcourt J, Bergl S. A survey of pain in patients with advanced cancer. Journal of Pain and Symptom Management. 1996;12(5):273-82.

64. Aaron LA, Herrell R, Ashton S, Belcourt M, Schmaling K, Goldberg J, et al. Comorbid Clinical Conditions in Chronic Fatigue. J Gen Intern Med. 2001;16(1):24-31.

65. Satariano WA. Comorbidity and functional status in older women with breast cancer: implications for screening, treatment, and prognosis.

66. Greenfield S, Blanco DM, Elashoff RM, Ganz PA. Patterns of care related to age of breast cancer patients. JAMA. 1987;257(20):2766-70.

67. DiMatteo MR, Lepper HS, Croghan TW. Depression Is a Risk Factor for Noncompliance With Medical Treatment Meta-analysis of the Effects of Anxiety and Depression on Patient Adherence. Am Med Assoc; 2000. p. 2101-7.

68. Ciaramella A, Poli P. Assessment of depression among cancer patients: the role of pain, cancer type and treatment. Psycho-Oncology. 2001;10(2):156-65.

69. Rao A, Cohen HJ. Symptom Management in the Elderly Cancer Patient: Fatigue, Pain, and Depression. JNCIMONO. 2004;2004(32):150-7.

70. Barsevick AM, Dudley WN, Beck SL. Cancer-related fatigue, depressive symptoms and functional status: A mediation model. Nursing Research. 2006;55(5):366-72.

71. Visser MRM, Smets EMA. Fatigue, depression and quality of life in cancer patients: how are they related? Supportive Care in Cancer. 1998;6(2):101-8.

72. Hopwood P, Stephens RJ. Depression in Patients With Lung Cancer: Prevalence and Risk Factors Derived From Quality-of-Life Data. Journal of Clinical Oncology. 2000;18(4):893.

73. Newell S, Sanson-Fisher RW, Girgis A, Ackland S. The physical and psychosocial experiences of patients attending an outpatient medical oncology department: a cross-sectional study. European Journal of Cancer Care. 1999;8(2):73-82.

74. Jacobsen PB. Assessment of Fatigue in Cancer Patients. JNCIMONO. 2004;2004(32):93-7.

75. Jacox A, Carr DB, Payne R. Management of Cancer Pain. Clinical Practice Guideline No. 9. Rockville, MD: US Department of Health and Human Service March. 1994:94-052.

76. Cleeland CS. Measurement of pain by subjective report. Advances in pain research and therapy. 1989;12:391?03.

77. Farrar JT, Young Jr JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149-58.

78. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain. 2000;88(3):287-94.

79. Cepeda MS AJ, Polo R, Alcala R, Carr DB. What decline in pain intensity is meaningful to patients with acute pain? Pain. 2003;105(1-2):151-7.

80. Serlin RC, Mendoza TR, Nakamura Y, Edwards KR, Cleeland CS. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. Pain. 1995;61(2):277-84.

81. Jensen MP. The validity and reliability of pain measures in adults with cancer. Journal of Pain. 2003;4(1):2-21.

82. Huskisson EC, Scott J. How double blind is double blind? And does it matter? Br J Clin Pharmacol. 1976;3(2):331-2.

83. Lasagna L. The clinical measurement of pain. Ann NY Acad Sci. 1960;86:28-37.

84. Cleeland CS. The impact of pain on the patient with cancer. Cancer. 1984;54(11 Suppl):2635-41.

85. Reyes-Gibby CC, Mendoza TR, Wang S, Anderson KO, Cleeland CS. Pain and Fatigue in Community-Dwelling Adults. Pain Medicine. 2003;4(3):231-7.

86. Jacox A, Carr DB, Payne R, Berde CB, Brietbart W, Cain JM, et al. Management of Cancer Pain. Clinical Practice Guideline No. 9. AHCPR Publication No. 94-0592. Rockville, MD. Agency for Health Care Policy and Research, US Department of Health and Human Services. Public Health Service. 1994;257.

87. Cleeland CS, Reyes-Gibby CC. When is it justified to treat symptoms? Measuring symptom burden. Oncology (Williston Park). 2002;16(9 Suppl 10):64-70.

88. Cleeland C. Assessment of pain in cancer: measurement issues. In: Foley K, Bonica, JJ., Ventafridda V., , editor. Advances in pain research and therapy. New York: Raven Press; 1990. p. 47-55.

89. Daut RL, Cleeland CS, Flanery RC. Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases. Pain. 2002;1983:197-210.

90. Wells N, Murphy B, Wujcik D, Johnson R, editors. Pain-Related Distress and Interference With Daily Life of Ambulatory Patients With Cancer With Pain; 2003. Onc Nurs Society.

91. Miaskowski C, Dodd M, West C, Paul SM, Schumacher K, Tripathy D, et al. The use of a responder analysis to identify differences in patient outcomes following a self-care intervention to improve cancer pain management. Pain. 2007;129(1-2):55-63.

92. Paul SM, Zelman DC, Smith M, Miaskowski C. Categorizing the severity of cancer pain: further exploration of the establishment of cutpoints. Pain. 2005;113(1-2):37-44.

93. Jensen MP, Smith DG, Ehde DM, Robinsin LR. Pain site and the effects of amputation pain: further clarification of the meaning of mild, moderate, and severe pain. Pain. 2001;91(3):317-22.

94. Hwang SS, Chang VT, Cogswell J, Kasimis BS. Clinical relevance of fatigue levels in cancer patients at a Veterans Administration Medical Center. Cancer. 2002;94(9):2481-9.

95. Dionne RA, Bartoshuk L, Mogil J, Witter J. Individual responder analyses for pain: does one pain scale fit all? Trends in Pharmacological Sciences. 2005;26(3):125-30.

96. Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MAG, Fayers PM. The clinical significance of adaptation to changing health: A meta-analysis of response shift. Quality of Life Research. 2006;15(9):1533-50.

97. Last JM. A Dictionary of Epidemiology. Nueva York. Oxford University Press; 2001.

98. Irwin JO. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. J Hygiene. 1949;47:188-9.

99. Karrison TG. Use of Irwin's restricted mean as an index for comparing survival in different treatment groups - interpretation and power considerations. Controlled Clinical Trials. 1997;18(2):151-67.

100. Andersen PK, Hansen MG, Klein JP. Regression Analysis of Restricted Mean Survival Time Based on Pseudo-Observations. Lifetime Data Analysis. 2004;10(4):335-50.

101. Zucker DM. Restricted Mean Life with Covariates: Modification and Extension of a Useful Survival Analysis Method. Journal of the American Statistical Association. 1998;93(442).

102. Mantel N, Haenszel W. Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. Journal of the National Cancer Institute. 1959;22:719-48.

103. Cox DR. Regression models and life tables (with discussion). JR Stat Soc B. 1972;34(2):187-220.

104. Breslow NE. Covariance analysis of censored survival data. Biometrics. 1974;30(1):89-99.

105. Collett D. Modelling Survival Data in Medical Research. CRC Press; 2003.

106. Schemper M. Cox analysis of survival data with non-proportional hazard functions. The Statistician. 1992;41:455-65.

107. Lin X, Wang H. A New Testing Approach for Comparing the Overall Homogeneity of Survival Curves. Biometrical Journal. 2004;46(5):489-96.

108. Moreau T, Le Minor M, Myquel P, Lellouch J. Estimation of the hazards ratio in two grouped samples. Biometrics. 1985;41(1):245-52.

109. Esteve J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: elements for further discussion. Stat Med. 1990;9(5):529-38.

110. Gehan EA. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples*. Biometrika. 1965;52(1-2):203-23.

111. Prentice RL. Linear rank tests with right censored data. Biometrika. 1978;65(1):167-79.

112. Heinze G. WCM: A SAS macro for weighted parameter estimation for Cox model. [cited 2008 June 1st]; Available from: http://www.meduniwien.ac.at/msi/biometrie/programme/frameset_programme_en.htm.

113. Rahbar M, Jeon S, Gardiner J. A Nonparametric Test for Homogeneity of Survival Means. Presented at the Joint Statistical Meeting in Salt Lake City. 2007.

114. Rahbar MH, Gardiner JC. Nonparametric estimation of regression parameters from censored data with a discrete covariate. Statistics and Probability Letters. 1995;24(1):13-20.

115. Rahbar MH, Gardiner JC. Nonparametric Modeling of the Mean Survival Time in a Multi-factor Design Based on Randomly Right-Censored Data. Biometrical Journal. 2004;46(5):497-502.

116. Rahbar MH, Talebi H, Jeon S. Nonparametric assessment of interactions in a multi-factor design based on randomly right-censored data. Submitted for Statistics and Probability Letters. 2008.

117. Odell PM, Anderson KM, D'Agostino RB. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. Biometrics. 1992;48(3):951-9.

118. Dorey FJ, Little RJ, Schenker N. Multiple imputation for threshold-crossing data with interval censoring. Stat Med. 1993;12(17):1589-603.

119. Fastclus S. SAS/STAT 9.1 User's Guide. SAS Institute Inc. 2004.

120. Hern MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. Pharmacoepidemiology and Drug Safety. 2005;14:477-91.

121. Wei LJ, Lin DY, Weissfeld L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. Journal of the American Statistical Association. 1989;84(408).

122. Diggle P. Analysis of Longitudinal Data. Oxford University Press; 2002.

123. Given B, Given CW, Sikorskii A, Jeon S, McCorkle R, Champion V, et al. Establishing Mild, Moderate, and Severe Scores for Cancer-Related Symptoms: How Consistent and Clinically Meaningful Are Interference-Based Severity Cut-Points? Journal of Pain and Symptom Management. 2008;35(2):126-35.

124. Sikorskii A, Given CW, Given B, Jeon S, Decker V, Decker D, et al. Symptom Management for Cancer Patients: A Trial Comparing Two Multimodal Interventions. Journal of Pain and Symptom Management. 2007;34(3):253-64.

125. Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement. 1977;1(3):385.

126. Comstock GW, Helsing KJ. Symptoms of depression in two communities. Psychol Med. 1976;6(4):551-63.

127. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can Comorbidity Be Measured By Questionnaire Rather than Medical Record Review? Medical Care. 1996;34(1):73.

128. Srivastava MS, Carter EM. An Introduction to Applied Multivariate Statistics. New York: North Holland; 1983.

129. Dobson AJ. An Introduction to Generalized Linear Models. Chapman & Hall/CRC; 2002.

130. SAS software. In: Cary, editor. Version 9 of the SAS System for Window ed. NC: SAS Institute Inc; 2002.

131. McCullagh P, Nelder JA. Generalized Linear Models. Chapman & Hall/CRC; 1989.

132. Wang XS, Fairclough DL, Liao Z, Komaki R, Chang JY, Mobley GM, et al. Longitudinal Study of the Relationship Between Chemoradiation Therapy for Non-Small-Cell Lung Cancer and Patient Symptoms. Journal of Clinical Oncology. 2006;24(27):4485.

133. Cox DR, Snell EJ. Analysis of Binary Data. Chapman & Hall/CRC; 1989.

134. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis. 1985;38(1):27-36.

135. Given C, Given B, Rahbar M, Jeon S, McCorkle R, Cimprich B, et al. Effect of a Cognitive Behavioral Intervention on Reducing Symptom Severity During Chemotherapy. Journal of Clinical Oncology. 2004;22(3):507.

136. Hess KR. Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression. Statistics in Medicine. 1995;14:1707-.

137. Gill R, Schumacher M. A simple test of the proportional hazards assumption. Biometrika. 1987;74(2):289-300.

138. Ng'andu NH. An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model. Statistics in Medicine. 1997;16(6):611-26.

139. McCall BP. Testing the Proportional Hazards Assumption in the Presence of Unmeasured Heterogeneity. JOURNAL OF APPLIED ECONOMETRICS. 1994;9:321-.

140. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-26.

141. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986;51(6):1173-82.

142. Tein JY, MacKinnon DP. Estimating mediated effects with survival data. New Developments in Psychometrics: Psychometric Society Proceedings. 2003:405-12.

143. Symptom Assessment Tools. MD Anderson Cancer Center; [cited 2008 July 15]; Available from: http://www.mdanderson.org/departments/prg/dIndex.cfm?pn=0EE78204-6646-11D5-812400508B603A14.