INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
Modeling Longitudinal Depression in Cardiac Patients in the Presence of Missing Data

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Amy K. Ferketich, M.A.

* * * * *

The Ohio State University

2000

Dissertation Committee:

Dr. Melvin L. Moeschberger, Adviser
Dr. Michael Browne
Dr. Deborah Burr
Dr. David J. Frid
Dr. Elizabeth Stasny

Approved by

Melvin L. Moeschberger
Adviser
School of Public Health
ABSTRACT

A growing body of evidence supports a rather strong link between depression and coronary heart disease, both incident events and complications in patients with existing disease. Depression has been examined in cardiac patients and the results from several studies show that depression in the hospital affects the prognosis and survival during the post-hospitalization period in patients with heart disease. One area of research that has not received much attention is the longitudinal change in depression following hospitalization in cardiac patients. The objective of this study was to examine depression during the two-year period following a hospitalization for heart disease. Following the initial hospitalization, patients were contacted at months 3, 6, 9, 12, 18, and 24 post-hospitalization. A total of 426 patients were included in the analysis. The change in depression level in this sample was difficult to estimate because close to half of the patients dropped out of the study over the two-year period. In addition, some patients had intermittently missing data. To handle the missing data, two multiple imputation procedures were used. Multiple imputation involves imputing $m$ values, instead of just one. The $m$ imputations for the missing values form $m$ completed data sets, which are analyzed using complete-data methods. The multiple imputation methods used in this analysis resulted in 20 completed data sets, and each data set was analyzed using a linear mixed effects model. The results from multiple imputation were compared to other methods of estimating the change in
depression that are often used when missing data are present. These methods include complete case analysis, last observation carried forward, and maximum likelihood on the observed data. The point estimates from multiple imputation were similar to those from the other methods, however the standard errors of the estimates from multiple imputation were much larger than those from the other methods. A simulation study was conducted to compare multiple imputation to maximum likelihood on the observed data under four different models for missing data. In all cases under study, maximum likelihood was less variable than multiple imputation, but within multiple imputation, as \( m \) increased, the variability decreased.
To Mario
ACKNOWLEDGMENTS

I would like to thank Dr. Melvin Moeschberger for his guidance and support throughout this process. I would also like to thank my other committee members, Drs. Michael Browne, Deborah Burr, David Frid, and Elizabeth Stasny, for the time they took to help me with this project and the thoughtful suggestions they provided at each meeting. My committee members were all extremely helpful, and without their insight and expertise I could not have completed this project. I would like to thank my parents, for their words of encouragement and support. Finally, I thank my husband, Mario, for the many times that he has lifted my spirits over the past five years while I have been pursuing this degree.
VITA

July 16, 1969 ................................. Born - Sellersville, PA

1992 .............................................. B.S. Nutrition and Dietetics

1992 .............................................. B.S. Physical Education

1994 .............................................. M.S. Exercise Physiology

1994-1996 ................................. Exercise Physiologist, Riverside Methodist Hospital

1996-present ................................. Graduate Research Associate, The Ohio State University.

PUBLICATIONS

Research Publications


**FIELDS OF STUDY**

Major Field: Public Health
   Epidemiology
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Literature Review</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Etiology and Pathophysiology</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1 Neurobiologic theories</td>
<td>5</td>
</tr>
<tr>
<td>2.1.2 Neurohumoral theories</td>
<td>6</td>
</tr>
<tr>
<td>2.1.3 Depression and Immune System Function</td>
<td>7</td>
</tr>
<tr>
<td>2.1.4 Cytokines and depression</td>
<td>11</td>
</tr>
<tr>
<td>2.1.5 Psychosocial Theories</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Diagnosis, Clinical Course and Treatment of Depression</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1 Factors Related to Depression</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Screening for depression</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Depression and Heart Disease</td>
<td>18</td>
</tr>
<tr>
<td>2.4.1 Depression Following Coronary Heart Disease (CHD) Diagnosis</td>
<td>18</td>
</tr>
<tr>
<td>2.4.2 Clinical Course and Outcome</td>
<td>22</td>
</tr>
</tbody>
</table>
2.4.3 Physiologic Mechanisms Linking Depression to Morbidity and/or Mortality in CHD Patients ...................................................... 29
2.5 Summary .................................................................................................. 35

3. Statistical Methods for Missing Data ............................................................. 37
3.1 Missing data mechanisms ...................................................................... 37
3.2 Imputation ............................................................................................... 39
  3.2.1 Proper multiple imputation ....................................................... 41
  3.2.2 Explicit versus implicit multiple imputation ................................ 42
  3.2.3 Inference from multiple imputation ........................................... 44
3.3 Multiple imputation for longitudinal data ......................................... 47
  3.3.1 Approximate Bayesian Bootstrap Methods using Propensity scores ............................................................................................. 47
  3.3.2 Multiple imputation for monotone missing data patterns using a longitudinal data model ................................................... 49

4. Methods ................................................................................................................. 52
  4.1 Data ............................................................................................................ 52
  4.2 Descriptive Statistics ................................................................................ 53
  4.3 Multiple Imputation ................................................................................ 54
    4.3.1 Multiple imputation for the intermittently missing data .......... 54
    4.3.2 Sequential multiple imputation for monotone missing data using a longitudinal data model ................................................... 59
  4.4 Complete-data model ............................................................................ 61
  4.5 Combining the complete-data estimates ............................................ 65

5. Results .................................................................................................................... 67
  5.1 Sample description ................................................................................... 67
  5.2 Multiple imputation results ................................................................... 76

6. Simulation Study ................................................................................................. 81
  6.1 Simulation Study Methods ...................................................................... 81
  6.2 Results ...................................................................................................... 86

7. Discussion ............................................................................................................ 93
  7.1 Study Limitations ................................................................................... 100
  7.2 Clinical Implications ................................................................................ 101
  7.3 Summary and Conclusions ..................................................................... 103

ix
7.4 Future Research ........................................... 104

Appendices:

A. Figures ......................................................... 106
B. Simulation Program ........................................ 113
C. List of Abbreviations ...................................... 133

Bibliography ...................................................... 135
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Baseline characteristics of sample and baseline BDI score by characteristic.</td>
<td>70</td>
</tr>
<tr>
<td>5.2 Drop-outs and intermittently missing scores at each evaluation time.</td>
<td>72</td>
</tr>
<tr>
<td>5.3 Parameter estimates and standard errors for the intercept, time and $\text{time}^2$ under the various models.</td>
<td>77</td>
</tr>
<tr>
<td>5.4 95% Confidence Intervals and $p$-values. These estimates were computed using Rubin's rules for combining estimates from multiple imputation.</td>
<td>78</td>
</tr>
<tr>
<td>6.1 Distribution characteristics of the intercept coefficient estimates.</td>
<td>91</td>
</tr>
<tr>
<td>6.2 Distribution characteristics of the time coefficient estimates.</td>
<td>92</td>
</tr>
<tr>
<td>6.3 Distribution characteristics of the $\text{time}^2$ coefficient estimates.</td>
<td>92</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Plot of Square-root Transformed BDI score averages and 95% confidence intervals at each evaluation time.</td>
<td>72</td>
</tr>
<tr>
<td>5.2 Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by drop-out status. Circles represent patients who dropped out between time $t$ and $t + 1$, triangles represent patients who did not drop out between time $t$ and $t + 1$.</td>
<td>73</td>
</tr>
<tr>
<td>5.3 Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by drop-out status, gender and smoking status. Circles represent patients who dropped out between time $t$ and $t + 1$, triangles represent patients who did not drop out between time $t$ and $t + 1$.</td>
<td>74</td>
</tr>
<tr>
<td>5.4 Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by intermittently missing status. Circles represent patients who were intermittently missing at time $t - 1$, triangles represent patients who were not intermittently missing at time $t - 1$.</td>
<td>75</td>
</tr>
<tr>
<td>5.5 Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by intermittently missing status. Circles represent patients who were intermittently missing at time $t + 1$, triangles represent patients who were not intermittently missing at time $t + 1$.</td>
<td>75</td>
</tr>
<tr>
<td>5.6 Plot of the estimated change in BDI scores from the four models.</td>
<td>80</td>
</tr>
</tbody>
</table>
6.1 Histograms of the intercept estimates from 1000 simulations under each model. MCAR refers to missing completely at random, MAR refers to missing at random, N1 is the nonignorable drop-out model that drops patients based on the score at time $t$, and N2 is the nonignorable model that drops patients based on scores from $t = 1, \ldots, T$. ML is the maximum likelihood estimate; $M=20$, $M=10$, and $M=5$ refer to the number of multiply imputed data sets that were combined to estimate the coefficient. The dotted line represents the complete-case estimate.

6.2 Histograms of the time estimates from 1000 simulations under each model. MCAR refers to missing completely at random, MAR refers to missing at random, N1 is the nonignorable drop-out model that drops patients based on the score at time $t$, and N2 is the nonignorable model that drops patients based on scores from $t = 1, \ldots, T$. ML is the maximum likelihood estimate; $M=20$, $M=10$, and $M=5$ refer to the number of multiply imputed data sets that were combined to estimate the coefficient. The dotted line represents the complete-case estimate.

6.3 Histograms of the $time^2$ estimates from 1000 simulations under each model. MCAR refers to missing completely at random, MAR refers to missing at random, N1 is the nonignorable drop-out model that drops patients based on the score at time $t$, and N2 is the nonignorable model that drops patients based on scores from $t = 1, \ldots, T$. ML is the maximum likelihood estimate; $M=20$, $M=10$, and $M=5$ refer to the number of multiply imputed data sets that were combined to estimate the coefficient. The dotted line represents the complete-case estimate.

A.1 Individual patient plots of BDI scores at each evaluation time.

A.2 Histograms of untransformed BDI scores at each evaluation time.

A.3 Histograms of square-root transformed BDI scores at each evaluation time.

A.4 Plot of the residuals versus predicted values for each time point using the values from the model fit to one of the multiply imputed data sets.

A.5 Normal scores plot of the residuals for each time point using the residuals from the model fit to one of the multiply imputed data sets.
A.6 Normal scores plots of the random effects for one of the multiply imputest data sets. ............................................. 112

A.7 Plot of the observed data versus the predicted values for one of the multiply imputed data sets. ................................. 112
CHAPTER 1

INTRODUCTION

An estimated 7 million Americans suffer from coronary heart disease (CHD), a type of heart disease that is caused by a narrowing of the arteries that supply blood to the heart. Each year, CHD takes the lives of more men and women than any other disease. Not only is CHD a problem in the U.S., but around the world as well. Data from the World Health Organization's MONICA project (monitoring trends in cardiovascular diseases) indicate that the leading cause of death in the world is cardiovascular disease, mainly due to heart attacks and strokes (WHO, 1999).

While studies such as the Framingham Heart Study have provided valuable information on cardiac risk factors for the medical community and the general public alike, not all cases of heart disease can be attributed to the standard risk factors. For example, half of all myocardial infarctions occur in patients for whom high cholesterol is not a problem (Ridker, Hennekens, Buring, and Rifai, 2000). Clearly, there must be additional factors that are responsible for heart disease.

A growing body of evidence supports a rather strong link between depression and CHD, both incident CHD and complications in patients with existing disease. Depression is widespread in the United States, with an estimated 19 million people
affected in 1998 (National Institute of Mental Health, 2000). Worldwide, it is perhaps the leading cause of disability.

Depression has been examined in cardiac patients. Results from several studies show that depression in the hospital affects the prognosis and survival during the post-hospitalization period in patients with CHD. The majority of studies in the literature were designed to measure depression with existing screening tools, and then classify individuals as either depressed or not depressed based on suggested cut-points. This method is problematic because the chance of misclassification is great for patients who score around the cut-point. In epidemiology, this is referred to as misclassification bias. Misclassification bias can alter effect estimates, biasing them either toward or away from the null. It therefore may be more desirable to use the scores as continuous variables.

Another problem with many of the studies reported in the literature is attrition. Depressed individuals may be more susceptible to loss to follow-up in longitudinal studies, because presumably depression would make patients less willing to comply with repeated evaluations. None of the studies reviewed in Chapter 2 have addressed this problem in the analysis. Some of the authors provided information on the number lost to follow-up and potential differences between those lost and the completers, however none of the analyses accounted for the uncertainty due to the missing data.

One area of research that has not been advanced is the longitudinal change in depression over time in cardiac patients. For example, it is not clear if the depression experienced in the hospital continues for an extended period of time following hospitalization, or if it diminishes quickly upon discharge.
This study will provide information on the longitudinal change in depression in a group of hospitalized cardiac patients. Over four hundred patients are included in the two-year follow-up study. Also, the analysis will address the problem of attrition. Close to half of the cohort dropped out by the end of the two year period. Results from an analysis that only included the complete cases may be severely biased, especially if the drop-outs are more likely to be depressed. Not only are there drop-outs in this sample, but many of the patients missed one or more of the examinations. Statistical methods for modeling drop-out in longitudinal studies exist, but the problem of intermittently missing data points was not specifically addressed in the literature.

In this report, methods for modeling the intermittently missing data and drop-outs will be proposed. Then, these methods will be applied to a longitudinal depression data set. Finally, the methods will be evaluated using simulation studies.
CHAPTER 2

LITERATURE REVIEW

2.1 Etiology and Pathophysiology

Depression is a condition characterized by a loss of interest and pleasure in activities, and by mood swings that range from euthymia (normal state of mood) to depression to recovery (Gruenberg and Goldstein, 1997). Individuals with depression experience feelings of sadness, irritability, dejection, despair, or loss of pleasure. Additional symptoms may be sleep impairment, loss of appetite, low energy levels, decreased libido or diminished psychomotor activity. Further, some individuals experience cognitive distortions about oneself, the surrounding world, and the future, accompanied by feelings of self-blame and uncertainty.

Sleep disturbances have been consistently reported in major depressive disorder (MDD) (Gruenberg and Goldstein, 1997). Sleep studies have shown that the progression from non-rapid eye movement to rapid eye movement sleep is disrupted, resulting in a reduced time to onset of rapid eye movement sleep. Also, the frequency of eye movements during rapid eye movement sleep is greater in MDD. Other
reported abnormalities include increased awakening during sleep, increased wakefulness, decreased arousal threshold, and early morning awakening (Kupfer and Thase, 1983).

Several theories have emerged to describe the etiology and pathophysiology of depressive disorders, particularly MDD. Some researchers have studied the effects of genetics and the environment on MDD. It is estimated that the heritability of MDD is about 50 percent (Gruenberg and Goldstein, 1997). Other risk factors for MDD include female gender, limited social support, stressful life events, and dependent, self-critical, and neurotic personality traits. One theory is that stressful life events are related to the initial or second episode of depression, however subsequent episodes can be explained by neurobiologic factors.

2.1.1 Neurobiologic theories

Neurobiologic theories have been advanced since the 1950's, however the interdependence of neurochemical systems is not understood well (Gruenberg and Goldstein, 1997). The monoamine neurotransmitter system has been studied extensively, and it is believed to play a central role in MDD (Gelder, Gath, Mayou and Cowen, 1996). This system involves dopamine, norepinephrine, and 5-hydroxytryptamine (5-HT) receptors. The findings on dopamine and norepinephrine have been inconsistent. However, depressed patients do appear to have impaired brain 5-HT function, which reverses in clinical recovery. The impairment may be due to the proinflammatory cytokines IL-1 and IFNγ (to be discussed in Section 2.1.3), which induce indoleamine-2,3-dioxygenase, an enzyme responsible for converting the amino acid L-tryptophan to 5-HT (Maes, 1999). Depressed patients have lower levels of plasma tryptophan
and a lower ratio of tryptophan to amino acids that compete with tryptophan in the brain (Maes and Meltzer, 1995).

Serotonin is another neurotransmitter which has been studied in relation to depression. Goodwin and Jamison (1990) reported lower concentrations of 5-hydroxyindoleacetic acid, the principle metabolite of serotonin, in the cerebrospinal fluid of suicide victims.

2.1.2 Neurohumoral theories

Neurohumoral theories have tried to establish a link between depression and hypothyroidism and hypercortisolism. Goodwin and Jamison (1990) reported a blunted thyroid-stimulating hormone (TSH) response to thyrotropin-releasing hormone (TRH) in about 30 percent of depressed patients. This accounts for the mild hypothyroidism associated with depression. Interestingly, TRH levels appear to be increased in patients with depression, and it is thought that TRH has a role in brain neurotransmission (Gelder et al., 1996). TRH is found in the brain in areas with monoamine transmitters, including 5-HT. Thus, the link between thyroid function and depression may be TRH, with hypersecretion of TRH leading to changes in monoamine transmitters.

One consistent finding in MDD has been elevated cortisol levels, likely due to hypothalamic-pituitary-adrenal axis hyperactivity (Gruenberg and Goldstein, 1997). Cortisol is released from the adrenal gland in response to corticotropin-releasing factor (CRH), which is produced in and released from the hypothalamus. In MDD, the rise in CRH levels is believed to be responsible for the associated hypercortisolemia. Several neurotransmitters, including norepinephrine, serotonin, and acetylcholine affect the activity of CRH. The elevation of CRH may be the result of stress and/or
neurotransmitter dysregulation. Hypersecretion of CRH could also arise from an altered feedback inhibition of CRH release by endogenous glucocorticoids, which normally act as negative regulators of the synthesis and release of CRH (Miller, Pariante and Pearce, 1999). Research showing nonsuppression of cortisol secretion following administration of dexamethasone, a synthetic glucocorticoid, in depressed patients suggests that glucocorticoid resistance is a very likely pathway by which depression is linked to hypercortisolemia. Cytokines (discussed in Section 2.1.3) are elevated in depressed individuals, and there is some evidence that they also have direct effects on glucocorticoid receptor function, leading to glucocorticoid resistance.

In the limbic regions of the brain, CRH is involved in the regulation of biochemical and behavioral responses to stress (Gelder et al., 1996). Thus, CRH may lead to depression directly through its neurotransmitter role. Animal studies have revealed that depressive-like symptoms appear after administration of CRH. In addition, some humans experience mood changes, either depression or mania, after receiving exogenous corticosteroids. Corticosteroids also regulate the function of monoamine receptors in the brain, leading to a decrease in the number of postysynaptic 5-HT receptors in the hippocampus (Deakin, 1991). Thus, excess cortisol secretion could possibly induce depression by decreasing 5-HT neurotransmission.

2.1.3 Depression and Immune System Function

Recently, much attention has been devoted to the study of depression and immune response. In a recent review on immune correlates of depression, Irwin (1999) reported that the earliest research findings suggested that the total numbers of white blood cells and neutrophils were elevated and the number of lymphocytes was lower among
depressed subjects. Two meta-analyses reported an association between lower lymphocyte proliferation and depression (Herbert and Cohen, 1993; Zorrilla, Luborsky, McKay, Rosenthal, Houldin, Tax, et al., 1998). Lymphoproliferation is measured by the degree to which lymphocytes divide in response to mitogens, with more proliferation associated with better immune functioning. Another finding reported in both analyses was a relation between depression and a decline in natural killer cell activity.

In a recent review, Maes (1999) summarized the findings on depression and the inflammatory response system. The inflammatory response is a reaction to one or a combination of the following: invasion of pathogenic microorganisms, a consequence of hypersensitivity reactions, invasion by malignant neoplasms, or tissue damage resulting from chemical or physical trauma (Pugh-Humphreys and Thomson, 1994). The reaction involves mobilization of leukocytes to the traumatized tissue. Depression is related to several indicators of inflammatory response system activation. For example, depressed patients tend to have a greater number of leukocytes, monocytes, neutrophils, activated T-lymphocytes, increased secretion of neopterin and prostaglandins. Two additional factors related to inflammatory response system activation in depressed patients are a higher concentration of soluble interleukin 2 receptor and an acute phase proteins. Maes (1999) states that the research findings indicate a relationship between depression and higher concentrations of the proinflammatory compounds and depression and lower concentrations of the anti-inflammatory compounds.

Depression is also associated with elevations in cytokine levels, specifically IL-6, IL-8, IL-1β, and IFN-γ. Cytokines are regulatory proteins released by white blood
cells and other cells in the body (Vilcek and Le, 1994). They are involved in regulating the inflammatory response system and their effects are seen on many different types of cells. An increased production of the proinflammatory cytokines IL-6, IL-1β, and IFNγ lead to immune activation (i.e. a rise in the number of peripheral blood mononuclear cells and acute phase proteins), and this is the most likely link between depression and inflammatory response system activation.

**Factors related to depression and immune function**

In his review on depression and immune function, Irwin (1999) describes the role of several potential confounding and/or effect modifying variables in the relation between depression and immune function. Age is one such variable. Older patients were, in some studies, at greater risk of depressed immune functioning compared to their younger counterparts (Herbert and Cohen, 1993). Other studies have failed to find an interaction between age and depression status on immune function (Irwin, Patterson, Smith, Caldwell, Brown, Gillin and Grant, 1990).

Gender is a factor that is related to depression (described below), with women at greater risk for developing depression compared to men. Therefore, it was hypothesized that perhaps the relation between depression and immune system function differed in men and women. Natural killer cell activity has been studied and the results have been mixed, with some results indicating an effect of depression on natural killer cell activity in men and no effect in women. Other studies have found the opposite: that is, depressed women exhibit reduced natural killer cell activity, whereas depressed men have similar activity compared to nondepressed men (Irwin, 1999).
Stress has an independent effect on immune system functioning, and because stress and depression are often experienced together, it is possible that the effect of depression on immune function is in fact due to psychological stress. Psychological stress can lead to increased concentration of circulating IL-6 (Zhou, Kusnecov, Shurin, DePaoli, and Rabin, 1993). IL-6 is a potent stimulator of C-reactive protein (CRP) synthesis in the liver (Yudkin, Kumari, Humphries, and Mohamed-Ali, 2000), and there is a positive relationship between plasma IL-6 and serum C-reactive protein when under stress (Dugue, Leppanen, Teppo, Fyrquist and Grasbeck 1993). University students who had a perception of stress while undergoing examination had higher levels of TNFα, IFNγ, and IL-6 compared to their less-stressed counterparts (Maes, Song, Lin, Gabriels, DeJongh, VanGastel et al., 1998). Irwin et al. (1990) failed to find evidence to support the hypothesis that the effect of depression on immune function is due to psychological stress. They found that the immune system changes in depressed patients could not be explained by the presence of severe life stress.

Specific symptoms of depression, such as melancholia, anxiety, weight loss, cognitive disturbance, diurnal variation (shifts in circadian rhythms) and sleep disturbance may account for some of the variance in the relationship between depression and immunity. The two symptoms that have been found to be more strongly related to natural killer cell activity are retardation and sleep disturbance (Cover and Irwin, 1994). Irwin and colleagues studied depressed patients and healthy controls, and in both groups the total amount of sleep and sleep efficiency (ratio of sleep to amount of time spent in bed) were positively correlated with natural killer cell activity (Irwin, Lacher and Caldwell, 1992). Thus, sleep disturbances may be a crucial link in the relation between depression and immune function. Experimental manipulation of sleep
deprivation has also resulted in changes in natural killer cell activity. Participants who are deprived of sleep during either the early part of the night or the latter part of the night/early morning, experienced declines in natural killer cell activity, lymphokine activated killer cell activity, and stimulated IL-2 production by peripheral blood mononuclear cells (Irwin, 1999).

2.1.4 Cytokines and depression

Although the literature suggests that elevated levels of proinflammatory cytokines result from depression, it is possible that in fact an overproduction of cytokines leads to depression (Maes, 1999). Rats that receive IL-1 and IL-6 exhibit a symptom cluster similar to that seen in depression: anorexia, weight loss, sleep disorders, suppression of social, locomotor, and exploratory behavior. Cytokine administration results in mood and behavioral disturbances similar to those seen in depression. This general syndrome has been termed 'sickness behavior'. Cytokines such as IFNa and IL-2 are becoming common treatments for certain cancers and viral infections. Valentine and co-authors reported that there is a risk of depression among patients undergoing IFNa therapy, and that the risk increases with longer duration of treatment, high-dose therapy, and history of psychiatric illness (Valentine, Meers, Kling, Richelson and Hauser, 1998). Cytokine administration may lead to the production of secondary cytokines that alter neuronal function, the release of stress hormones that cause mood disorders, and/or the development of thyroid disease (Meyers, 1999).

2.1.5 Psychosocial Theories

In addition to the biological-based theories, several psychosocial theories explaining the onset of MDD have been advanced. Psychoanalytic theory, put forth by Freud
(1957), connects mourning and melancholia, through the loss of an ambivalently held object. The theory maintains that when the object is lost, the melancholic person feels despair and, at the same, feelings of hostility attached to the object are redirected against oneself. The theory also maintains that self-esteem, developed early on, plays a central role in MDD. Attachment problems early in life and the impact of separation and loss of loved ones have also been proposed as factors that predispose the individual to depression later in life (Gruenberg and Goldstein, 1997).

Behavioral theory links social support to depression, claiming that a loss of social support leads to depression through feelings of isolation on the part of the patient (Gruenberg and Goldstein, 1997). The loss of social support is believed to be the result of negative responses, evoked by the depressed patient, from significant others. Behavioral treatments then focus on social skills training and self-monitoring of mood.

Cognitive theory focuses on cognitive distortions experienced by the depressed individual (Gruenberg and Goldstein, 1997). These distortions involve pessimistic views of one's self, one's world and current situation, and the future. Learned helplessness models and hopelessness theory further detail the cognitive perspective (Abramson, Metalsky, and Alloy, 1989; Seligman, 1975). In this view, stressful events are believed to be permanent and affecting one's entire life, rather than a specific aspect of one's life. Thus, personal factors explain the depressive disorder.

Interpersonal theory asserts that depression is the result of one's difficulty in adapting to the interpersonal environment (Gruenberg and Goldstein, 1997). The four potential problem areas relating to depression include grief, interpersonal role disputes, role transitions, and interpersonal deficiencies.
2.2 Diagnosis, Clinical Course and Treatment of Depression

Depression is diagnosed when there is a loss of interest and pleasure in activities or a mood disturbance that lasts for more than two weeks. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychological Association, 1994) lists several criteria which must be met for a diagnosis of MDD. Some of the more prominent symptoms are sadness, agitation, irritability, frustration, somatic preoccupation, reduced energy, decreased interest in sexual activity, weight loss due to loss of appetite, loss in the ability to concentrate, guilty preoccupation, and suicidal thoughts. Severity ratings are determined by the number and degree of symptoms present. Mild depression meets the criteria for MDD, however the individual only experiences mild impairment in psychosocial function. Moderate depression is diagnosed when all of the criteria are met and there is a moderate impairment in psychosocial function. Finally, severe depression occurs when the individual experiences symptoms in excess of the criteria for MDD, and the psychosocial impairment is judged to be severe.

The symptoms of depression are quick to develop, often within a period of days or weeks (Gruenberg and Goldstein, 1997). The early signs include anxiety, sleeplessness, and worry, before the final symptom of apparent sadness. Depression appears to be a recurrent condition: one episode carries a 50 percent risk of a second episode, two episodes carry a 70 percent risk of a third, and three or more episodes carry a 90 percent chance of future occurrences.

Treatment goals are to reduce or eliminate the symptoms of depression and to return the patient to full psychosocial functioning. Once these two goals are achieved, the patient will undergo a second phase of treatment which is directed at preventing
relapse via education counseling, pharmacotherapy, and psychotherapy. Finally, there is a maintenance phase which focuses on preventing future occurrences based on clinical knowledge of the events leading to recurrence (Gruenberg and Goldstein, 1997). In general, the maintenance phase is recommended only for those individuals with a history of two or more episodes of MDD, since this therapy calls for the use of mood-stabilizing drug treatment.

2.2.1 Factors Related to Depression

In most population studies, the prevalence of depression among women is twice that of men (Meagher and Murray, 1997). One explanation is that the reproductive cycle leads to physiological changes that promote depression. Childbirth has been linked to both affective psychosis and non-psychotic depression, however the underlying mechanism may be different in each condition. It is believed that dopamine sensitivity is related to affective psychosis, whereas psychosocial stressors are associated with non-psychotic depression. Another way in which gender may impact the risk of depression is through the immune system. In his review, Maes (1999) states that women have a greater immune responsiveness compared to men. Also, women taking oral contraceptives have a greater stress-induced immune response compared to women not taking such compounds.

Social factors are also believed to play a role in depression, for both men and women. In particular, intimate relationships, or lack of such relationships, may make one vulnerable to depression (Meagher and Murray, 1997). Low self-esteem is another risk factor for depression, and females are more likely than males to display this trait.
Early parental loss, rearing style of parents, unemployment, and degree of social support are all related to low self-esteem (Roy, Neale, and Kendler, 1995).

Coping styles, often quite different in men and women, may be related to depression. It is unclear if women rate stressful situations similarly as men do, and just respond to these situations with greater intensity, or if women rate the impact of stressful situations more severely (Meagher and Murray, 1995). Indeed, there is some evidence that men and women respond to stress differently. Men are more likely to respond with antisocial behavior or alcohol abuse, whereas women will more likely become depressed. This coincides with the finding that overall rates of psychiatric morbidity are similar in men and women, however depression is more common in women.

2.3 Screening for depression

Screening tools have been developed for use in the general population because in research, in particular, it would be very expensive to conduct psychological interviews to diagnose depression. These tools have the property of measuring depressive symptomology rather well, and they are easy to administer and cost relatively little. The Beck Depression Inventory (BDI) is the scale used in the present study, therefore its design and psychometric properties will be discussed below.

The BDI, which measures the depth of depression in individuals, was developed to be used in combination with clinical ratings for some studies and as the sole measure of depression for others (Beck et al., 1961; Beck, 1967). The inventory was designed to approximate clinical judgments of the intensity of depression, thus providing a standardized, consistent measure that is easy to administer and, the results of which,
could be compared to other quantitative data. The scale includes questions on all symptoms that define the depressive state, and each symptom is graded on the degree of severity experienced during the past week. A number is assigned to each symptom and the total score is the sum of the individual items. Thus, the score reflects a combination of the number and severity of symptoms.

The items in the inventory were clinically derived, from Dr. Beck's observations of depressed patients. Twenty-one categories are included, with each describing a specific behavioral manifestation of depression (Beck, 1967). Respondents are asked to grade each item based on feelings over the past week. All items are scaled from zero to three, with zero indicating that the symptom was rarely experienced and three corresponding to the symptom being experienced to a great extent. The scores, therefore, can range from 0 to 63. A score of 10 or higher has been selected to define depression, however this may lead to a great number of false positives, especially in clinical populations. Nielson and Williams (1980) determined, in their studies in three outpatient groups, that the optimal cut-point may be 13 for mild depression.

The initial testing was conducted on patients who were admitted to two psychiatric outpatient departments and a psychiatric inpatient service (Beck, 1967). Patients were given the Inventory and then seen by a psychiatrist. In half of the cases, the Inventory was administered first and in the other half, the patient saw the psychiatrist first. The psychiatrist (one of four who had reached a consensus on the criteria for categorizing depression severity prior to the start of the study) rated the patient on a four-point scale of the severity of depression. The split-half reliability test yielded a reliability of 0.86. Test-retest reliability was difficult to assess in this sample, as psychiatric patients experience fluctuations in depth of depression (Beck, 1967). Instead
of the usual test-retest reliability, a method of reliability was used that included, in addition to the Inventory, a clinical exam. The evaluations were conducted twice, separated by 2 - 6 weeks. Change in score on the Inventory paralleled changes in the clinical ratings.

Concurrent validity was assessed by comparing scores across individuals, grouped by degree of depression severity, to clinical ratings and to other depression scales. The mean scores were found to be different between the groups of depression severity (none, mild, moderate, and severe) (Beck, 1967). The Inventory scores were also found to be correlated with clinical judgements (r=0.65). Clinical ratings of change were related to change in depression score, with the agreement being 85 percent. Finally, BDI scores were correlated with other depression scales and these correlations were in the range of 0.40 to 0.66.

Construct validity was assessed by testing the theory that the depressive syndrome leads to high scores on the Inventory due to life experiences early on that predispose individuals to react to stress later by displaying depressive symptoms (Beck, 1967). This hypothesis was supported with research, as depressed individuals scored higher on negative self-concept scales, viewed themselves as the "loser", were more likely to experience childhood bereavement, and had pessimistic predictions.

The BDI has been tested in cardiac patients (Freedland, Lustman, Carney and Hong, 1992). In this study, depression was classified according to the National Institute of Mental Health Diagnostic Interview Schedule (DIS) (Robins, Helzer, Croughan and Ratcliff, 1981). The DIS is a comprehensive diagnostic instrument that can be administered either by lay interviewers or clinicians. The BDI scores of depressed and nondepressed patients were compared. The mean BDI scores among the depressed
and nondepressed were 13.6 (sd=5.4) and 6.2 (sd=4.0), respectively. The number of non-zero responses was also examined, and the average was 10.1 (sd=3.6) among the depressed and 4.8 (sd=2.5) among the nondepressed. Only 10 of the 21 symptoms included in the Inventory occurred more frequently in depressed patients, and 9 symptoms were experienced by at least 20 percent of patients in both groups. Further, when the symptoms were ranked according to frequency, the symptoms most common among the depressed were also most common among the non-depressed (Spearman correlation = 0.91). Finally, the authors reported that the symptoms in the depressed group were mild and nonspecific, as evidenced by the relatively low BDI score among those who were depressed according to the DIS. The authors noted that none of the ten symptoms occurring more frequently among the depressed were either highly sensitive or highly selective to major depression. Thus, they suggest that a more sensitive and specific depression screening tool be developed for use in cardiology and primary care practice.

2.4 Depression and Heart Disease

2.4.1 Depression Following Coronary Heart Disease (CHD) Diagnosis

Depression appears to be a common occurrence after a myocardial infarction is experienced. Fielding (1991) reviewed the literature and reported that depressive symptoms appear 2-3 days after an acute myocardial infarction. It is difficult to determine whether depression during hospitalization is due to true underlying depression, or just a mere reaction to being hospitalized. The feelings of helplessness that arise from being confined to the hospital may lead to depression. However, the
studies reviewed below provide evidence that depression following myocardial infarction may not be solely a reaction to the hospitalization. Rather, there is strong evidence that depression is more likely a physiological response to the myocardial infarction.

Schleifer and colleagues interviewed 283 patients 8 to 10 days after myocardial infarction, and then conducted follow-up interviews at 3 months post myocardial infarction (Schleifer, Macari- Hinson, Coyle, Slater, Kahn, Gorlin and Zucker, 1989). They used the Schedule for Affective Disorders and Schizophrenia (Endicott and Spitzer, 1978) to assess depression, which a structured interview assessment. Diagnoses of major and minor depression were made according to the Research Diagnostic Criteria (Spitzer, Endicott and Robins, 1978). These criteria require a depressed mood state for 1 to 2 weeks and the presence of associated symptoms such as sleep disturbance, loss of appetite, fatigue, and feelings of guilt or suicide. Eight to ten days post myocardial infarction, 45 percent of the sample met the criteria for either major or minor depression, with 18 percent meeting the criteria for major depression. The depressed patients reported having more stressful life events during the year prior to the hospitalization compared to the nondepressed. In addition, the depressed also reported a decreased sense of being in control of their lives, lower self-esteem, and elevated Framingham Type A scores than the nondepressed. Women were also more likely to meet the criteria for major depressive disorder than men. Three months post myocardial infarction, 83 percent of the nondepressed at baseline remained nondepressed, whereas 11 and 6 percent developed minor and major depression, respectively. Among the patients with minor depression at baseline, 64 percent were no longer depressed at the follow-up interview, and 22 and 14 percent were diagnosed with minor and major
depression, respectively. Those patients who had major depression at baseline were more likely to have major depression at three months than to remit: 44 percent remained in the category of major depression, whereas only 23 percent were no longer depressed and 33 percent had a diagnosis of minor depression. These latter results should be interpreted with some caution, since there were only 30 patients who had major depression at baseline.

Lesperance and co-workers followed 222 patients hospitalized with myocardial infarction for 18 months following their hospitalization (Lesperance, Frasure-Smith and Talajic, 1996). The DIS (Robins et al., 1981) was used to measure current and lifetime diagnoses of major depression, and the BDI was used to measure the severity of depression. Depression was diagnosed according to Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM-III-R) criteria, which requires a period of at least two weeks of depressed mood or loss of interest combined with symptoms of depression. Lifetime depression was assessed using a modified DIS. Patients had a positive history of depression if they reported at least one period where they felt sadness or lack of interest for at least two weeks. Sixteen percent of the patients were diagnosed with major depression in the hospital. During the one-year follow-up period, 21 percent of the patients not depressed at baseline became depressed, with the greatest risk of depression being during the first 6 months after the myocardial infarction. Twenty-eight percent of the total sample had a history of at least one episode of major depression prior to the hospitalization, but only 8 percent experienced an episode during the year preceding the myocardial infarction. Among the patients with depression in the hospital, 43 percent had a history of major depression. This is in contrast to the nondepressed patients: 25 percent of these patients had a
history of major depression. Thus, patients with a history of major depression are more likely to experience depression while in the hospital than patients without such a history.

Depression appears to be related to cardiac surgery. Burker and colleagues examined male and female patients undergoing cardiac surgery to determine the prevalence of depression, both before and after surgery (Burker, Blumenthal, Feldman, Burnett, White, Smith, et al., 1995). Prior to surgery, 46 percent of all patients were classified as depressed, according to scores on the Center for Epidemiologic Studies Depression scale (Radloff, 1977). One day prior to discharge, 61 percent of the percent of patients were depressed. Thus, this patient population may be at greater risk of developing depression than patients admitted for myocardial infarction.

Individuals diagnosed with CHD, but who have not yet had an event, may be at risk for developing depression. Hance and co-authors evaluated the course and outcome of depression in patients with CHD diagnosed by cardiac catheterization (Hance, Carney, Freedland and Skala, 1996). They used the DIS to evaluate depressive symptoms, and the DSM-IV criteria to diagnose major and minor depression. Of the 200 patients in the sample, 17 percent were diagnosed with major depression and 17 percent with minor depression. Among patients with major depression, 23 percent experienced full remission of depression during the one-year follow-up period, whereas 50 percent remained depressed. Fifty percent of the patients with minor depression fully recovered, whereas 42 percent of these patients progressed to major depression at one year. Patients who recovered were older and had less severe depression at baseline compared to those who remained depressed. The patients with minor depression who subsequently developed major depression were younger and had less severe CHD than
those who remitted. The authors concluded that depression tends to be persistent in this population, and not just a transient adjustment to the diagnosis of CHD.

There is some evidence of gender differences in the relation between CHD and depression. Freedland and co-authors reported that among patients with angiographically diagnosed CHD and major depression, women were more likely to report a history of major depression (Freedland, Carney, Lustman, Rich and Jaffe, 1992). Similarly, Lesperance and colleagues (Lesperance, et al. 1996) found that patients with a history of depression in their sample were more likely to be female. In the surgical sample, Burker et al. (1995) reported that 67 percent of the women were depressed prior to surgery and 76 percent were depressed post-surgery. This was higher than the estimates among the men in the sample, which were 40 and 55 percent, respectively. Neither Hance et al. (1996) nor Schleifer et al. (1989) reported gender differences in the course of depression following a CHD diagnosis.

2.4.2 Clinical Course and Outcome

Depression after myocardial infarction is associated with a poor outcome, leading to an increased risk in both fatal and nonfatal second myocardial infarction. Frasure-Smith and co-authors examined the effect of depression on survival following an myocardial infarction (Frasure-Smith, Lesperance and Talajic, 1993). Patients were classified as depressed if they met the DSM-III-R criteria for depression after completing the DIS. After six months, the relative risk (RR) estimate of CHD mortality associated with depression was 5.74 (95 percent CI 4.61 - 6.87). After adjusting for warfarin use (a blood thinner), lack of close friends, Killip class (a clinical estimate
of the degree of ventricular dysfunction), and previous myocardial infarction, the estimate decreased to 3.44 (95 percent CI 2.25 - 4.63). Later, the authors reported the results from the 18-month follow-up of this cohort (Frasure-Smith, Lesperance and Talajic, 1995). Unfortunately, they used a different model to estimate the effect of depression on mortality. In the latter study, they used a multiple logistic regression model, whereas in the former study a Cox proportional hazards regression model was employed. In any event, the adjusted odds ratio of mortality associated with depression during the 18-month period following myocardial infarction was 2.68 (95 percent CI 0.77 - 9.31). Interestingly, when they used scores on the BDI to classify depression, there was an effect of depression on 18-month cardiac mortality. After adjusting for previous myocardial infarction, Killip class, and arrhythmia, the odds ratio of fatal CHD among the depressed was 6.64 (95 percent CI 1.76 - 25.09). The authors noted that the in-hospital BDI scores may have predicted post myocardial infarction depression, whereas the DIS measures current depression. This would explain the association between elevated BDI scores and 18-month cardiac mortality. The DIS diagnosis was only associated with 6-month mortality, and the authors speculated that the deaths during the early post-myocardial infarction period may have been due to the in-hospital episode of depression.

Not only does depression adversely affect the prognosis in myocardial infarction patients, but it appears to be related to poor outcomes in patients with unstable angina as well. Carney and co-authors followed 52 males diagnosed with CHD for one year (Carney, Rich, Freedland, Saini, Te Velde, Simeone, and Clark, 1988). The DIS was used to assess the presence of symptoms of depression, and DSM-III criteria were used to diagnose major depressive disorder. Nine patients were diagnosed with
depression, and 7 of these patients experienced a cardiac event (either myocardial infarction, cardiac surgery, angioplasty, or death) during the follow-up period. In the nondepressed group, 15 out of 43 patients experienced an event. The proportion of events was significantly greater in the depressed group. This study provides some indication that depression may lead to events in patients with CHD, however due to the small sample size, the numbers should be interpreted with some caution. In a more recent and larger study, Lesperance and co-workers followed 430 patients diagnosed with unstable angina for one year (Lesperance, Frasure-Smith, Juneau, and Theroux, 2000). At the baseline assessment, which took place in the hospital, the BDI was administered. One year later, follow-up of this cohort was performed using telephone interviews. Patients who were depressed had an elevated odds of suffering a major cardiac event (cardiac death or myocardial infarction) compared to their nondepressed counterparts. The odds ratio was 6.73 (95 percent CI 2.43 - 18.64), and this was adjusted for electrocardiographic evidence of ischemia, left ventricular ejection fraction of 0.45 or lower, and the number of diseased vessels.

The effect of depression on mortality in CHD patients may extend beyond 12 - 18 months, which is the standard length in follow-up studies. Barefoot and co-authors reported the results from a study that followed patients with CHD, diagnosed by cardiac catheterization, for up to 19.4 years (Barefoot, Helms, Mark, Blumenthal, Califf, Haney, et al., 1996). The Zung Self-Rating Depression Scale was used to measure depression (Zung, 1965). This scale includes 20 items that measure depressive symptoms, and each symptom is rated according to how frequently it is experienced. To assess the overall effect of moderate/severe and mild depression on mortality, a
multiple logistic regression model was used. The authors were also interested in examining the differences in the effect of depression (moderate/severe and mild) over time. Therefore, they used a time-dependent Cox proportional hazards model to evaluate the risk of mortality due to depression during the following time periods: 1, 2 - 5, 5 - 10, and > 10 years. Compared to the nondepressed, those with depression had a reduced odds of survival during the follow-up period. The adjusted odds of mortality due to moderate/severe and mild depression were 1.69 and 1.38, respectively. The effect of depression was stable over time as well. Compared to the nondepressed, the patients classified as moderately/severely depressed had a 66 percent greater risk in the first year, an 84 percent greater risk during years 5 to 10, and a 72 percent greater risk after 10 years. The effect of depression was not significant during years 2 to 5. However, a test indicated that the effect of depression did not differ between the time periods. The authors concluded that this was the first study to demonstrate an effect of depression on long-term mortality. They also speculated that perhaps two different mechanisms are responsible for the short-term and long-term relations between depression and mortality. Short-term risk may be associated with physiologic processes that trigger cardiac events, whereas long-term risk may be due to processes that enhance progression of atherosclerosis, such as altered neuroendocrine function and poor compliance to treatment.

In addition to second cardiac events, depression may be related to reduced functional status in patients diagnosed with CHD by heart catheterization. Sullivan and colleagues followed 198 patients diagnosed with CHD for one year after the cardiac catheterization (Sullivan, LaCroix, Baum, Grothaus, and Katon, 1997). Patients completed a functional status questionnaire and depression was measured with the
Hamilton Rating Scale for Depression (Hamilton, 1960). The Hamilton Scale is a 24-item interviewed-administered questionnaire that assesses the current severity of depressive symptoms. Scores were divided into quartiles (1st quartile was the least depressed) and a repeated measures analysis of covariance was used to assess the effect of depression on functional status at baseline, 6, and 12 months following the CHD diagnosis. The results indicated that there were significant differences in physical function scores over time across depression quartiles, however all groups improved physical functioning at a similar rate. At baseline, the first quartile had significantly higher scores on the physical functioning measure compared to the other three quartiles, and at 12 months, this quartile only differed from the third and fourth quartiles. This study suggests that the importance of depression extends beyond the risk of death and cardiac events. It also has an impact on the daily lives of cardiac patients, and the authors suggest that management of depression should be incorporated into routine CHD care.

An indicator of recovery after an myocardial infarction is pain-free return to work and social activities. Ladwig and co-authors studied the effect of depression on post myocardial infarction recovery (Ladwig, Roll, Breithardt, Budde, and Borggrefe, 1994). They followed 522 male myocardial infarction patients for 6 months following their hospitalization. Depression was assessed with a scale that was designed and validated by the authors, and patients were classified as having either severe, moderate or low levels of depression. The 6-month follow-up evaluation included a standard 24-hour ambulatory ECG recording, a signal average ECG, an assessment of non-fatal events, and an assessment of vocational and social status. Only 377 patients were available for the follow-up interview. The authors reported that the patients lost to
follow-up were more likely to be blue-collar workers. The patients who were severely depressed had an increased odds of angina during the follow-up period, compared to the non-depressed (odds ratio 2.31, 95 percent CI 1.11 - 4.80). Smoking was more prevalent among the severely depressed during the follow-up period, the odds ratio of smoking among the depressed compared to the nondepressed was 2.84 (95 percent CI 1.22 - 6.63). Finally, both the moderately and severely depressed were more likely to suffer from anxiety during the 6-month period following the hospitalization, compared to the nondepressed. The odds ratio estimates (95 percent CI) were 1.87 (1.07 - 3.27) and 4.61 (2.32 - 9.18), respectively. The finding of the relation between depression and post myocardial infarction angina is particularly important, as the authors state that follow-up angina is a significant risk factor for recurrent infarction and negatively influences the quality of life in patients.

Rehospitalization is a proxy for post myocardial infarction morbidity. Levine and colleagues followed 210 cardiac patients who were admitted to the hospital for either myocardial infarction, angioplasty, or bypass surgery (Levine, Covino, Slack, Safran, Safran, Boro, et al., 1996). Depression was assessed with the BDI, and a score above 10 was used to classify mild depression. The outcome measure was the sum of the number of days the patient was rehospitalized for cardiac reasons during the 6-month period after the initial hospitalization. Depression was significantly related to days of rehospitalization, even after controlling for disease severity. In fact, there was a significant interaction between disease severity (as measured by left ventricular ejection fraction) and depression. The effect of depression was greater among patients with a lower ejection fraction than among patients in the normal range. Thus, this
study provides more evidence that depression predicts morbidity in post myocardial infarction patients.

Not only do depressed patients tend to be rehospitalized more than their non-depressed counterparts, but the costs associated with hospitalization appear to be higher as well. Frasure-Smith and co-workers examined the costs during the first admission and during the 12 months following the first admission in depressed and nondepressed patients, as defined by the BDI (Frasure, Smith, Lesperance, Gravel, Masson, Juneau, Talajic, and Bourassa, 2000). During the initial hospitalization, the cost for treating depressed patients was on average 11 percent higher than the cost for treating nondepressed patients. This increase was not due to a different distribution of major procedures. Rather, depressed patients remained in the hospital longer than nondepressed patients. During the one-year follow-up period, the costs associated with medical care among the patients depressed at baseline were 41 percent higher compared to the costs among the patients who were not depressed at baseline. Again, this was not due to different rates of major procedures such as surgery, angioplasty or catheterization. The depressed, compared to the nondepressed, were more likely to be readmitted to the hospital, visit the emergency room, and visit physicians.

Finally, depressed patients are perhaps less likely to follow behavioral change recommendations after myocardial infarction. Ziegelstein and co-authors measured depression with the BDI in 276 patients hospitalized for myocardial infarction (Ziegelstein, Fauerbach, Stevens, Romanelli, Richter, and Bush, 2000). Four months later, 204 patients from the original sample were reinterviewed. Patients who were classified as depressed in the hospital were less likely to adhere to a low-fat diet, regular exercise, and stress reduction techniques compared to patients who were not depressed.
at baseline. Furthermore, the depressed patients socialized less often and were less likely to adhere to prescribed medications than the nondepressed patients. The authors report that the latter findings cannot be explained by differences in the types of cardiac medications prescribed between depressed and nondepressed patients.

2.4.3 Physiologic Mechanisms Linking Depression to Morbidity and/or Mortality in CHD Patients

While the studies described in the preceding section point to an association between depression and adverse outcomes in CHD patients, the pathway by which depression leads to these outcomes is not entirely clear. Reduced heart rate variability (HRV), a measure of autonomic nervous system modulation, and depression have been associated (see below). Reduced vagal tone (impulses from the vagus nerve producing inhibition of the heart beat) results in decreased HRV, and it predisposes the myocardium to ventricular fibrillation in experimental animals (Magid, Eckberg and Sprendle, 1983). Decreased HRV has been found to predict mortality after myocardial infarction (Kleiger, Miller, Bigger and Moss, 1987) and in patients with congestive heart failure (Frey, Binder, and Teufelsbauer, 1993).

Carney and colleagues compared the HRV in depressed and nondepressed CHD patients (Carney, Saunders, Freedland, Stein, Rich and Jaffe, 1995). The primary measure of HRV was the standard deviation of normal RR intervals. An RR interval is the time between two heart beats, and it is measured as the time between two R waves of the QRS complex. The authors also looked at other indices of HRV for exploratory analyses, including measures of short-term variability that reflect vagal influences (the proportion of adjacent cycles differing by >50 ms, and the root-mean-square of successive differences), an index of intermediate-term variability (the
average of the standard deviations of all normal-to-normal intervals calculated for each 5-minute interval), and an index of long-term variability reflecting circadian and other rhythms (the standard deviation of the 5-minute averages of all normal RR intervals). The DIS was used to measure depression, and patients were diagnosed as depressed if they met the DSM-IV criteria for depression. Nineteen depressed CHD patients were compared to nineteen nondepressed patients who were matched on age, gender, and smoking status. Depressed patients had a significantly lower standard deviation of normal RR intervals, the primary measure of HRV used in this study, compared to the nondepressed patients. Also, the depressed patients had a significantly lower standard deviation of the 5-minute averages of all normal RR intervals, which was the measure of long-term variability. This latter measure may reflect increased sympathetic or decreased autonomic activity in depressed patients, and is the measure related to mortality in congestive heart failure patients (Frey et al., 1993). The primary measure, standard deviation of normal RR intervals, is related to mortality in myocardial infarction patients (Kleiger et al., 1987). The authors concluded that depression is associated with altered cardiac autonomic tone in CHD patients, but that prospective studies are needed to determine whether such alterations explain the increased morbidity and mortality seen in depressed patients.

Krittayaphong and co-authors examined the relation between HRV during daily life and depressed mood in CHD patients (Krittayaphong, Cascio, Light, Sheffield, Golden, Finkel, et al., 1997). Forty-two patients were enrolled in the study, and each patient underwent two 24-hour periods of ambulatory ECG monitoring. The standard deviation of the normal RR interval and average heart rate were the primary measures
in this study. Depression was measured with the Minnesota Multiphasic Personality Inventory, which includes a 60-item depression scale (MMPI-D). The MMPI-D is designed to measure depressed mood, and patients were classified by a median split of MMPI-D scores. Patients with high depression scores had a significantly lower standard deviation of normal RR intervals and significantly higher heart rate compared to the patients with low depression scores. The authors concluded that both of these measures could be related to elevated sympathetic activity, which promotes coronary vasospasm, leading to a decrease in oxygen supply to the myocardial tissue. Myocardial ischemia could result from the combination of the increased demand and decreased supply of oxygen. Furthermore, elevated sympathetic activity may alter the ventricular fibrillation threshold, making the myocardial tissue susceptible to life-threatening arrhythmias.

In a more recent study, further evidence of a relation between reduced heart rate variability and depression was reported (Stein, Carney, Freedland, Skala, Jaffe, Kleiger and Rottman, 2000). In their study, patients with CHD, but not a recent myocardial infarction, were evaluated. Depression was assessed with the BDI and HRV was measured with a 24-hour ambulatory monitor. All patients temporarily discontinued β-adrenergic antagonists and antidepressants for 3 days prior to the assessment of HRV. A total of 40 depressed and 30 nondepressed patients were examined. The depressed patients had higher heart rates and lower HRV on nearly all indices of HRV compared to the nondepressed patients. The authors reported that these results extend the findings from previous work in their lab to patients with stable disease who had temporarily discontinued all cardiac medications.
Another hypothesis for the relation between depression and CHD events has to do with platelet function. Two studies have reported a link between depression and abnormal platelet function. Musselman and colleagues studied platelet reactivity in depressed and nondepressed individuals who were free of CHD (Musselman, Tomer, Manatunga, Knight, Porter, Kasey, et al., 1996). They found that depressed patients exhibit enhanced platelet activation and responsiveness following a resting period and in response to a mild cardiovascular challenge.

Laghrissi-Thode and co-authors studied the effect of depression on platelet activation in cardiac patients (Laghrissi-Thode, Wagner, Pollock, Johnson, and Finkel, 1997). They compared the levels of platelet factor-4 and β-thromboglobulin levels among depressed cardiac patients to their nondepressed counterparts and nondepressed healthy adults. Platelet factor-4 and β-thromboglobulin normally occur in plasma in trace amounts, however in response to platelet activation the levels are elevated. Platelet activation may result from interactions with atherosclerotic plaques, immune complexes or thrombin generated by a hypercoagulable state. The depressed patients had significantly higher levels of both platelet factor-4 and β-thromboglobulin compared to nondepressed patients and healthy controls. There was no difference between the nondepressed patients and healthy controls, suggesting that this is not a manifestation of heart disease alone.

There is a growing body of evidence that the inflammatory response system and CHD are related. C-reactive protein is a marker of chronic inflammation and, as stated above, it is released in response to IL-1 and IL-6. Tataru and co-workers studied the relation between C-reactive protein levels and severity of atherosclerosis in myocardial infarction patients (Tataru, Junker, Schulte, von Eckardstein, Assmann,
and Koehler, 2000). Their sample included 1112 male and 299 female patients, and 326 male and 138 female controls who were matched to the cases on age. The patients had significantly higher C-reactive protein levels compared to the controls. Also, within the patient sample, the concentration of C-reactive protein increased with the number of blocked vessels.

The compounds IL-6 and C-reactive protein appear to be antecedents to CHD. A nested case-control study was performed using participants in the Physicians' Health Study (Ridker, Rifai, Stampfer, and Hennekens, 2000). This study is a randomized, double-blind, placebo-controlled trial of aspirin and β-carotene in the primary prevention of CHD and cancer. All participants are males, aged 40 to 84 years. Baseline blood samples were obtained prior to randomization, and these samples were thawed and IL-6 and C-reactive protein levels (as well as some other markers of immune function) were examined for this study. Participants who experienced an event during the 6-year follow-up period were matched on age, study length, and smoking status to one control subject. The baseline levels of IL-6 were higher among the men who had MI's during the follow-up period compared to the controls. The odds of experiencing an myocardial infarction during the follow-up period were 2.3 times higher among the men in the highest quartile of IL-6 concentration compared to those in the lowest quartile (95 percent CI 1.3 - 4.3).

Women with elevated levels of C-reactive protein are at greater risk of developing CHD. A similar nested case-control study was performed among women enrolled in the Women's Health Study (Ridker, Hennekens, Buring and Rifai, 2000). This study is an ongoing trial of aspirin and vitamin E for the primary prevention of cardiovascular disease and cancer. Cases were women who experienced a cardiovascular event during
a three-year period. Two controls were matched to each case on age and smoking status. C-reactive protein was found to be a strong predictor of cardiovascular events. The adjusted odds ratio of an event among those women in the highest quartile of C-reactive protein level compared to those in the lowest quartile was 1.5 (95 percent CI 1.1 - 2.1).

Danesh and colleagues conducted a nested case-control study using men enrolled in a prospective study in Britain (Danesh, Whincup, Walker, Lennon, Thomson, Appleby, et al., 2000). Between the beginning of the study, 1978-80, and 1995 there were 507 CHD events in men for whom baseline blood samples were available. Controls were frequency matched to cases on residence and age. The odds of a CHD event were 2.13 (95 percent CI 1.38 - 3.28) higher among men in the highest tertile of baseline C-reactive protein concentration compared to those in the lowest tertile, and this estimate was adjusted for known confounders. The authors stated that the low grade inflammation was not related to chronic infection, homocysteine concentration, or classic risk factors for CHD. The results from a study that specifically addressed the effect of 3 major infectious agents on inflammatory markers in patients hospitalized for either minor myocardial infarction or angina were recently reported (Choussat, Montalescot, Collet, Jardel, Ankri, Fillet, et al., 2000). Eighty-one patients were included in the sample. The inflammatory markers C-reactive protein, IL-6, fibronogen and serum amyloid A protein were elevated on admission to the hospital, and they increased further over the 48 hour period following admission. Neither the admission levels nor the levels after 48 hours of the inflammatory markers were related to the presence of one of the infectious agents studied.
Currently, there are no reports in the literature that link depression and immune function to CHD. Recall from above that depression and IL-6 are related, however it is not known whether elevated IL-6 causes depression or whether depression results in a rise in IL-6. Thus, the missing link between depression and CHD could quite possibly be immune function, specifically IL-6 and C-reactive protein. Indeed, much more research is necessary to determine just what mechanism mediates the path from depression to CHD.

2.5 Summary

This review was meant to provide an introduction to the vast body of research that has been performed in the area of depression and heart disease. Many questions have not yet been answered. For example, only theories have been advanced to describe the physiological mechanisms linking depression to heart disease. There are no large scale epidemiologic studies that have simultaneously studied all of the proposed factors linking depression to incident heart disease, or even depression to second events. Until such studies exist, clinical trials designed to reduce the risk of CHD among depressed patients cannot even begin.

Not only is the evidence on physiological mechanisms lacking, but also the evidence on how depression changes over time in CHD patients is not well defined. Two of the reviewed studies examined change in a crude way (Schleifer et al., 1989; Lesperance et al., 1996). In both studies, depression was assessed at baseline and then again at either 3 months (Schleifer et al., 1989) or 12 months (Lesperance et al., 1996). Patients were classified at both times according to cut-points. In general, cut-points are not a good way to model depression because there can be a great
deal of misclassification, especially around the defining point. Also, most of these cut-points have been established using samples of psychiatric patients or community samples, neither of which describes cardiac patients. Furthermore, males and females were grouped together in the samples and the cut-point may not be ideal for both genders. For example, women tend to score higher on many depression scales and it is not entirely clear why this happens. Women may be more depressed, or they may just respond differently to the questions.

Another limitation in interpreting the findings on depression and CHD is the failure of the investigators to account for missing data. In the Ladwig et al. (1994) study, 552 patients were included in the original sample, but only 377 could be examined at six months. Twelve died, and 163 refused to be interviewed. No account of the uncertainty in the estimates due to the missing data was made, or even considered. Similarly, in the Lesperance et al. (1996) study, patients were lost to follow-up and the authors even noted that more of the depressed patients were lost compared to the nondepressed. However, this was not accounted for in the analysis. If depressed patients are more likely to drop out than nondepressed patients, then the estimated change in depression over time may not reflect the truth if the analysis uses only complete cases or just the observed data is included in the analysis.
CHAPTER 3

STATISTICAL METHODS FOR MISSING DATA

In this chapter, missing data mechanisms and the methods that are currently being used to model incomplete data will be reviewed. Multiple imputation techniques will be the focus of the discussion. Section 1 will focus on missing data mechanisms. In Section 2, multiple imputation will be described, focusing on the work of Rubin (1987) and Schafer (1997). In Section 3, multiple imputation methods for longitudinal data analysis will be reviewed.

Throughout Chapters 3 and 4, the following notation will be used. A column vector will be denoted as an underlined letter, \( \underline{x} \), and a row vector as \( \underline{x}^T \). A matrix will be denoted as \( X \), a boldface uppercase letter, and the transpose of a matrix as \( X^T \).

### 3.1 Missing data mechanisms

Missing data fall into three general models, originally defined by Rubin (1976). Using the notation of Little and Rubin (1987), let \( Y \) denote the data that would occur in the absence of missing values. The data \( Y \) can be partitioned as \( Y = (Y_{\text{obs}}, Y_{\text{mis}}) \), where \( Y_{\text{obs}} \) denotes the observed values and \( Y_{\text{mis}} \) denotes the missing values. The joint distribution of \( Y_{\text{obs}} \) and \( Y_{\text{mis}} \) is denoted as \( f(Y|\theta) \equiv f(Y_{\text{obs}}, Y_{\text{mis}}|\theta) \). The marginal
distribution of \( Y_{\text{obs}} \) can be obtained by integrating out the missing data \( Y_{\text{mis}} \):

\[
f(Y_{\text{obs}}) = \int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta) dY_{\text{mis}}.
\]

For each component of \( Y \), define an indicator variable, \( R \), taking the value of 1 if the component is observed, and 0 if it is missing. If \( Y \) is an \((n \times K)\) matrix of \( n \) observations for \( K \) variables, the response indicator \( R_{ij} \) will equal one if \( Y_{ij} \) is observed, and 0 if missing, where \( i \) is the \( i \)th row and \( j \) is the \( j \)th column. Treating \( R \) as a random variable, the joint distribution of \( R \) and \( Y \) can be specified. The joint density is equal to the product of the density of the distribution of \( Y \) and the conditional distribution of \( R \) given \( Y \),

\[
f(Y, R|\theta, \psi) = f(Y|\theta)f(R|Y, \psi).
\]

The conditional distribution of \( R \) given \( Y \) has parameter \( \psi \). This conditional distribution is the distribution for the missing data. The distribution for \( Y_{\text{obs}} \) is obtained by integrating \( Y_{\text{mis}} \) out of the joint density \( Y = (Y_{\text{obs}}, Y_{\text{mis}}) \) and \( R \). That is,

\[
f(Y_{\text{obs}}, R|\theta, \psi) = \int f(Y_{\text{obs}}, Y_{\text{mis}}|\theta)f(R|Y_{\text{obs}}, Y_{\text{mis}}, \psi)dY_{\text{mis}}.
\]

If the distribution of the missing-data mechanism is independent of the missing values \( Y_{\text{mis}} \), that is, if

\[
f(R|Y_{\text{obs}}, Y_{\text{mis}}, \psi) = f(R|Y_{\text{obs}}, \psi),
\]

then,

\[
f(Y_{\text{obs}}, R|\theta, \psi) = f(R|Y_{\text{obs}}, \psi)f(Y_{\text{obs}}|\theta).
\]

The joint parameter space of \((\theta, \psi)\) factors into the parameter space of \( \theta \) and the parameter space of \( \psi \). If (3.5) holds, then the data are defined as missing at random.
MAR. MAR means that the probability that a component of $Y$ is missing does not depend on the value of that component when it is missing.

If the missing data are not at all related to the variables under study, then the missing data are said to be missing completely at random (MCAR) (Little and Rubin, 1987). Data are missing completely at random if the missing data are independent of the observed data.

The third missingness mechanism described in the literature is known as nonignorable nonresponse (NI) (Little and Rubin, 1987; Rubin, 1987). As the name implies, missing data are nonignorable if the probability that an observation is missing depends on the value of the missing observation.

A sampling distribution inference is one which compares the observed value of a statistic to the sampling distribution of the statistic under a hypothesized distribution. A likelihood inference involves evaluating ratios of the likelihood function for hypothesized values of the parameter. Rubin (1976) shows that if the missing data mechanism is MCAR, then the sampling distribution is only correct if the process that causes the missing data is completely known. Direct likelihood inference is correct even if the process that causes the missing data is ignored and the nonresponse mechanism is MAR. Direct likelihood inference only requires that the missing data are MAR and that $\psi$ and $\theta$ are distinct.

### 3.2 Imputation

Single imputation is a method for handling nonresponse that involves filling in one value for each missing data point. The benefits of imputation are that complete-data methods of analysis can be used on the data set containing the filled in values
and, in the setting of public-use data bases, the data collector's knowledge can be incorporated into the imputation procedure and the imputations only need to be carried out once (Rubin and Schenker, 1991). However, there are disadvantages to single imputation. Specifically, the single value being imputed does not incorporate uncertainty about the actual value. Even if the mechanism creating nonresponse is known, inferences based on the imputed data set will not reflect the extra variability due to the unknown missing values. In addition, correlations that depend on variabilities may be biased. Finally, in the case where the nonresponse mechanism is not known, no account is being taken for the uncertainty arising from not knowing which nonresponse models for imputation are appropriate (Rubin, 1987).

Multiple imputation shares the advantages of single imputation and corrects its problems (Rubin, 1987). The procedure involves imputing \( m \) values, instead of just one, and these \( m \) values are ordered so that the first set of values form the first completed data set, the second set form the second completed data set, and so on. The imputation procedure used for multiple imputation must be a stochastic method, rather than deterministic method. The \( m \) imputations for the missing values form \( m \) completed data sets which can be analyzed using complete-data methods. The \( m \) complete-data estimates are combined, and the variability among the results of the \( m \) analyses provides a measure of the uncertainty due to the missing data. This between-imputation variability is combined with the usual sample variance to make inferential statements about the parameters of interest. Rubin (1987) was the first to describe multiple imputation in the context of nonresponse in surveys.
3.2.1 Proper multiple imputation

In multiple imputation, $m$ typically does not have to be very large and the claim is that good results can be obtained with $m$ as small as 3-5. Rubin (1987) showed that even when the fraction of missing information is large, efficient estimates can be obtained when $m=5$. For example, when the fraction of missing information is 90%, an imputation procedure with $m=5$ is 92% as efficient, in terms of units of standard deviations, as using infinite $m$.

A multiple imputation procedure is 'proper' if the imputations created properly reflect sampling variability under a specific model (Rubin, 1987). Proper imputations lead to valid inferences. This concept becomes clear after considering an example. Suppose non-respondents and respondents with the same value of $X$ have $Y$ values that differ only randomly from each other (ignorable non-response). If values for $Y_{mis}$ are simply drawn from the $Y_{obs}$ values, then the variability will be underestimated because this method is not taking into account the sampling variability that arises from the fact that the sampled respondents' $Y$ values at $X$ differ randomly from the population of $Y$ values at $X$ (Rubin and Schenker, 1991). This variability must be properly accounted for in order to obtain valid inferences from the repeated-imputations under the assumed imputation model. Fully Bayesian methods are automatically proper.

Schafer (1997) describes imputations as Bayesianly proper if they are independent realizations of $P(Y_{mis}|Y_{obs})$, the posterior predictive distribution of the missing data under a complete-data model and prior. Factoring this conditional distribution yields

$$P(Y_{mis}|Y_{obs}) = \int P(Y_{mis}|Y_{obs}, \theta)P(\theta|Y_{obs})d\theta,$$
which is the conditional predictive distribution of $Y_{mis}$ given $\theta$, averaged over the observed-data posterior of $\theta$. Thus, Bayesianly proper imputations reflect the uncertainty about $Y_{mis}$ given the parameters of the complete-data model, as well as the uncertainty about the unknown model parameters. Because $P(Y_{mis}|Y_{obs})$ does not depend on $R$, the multiple imputations are only valid under the assumption of ignorable nonresponse mechanisms.

3.2.2 Explicit versus implicit multiple imputation

Multiple imputation models fall into two general classes: explicit and implicit models. An explicit model will estimate the distribution of $Y_{mis}$ under a Bayesian model (Rubin, 1987). The model should accurately represent the data so that a distribution for the parameters can be estimated. A random draw of the parameters is then made and, conditional on the drawn parameters, a distribution for $Y_{mis}$ is created. Imputations are then assigned as random draws from this conditional distribution of $Y_{mis}$. The procedure is repeated $m$ times for each missing value under the explicit model.

One example of an explicit model is a multiple linear regression model, where the univariate observations, $Y_i$, are independently distributed according to a $N(x_i^T \beta, \sigma^2)$ distribution. Here, $Y_i$ is the dependent variable, $x_i^T$ is the $i$th row of the $n \times p$ design matrix $X$, and $\beta$ is the $p \times 1$ vector of parameters. Box and Tiao (1973) show that the marginal posterior distribution of $\sigma^2$ is $(n-p)\hat{\sigma}^2/\chi_{n-p}^2$ and that the conditional distribution of $\beta$ given $\sigma^2$ is normal with mean $\hat{\beta}$ and covariance matrix $\sigma^2V$. Here $V = (X^TX)^{-1}$. Therefore, an explicit imputation method can work as follows (Rubin, 1987).
1. Draw $g$, a random observation from a $\chi^2_{n-p}$ distribution, and let $\sigma^2_* = \hat{\sigma}^2(n - p)/g$.

2. Draw $p$ independent $N(0,1)$ random variables to create a $p$-component vector $z$ and let $\beta_* = \hat{\beta} + \sigma_* [V]^{1/2} z$, where $[V]^{1/2}$ is the square root of $V$.

3. Impute the elements of $Y_{mis}$ as $Y_{i*} = x_i^T \beta_* + z_i \sigma_*$, where $z_i$ is a drawn $N(0,1)$ random variable.

These steps are repeated $m$ independent times to create $m$ imputations.

Implicit models work differently. First, respondents who are 'close' to nonrespondents are identified (Rubin, 1987). The respondents are then used as the basis for the distribution of $Y_{mis}$. Two implicit techniques that produce proper imputations are the Bayesian bootstrap (BB) and the approximate Bayesian bootstrap (ABB) (Rubin, 1987). The BB involves two steps. Let $n = n_0 + n_1$, where $n_1$ is the number of observations in $Y_{obs}$, and $n_0$ is the number in $Y_{mis}$. The first step involves drawing $n_1 - 1$ U[0,1] random numbers, say $a_1, a_2, \ldots, a_{n_1-1}$. These random numbers are then ordered from $a_1, \ldots, a_{n_1-1}$, letting $a_0 = 0$ and $a_{n_1} = 1$. At the second step, the $n_0$ missing values to impute for $Y_{mis}$ are drawn from $Y_1, \ldots, Y_{n_1}$ with probabilities $(a_1 - 0), (a_2 - a_1), \ldots, (1 - a_{n_1-1})$. For each missing unit in $Y_{mis}$, a random $U[0,1]$ is drawn. If this random number is between $a_{i-1}$ and $a_i$, then the value $Y_i$ from $Y_{obs}$ is imputed. For example, suppose $Y_{obs}$ consists of five numbers, and $Y_{mis}$ has one value missing. If $a_1 = 0.20, a_2 = 0.35, a_3 = 0.55, \text{and } a_4 = .78$, then $Y_1$ will be drawn with probability $(0.20-0)$, $Y_2$ with probability $(0.35-0.20)$, $Y_3$ with probability $(0.55-0.35)$, and so on. To impute the value for $Y_{mis}$, suppose the drawn random $U[0,1] = 0.65$. The imputed value will be $Y_4$, since 0.65 is between 0.55 and 0.78.
The ABB works as follows. First, among a total of \( n \) units with the same value of \( X \), there are \( n_1 \) respondents and \( n_0 \) nonrespondents. The ABB creates \( m \) repeated imputations by first drawing \( n_1 \) possible values with replacement from the \( n_1 \) values. Then, one value for each unit in \( n_0 \) is drawn at random from this sample of potential values. The drawing of the missing values from a set of potential values rather than the observed values ensures proper between-imputation variability, since the ABB step is roughly equivalent to choosing the parameters for the posterior predictive distribution from their posterior distribution (Rubin and Shenker, 1991). Non-ignorability can be built into this model for imputations by selecting the potential sample by some non-random mechanism.

### 3.2.3 Inference from multiple imputation

Using the notation of Rubin (1987), denote a scalar estimand as \( Q \). In multiple imputation inference, \( Q \) may be a function of the parameters in the imputation model or, in the case of sample surveys, \( Q \) may be a function of the data from a finite population. Let \( \hat{Q} \) be the complete-data point estimate for \( Q \); that is, \( \hat{Q} \) is the estimate that would have been obtained had there been no missing data. Let \( U \) be the variance estimate associated with \( \hat{Q} \), so that \( \sqrt{U} \) is the complete-data standard error. Both \( \hat{Q} \) and \( U \) are functions of \( Y = (Y_{obs}, Y_{mis}) \), therefore we can write \( \hat{Q}(Y_{obs}, Y_{mis}) \) and \( U(Y_{obs}, Y_{mis}) \). Multiple imputation inference assumes that \( \hat{Q} \) and \( U \) are first-order approximations to a posterior mean and variance of \( Q \) under an appropriate complete-data model and prior. Multiple imputation inference also assumes that the complete-data are sufficiently regular and the sample size sufficiently large for the
asymptotic normal approximation to be reasonable.

\[ U^{(-1/2)}(Q - \bar{Q}) \sim N(0, 1) \]  (3.6)

**Rules for combining estimates**

Rubin (1987) gives the rules for combining the \( m \) repeated complete-data estimates and associated complete-data variances for \( Q \) under the assumption of a single model for nonresponse. Let

\[ \bar{Q}_m = \frac{\sum_{i=1}^{m} \hat{Q}_{i}}{m} \]  (3.7)

be the average of the \( m \) complete-data estimates and

\[ \bar{U}_m = \frac{\sum_{i=1}^{m} U_{i}}{m} \]  (3.8)

be the average of the \( m \) complete-data variances, and

\[ B_m = \frac{\sum_{i=1}^{m} (\hat{Q}_{i} - \bar{Q}_m)^2}{m - 1} \]  (3.9)

be the variance between the \( m \) complete-data estimates. The total variance of \( (Q - \bar{Q}_m) \), \( T_m \), is

\[ T_m = \bar{U}_m + (1 + m^{-1})B_m. \]  (3.10)

Inferences are based on the approximation

\[ T_m^{-1/2}(Q - \bar{Q}) \sim t_\nu, \]  (3.11)

where the degrees of freedom, \( \nu \), are given by

45
\[ \nu = (m - 1)(1 + r_m^{-1})^2, \quad (3.12) \]

and \( r_m \) is the relative increase in variance due to nonresponse,

\[ r_m = (1 + m^{-1})B_m/\bar{U}_m. \quad (3.13) \]

Therefore, a 100(1 - \( \alpha \))% confidence interval estimate of \( Q \) is given by

\[ \bar{Q}_m \pm t_{\nu, \alpha/2}T_m^{1/2}. \quad (3.14) \]

The significance level associated with the null value \( Q_0 \) is

\[ P[F_{1, \nu} \geq (Q_0 - \bar{Q}_m)^2/T_m], \quad (3.15) \]

where \( F_{1, \nu} \) is an \( F \) random variable with 1 and \( \nu \) degrees of freedom.

The fraction of missing information about \( Q \) due to nonresponse is given by

\[ \gamma_m = \frac{r_m + 2/(\nu + 3)}{r_m + 1}. \quad (3.16) \]

Calculation of \( r_m \) and \( \gamma_m \) is recommended because they are helpful in assessing the effect of missing data on uncertainty about \( Q \) (Schafer, 1997).

When \( Q \) is a \( k \times 1 \) vector, hypotheses about \( Q \) can be tested and p-values computed. Often there is an interest in testing two models for the data, for example a full model \( M_F \) and a reduced model \( M_R \), where \( M_F \) reduces to \( M_R \) when \( Q = Q_0 \).

The test statistic for testing \( Q = Q_0 \) is

\[ D_m = (\bar{Q} - Q_0)^T \tilde{T}^{-1}(\bar{Q} - Q_0)/k, \quad (3.17) \]
where

\[ \tilde{T} = (1 - r_m) \tilde{U}. \]

When \( Q \) is a \( k \times 1 \) vector, \( r_m \) is defined as

\[ r_m = (1 + m^{-1}) \text{tr}(B \tilde{U}^{-1})/k. \]

The p-value for testing \( Q = Q_0 \) is

\[ p = P(F_{k, \nu} \geq D_m), \]

where \( F_{k, \nu} \) is an \( F \) random variable with \( k \) and \( \nu \) degrees of freedom. When \( Q \) is a vector, \( \nu \) is defined as

\[ \nu = 4 + (t - 4)[1 + (1 - 2t^{-1})r_m^{-1}]^2, \]

and

\[ t = k(m - 1). \]

The test statistic (3.13) assumes that the fractions of missing information are the same for all components of \( Q \). Even if this assumption is violated the procedure may be valid, but result in conservative p-values. Schafer (1997) states that for the procedure to work well, a large sample and an appropriate scale for \( Q \) is needed.

3.3 Multiple imputation for longitudinal data

3.3.1 Approximate Bayesian Bootstrap Methods using Propensity scores

Lavori, Dawson and Shera (1995) describe a method for imputing data in the setting of intent-to-treat analysis that relies on the propensity score. The propensity
score is the conditional probability of assignment to a particular treatment, when assignment to treatment is not random, given a vector of observed covariates (Rosenbaum and Rubin, 1984). Propensity scores are balancing scores that are useful in observational studies when the objective is to compare individuals who receive a given treatment to those who do not receive the treatment. A balancing score, \( b(x) \), is a function of the observed covariates \( x \) such that the conditional distribution of \( x \) given \( b(x) \) is the same for treated and control units. The finest balancing score is \( b(x) = x \), and the coarsest is the propensity score. Here, coarsest means that functions of \( x \) that are less fine than the propensity score are not considered to be balancing scores.

The Lavori et al. (1995) method is for monotone patterns of missingness. The propensity to remain on study through week \( t \) is estimated, where \( t \) is the first week where subjects drop out. A logistic regression model is used to estimate this propensity given the patient's observed trajectory through week \( t-1 \) and baseline covariates. The propensity scores are then stratified into quintiles, and within each quintile the approximate Bayesian bootstrap (ABB) method is applied. That is, within each quintile, a sample of observed scores is drawn with replacement from \( Y_{obs} \), and then one random draw for each value in \( Y_{mis} \) is taken from the drawn sample of scores.

When there is a monotone pattern of missing data, the data at week \( t + 1 \) can be filled in using the observed data and imputed values in the \( m \) completed data sets up to time \( t \). The propensity to remain on study through week \( t + 1 \) can be estimated separately in the \( m \) data sets and used to impute the missing week \( t + 1 \) responses by ABB draws within quintiles of propensity score. This procedure continues sequentially until all weeks are filled in.
3.3.2 Multiple imputation for monotone missing data patterns using a longitudinal data model

Little and Yau (1996) describe a method for multiple imputation that is based on a repeated-measures model. The motivation for their method was a need for an alternative to traditional intent-to-treat analyses, which include last observation carried forward and complete case analyses. The data they used were taken from a clinical trial that was designed to evaluate the effectiveness of a drug in the treatment of Alzheimer's disease. This was a multi-center trial and patients were randomized to one of three groups: a placebo group, a low-dose group, and a high-dose group. The objective was to see if the drug could prevent cognitive decline in these patients.

The multiple imputation model proposed by Little and Yau (1996) is as follows. Let \( y_i = (y_{i1}, y_{i2}, \ldots, y_{iT}) \) denote the vector of repeated measures for subject \( i \); \( T \) is the number of observation times in the study and \( y_{it} \) is the change in score between time \( t \) and baseline. The distribution of \( y_{it} \), given treatment randomization group indicators \( r_i \) and other covariates \( x_i \), is modeled as multivariate normal with mean \( \mu_{it} \) and covariance matrix \( \Sigma \), where

\[
\mu_{it} = E(y_{it} | r_i, x_i) = \mu_{it}^{(0)} + \beta_{t}^T r_i.
\]

Here \( \mu_{it}^{(0)} \) represents the mean of \( y_{it} \) for subjects in the reference (control) group, and \( r_i \) represents a \((K-1) \times 1\) vector of randomization group indicators for subject \( i \), with \( K \) being the number of treatment groups in the study. In Little and Yau's (1996) example, \( K=3 \), \( r_{i1} = 1 \) if the individual is assigned to the low-dose group and 0 otherwise, \( r_{i2} = 1 \) if the individual is assigned to the high-dose group and 0
otherwise, and $\delta^T = (\delta_{t1}, \delta_{t2})$ represent the treatment effects. The means $\mu_{it}^{(0)}$ are modeled as linear functions of other covariates $\mathbf{x}_i$. Thus,

$$\mu_{it}^{(0)} = \beta_0 + \tau_i + \lambda_{s(i),t} + \alpha_t \mathbf{x}_{it0} \quad (3.19)$$

where $\tau_i$ represents a time effect, $s(i)$ is the treatment site for subject $i$ so that $\lambda_{s(i),t}$ represents a site effect, and $\alpha_t$ is a linear regression coefficient on the baseline score.

When subjects drop out of the study, the treatments actually administered is unknown. The imputation model that Little and Yau (1996) propose conditions on the dose received after drop-out, but because this is not known to the investigator, the analysis becomes a sensitivity analysis. Thus, several models are used and the results under each are compared to determine how sensitive the results are to the dose after drop-out assumption. The model is as follows

$$(y_{it}, y_{i0}, y_{i1}, \ldots, y_{it-1}, \mathbf{d}_t, \mathbf{x}_{it}) \sim N(\mu_{it}, \sigma_{it,12\ldots,t-1}), \quad (3.20)$$

where $y_{i0}$ is the baseline score and $\mathbf{d}_t$ represents the indicators for the actual treatment for subject $i$ at time $t$. Equation (3.20) leads to the following,

$$\mu_{i1} = \beta_{01} + \lambda_{s(i),1} + \alpha_{1} \mathbf{x}_{i0} + \gamma_{1}^{T} \mathbf{d}_{i1},$$

$$\mu_{it} = \beta_{0t} + \alpha_{t} \mathbf{x}_{i0} + \gamma_{t}^{T} \mathbf{d}_{it} + \sum_{j=1}^{t-1} \beta_{tj} y_{ij},$$

for $t = 2, \ldots, T$. The vector $\gamma_{t}^{T} = (\gamma_{t1}, \gamma_{t2}, \gamma_{t3})$ represents dose-specific treatment effects for treatments 1, 2, and 3 at time $t$. The coefficients $\beta_{tj}$ relate the score at time $t$ with the scores at previous time points. The means are therefore represented
as linear functions of the baseline score $x_{i0}$ and site-specific effects, although site is an effect at time 1 only.

Little and Yau (1996) created imputations based on this model sequentially using the methods described above for explicit models (specifically, the linear regression example in Section 3.2.2). Thus, missing values at time 2 were imputed as draws from the predictive distribution of $y_{i2}$ given $y_{i1}$. Then, values at time 3 were imputed as draws from the predictive distribution of $y_{i3}$ given observed or imputed values of $y_{i1}, y_{i2}$. Values at time 4 were imputed similarly. Model parameters were drawn from their predictive distribution, which was created using the available data up to that time point. A residual variance was drawn as the residual sum of squares divided by a randomly drawn chi-square with degrees of freedom (df) equal to the residual df. The regression coefficients were drawn from a multivariate normal distribution centered at the parameter estimates with covariance matrix given by $(X^TX)^{-1}$ multiplied by the drawn residual variance (Rubin, 1987). Then a predictive distribution of missing values conditional on the drawn parameters was created.
CHAPTER 4

METHODS

4.1 Data

The data for this analysis come from a longitudinal study that was conducted at The Ohio State University Medical Center. The objective of the study was to assess psychosocial functioning in cardiac patients over time. Between August 31, 1995 and November 26, 1996, 666 patients were recruited to participate in the study. All patients were hospitalized for one of the following cardiac conditions: myocardial infarction, unstable angina, dysrhythmia, valve replacement, congestive heart failure, or heart transplant (a hospitalization after the transplant). Patients were not eligible to participate if they were scheduled for coronary artery bypass surgery. The subjects were not recruited randomly. Rather, the nurses approached patients if they were in their rooms during the recruiting time each day (which varied from day to day). The nurses approached the patients and explained the details of the study. Then, the patients decided whether they wished to participate, and if so, the baseline questionnaire was left with them to complete. Patients could either complete the baseline questionnaire in the hospital, or return it in the mail shortly after discharge.
The follow-up evaluations began three months after discharge. Mail questionnaires were sent at months 3, 6, 9, 12, 18, and 24 post-discharge. The questionnaire consisted of items relating to behaviors (such as smoking, alcohol use, physical activity, cardiac rehabilitation, and nutrition), medication use, functional status, symptoms, and a battery of psychosocial instruments. One instrument was the Beck Depression Inventory (BDI). In the present study, the BDI scores are the dependent variables.

Only patients who completed the baseline evaluation and at least one follow-up questionnaire were eligible to participate. This left 444 patients for the analysis.

4.2 Descriptive Statistics

Descriptive statistics for the sample will be calculated. Depression scores at each time will be presented by diagnosis category, gender, race, education level, history of heart disease, baseline smoking status, baseline physical activity status, and baseline alcohol consumption. Plots for depression scores over time will also be presented.

The pattern of drop-out will be evaluated next. First, the number of patients who are intermittently missing and who drop out at each time point will be calculated. Plots of scores at each time point (months 3, 6, 9, 12, and 18) for patients who did not drop out immediately after that time point and for patients who did drop out will be compared. These plots will be constructed for the entire group of patients, and by gender and smoking status. The information will be useful for detecting differences between drop-outs and completers at each time point. If the BDI scores at time $t$ are similar for patients who did drop out compared to those who did not drop out at time $t + 1$, then missing completely at random (MCAR) can be assumed. However, if the scores at time $t$ are consistently higher (or lower) for patients who dropped out
at time \( t + 1 \) compared to those patients who did not drop out, then either missing at random (MAR) or nonignorable nonresponse (NI) is the missing data mechanism.

4.3 Multiple Imputation

4.3.1 Multiple imputation for the intermittently missing data

The missing data pattern in this data set presents a problem because the pattern is not one of monotone missingness. Intermittently missing data, or nonmonotone missingness, creates a problem because an explicit model can no longer be used easily. For example, suppose there are two variables of interest that are subject to missingness, say \( Y_1 \) and \( Y_2 \), and one variable that is observed for all subjects, say \( X \). If the pattern of missingness is monotone, that is, if \( Y_1 \) is missing then \( Y_2 \) is as well, then a linear regression of the observed \( Y_1 \) on \( X \) can be used to impute the unobserved \( Y_1 \) values. Similarly, the observed \( Y_2 \) could be regressed on \( X \) and \( Y_1 \) to provide information for imputing unobserved \( Y_2 \). This type of modeling is not possible when some subjects have observed \( X \) and \( Y_2 \), but not \( Y_1 \) (Rubin, 1987).

One solution would be to discard the patients with intermittently missing data. In this data set, that would mean dropping 96 patients. The results could be biased if they were removed from the analysis. A different solution therefore had to be found in order to keep these patients in the analysis because they provided information about the effect of time after hospitalization on depression. They just did not provide information continuously up to the point of drop-out. One of Rubin's (1987) suggestions is to use a computationally convenient explicit model to impute the intermittently missing data. This is the method that will be used for the intermittently missing
BDI scores, with one modification. Instead of using an explicit multiple imputation model, an implicit model will be used.

The intermittently missing values will be imputed using an approximate Bayesian bootstrap (ABB) method. Recall from Chapter 3, the ABB is an implicit model that involves randomly drawing scores for imputation from a distribution of potential scores that is created from $Y_{obs}$.

It is unlikely that there will be adequate numbers of patients who are present and who are missing at each value of $x$, therefore an approximation to $x$ must be used. The survival probability at each time point is one such approximation. The survival probability is given by $S(t) = P(T > t)$, or the probability of an individual surviving beyond time $t$. The survival probability can be estimated from a Cox proportional hazards model (Klein and Moeschberger, 1997). A repeated event survival model will be fit to the data because each subject is allowed to come in and out of the study. Failure to complete a questionnaire will be considered an event, therefore each subject can have 0-5 events in this model. The model described by Andersen and Gill (1982) will be used to estimate the coefficients $\hat{\beta}$. This model is based on a multivariate counting process $N = (N_1, N_2, \ldots, N_n)$, where $N_i$ counts the observed events in the life of patient $i$. The events are treated as time-ordered outcomes. The model uses all of the data, however the observations are correlated so therefore the variance may be underestimated.

This Cox proportional hazards model is given by

$$h(t|x) = h_0(t) \exp \left( \sum_{k=1}^{p} \beta_k x_k \right),$$

where $h_0(t)$ is the baseline hazard rate and the sum of the coefficients $\sum_{k=1}^{p} \beta_k x_k$ accounts for the effects of the covariates (Klein and Moeschberger, 1997). A full model

55
will be fit to the data set that contains all of the patients, using all variables that are believed to be related to drop-out. The following variables will be included in the model: age, gender, race, diagnosis, education level, activity level at baseline, smoking status at baseline, history of heart disease, diabetes status, hypertension status, and baseline depression score.

Because each subject has the opportunity to experience more than one event under this model, the events are not independent. To account for this, a robust variance estimate must be computed. A grouped jackknife method is one such way to compute a robust variance. The method involves leaving out one subject at a time rather than one observation at a time. In S-plus (the program used to fit the model), the grouped jackknife variance estimate is computed the following way (S-Plus 2000 Guide to Statistics, 1999). First a leverage matrix is defined

\[ L = UI^{-1} \]

where \( L_{ij} \) is the approximate change in \( \hat{\beta}_j \) when group \( i \) is left out of the data set, \( U \) is the matrix of score residuals and \( I \) is the information matrix. The column sums of \( L = UI^{-1} \) are equal to the Newton-Raphson iteration step

\[ \Delta \beta = 1'(UI^{-1}) \]

The following approximation to the jackknife is used:

1. Treat the information matrix \( I \) as fixed.
2. Remove group \( i \).
3. Beginning at the full data solution \( \hat{\beta} \), do one Newton-Raphson iteration.
This procedure is the same as removing one row from $L$ and using the new column sums as the increment. The column sums of $L(\hat{\beta})$ are equal to zero, therefore $\Delta \beta = -L_i$. The rows of $L$ are an approximation to the jackknife, and the estimate $L'L$ is an approximation to the jackknife estimate of variance.

The event times were computed the following way. If a patient failed to complete a questionnaire at 3 months, then the event time was assumed to be 4.5 months, which is the midpoint of the interval between 3 and 6 months. If the true event time is unknown, it is common practice in epidemiology to assign the event time to the midpoint of the interval between two evaluation times. Similarly, missing questionnaires at months 6, 9, 12, and 18 were given event times at months 7.5, 10.5, 15, and 21, respectively. The last event time remained at 24 months, the time corresponding to the end of the study.

The Anderson and Gill (1982) model assumes that the multiple events for a given patient are independent. Therefore, no strata are included in the model to allow the baseline hazard to increase with increasing events. This is an obvious limitation if we assume that the risk for not completing a questionnaire increases if previous questionnaires are not completed (i.e. after a subject drops out). The problem with including a stratification variable is that the expected survival becomes rather difficult to estimate. To correctly estimate the survival probability at time $t$, it is necessary to use the baseline hazard from the appropriate strata. Thus, patients with several events will have survival estimates using different baseline hazard rates.

The survival probability at each time can be estimated from this survival model. After the model is fit to the data and estimates of $\hat{\beta}$ and $\hat{V}(\beta)$ (the covariance matrix
of the coefficients) are obtained, the baseline hazard rate is estimated using Breslow's estimator,

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{d_i}{\sum_{j \in R(t_i)} \exp(\hat{\beta}^T x_j)}$$  \hspace{1cm} (4.2)

where \(d_i\) is the number of events at time \(t_i\) and \(R(t_i)\) is the set of all individuals at risk just prior to \(t_i\). The estimator of the baseline survival function is given by,

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t)].$$  \hspace{1cm} (4.3)

For an individual with covariate vector \(x = x_0\) the survival function is given by

$$\hat{S}(t|x = x_0) = \hat{S}_0(t)^{\exp(\hat{\beta}^T x_0)}$$  \hspace{1cm} (4.4)

The Efron approximation (Efron, 1982) is used to estimate the probability when there are tied events. If there are two events at \(U\), then the term

$$\frac{2}{\sum_{j \in R(t_i)} \exp(\hat{\beta}^T x_j)}$$

in (4.2) would be replaced by

$$\frac{1}{\sum_{j \in R(t_i)} \exp(\hat{\beta}^T x_j)} + \frac{1}{\sum_{j \notin R(t_i)} \exp(\hat{\beta}^T x_j) + \frac{k-1}{d_i} \sum_{k \in d_i} \exp(\hat{\beta}^T x_k)}.$$

At each time point, the survival probability estimates for the entire sample will be stratified into quintiles and the intermittently missing and observed values in each quintile will be examined. Assuming that the intermittently missing values prior to drop-out are missing at random (MAR), in the sense of Rubin (1987), the participants who have observations at times prior to drop-out can be used as the basis for the ABB. For example, some individuals will have intermittently missing data at
3 months. Thus, they are missing at 3 months but return and then may drop out sometime following their return. Therefore, the subjects who have data at 3 months can be used as the basis for the ABB within a particular quintile for the subjects who do not have data within that quintile. A sample drawn with replacement from the observed values will serve as the sample of potential values from which draws for the missing values will be made.

This imputation procedure will be carried out \( m \) times. First, \( m=10 \) will be tried. If the parameter estimates are not stable with this value (after the full imputation procedure is performed), then \( m=20 \) will be examined. The result of the ABB method will be \( m \) data sets with a pattern of monotone missing responses. The next step will be to fill in the data after drop-out using a different multiple imputation model, described below, creating \( m \) complete data sets.

4.3.2 Sequential multiple imputation for monotone missing data using a longitudinal data model

The depression scores after drop-out will imputed using an explicit model. It is based on the work by Little and Yau (1996) that was reviewed in Chapter 3. Although the method was developed for intent-to-treat analysis, it can be generalized to observational data.

Let \( y_i = (y_{i1}, y_{i2}, \ldots, y_{iT}) \) denote the vector of repeated measures for subject \( i \); \( T \) is the number of observation times in the study and \( y_{it} \) is the score for patient \( i \) at time \( t \). This is a modification of Little and Yau’s method, where \( y_{it} \) was the change from baseline. Rather than imputing the change and adding that imputed change to the baseline score to get a value for time \( t \), the actual score at time \( t \) will be imputed.
The imputation model assumes the following:

\[(y_{it}|y_{i1}, y_{i2}, \ldots, y_{i(t-1)}, z_i) \sim N(\mu_{it}, \sigma_{it}^2; \mathbf{x}_{t-1}), \tag{4.5}\]

where

\[
\mu_{it} = \beta_0 + \sum_{j=1}^{t-1} \beta_{tj} y_{ij} + \sum_{k=1}^{p} \beta_k x_{ik},
\]

for \( t = 3, 4, 5, 6, 7 \). The range for \( t \) is 3-7 because the criteria for inclusion in the analysis was that the baseline and at least one follow-up questionnaire had to be completed. Therefore, the earliest time for drop-out was time 3, or 6 months. The coefficients \( \beta_{tj} \) relate the score at time \( t \) with the scores at previous time points and coefficients \( \beta_k \) represent effects for covariates \( z_i \). The means are therefore represented as linear functions of previous scores and covariates.

The imputation method for the scores after drop-out will be carried out using Rubin’s methods (Rubin, 1987).

1. Model parameters \( \beta_{tj} \)'s and \( \beta_k \)'s at each time point will be estimated using the complete data up to that time point. They will be estimated as

\[
\hat{\beta} = (X^TX)^{-1}X^Ty
\]

2. Next, parameters will be randomly drawn from the distribution of possible parameter estimates. First, a residual variance will be drawn as

\[
r_v = \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{\chi_{n-p}}
\]

where the numerator is the residual sum of squares and the denominator a randomly drawn chi-square with degrees of freedom (df) equal to the residual \( \text{df}=n - p \). The regression coefficients will be drawn as

\[
\hat{\beta}_{\text{drawn}} = \hat{\beta} + \text{chol}((X^TX)^{-1} * r_v) * z
\]
where $z$ is a $p \times 1$ vector of $N(0,1)$ random variables, $p$ is the number of variables in the model, and $chol$ refers to the Cholesky factorization.

3. Then a predictive distribution of missing values conditional on the drawn parameters will be created. Thus, each missing score will be imputed as

$$y_i = x_i^T \hat{\beta}_{\text{drawn}} + rv = z.$$ 

where $z$ is a drawn $N(0,1)$ random variable.

Imputations will be created sequentially. At each time, missing scores will be imputed as draws from the distribution of possible scores at that time, given the drawn parameter estimates, covariates, and previous scores.

The result of this step will be $m=10$ or $m=20$ completed data sets. Again, this will depend on the stability of the complete-data estimates. In the next section the model we used for the change in depression score over time will be described.

### 4.4 Complete-data model

A linear mixed effects model will be fit to each multiply imputed data set. This is a flexible model that allows for the estimation of both random and fixed effects. In this group of patients, the pattern of BDI scores over time varies from subject to subject. Some subjects start out with high scores and remain high whereas others start out high and decrease over time. Other subjects start out low and remain there for the duration of the observation period. Figure A.1 in Appendix A contains plots for 12 randomly selected patients. This is presented to illustrate the different patterns of change that were observed in the full sample. Because this was noted early on, a
linear mixed effects model was chosen because of its flexibility in incorporating these individual differences.

The general linear mixed effects model assumes that there are both subject level and observation level covariates. For example, in a designed experiment such as a multicenter clinical trial, the subject level covariates may be treatment and center. Each observation (within subjects) has covariates as well. In a clinical trial, these covariates may be time and functions of time. Subject level covariates thus remain constant for all the repeated observations on a single subject, whereas observation level covariates vary as the repeated observations are made (Cnaan, Laird, and Slasor, 1997). The repeated observations are often assumed to be correlated. This implies that the observation level covariates will be estimated with greater precision compared to the subject level covariates. In a longitudinal study the effect of time is usually the focus of the analysis, as well as the effects of subject level covariates on changes over time.

In the present study, the observation level covariates will be time and a quadratic function of time. Including a linear and quadratic term for time can be interpreted as a decrease in BDI score linearly with time, and the decrease is larger early on and then tends to level off. The leveling off is captured with the quadratic term. It was noted early on that the BDI scores tended to decrease during the first 6-9 months following hospitalization for heart disease, and after that time, the scores no longer decreased. The subject level covariates will include age, gender, history of heart disease, smoking at baseline, diagnosis of diabetes, diagnosis of hypertension, education status, and work status at baseline. These covariates were chosen because they are known confounders for depression and heart disease.
Model parameters for the general linear mixed effects model are estimated using a two-stage process (Cnaan et al., 1997). Let $Z_i$ be a $n_i \times q$ matrix of observation level covariates for subject $i$, where $n_i$ is the number of times subject $i$ is observed. Because the multiple imputation procedures will result in $m$ completed data sets, each patient will have the same number of observation times. Thus, $n_i = 7$ for all patients. In the first stage the following model is fit to the data

$$y_{ij} = z_{ij}^T \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, N$$

(4.6)

where $\alpha_i$ is the $3 \times 1$ vector of regression coefficients for subject $i$ (coefficients for the intercept, time, and $\text{time}^2$), $N$ is the total number of subjects, $z_{ij}^T$ is the $j$th row of $Z_i$, and the $\epsilon_{ij}$ are error terms assumed to have mean zero and variance $\sigma^2$. Correlation among the $\epsilon_{ij}$ can be modeled if the variation is believed to occur in a systematic way.

In stage 2, the $\alpha_i$ are regarded as $N$ independent 3-dimensional random vectors (Cnaan et al., 1997). The mean of the $\alpha_i$ is modeled as depending on subject level characteristics. Thus,

$$E(\alpha_i) = A_i \beta$$

(4.7)

where $A_i$ is a block diagonal matrix with diagonal elements equal to $(\beta_{1i}^T, \beta_{2i}^T, \beta_{3i}^T)$ and $\beta^T = (\beta_{1i}^T, \beta_{2i}^T, \beta_{3i}^T)$. The $\beta_k$'s are regression parameter vectors and their length depends on the number of subject level covariates that affect the $k$th coefficient. Very often, the number of predictor covariates is the same for each for each component of $\alpha_i$. In the present study, this is the case. Another assumption at this stage is that $\text{var}(\alpha_i) = D$, a $3 \times 3$ matrix. The diagonal elements of $D$ model the between-subject variance. Thus, they indicate the variability in the $\alpha_i$ after adjusting for subject level variance.
covariates. The following linear model can be used to specify $\alpha_i$,

$$\alpha_i = A_i\beta + b_i$$  \hspace{1cm} (4.8)

where the $b_i$ are i.i.d. random variables with zero mean and variance $D$. The $b_i$ are often described as random deviations from the mean. Thus, the vector $y_i$ can be written as

$$y_i = Z_i\alpha_i + \epsilon_i$$

$$= Z_i A_i \beta + Z_i b_i + \epsilon_i$$  \hspace{1cm} (4.9)

Equation (4.9) implies that $E(y_i) = Z_i A_i \beta$ and $\text{var}(y_i) = Z_i D Z_i^T + R_i$, where $R_i = \text{var}(\epsilon_i)$. Using this specification for the model, the vector $\beta_i$ is the vector of fixed effects, whereas the $b_i$'s are the random coefficients. The design matrix in (4.9) assumes that all observation level variables are specified as random effects. If this is not the case, for example if there are observation level covariates that are indicator variables as there are in cross-over trials, then the model must be specified differently. One possible specification involves setting some components of $D$ to be zero. Another method that is often used involves replacing $Z_i A_i$ with an arbitrary design matrix $X_i$ leading to

$$y_i = X_i \beta + Z_i b_i + \epsilon_i$$  \hspace{1cm} (4.10)

$$b_i \sim N(0, D),$$

$$\epsilon_i \sim N(0, R_i),$$

$$i = 1, \ldots, N.$$

The model assumes that the distribution of the outcome, BDI, is normal. To prevent problems with model misspecification, the distribution of the raw BDI scores
will be assessed by constructing a histogram of the depression scores at each time. The scores will likely be skewed, given previous experience with depression scores. A transformation will be applied to the scores in order to meet this assumption. The square-root transformation will be tried first, since there are some scores that are zero, which prohibits the natural log transformation.

The correlation between the errors will be modeled using a first-order autocorrelation structure. Thus,

$$\text{corr}(\epsilon_{ij}, \epsilon_{ij'}) = h(k, \phi) = \phi^k,$$

where \( k=(0,1,\ldots) \) represents the distance between measurements, and \( \phi \) is the single correlation parameter. In the present study, the times are not equally spaced after 12 months. Therefore, the first-order autocorrelation structure is only an approximation to the true correlation structure.

The complete data model will be fit using the NLME library in S-plus (Pinheiro and Bates, 2000). This library has several built-in features that automatically plot residuals to assess the assumptions of the errors terms and the random effects.

### 4.5 Combining the complete-data estimates

Each completed data set will produce parameter estimates \( \hat{\beta}(1), \ldots, \hat{\beta}(m) \) and variances \( \hat{\sigma}^2(1), \ldots, \hat{\sigma}^2(m) \). The estimates will be combined using the rules defined by Rubin (1987). The covariates \( \text{time} \) and \( \text{time}^2 \) were of major interest because the coefficients corresponding to these covariates indicate how the depression score is changing over time. The \text{intercept} is also of some interest because it represents the baseline depression score.
The estimates from the multiple imputation models will be computed, along with the estimates from models that used last observation carried forward (LOCF), complete case (CC) only, and the maximum likelihood method (ML) which analyzes observed data as incomplete. The LOCF method is a single imputation technique that fills in missing data using the most recent score. Thus, if a patient drops out, all scores after drop-out are assumed to be the score that was observed just prior to drop-out. Clearly, this method will underestimate the variance since the imputed scores do not change. The complete case analysis drops all patients with missing data. This method assumes that the missing data mechanism is MCAR. The ML method fits the model using all observed data. Missing data are easy to handle when using a linear mixed effects model. Recall, the \( n_i \), or the number of observation times for patient \( i \), do not have to be equal, and there is no requirement that the patients have to be measured on the same set of occasions. There is one assumption that has to be met however, and that is the missing data mechanism is MAR.
CHAPTER 5

RESULTS

5.1 Sample description

A total of 444 cardiac patients completed the baseline and at least one follow-up questionnaire. During the study, 18 patients were known to have died. The sample size for this analysis is therefore 426 patients. The patients who died had an average baseline score and standard deviation equal to 13.8 and 10.2, respectively. A description of the sample included in the analysis is presented in Table 5.1. The mean BDI score and standard deviation of the score at baseline for the entire sample were 10.3 and 7.7, respectively. The scores ranged from 0-40 and the median score was 8.

In this sample, there were more males than females, however the females had higher baseline BDI scores. This was expected since more men than women are hospitalized with heart disease, and depression is more prevalent among women. In this sample, the BDI score is inversely related to education, and the scores are higher among the working compared to the retired and unemployed. The BDI scores were also higher among nonwhites compared to whites. In regard to diagnosis, the patients hospitalized with heart failure have the highest scores and those with a history of
CHD have higher scores compared to those patients without such a history. Patients with heart failure tend to be older and more ill compared to patients in the other diagnosis groups. This could be the reason why they score higher on the BDI. The mean score among smokers was higher than the mean score among nonsmokers. The median scores among smokers and nonsmokers were equal however. This suggests that the distribution of scores for smokers was more skewed than the distribution for nonsmokers. The scores were higher among hypertensives compared to normotensives, and patients with diabetes compared to those without. This pattern is not unusual, and similar findings have been reported in other studies. Activity level and BDI score were inversely related. Finally, the scores among nondrinkers were higher than the scores among patients who consumed alcohol. Interestingly, but perhaps not surprisingly, patients who drank up to 14 drinks/week had the lowest BDI score. The recommendation for alcohol consumption is for up to two drinks per day. Thus, patients who drank more than this recommended amount scored higher, on average, on the BDI.

The raw BDI scores are highly skewed, which might lead to non-normal error terms after the model is fit. A square-root transformation was found to make the histograms appear closer to normal (see Figures A.2 and A.3 in Appendix A). The mean square-root transformed BDI scores and 95% confidence intervals are presented in Figure 5.1 for each evaluation time. Only the patients who completed the survey at each time are represented, thus a different number of patients are included in each calculation. This plot therefore does not accurately represent the change in depression in the entire sample since the BDI scores for the drop-outs and intermittently missing patients are not included. The plot does provide a rough estimate of the change in
depression scores over time. It shows that initially there is a decrease in the BDI scores, followed by a slower decline until the end of the study, 24 months.

The pattern of patient drop-out is presented in Table 5.2. A similar number of patients dropped out between 3 and 6 months, 6 and 9 months, and 9 and 12 months. A fairly large number of patients dropped out between 12 and 18 months. This could be due to the fact that the original study design involved follow-up until one-year post-hospitalization. After many subjects had been recruited, the research team decided to add on another year. Patients received a mail questionnaire 18 months following their hospitalization and a letter explaining that more data would be collected. It is possible that some patients made a one-year commitment and were not interested in anything beyond that.

The number of intermittently missing questionnaires at each time is also presented in Table 5.2. The greatest number occurred at 3 months. This is another function of the study design. The follow-up questionnaire was not ready for some of the subjects who were recruited early on. Therefore, they never received the first follow-up questionnaire.

The square-root transformed BDI scores and 95% confidence intervals among patients who dropped and did not drop out between time $t$ and $t + 1$ are presented in Figure 5.2. Except for the scores at 6 months, patients who dropped out between time $t$ and $t + 1$ had higher BDI scores at time $t$ compared to patients who did not drop out of the study between $t$ and $t + 1$. This is strong indication that the missing data due to drop-outs are not missing completely at random (MCAR). Figure 5.3 compares drop-outs between time $t$ and $t + 1$ to patients who did not drop out between these
<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>BDI</th>
<th>SD(BDI)</th>
<th>Median BDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>275</td>
<td>9.1</td>
<td>7.1</td>
<td>8</td>
</tr>
<tr>
<td>Females</td>
<td>151</td>
<td>12.4</td>
<td>8.2</td>
<td>11</td>
</tr>
<tr>
<td>Nonwhites</td>
<td>45</td>
<td>13.2</td>
<td>8.6</td>
<td>11</td>
</tr>
<tr>
<td>Whites</td>
<td>381</td>
<td>10.0</td>
<td>7.5</td>
<td>8</td>
</tr>
<tr>
<td>CHD History Absent</td>
<td>183</td>
<td>9.4</td>
<td>7.8</td>
<td>7</td>
</tr>
<tr>
<td>CHD History Present</td>
<td>243</td>
<td>11.0</td>
<td>7.5</td>
<td>9</td>
</tr>
<tr>
<td>Smokers</td>
<td>70</td>
<td>12.3</td>
<td>9.3</td>
<td>8</td>
</tr>
<tr>
<td>Nonsmokers</td>
<td>356</td>
<td>9.9</td>
<td>7.3</td>
<td>8</td>
</tr>
<tr>
<td>Hypertension Absent</td>
<td>182</td>
<td>9.7</td>
<td>7.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Hypertension Present</td>
<td>244</td>
<td>10.8</td>
<td>7.7</td>
<td>9</td>
</tr>
<tr>
<td>Diabetes Absent</td>
<td>318</td>
<td>9.8</td>
<td>7.7</td>
<td>8</td>
</tr>
<tr>
<td>Diabetes Present</td>
<td>108</td>
<td>11.8</td>
<td>7.5</td>
<td>10</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than HS</td>
<td>70</td>
<td>12.7</td>
<td>8.7</td>
<td>11</td>
</tr>
<tr>
<td>HS Diploma</td>
<td>149</td>
<td>10.2</td>
<td>7.2</td>
<td>8</td>
</tr>
<tr>
<td>Some College</td>
<td>207</td>
<td>9.6</td>
<td>7.5</td>
<td>8</td>
</tr>
<tr>
<td>Work Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currently working</td>
<td>158</td>
<td>12.2</td>
<td>8.4</td>
<td>10</td>
</tr>
<tr>
<td>Retired</td>
<td>176</td>
<td>9.2</td>
<td>7.3</td>
<td>7</td>
</tr>
<tr>
<td>Unemployed</td>
<td>92</td>
<td>9.2</td>
<td>6.6</td>
<td>8</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial Infarction</td>
<td>145</td>
<td>9.2</td>
<td>7.2</td>
<td>7</td>
</tr>
<tr>
<td>Unstable Angina</td>
<td>162</td>
<td>10.7</td>
<td>8.0</td>
<td>9</td>
</tr>
<tr>
<td>Dysrhythmia</td>
<td>53</td>
<td>10.6</td>
<td>7.7</td>
<td>8</td>
</tr>
<tr>
<td>Valve Replacement</td>
<td>13</td>
<td>8.2</td>
<td>5.9</td>
<td>6</td>
</tr>
<tr>
<td>Transplant</td>
<td>10</td>
<td>10.1</td>
<td>7.3</td>
<td>9.5</td>
</tr>
<tr>
<td>Heart Failure</td>
<td>43</td>
<td>13.0</td>
<td>7.9</td>
<td>11</td>
</tr>
<tr>
<td>Activity Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>195</td>
<td>11.6</td>
<td>7.9</td>
<td>10</td>
</tr>
<tr>
<td>Moderate</td>
<td>109</td>
<td>9.7</td>
<td>7.4</td>
<td>8</td>
</tr>
<tr>
<td>Active</td>
<td>122</td>
<td>8.7</td>
<td>7.4</td>
<td>7</td>
</tr>
<tr>
<td>Alcohol Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Drinks/Week</td>
<td>313</td>
<td>10.9</td>
<td>7.6</td>
<td>10</td>
</tr>
<tr>
<td>&gt; 0 to 14 Drinks/Week</td>
<td>94</td>
<td>8.4</td>
<td>7.7</td>
<td>6</td>
</tr>
<tr>
<td>15+ Drinks/Week</td>
<td>19</td>
<td>9.3</td>
<td>8.3</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 5.1: Baseline characteristics of sample and baseline BDI score by characteristic.
times, by gender and smoking status. The nonsmokers and males demonstrate a pattern that is similar to the entire data set. That is, other than the scores at 6 months, drop-outs tend to have higher scores at the time just prior to drop-out. This is not surprising, since the majority of the patients are male and do not smoke. Among females and smokers, the most striking difference is seen at 3 months. Patients in these groups who drop out at 6 months have much higher average scores than their counterparts who did not drop out at 6 months.

Figures 5.4 and 5.5 display the scores for patients by intermittently missing status. In Figure 5.4, average scores for patients who were intermittently missing at time $t-1$ are plotted against those for patients who were not missing at $t-1$. In Figure 5.5, average scores for patients who were intermittently missing at time $t+1$ are displayed in a similar way. There is no consistent pattern in scores among patients who were intermittently missing, compared to those who were not. The scores among intermittently missing patients are similar to the scores among all other patients, except for the intermittently missing at 12 months. The plot in Figure 5.5 indicates that the average BDI score at 9 months are higher among patients who were intermittently missing at 12 months compared to those who were not missing.
Table 5.2: Drop-outs and intermittently missing scores at each evaluation time.

<table>
<thead>
<tr>
<th>Time</th>
<th>Dropped between $t - 1$ and $t$</th>
<th>Intermittently missing at $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td>18</td>
<td>78</td>
<td>4</td>
</tr>
<tr>
<td>24</td>
<td>37</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 5.1: Plot of Square-root Transformed BDI score averages and 95% confidence intervals at each evaluation time.
Figure 5.2: Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by drop-out status. Circles represent patients who dropped out between time $t$ and $t + 1$, triangles represent patients who did not drop out between time $t$ and $t + 1$. 
Figure 5.3: Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by drop-out status, gender and smoking status. Circles represent patients who dropped out between time $t$ and $t + 1$, triangles represent patients who did not drop out between time $t$ and $t + 1$. 
Figure 5.4: Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by intermittently missing status. Circles represent patients who were intermittently missing at time $t - 1$, triangles represent patients who were not intermittently missing at time $t - 1$.

Figure 5.5: Plot of square-root transformed BDI score averages and 95% confidence intervals at each time by intermittently missing status. Circles represent patients who were intermittently missing at time $t + 1$, triangles represent patients who were not intermittently missing at time $t + 1$.  

75
5.2 Multiple imputation results

The multiple imputation procedures described in Chapter 3 were carried out using $m=20$ imputations. The coefficients were still quite variable, however. The simulation study results (described in the next Chapter) indicated that $m=20$ was better, in terms of lower variability, compared to $m=10$. Perhaps in this case $m=20$ will result in a better estimate than $m=10$. The linear mixed effects model described in Chapter 3 was fit to each of the 20 'complete' data sets. The covariates $time, time^2, gender, age, race, CHD history, baseline smoking status, diagnosis of diabetes, diagnosis of hypertension, education level and work status$ were included in the model. Each covariate is known to be related to both depression and heart disease. Other covariates were investigated as well. Baseline alcohol consumption and diagnosis were two such variables. It was thought that perhaps patients who consumed moderate amounts of alcohol would have lower BDI scores. This was not the case, however, since alcohol consumption was not significant in the model. Most of the patients were hospitalized for myocardial infarction or unstable angina. While the patients who were hospitalized for congestive heart failure had higher BDI scores at baseline, and are in fact known to be more depressed in general, the 5 degree of freedom variable $diagnosis$ (there were 6 diagnoses in all) was not significant in the model. Interactions between the functions of time and the other covariates included in the model were also examined. None were statistically significant, so therefore the model only main effects.

The model fits appeared to be fine. Residual plots are presented in Appendix A. Figure A.4 includes the plots of the residuals versus fitted values for each time point separately. The equal variances assumption does not appear to be violated. The QQ
plots for the residuals at each time point are presented in Figure A.5. The normality assumption on the error seems to be fine. The assumption of normality of the random effects appears to be met as well, the QQ plots for the random effects are presented in Figure A.6. Finally, Figure A.7 is a plot of the observed values versus fitted values. The predicted values and observed data agree fairly well.

The coefficients for the intercept, time and time$^2$ were estimated using Rubin’s rules for combining complete-data statistics (Rubin, 1987). Table 5.3 contains the parameter estimates from the multiple imputation (MI) procedure, as well as the estimates from the model that used the last observation carried forward (LOCF) method to impute missing BDI scores, the model that used the complete cases (CC) only, and the maximum likelihood (ML) method that modeled the observed data only. The estimate for the intercept and coefficient for time from the MI method are larger compared to the estimates from the other models. The time$^2$ coefficient from the MI method is the smallest among the four models.

The 95% confidence intervals and p-values for the three coefficients estimated with multiple imputation are presented in Table 5.4. These were computed using Rubin’s rules outlined in Chapter 3. Note the estimates for the time and time$^2$ coefficients.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\beta}_{int}$</th>
<th>$\hat{\sigma}_{int}$</th>
<th>$\hat{\beta}_{time}$</th>
<th>$\hat{\sigma}_{time}$</th>
<th>$\hat{\beta}_{time^2}$</th>
<th>$\hat{\sigma}_{time^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>3.2275</td>
<td>0.3738625</td>
<td>-0.013076</td>
<td>0.0128678</td>
<td>0.000469</td>
<td>0.0016057</td>
</tr>
<tr>
<td>LOCF</td>
<td>3.1964</td>
<td>0.3585913</td>
<td>-0.009384</td>
<td>0.0024296</td>
<td>0.000618</td>
<td>0.0002258</td>
</tr>
<tr>
<td>CC</td>
<td>3.1853</td>
<td>0.5584019</td>
<td>-0.012410</td>
<td>0.0032337</td>
<td>0.000919</td>
<td>0.0003453</td>
</tr>
<tr>
<td>ML</td>
<td>3.1595</td>
<td>0.3572464</td>
<td>-0.011208</td>
<td>0.0025417</td>
<td>0.000637</td>
<td>0.0002766</td>
</tr>
</tbody>
</table>

Table 5.3: Parameter estimates and standard errors for the intercept, time and time$^2$ under the various models.
The 95% confidence intervals include 0, suggesting that there is not a significant change over time. Recall from Chapter 3, the significance level associated with the null value $Q_0$ is $P[F_{1,v} \geq (Q_0 - \bar{Q}_m)^2/T_m]$, where $v$ is calculated as $v = (m - 1)(1 + r_m^{-1})^2$, and $r_m$ is the relative increase in variance due to nonresponse, $r_m = (1 + m^{-1})B_m/U_m$. The fraction of information missing about $Q$ is equal to $\gamma_m = \frac{r_m^{m+2}/(v+3)}{r_m+1}$. These components should be examined, as they are useful diagnostic tools for assessing how the missing data contribute to the inferential uncertainty about the coefficient (Schafet, 1997). In this study, the relative increase in variance due to nonresponse for the intercept is 0.239, and for the coefficients time and time$^2$ the estimates are 29.03 and 34.56, respectively. The fraction of missing information for the intercept is 0.175, and for time and time$^2$ the estimates are 0.969 and 0.974, respectively. Clearly, the missing data presents a greater challenge for estimating the trajectory of depression scores over time than it does for estimating the baseline mean (the intercept). Intuitively this makes sense because there is no missing data at baseline.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.4933</td>
<td>3.9616</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Time</td>
<td>-0.0399</td>
<td>0.0137</td>
<td>0.3215</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>-0.00288</td>
<td>0.00382</td>
<td>0.7732</td>
</tr>
</tbody>
</table>

Table 5.4: 95% Confidence Intervals and p-values. These estimates were computed using Rubin's rules for combining estimates from multiple imputation.

A plot of the estimated change in the depression score over time from the four models is presented in Figure 5.6. The plot shows that the MI method estimates the
greatest decrease in BDI score over time, and because the quadratic term is so small, there is more leveling off. The other curves are starting to increase after 18 months.
Figure 5.6: Plot of the estimated change in BDI scores from the four models.
CHAPTER 6

SIMULATION STUDY

6.1 Simulation Study Methods

This method of sequentially imputing longitudinal missing data has been presented in the literature (Little and Yau, 1996; Lavori et al., 1995), however the properties of the estimates have not been reported. There is potentially a problem because an implicit model is used to fill in the intermittently missing data, and then an explicit model is used to fill in the data after a monontone missing pattern is created. Furthermore, imputed values are being treated as the truth because previous observed and imputed values are being used to predict the scores after drop-out under the explicit model. Therefore, another part of this analysis will involve evaluating the properties of the proposed methods. The purpose here is to measure how close the estimated change in depression score over time, estimated from multiple imputation, is to what would have been observed had there been no missing data.

The properties of the estimates obtained following multiple imputation will be studied using the 195 complete cases. This will serve as the population or "truth" to which the results from each simulation study will be compared. Missing data will be
simulated from the complete cases under several models. Then, the missing data will be imputed using the methods proposed using \( m = 3, 5, 10, \) and 20.

The drop-outs will be simulated under missing completely at random (MCAR), missing at random (MAR) and nonignorable missingness (NI) patterns. The missingness patterns will be created as follows. First, a \( N(0, 1) \) random variable will be drawn for each patient. Then, drop-outs will be assigned according to the following rules (which will be elaborated on in future paragraphs)

1. \( MCAR : p(y_{it} \text{ missing}) = \Phi(\beta_0(t)) \)

2. \( MAR : p(y_{it} \text{ missing}) = \Phi(\beta_0(t) - \beta_1(t)(\sum_{k=1}^{T-1} y_{ik})) \)

3. \( NI_1 : p(y_{it} \text{ missing}) = \Phi(\beta_0(t) - \beta_1(t)y_{it}) \)

4. \( NI_2 : p(y_{it} \text{ missing}) = \Phi(\beta_0(t) - \beta_1(t)(\sum_{k=1}^{T} y_{ik})) \)

Here, \( T \) is the last observation time in the study (24 months), \( \Phi \) is the cumulative normal distribution function, and \( \beta_0(t) \) and \( \beta_1(t) \) will be chosen so that 5% of the participants will drop-out at immediately after 3 months, 5% after 6 months, 10% after 9 months, 15% after 12 months, and 10% after 18 months. Thus by the end of the study, slightly less than half of the subjects will have dropped out. This is similar to the pattern seen in the data described in this report.

The algorithm to drop patients according to MCAR will proceed the following way. First, each patient will receive a randomly drawn \( N(0, 1) \) variate. If this value is less than \( \Phi(\beta_0(t)) \), then that patient will be dropped at time \( t \). The following values will be used: \( \Phi(\beta_0(3)) = \Phi(\beta_0(4)) = -1.645, \Phi(\beta_0(6)) = -1.28, \Phi(\beta_0(6)) = -1.04, \) and \( \Phi(\beta_0(7)) = -1.28. \)
For the MAR simulations, $\Phi(\beta_0(t))$ and $\Phi(\beta_1(t))$ will be estimated using a linear regression model. Each patients' $y$ will be set to $z_\alpha$. The standard normal percentile, $z_\alpha$, will be chosen to correspond to the percentage of dropped subjects at each time. The $x$ in the regression model will be the sum of the previous BDI scores. For example, to drop out immediately after 3 months, each patient will be assigned $y = z_{0.05} = -1.28$, and the $x$ in the linear regression will be equal to the sum of the BDI scores at baseline and 3 months. To drop out immediately following 6 months, $y = z_{0.05} = -1.28$, and $x$ will equal the sum of the BDI scores at baseline, 3 months and 6 months. Patients will be dropped immediately following the 9, 12, 18 month evaluation similarly. Next, at each time, predicted values will be determined for each patient, $\hat{y}_i = \hat{\beta}_0(t) - \hat{\beta}_1(t)x$, using the parameter estimates from the linear regression model and the observed $x$. Patients will then be assigned a $N(0,1)$ variate at each time and if this value is less than $\hat{y}_i$ at time $t$, and if the patient has not already been dropped, then he/she will be dropped at time $t$.

The NI drop-out patterns will be created according to an algorithm that is similar to the MAR algorithm. Again, $\beta_0(t)$ and $\beta_1(t)$ will be estimated using a linear regression model, where each patient's $y$ will be equal to $z_\alpha$. The standard normal percentile, $z_\alpha$, will be chosen to correspond to the percentage of dropped subjects at a given time. The $x$ will be a function of the BDI scores observed after drop-out. That is, the missing data will depend on the unobserved scores. The $NI_1$ model will drop patients with probability proportional to the value at time $t$. So to drop immediately after 3 months, each patient will again be assigned $y = z_{0.05} = -1.28$. Then, $y$ will be regressed on $x$, the depression score at 6 months. Under the $NI_2$ model, $y$ will be regressed on $x$, where $x$ is the sum of all scores from 6 months until the end of the
study. Under both NI models, \( \hat{y} \) will be calculated for each patient. Then, a \( N(0,1) \) random variable will be drawn for each patient at each time. If the value of the drawn variate is less than \( \hat{y} \), the patient will be dropped.

After the drop-outs are assigned, intermittently missing data points will be simulated. At months 3, 6, 9, 12 and 18, approximately 15% of the observations will be set to intermittently missing. These observations will be set to missing under a MCAR mechanism. This estimate is slightly higher than what was observed in the present study.

For each study, 1000 repetitions will be performed. Each loop will begin by dropping patients and then creating intermittently missing data. Next, the linear mixed effects model that was described in Chapter 4 will be fit to the unbalanced data that results. This step is included so that the maximum likelihood (ML) method results can be compared to multiple imputation. Only time and \( \text{time}^2 \) will be included in the linear mixed effects model. This is because some of the models with other covariates did not converge. For some loops, too many patients in one particular group were dropped (for example, smokers, nonwhites, females) and therefore there were not enough observations to estimate the parameter for that particular group. Therefore, the simple model with time and \( \text{time}^2 \) was fit in each study to avoid this problem.

The next step will be to impute the intermittently missing observations using the ABB method described above. The covariates age, female, smoke, and CHD history will be included in the survival model. Again, fewer were included in the simulation study compared to the main analysis because of problems with models not converging for some parameters. Then, after a monotone missing data pattern is created, the observations after drop-out will be imputed using the generalization to Little and
Yau's (1996) method. All of the covariates listed in Chapter 4 will be included in the imputation model. These parameters were estimated using least squares, therefore there was no concern about models not converging. For each completed data set that is created, a linear mixed effects model will be fit to the data. The number of completed data sets that are created will depend on the value that \( m \) is set to. For each missingness model, \( m = (3, 5, 10, 20) \) will be studied. The multiply imputed intercept and the coefficients for the intercept, time and \( \text{time}^2 \) will be estimated using the rules for combining multiple imputation estimates. All of the simulation studies will be run in S-Plus.

The coefficients (intercept, time and \( \text{time}^2 \)) from each study will be presented in a histogram. For each study, the following will be calculated:

1. Bias: \( b = \beta - \frac{1}{1000} \sum_{i=1}^{1000} \hat{\beta}_i = \beta - \bar{\beta} \), where \( \hat{\beta}_i \) is the parameter estimate from the \( i \)th loop, \( \bar{\beta} \) is the average parameter estimate from the 1000 simulations, and \( \beta \) is the complete-data estimate.

2. Standard error: \( SE = \left( \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta}_i - \bar{\beta})^2 \right) ^{1/2} \).

3. Square root of the mean squared error: \( (\hat{\sigma}^2 + b^2)^{1/2} \), where \( \hat{\sigma}^2 \) is the estimated variance of the distribution of 1000 parameter estimates and \( b^2 \) is the square of the bias.

Each study will produce 1000 estimates for the intercept, time, and \( \text{time}^2 \) coefficients from the multiple imputation method and from the ML method which analyzes the unbalanced data before imputation. The squared deviations of these coefficients from the complete case coefficient estimate will be computed for the maximum likelihood and multiple imputation estimates. The difference of the squared deviations
will be compared using the Wilcoxon signed rank test. This is the appropriate paired
test because the distribution of the difference of the squared deviations is skewed.

6.2 Results

The multiple imputation procedures were evaluated using a simulation study. Missing data were simulated from the 195 complete cases under four models: \( MCAR \), \( MAR \), nonignorable drop-out \( NI1 \) where patients were dropped at time \( t \) with probability proportional to the score at time \( t \), and nonignorable dropout \( NI2 \) where patients were dropped at time \( t \) with probability proportional to the scores from time= \( t, \ldots, T \). Under each missingness model, five estimates were obtained for each coefficient: the maximum likelihood estimate from the model that used only the observed data and multiple imputation estimates using \( m = 3, 5, 10, 20 \). The coefficients for the intercept, time and time\(^2\) were examined.

The histograms of the 1000 coefficient estimates from the studies using \( m=5,10,20 \) and ML are presented in Figures 6.1-6.3. The histograms from the studies with \( m=3 \) are not presented in the trellis plot because their presence makes the other histograms too small to detect differences. The histograms of the estimates from the models with \( m=3 \) have heavier tails than the ones from the models with larger \( m \). From the trellis plot, it is clear that all of the distributions appear to be fairly normally distributed and centered at the point estimate, which is indicated by the dashed line. Clearly, the ML method produced parameter estimates that were less variable than the multiple imputation estimates. Within multiple imputation, the variability of the estimates decreased as \( m \) increased.
The summary statistics from each study for the coefficients of interest are presented in Tables 6.1-6.3. Bias is defined as the deviation from the complete-case estimate. For the intercept coefficient, there is no clear pattern of the bias estimates. There is a consistent finding for the standard error (SE) estimates and the square-root of the mean square (MSE$^{1/2}$) estimates. First, the ML estimates have lower SEs and MSE$^{1/2}$s compared to the multiple imputation estimates. Second, within multiple imputation, the SEs and MSE$^{1/2}$s decrease as $m$ increases. This is consistent with the interpretation from the histograms of the distributions of parameter estimates.

The characteristics of the time coefficient estimates are presented in Table 6.2. The results are similar to those for the intercept coefficient. That is, there is no consistent pattern for bias. However, for SE and MSE$^{1/2}$, the ML produces estimates with lower SE and MSE$^{1/2}$ compared to multiple imputation. Within multiple imputation estimates, SE and MSE$^{1/2}$ decrease as $m$ increases. The same statements can be made for the time$^2$ coefficient. All of the p-values from the Wilcoxon signed rank test are very close to zero. This is further evidence that the ML method resulted in less variable estimates.

87
Figure 6.1: Histograms of the intercept estimates from 1000 simulations under each model. MCAR refers to missing completely at random, MAR refers to missing at random, N1 is the nonignorable drop-out model that drops patients based on the score at time $t$, and N2 is the nonignorable model that drops patients based on scores from $t = 1, \ldots, T$. ML is the maximum likelihood estimate; $M=20$, $M=10$, and $M=5$ refer to the number of multiply imputed data sets that were combined to estimate the coefficient. The dotted line represents the complete-case estimate.
Figure 6.2: Histograms of the time estimates from 1000 simulations under each model. MCAR refers to missing completely at random, MAR refers to missing at random, N1 is the nonignorable drop-out model that drops patients based on the score at time t, and N2 is the nonignorable model that drops patients based on scores from t = 1, ..., T. ML is the maximum likelihood estimate; M=20, M=10, and M=5 refer to the number of multiply imputed data sets that were combined to estimate the coefficient. The dotted line represents the complete-case estimate.
Figure 6.3: Histograms of the $\text{time}^2$ estimates from 1000 simulations under each model. MCAR refers to missing completely at random, MAR refers to missing at random, N1 is the nonignorable drop-out model that drops patients based on the score at time $t$, and N2 is the nonignorable model that drops patients based on scores from $t = 1, \ldots, T$. ML is the maximum likelihood estimate; $M=20$, $M=10$, and $M=5$ refer to the number of multiply imputed data sets that were combined to estimate the coefficient. The dotted line represents the complete-case estimate.
<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>M=20</th>
<th>M=10</th>
<th>M=5</th>
<th>M=3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ML Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-9.4x10^{-4}</td>
<td>1.14x10^{-3}</td>
<td>8.87x10^{-4}</td>
<td>1.80x10^{-3}</td>
<td>4.16x10^{-4}</td>
</tr>
<tr>
<td>SE</td>
<td>5.00x10^{-4}</td>
<td>8.79x10^{-4}</td>
<td>1.04x10^{-3}</td>
<td>1.32x10^{-3}</td>
<td>1.56x10^{-3}</td>
</tr>
<tr>
<td>MSE^{1/2}</td>
<td>1.58x10^{-2}</td>
<td>2.78x10^{-2}</td>
<td>3.29x10^{-2}</td>
<td>4.18x10^{-2}</td>
<td>4.93x10^{-2}</td>
</tr>
<tr>
<td><strong>MAR Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-1.01x10^{-3}</td>
<td>5.6x10^{-4}</td>
<td>1.35x10^{-3}</td>
<td>4.22x10^{-4}</td>
<td>9.87x10^{-4}</td>
</tr>
<tr>
<td>SE</td>
<td>4.93x10^{-4}</td>
<td>8.6x10^{-4}</td>
<td>1.02x10^{-3}</td>
<td>1.34x10^{-3}</td>
<td>1.57x10^{-3}</td>
</tr>
<tr>
<td>MSE^{1/2}</td>
<td>1.56x10^{-2}</td>
<td>2.73x10^{-2}</td>
<td>3.25x10^{-2}</td>
<td>4.23x10^{-2}</td>
<td>4.95x10^{-2}</td>
</tr>
<tr>
<td><strong>N1 Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>1.18x10^{-4}</td>
<td>7.83x10^{-4}</td>
<td>1.46x10^{-3}</td>
<td>9.91x10^{-4}</td>
<td>1.22x10^{-3}</td>
</tr>
<tr>
<td>SE</td>
<td>4.87x10^{-4}</td>
<td>8.40x10^{-4}</td>
<td>1.01x10^{-3}</td>
<td>1.24x10^{-3}</td>
<td>1.66x10^{-3}</td>
</tr>
<tr>
<td>MSE^{1/2}</td>
<td>1.54x10^{-2}</td>
<td>2.66x10^{-2}</td>
<td>3.20x10^{-2}</td>
<td>3.93x10^{-2}</td>
<td>5.25x10^{-2}</td>
</tr>
<tr>
<td><strong>N2 Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>4.09x10^{-4}</td>
<td>5.61x10^{-4}</td>
<td>7.79x10^{-4}</td>
<td>2.51x10^{-3}</td>
<td>-9.83x10^{-4}</td>
</tr>
<tr>
<td>SE</td>
<td>4.84x10^{-4}</td>
<td>8.96x10^{-4}</td>
<td>9.95x10^{-4}</td>
<td>1.27x10^{-3}</td>
<td>1.54x10^{-3}</td>
</tr>
<tr>
<td>MSE^{1/2}</td>
<td>1.53x10^{-2}</td>
<td>2.83x10^{-2}</td>
<td>3.15x10^{-2}</td>
<td>4.40x10^{-2}</td>
<td>4.88x10^{-2}</td>
</tr>
</tbody>
</table>

Table 6.1: Distribution characteristics of the intercept coefficient estimates.

91
<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>M=20</th>
<th>M=10</th>
<th>M=5</th>
<th>M=3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCAR Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-1.34x10^{-4}</td>
<td>-1.25x10^{-4}</td>
<td>-2.1x10^{-4}</td>
<td>-3.62x10^{-5}</td>
<td>-1.42x10^{-4}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>2.09x10^{-6}</td>
<td>8.60x10^{-5}</td>
<td>1.05x10^{-4}</td>
<td>1.43x10^{-4}</td>
<td>1.74x10^{-4}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.45x10^{-3}</td>
<td>2.72x10^{-3}</td>
<td>3.32x10^{-3}</td>
<td>4.51x10^{-3}</td>
<td>5.50x10^{-3}</td>
</tr>
<tr>
<td><strong>MAR Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-9.22x10^{-5}</td>
<td>-4.86x10^{-5}</td>
<td>-4.66x10^{-5}</td>
<td>3.00x10^{-5}</td>
<td>-2.12x10^{-4}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>4.75x10^{-5}</td>
<td>8.53x10^{-5}</td>
<td>1.07x10^{-4}</td>
<td>1.42x10^{-4}</td>
<td>1.76x10^{-4}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.50x10^{-3}</td>
<td>2.69x10^{-3}</td>
<td>3.38x10^{-3}</td>
<td>4.49x10^{-3}</td>
<td>5.56x10^{-3}</td>
</tr>
<tr>
<td><strong>N1 Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-4.81x10^{-6}</td>
<td>-1.17x10^{-5}</td>
<td>-2.98x10^{-5}</td>
<td>-4.13x10^{-5}</td>
<td>-3.66x10^{-5}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>4.87x10^{-6}</td>
<td>9.77x10^{-6}</td>
<td>1.24x10^{-5}</td>
<td>1.67x10^{-5}</td>
<td>2.06x10^{-5}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.54x10^{-4}</td>
<td>3.09x10^{-4}</td>
<td>3.95x10^{-4}</td>
<td>5.29x10^{-4}</td>
<td>6.53x10^{-4}</td>
</tr>
</tbody>
</table>

Table 6.2: Distribution characteristics of the time coefficient estimates.

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>M=20</th>
<th>M=10</th>
<th>M=5</th>
<th>M=3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCAR Time^2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>6.13x10^{-6}</td>
<td>-3.68x10^{-5}</td>
<td>-3.55x10^{-5}</td>
<td>-4.41x10^{-5}</td>
<td>-2.24x10^{-5}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>5.18x10^{-6}</td>
<td>9.76x10^{-6}</td>
<td>1.29x10^{-5}</td>
<td>1.66x10^{-5}</td>
<td>2.07x10^{-5}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.64x10^{-4}</td>
<td>3.11x10^{-4}</td>
<td>4.09x10^{-4}</td>
<td>5.26x10^{-4}</td>
<td>6.55x10^{-4}</td>
</tr>
<tr>
<td><strong>MAR Time^2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>2.43x10^{-6}</td>
<td>-1.82x10^{-5}</td>
<td>-3.18x10^{-5}</td>
<td>-9.07x10^{-6}</td>
<td>-3.26x10^{-5}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>5.07x10^{-6}</td>
<td>1.00x10^{-5}</td>
<td>1.23x10^{-5}</td>
<td>1.66x10^{-5}</td>
<td>2.08x10^{-5}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.60x10^{-4}</td>
<td>3.17x10^{-4}</td>
<td>3.90x10^{-4}</td>
<td>5.24x10^{-4}</td>
<td>6.57x10^{-4}</td>
</tr>
<tr>
<td><strong>N11 Time^2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>-4.81x10^{-6}</td>
<td>-1.17x10^{-5}</td>
<td>-2.98x10^{-5}</td>
<td>-4.13x10^{-5}</td>
<td>-3.66x10^{-5}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>4.87x10^{-6}</td>
<td>9.77x10^{-6}</td>
<td>1.24x10^{-5}</td>
<td>1.67x10^{-5}</td>
<td>2.06x10^{-5}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.54x10^{-4}</td>
<td>3.09x10^{-4}</td>
<td>3.95x10^{-4}</td>
<td>5.29x10^{-4}</td>
<td>6.53x10^{-4}</td>
</tr>
<tr>
<td><strong>N12 Time^2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>1.11x10^{-5}</td>
<td>-1.07x10^{-5}</td>
<td>4.00x10^{-7}</td>
<td>-1.88x10^{-5}</td>
<td>1.56x10^{-5}</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>5.23x10^{-6}</td>
<td>1.05x10^{-5}</td>
<td>1.25x10^{-5}</td>
<td>1.63x10^{-5}</td>
<td>2.14x10^{-4}</td>
</tr>
<tr>
<td><strong>MSE^{1/2}</strong></td>
<td>1.66x10^{-4}</td>
<td>3.33x10^{-4}</td>
<td>3.94x10^{-4}</td>
<td>5.16x10^{-4}</td>
<td>6.76x10^{-4}</td>
</tr>
</tbody>
</table>

Table 6.3: Distribution characteristics of the time^2 coefficient estimates.
CHAPTER 7

DISCUSSION

The results from this analysis suggest that there is a decrease in depression, measured by the BDI, over the two-year period following a hospitalization for heart disease. This is the first study to evaluate the time course, in detail, of depression following hospitalization. Other studies have examined depression following hospitalization for heart disease, however these studies described the percentages of patients depressed at baseline and follow-up. They did not model the change over time. Schleifer et al. (1989), using a sample of 283 patients, reported that 83 percent of patients who were not depressed in the hospital remained nondepressed 3 months post-myocardial infarction. Among patients with minor depression, 64 percent were no longer depressed at follow-up, whereas 22 and 14 percent were diagnosed with minor and major depression, respectively. Similarly, Lesperance and co-authors (1996) used cut-points to classify depression both in the hospital and during the follow-up period. In their sample of 222 patients, 16 percent were depressed in the hospital, and among those not depressed in the hospital, 21 percent became so during the one-year follow-up period. They only had two follow-up evaluation times: 6 months and 12 months.
The present study went beyond these simple descriptives by evaluating a large group of patients with heart disease, the majority of which had either myocardial infarction or unstable angina. The average change in depression level, measured by the BDI, was modeled and therefore statements can be made about how depression changes over time in a large sample of patients. Modeling the scores in their raw form, as opposed to dichotomizing scores according to some criterion, allows one to see how the level of depression changes. For example, a pattern exhibiting a sharp decrease in scores within the first few months post-hospitalization followed by a more gradual decline and then a leveling off can be modeled using the scores as a continuous variable. Such a pattern may not be distinguishable if the scores are categorized because the scores may be close to the cut-point at baseline, or in the hospital in this study. Over time, while the average score may drop below the selected cut-point, and then continue to decline, the dichotomized classification would not reflect this pattern. Rather, we would only see that a majority of patients fall below the depressed level. We would not know whether they were close to non-diseased patients.

The results from this study must be interpreted with caution, however. There was a lot of missing data, and it is not entirely clear that the multiple imputation results accurately reflected what happened in the sample. Recall, the fraction of missing information for the coefficients $time$ and $time^2$ were both close to 0.97. This suggests that the multiple imputation methods did not do a good job at creating imputations that were consistent. The between-imputation coefficient estimates were quite variable, therefore the total variance was large because the total variance is a function of the within-imputation and the between-imputation variances. These findings suggest that the imputation model did not produce good predictions. Perhaps
there were important variables missing from the imputation model. For example, symptoms such as chest pain could perhaps predict depression well since the presence of chest pain means that the patient is not recovering well. If a patient does not recover well after a hospitalization, then he/she may be prone to depression.

Another set of variables that may predict depression well are perceived social support and stress. Recently, Bosworth and colleagues (Bosworth, Steffens, Kuchibhatla, Jiang, Arias, O'Connor, and Krishnan, 2000) reported that a lack of perceived social support increased the likelihood of depressive symptoms among patients with CHD. Lack of perceived social support was found to be predictive of depression one month following hospitalization among nondepressed patients hospitalized for heart disease (Brummett, Babyak, Barefoot, Bosworth, Clapp-Channing, Siegler, et al., 1998). Brummett and colleagues also found that stressful life events were predictive of depression among CHD patients. Ziegelstein and co-authors (Ziegelstein et al., 2000) reported that patients with depression following an acute myocardial infarction are less likely to adhere to stress reduction practices. Therefore, these two studies suggest that depression and stress are somehow related. It is not clear if depression causes stress, or vice-versa. In any event, current level of stress may be a good predictor for depression score.

Body mass index and smoking are other variables that could possibly be good predictors of BDI score. Body weight and height were not collected, either at baseline or at follow-up. These two variables are needed to calculate body mass index. Istvan, Zavela, and Weidner (1992) reported that, in a national sample, obesity was associated with elevated reports of depressive symptoms in women. Women in the highest quintile of body mass index had an elevated odds of being depressed compared to
women in the lower quintiles (odds ratio 1.38, 95% confidence interval 1.07-1.69). In a random sample of residents of St. Louis, depressed subjects were more likely to have a history of smoking and to be less successful in their attempt to quit smoking compared to their nondepressed counterparts (Glassman, Helzer, Covey, Cottler, Stetner, Tipp and Johnson, 1990). In another study, depressed smokers were 40 percent less likely to quit over a nine-year follow-up period compared to the nondepressed smokers (Anda, Williamson, Escovedo, Mast, Giovino, and Remington, 1990). These two studies suggest that depression and smoking are related in the general population. It is likely that the two are related in cardiac patients as well. Therefore, smoking could have been used in the imputation model using the responses at each follow-up.

Rehospitalization is also likely to be a good predictor of depression. Obviously, if depression scores were high at baseline in this sample, then they would likely be high again if a patient were to be rehospitalized. In fact, in the present data set a general pattern was seen when the depression scores at time $t$ for patients hospitalized between times $t$ and $t+1$ were plotted against patients not hospitalized during that time, and when the scores for patients hospitalized between time $t-1$ and $t$ were plotted against patients not rehospitalized between those two points. At all times, the patients who were rehospitalized had higher scores compared to those not rehospitalized.

Clearly, these covariates could not be included in the imputation model because they are time-dependent covariates. That is, their value changed over time and the value at any time depended on the responses to the survey questions. Therefore, if a depression score was missing, then so was the information on the time-dependent covariates. Therefore, they could not have been used in the model to create imputations. If the researchers had obtained information on rehospitalization from a source
other than the patient, for example from the patient's doctor, then this could have been used in the imputation model. Information from the doctor would perhaps be more accurate than the self-reported information gathered from the questionnaire. Unfortunately, the doctors were not contacted in the study.

Only fixed-time covariates could be used in the imputation model. It would have been possible to include covariates, such as social support and rehospitalization, at the three month evaluation if there had not been intermittently missing values at that time. The missing data at three months prohibited the use of time-dependent covariates up to three months because there was not complete data to estimate the parameters for these variables. Therefore, the values for the time-dependent covariates would have had to have been imputed for the patients with intermittently missing data. Clearly, this would have complicated the analysis if covariate values had been imputed. This is not to say, however, that this technique cannot be used. In the future I plan to look at methods that impute values for time-dependent covariates to see if the end result is a better imputation model for depression.

Increasing the number of imputed data sets ($m$) could have been another way to reduce the standard errors. The simulation study results suggests that the estimates become less variable as $m$ increases. It was not possible, however, to run the simulation study with a value of $m$ much larger than 20. For the studies with $m=20$, each loop took approximately 6 minutes to run. Of course, this time increased with each loop due to the nature of S-plus. Only 150 loops could be run at at time, otherwise the program would have slowed down too much. Without a simulation study, it can however be assumed that increasing $m$ would reduce the variability of the estimates.
The simulation study results indicate that the variability associated with multiple imputation was much greater than that associated with maximum likelihood estimation on the unbalanced data. However, the histograms demonstrated that the estimates from both methods had distributions centered very close to the complete-case estimate. The fact that both methods, and multiple imputation in particular, had small bias, as measured by the deviation from the complete-case estimate, is reassuring.

The design of the simulation study had some limitations, which could have affected the results. First, data were not simulated from a distribution whose properties were known. Rather, missing data were simulated from the 195 complete cases. The maximum likelihood estimate obtained from the model fit to these 195 cases was assumed to be the truth. Of course, this value is one of many possible estimates of the true value. The fact that this estimate was treated as the 'truth' could have caused the maximum likelihood estimates from the model fit to the unbalanced data to appear better than the estimates from the multiple imputation method. There was, however, no consistent finding in terms of bias. The maximum likelihood method had smaller bias in some cases, and larger bias in others.

The variability in the maximum likelihood estimates was substantially smaller than that in the multiple imputation estimates. In fact, the results from the signed rank test of the error mean squares from the two methods were highly significant in all cases. This means that in every simulation study, the maximum likelihood estimates fit to the unbalanced data, before the missing values were filled-in using multiple imputation, were significantly less variable than the estimates obtained after the data were imputed. Again, this suggests that the multiple imputation model did
not produce consistent estimates for the missing data. This was observed in fact. On several occasions, the multiply imputed estimates were examined for individuals. The values appeared to range greatly. For example, it was not unusual to see one value imputed as 0 and another as 15. Recall how the values were imputed. A residual variance was drawn as the residual sum of squares divided by a randomly drawn chi-square. The drawn residual variance was used in the process of randomly drawing the parameters, and then again used to impute a data value. If the residual sum of squares from the model is large, then there will be a great deal of variability from one imputed value to the next. This is because the imputed value is drawn as

\[ y_i = x_i^T \hat{\beta} + \tau v \ast z. \]

Thus, if the residual variance is large, then the range of drawn values will be as well.

It is not clear whether the multiple imputation estimates for time and time\(^2\) overestimated the standard errors or the maximum likelihood estimates from the observed data underestimated the standard errors. The estimated standard error for time from the multiple imputation method was 5 times larger than the standard error from the maximum likelihood method fit to the observed data. The estimated standard error for time\(^2\) was 5.8 times larger using multiple imputation. The standard errors from the maximum likelihood method fit to the observed data were similar to those estimates obtained from the last observation carried forward method. Indeed, this latter method underestimates the variance because the scores after drop-out do not vary at all, by definition. Likely, both possibilities are true, to some extent. That is, multiple imputation overestimated and maximum likelihood underestimated the standard errors. For the reasons stated above, the multiple imputation method
created widely variable imputed scores for the missing observations. Of course, the idea behind multiple imputation is to add uncertainty to the estimates. However, quite possibly, too much uncertainty was added in this case.

7.1 Study Limitations

There are several limitations in the study design and implementation that must be considered when interpreting the findings. First, and most importantly, this was not a randomly selected sample of patients. Rather, the patients were approached if they were in their rooms during the recruiting time each day. A much better design would have been to randomly select patients from a list of all patients admitted for one of the diagnoses included in the study criteria. The nonrandom selection is problematic because not all patients were given an opportunity to participate. It is quite possible that the sicker patients were not in their rooms during study recruitment because they were away having procedures performed. Since there is no information on the patients who were possibly eligible but not approached because they were not in their rooms, the differences between these patients and the ones that participated can only be speculated.

Another limitation has to do with the three-month evaluation for the first group of patients that entered the study. The questionnaire was not ready for them complete, so therefore no information is available for that select group at three months.

Almost all of the data in this study were obtained via self-report. This could potentially affect the quality of the data. For example, patients who are depressed may be in denial. Therefore, the responses to the questions on depressive symptoms may not be accurate. Again, no data were collected from the patients' physicians. The
physician could have provided information on the depression status of the patient. Other health care professionals could have provided more accurate information on depression status by using one of the interview-based depression instruments. These tools are less biased than the self-report questionnaires.

It is possible that the repeated administration of the BDI affected the responses. The patients received the same questionnaire every three months. Likely, many of them remembered the questions after seeing them several times. It is not unreasonable to think that some patients also remembered how they responded in the past, and were affected by past responses. For example, a patient might remember that he/she rated one symptom a two at the six month evaluation, and at nine months use this response again just to be consistent. Of course, there is no way to know if this happened in the present study.

7.2 Clinical Implications

The finding of decreasing depression score following hospitalization is encouraging for cardiac patients. Recall, in Chapter 2 the findings of the studies that have examined the relation between depression and post-hospitalization events were reviewed. Depression is associated with mortality in myocardial infarction patients (Frasure-Smith et al., 1995; Frasure-Smith et al., 1993) and patients with unstable angina (Lesperance et al., 2000). Complete recovery is also compromised in depressed patients (Ladwig et al., 1994). Depressed myocardial infarction patients have an elevated odds of experiencing angina compared to their nondepressed counterparts, and post-myocardial infarction angina has been related to recurrent events and reduced
quality of life. Given these findings, the decrease in depression following hospitaliza-
tion is likely beneficial to the well-being of patients.

The mean baseline score was 10.3, which is, on average, above the cut-point for
defining mild depression with the BDI. Among the 426 patients, 44.8 percent were
considered depressed according to this cut-point at baseline. Three months post-
hospitalization, the mean score was 9.9, still close to the cut-point, but only 40
percent of the patients were classified as depressed. It is difficult to truly know what
happened in this sample beyond three months, because that is when patients began
to drop out.

The results from the present study cannot be generalized to the entire population
of cardiac patients, however. This sample was obtained from an academic medical
center. Patients who attend these types of hospitals are different in at least two
ways from patients who attend community hospitals. One, they are usually of lower
socioeconomic status because these types of institutions are often located in low-
income sections of the city and they accept more patients without insurance compared
to other types of hospitals. Two, patients who attend an academic medical center
may be sicker because often patients are transferred in from outlying hospitals for
specialized care. The generalizability of the results is also affected by the nonrandom
sample. Because this was a volunteer sample, in some sense, the results cannot even
be generalized to all patients at The Ohio State University Medical Center (OSUMC).
Hopefully, the patients in this sample are not too different from all patients in the
cardiac unit at OSUMC. If this sample differs only randomly from the other patients
not included in the sample, then the results can be generalized to OSUMC cardiac
patients.
7.3 Summary and Conclusions

The multiple imputation procedure estimated a change in BDI score over time that was similar to the estimates obtained from the complete-case analysis, the last observation carried forward analysis, and the maximum likelihood method. The linear terms were in close agreement, however the multiple imputation procedure did not estimate as large a quadratic effect as the other methods. This resulted in an estimated change in BDI score that continued to decline for the duration of the study. The larger quadratic term estimated using the other methods made it appear as though the BDI score actually increased after 18 months. There is no logical reason to explain why the score may increase at 18 months.

There was a problem with multiple imputation, however, and that was related to the large standard errors associated with the estimates. The standard errors were much larger using multiple imputation than they were for the complete-case, last observation carried forward, and maximum likelihood analyses. The large variability was due to the between-imputation variance. The individual estimates for \( time \) and \( time^2 \) were quite variable, and as a consequence the fraction of missing information was very large. Likely, the imputation model did not produce good imputations. Rather, they were quite variable for each completed data set that was created. The simulation study supported the finding that was observed in the main analysis. That is, the estimates from multiple imputation were more variable than those obtained from maximum likelihood fit to the observed data.
7.4 Future Research

Future work should focus on developing good methods for multiple imputation in longitudinal data analysis. In particular, multiple imputation methods for intermittently missing data. Ideally, the imputation procedure should use information collected prior to and following the intermittently missing event. Currently, no methods that incorporate all of this information on the individual are presented in the literature.

The properties of multiple imputation should also be examined more closely. That is, future simulation studies should be performed to determine if and when multiple imputation outperforms maximum likelihood. Only four missingness models were examined, therefore no definitive conclusions can be made regarding multiple imputation versus maximum likelihood. Other patterns of missing data should be examined. For example, it is possible that if more subjects drop out later on, compared to early on, then multiple imputation will be better, in terms of variability, because fewer observations will be missing. Also, perhaps if the total number of participants who drop out of the study is less than what was examined in the simulation studies reported here, then multiple imputation will produce less variable estimates compared to maximum likelihood.

Methods to retain study samples should be investigated more closely. Less than half of the sample in this study had complete data. Perhaps there are ways of motivating participants to complete the study, such as providing incentives. Monetary incentives or offering a service such as a free physical examination performed by a physician might help to keep participants enrolled in the study. The use of incentives
in getting participants to provide reasons for drop-out is also worth investigating. Information related to drop-out could be used in the imputation model and this might lead to better imputations.
APPENDIX A

FIGURES
Figure A.1: Individual patient plots of BDI scores at each evaluation time.
Figure A.2: Histograms of untransformed BDI scores at each evaluation time.
Figure A.3: Histograms of square-root transformed BDI scores at each evaluation time.
Figure A.4: Plot of the residuals versus predicted values for each time point using the values from the model fit to one of the multiply imputed data sets.
Figure A.5: Normal scores plot of the residuals for each time point using the residuals from the model fit to one of the multiply imputed data sets.
Figure A.6: Normal scores plots of the random effects for one of the multiply imputed data sets.

Figure A.7: Plot of the observed data versus the predicted values for one of the multiply imputed data sets.
APPENDIX B

SIMULATION PROGRAM

#These two statements initialize the matrices that will contain the 
#output from the simulation studies: the coefficients and variances,
#and the means and variances.

mar.coef.mat.mi5 <- matrix(0,nrow=1000,ncol=12)
mar.mean.mat.mi5 <- matrix(0,nrow=1000,ncol=14)

#Sets the seed for the random number generator
set.seed(154475)

#Omits all missing observations from the main data set
bdi.nona <- na.omit(BDIsplus)

#Sets contrasts to avoid the default Helmert contrasts
options(contrasts=c("contr.treatment","contr.poly"))

#Opens up NLME library
library(nlme)

#Increases default object size to handle the coxph object
options("object.size"=6000000)

#Sets max number of subjects

113
nsubj <- 195
#Number of BDI scores after baseline
nscore <- 6
score.subj <- nsubj*nscore
#Number of BDI scores to be imputed.
p <- 6
#Sets 'm'
impute.ssize <- 5
#Creates time vectors that will be used to assign times of drop-out
months <- matrix(c(0,4.5,7.5,10.5,15,21,24),nrow=1,ncol=7)
mon.new <- matrix(c(4.5,7.5,10.5,15,21,24),nrow=1,ncol=6)

#The next group of statements (y.time, x.time, reg.time, and
#prob.time) are the statements for the linear regression that is used
#to drop patients. y.time is the value of the alpha percentile of the
#N(0,1) distribution; x.time is the sum of BDI scores from baseline to
#time t-1; reg.time is the regression model; and prob.time creates the
#vector of predicted values.
y.time3 <- rep(qnorm(0.05,mean=0,sd=1),195)
x.time3 <- bdi.nona$d1+bdi.nona$d2
reg.time3 <- lm(y.time3 ~ x.time3)
prob.time3 <- reg.time3$coeff[1] - 
  (reg.time3$coeff[2]*(bdi.nona$d1+bdi.nona$d2))

114
y.time4 <- rep(qnorm(0.05, mean=0, sd=1), 195)
x.time4 <- bdi.nona$d1+bdi.nona$d2+bdi.nona$d3
reg.time4 <- lm(y.time4 ~ x.time4)
prob.time4 <- reg.time4$coeff[1] -
            (reg.time4$coeff[2]*(bdi.nona$d1+bdi.nona$d2+bdi.nona$d3))

y.time5 <- rep(qnorm(0.1, mean=0, sd=1), 195)
x.time5 <- bdi.nona$d1+bdi.nona$d2+bdi.nona$d3+bdi.nona$d4
reg.time5 <- lm(y.time5 ~ x.time5)
prob.time5 <- reg.time5$coeff[1] -
            (reg.time5$coeff[2]*(bdi.nona$d1+bdi.nona$d2+bdi.nona$d3
            +bdi.nona$d4))

y.time6 <- rep(qnorm(0.15, mean=0, sd=1), 195)
x.time6 <- bdi.nona$d1+bdi.nona$d2+bdi.nona$d3+bdi.nona$d4
            +bdi.nona$d5
reg.time6 <- lm(y.time6 ~ x.time6)
prob.time6 <- reg.time6$coeff[1] -
            (reg.time6$coeff[2]*(bdi.nona$d1+bdi.nona$d2+bdi.nona$d3
            +bdi.nona$d4+bdi.nona$d5))

y.time7 <- rep(qnorm(0.1, mean=0, sd=1), 195)
x.time7 <- bdi.nona$d1+bdi.nona$d2+bdi.nona$d3+bdi.nona$d4
+ bdi.nona$d5+bdi.nona$d6
reg.time7 <- lm(y.time7 ~ x.time7)
prob.time7 <- reg.time7$coeff[1] -
(reg.time7$coeff[2]*(bdi.nona$d1+bdi.nona$d2+bdi.nona$d3
+ bdi.nona$d4+bdi.nona$d5+bdi.nona$d6))

# Creates a matrix of predicted values
prob <- matrix(c(rep(0,195),cbind(prob.time3,prob.time4,prob.time5,
prob.time6,prob.time7)),ncol=nscore,nrow=nsubj)

# Matrix of times for LME model
lme.mat <- cbind(rep(bdi.nona$subj,7),c(rep(-10.3,195),rep(-7.3,195),
rep(-4.3,195),rep(-1.3,195),rep(1.7,195),rep(7.7,195),
rep(13.7,195)),c(rep((-10.3)^2,195),rep((-7.3)^2,195),
rep((-4.3)^2,195),rep((-1.3)^2,195),rep(1.7^2,195),rep(7.7^2,195),
rep(13.7^2,195)))

# Begins simulation
for (s in 1:1000)
{
# Creates a matrix of random N(0,1) variates; one for each patient at
# months 3, 6, 9, 12, 18, and 24
random.num <- matrix(rnorm(n=1170,mean=0,sd=1),ncol=nscore,nrow=nsubj)
#Creates matrix of BDI scores at months 3, 6, 9, 12, 18, and 24
dep.scores <- matrix(c(bdi.nona$d2, bdi.nona$d3, bdi.nona$d4,
                     bdi.nona$d5, bdi.nona$d6, bdi.nona$d7), ncol=nscore, nrow=nsubj)

#Initializes indicator matrix for intermittently missing scores
miss.list <- matrix(0, ncol=nscore, nrow=nsnbj)

#Initializes indicator matrix for drop-out
drop.list <- matrix(0, ncol=nscore, nrow=nsubj)

#Initializes time vector, for time to drop-out (or time to end of study)
time <- matrix(0, nrow=nsubj, ncol=1)

#Initializes status vector
status <- matrix(0, nrow=nsubj, ncol=1)

#Starts loop i for creating drop-outs and intermittently missing data
for (i in 2:6){
  #Adds row entries in drop.list; if zero, then subject has not yet
dropped
  cumsum.test <- t(apply(drop.list[, 1:i], 1, cumsum))
  #Assigns drop.list element 1 if the drawn \( N(0,1) \) < predicted \( y \) and
  #if subject has not yet been dropped
  drop.list[, i][random.num[, i] < prob[, i] & cumsum.test[, (i-1)] == 0] <- 1
  #Next two commands set the BDI at drop-out, and all scores after
  #drop-out, to missing
  dep.scores[, i][drop.list[, i] == 1] <- NA
### Code Snippet

```r
dep.scores[,i][dep.scores[,i-1] == 'NA'] <- NA

# Creates intermittently missing data
miss.list[,i-1][random.num[,i-1] < -1.036433 & dep.scores[,2] != 'NA' & dep.scores[,i] != 'NA'] <- 1

# Sets BDI missing if the intermittently missing indicator=1
dep.scores[,i-1][miss.list[,i-1] == 1] <- NA

# Sets drop-out time
   time[,][drop.list[,i] ==1] <- months[,i]
}

# Closes i loop

# Replaces the time value to 24 if patient is not dropped in simulation
   time[time[,] == 0] <- 24

# Assigns status indicator if subject drops out
   status[,][time[,] < 24] <- 1

# Creates data frame for LME model fit to unbalanced data
BDI.unbal.lme.df <- data.frame(cbind(lme.mat), c(bdi.nona$d1, dep.scores[,1], dep.scores[,2], dep.scores[,3], dep.scores[,4], dep.scores[,5], dep.scores[,6]))
names(BDI.unbal.lme.df) <- c("subj", "mon", "mon_sq", "bdi")

# Omits missing observations to create an unbalanced data frame
BDI.unbal.lme.df.nona <- na.omit(BDI.unbal.lme.df)
```
# Creates a grouped data object for the LME program

```r
bdi.unbal.grpd <- groupedData(sqrt(bdi) ~ mon + mon.sq | subj,
   data=BDI.unbal.lme.df.nona)
```

# LME model for unbalanced data

```r
bdi.unbal.lme <- lme(sqrt(bdi) ~ mon + mon.sq, data=bdi.unbal.grpd,
   random=!mon + mon.sq | subj, control = list(apVar = F))
```

# The next set of commands creates a data set for the repeated event
# survival model; stat.mat initializes the status matrix; enum.mat
# initializes the enumeration matrix; stop.mat and start.mat initialize
# the start and stop times matrices

```r
stat.mat <- matrix(0, ncol=nscore, nrow=nsubj)
enum.mat <- matrix(1, ncol=nscore, nrow=nsubj)
stop.mat <- matrix(NA, ncol=nscore, nrow=nsubj)
start.mat <- matrix(0, ncol=nscore, nrow=nsubj)
```

# Starts i loop for assigning drop indicators for repeated event
# survival model

```r
for (i in 1:5) {
  # If the ith values is intermittently missing or dropped or if the
  # drop time is less than or equal to the ith value in mon.new, then
  # stop.mat is assigned the ith value in mon.new
  stop.mat[,i][miss.list[,i] == 1 | drop.list[,i] == 1 | time[,] <=
```
mon.new[,i] <- mon.new[,i]

# Status is assigned 1 if ith element is an event
stat.mat[,i][miss.list[,i] == 1 | drop.list[,i] == 1 | time[,] <=
mon.new[,i]] <- 1
}

# Closes i loop

# Assigns the last column in stop.mat the final observation time
stop.mat[,6] <- 24

# Assigns the last column in stat.mat the status vector
stat.mat[,6] <- status

# Loop for assigning numbers to enumeration matrix.
for(i in 2:6){ enum.mat<- t(apply(stat.mat[,1:i],1,cumsum)) }
enum.mat[,1:6][enum.mat[,1:6]==0] <- 1

# Loop to assign start times for stop times not equal to zero at
# previous times
for (i in 1:6){
    start.mat[,i][stop.mat[,i-1] == 4.5] <- 4.5
    start.mat[,i][stop.mat[,i-1] == 7.5] <- 7.5
    start.mat[,i][stop.mat[,i-1] == 10.5] <- 10.5
    start.mat[,i][stop.mat[,i-1] == 15] <- 15
    start.mat[,i][stop.mat[,i-1] == 21] <- 21
}
#Loop to assign start times when start times are not equal to zero
#at previous times
for (i in 1:6) {
  start.mat[,i][start.mat[,i] == 0 & start.mat[,i-1] == 4.5] <- 4.5
  start.mat[,i][start.mat[,i] == 0 & start.mat[,i-1] == 7.5] <- 7.5
  start.mat[,i][start.mat[,i] == 0 & start.mat[,i-1] == 10.5] <- 10.5
  start.mat[,i][start.mat[,i] == 0 & start.mat[,i-1] == 15] <- 15
  start.mat[,i][start.mat[,i] == 0 & start.mat[,i-1] == 21] <- 21
}

#Creates a new data frame for BDI with one row for each subject
BDI.df.new <- data.frame(bdi.nona$subj,bdi.nona$age,bdi.nona$d1, dep.scores,miss.list)
names(BDI.df.new) <- c("subj","age","d1","d2","d3","d4","d5","d6",
  "d7","miss3","miss6","miss9","miss12","miss18","miss24")

#Creates a data frame for the repeated event survival model; one line
#per event
BDI.mult <- data.frame(cbind(rep(bdi.nona$subj,6),rep(bdi.nona$age,6),
  rep(bdi.nona$smoke,6),rep(bdi.nona$chd.hx,6),rep(bdi.nona$female,6),
  rep(bdi.nona$d1,6),rep(dep.scores[,1],6),rep(dep.scores[,2],6),
  rep(dep.scores[,3],6),rep(dep.scores[,4],6),rep(dep.scores[,5],6),
  rep(dep.scores[,6],6),c(start.mat[,1:6]),c(stop.mat),c(stat.mat),
  rep(miss.list[,1],6),rep(miss.list[,2],6),rep(miss.list[,3],6),
  rep(miss.list[,4],6),rep(miss.list[,5],6),rep(miss.list[,6],6),
  121
c(enum.mat))

names(BDI.mult) <- c("subj","age","smoke","chd.hx","female","d1","d2",
"d3","d4","d5","d6","d7","time1","time2","status","miss3","miss6",
"miss9","miss12","miss18","miss24","enum")

# Omits observations with missing time 2 values
BDI.mult.naomit <- BDI.mult[BDI.mult$time2 != "NA",]

# Cox model using repeated events and cluster statement
BDI.coxph <- coxph(Surv(time1,time2,status) ~ age + female + smoke +
chd.hx + d1 + cluster(subj), data=BDI.mult.naomit)

# Cox model with repeated events and no cluster statement
BDI.coxph1 <- coxph(Surv(time1,time2,status) ~ age + female + smoke +
chd.hx + d1, data=BDI.mult.naomit)

#'Trick' to get the right variance when computing survival probabilities
BDI.coxph1$var <- BDI.coxph$var

# Estimates survival probabilities for each 'event' (some subjects
# appear more than once if they have more than one event)
BDI.coxph$survfit <- survfit(BDI.coxph1, newdata=BDI.mult.naomit)

# Puts survival probabilities and ID into one matrix
surv.rbind <- matrix(rbind(BDI.coxph$survfit$surv,
BDI.mult.naomit$subj), nrow=7)

# Deletes duplicated observations; creating a matrix with 195 columns
surv.matrix.unq <- surv.rbind[,(! duplicated(surv.rbind[7,])]

122
# Sorts elements in terms of ID

```r
sorted.id <- sort(surv.matrix.unq[7,])
```

# Orders elements in order of ID

```r
order.id <- order(surv.matrix.unq[7,])
```

# Orders all probabilities in order of ID

```r
order.r1 <- surv.matrix.unq[1,][order.id]
order.r2 <- surv.matrix.unq[2,][order.id]
order.r3 <- surv.matrix.unq[3,][order.id]
order.r4 <- surv.matrix.unq[4,][order.id]
order.r5 <- surv.matrix.unq[5,][order.id]
order.r6 <- surv.matrix.unq[6,][order.id]
```

# Puts ordered elements into one matrix

```r
surv.matrix <- matrix(rbind(order.r1,order.r2,order.r3,order.r4,
                          order.r5,order.r6,sorted.id),nrow=7,ncol=195)
```

# Deletes ID in surv.matrix

```r
surv.matrix <- surv.matrix[1:6,]
```

# Attaches BDI.df.new for MI methods

```r
attach(BDI.df.new)
```

# Creates matrix with the BDI at time 2 repeated impute.ssize times

```r
d2.impute <- matrix(rep(d2,impute.ssize),ncol=impute.ssize,byrow=F)
```

# Breaks up survival probabilities into quintiles

```r
quintiles <- apply(surv.matrix,1,quantile,c(0,1,2,3,4,5)/5)
```
#Subtracts a small number from the first quintile

quintiles[1,] <- quintiles[1,] - 1.0E-10

#Starts j loop for the ABB method (for j in each quintile)

for(j in seq(1,5)){

#Separates patients with missing scores and those with observed scores in each quintile

idx.miss <- (quintiles[j,1] < surv.matrix[1,] & surv.matrix[1,] <= quintiles[(j+1),1] & is.na(d2))

idx.not.miss <- (quintiles[j,1] < surv.matrix[1,] & surv.matrix[1,] <= quintiles[(j+1),1] & (is.na(d2) == F))

#Selects the 3-month BDI scores from those not missing in each quintile

impute.distr.2 <- d2[idx.not.miss]

#Starts i loop for the ABB method (for i in each value of m)

for(i in seq(1:impute ssize)){

#Samples from the observed values with replacement

sampling.distr <- sample(impute.distr.2,length(impute.distr.2), replace=T)

#Draws one value for each missing element from the sample of observed scores

d2.impute[idx.miss,i] <- sample(sampling.distr,sum(idx.miss), replace=T)

} #Closes i loop

124
The above procedure is repeated for scores at months 6, 9, 12, 18, and 24; all are done in one loop.

d3.impute <- matrix(rep(d3, imputessize), ncol=imputessize, byrow=F)
d4.impute <- matrix(rep(d4, imputessize), ncol=imputessize, byrow=F)
d5.impute <- matrix(rep(d5, imputessize), ncol=imputessize, byrow=F)
d6.impute <- matrix(rep(d6, imputessize), ncol=imputessize, byrow=F)
d.impute <- list(d2.i=d2.impute, d3.i=d3.impute, d4.i=d4.impute,
d5.i=d5.impute, d6.i=d6.impute)
imiss.names <- c("miss6", "miss9", "miss12", "miss18")
d.names <- c("d3", "d4", "d5", "d6")

for(kk in seq(2,5)){
  for(j in seq(1, 5)){
    idx.miss <- (quintiles[j, kk] < surv.matrix[kk,] &
                 surv.matrix[kk,] <= quintiles[(j+1), kk] &
                 is.na(get(d.names[kk-1])) & get(imiss.names[kk-1]) == 1)
    idx.impute.distr <- (quintiles[j, kk] < surv.matrix[kk,] &
                          surv.matrix[kk,] <= quintiles[(j+1), kk] &
                          (is.na(get(d.names[kk-1])) == F))
    impute.distr <- get(d.names[kk-1])[idx.impute.distr]
    for(i in seq(1:impute ssize)){
      sampling.distr <- sample(impute.distr, length(impute.distr),
replace=T)
d.impute[[kk]][idx.miss,i] <-
sample(sampling.distr,sum(idx.miss),
replace=T)

}
}
#Creates a matrix to store the m means and m variances to be combined
#using the multiple imputation rules
mean.var.mat <- matrix(0,nrow=impute.ssize,ncol=14)
#Creates a matrix to store the m coefficients and m variances of the
#estimates for the intercept, time and time^2; they will be combined
#using the multiple imputation rules
coeff.mat <- matrix(0,nrow=impute.ssize,ncol=6)

#Starts jj loop to impute the scores after drop-out (jj for each value
#in impute.ssize)
for (jj in 1:impute.ssize)
{
  #Creates a data frame containing the intermittently missing imputed
  #values or observed values and covariates; has a monotone missing
  #pattern
  BDI.df.imp <- data.frame(bdi.nona$subj,rep(1,nsubj),bdi.nona$age,
                           bdi.nona$d1, d.impute[[1]][,jj],d.impute[[2]][,1],
                           ...
d.impute[[3]][,jj], d.impute[[4]][,jj], d.impute[[5]][,jj],
dep.scores[,6], bdi.nona$female, bdi.nona$white, bdi.nona$no.hs,
bdi.nona$hs.ed, bdi.nona$smoke, bdi.nona$worknow, bdi.nona$retired,
bdi.nona$evhosp, bdi.nona$chd.hx, bdi.nona$diabcrf, bdi.nona$htncrf)

names(BDI.df.imp) <- c("subj", "int", "age", "d1", "d2", "d3", "d4", "d5",
"d6", "d7", "female", "white", "no.hs", "hs.ed", "smoke", "worknow",
"retired", "evhosp", "chd.hx", "diabcrf", "htncrf")

# Creates a matrix containing the monotone missing BDI scores
BDI.mat <- data.matrix(BDI.df.imp[,5:10])

# Creates a list containing a design matrix for each time that BDI
# scores will be imputed; contains scores prior to each time and
# covariates; includes only those patients with complete data

design.list <- list(x1 <- data.matrix(BDI.df.imp[d2 != 'NA', c(2:4, 11:21)]),
x2 <- data.matrix(BDI.df.imp[d3 != 'NA', c(2:5, 11:21)]),
x3 <- data.matrix(BDI.df.imp[d4 != 'NA', c(2:6, 11:21)]),
x4 <- data.matrix(BDI.df.imp[d5 != 'NA', c(2:7, 11:21)]),
x5 <- data.matrix(BDI.df.imp[d6 != 'NA', c(2:8, 11:21)]),
x6 <- data.matrix(BDI.df.imp[d7 != 'NA', c(2:9, 11:21)])

# Creates a list containing the vectors of y values at each time that
# BDI scores will be imputed; includes only those patients with complete
# data
y.list <- list(y.d2<-data.matrix(BDI.df.imp[d2 != 'NA',5]),
y.d3<-data.matrix(BDI.df.imp[d3 != 'NA',6]),
y.d4<-data.matrix(BDI.df.imp[d4 != 'NA',7]),
y.d5<-data.matrix(BDI.df.imp[d5 != 'NA',8]),
y.d6<-data.matrix(BDI.df.imp[d6 != 'NA',9]),
y.d7<-data.matrix(BDI.df.imp[d7 != 'NA',10])

#Creates a list containing the vectors of parameter estimates at each
time that BDI scores will be imputed; includes estimates for the
#intercept, previous scores and covariates
beta.list <- list(fbeta1<-solve(t(x1)%*%x1) %*% y.d2,
fbeta2<-solve(t(x2)%*%x2) %*% y.d3,
fbeta3<-solve(t(x3)%*%x3) %*% y.d4,
fbeta4<-solve(t(x4)%*%x4) %*% y.d5,
fbeta5<-solve(t(x5)%*%x5) %*% y.d6,
fbeta6<-solve(t(x6)%*%x6) %*% y.d7)

#Creates a list containing the (X~TX)^{-1} matrices at each time
sg.list <- list(sg1<-solve(t(x1)%*%x1),sg2<-solve(t(x2)%*%x2),
sg3<-solve(t(x3)%*%x3),sg4<-solve(t(x4)%*%x4),sg5<-solve(t(x5)%*%x5),
sg6<-solve(t(x6)%*%x6))

#Initializes a vector that will contain the residual sum of squares
sumsq <- matrix(0,nrow=1,ncol=6)
#Initializes a list that will contain the randomly drawn parameter estimates
rbeta.list <- list(rbi<- matrix(0,nrow=nrow(fbeta1),ncol=1),
                  rb2<- matrix(0,nrow=nrow(fbeta2),ncol=1),
                  rb3<- matrix(0,nrow=nrow(fbeta3),ncol=1),
                  rb4<- matrix(0,nrow=nrow(fbeta4),ncol=1),
                  rb5<- matrix(0,nrow=nrow(fbeta5),ncol=1),
                  rb6<- matrix(0,nrow=nrow(fbeta6),ncol=1))

#Creates vectors that store values of 'n' and 'p' at each time; where '
#'n' is the number of observations and 'p' the number of covariates;
#these are used to create the appropriate degrees of freedom for the
#randomly drawn chi-square
n.mat <- matrix(c(nrow(x1),nrow(x2),nrow(x3),nrow(x4),nrow(x5),
                 nrow(x6)),nrow=6,ncol=1)
p.mat <- matrix(c(nrow(fbeta1),nrow(fbeta2),nrow(fbeta3),nrow(fbeta4),
                 nrow(fbeta5),nrow(fbeta6)),nrow=6,ncol=1)

#Initializes a vector that will store the appropriate df for the drawn
#chi-square
df <- matrix(0,nrow=p,ncol=1)

#Initializes a vector that will store the drawn residual sum of
#squares
rsumsq <- matrix(0,nrow=p,ncol=1)

#Starts loop to draw parameters and missing BDI scores (i in each
#time that subjects can drop out)
for (i in 2:6)
{

  #Calculates df for drawn chi-square
  df[i,] <- n.mat[i,] - p.mat[i,]

  #Calculates the residuals
  tmp <- y.list[[i]] - (design.list[[i]] %*% beta.list[[i]])

  #Calculates the residual sum of squares
  sumsq[,i] <- t(tmp) %*% tmp

  #Draws a residual sum of squares
  rsumsq[i,] <- sumsq[,i]/rchisq(1, df[i,])

  #Draws parameters
  rbeta.list[[i]] <- t(chol(rsumsq[i,]*(0.5*(sg.list[[i]] +
                                 t(sg.list[[i]]))))) %*% matrix(rnorm(n=p.mat[i,],mean=0,sd=1),
                                nrow=p.mat[i,],ncol=1) + beta.list[[i]]

  #Imputes values for missing observations using drawn parameters
  BDI.df.imp[BDI.mat[,i]=='NA',(i+4)] <-
    round(data.matrix(BDI.df.imp[BDI.mat[,i]=='NA',
                     c(2:(i+3),11:21)]) %*%rbeta.list[[i]] + (sqrt(rsumsq[i,]) *
                               rnorm(1,mean=0,sd=1)),digits=1)
}

#Assigns BDI to zero if imputed score is negative

BDI.df.imp[,5:10][BDI.df.imp[,5:10]<0] <- 0
# Creates a data frame containing 'completed data' from multiple
# imputation and times for the LME model

BDI.bal.lme.df <- data.frame(cbind(lme.mat), c(bdi.nona$d1,
                                BDI.df.imp$d2, BDI.df.imp$d3, BDI.df.imp$d4, BDI.df.imp$d5,
                                BDI.df.imp$d6, BDI.df.imp$d7))

names(BDI.bal.lme.df) <- c("subj", "mon", "mon.sq", "bdi")

# Creates a grouped data object for the LME

bdi.bal.grpd <- groupedData(sqrt(bdi) ~ mon + mon.sq | subj,
                             data = BDI.bal.lme.df)

# Fits the LME

bdi.bal.lme <- lme(sqrt(bdi) ~ mon + mon.sq, data = bdi.bal.grpd,
                   random = ~mon + mon.sq | subj, control = list(apVar = F))

# Stores coefficients and variances for the intercept, time and time^2
# from the LME in coeff.mat (created earlier)

coeff.mat[, 1:3] <- bdi.bal.lme$coef[, 1]

coeff.mat[, 4:6] <- diag(bdi.bal.lme$varFix)

# Stores means and variances from multiple imputation in mean.var.mat
# (created earlier)

mean.var.mat[, 1:7] <- apply(BDI.df.imp[, 4:10], 2, mean)

mean.var.mat[, 8:14] <- apply(BDI.df.imp[, 4:10], 2, var)

} # Closes jj loop

131
# Combines parameter estimates to obtain point estimate and the total variance using the rules for combining multiple imputation estimates

mar.coef.mat.mi5[s,1:3] <- apply(coeff.mat[1:impute ssize,1:3],2,mean)
mar.coef.mat.mi5[s,4:6] <- (1+1/impute ssize) *
apply(coeff.mat[1:impute ssize,1:3],2,var) +
apply(coeff.mat[1:impute ssize,4:6],2,mean)

# Adds the values to the matrix from the LME fit to the unbalanced data
mar.coef.mat.mi5[s,7:9] <- bdi.unbal.lme$coef[[1]]
mar.coef.mat.mi5[s,10:12] <- diag(bdi.unbal.lme$varFix)

# Combines the means to obtain a point estimate and the total variance using the rules for combining multiple imputation estimates
mar.mean.mat.mi5[s,1:7] <- apply(mean.var.mat[1:impute ssize,1:7],2,mean)
mar.mean.mat.mi5[s,8:14] <- (1+1/impute ssize) *
apply(mean.var.mat[1:impute ssize,1:7],2,var) +
apply(mean.var.mat[1:impute ssize,8:14],2,mean)
detach("BDI.df.new")

# Prints sth iteration in loop
print(s)

# Prints time and date that sth loop ended
stamp()
} # closes s loop
APPENDIX C

LIST OF ABBREVIATIONS

The following abbreviations are used throughout the document.

1. BDI - Beck Depression Inventory
2. CC - complete-case analysis
3. CHD - coronary heart disease
4. CRH - corticotropin releasing hormone
5. DIS - Diagnostic Interview Schedule
6. DSM-IV - Diagnostic and Statistical Manual of Mental Disorders Fourth Edition
7. ECG - electrocardiogram
8. 5-HT - 5-hydroxytryptamine
9. HRV - heart rate variability
10. IFNα - alpha interferon
11. IFNγ - gamma interferon
12. IL-1 - interleukin 1

13. IL-1β - interleukin 1-beta

14. IL-2 - interleukin 2

15. IL-6 - interleukin 6

16. IL-8 - interleukin 8

17. LOCF - last observation carried forward

18. $m$ - the number of completed data sets formed for multiple imputation

19. MAR - missing at random

20. MCAR - missing completely at random

21. MDD - major depressive disorder

22. MI - multiple imputation

23. ML - maximum likelihood

24. MMPI-D - Minnesota Multiphasic Personality Inventory - Depression subscale

25. NI - nonignorable nonresponse

26. TRH - thyrotropin-releasing hormone

27. TSH - thyroid-stimulating hormone
BIBLIOGRAPHY


[87] Ziegelstein, RC, Fauerbach, JA, Stevens, SS, Romanelli, J, Richter, DP, and Bush, DE. Patients with depression are less likely to follow recommendations to reduce cardiac risk during recovery from a myocardial infarction. *Archives of Internal Medicine* 1999;160:1818-1823.
